AN INTRODUCTION TO SET THEORY AND TOPOLOGY

Ronald C. Freiwald

An Introduction to Set Theory and Topology

Ronald C. Freiwald
Washington University in St. Louis

These notes are dedicated to all those who have never dedicated a set of notes to themselves

An Introduction to Set Theory and Topology

By Ronald C. Freiwald
2014
Washington University in St. Louis
Saint. Louis, Missouri

doi: 10.7936/K7D798QH
http://doi.dx.org/10.7936/K7D798QH
ISBN: 978-1-941823-10-1

Cover design by Clayton Petras, MFA 2017, Sam Fox School of Design & Visual Arts

Introduction

These notes are an introduction to set theory and topology. They are the result of teaching a two-semester course sequence on these topics for many years at Washington University in St. Louis. Typically the students were advanced undergraduate mathematics majors, a few beginning graduate students in mathematics, and some graduate students from other areas that included economics and engineering.

Over time my lecture notes evolved into written outlines for students, then written versions of the more involved proofs. The full set of notes was a project completed during the years 2003-2007 with small revisions thereafter.

The usual background for the material is an introductory undergraduate analysis course, mostly because it provides a solid introduction to Euclidean space $\mathbb{R}^n$ and practice with rigorous arguments – in particular, about continuity. Strictly speaking, however, the material is mostly self-contained. Examples are taken now and then from analysis, but they are not logically necessary for the development of the material. The only real prerequisite is the level of mathematical interest, maturity and patience needed to handle abstract ideas and to read and write careful proofs. A few very capable students have taken this course before introductory analysis (even, rarely, outstanding university freshmen) and invariably they have commented later on how material eased their way into analysis.

The material on set theory is not done axiomatically. However, we do try to provide some informal insights into why an axiomatization of the subject might be valuable and what some of the most important results are. A student with a good grasp of the set-theoretic material – scattered throughout the notes, but heavily concentrated in Chapters I and VIII – will know all the informal set theory that most mathematicians ever need and will be in a strong position to continue on to a study of axiomatic set theory.

The topological material is lies within the area traditionally labeled "general topology." No topics from algebraic topology are included. This was a conscious choice that reflects my own training and tastes, as well as a conviction that students are usually rushed too quickly through the basics of topology in order to get to "where the action is." It is certainly true that general topology has not been the scene of much research for several decades, and most of the research that does still continue is closely related to set theory and mathematical logic. Nevertheless, general topology contains a set of tools that most mathematicians need, whether for work in analysis or other parts of topology.

Many of those basic tools (such as "compactness" and the "product topology") seem very abstract when a student first meets them. It takes time to develop an ownership of these tools. This includes a sense of their significance, an appropriate "feel" for how they behave, and good technique – in short, all the things necessary to make using a compactness argument, say, into a completely routine tool. I believe this "absorption" process is often short-circuited in the rush to move students along to algebraic topology. The result then can be an introduction to algebraic topology where many tedious details are (appropriately) omitted and the student is ill-equipped to fill them in – or even to feel confident that the omissions are genuinely routine. When that happens, a student can begin to feel that the subject has a vague, hand-waving quality about it.

These notes are designed to give the student the necessary practice and build up intuition. They begin with the more concrete material (metric spaces) and move outward to the more general ideas. The basic notions about topological spaces are introduced in the middle of the study of metric spaces to illustrate the idea of increasing abstraction and to highlight some important properties of metric spaces against a background where these properties fail. The result is an exposition that is not as efficient as it could be if the more general definitions were stated in the first place. In particular, many of the basic ideas about metric spaces (Chapter II) are revisited in the introductory chapter on topological spaces (Chapter III).

Just as in any mathematics course, solving problems is essential. There are many exercises in the notes, particularly in the early chapters. They vary in difficulty but it is fair to say that a majority of the problems require some thought. Few, if any, could be genuinely called "trivial." For example, in Chapter I (Sets) there are no problems of the sort: prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. It is assumed that students are sufficiently sophisticated not to need that sort of drill.

There are "Chapter Reviews" at the end of each chapter. A review consists of a list of statements, each of which requires either an explanation or a counterexample. Presented with statements whose truth is uncertain, students can develop confidence and intuition, learn to make thoughtful connections and guesses, and build a tool chest of examples and counterexamples. Nearly every review statement requires only an insight, a use of an earlier result in a new situation, or the application of a more abstract result to a concrete situation. For almost every true statement, an appropriate justification consists of at most a few sentences.

These notes were a "labor of love" over many years and are intended as an aid for students, not as a work for publication. Such originality as there is lies in the selection of material and its organization. Many proofs and exercises have been refashioned or polished, but others are more-or-less standard fare drawn from sources some of which are now forgotten. Readers familiar with the material will probably recognize overtones of my predecessors and contemporaries such as Arthur H. Stone, Leonard Gillman, Robert McDowell and Stephen Willard. My thanks to them for all their insights and contributions, and to a few hundred students who have worked with various parts of these notes over the years. Of course, any errors are my own.

The notes are organized into ten chapters (I, II,..., X) and each chapter is divided into sections (1, 2,..., n). Definitions, theorems, and examples are numbered consecutively within each of these sections $-$ for example, Definition 4.1, Theorem 4.2, Theorem 4.3, Example 4.4, ... For example, a reference to Theorem 6.4 refers to the 4^{th} numbered item in Section 6 of the current chapter. A reference to an item outside the current chapter would include the chapter number: for example, Theorem III.6.4 means the 4^{th} item in Section 6 of Chapter III.

Exercises are numbered consecutively within each chapter: E1, E2, ... A reference to an exercise outside the current chapter would include the chapter number $-$ for example, Exercise III.E8.

Ronald C. Freiwald
St. Louis, Missouri
May 2014

Table of Contents

Chapter I Sets

Chapter II Pseudometric Spaces

Chapter III Topological Spaces

Chapter IV Completeness and Compactness

Chapter V Connected Spaces

Chapter VI Products and Quotients

Chapter VII Separation Axioms

Chapter VIII Ordered Sets, Ordinals and Transfinite Methods

Chapter IX Convergence

Chapter X Compactifications

Chapter I
The Basics of Set Theory

1. Introduction

Every mathematician needs a working knowledge of set theory. The purpose of this chapter is to provide some of the basic information. Some additional set theory will be discussed in Chapter VIII.

Sets are a useful vocabulary in many areas of mathematics. They provide a language for stating1 interesting results. For example, in analysis: "a monotone function from $\mathbb{R}$ to $\mathbb{R}$ is continuous except, at most, on a countable set of points." In fact, set theory had its origins in analysis, with work done in the late 19th century by Georg Cantor (1845-1918) on Fourier series. This work played an important role in the development of topology, and all the basics of the subject are written in the language of set theory. However sets are not just a tool; like many other mathematical ideas, "set theory" has grown into a fruitful research area of its own.

In addition, on the philosophical side, most mathematicians accept set theory as a foundation for mathematics – this means that the notions of "set" and "membership in a set" can be taken as the most primitive notions of mathematics, in terms of which all (or nearly all) others can be defined. From this point of view, "every object in mathematics is a set." To put it another way, most mathematicians believe that "mathematics can be embedded in set theory."

So, you ask, what is a set? There are several different ways to try to answer. Intuitively – and this is good enough for most of our work in this course – a set is a collection of objects, called its elements or members. For example, we may speak of "the set of United States citizens" or "the set of all real numbers." The idea seems clear enough. However, this is not really a satisfactory definition of a set: to say "a set is a collection of objects..." seems almost circular. After all, what is a "collection"?

In the early days of the subject, writers tried to give definitions of "set," just as Euclid attempted to give definitions for such things as "point" and "straight line" ("a line which lies evenly with the points on itself"). And, as in Euclid's case, these attempts did not really clarify things very much. For example, according to Cantor

> *Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten in unserer Anschauung oder unseres Denkens (welche die Elemente von M genannt werden) zu einem Ganzen* [By a set we are to understand any collection into a whole M of definite and separate objects (called the elements of M) of our perception or thought.] (*German seems to be a good language for this kind of talk.*)
>
> More compactly, Felix Hausdorff, around 1914, stated that a set is "a plurality thought of as a unit."

So there are several ways we could proceed. One possibility is simply to use our intuitive, informal notion of a set, move on from there and ignore any more subtle issues – just as we would not worry about having definitions for "point" and "line" when we begin to study geometry. Another option might be to try to make a formal definition of "set" in terms of some other mathematical objects (assuming, implicitly, that these objects are "more fundamental" and intuitively understood). As a third approach, we could take the notions of "set" and "set membership" as primitive undefined terms and simply write down a collection of formal axioms that prescribe how "sets" behave.

The first approach is sometimes called naive set theory. ("Naive" refers only to the starting point – naive set theory gets quite complicated.) Historically, this is the way set theory began. The third option would take us into the subject of axiomatic set theory. Although an enormous amount of interesting and useful naive set theory exists, almost all research work in set theory nowadays requires the axiomatic approach (as well as some understanding of mathematical logic).

We are going to take the naive approach. For one thing, the axiomatic approach is not worth doing if it isn't done carefully, and that is a whole course in itself. In addition, axiomatic set theory isn't much fun unless one has learned enough naive set theory to appreciate why some sort of axiomatization is important. It's more interesting to try to make things absolutely precise after we have a good overview. However as we go along, we will add some tangential comments about the axiomatic approach to help keep things in a more modern perspective.

2. Preliminaries and Notation

Informal Definition 2.1 A set is a collection of objects called its elements (or members). If A is a set and x is an element of A, we write $x \in A$. Otherwise, we write $x \notin A$.

As the informal definition suggests, we may also use the word "collection" (or other similar words such as "family") in place of "set." Strictly speaking, these words will be viewed as synonymous. We sometimes interchange these words just for variety, and often we will switch these words for emphasis: for example, we might refer to a set whose elements are also sets as a "collection of sets" or a "family of sets," rather than a "set of sets" – even though these all mean the same thing.

Here are two ways to describe sets:

By listing the elements: this is most useful for a small finite set or an infinite set whose elements can be referred to using an ellipsis "..."

For example: $A = \{1, 2\}$
$\mathbb{N} = \{1, 2, 3, \dots\}$, the set of natural numbers
$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, the set of integers

By abstraction: this means that we specify a property that describes exactly what elements are in the set. We do this by writing something like $\{x : x$ has a certain property $\}$.

For example: $\mathbb{R} = \{x : x$ is a real number$\}$, the set of all real numbers
$\mathbb{Q} = \{\frac{p}{q} : p \in \mathbb{Z},\ q \in \mathbb{Z}$ and $q \neq 0\}$, the set of rational numbers
$\mathbb{P} = \{x : x \in \mathbb{R}$ and $x \notin \mathbb{Q}\}$, the set of irrational numbers

Suppose we write $\{x : x \in \mathbb{R} \text{ and } x^2 = -1\}$ or $\{x : x \in \mathbb{R} \text{ and } x \neq x\}$. No real number is actually a member of either set – both sets are empty. The empty set is usually denoted by the symbol $\emptyset$; it is occasionally also denoted by $\{\ \}$. Sometimes the empty set is also called the "null set," although that term is more often used as a technical term for a certain kind of set in measure theory. (*By the way, $\emptyset$ is a Danish letter, not a Greek phi $= \Phi$ or ϕ*).

It may seem odd to talk about an empty set and even to give it a special symbol; but otherwise we would need to say to say that $\{x : x \in \mathbb{R} \text{ and } x^2 = -1\}$, which looks perfectly well-formed, is not a set at all. Even worse: consider $S = \{x : x \in \mathbb{Q} \text{ and } x = \alpha^\beta$, where α and β are irrational $\}$. Do you know whether or not there are any such rational numbers x? If you're not sure and if we did not allow an empty set, then you would not be able to decide whether or not S is a set! It's much more convenient simply to agree that $\{x : x \in \mathbb{Q} \text{ and } x = \alpha^\beta$, where α and β are irrational$\}$ is a set and allow the possibility that it might be empty.

In our informal approach, a member of a set could be any object. In mathematics, however, it's not likely that we would be interested in a set whose members are aardvarks. We will only use sets that contain various mathematical objects. For example, a set of functions

$$C[a,b] = \{f : f \text{ is a continuous real-valued function with domain } [a,b]\}$$

or a set of sets such as

$$\{\{1\}, \{1,2\}\}, \text{ or } \{\emptyset\}, \text{ or } \{\emptyset, \{\emptyset\}\}.$$

Of course, if "everything object in mathematics is a set," then all sets in mathematics can only have other sets as members (because nothing else is available to be a member).

> Once we start thinking that everything in mathematics is a set, then an interesting thought comes up. If x is a set, then either $x = \emptyset$ or there is an element $x_1 \in x$. Since x_1 is a set, either $x_1 = \emptyset$ or there is a set $x_2 \in x_1$, and so on. Is it possible to find a set for which there is an "infinite descending chain" of members
>
> $$\ldots \in x_n \in x_{n-1} \in \ldots \in x_1 \in x \ ?$$

We say that $A = B$ if for all x, $x \in A \Leftrightarrow x \in B$: that is, two sets A and B are equal if A and B have precisely the same elements. For example, $\{x, y\} = \{y, x\}$ and $\{x, x\} = \{x\}$. Two sets whose descriptions look very different on the surface may turn out, on closer examination, to have exactly the same elements and therefore be equal. For example, $\{x : x \in \mathbb{R} \text{ and } x^2 = -1\}$ $= \{x : x \in \mathbb{R} \text{ and } x \neq x\}$; and a scrupulous reader could verify that

$$\{x : x \in \mathbb{R} \text{ and } x^5 + 5x^4 - 29x^3 - 109x^2 - 8x + 140 = 0\} = \{-7, -2, 1, 5\}.$$

We say that A is a subset of B, and write $A \subseteq B$ provided that each element of A is also a member of B, that is, for all x, $x \in A \Rightarrow x \in B$. If $A \subseteq B$ but $A \neq B$, then we say A is a proper subset of B. Clearly, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$ are both true. (*Note: "if and only if" is often abbreviated by "iff".*)

Note that $A \subseteq B$ and $B \subseteq C$ implies that $A \subseteq C$.

You should look carefully at each of the following true statements to be sure the notation is clear:

$x \in A$ iff $\{x\} \subseteq A$

$\mathbb{N} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ (the set of complex numbers)

$\emptyset \neq \{\emptyset\}$ $\quad$ $\emptyset \subseteq \{\emptyset\}$ $\quad$ $\emptyset \in \{\emptyset\}$

$\emptyset \subseteq \emptyset$ $\quad$ $\emptyset \notin \emptyset$ $\quad$ $\emptyset \subseteq A$ for any set A

$\emptyset \in \{\emptyset\} \in \{\{\emptyset\}\}$, but $\emptyset \notin \{\{\emptyset\}\}$ $\quad$ (so $A \in B \in C$ doesn't imply $A \in C$)

Note that $\emptyset \neq \{\emptyset\}$. The set on the left is empty, while the set on the right has one member, namely the set $\emptyset$. This might be clearer with the alternate notation: $\{\,\} \neq \{\{\,\}\}$. The set on the left is analogous to an empty paper bag, while the set on the right is analogous to a bag with an empty bag inside it.

We define the power set of a set A, written as $\mathcal{P}(A)$, to be the set of all subsets of A. In symbols, $\mathcal{P}(A) = \{B : B \subseteq A\}$.

Since $\emptyset$ is a subset of every set, we have $\emptyset \in \mathcal{P}(A)$ for every set A.

Since $A \subseteq A$, we also have $A \in \mathcal{P}(A)$ for every set A.

$\mathcal{P}(\{1,2\}) = \{\emptyset\,, \{1\}, \{2\}, \{1,2\}\}$

$\mathcal{P}(\{1\}) = \{\emptyset\,, \{1\}\}$

$\mathcal{P}(\emptyset) = \{\emptyset\,\}$

The last three examples suggest that a set A with n elements has 2^n subsets (*Why? To define a subset B of A we must decide, for each $x \in A$, whether or not to put x into the subset B. So for each element x in A, there are two choices – "yes" or "no." So there are 2^n ways to pick the elements to form a subset B.*)

Exercises

E1. Use induction to prove that if a set X has n elements, then $\mathcal{P}(X)$ has 2^n elements.

E2. Use Exercise 1 to explain the meaning of the identity

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n \qquad (n = 0, 1, 2 \ldots)$$

3. Paradoxes

The naive approach to sets seems to work fine until someone really starts trying to cause trouble. The first person to do this was Bertrand Russell who discovered Russell's Paradox in 1901:

> It makes sense to ask whether a set is one of its own members – that is, for a given set A, to ask whether $A \in A$ is true or false. The statement $A \in A$ is false for the sets you immediately think of: for example $\{1,2\} \notin \{1,2\}$. However, is the infinite set
>
> $A = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}, \ldots\}$
>
> a member of itself? How about an even more complicated infinite set? Could it happen that $A \in A$? Whatever the answer, it makes sense to ask.
>
> If we simply follow our naive approach, we can define a set $\mathfrak{A}$ by writing $\mathfrak{A} = \{A : A \notin A\}$, so that $\mathfrak{A}$ is the set of all sets which are not members of themselves. Then, we ask, is $\mathfrak{A} \in \mathfrak{A}$ true or false? If $\mathfrak{A} \in \mathfrak{A}$, then $\mathfrak{A}$ must satisfy the requirement for being a member of $\mathfrak{A}$, that is, $\mathfrak{A} \notin \mathfrak{A}$. On the other hand, if $\mathfrak{A} \notin \mathfrak{A}$, then $\mathfrak{A}$ does meet the membership requirement for $\mathfrak{A}$, so $\mathfrak{A} \in \mathfrak{A}$. Thus, each of the only two possibilities about the set $\mathfrak{A}$ (that $\mathfrak{A} \in \mathfrak{A}$ or $\mathfrak{A} \notin \mathfrak{A}$) leads to a contradiction!

Russell's Paradox illustrates that dilemmas can arise if we use the method of abstraction too casually to a define set. One way out is to refuse to call $\mathfrak{A}$ a set. To do that, in practice, we will insist that whenever we define a set by abstraction, we only form subsets of sets that already exist. That is, in defining a set by abstraction, we should always write $\{x : \underline{x \in X} \text{ and } \ldots\}$ or, for short, $\{x \in X : \ldots\}$ where X is some set that we already have. The result is a subset of X. Since the preceding definition of $\mathfrak{A}$ doesn't follow this form, $\mathfrak{A}$ is not guaranteed to be a set.

This is the route taken in axiomatic set theory, and we see that it eliminates Russell's paradox. If X is a set and we consider the set $\mathfrak{A} = \{A \in X : A \notin A\}$, the dilemma vanishes : $\mathfrak{A} \notin \mathfrak{A}$ means either that $\mathfrak{A} \in \mathfrak{A}$ (a contradiction) or that $\mathfrak{A} \notin X$ — an alternative conclusion that we can live with.

Russell's Paradox has the same flavor as many "self-referential" paradoxes in logic. For example, some books mention themselves – in the preface the author might say, "In this book, I will discuss ... ". Other books make no mention of themselves. Suppose Library X wants to make a book listing all books that do not mention themselves. Should the new book list itself? That is Russell's Paradox: if it doesn't mention itself, then it should; and if it does mention itself, then it shouldn't. The resolution of the paradox is that the new book really is intended to list "all books in $\underline{X}$ which do not mention themselves"— that is, in forming the new book, one is restricted to considering only those books already in the collection X. With this additional qualification, the paradox disappears. Think about the same paradox in the dedication at the front of these notes: are the notes dedicated to Freiwald?

In everyday mathematics, we usually don't have to worry about this kind of paradox. Almost always, when we form a new set, we have (at least in the back of our minds) a larger set X of which it is a subset. Therefore, indulging in a bit of sloppiness, we may sometimes write $\{x : \ldots\}$ rather than the more correct $\{x \in X : \ldots\}$ simply because the set X could be supplied on demand, and the notation is simpler.

There is another kind of difficulty we can get into when defining sets by abstraction. It arises from the nature of the description " ... " when we write $\{x \in X : \text{ ... }\}$. This is illustrated by Richard's Paradox:

> Consider $A = \{x \in \mathbb{N} : \ x \text{ is definable in English using less than 10000 characters}\}$. There are only finitely many English character strings with length < 10000 (a very large number, but finite). Most of these character strings are gibberish, but some of them define a positive integer. So the set A is finite. Therefore there must be natural numbers not in A, and we can pick the smallest such natural number : call it m.
>
> But if A is a well-defined set, then the preceding paragraph has precisely defined m, using fewer than 10000 characters (count them!), so m is in A !

To resolve this paradox carefully involves developing a little more formal machinery than we want to bother with here, but the idea is easy enough. The idea involves requiring that only a certain precise kind of description can be used for the property " ... " when defining a set $\{x \in X : \text{ ... }\}$.

Roughly, these are the descriptions we can form using existing sets, $\in$, the logical quantifiers $\forall$ and $\exists$, and the familiar logical connectives $=$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ and $\neg$. When all this is made precise (using first order predicate calculus), the "set" A above is not allowed as a set – because the description " ... " we used for A is not a "legal" description. Fortunately, the sets we want to write down in mathematics can be described by "legal" expressions. For example, we could define the set of positive real numbers by

$$\mathbb{R}^+ = \{x \in \mathbb{R} : \neg(x = 0) \wedge \exists y\,(y \in \mathbb{R} \wedge x = y^2)\}$$

To summarize: there are dangers in a completely naive, casual formation of "sets." One of the reasons for doing axiomatic set theory is to avoid these dangers by giving a list of axioms that lay out precise initial assumptions about sets and how they are formed. But fortunately, with some common sense and a little feeling acquired in practice, such dangerous situations rarely arise in the everyday practice of mathematics.

4. Elementary Operations on Sets

We want to have operations that we can use to combine old sets into new ones. The simplest operations are union and intersection.

Informally, the union of two sets is the set consisting of all elements in A or in B. (*Note: when a mathematician says "either p or q", this means "either p or q or both." This is called the "inclusive" use of the word "or."*)

The intersection of two sets is the set of all elements belonging to both A and B. In symbols,

the union of A and B is $\quad A \cup B = \{x : \ x \in A \text{ or } x \in B\}$, and

the intersection of A and B is $\quad A \cap B = \{x : x \in A \text{ and } x \in B\}$.

Note: You might expect, after the discussion of paradoxes, that the definition of union should read: $A \cup B = \{x \in X : x \in A$ or $x \in B\}$ and then ask "given A and B, what is X?"

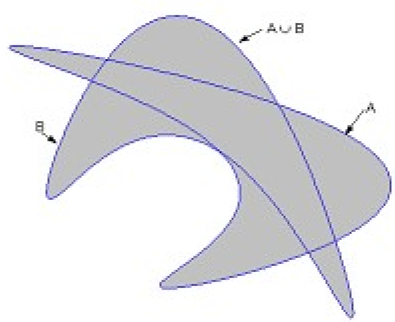
In practice, the sets A and B that we combine are always subsets of some larger set X. Then there is no need to worry because $A \cup B$, we understand, could be written more properly as $\{x \in X : x \in A$ or $x \in B\}$. But to cover all possible situations, axiomatic set theory adds a separate axiom (see A5 on p. 12) to guarantee that unions always exist.

This issue doesn't come up for intersections. For example, given sets A and B, we can always write $A \cap B = \{x \in A : x \in B\}$.

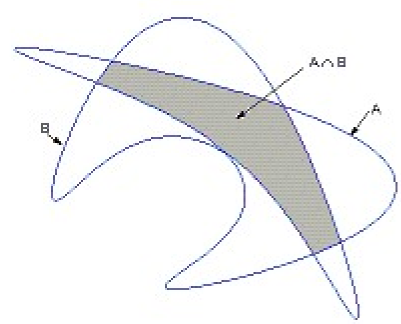
Examples 4.1 $\{1,2\} \cup \{2,3\} = \{1,2,3\}$ $\{1,2\} \cap \{2,3\} = \{2\}$

$\mathbb{P} \cup \mathbb{Q} = \mathbb{R}$ $\mathbb{P} \cap \mathbb{Q} = \emptyset$

The union and intersection of two sets can be pictured with "Venn diagrams" :



$A \cup B$



$A \cap B$

We want to be able to combine more than just two sets, perhaps even infinitely many. To express this, we use the idea of an <u>index set</u>.

Definition 4.2 Suppose that for each λ in some set Λ, a set A_λ is given. We then say that the collection $\mathfrak{A} = \{A_\lambda : \lambda \in \Lambda\}$ is <u>indexed</u> by Λ. We might write $\mathfrak{A}$ more informally as $\{A_\lambda\}_{\lambda \in \Lambda}$ or even merely as $\{A_\lambda\}$ if the index set Λ is clearly understood.

Examples 4.3 1) $\mathfrak{A} = \{A_1, A_2\}$ is indexed by the set $\Lambda = \{1, 2\}$

2) Let I_x be the interval $[0, x]$ of real numbers. Then $\mathfrak{A} = \{I_x : x \in \mathbb{R}, x \geq 0\}$ is indexed by $\Lambda =$ the set of nonnegative real numbers.

3) $\mathfrak{A} = \{A_1, A_2, ..., A_n, ...\}$ is indexed by $\Lambda = \mathbb{N}$

4) $\mathfrak{A} = \emptyset$ iff $\mathfrak{A}$ can be indexed by $\Lambda = \emptyset$

Definition 4.4 Suppose $\mathfrak{A} = \{A_\lambda : \lambda \in \Lambda\}$. The union of the family $\mathfrak{A}$ is the set $\{x : \text{for some } \lambda_0 \in \Lambda,\, x \in A_{\lambda_0}\}$, also written $\bigcup\{A_\lambda : \lambda \in \Lambda\}$ or, more simply, just as $\bigcup \mathfrak{A}$. (So $x \in \bigcup \mathfrak{A}$ iff x is an element of an element of $\mathfrak{A}$.)

The intersection of the family $\mathfrak{A}$ is the set $\{x : x \in A_\lambda \text{ for all } \lambda \in \Lambda\}$, also written as $\bigcap\{A_\lambda : \lambda \in \Lambda\}$ or, more simply, just as $\bigcap \mathfrak{A}$.

When the specific set Λ is understood or irrelevant, we may ignore the "$\lambda \in \Lambda$" and just write $\bigcup A_\lambda$ or $\bigcap A_\lambda$. When $\Lambda = \mathbb{N}$, we might also write $\bigcup \mathfrak{A} = \bigcup_{n=1}^{\infty} A_n$ and $\bigcap \mathfrak{A} = \bigcap_{n=1}^{\infty} A_n$.

Examples 4.5 1) Suppose $\mathfrak{A} = \{A_1, A_2\}$. We can write the union of this family of sets in several different ways: $\bigcup \mathfrak{A} = \bigcup_{i=1}^{2} A_i = A_1 \cup A_2 = \bigcup\{A_i : i = 1, 2\}$.

2) If $I_x = [0, x) \subseteq \mathbb{R}$, then $\bigcup\{I_x : x \geq 0\} = [0, \infty)$ and $\bigcap\{I_x : x \geq 0\} = \emptyset$, and $\bigcap\{I_x : x > 0\} = \{0\}$.

3) Let $\mathfrak{A} = \{A_n : n \in \mathbb{N}\}$, where $A_n = \left[-\frac{1}{n}, 1 + \frac{1}{n}\right] \subseteq \mathbb{R}$.
Then $\bigcup \mathfrak{A} = \bigcup\{A_n : n \in \mathbb{N}\} = \bigcup_{n=1}^{\infty} A_n = [-1, 2]$ and
$\bigcap \mathfrak{A} = \bigcap\{A_n : n \in \mathbb{N}\} = \bigcap_{n=1}^{\infty} A_n = [0, 1]$

4) If $B_n = [n, \infty) \subseteq \mathbb{R}$, then $\bigcup\limits_{n=1}^{\infty} B_n = [1, \infty)$ and $\bigcap\limits_{n=1}^{\infty} B_n = \emptyset$

5) Suppose each set $A_\lambda \subseteq X$. If $\Lambda = \emptyset$, then $\mathfrak{A} = \{A_\lambda : \lambda \in \Lambda\}$ is an empty family. Then $\bigcup \mathfrak{A} = \emptyset$ and $\bigcap\{A_\lambda : \lambda \in \Lambda\} = X$. *For the intersection: if $x \in X$, then x is in every A_λ (can you name an A_λ that doesn't contain x ?) so $x \in \bigcap\{A_\lambda : \lambda \in \Lambda\}$.*
If this argument bothers you, then you can consider the statement $\bigcap\{A_\lambda : \lambda \in \Lambda\} = X$ to be a just a convention about the intersection of an empty family—motivated by the idea that "the fewer sets in the family, the larger the intersection should be, so that when $\Lambda = \emptyset$, the intersection should be as large as possible."

Theorem 4.6 1) $A \cup B = B \cup A$, and $A \cap B = B \cap A$.
2) $A \cup (B \cup C) = (A \cup B) \cup C$, and $A \cap (B \cap C) = (A \cap B) \cap C$. More generally,

$$\bigcup_{\lambda \in \Lambda} A_\lambda \cup \Big(\bigcup_{\mu \in M} A_\mu \cup \bigcup_{\nu \in N} A_\nu\Big) = \Big(\bigcup_{\lambda \in \Lambda} A_\lambda \cup \bigcup_{\mu \in M} A_\mu\Big) \cup \bigcup_{\nu \in N} A_\nu,$$

which we could also write in an alternate form

$$\bigcup_{\lambda \in \Lambda} A_\lambda \cup \Big(\bigcup_{\mu \in M} A_\mu \cup \bigcup_{\nu \in N} A_\nu\Big) = \bigcup_{\alpha \in \Lambda \cup M \cup N} A_\alpha$$

The same equations hold with "$\cap$" replacing "$\cup$" everywhere.

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

More generally,

$$\Big(\bigcup_{\lambda \in \Lambda} A_\lambda\Big) \cap \Big(\bigcup_{\mu \in M} B_\mu\Big) = \bigcup_{\lambda \in \Lambda,\, \mu \in M} (A_\lambda \cap B_\mu) \quad \text{and}$$

$$\Big(\bigcap_{\lambda \in \Lambda} A_\lambda\Big) \cup \Big(\bigcap_{\mu \in M} B_\mu\Big) = \bigcap_{\lambda \in \Lambda,\, \mu \in M} (A_\lambda \cup B_\mu)$$

Proof To prove two sets are equal we show that they have the same elements; the most basic way to do this is to show that if x is in the set on the left hand side (LHS) of the proposed equation, then x must also be in the set on the right hand side (RHS) (*thereby proving LHS $\subseteq$ RHS*) and vice-versa. All parts of the theorem are easy to prove; we illustrate by proving the last equality:

$$\left(\bigcap_{\lambda\in\Lambda} A_\lambda\right) \cup \left(\bigcap_{\mu\in M} B_\mu\right) = \bigcap_{\lambda\in\Lambda,\,\mu\in M} (A_\lambda \cup B_\mu)$$

If $x \in$ LHS, then $x \in \bigcap_{\lambda\in\Lambda} A_\lambda$ or $x \in \bigcap_{\mu\in M} B_\mu$.

> If $x \in \bigcap_{\lambda\in\Lambda} A_\lambda$, then $x \in A_\lambda$ for every $\lambda \in \Lambda$, so $x \in A_\lambda \cup B_\mu$ for every $\lambda \in \Lambda$ and every $\mu \in M$, so $x \in$ RHS.
> If $x \in \bigcap_{\mu\in M} B_\mu$, then $x \in B_\mu$ for every $\mu \in M$, so $x \in A_\lambda \cup B_\mu$ for every $\lambda \in \Lambda$and every $\mu \in M$, so $x \in$ RHS.

Therefore LHS $\subseteq$ RHS.

Conversely, suppose Λ and M are nonempty. If $x \notin$ LHS, then $x \notin \bigcap_{\lambda\in\Lambda} A_\lambda$ and $x \notin \bigcap_{\mu\in M} B_\mu$, so there exist indices λ_0 and μ_0 such that $x \notin A_{\lambda_0}$ and $x \notin B_{\mu_0}$. Then $x \notin A_{\lambda_0} \cup B_{\mu_0}$, so $x \notin$ RHS.

Therefore RHS $\subseteq$ LHS, so RHS $=$ LHS. (*What happens if* $\Lambda = \emptyset$ *or* $M = \emptyset$?) •

Remarks Part 1) of the theorem states the <u>commutative</u> laws for union and intersection.
Part 2) states the <u>associative</u> laws.
Part 3) states the <u>distributive</u> laws.

Exercise Try to write a generalization of the commutative laws for infinite families.

Definition 4.7 If A and B are sets, then $A - B = \{x \in A : x \notin B\}$ is called the <u>complement of B in A</u> . If the set A is clearly understood, we might simply refer to $A - B$ as the <u>complement of B</u>, sometimes written as B^c or $-B$ or $\widetilde{B}$.

Theorem 4.8 (DeMorgan's Laws) For sets A and B_λ,

1) $A - \bigcup\{B_\lambda : \lambda \in \Lambda\} = \bigcap\{A - B_\lambda : \lambda \in \Lambda\}$, and
2) $A - \bigcap\{B_\lambda : \lambda \in \Lambda\} = \bigcup\{A - B_\lambda : \lambda \in \Lambda\}$

Proof Exercise (*DeMorgan's Laws are very simple but <u>important</u> tools for manipulating sets.*)

Definition 4.9 The set $A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$ is called the <u>product</u> of A and B. This definition can be extended in an obvious way to any finite product of sets: for example $A \times B \times C = \{(a, b, c) : a \in A,\ b \in B,\ c \in C\}$.

If $A = B$ then we sometimes write A^2 for $A \times A$. For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Note that $A \times B$ and $B \times A$ are usually not the same. (*In fact,* $A \times B = B \times A$ *iff ... ?*)

$A \times B$ is a set of ordered pairs, and (if every object in mathematics is a set) an ordered pair should itself be defined as a set. The characteristic behavior of ordered pairs is that $(a, b) = (c, d)$ iff $a = c$ and $b = d$. So we want to define (a, b) to be a set having this property. Any set that behaves in this way would be as good as any other to use as the official definition of an ordered pair; defining $(a, b) = \{a, b\}$ would not work. The most commonly used definition is due to the Polish topologist Kazimierz Kuratowski (1896-1980): we define

$$(a, b) = \{\{a\}, \{a, b\}\}$$

If we use this definition then we can prove that $(a, b) = (c, d)$ iff $a = c$ and $b = d$. (*Try it – and remember, in your argument, that a, b, c, d may not be distinct!*)

Exercise: There are other ways one could define an ordered pair. For example, a possible alternate definition is due to the American mathematician Norbert Wiener (1894-1964): we could define

$$(a, b) = \{\{\{a\}, \emptyset\}, \{\{b\}\}\}.$$

Using this definition, prove that $(a, b) = (c, d)$ iff $a = c$ and $b = d$.

> ***Note*** *In the discussion of paradoxes, we stated that sets should only be defined as subsets of other sets. Therefore, a careful definition of $A \times B$ should read:*
>
> $$A \times B = \{(a, b) \in X : a \in A \textit{ and } b \in B\}.$$
>
> *So, if we were doing axiomatic set theory, we would have to provide, as part of the definition, a set X which we know exists and define $A \times B$ to be a subset of X. Given A and B, how in general would we supply X? To give the flavor of how axiomatic set theory proceeds, we'll elaborate on this point.*
>
> > *Axiomatic set theory begins with some axioms: the variables $x, y, \ldots$ refer to sets and $\in$ refers to the membership relation. In axiomatic set theory, everything is a set, so members of sets are other sets. Therefore we see notation like $x \in y$ in the axioms: "the set x is a member of the set y." A partial list of these axioms includes:*
> >
> > *A1)* ***two sets are equal iff they have exactly the same elements****: more formally,*
> >
> > $$\forall x \, \forall y \, (x = y \Leftrightarrow \forall z \, (z \in x \Leftrightarrow z \in y))$$
> >
> > *A2)* ***there is an empty set****: more formally,*
> >
> > $$\exists x \, (\forall y \; \neg(y \in x))$$
> >
> > (*Notice, by A1, this set x is unique, so we can give this set a name: $\emptyset$*)
> >
> > *A3)* ***any two sets x and y can be "paired" to create a new set $\{x, y\}$****:*
> > *more formally,*
> >
> > $$\forall x \, \forall y \, \exists z \forall w \, (w \in z \Leftrightarrow (w = x \vee w = y))$$
> >
> > *A4)* ***every set x has a power set y****: more formally,*
> >
> > $$\forall x \, \exists y \, \forall z \, (z \in y \Leftrightarrow (\forall w \, (w \in z \Rightarrow w \in x))$$

A5) ***the union u of any set x exists****: more formally,*

$$\forall x\, \exists u\, \forall z\ (z \in u \Leftrightarrow (\exists y\, (\, z \in y\ \wedge\ y \in x\,)))$$

The last axiom to be mentioned here, A6, is a little harder to write precisely, so we will simply state an informal version of it:

A6) ***for any set x, there is a subset y of x whose members are all the elements in x that satisfy an appropriate "legal" description "..." .*** *Roughly,*

$$\forall x\ \exists y\ \forall z\ \ (\, z \in y \Leftrightarrow (z \in x\ \wedge\ \text{"..."})\,)$$

A6) is a fancier (but still somewhat vague) version of saying that we are allowed to define a set y by writing:

$$y = \{z \in x :\ \text{"..."}\ \}$$

These axioms and three others A7) – A9) – which we will not state here – are called the *<u>Zermelo-Fraenkel Axioms</u>, or ZF for short. (One of the additional axioms, for example, implies that no set is an element of itself.) The resulting system (the axioms and all the theorems that can be proved from the axioms) is called <u>ZF set theory</u>.*

In ZF, by using these axioms, we can make a complete definition of $A\ \times B$. Suppose $a \in A$ and $b \in B$. By axiom A3), the sets $\{A, B\}$, $\{a, b\}$ and $\{a, a\} = \{a\}$ exist so, by axiom A3) again, the set $\{\{a\}, \{a, b\}\} = (a, b)$ also exists.

By axiom A5) $A \cup B =$ the union of the set $\{A, B\}$ exists, and by axiom 4), the sets $\mathcal{P}(A \cup B)$ and $\mathcal{P}(\mathcal{P}(A \cup B))$ exist. Using these sets, we then notice that

$a \in A \cup B$ *and* $b \in A \cup B,$	*so*
$\{a\} \subseteq A \cup B$ *and* $\{a, b\} \subseteq A \cup B,$	*so*
$\{a\} \in \mathcal{P}(A \cup B)$ *and* $\{a, b\} \in \mathcal{P}(A \cup B),$	*so*
$\{\{a\}, \{a, b\}\} \subseteq \mathcal{P}(A \cup B),$	*so*
$\{\{a\}, \{a, b\}\} = (a, b) \in \mathcal{PP}(A \cup B)$	

Therefore each pair $(\, a,\, b\,)$ that we want to put in the set to be called $A \times B$ happens to be a member of the set $X = \mathcal{PP}(A \cup B)$, whose existence follows from the axioms. Of course, not every element of $\mathcal{PP}(A \cup B)$ is an ordered pair. Putting all this together, we could then make the definition

$$A \times B = \{z \in \mathcal{PP}(A \cup B) \colon\ (\exists a \in A)(\exists b \in B)\ z = (a, b)\}.$$

In the rest of these notes we will usually not refer to the ZF axioms. However, we will occasionally make comments about the axioms when something interesting is going on.

Exercises

E3. Which of the following are true when " $\in$ " is inserted in the blank space? Which are true
when " $\subseteq$ " is inserted?

a) $\{\emptyset\}$ ___ $\{\emptyset, \{\emptyset\}\}$
b) $\{\emptyset\}$ ___ $\{\emptyset, \{\{\emptyset\}\}\}$
c) $\{\{\emptyset\}\}$ ___ $\{\emptyset, \{\emptyset\}\}$
d) $\{\{\emptyset\}\}$ ___ $\{\emptyset, \{\{\emptyset\}\}\}$
e) $\{\{\emptyset\}\}$ ___ $\{\emptyset, \{\emptyset, \{\emptyset\}\}\}$

E4. a) Suppose A and $\mathcal{B}$ are sets and $A \in \mathcal{B}$. Prove, or give a counterexample, to each of the following statements:

i) $\mathcal{P}(A) \in \mathcal{P}(\mathcal{B})$
ii) $\mathcal{P}(A) \in \mathcal{P}(\bigcup\mathcal{B})$
iii) $\mathcal{P}(A) \subseteq \mathcal{P}(\bigcup\mathcal{B})$

b) Let $A = \mathcal{P}(\emptyset)$, $B = \mathcal{P}(A)$ and $C = \mathcal{P}(B)$. What is $A \cap B \cap C$?

c) Give an example of a set A which has more than one element and such that whenever $x \in A$, then $x \subseteq A$. (Such a set is called transitive.)

E5. Explain why the following statement is or is not true:

If $A = \{\{\emptyset\}, \{\{\emptyset\}\}\}$, then $\bigcup\{B : B \in \mathcal{P}(A)\} = \{\{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}\}$

E6. a) Prove that $\mathcal{P}(A - B) \cup \mathcal{P}(B) = \mathcal{P}(B - A) \cup \mathcal{P}(A)$ if and only if ($A = B$ or $A \cap B = \emptyset$).

b) State and prove a theorem of the form:

$\mathcal{P}(A - B) - \{\emptyset\} = \mathcal{P}(A) - \mathcal{P}(B)$ if and only if ...

c) For any sets A and B it is true that $\mathcal{P}(A) \cap \mathcal{P}(B) = \mathcal{P}(A \cap B)$ and that $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$.

State and prove a theorem of the form:

$\mathcal{P}(A) \cup \mathcal{P}(B) = \mathcal{P}(A \cup B)$ if and only if ...

E7. Suppose $A, B, C,$ are sets with $A \neq \emptyset$, $B \neq \emptyset$ and $(A \times B) \cup (B \times A) = C \times C$. Prove that $A = B = C$.

E8. Suppose $A, B, C,$ and D are sets, with $A \neq \emptyset$ and $B \neq \emptyset$. Show that if

$$(A \times B) \cup (B \times A) = (C \times D) \cup (D \times C),$$

then either $(A = C$ and $B = D)$ or $(A = D$ and $B = C)$.

E9. Let A_1, A_2,..., A_n,... be subsets of A. Define $\underline{\lim}\ A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

($\underline{\lim}$ is read "lim inf"). In expanded form, the notation means that

$$\underline{\lim}\ A_n = (A_1 \cap A_2 \cap A_3 \cap ...) \cup (A_2 \cap A_3 \cap A_4 \cap ...) \cup (A_3 \cap A_4 \cap A_5 \cap ...) \cup ... \ .$$

Similarly, define $\overline{\lim}\ A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ ($\overline{\lim}$ is read "lim sup").

a) Prove that $\underline{\lim}\ A_n = \{x : x$ is in all but at most finitely many $A_n\}$ and that
$\overline{\lim}\ A_n = \{x : x$ is in infinitely many of the $A_n\}$.

b) Prove that $\bigcap_{n=1}^{\infty} A_n \subseteq \underline{\lim}\ A_n \subseteq \overline{\lim}\ A_n \subseteq \bigcup_{n=1}^{\infty} A_n$

c) Assume that all the A_n's are subsets of a set X. Prove that

$$\underline{\lim}\,(X - A_n) = X - \overline{\lim}\ A_n$$

d) Prove that $\underline{\lim}\ A_n = \overline{\lim}\ A_n$ if either $A_1 \subseteq A_2 \subseteq A_3 \subseteq$ ... or $A_1 \supseteq A_2 \supseteq A_3 \supseteq$ ... .

E10. Suppose $A,\ B, C,$ and D are nonempty sets satisfying

$$A \neq B,\ C \neq D, \text{ and } \{A, B\} \neq \{C, D\}.$$

Prove that

~

$$(A \times (B - A)) \cup (B \times (A - B)) \neq (C \times (D - C)) \cup (D \times (C - D))$$

5. Functions

Suppose X and Y are sets. Informally, a function (or mapping, or map) f from X into Y is a rule that assigns to each element x in X a unique element y in Y. We call y the image of x and call x a preimage of y. We write $y = f(x)$, and we denote the function by $f : X \to Y$. This informal definition is usually good enough for our purposes.

X is called the domain of $f = \text{dom}(f)$. The range of $f = \text{ran}(f)$ is the set

$$\{y \in Y\colon y = f(x) \text{ for some } x \in X\}.$$

We can think of giving an "input" x to f and the corresponding "output" is $y = f(x)$. Then dom(f) is the set of "all allowed inputs" and ran(f) is the "set of all outputs" corresponding to those inputs.

Y is sometimes referred to as the codomain of f. Of course, $\text{ran}(f) \subseteq Y$, but $\text{ran}(f)$ may not be the whole codomain Y. When $\text{ran}(f) =$ the codomain Y, we say that f is a function from X onto Y, or simply that f is onto. In some books, an onto function is also called a surjection.

If different inputs always produce different outputs, we say that f is a one-to-one (or 1-1) function. More formally: f is one-to-one if, whenever $a, b \in \text{dom}(f)$ and $a \neq b$, then $f(a) \neq f(b)$. In some books, a one-to-one function is also called an injection.

If f is both one-to-one and onto, we call f a bijection between X and Y. A bijection sets up a perfect one-to-one correspondence between the elements of the two sets X and Y. Intuitively, a bijection can exist if and only if X and Y have the same number of elements.

Examples 5.1 (verify the details as needed)

1) Suppose $X = Y$ and $f : X \to X$ is given by $f(x) = x$. This bijection is called the identity map on X, sometimes denoted by $f = i_X$.

2) If $f : X \to Y$ and $A \subseteq X$, then we can define a new function $g : A \to Y$ by $g(x) = f(x)$ for $x \in A$; g is called the restriction of f to A, denoted $g = f|A$. If f is one-to-one then so is g; but g may not be onto even if f is onto. Note that we consider f and g to be different functions when $A \neq X$ – because $\text{dom}(f) \neq \text{dom}(g)$.

For example, $g = \sin|[-\frac{\pi}{2}, \frac{\pi}{2}]$ is a bijection between $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[-1, 1]$.

3) Let $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by the rule $f(m, n) = 2^m 3^n$. Then f is one-to-one by the Fundamental Theorem of Arithmetic (which states that each natural number has a prime factorization that is unique except for order of the factors). But f is not onto because, for example, $35 \notin \text{ran}(f)$.

4) Let $X =$ the interval $(1, \infty) \subseteq \mathbb{R}$ and $Y = \mathbb{R}$. Define $f(x) = \int_1^\infty \frac{1}{t^x}\, dt$. (*This definition makes sense since the improper integral converges for each* $x > 1$.) For example, $f(2) = \int_1^\infty \frac{1}{t^2}\, dt = 1$. Then $f : X \to Y$ is one-to-one but not onto (why?).

5) Let $f : \{2, 3, 4, ...\} \to \mathbb{N}$. where $f(x) =$ "the least integer ≥ 2 which divides x."
For example, $f(2) = 2$, $f(3) = 3$, $f(4) = 2$, $f(5) = 5$, $f(6) = 2$, and $f(9) = 3$. The function f is not one-to-one because $f(2) = f(4)$, and f is not onto because $\mathbb{N} \neq \text{ran}(f) =$ the set of prime numbers.

6) Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) =$ the n^{th} prime number. For example $f(1) = 2$, $f(2) = 3$, $f(3) = 5$, and $f(4) = 7$. The function f is clearly one-to-one but not onto.

7) Let $\mathbb{S}$ be the set of prime numbers and $f : \mathbb{N} \to \mathbb{S}$ using the same rule as in Example 6.
Then f is a bijection between $\mathbb{N}$ and $\mathbb{S}$. This illustrates that whether $f : X \to Y$ is onto depends on the codomain Y. If you know the domain of f and a rule that defines f, then these completely <u>determine the range</u> of f, but they <u>do not determine the codomain</u> of f. So without naming the set Y, we cannot say whether a function f is onto.

8) Define $\pi : \mathbb{R} \to \{z \in \mathbb{Z} : z \geq 0\}$ by $\pi(x) =$ "the number of primes $\leq x$". Thus, $\pi(1) = 0$, $\pi(2) = 1$, $\pi(2.5) = 1$, and $\pi(5.6) = 3$. The function π is onto (why?) but not one-to-one.
This function appears in the famous Prime Number Theorem which states that

$$\lim_{x \to \infty} \frac{\pi(x) \ln(x)}{x} = 1.$$

A proof of the Prime Number Theorem is quite difficult. However, a much simpler fact is that $\lim_{x \to \infty} \pi(x) = \infty$. *Why is this simpler fact true? And does it also follow <u>from</u> the Prime Number Theorem?*

9) Define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. This function is one-to-one but not onto. (Why?)

10) Let A be a set and define $f : A \to \mathcal{P}(\mathcal{P}(A))$ by

$$f(a) = \text{“the set of all subsets of } A \text{ containing } a\text{”} = \{B \in \mathcal{P}(A) : a \in B\}.$$

For example, if $A = \{1, 2, 3\}$, then $f(2) = \{\{2\}, \{1, 2\}, \{2, 3\}, \{1, 2, 3\}\}$.

This function f is one-to-one. To see this, suppose that $a \neq b$; then $\{a\} \in f(a)$ and $\{a\} \notin f(b)$, so $f(a) \neq f(b)$. Is the function onto?

11) Define $\pi : X \times Y \to Y$ by $\pi(x, y) = y$, called the <u>projection of $X \times Y$ to Y</u>." When is this projection function $1 - 1$? Must π be onto?

12) Let $X = \{C : C$ is a rectifiable curve in the plane$\}$. (*A rectifiable curve is one for which an arc-length is defined.*) Let $f : X \to [0, \infty)$ by $f(C) =$ "the length of the curve C." Is f one-to-one? onto?

Our informal definition of function is good enough for most purposes. However in the spirit of "every object in mathematics is a set," we should be able to give a more precise definition of a "function f from X into Y" as a set. To do this, we begin by using sets to define a relation.

Definition 5.2 A relation R from X to Y (or between X and Y) is a subset of $X \times Y$. If $(x, y) \in R$, we write xRy, meaning that "x is related to y" (by the relation R).

Example 5.3 Consider $\{(x, y) \in \mathbb{R} \times \mathbb{R} :$ for some $b \in \mathbb{R},\ y = x + b^2\}$. This set of ordered pairs is a relation from $\mathbb{R}$ to $\mathbb{R}$. We can call this relation R, but it actually has a more familiar name, $\leq$:

$$\leq\ = \{(x, y) \in \mathbb{R} \times \mathbb{R} : \text{for some } b \in \mathbb{R},\ y = x + b^2\}$$

The relation $\leq$ is a set. The statement "$x \leq y$" means that the pair $(x, y) \in\ \leq$. Notice that for each x there are many y's for which $(x, y) \in\ \leq$: for example $(1, 2) \in\ \leq$, $(1, \pi) \in\ \leq$, ...

> **Exercise:** The relation $\leq$ is a set. What sets are the familiar relations $>$, $<$, and $\geq$ between $\mathbb{R}$ and $\mathbb{R}$?

A function f is a special kind of relation from X to Y: one in which every x is related by f to only one y in Y. The relation $\leq$ in example 5.3 is not a function.

Definition 5.4 A function f from X into Y, denoted $f : X \to Y$ is a set $f \subseteq X \times Y$ (*so f is a relation from X to Y*) with the additional properties that:

a) for every $x \in X$, there is a $y \in Y$ such that $(x, y) \in f$, and

b) if $(x, y) \in f$ and $(x, z) \in f$, then $y = z$.

X is called the domain of f and the range of f is the set

$$\{y \in Y : \text{there exists an } x \in X \text{ for which } (x, y) \in f\}$$

If f is a function from X to Y and $(x, y) \in f$, then we will use standard function notation and write $y = f(x)$ rather than use the relation notation xfy.

Condition a) states that a function f from X into Y is defined at each point of X, and condition b) states that f is single-valued – f can't have more than one value for a particular x. For functions $f : \mathbb{R} \to \mathbb{R}$, condition b) just states the familiar precalculus condition that a vertical line cannot intersect the graph of a function f in more than one point.

Example 5.5 Let $f : \mathbb{R} \to \mathbb{R}$ be the squaring function, $f(x) = x^2$. By our formal definition, this function is a subset of $\mathbb{R}^2$: $f = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R}^2$. Picturing this set of ordered pairs in $\mathbb{R}^2$ gives the parabola: $y = x^2$. In precalculus, this set is called the graph of the function, but in our formal definition, the set literally is the function f : that is, f is defined to be the set of pairs which, in precalculus, is called its graph.

Example 5.6 For sets X and Y, we write Y^X to represent the set of all functions from X to Y. For example, $\mathbb{R}^{\mathbb{R}}$ is the set of all real-valued functions of a real variable.

<u>If $X = \emptyset$, then there is exactly one function in</u> Y^X : <u>the empty function</u> $\emptyset$.
(Check: $\emptyset \subseteq \emptyset \times Y$ and $\emptyset$ satisfies the conditions a) and b) in Definition 5.4.)
So $Y^X = \{\emptyset\}$.

<u>If $X \neq \emptyset$ and $Y = \emptyset$, then there are no functions in</u> Y^X, so $Y^X = \emptyset$.
(Since $X \times \emptyset = \emptyset$, the only relation from X to Y is $\emptyset$, so $\emptyset$ is the only "candidate" for a function from X to Y. But since $X \neq \emptyset$, $\emptyset$ fails to satisfy part a) in Definition 5.4.)

<u>If X has n elements and Y has m elements ($n, m \in \mathbb{N}$), then there are</u> m^n <u>functions in</u> Y^X. (Why? For each $x \in X$, how many choices are there for $f(x)$?)

It is usually not necessary to think of a function as a set of ordered pairs; the informal view of a function as a "rule" usually is good enough. The formal definition is included partly to reiterate the point of view that "every object in mathematics is a set." But sometimes this point of view is also useful notationally: if f and g are functions then, since f and g are sets, it makes sense to form the sets $f \cup g$ and $f \cap g$. Sometimes these sets are new functions.

1) Suppose $f : (-\infty, 0] \to \mathbb{R}$ and $g : [0, \infty) \to \mathbb{R}$ are given by $f(x) = -x$ and $g(x) = x$. As sets, $f = \{(x, -x) : x \leq 0\}$ and $g = \{(x, x) : x > 0\}$. Then the set $h = f \cup g$ is a function with domain $\mathbb{R}$. For $x \in \mathbb{R}$, $h(x) =$?
Is the set $k = f \cap g$ a function? If so, what is its domain and what is a formula for the function?

2) In general, if f and g are functions, must $f \cup g$ and $f \cap g$ <u>always</u> be functions?
If not, then what conditions must f and g satisfy so that $f \cup g$ and $f \cap g$ are functions?
If they are functions, what is the domain of each?

Example 5.7 As another illustration of how "every object in mathematics is a set," consider <u>the real number system</u> $-$ it consists of the set $\mathbb{R}$, two operations called addition ($+$) and multiplication ($\cdot$), and an order relation $\leq$. (*Subtraction and division are defined in terms of addition and multiplication.*) Notice that the operations $+$ and $\cdot$ are <u>functions</u> $+ : \mathbb{R}^2 \to \mathbb{R}$ and $\cdot : \mathbb{R}^2 \to \mathbb{R}$.

Therefore, $+ \subseteq \mathbb{R}^2 \times \mathbb{R}$ and, for example, $((2,3), 5) \in +$. In function notation this would be written $+(2,3) = 5$, but instead we usually write: $2 + 3 = 5$.

The functions $+$ and $\cdot$ are required to obey certain axioms such as $(x, y) \in +$ iff $(y, x) \in +$, that is, $x + y = y + x$. The order relation $\leq$ is also just a certain subset of $\mathbb{R}^2$ (see Example 5.3).

Therefore we can think of the real number <u>system</u>, $-$ its elements and all its ordering and arithmetic $-$ as being "captured" in the 4 sets: $\mathbb{R}, +, \cdot$, and $\leq$, and we can gather all this together into a single object: the 4-tuple of sets: $(\mathbb{R}, +, \cdot, \leq)$

But this 4-tuple can be viewed as just a complicated ordered pair: $(((\mathbb{R}, +), \cdot), \leq)$. Since each ordered pair is just a set, we conclude: the whole real number system, with all its ordering and operations, can be thought of as one single (complicated) set.

6. More About Functions

Suppose $f : X \to Y$ and that $A \subseteq X$, $B \subseteq Y$. The image of the set A is denoted $f[A]$; it is the set of all images of the elements of A. More precisely, $f[A] = \{y \in Y : \ y = f(a) \text{ for some } a \in A\}$. A little less formally we could also write $f[A] = \{f(a) : a \in A\}$.

The inverse image set of B is denoted $f^{-1}[B]$; it is the subset of X consisting of all preimages of all the elements from B. More precisely, $f^{-1}[B] = \{x \in X : f(x) \in B\}$.

Example 6.1 Suppose $f : \mathbb{R} \to \mathbb{R}$ is the squaring function, $f(x) = x^2$. Then

$$f[\,[0,2]\,] = [0,4] \qquad f^{-1}[\,[0,9]\,] = [\,-3,3]$$

$$f[\,\{2\}] = \{4\} \qquad f^{-1}[\,\{4\}] = \{\,-2,2\}$$

$$f[\,[\,-3,\,-2]\,] = [4,9] \qquad f^{-1}[\,[\,-5,\,-4]\,] = \emptyset$$

The function f is not onto since $f[\mathbb{R}] \neq \mathbb{R}$.

When B is a one point set such as $\{4\}$, we will often write $f^{-1}[4]$ or even $f^{-1}(4)$ rather than the more formal $f^{-1}[\{4\}]$. Be careful when using this more informal notation in which $f^{-1}(4) = \{\,-2,2\}$ is a set, not a number.

The next theorem gives some properties of image and inverse image sets that are frequently used.

Theorem 6.2 Suppose $f\colon X \to Y$, that $A_\lambda \subseteq X$ and $B_\lambda \subseteq Y$ $\quad (\lambda \in \Lambda)$

1) $f[\emptyset] = \emptyset$ and $f[X] \subseteq Y$ $\qquad$ 1′) $f^{-1}[\emptyset] = \emptyset$ and $f^{-1}[Y] = X$

2) if $A_1 \subseteq A_2$, then $f[A_1] \subseteq f[A_2]$ $\qquad$ 2′) if $B_1 \subseteq B_2$, then $f^{-1}[B_1] \subseteq f^{-1}[B_2]$

3) $f[\bigcup A_\lambda] = \bigcup f[A_\lambda]$ $\qquad$ 3′) $f^{-1}[\bigcup B_\lambda] = \bigcup f^{-1}[B_\lambda]$

4) $f[\bigcap A_\lambda] \subseteq \bigcap f[A_\lambda]$ $\qquad$ 4′) $f^{-1}[\bigcap B_\lambda] = \bigcap f^{-1}[B_\lambda]$

Proof The proof of each part is extremely simple. As an example, we prove 4′).

Let LHS and RHS denote the left and right sides of 4′). Then $x \in$ LHS iff $f(x) \in \bigcap B_\lambda$ iff $f(x) \in B_\lambda$ for every $\lambda \in \Lambda$ iff $x \in f^{-1}[B_\lambda]$ for every $\lambda \in \Lambda$ iff $x \in$ RHS. Thus LHS and RHS have the same members and are equal. ●

Example 6.3 In part 4) of Theorem 6.2, LHS = RHS <u>may be false</u>. For example, suppose $A_1 = \{x \in \mathbb{R} : x < 1\}$, $A_2 = \{x \in \mathbb{R} : x > 1\}$ and let f be the constant function $f(x) = 1$. Then $f[A_1] = f[A_2] = \{1\}$, so $f[A_1] \cap f[A_2] = \{1\}$, but $f[A_1 \cap A_2] = f[\emptyset] = \emptyset$.

Can you think of an additional hypothesis about the function f that would guarantee that " $=$ " holds in part 4) ?

Definition 6.4 Suppose $f : X \to Y$ and $g : Y \to Z$ are functions. Their <u>composition</u> $g \circ f$ is a function from X to Z formed by first applying f, then g. More precisely, the <u>composition</u> $(g \circ f) : X \to Z$ is defined by $(g \circ f)(x) = g(f(x))$.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & Y \arrow[r, "g"] & Z \end{tikzcd}$$

If we have another function $h : Z \to W$, we can form the "triple compositions" $h \circ (g \circ f)$ and $(h \circ g) \circ f$. It is easy to check that these functions $X \to W$ are the same. In other words, the associative law holds for composition of functions and we may write $h \circ g \circ f$ without worrying about parentheses.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rrr, bend right, "h \circ g \circ f"'] & Y \arrow[r, "g"] & Z \arrow[r, "h"] & W \end{tikzcd}$$

Here is a useful observation about compositions and inverse image sets: if $A \subseteq Z$, then

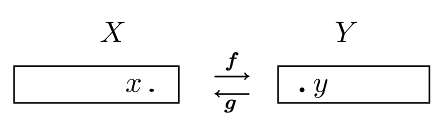
$$(g \circ f)^{-1}[A] = f^{-1}[g^{-1}[A]]$$

To check this, note that $x \in$ RHS iff $f(x) \in g^{-1}[A]$ iff $g(f(x)) \in A$ iff $(g \circ f)(x) \in A$ iff $x \in$ LHS.

If $f : X \to Y$ is a <u>bijection</u>, then f gives a perfect one-to-one correspondence between the elements of X and Y. In that case, we can define a function $g : Y \to X$ as follows.

> Suppose $y \in Y$. Then $y = f(x)$ for some $x \in X$ (*because f is onto*) and there is only one such x (*because f is one-to-one*). Define $g(y) = x$. Then $g : Y \to X$ and $g(y) = x$ if and only if $f(x) = y$.

More formally, $g = \{(y, x) \in Y \times X : x \in f^{-1}(y)\}$. Because f is a bijection, the inverse image set $f^{-1}(y)$ contains exactly one member, x – so b) in Definition 5.4 is true.



It follows that $f(g(y)) = y$ and $g(f(x)) = x$, that is, $f \circ g = i_Y : Y \to Y$ <u>and</u> $g \circ f = i_X : X \to X$. The function g defined above is clearly the only possible function with both of these properties.

Of course, g is also a bijection and the whole discussion could be carried out starting with g and using it to "get back" the function f, which has the property that $f \circ g = i_Y : Y \to Y$.

Definition 6.5 Suppose $f : X \to Y$ and $g : Y \to X$ and that $g \circ f = i_X$ and $f \circ g = i_Y$. Then we say f and g are inverse functions to each other. We write $g = f^{-1}$ and $f = g^{-1} = (f^{-1})^{-1}$.
(*The idea is that inverse functions "undo" each other.*)

The preceding discussion proves the first part of the following theorem.

Theorem 6.6 Suppose $f: X \to Y$. Then

1) If f is a bijection, there is a unique bijection $g : Y \to X$ for which $g \circ f = i_X$ and $f \circ g = i_Y$.

2) If f has an inverse function g, then f is a bijection.

Proof To prove part 2), notice that if $f(a) = f(b)$, then $g(f(a)) = g(f(b))$. But $g \circ f = i_X$, so this implies that $a = b$, so f is one-to-one. Also, for any $y \in Y$ we have $g(y) = x \in X$, and since $f \circ g = i_Y$, we get that $f(g(y)) = f(x) = y$. Therefore y is the image of the point $x \in X$, so f is onto. ●

Comment on notation: For any function $f : X \to Y$ and $B \subseteq Y$, we can write the inverse image set $f^{-1}[B]$, whether or not f is bijective and has an inverse function. In particular, for any function f we might write $f^{-1}(y)$ as a shorthand for $f^{-1}[\{y\}]$ = the inverse image set of the one-point set $\{y\}$.

If f happens to be bijective, then the notation $f^{-1}(y)$ is ambiguous: "f^{-1}" might mean the inverse image operation for which $f^{-1}(y) = \{x\}$, or "f^{-1}" might mean the inverse function, in which case $f^{-1}(y) = x$.

This double use of the symbol f^{-1} to denote the inverse image set operation (defined for any function) and also to denote the inverse function (when f is bijective) usually causes no confusion in practice – the intended meaning is clear from the context.

In a situation where the ambiguity might cause confusion, we simply avoid the shorthand and use the more awkward notation $f^{-1}[\{y\}]$ for the inverse image set..

As a simple exercise in notation, convince yourself that f is bijective iff $f^{-1}[\{y\}]$ is a one-element subset of X for every $y \in Y$.

Examples 6.7 (In each part, verify the details.)

1) Let $\ln : (0, \infty) \to \mathbb{R}$ be defined by $\ln x = \int_1^x \frac{1}{t}\, dt \quad (x > 0)$. This function is bijective (why?) and therefore has an inverse function $\ln^{-1} : \mathbb{R} \to (0, \infty)$. The function $\ln^{-1}$ is often called exp, so $\exp \circ \ln : (0, \infty) \to (0, \infty)$ and $\exp(\ln x) = x$ for every $x > 0$. Similarly, $\ln \circ \exp : \mathbb{R} \to \mathbb{R}$ and $\ln(\exp x) = x$ for each $x \in \mathbb{R}$. (*It turns out, of course, that $\exp x = e^x$, although that requires proof.*)

2) The function $f : \mathbb{N} \to \{2, 4, 6....\} = \mathbb{E}$ given by $f(n) = 2n$ is bijective; its inverse function is $g : \mathbb{E} \to \mathbb{N}$ given by $g(e) = \frac{e}{2}$.

3) A linear function $f : \mathbb{R}^n \to \mathbb{R}^n$ can be expressed as multiplication by some $n \times n$ matrix M: $f(x) = M \cdot x$. Then f is bijective iff det(M) $\neq 0$ and, in that case, $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is multiplication by the inverse matrix M^{-1} and

$$f^{-1}(f(x)) = M^{-1} \cdot M \cdot x = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & & \vdots & \\ 0 & 0 & \ldots & 1 \end{bmatrix} \cdot x = x$$

These are facts you should know from linear algebra.

> **Exercise** (*if you have had an analysis course*) Prove that if $f : \mathbb{R} \to \mathbb{R}$ is continuous and onto, then f has an inverse iff f is strictly monotone.
> *Hint: Remember the Intermediate Value Theorem. If f is not strictly monotone, then there must exist three points $a < c < b$ such that $f(c)$ is either $\leq$ or $\geq$ both of $f(a)$ and $f(b)$.*

If we look carefully, we see that part 1) of Theorem 6.8 can be broken into two pieces.

Theorem 6.8 1) $f: X \to Y$ is one-to-one iff there exists a function $g : Y \to X$ such that $g \circ f = i_X$ (unless $X = \emptyset$, $Y \neq \emptyset$, and $f = \emptyset$; why?)

2) $f: X \to Y$ is onto iff there exists a function $g : Y \to X$ such that $f \circ g = i_Y$.

Proof The ideas are quite similar to the argument establishing the existence of an inverse for a bijection. The proof of 1) is left as an exercise. We will illustrate by proving part 2), which is of special interest because of a subtle point that comes up.

If such a g exists, then for each $y \in Y$ we have $f(g(y)) = y$, so y is the image of the point $x = g(y) \in X$. Therefore f is onto.

Conversely suppose f is onto so that, for each $y \in Y$, we know that $f^{-1}(y) \neq \emptyset$. Choose an arbitrary element $x \in f^{-1}(y)$ and define $g(y) = x$. Clearly, $f \circ g = i_Y$. (Note that if $Y = \emptyset$, then we must have $f = \emptyset$ and $X = \emptyset$. The description above then defines $g = \emptyset$, and $g \circ f = i_Y = \emptyset$ is still true.) •

The subtle point of the proof lies in the definition of g which, in set notation, reads:

$$g = \{(y, x) \in Y \times X : \ x \textit{ is an element } \underline{\textit{arbitrarily}}\ \underline{\textit{chosen}} \textit{ from } f^{-1}(y)\}$$

In discussing paradoxes, we stated that a set g could only be formed by abstraction as $g = \{(y, x) \in Y \times X$: ... "some legal description"... $\}$. If one is doing axiomatic set theory, the phrase "an arbitrarily chosen element" is not legal; it is "too vague." More precisely, the description cannot be properly written in the language of first order predicate calculus and therefore the "definition" of the function g would be invalid in ZF set theory.

The axioms for set theory (ZF) tell us that certain sets exist ($\emptyset$, for example) and give methods to create new sets from old ones. Roughly, these methods are of two types:

i) "from the top down" – forming a subset of a given set
ii) "from the bottom up" – somehow piecing together a new set from old ones (for example, by union, or by pairing). It is understood that the "piecing together" must be

done in a finite number of steps – for example, we cannot say "apply the Pairing Axiom infinitely many times to get the set z ..."

In the present situation ,

i) we tried to define g ("from the top down") by describing g as a certain subset of $Y \times X$; the problem is that we can't say precisely how to pick each x and therefore we can't use the Subset Axiom A6.

ii) if we tried to define g "from the bottom up," we could begin by using the fact that f is onto: this means that

$$\forall y \in Y \ \exists x \ x \in f^{-1}(y)$$

Therefore for any particular $y \in Y$, we can form an ordered pair (y, x) with $y \in f^{-1}(x)$. But if there are infinitely many pairs (y, x), we have no axiom that allows us to "gather" these pairs into a single set g.

Also, none of the ZF axioms A7) – A9) (which we didn't state) is any help here, so we seem to be stuck. But in spite of all this, our informal description of g seems intuitively sound, so another axiom called the Axiom of Choice (AC, for short) is usually added to the set theory axioms ZF to justify our intuitive argument. The system ZF together with the Axiom of Choice is referred to as ZFC set theory for short.

In one of its many equivalent forms, the axiom of choice reads:

[AC] If $\mathcal{A} = \{A_y : y \in Y\}$ is a family of nonempty sets, then there exists a function $h : Y \to \bigcup\{A_y : y \in Y\}$ such that, for each y, $h(y) \in A_y$.

AC guarantees the existence of a function (set) h that "chooses" an element $h(y)$ from each set A_y and that's just what we need here. If we use the sets $A_y = f^{-1}(y)$, then AC gives us the "choice function" h, and we can then use h to define

$$g = \{(y, x) \in Y \times X : y \in Y \wedge x = h(y)\}$$

Comment: Defining the function g is not always such a delicate matter. In some special cases, we can avoid the whole problem and don't need AC. For example:

i) if $X = \mathbb{N}$, we could quite specifically define g by saying "let x be the smallest element in $g^{-1}(y)$." In other words, we could write, without AC, a perfectly precise definition of g "from the top down" using the language of first order predicate calculus:

$$g = \{(y, x) \in Y \times X : \ x \in f^{-1}(y) \ \wedge \ \forall z \, (z \in f^{-1}(y) \Rightarrow z \geq x) \ \}.$$

ii) if Y were finite, we could index its elements so that $Y = \{y_1, y_2, ..., y_n\}$ for some $n \in \mathbb{N}$. Then we could proceed "from the bottom up" to write the finite list of statements

$$\begin{array}{l} \exists x_1 \in f^{-1}(y_1) \\ \exists x_2 \in f^{-1}(y_2) \\ \quad \vdots \\ \exists x_n \in f^{-1}(y_n) \end{array}$$

Using these $x_1, ..., x_n$ we can write a definition of g :

$$g = \{(y, x) \in Y \times X : (y = y_1 \wedge x = x_1) \vee ... \vee (y = y_n \wedge x = x_n)\}\}$$
(The description is "legal" since it involves only finitely many terms.)

Generally speaking, AC is required to create a set when it is necessary to choose an element from each nonempty set in an infinite family and there is no way to describe precisely which element to choose from each set. When the choice can be explicitly stated (as when $X = \mathbb{N}$ above), AC is not necessary.

To borrow a non-mathematical example from the philosopher Bertrand Russell: if you have an infinite collection of pairs of socks, you need AC to create a set consisting of one sock from each pair, but if you have an infinite collection of pairs of shoes, you don't need AC to create a set containing one shoe form each pair— because you can precisely describe each choice: "from each pair, pick the left shoe."

The Axiom of Choice, when added to the other axioms of set theory, makes it possible to prove some very nice results. For example, in real analysis, AC can be used to show the existence of a "nonmeasurable" set of real numbers. AC can also be used to show that every vector space (even one that's not finite dimensional) has a basis. AC is equivalent to a mathematical statement called Zorn's Lemma (see Chapter VIII) which you may have met in another course.

The cost of adding AC to the axioms ZF is that it also makes it possible to prove some very counter-intuitive results about infinite sets. Here is a famous example:

> *The Banach-Tarski paradox (1924) states that it is possible to divide a solid ball into six pieces which can be reassembled by rigid motions to form two balls of the same size as the original. The number of pieces was subsequently reduced to five by R. M. Robinson in 1944, although the pieces are extremely complicated. (Actually, it can be done with just four pieces if the single point at the center of the ball is ignored..)*

Most mathematicians are content to include AC with the other along with the other axioms and simply to be amused by some of the strange results it can produce. This is because AC seems intuitively very plausible and because many important mathematical results rely on it. We will adopt this attitude and use AC freely when it's needed (and perhaps even when it isn't!), usually without calling attention to the fact.

Definition 6.9 A sequence in a set X is a function $f : \mathbb{N} \to X$. The terms of the sequence are $f(1) = x_1$, $f(2) = x_2$, ... , $f(n) = x_n$, ... We often denote a sequence informally by the notation (x_n).

For example, the function $f(n) = 2n + 1$ defines the sequence whose terms are $3, 5, 7, 9, ...$ The n^{th} term of the sequence is $x_n = 2n + 1$, and we might refer to the sequence as (x_n) or $(2n + 1)$.

In the spirit of "every object in mathematics is a set": since a sequence in X is a function from $\mathbb{N}$ to X, a sequence (formally) is just a set $-$ in this case, a special subset of $\mathbb{N} \times X$. Of course, we usually think of a sequence, informally, as an infinite list of objects: $x_1, x_2, ..., x_n, ...$. And usually, this is good enough.

Exercises

E11. a) Show that if $(x, y) \in A$, then $x \in \bigcup\bigcup A$ and $y \in \bigcup\bigcup A$

b) Show that if f is a function, then dom f and ran f $\subseteq \bigcup\bigcup f$

c) Show that if $f : A \to B$, then $f \subseteq \mathcal{P}\mathcal{P}(A \cup B)$

E12. Let $\mathbb{R}$ denote the real numbers and $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 3x^2 - 2x + 1$ and $g(x) = |2x - 1|$. Find the range of $f \circ f$, $f \circ g$, $g \circ f$, and $g \circ g$.

E13. a) How many bijections exist from the set $X = \{1, 3, 5, ..., 99\}$ to the set $Y = \{2, 4, 6, ..., 100\}$?

b) How many 1-1 maps are there from $X = \{1, 3, 5..., 99\}$ into $Y = \{2, 4, 6, ..., 100\}$?

c) How many 1-1 maps are there from $X = \{1, 3, 5..., 99\}$ into $Y = \{1, 2, 3, ..., 100\}$?

E14. Define $f : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathcal{P}(X \times Y)$ by $f((A, B)) = A \times B$. Prove that f is onto if and only if one of the sets X or Y contains no more than one point.

E15. Let $C[a, b]$ be the set of all continuous functions $f : [a, b] \to \mathbb{R}$. Define a function $\Gamma : C[a, b] \to \mathbb{R}$ by $\Gamma(f) = f(\frac{a+b}{2})$. Is Γ one-to-one? onto?

E16. Suppose $p(t) = (r(t), s(t))$ is a one-to-one map from $\mathbb{R}$ into $\mathbb{R}^2$. Define a new map $q : \mathbb{R} \to \mathbb{R}^3$ by $q(t) = (r(t), r(s(t)), s(s(t)))$. Prove that q is one-to-one.

E17. Let $\mathcal{L}$ be the set of straight lines in $\mathbb{R}^2$ which do not pass through the origin $(0, 0)$. Describe geometrically a bijection $f : \mathcal{L} \to \mathbb{R}^2 - \{(0, 0)\}$.

E18. Let $f : X \to X$ and let $f^n : X \to X$ denote the result of composing f with itself n times. Suppose that for every $x \in X$, there exists an $n \in \mathbb{N}$ such that $f^n(x) = x$ (note that n may depend on x). Prove that f is a bijection.

E19. Let $C(\mathbb{R})$ denote the set of all continuous real-valued functions with domain $\mathbb{R}$, that is, $C(\mathbb{R}) = \{f : f \in \mathbb{R}^{\mathbb{R}}$ and f is continuous$\}$. Define a map $I : C(\mathbb{R}) \to C\mathbb{R})$ as follows:

for $f \in C(\mathbb{R})$, $I(f)$ is the function given by $I(f)(x) = \int_0^x f(t)\,dt$.

Is I one-to-one? onto? (*Hint: The Fundamental Theorem of Calculus is useful here.)*

E20. Let f be the bijection between the set of nonnegative integers and the set $\mathbb{Z}$ of all integers defined by

$$f(n) = \begin{cases} \frac{n}{2}\,, & \text{if } n \text{ is even} \\ -\frac{(n+1)}{2}\,, & \text{if } n \text{ is odd} \end{cases}$$

Now define a mapping $g : \mathbb{Z} \to \mathbb{Q}$ as follows:

> For any integer $m > 1$ we can factor m in a unique way into a product of primes $m = \prod_{i=1}^{k} p_i^{n_i}$, where $p_i < p_{i+1}$ and each n_i is a positive integer. For each $m > 1$, let
>
> $$g(m) = \prod_{i=1}^{k} p_i^{f(n_i)}$$
>
> Define $g(1) = 1$, $g(0) = 0$ and, for negative integers k, define $g(k) = -g(-k)$.

Prove that g is a bijection between $\mathbb{Z}$ and $\mathbb{Q}$.

E21. Let $a_1 = 0$ and for $n = 2, 3, \ldots$ define $a_n = \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}$. Show that the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(m, n) = a_{m+n-1} + n$ is a bijection.

E22. Let $\mathcal{B}$ be the collection of infinite subsets of $\mathbb{N}$. Define $f : \mathcal{B} \to (0, 1]$ as follows: for $B \in \mathcal{B}$, $f(B) =$ the " binary decimal" $0.x_1 x_2 x_3 \ldots x_n \ldots$ where $x_n = 1$ if $n \in B$ and $x_n = 0$ if $n \notin B$.
For example, $B = \{2, 4, 6, \ldots\} \in \mathcal{B}$ *and* $f(B) = 0.010101\ldots_{\text{base } 2}$

Prove or disprove that f is onto.

E23. In elementary measure theory, <u>measurable sets in $\mathbb{R}$</u> are defined. These lead to the Lebesgue integral – a kind of integration that is more general than the Riemann integration used in beginning calculus and analysis. Each measurable subset X is assigned a number $\mu(X)$ that "measures" the size of the set and satisfies certain rules. For example, $\mu([a, b]) = b - a$, and if X is a measurable subset of $[0, 1]$, then $0 \leq \mu(X) \leq 1$.

In fact, there must exist <u>nonmeasurable</u> sets – not every subset of $\mathbb{R}$ can be measurable – but proving this is nontrivial. However, the following argument <u>claims</u> to be a very simple proof that a nonmeasurable subset of $[0, 1]$ must exist. Find the error in the argument. (*Of course, the argument should look very suspicious because it seems like it doesn't matter that you don't know the definition of "measurable set."*)

> Assume that every subset X of $[0, 1]$ is measurable and has Lebesgue measure $\mu(X)$. Then $\mu(X) \in [0, 1]$ and the number $\mu(X)$ may or may not be in the set X. Let $B = \{\mu(X) : X \subseteq [0, 1] \text{ and } \mu(X) \notin X\}$. Since $B \subseteq [0, 1]$, our assumption tells us that B is measurable.
>
> But $\mu(B) \in B$ if and only if $\mu(B) \notin B$, which is impossible. Therefore not every $X \subseteq [0, 1]$ can be measurable. •

7. Infinite Sets

We can classify sets as either finite or infinite and, as we will see later, infinite sets can be further classified into different "sizes." We have already used the words "finite" and "infinite" informally, but now we want to make a more careful definition.

Definition 7.1 The sets A and B are equivalent, written $A \sim B$, if there exists a bijection $f\colon A \to B$.

Clearly, $A \sim A$; $A \sim B$ implies $B \sim A$; and if $A \sim B$ and $B \sim C$, then $A \sim C$.

Definition 7.2 The set A is called infinite if there is a one-to-one map $f : \mathbb{N} \to A$. A is called finite if A is not infinite.

If A is infinite, then $f[\mathbb{N}]$ is a subset of A equivalent to $\mathbb{N}$, so we could say that A is infinite iff A contains a "copy" of the set $\mathbb{N}$.

Since we have defined "finite" as "not infinite," we should prove some things about finite sets using that definition. For example:

A is finite iff $A = \emptyset$ or A is equivalent to $\{1, 2, ..., n\}$ for some $n \in \mathbb{N}$.
A subset of a finite set is finite.
A union of a finite number of finite sets is finite.
The image of a finite set under a map f is finite.
The power set of a finite set is finite.

These statements are easy to believe and the proofs are not really difficult but they are tedious induction arguments and will be omitted.

Examples 7.3

1) $A \times B$ is equivalent to $B \times A$. To see this, consider the function given by $f(a, b) = (b, a)$.

2) The mapping $f : \mathbb{N} \to \mathbb{E} = \{2, 4, 6, ...\}$ given by $f(n) = 2n$ is a bijection, so $\mathbb{N} \sim \mathbb{E}$. Thus, an infinite set may be equivalent to a proper subset of itself. (*This fact was noted by Galileo in the 17th century.*) In fact, the next theorem shows that this property actually characterizes infinite sets.

Theorem 7.4 A is infinite if and only if A is equivalent to a proper subset of itself.

Proof Suppose A is infinite. Then there is a one-to-one map $f : \mathbb{N} \to A$. Let $f(n) = a_n$. Break A into two pieces: $A = f[\mathbb{N}] \cup (A - f[\mathbb{N}])$. Then define

$$g : A \to (A - \{a_1\}) \text{ by } g(x) = \begin{cases} a_{n+1} & \text{if } x = a_n \in f[\mathbb{N}] \\ x & \text{if } x \in A - f[\mathbb{N}] \end{cases}$$

This g is a bijection between A and $A - \{a_1\}$, so A is equivalent to a proper subset of itself.

Conversely, suppose there is a bijection $f : A \to B$, where $B \subseteq A$ but $B \neq A$. Then $A - B \neq \emptyset$ so we can pick a point $a_1 \in A - B$. Starting with a_1, apply f repeatedly to get a sequence $a_1,\ f(a_1),\ f(f(a_1)) = f^2(a_1), ..., f^n(a_1), ...$.

All the terms $f^n(a_1)$ must be different. To see this, suppose that two of these points are equal, say $f^n(a_1) = f^{n+j}(a_1)$ for some n, j. Then (because f is one-to-one) we have $f^{n-1}(a_1) = f^{n+j-1}(a_1)$, $f^{n-2}(a_1) = f^{n+j-2}(a_1)$, ..., and so on until we get to $a_1 = f^j(a_1)$. But this is impossible because $f^j(a_1)$ is in B and a_1 is not. Therefore the map $g : \mathbb{N} \to A$ given by $g(n) = f^n(a_1)$ is one-to-one, so A is infinite. •

Definition 7.5 The set A is called countable if there exists a one-to-one map $f \colon A \to \mathbb{N}$.

Thus, a countable set is one which is equivalent to a subset of $\mathbb{N}$ (namely, the range of f). A countable set may be finite or infinite. (*In some books, the word "countable" is defined to mean "countable and infinite."*)

Examples 7.6

1) If $A \subseteq \mathbb{N}$, then A is countable (because $A \sim A \subseteq \mathbb{N}$).

2) Any countable set A is either finite or equivalent to $\mathbb{N}$.

Proof: Suppose $A \sim J \subseteq \mathbb{N}$, where J is infinite. Let j_1 be the least element of J, let j_2 be the least element in $J - \{j_1\}$, ... , and in general let $j_n =$ the least element in the set $J - \{j_1, \ldots, j_{n-1}\}$. Then the map $f : \mathbb{N} \to J$ given by $f(n) = j_n$ is a bijection, so $J \sim \mathbb{N}$. Therefore $A \sim \mathbb{N}$.

By (2), A is countable and infinite if and only if there is a bijection $g : \mathbb{N} \to A$. The function g is a one-to-one sequence whose range is A. Therefore a countable infinite set is one whose elements can be listed as a sequence: $a_1 = g(1),\ a_2 = g(2), \ldots, a_n = g(n)$, ... with no repeated terms.

3) The set $\mathbb{Q}^+ = \{x \in \mathbb{Q} : x > 0\}$ is countable. This is the first really surprising example: there is a bijection $g : \mathbb{N} \to \mathbb{Q}^+$ so $g(1) = q_1, g(2) = q_2, ..., g(n) = q_n, ...$ This means that we can list all the positive rationals in a sequence. It is not at all clear, at first, how this can done. In the usual order on $\mathbb{R}$, there is a third rational between any two rationals so we can't simply list the rationals, for example, in order of increasing size. But the definition of "countable" doesn't require that our list have any connection to size.

One way to create the list is to use the "diagonal argument" invented by Cantor. Begin by imagining all the positive rationals arranged into the following "infinite matrix" :

$$
\begin{array}{cccccccccccccccc}
\frac{1}{1} & \to & \frac{2}{1} & & \frac{3}{1} & \to & \frac{4}{1} & & \frac{5}{1} & & \frac{6}{1} & & \frac{7}{1} & & \frac{8}{1} & \cdots\cdots \\
 & \swarrow & & \nearrow & & \swarrow & & & & & & & & & & \\
\frac{1}{2} & & \frac{2}{2} & & \frac{3}{2} & & \frac{4}{2} & & \frac{5}{2} & & \frac{6}{2} & & \frac{7}{2} & & \frac{8}{2} & \cdots\cdots \\
\downarrow & \nearrow & & \swarrow & & & & & & & & & & & & \\
\frac{1}{3} & & \frac{2}{3} & & \frac{3}{3} & & \frac{4}{3} & & \frac{5}{3} & & \frac{6}{3} & & \frac{7}{3} & & \frac{7}{3} & \cdots\cdots \\
 & \swarrow & & & & & & & & & & & & & & \\
\frac{1}{4} & & \frac{2}{4} & & \frac{3}{4} & & \frac{4}{4} & & \frac{5}{4} & & \frac{6}{4} & & \frac{7}{4} & & \frac{8}{4} & \cdots\cdots \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots &
\end{array}
$$

Then create a bijection $g : \mathbb{N} \to \mathbb{Q}^+$ by moving back and forth along the diagonals (skipping over a rational from the matrix if it has been listed previously): the sequence begins $g(1) = \frac{1}{1}$, $g(2) = \frac{2}{1}$, $g(3) = \frac{1}{2}$, $g(4) = \frac{1}{3}$, $g(5) = \frac{3}{1}$, $g(6) = \frac{4}{1}$, $g(7) = \frac{3}{2}$, $g(8) = \frac{2}{3}$, $g(9) = \frac{1}{4}$, $g(10) = \frac{1}{5}$, ...

Of course, we can't picture the whole "infinite matrix" and we don't have a formula for $g(n) = ...$. However we have described a definite function using a computational procedure: you would have no trouble finding $g(15)$, and you could find $g(9999)$ with enough time and patience. The function g is clearly one-to-one and onto (for example, with some effort you could find the n for which $g(n) = \frac{379}{211}$).

Those who prefer formulas can consider the following alternate approach. Each natural number has a unique representation in base 11, using the numerals 0,1,2,3,...,9, / (*with the symbol* / *representing* 10). *For a positive rational* p/q *reduced to lowest terms, we may reinterpret the symbol string "*p/q*" as a natural number by thinking of it as a base 11 numeral. For example, 23/31 would be interpreted in base* 11 *as* $2(11)^4 + 3(11)^3 + 10(11)^2 + 3(11)^1 + 1(11)^0 = 34519$ *(in base 10). In this way, we define a one-to-one function* f *from* $\mathbb{Q}^+$ *into* $\mathbb{N}$. *For example,* $f(23/31) = 34519$. *So the set* $\mathbb{Q}^+$ *is countable.*

We can show that the set of negative rationals, $\mathbb{Q}^-$, is countable using a similar diagonal argument. Or, we can notice that the function $f : \mathbb{Q}^+ \to \mathbb{Q}^-$ given by $f(x) = -x$ is a bijection; since $\mathbb{Q}^-$ is equivalent to a countable set, $\mathbb{Q}^-$ is also countable.

Of course, introducing the term "countable" would be a waste of words if all sets were countable. Cantor also proved that uncountable infinite sets exist. This means that not all infinite sets are equivalent. Some are "bigger" than others!

Theorem 7.7 The interval of real numbers $(0, 1)$ is uncountable.

The heart of the proof is another "diagonal" argument. A technical detail in the proof depends on the following fact about decimal representations of real numbers:

Sometimes two different decimal expansions can represent the same real number. For example, 0.10000... = 0.09999... (why?). However, two different decimal expansions can represent the same real number only if one of the expansions ends in an infinite string of 0's and the other ends in an infinite string of 9's.

Proof Consider any function $f : \mathbb{N} \to (0, 1)$. We will show f cannot be onto and therefore no bijection exists between $\mathbb{N}$ and $(0, 1)$. To begin, write decimal expansions of all the numbers $f(1)$, $f(2)$, ...$f(n)$, ... in ran(f):

$$\begin{aligned} r_1 &= f(1) = 0.x_{11}x_{12}x_{13}\ldots x_{1n}\ldots \\ r_2 &= f(2) = 0.x_{21}x_{22}x_{23}\ldots x_{2n}\ldots \\ r_3 &= f(3) = 0.x_{31}x_{32}x_{33}\ldots x_{3n}\ldots \\ &\vdots \\ r_n &= f(n) = 0.x_{n1}x_{n2}x_{n3}\ldots x_{nn}\ldots \\ &\vdots \end{aligned}$$

where each x_{ij} is one of the digits $0, 1, ..., 9$. Now define a real number $y = 0.y_1y_2y_3 \ldots y_n \ldots$ by

$$\begin{cases} y_n = 1 & \text{if } x_{nn} \neq 1 \\ y_n = 2 & \text{if } x_{nn} = 1 \end{cases}$$

Then $y \in (0,1)$ and y is not equal to any of the numbers $r_n = f(n)$. To see that $y \neq r_n$, notice that

i) by construction, the decimal expression for y differs from r_n in the n^{th} decimal place; and
ii) y is not an alternate decimal representation of r_n because y does not end in an infinite string of 0's or 9's.

Therefore $y \notin \text{ran}(f)$, so f is not onto. •

Of course, we could start over, adding y to the original list. But then the same construction could be repeated. The list can never be complete, that is, $\text{ran}(f) = (0,1)$ is impossible.

A different way to try to dodge the technical difficulty about non-uniqueness of representations might be: whenever a number $f(n)$ has two different decimal representations, include both in the list and then define y. Is there any problem with that approach?

The next theorem tells us some important properties of countable sets.

Theorem 7.8 1) Any subset of a countable set is countable.

2) If A_n is countable for each $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n$ is countable.

3) If $A_1, A_2, ..., A_n$ are countable, then $A_1 \times A_2 \times ... \times A_n$ is countable.

Proof To show that a set is countable, we need to produce a one-to-one map g of the set into $\mathbb{N}$.

1) If A is countable, we have, by definition, a one-to-one map $f : A \to \mathbb{N}$. If $B \subseteq A$, then $g = f \mid B : B \to \mathbb{N}$ is also one-to-one, so B is also countable.

2) Consider the sets $B_1 = A_1$, $B_2 = A_2 - A_1$, ... , $B_n = A_n - (A_1 \cup ... \cup A_{n-1})$, ... By part 1), each B_n is countable, and for any $m \neq n$, $B_n \cap B_m = \emptyset$. It's easy to check that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$, so we need to prove that $\bigcup_{n=1}^{\infty} B_n$ is countable. All this amounts to saying that we will not lose any generality if we just assume, at the beginning of the proof, that the A_n's are disjoint from each other.

For each n, there is a one-to-one function $f_n : A_n \to \mathbb{N}$. Let p_n denote the n^{th} prime number and define $g : \bigcup_{n=1}^{\infty} A_n \to \mathbb{N}$ as follows:

if $x \in \bigcup_{n=1}^{\infty} A_n$, then x is in exactly one set A_n and using that n, let $g(x) = p_n^{f_n(x)}$.

Then g is well-defined because the A_n's are disjoint, and g is one-to-one (*why?*).

3) Let p_1, ... , p_n be the first n primes and, as before, let f_n be a one-to-one map from A_n into $\mathbb{N}$. Define $g : A_1 \times A_2 \times \ ... \ \times A_n \to \mathbb{N}$ by

$$g(a_1, ..., a_n) = p_1^{f_1(a_1)} \cdot p_2^{f_2(a_2)} \cdot \ . \ . \ . \ \cdot p_n^{f_n(a_n)}$$

The function g function is one-to-one (*why?*). •

Example 7.9

1) A union of a finite number of countable sets $A_1, \ldots, A_k$ is countable: just let $A_{k+1} = A_{k+2} = \ldots = \emptyset$ and then use part 2) of Theorem 7.8 to conclude that $\bigcup_{i=1}^{k} A_i = \bigcup_{i=1}^{\infty} A_i$ is countable.

We can combine this result with Theorem 7.8.2 to state: if Λ is a countable index set and A_λ is a countable set (for each $\lambda \in \Lambda$), then $\bigcup_{\lambda \in \Lambda} A_\lambda$ is countable.. Stated more informally: <u>a union of countably many countable sets is countable</u>.

2) The set $\mathbb{Q} = \mathbb{Q}^+ \cup \{0\} \cup \mathbb{Q}^-$ is countable because it is the union of three countable sets.

3) If $B \subseteq A$ and B is uncountable, then A is uncountable – because if A were countable, then its subset B would be countable by Theorem 7.8.1.

For example, the set $\mathbb{R}$ of real numbers is uncountable because its subset $(0,1)$ is uncountable. This means that you <u>cannot</u> index the real numbers using $\mathbb{N}$: you <u>cannot</u> sensibly write "Let $\mathbb{R} = \{r_1, r_2, r_3, \ldots\}$."

4) $\mathbb{R} = \mathbb{P} \cup \mathbb{Q}$, where $\mathbb{P}$ is the set of irrational numbers. Since $\mathbb{R}$ is uncountable and $\mathbb{Q}$ is countable, the set $\mathbb{P}$ must be uncountable. Thus there are "more" irrational numbers than rationals.

5) $\mathbb{R}^n$ $(n \geq 1)$ is uncountable, because it contains an uncountable subset: for example, the "x_1-axis" $\mathbb{R} \times \{0\} \times \ldots \times \{0\}$, a subset which is clearly equivalent to $\mathbb{R}$.

6) Suppose $A = \{a_1, a_2, \ldots, a_n, \ldots\} \subseteq \mathbb{R}$ and that $\epsilon > 0$. Let I_n be the open interval centered at a_n with length $\frac{\epsilon}{2^{n+1}}$, that is, $I_n = (a_n - \frac{\epsilon}{2^{n+2}}, a_n + \frac{\epsilon}{2^{n+2}})$. Then $A \subseteq \bigcup_{n=1}^{\infty} I_n$, and the total length of all the intervals I_n is $\sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+1}} = \frac{\epsilon}{2} < \epsilon$. This shows that <u>any countable subset of $\mathbb{R}$ can be "covered" by a sequence of open intervals whose total length is arbitrarily small</u>. (If this does not seem surprising, suppose $A = \mathbb{Q}$; and if you think that $\bigcup_{n=1}^{\infty} I_n$ must be $\mathbb{R}$, try to prove that $\sqrt{2}$ must be in $\bigcup_{n=1}^{\infty} I_n$.)

More informally: take a piece of string with length $\frac{\epsilon}{2}$. Cut it in half and lay one half on $\mathbb{R}$ with its center at a_1. Cut the remaining piece in half again and lay one half on $\mathbb{R}$ with center at a_2. Continue in this way: cut the remaining piece of string in half and lay it down on $\mathbb{R}$ so that its center is on top of the next element in the list of elements of A. (Of course, some of the pieces of string, when laid down, may overlap with earlier pieces.) In the end, all of A is covered with pieces of the original string.

Exercises

E24. Prove or give a counterexample:

a) if $A \sim B$ and $C \sim D$, then $A \cap C \sim B \cap D$.
b) if $A \sim B$ and $C \sim D$, then $A \cup C \sim B \cup D$.
c) if $A \sim B$, then $A - B \sim B - A$.
d) if $A - B \sim B - A$, then $A \sim B$.
e) if A, B, and C are nonempty sets and $A \times B \sim A \times C$, then $B \sim C$.
f) if A is infinite and B is countable, then $A \cup B \sim B$.
g) if A is infinite and B is countable, then $A \cup B \sim A$.

E25. Prove that if A is uncountable and B is countable, then $A \sim A - B$.

E26. a) Give an explicit formula for a bijection between the intervals $(-\infty, 7)$ and $(0, \infty)$.
b) Give an explicit formula for a bijection between the sets $[0,1]$ and $(0,1) \cup \{2\}$.

E27. Prove that the set of all real numbers in the interval $(0,1)$ that have a decimal expansion using only even digits is uncountable.

E28. Is the following statement true? If not, what additional assumptions about A and B will make it true?

$A \sim B$ iff there is a set $f = \{(a,b) : a \in A, b \in B\}$ such that each element of A and each element of B occur in exactly one pair (a,b).

E29. A subset B of X is called <u>cocountable</u> if $X - B$ is countable and <u>cofinite</u> if $X - B$ is finite. (*The names are shorthand:* <u>cocountable</u> $=$ <u>co</u>*mplement is* <u>countable</u>.)

a) Prove that if B and C are cocountable subsets of X, then $B \cup C$ and $B \cap C$ are also cocountable. Is the analogous result true for cofinite sets? For infinite unions and intersections?
b) Show how to write the set of irrational numbers, $\mathbb{P}$, as an intersection of countably many cofinite subsets of $\mathbb{R}$.

E30. A collection $\mathcal{A}$ of sets is called <u>pairwise disjoint</u> if whenever $A, B \in \mathcal{A}$ and $A \neq B$, then $A \cap B = \emptyset$. For each statement, provide a proof or a counterexample:

a) If $\mathcal{A}$ is a collection of pairwise disjoint circles in the plane, then $\mathcal{A}$ is countable.
b) If $\mathcal{A}$ is a collection of pairwise disjoint circular disks in the plane, then $\mathcal{A}$ is countable.

E31. Prove that there cannot exist an uncountable collection of pairwise disjoint open intervals in $\mathbb{R}$.

E32. Let ℓ be a straight line in the plane. Prove that the set $\ell \cap (\mathbb{Q} \times \mathbb{Q})$ is either empty, contains exactly one point, or is countably infinite. In each case, give an equation for a specific line ℓ to illustrate.

E33. Suppose that $f : [0,1] \to \mathbb{R}$ has the following property:

> there is a fixed constant, *M,* such that for every finite set $\{x_1, x_2, \ldots, x_n\} \subseteq [0,1]$, the following inequality is true:
>
> $$|\ f(x_1) + \ldots + \ f(x_n)\ | < M$$

a) Give an example of such a function *f* where ran (f) is infinite.
b) Prove that for such a function f, $\{x \in [0,1] : f(x) \neq 0\}$ is countable.

E34. What is wrong with the following argument?

For each irrational number $p \in \mathbb{P}$, pick an open interval (a_p, b_p) with rational endpoints and centered at p. Since the interval (a_p, b_p) is centered at p, the function $\Phi : \mathbb{P} \to \mathbb{Q} \times \mathbb{Q}$ given by $\Phi(p) = (a_p, b_p)$ is one-to one. There are only countably many possible pairs (a_p, b_p) because $\mathbb{Q} \times \mathbb{Q}$ is countable. Therefore $\mathbb{P}$ is equivalent to a subset of $\mathbb{Q} \times \mathbb{Q}$, so $\mathbb{P}$ is countable.

E35. a) A sequence s in $\mathbb{N}$ is called an arithmetic progression if $\exists d \in \mathbb{N}$ such that $s_{n+1} = s_n + d$ for every $n \in \mathbb{N}$. Prove that the set of arithmetic progressions in $\mathbb{N}$ is countable.

b) A sequence s in $\mathbb{N}$ is called eventually constant if $\exists k, l \in \mathbb{N}$ such that $s_n = l$ for all $n \geq k$. Prove that the set of eventually constant sequences in $\mathbb{N}$ is countable.

c) A sequence s in $\mathbb{N}$ is called eventually periodic if $\exists k, p \in \mathbb{N}$ such that $s_n = s_{n+p}$ for all $n \geq k$. Prove that the set of all eventually periodic sequences in $\mathbb{N}$ is countable. *Note: the set $\mathbb{N}^{\mathbb{N}}$ of all sequences in $\mathbb{N}$ is uncountable, but we have not proved that yet.*

E36. For a set A, let $\mathcal{W}(A) = \{f \in A^{\{0,1,\ldots,n\}} : n \in \mathbb{N}\}$. (*We can think of A as an "alphabet" and $\mathcal{W}(A)$ as the set of all "finite sequences" or "words" that can be formed from this alphabet.*)

Prove that if A is countable, then $\mathcal{W}(A)$ is countable.

E37. Let A be an uncountable subset of $\mathbb{R}$. Prove that there is a sequence of distinct elements $a_1, a_2, \ldots a_n, \ldots$ in A such that $\sum_{n=1}^{\infty} a_n$ diverges.

E38. Let Y be a set and, for each $n \in \mathbb{N}$, suppose $f_n : \mathbb{R} \to Y$. Let g be a function $g : \mathbb{R} \to Y$ such that, for every $n \in \mathbb{N}$, the set $\{x \in \mathbb{R} : g(x) = f_n(x)\}$ is countable. Prove that there is a point $x_0 \in \mathbb{R}$ such that for all n, $g(x_0) \neq f_n(x_0)$.

What property of $\mathbb{R}$ makes the proof work?

E39. Suppose $S \subseteq \mathbb{R}$ and $t \in \mathbb{R}$. Let $S + t$ denote the "translated set" $\{s + t : s \in S\}$.

a) Show that if S is countable, then $\exists t \in \mathbb{R}$ such that $S + t \subseteq \mathbb{P}$.

b) For sequences s and t in $\mathbb{R}$ (that is, for $s, t \in \mathbb{R}^{\mathbb{N}}$), let $s + t \in \mathbb{R}^{\mathbb{N}}$ be the sequence defined by $(s + t)(n) = s(n) + t(n)$. Show that if $\mathcal{S}$ is a countable subset of $\mathbb{R}^{\mathbb{N}}$, then $\exists\ t \in \mathbb{R}^{\mathbb{N}}$ such that $s + t \in \mathbb{P}^{\mathbb{N}}$ for every $s \in \mathcal{S}$.

E40. Give a proof or a counterexample for the following statement:

> If $\mathcal{O} = \{O_\lambda : \lambda \in \Lambda\}$ is a collection of open intervals in $\mathbb{R}$ with the property that $\mathbb{Q} \subseteq \bigcup \mathcal{O}$, then $\bigcup \mathcal{O} = \mathbb{R}$.

E41. a) Show how to write $\mathbb{N}$ as the union of infinitely many pairwise disjoint infinite sets.

b) Show how to write $\mathbb{N}$ as the union of uncountably many sets with the property that, given any two of them, one is a subset of the other.

c) Show how to write $\mathbb{N}$ as the union of uncountably many sets with the property that any two of them have finite intersection. (*Such sets are called almost disjoint.*)

(*Hint: These statements that actually are true if* $\mathbb{N}$ *is replaced by any infinite countable set. For parts b) and c), you may find it easier to solve the problem first for* $\mathbb{Q}$*, and then use a bijection to "convert" your solution for* $\mathbb{Q}$ *into a solution for the set* $\mathbb{N}$.)

E42. a) Imagine an infinite rubber stamp which, when applied to the plane, inks over all concentric circles with irrational radii around its center. What is the minimum number of stampings necessary to ink over the whole plane?

b) What if the stamp, instead, inks over all concentric circles with rational radii around its center?

E43. Suppose $f : \mathbb{R} \to \mathbb{R}$. Let $A = \{a \in \mathbb{R} : \lim_{x \to a} f(x)$ exists but is not equal to $f(a)\}$. Prove that A is countable.

(*Hint: One way to start is to define, for* $r \in \mathbb{Q}$, $A_r^- = \{a \in A : f(a) < r < \lim_{x \to a} f(x)\}$ *and* $A_r^+ = \{a \in A : \lim_{x \to a} f(x) < r < f(a)\}$. *Then* $A = \bigcup_{r \in \mathbb{Q}} (A_r^- \cup A_r^+)$ *(why?). Prove that* A_r^- *and* A_r^+ *are countable.*)

E44. Let D be a countable set of points in the plane, $\mathbb{R}^2$. Prove there exist sets A and B such that $D = A \cup B$, where the set A has finite intersection with every horizontal line in the plane and B has finite intersection with every vertical line in the plane.

Notes: 1) This problem is fairly hard. You might get an idea by starting the with an easy special case: $D = \mathbb{N} \times \mathbb{N}$

2) The statement that "every $D \subseteq \mathbb{R}^2$ can be written as $D = A \cup B$, where A has countable intersection with every horizontal line and B has countable intersection with every vertical line" is actually equivalent to the continuum hypothesis (*see p. 40 and Exercise VIII.E.26*).

E45. We say that a function $f : \mathbb{R} \to \mathbb{R}$ has a local maximum at x if there exists an open interval (a, b) containing x such that $f(x) \geq f(y)$ for all $y \in (a, b)$ – in other words, $f(x)$ is the (absolute) maximum value of f on the interval (a, b).

a) Give an example of a nonconstant f which has a local maximum at every point. Then modify your example, if necessary, to get an example where $\mathrm{ran}(f)$ is infinite.

Is it possible that $\mathrm{ran}(f) =$ the set of positive rational numbers? the set of all rational numbers? Is it possible, for every countable $S \subseteq \mathbb{R}$, to find an f having a local maximum at every point $x \in \mathbb{R}$ and $\mathrm{ran}(f) = S$?

b) Suppose f has a local maximum at every point $x \in \mathbb{R}$. Prove that $\mathrm{ran}(f)$ is countable.

(*Hint: if $f(x) = y \in$ ran(f), pick an interval (a, b) with rational endpoints containing x and such that y is the maximum of f on (a, b).*)

8. Two Mathematical Applications

The relatively simple facts that we know about countable and uncountable sets are enough to prove an interesting fact about the real numbers.

Definition 8.1 A real number r is called <u>algebraic</u> if r is a root to a nonconstant polynomial equation

$$a_n x^n + a_{n-1}x^{n-1} + \ \dots\ + a_1 x + a_0 = 0 \quad (*)$$

where the coefficients $a_0, \ \dots\ , a_n$ are all integers. (*It is clearly equivalent to require that the coefficients be rational numbers – because we could multiply both sides by some integer to "clear the fractions" and arrive at a polynomial equation that has integer coefficients and exactly the same roots.)*

(*<u>Complex roots</u> of (*) are called <u>complex algebraic numbers</u>.)*

A real number is rational if and only if it is the root of a <u>first degree</u> polynomial equation with integer coefficients: the rational $\frac{p}{q}$ is a root of the equation $qx - p = 0$. Therefore rational numbers are algebraic. The algebraic numbers are a natural generalization to include certain irrationals: for example, $\sqrt{2}$ is algebraic, since it is a root of the quadratic equation $x^2 - 1 = 0$; and $-1 + 2\sqrt{2}$ is algebraic because it is a root of $x^3 + x^2 - 9x + 7 = 0$. A reasonable question would be: are there any nonalgebraic real numbers?

Theorem 8.2 The set $\mathbb{A}$ of all algebraic real numbers is countable.

Proof For each $n \geq 1$, let $\mathcal{P}_n$ be the set of all polynomials with integer coefficients and degree $\leq n$.

Suppose $P \in \mathcal{P}_n$. Let $(a_n, a_{n-1}, \ \dots\ , a_0)$ be the $(n+1)$-tuple of coefficients of P. Then $(a_n, a_{n-1}, \ \dots\ , a_0)$ is in the set $\mathbb{Z}^{n+1}$, which is countable. Since the function $f : \mathcal{P}_n \rightarrow \mathbb{Z}^{n+1}$ given by $f(P) = (a_n, a_{n-1}, \ \dots\ , a_0)$ is a bijection, $\mathcal{P}_n$ is countable.

For each polynomial P, let $Z(P)$ be the set of real roots of the equation $P(x) = 0$. $Z(P)$ is a finite set. Then $R_n = \bigcup\{Z(P) : P \in \mathcal{P}_n\}$ is a union of countably many finite sets (each with at most n elements), so R_n is countable. (*For any particular $n_0 \in \mathbb{N}$, R_{n_0} is the set of all algebraic numbers that are roots of an equation (*) that has degree $\leq n_0$; for example, $R_1 = \mathbb{Q}$.*)

By Theorem 7.8(2), $\mathbb{A} = \bigcup_{n=1}^{\infty} R_n$ is countable. •

Definition 8.3 A real number which is not algebraic is called <u>transcendental</u>.
(*Euler called these numbers "transcendental" because they "transcend the power of algebraic methods." To be more politically correct, we might call these numbers "polynomially challenged."*)

Corollary 8.4 Transcendental numbers exist.
Proof Let $\mathbb{T}$ be the set of transcendental numbers. Since $\mathbb{R} = \mathbb{A} \cup \mathbb{T}$ and $\mathbb{A}$ is countable, $\mathbb{T}$ cannot be empty. •

In fact, this short proof shows much more: not only is $\mathbb{T}$ nonempty, but $\mathbb{T}$ must be uncountable! There are "more" transcendental numbers than algebraic numbers. This is an example of a "pure existence proof" – meaning that the proof does not give us any particular transcendental numbers, nor does it give us a way to construct one. To do that is harder. Transcendental numbers were first shown to exist (using different methods) by Liouville in 1844. The numbers e (Hermite, 1873) and π (Lindemann, 1882) are transcendental. One method for producing transcendental numbers is contained in a theorem of Gelfand (1934) from algebraic number theory; it states that

> ff α is an algebraic number, $\alpha \neq 0, 1$, and β is an algebraic irrational, then α^{β} is transcendental.

For example, the theorem implies that $\sqrt{2}^{\sqrt{2}}$ is transcendental. The number e^{π} is also transcendental. This follows from Gelfand's Theorem (which also allows complex algebraic numbers) because:

> $e^{\pi} = e^{-i^2\pi} = (e^{i\pi})^{-i}$ and $e^{i\pi} = \cos\pi + i\sin\pi = -1$. So $e^{\pi} = (-1)^{-i}$, which is transcendental by Gelfand's Theorem.

As a second application to a different part of mathematics, we will prove a simple theorem from analysis.

Theorem 8.5 A monotone function $f : [a, b] \to \mathbb{R}$ has at most countably many points of discontinuity. (S*ince* $[a, b]$ *is uncountable, this implies that a monotone function on* $[a, b]$ *must be continuous at "most" points*.)

Proof Assume $x \leq y$ implies $f(x) \leq f(y)$. (*If* f *is decreasing, simply apply the following argument to the increasing function* $-f$.) Therefore, at each point $c \in (a, b)$ we have

$$\lim_{x \to c^-} f(x) \leq f(c) \leq \lim_{x \to c^+} f(x) \quad (\textit{why must these one-sided limits exist ?})$$

Let $j(c)$ denote the "jump of f at c" $= \lim_{x \to c^+} f(x) - \lim_{x \to c^-} f(x)$. Since f is increasing, $j(c) \geq 0$, and f is discontinuous at c if and only if $j(c) > 0$.

Let $A_n = \{c \in (a, b) : j(c) > \frac{1}{n}\}$. The set A_n is finite because the sum of any set of jumps cannot be more than $f(b) - f(a)$. Furthermore, if $j(c) > 0$, then $j(c) > \frac{1}{n}$ for some sufficiently large n, so $c \in A_n$ for some n. Therefore every point c of discontinuity of f in (a, b) must be in the countable set $\bigcup_{n=1}^{\infty} A_n$. (The function might also be discontinuous at an endpoint a or b, but the set of discontinuities would still be countable.) •

Corollary 8.6 A monotone function $f : \mathbb{R} \to \mathbb{R}$ has at most countably many points of discontinuity.

Proof For every $k \in \mathbb{Z}$, the set D_k of discontinuities of f in the interval $[k, k+1]$ is countable, by Theorem 8.5. So $\bigcup_{k \in \mathbb{Z}} D_k$, the set of all discontinuities, is countable. •

As a sort of "converse," it is possible to prove that if A is any countable subset of $\mathbb{R}$, then there exists a monotone function $f : \mathbb{R} \to \mathbb{R}$ for which A is precisely the set of points of discontinuity. (You can find an argument in *A Primer of Real Functions* (Boas, p. 129), or see Exercise E65 below.)

9. More About Equivalent Sets

So far, we have seen two kinds of infinite set, countable and uncountable. In this section, we explore uncountable sets in more detail. Notice that we have no reason to assume that two uncountable sets must be equivalent.

There are many equivalent uncountable subsets of $\mathbb{R}$. The following theorem gives some examples. (*The proof could be made much easier by using the Cantor-Schroeder-Bernstein Theorem, which we will discuss on p. 42. However, at this stage, it is instructive to give a direct proof.*)

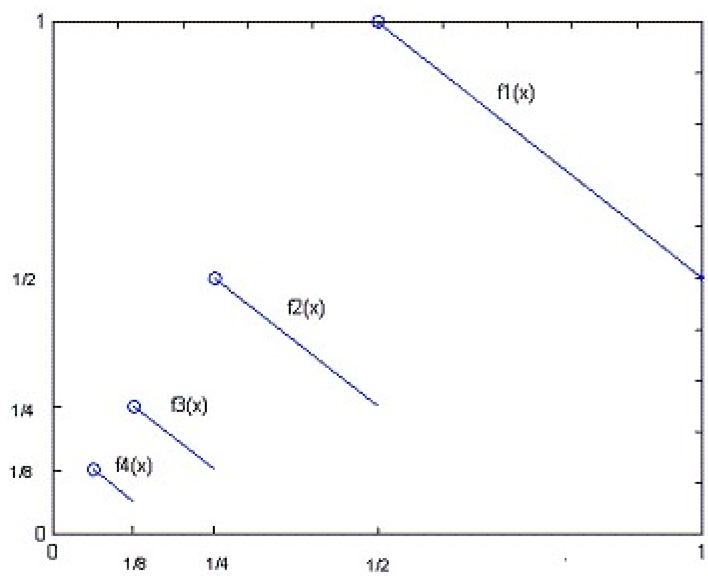
Theorem 9.1 Suppose $a < b$. The following intervals in $\mathbb{R}$ are all equivalent:

$$\begin{array}{lllll} \mathbb{R} = (-\infty, \infty) & \sim (a,b) & \sim (0,1) & \sim (a,b] & \sim (0,1] \\ \quad \sim [a,b) & \sim [0,1) & \sim [a,b] & \sim [0,1] & \sim [0,\infty) \\ \quad \sim [a,\infty) & \sim (0,\infty) & \sim (a,\infty) & \sim (-\infty,0] & \sim (-\infty,a] \\ \quad \sim (-\infty,0) \ \sim (-\infty,a). & & & & \end{array}$$

Theorem 9.1 says that all <u>infinite</u> intervals in $\mathbb{R}$ are equivalent. (*For technical reasons that we will discuss later, it is convenient to consider $\emptyset$ and all one-point sets $\{a\}$ as intervals. It turns out that these are the only finite intervals.*)

Proof The linear map $f(x) = (b-a)x + a$ can be used to show that
$(0,1) \sim (a,b)$ $(0,1] \sim (a,b]$ $\quad [0,1) \sim [a,b)$ $\quad [0,1] \sim [a,b]$.
The bijection $\tan : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ proves that $\mathbb{R} \sim \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.
To show that $(0,1] \sim (0,1)$, define functions:

$$f_n(x) = \frac{3}{2^n} - x \text{ for } \frac{1}{2^n} < x \leq \frac{1}{2^{n-1}} \text{ for each } n \in \mathbb{N}$$



Each f_n is a bijection from $\left(\frac{1}{2^n}, \frac{1}{2^{n-1}}\right]$ to $\left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)$, as illustrated above. Let $f = \bigcup_{n=1}^{\infty} f_n$.
(*The graph of f consists of all the separate pieces, shown above, taken together.*)

Then f is a function (*why?*) with $\mathrm{dom}(f) = \bigcup_{n=1}^{\infty} \mathrm{dom}(f_n) = (0,1]$ and $\mathrm{ran}(f) = \bigcup_{n=1}^{\infty} \mathrm{ran}(f_n)$ $= (0,1)$. It is easy to check that f is a bijection. (*Of course, f is not continuous, but that is not required in the definition of set equivalence.*)

We can <u>extend</u> the definition of f to prove that $[0,1] \sim [0,1)$. Specifically, use f to define a function $g : [0,1] \to [0,1)$ by

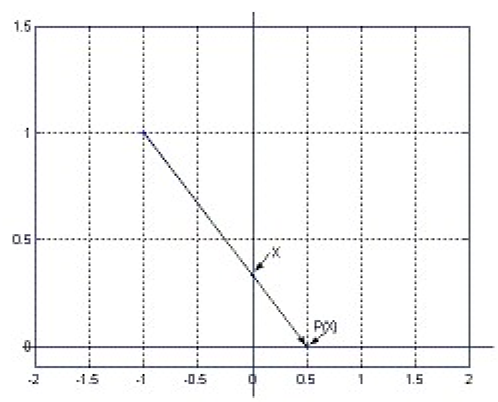
$$g(x) = \begin{cases} f(x) & \text{if } x \in (0,1] \\ 0 & \text{if } x = 0 \end{cases}$$

Since f is a bijection, so is g.

It is very simple to show that $[0,1) \sim (0,1]$, that $(-\infty,0] \sim [0,\infty)$, and that $(-\infty,0) \sim (0,\infty)$: just use the maps $h(x) = 1 - x$ and $k(x) = -x$.

The function $\ln : (0,\infty) \to \mathbb{R}$ proves that $(0,\infty) \sim \mathbb{R}$.

We can use a "projection" to show that $[0,1)$ is equivalent to $[0,\infty)$. Imagine a ray of light that emanates from $(-1,1)$, passes through the point x on the y-axis ($0 \le x < 1$) and then hits the x-axis at a point that we call $p(x)$. (*See the following illustration*). Then $p : [0,1) \to [0,\infty)$ is a bijection.



Those who prefer formulas can check that a formula for $p : [0,1) \to [0,\infty)$ is $p(x) = \frac{x}{1-x}$.

The few remaining equivalences such as $(a,\infty) \sim (0,\infty)$ are left for the reader to prove. ●

At this point, we might ask "<u>Are all uncountable sets are equivalen</u>t?"

Example 9.2 The set $\mathbb{R}^{\mathbb{R}}$ is not equivalent to $\mathbb{R}$. To see this, we use an argument whose flavor is similar to the "diagonal argument" used by Cantor to prove that $(0,1)$ is uncountable.

> Consider <u>any</u> function $\phi : \mathbb{R} \to \mathbb{R}^{\mathbb{R}}$. We will show that ϕ <u>cannot</u> be onto, and therefore no bijection can exist between $\mathbb{R}$ and $\mathbb{R}^{\mathbb{R}}$. If $x \in \mathbb{R}$, then $\phi(x) \in \mathbb{R}^{\mathbb{R}}$. Note that $\phi(x)$ is not a number: $\phi(x)$ is a <u>function</u> from $\mathbb{R}$ into $\mathbb{R}$; to emphasize this, we will temporarily write

$\phi(x) = \phi_x$. Then for $y \in \mathbb{R}$, it makes sense to evaluate the function ϕ_x at a point $y \in \mathbb{R}$ to get a number $\phi_x(y) \in \mathbb{R}$.

Now we define a function $f \in \mathbb{R}^{\mathbb{R}}$ by the formula

$$f(x) = \begin{cases} 1 & \text{if } \phi_x(x) = 0 \\ 0 & \text{if } \phi_x(x) \neq 0 \end{cases}$$

We claim that for each $x \in \mathbb{R}$, $\phi(x) = \phi_x \neq f$. This is because the two functions ϕ_x and f have different values at the point $x : \phi_x(x) \neq f(x)$ because of the way f was defined. Therefore $f \notin$ ran(ϕ), so ϕ is not onto.

It is also clear, then, that no subset of $\mathbb{R}$ can be equivalent to $\mathbb{R}^{\mathbb{R}}$. To see this, suppose $A \subseteq \mathbb{R}$ and that $\psi : A \to \mathbb{R}^{\mathbb{R}}$. If ψ were onto, we could use ψ to create an onto map ϕ from $\mathbb{R}$ to $\mathbb{R}^{\mathbb{R}}$ (which is impossible!) as follows: pick any function $f \in \mathbb{R}^{\mathbb{R}}$ and define

$$\phi(x) = \begin{cases} \psi(x) & \text{if } x \in A \\ f & \text{if } x \in \mathbb{R} - A. \end{cases}$$

Therefore $\mathbb{R}^{\mathbb{R}}$ is an uncountable set which is not equivalent to any subset of $\mathbb{R}$. In particular, $\mathbb{R}^{\mathbb{R}}$ is not equivalent to $\mathbb{N}$. Intuitively, $\mathbb{R}^{\mathbb{R}}$ is "bigger" than any subset of $\mathbb{R}$.

The fact that not all uncountable sets are equivalent leads to an interesting question. We have already seen (*Example 7.6(2)*) that every subset of $\mathbb{N}$ is either finite or equivalent to the whole set $\mathbb{N}$— there are no "intermediate-sized" subsets of $\mathbb{N}$. How about for $\mathbb{R}$? Could there be a subset of $\mathbb{R}$ which is uncountable but still not equivalent to $\mathbb{R}$? Intuitively, such an infinite set would be "bigger than $\mathbb{N}$ but smaller than $\mathbb{R}$." It turns out that this question is "undecidable" !

An old word for the real number line is "continuum," and the continuum hypothesis (or CH, for short) is a conjecture of Cantor that dates from about 1878. CH says that the answer to the question is "no"— that is, CH states that every subset of $\mathbb{R}$ is either countable or equivalent to $\mathbb{R}$. Cantor was unable to prove his conjecture. In an address to the International Congress of Mathematicians in 1900, David Hilbert presented a list of research problems he considered most important for mathematicians to solve in the new century and the very the first problem on his list was CH.

Now if one believes (philosophically) that the set of real numbers actually exists "somewhere out there" (in some Platonic sense, or, say, in the mind of God) then any proposition about these real numbers, such as CH, must be either true or false – we simply have to figure out which. But to prove such a proposition mathematically requires an argument based on some set of assumptions (axioms) and theorems in terms of which we have made our mathematical definition of $\mathbb{R}$. Since $\mathbb{R}$ is defined mathematically in terms of set theory, a proof of CH (or of its negation) needs to be a proof derived from the axioms of set theory.

The usual collection of axioms for set theory – we have seen some of these axioms – is called the Zermelo-Fraenkel system (ZF). If the Axiom of Choice is added for good measure, the axiom system is referred to as ZFC. The fact is that the standard axioms for set theory (ZFC) are not sufficient to prove either CH or its negation. Since CH cannot be proven from ZFC, we say CH is independent of ZFC. But, since the denial of CH also cannot be proven from these axioms, we say that CH is also consistent with ZFC. Taken together, these statements say that CH is undecidable in ZFC: one can add either CH or $\sim$ CH as an additional axiom to ZFC without fear of introducing a contradiction. The consistency of CH was proved by Kurt Gödel (1906-1978) in 1939, and in 1963 Paul J. Cohen (1934-2007) showed that ZFC could not prove CH $-$ the independence of CH.

> *"In the early 1960's, a brash, young and extremely brilliant Fourier analyst named Paul J. Cohen (people who knew him in high school assure me he was always brash and brilliant) chatted with a group of colleagues at Stanford about whether he would become more famous by solving a certain Hilbert problem or by proving that CH is independent of ZFC. This (informal) committee decided that the latter problem was the ticket. (To be fair, Cohen had been interested in logic and recursive functions for several years; he may have conducted this séance just for fun.) Cohen went off and learned the necessary logic and, in less than a year, had proved the independence. This is certainly one of the most amazing intellectual achievements of the twentieth century, and Cohen was awarded the Fields Medal for the work. But there is more.*
>
> *Proof in hand, Cohen flew off to Princeton to the Institute for Advanced Study to have his result checked by Kurt Gödel. Gödel was naturally skeptical, as Cohen was not the first person to claim to have solved the problem; and Cohen was not even a logician! Gödel was also at this time beginning his phobic period. (Toward the end of his life, Gödel became convinced that he was being poisoned, and he ended up starving himself to death.) When Cohen went to Gödel's home and knocked on the door, it was opened six inches and a hoary hand snatched the manuscript and slammed the door. Perplexed, Cohen departed. However, two days later, Cohen received an invitation for tea at Gödel's home. His proof was correct: the master had certified it."*
>
> *Mathematical Anecdotes*,
> Steven G. Krantz, Mathematical Intelligencer, v. 12, No. 4,
> 1990, pp. 35-36)

By a curious coincidence, Cohen was an analyst interested in Fourier series, and it was Cantor's work on Fourier series that led to his creation of set theory in the first place.

Therefore CH has a status, with respect to ZFC, like that of the parallel postulate in Euclidean geometry: the other axioms are not sufficient to prove or disprove it. To the other axioms for Euclidean geometry, you can either add the parallel postulate (to get Euclidean geometry) or, instead, add an axiom which denies the parallel postulate (to get some kind of non-Euclidean geometry). Likewise, to ZFC, one may consistently add CH as additional axiom, or get a "different set theory" by adding an axiom which denies CH.

Unlike the situation with AC, most mathematicians prefer to avoid assuming CH or its negation in a proof whenever possible. The reason is that CH (in contrast to AC) does not seem to command intuitive belief and, moreover, no central mathematical results depend on CH. When the use of CH use seems necessary, it is customary to call attention to the fact that it is being used.

To reiterate one last point: if you believe that, in some sense, the real numbers actually exist "out there" beyond ourselves, then you believe that the set $\mathbb{R}$, as defined mathematically within ZFC, is just a mathematical model of the "real" real number system. This model may be an inaccurate or incomplete fit to reality. You may therefore continue to believe that for the "real" real numbers, CH is either true or false as a matter of fact. People with this point of view would say that the undecidability of CH within ZFC simply reflects the inadequacy of the axiom system ZFC. Gödel himself seemed to feel that ZFC is inadequate, although for perhaps different reasons:

> *"I believe that ... one has good reason for suspecting that the role of the continuum problem in set theory will be to lead to the discovery of new axioms which will make it possible to disprove Cantor's conjecture."*
>
> Kurt Gödel, "What is Cantor's Continuum Problem?", *Philosophy of Mathematics*, ed. Benacerraf & Putnam, Prentice-Hall, 1964, p. 268

10. The Cantor-Schroeder-Bernstein Theorem

The Cantor-Schroeder-Bernstein Theorem (CSB for short) gives a way to prove that two sets are equivalent without the work of actually constructing a bijection between them. It states that two sets are equivalent if each is equivalent to a subset of the other.

Theorem 10.2 (CSB) Suppose there exist one-to-one functions $f : A \to B$ and $g : B \to A$. Then $A \sim B$.

Proof We divide the set A into three subsets in the following way. For a point $x \in A$, we say "x has ≥ 0 ancestors" if $g^{-1}(x) \subseteq B$ (*which is, of course, always true, so every x has ≥ 0 ancestors*). We say "x has $\geq$ 1 ancestor" if $g^{-1}(x) \neq \emptyset$, that "x has ≥ 2 ancestors" if $f^{-1}(g^{-1}(x)) \neq \emptyset$, and so on. In general, "x has $\geq n$ ancestors" if the inverse image set resulting from the alternating application of first g^{-1}, then f^{-1}, then g^{-1}, ..., n times produces a nonempty set. We say "x has exactly n ancestors" if x has $\geq$ n ancestors but x does not have $\geq$ $n+1$ ancestors. We say "x has infinitely many ancestors" if x has $\geq n$ ancestors for every $n \in \mathbb{N}$. Let

$$\begin{aligned} A_E &= \{x \in A : x \text{ has an even number of ancestors}\}, \\ A_O &= \{x \in A : x \text{ has an odd number of ancestors}\}, \quad \text{and} \\ A_I &= \{x \in A : x \text{ has an infinite number of ancestors}\} \end{aligned}$$

Clearly, these sets are disjoint and $A = A_E \cup A_O \cup A_I$. Define ancestors for a point $y \in B$ in a similar way, beginning the definition with an application of f^{-1} rather than g^{-1} and divide B into the sets B_E, B_O, and B_I. The maps $f \mid A_E : A_E \to B_O$ and $f \mid A_I : A_I \to B_I$ are bijections (why?).

The function $f \mid A_O$ maps A_O into B_E, but it may not be onto (why?). However, the map $g \mid B_E : B_E \to A_O$ is a bijection, so it has an inverse bijection $(g \mid B_E)^{-1} : A_O \to B_E$. We can now define a bijection h from A to B by piecing these maps together:

$$h = (f \mid A_E) \cup (f \mid A_I) \cup (g \mid B_E)^{-1}$$

More explicitly,

$$h(x) = \begin{cases} f(x) & \text{if } x \in A_E \cup A_I \\ (g|B_E)^{-1}(x) & \text{if } x \in A_O \end{cases}$$ •

Our previous examples of equivalent sets were exactly that: merely examples, and they were not used in proving the CSB Theorem. Therefore, without circular reasoning, we can now revisit some earlier examples and use the CSB Theorem to give simpler proofs of many of those equivalences. The following two examples illustrate this.

Examples 10.3

1) Suppose $A \subseteq \mathbb{N}$. Then $i : A \to \mathbb{N}$ by $i(x) = x$. If A is not finite, then the definition of "infinite" says that there is also a one-to-one map $f : \mathbb{N} \to A$. In that case, by CSB, we conclude that $A \sim \mathbb{N}$. So A is either finite or equivalent to $\mathbb{N}$.

2) $(0,1) \sim [0,1]$ since each set is equivalent to a subset of the other:

$(0,1) \sim (0,1) \subseteq [0,1]$ and
$[0,1]$ is easily seen to be equivalent to $\left[\frac{1}{4}, \frac{1}{2}\right] \subseteq (0,1)$ (*use a linear map*)

3) $(0,1) \sim (0,1)^2$. Why? $f(x) = (x, \frac{1}{2})$ is a one-to-one map of the interval into the square. In the opposite direction, we can easily define a one-to-one map in the opposite direction as follows:

for $(x, y) \in (0,1)^2$, write the binary expansions of x and y

$$x = 0.x_1x_2x_3...x_n..._{two} \quad \text{and} \quad y = 0.y_1y_2y_3...y_n..._{two}$$

choosing, in both cases, a binary expansion that does not end with an infinite string of 1's. Then define $g : (0,1)^2 \to (0,1)$ by "interlacing" the digits to create a base 10 decimal:

$$g(x, y) = 0.x_1y_1x_2y_2x_3y_3...x_ny_n...$$

The function g is one-to-one: suppose $(x, y) \neq (x', y')$. Then $g(x, y)$ and $g(x', y')$ have different decimal expansions. Neither decimal expansion ends in an infinite string of 9's, so $g(x, y)$ and $g(x', y')$ are different real numbers. (*What was the point of writing binary expansions of x and y?*)

4) It is very easy to show that if $A \sim B$ and $C \sim D$, then $A \times C \sim B \times D$ (*check!*). Then, because $(0,1) \sim [0,1] \sim [0,1) \sim \mathbb{R}$, we can immediately write down such equivalences as

$\mathbb{R} \sim (0,1) \sim (0,1)^2 \sim [0,1]^2 \sim (0,1) \times [0,1) \sim \mathbb{R}^2$. And since $\mathbb{R} \sim \mathbb{R}^2$, it follows that

$$\mathbb{R} \sim \mathbb{R} \times \mathbb{R} \sim \mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3.$$

Similarly, we can see that $\mathbb{R}^n \sim \mathbb{R}^m$ for any $m, n \in \mathbb{N}$.

Example 10.4 (A tangential comment) Although $\mathbb{R}^2 \sim \mathbb{R}$, there cannot be a continuous bijection between them. In fact, if $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous, then f cannot even be one-to-one. The proof is an exercise in the use of the Intermediate Value Theorem from calculus.

> Define a function ϕ of one variable by holding constant one variable in f: for convenience, say $\phi(x) = f(x, 0)$. Let $f(0,0) = \phi(0) = a$ and $f(1,0) = \phi(1) = b$. We can assume $a < b$ (if $a = b$, f is not one-to-one and we're done). Since f is continuous, so is ϕ. By the Intermediate Value Theorem, ϕ assumes all values between a and b; in particular, for some $c \in (0,1)$, $\phi(c) = \frac{a+b}{2}$.
>
> For that c, define $\psi(y) = f(c, y)$. Because ψ is continuous at 0, $\psi(y)$ is close to $\psi(0) = \phi(c) = \frac{a+b}{2}$ for y close to 0. Therefore, we know that $a < \psi(y) < b$ for all y in some sufficiently short interval $(-\delta, \delta)$. In particular, then $\psi(\frac{\delta}{2}) \in (a, b)$.
>
> But we know that ϕ assumes all values between a and b, so $\phi(z) = \psi(\frac{\delta}{2})$ for some z in $(0,1)$. Therefore $f(c, \frac{\delta}{2}) = \psi(\frac{\delta}{2}) = \phi(z) = f(z, 0)$. Since $(c, \frac{\delta}{2}) \neq (z, 0)$, f is not one-to-one. •
>
> *Note: we did not even use the full strength of the assumption that f is continuous. We needed only to know that ϕ and ψ were continuous – a weaker statement, as you should know from advanced calculus. This means there cannot even be a one-to-one map of the plane to the line which is continuous in each variable separately.*

A similar argument shows that there cannot be a one-to-one continuous map $f: [0,1]^2 \to [0,1]$. Therefore, in particular, there cannot be a continuous bijection from $[0,1]^2$ to $[0,1]$.

However, in the "opposite" direction, it is possible to construct a continuous map $f: [0,1] \to [0,1]^2$ which is onto. Such a map is called a space-filling curve, and such a map is often constructed in an beginning analysis course.

Is there be a continuous onto map from $\mathbb{R}^2 \to \mathbb{R}$?

11. More About Subsets

We have already seen that there is more than one "size" of infinite set: some are countable, others are uncountable. Moreover, we have seen uncountable sets of two different sizes: $\mathbb{R}$ and $\mathbb{R}^{\mathbb{R}}$. In fact, there are infinitely many different sizes. We will see this using the power set operation $\mathcal{P}$. But first we will prove a result about subsets of countable sets that frequently is useful to know.

Theorem 11.1 The collection $\mathcal{F}(A)$ of all finite subsets of a countable set A is countable.

Proof If A is finite, it has only finitely many subsets. So assume A is countable and infinite. For convenience, we assume A is the set of all prime numbers (*be sure you understand why there is no loss of generality in making this assumption!*). Then each finite $B \subseteq A$ is a set of primes and we can define $f : \mathcal{F}(A) \to \mathbb{N}$ by

$$f(B) = \begin{cases} 1 & \text{if } B = \emptyset \\ \text{the product of the primes in } B & \text{if } B \neq \emptyset \end{cases}$$

By the Fundamental Theorem of Arithmetic, f is one-to-one; so $\mathcal{F}(A)$ is countable. •

Recall that Y^A denotes the set of all functions from A into Y. When $Y = \{0, 1\}$, we will also use the notation 2^A for the set $\{0,1\}^A$. Thus, 2^A is the set of all "binary functions" with domain A. The following theorem gives us two important facts.

Theorem 11.2 For any set A,

1) $2^A \sim \mathcal{P}(A)$
2) A is not equivalent to $\mathcal{P}(A)$ (so $A \not\sim 2^A$)

Proof 1) For each $B \subseteq A$, define $\chi_B =$ the characteristic function of B

$$\text{by } \chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$$

Define $\phi : \mathcal{P}(A) \to 2^A$ by $\phi(B) = \chi_B$. The function ϕ pairs each subset of A with its characteristic function. The function ϕ is a bijection: it is clearly one-to-one because different subsets have different characteristic functions. And ϕ is also onto because every function $f : A \to \{0, 1\}$ is the characteristic function of a subset B of A : $f = \chi_B = \phi(B)$, where $B = \{x \in A : f(x) = 1\}$. So $\mathcal{P}(A) \sim 2^A$.

2) Consider any function $\psi : A \to \mathcal{P}(A)$. We will show that ψ cannot be onto. For each $x \in A$, $\psi(x)$ is a subset of A, so it makes sense to ask whether or not $x \in \psi(x)$. Let $B = \{x \in A : x \notin \psi(x)\}$. Of course, $B \in \mathcal{P}(A)$.

Suppose $x \in A$. If $x \in \psi(x)$, then $x \notin B$, so $B \neq \psi(x)$. But if $x \notin \psi(x)$, then $x \in B$ so, again, $B \neq \psi(x)$. Either way, $\psi(x) \neq B$, so $B \notin \text{ran}(\psi)$. •

We use the notation $\mathcal{P}^2(A)$ for $\mathcal{P}(\mathcal{P}(A))$ and, more generally, $\mathcal{P}^n(A)$ for $\mathcal{P}(\mathcal{P}^{n-1}(A))$. The preceding theorem says that no set is equivalent to its power set, so $\mathcal{P}^{n-1}(A)$ is not equivalent to $\mathcal{P}^n(A)$. We can prove even a bit more.

Corollary 11.3 No two of the sets in the sequence A, $\mathcal{P}(A)$, $\mathcal{P}^2(A)$, ... , $\mathcal{P}^n(A)$, ... are equivalent.

Proof To begin, notice that for each j, there is a one-to-one function $i_j : \mathcal{P}^j(A) \to \mathcal{P}^{j+1}(A)$ given by $i_j(a) = \{a\}$.

Now suppose there were a bijection $f : \mathcal{P}^j(A) \to \mathcal{P}^{j+k}(A)$ for some j, k. Then we would have the following maps:

$$\begin{tikzcd}
\mathcal{P}^j(A) \arrow[r, "i_j"] \arrow[rrrr, bend right, "f"'] & \mathcal{P}^{j+1}(A) \arrow[r, "i_{j+1}"] & \text{........} \arrow[r, "i_{j+k-2}"] & \mathcal{P}^{j+k-1}(A) \arrow[r, "i_{j+k-1}"] & \mathcal{P}^{j+k}(A)
\end{tikzcd}$$

Then $i_{j+k-1} : \mathcal{P}^{j+k-1}(A) \to \mathcal{P}^{j+k}(A)$ and $i_{j+k-2} \circ ... \circ i_j \circ f^{-1} : \mathcal{P}^{j+k}(A) \to \mathcal{P}^{j+k-1}(A)$ are both one-to-one. But then the CSB Theorem would imply that $\mathcal{P}^{j+k-1}(A) \sim \mathcal{P}^{j+k}(A)$, which is impossible by Theorem 11.2(2). •

Therefore, using repeated applications of the power set operation, we can generate an infinite sequence of sets no two of which are equivalent – for example, $\mathbb{N}$, $P(\mathbb{N})$, $\mathcal{P}^2(\mathbb{N})$, ... , $\mathcal{P}^n(\mathbb{N})$, ... In the "exponential" notation of Theorem 11.2(1), we can also write this sequence as $\mathbb{N}$, $2^{\mathbb{N}}$, $2^{2^{\mathbb{N}}}$, ... Thus we have infinitely many different "sizes" of infinite set.

Examples 11.4

1) $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}} \sim \mathcal{P}(\mathbb{N}) \not\sim \mathbb{N}$, so we conclude that the set of all binary sequences is uncountable.

2) Because $\mathcal{P}(\mathbb{N})$ is uncountable and because $\mathbb{N}$ has only countably many finite subsets (Theorem 11.1), we conclude that $\mathbb{N}$ has uncountably many infinite subsets. It follows that every infinite set has uncountably many infinite subsets (*explain the details!*).

3) Using the CSB Theorem, we can create another paradox similar to Russell's:

> Let $B = \{\{b\} : b \text{ is a set}\}$, so B is the "set of all one-element sets." Then the maps $\mathcal{P}(B) \to B$ and $B \to \mathcal{P}(B)$ given (in both directions!) by $v \to \{v\}$ are one-to-one. Then the CSB Theorem gives that $B \sim \mathcal{P}(B)$.

What's wrong? (*The paradox is dealt with in the same way as Russell's Paradox.*)

Since $2^{\mathbb{N}} \not\sim \mathbb{N}$, we can ask whether $2^{\mathbb{N}}$ is equivalent to some other familiar set. The next theorem answers this question.

Theorem 11.5 $2^{\mathbb{N}} \sim \mathbb{R}$

Proof We use the CSB Theorem. Define $\phi : 2^{\mathbb{N}} \to \mathbb{R}$ as follows: if $f \in 2^{\mathbb{N}}$, then $f : \mathbb{N} \to \{0,1\}$ so we can define $\phi(f) = \sum_{n=1}^{\infty} \frac{f(n)}{10^n} = \sum_{n=1}^{\infty} \frac{a_n}{10^n}$, where $a_n = f(n)$. In other words, $\phi(f)$ is the real number whose decimal expansion is $0.a_1a_2 \ldots a_n \ldots$. If $f \neq g \in 2^{\mathbb{N}}$, then $\phi(f)$ and $\phi(g)$ have different decimal expansions, and neither expansion ends in a string of 9's (*each* a_n *is either* 0 *or* 1). Therefore $\phi(f) \neq \phi(g)$, so $\phi : 2^{\mathbb{N}} \to \mathbb{R}$ is one-to-one.

To get a one-to-one function from $\mathbb{R}$ into $2^{\mathbb{N}}$, we begin by defining a function $\psi : \mathbb{R} \to \mathcal{P}(\mathbb{Q})$: for each $x \in \mathbb{R}$, let $\psi(x) = \{q \in \mathbb{Q} : q < x\}$ The function ψ is clearly one-to-one. Since $\mathbb{Q} \sim \mathbb{N}$, there is a bijection $h : \mathcal{P}(\mathbb{Q}) \to 2^{\mathbb{N}}$ (*see Theorem 11.2)*. Then $h \circ \psi : \mathbb{R} \to 2^{\mathbb{N}}$ is one-to-one.

By the CSB Theorem, $2^{\mathbb{N}} \sim \mathbb{R}$. ●

For $k > 1$, *it is fairly easy to generalize Theorem 11.5 replacing* $2^{\mathbb{N}}$ *with* $k^{\mathbb{N}}$, *where* $k^{\mathbb{N}}$ *is short for* $\{0, 1, \ldots, k-1\}^{\mathbb{N}}$.

12. Cardinal Numbers

To each finite set A we can assign, in principle, a number called the cardinal number or cardinality of A. It answers the question: "How many elements does A have?" The symbol $|A|$ represents the cardinal number of A. For example, $|\emptyset| = 0$, $|\{0\}| = 1$, and $|\{0,1\}| = 2$. In practice, of course, this might be difficult. For example, if $A = \{p \in \mathbb{N} : p \text{ is prime and } p \leq 10^{100}\}$, then $|A| = ?$
Of course, there must be a correct value for $|A|$, but actually finding it would be hard. (*You could make a rough estimate using the Prime Number Theorem – see Example 5.1(8), p. 16:* $|A| = \pi(10^{100}) \approx \frac{10^{100}}{\ln(10^{100})} \approx 4.3429 \times 10^{97}$.)

We have already stated informally that nonequivalent infinite sets have "different sizes." To make this more precise, we will assume that to each set A there is associated an "object" denoted $|A|$ and called the cardinal number of A (or cardinal of A for short), and that this is done in such a way that $|A| = |B|$ if and only if $A \sim B$.

> *Note: How to justify this assumption precisely doesn't matter right now. That question belongs in a more advanced set theory course. It turns out that* $|A|$ *can be precisely defined in ZFC. Like every object in mathematics,* $|A|$ *is itself a certain set and, of course,* $|A|$ *is defined in terms of the given set A. For these notes, though, it is enough to know that two sets have the same cardinal number if and only if the sets are equivalent.*

There are standard symbols for the cardinal numbers associated to certain everyday sets:

$$
\begin{array}{lll}
|\emptyset| & = 0 & \\
|\{0\}| & = 1 & (\text{Of course, } |\{a\}| = 1 \text{ for any } a, \text{ because } \{0\} \sim \{a\}) \\
|\{0,1\}| & = 2 & (\text{Of course, } |\{a,b\}| = 2 \text{ for any } a \neq b) \\
 & \vdots & \\
|\mathbb{N}| & = \aleph_0 & (\text{So } \aleph_0 \text{ is the cardinal number for every countable infinite set;} \\
 & & \text{for example, } |\mathbb{N} \times \mathbb{N} \times \mathbb{N}| = |\mathbb{Q}| = \ldots = \aleph_0) \\
|\mathbb{R}| & = c &
\end{array}
$$

The symbols $\aleph$ and c were chosen by Cantor in his original work. The symbol $\aleph$ is "aleph," the first letter of the Hebrew alphabet (*but Cantor was raised as a Lutheran*). We read $\aleph_0$ as "aleph-zero" (or the more British "aleph-nought" or "aleph-null.")

Cantor chose "c" for $|\mathbb{R}|$ since c is the first letter of the word "continuum" (*an old word for the real line*). Because we know that the following sets are equivalent, we can write, for example, that $|[0,1]| = |(0,1)| = |(0,1)^2| = |\mathbb{R}| = |\mathbb{R}^2| = |\mathbb{R}^3| = \ldots = c$.

What is $|\mathbb{P}|$? Why? (*You can give a definite answer for this, but be careful: you can't assume that an uncountable subset of $\mathbb{R}$ must be equivalent to $\mathbb{R}$. See the discussion of CH on pp. 40-42.*)

13. Ordering the Cardinals

We have talked informally about some infinite sets being "bigger" than others. We can make this precise by defining an order " $\leq$ " between cardinal numbers. Throughout this section, M, N, P are sets with $|M| = m$, $|N| = n$, and $|P| = p$ (N is not necessarily the set natural numbers, $\mathbb{N}$). We say that the sets $M, N,$ and P represent the cardinal numbers $m, n,$ and p.

Definition 13.1 1) $m \leq n$ means that M is equivalent to a subset of N. (*We also write $m \leq n$ as $n \geq m$.*)

2) $m < n$ (or $n > m$) means that $m \leq n$ but $m \neq n$. (*We also write $m < n$ as $n > m$.*)

Thus $m < n$ means that M is equivalent to a subset of N but M is not equivalent to N. According to the CSB Theorem, this is the same as saying that M is equivalent to a subset of N but N is not equivalent to a subset of M.

There is a detail to check. The relation $m \leq n$ is defined using the sets M and N. Another person might choose different sets, say $m = |M'|$ and $n = |N'|$ to represent the cardinal numbers. Would that person necessarily come to the same conclusion that $m \leq n$? We need to check that our definition is independent of which sets are chosen to represent the cardinals $m, n,$ and p or, in other words, that $\leq$ has been well-defined. This is easy to do.

Suppose M' and N' also represent m and n. Then $M \sim M'$ and $N \sim N'$, so there are bijections f, g as shown below. If M is equivalent to a subset of N, then there is also a one-to-one map h as pictured below:

$$\begin{array}{ccc} M & \xrightarrow{f} & M' \\ h \downarrow & & \downarrow k \\ N & \xrightarrow{g} & N' \end{array}$$

So $k = g \circ h \circ f^{-1}$ is a one-to-one map $k : M' \to N'$, so M' is equivalent to a subset of N'. Therefore the question "Is $m \leq n$?" does not depend on which representing sets we use – that is, the relation $\leq$ is well-defined.

Theorem 13.2 For any cardinal numbers $m, n,$ and p :

1) $m \leq m$
2) $m \leq n$ and $n \leq p$ implies $m \leq p$
3) $m \leq n$ and $n \leq m$ implies $m = n$
4) $m \leq n$ and $n < p$ implies $m < p$
5) at most one of the relations $m < n$, $m = n$, and $m > n$ holds
6*) at least one of the relations $m < n$, $m = n$, and $m > n$ holds.

Proof The proofs of 1) and 2) are obvious.

For 3), notice that we are given that each of M and N is equivalent to a subset of the other. Then $M \sim N$ (by the CSB Theorem), so $m = n$.

4) If $n < p$, then $n \leq p$ so, by part 2), $m \leq p$. If $p = m$, then $p \leq n$ and therefore $n = p$ by part 3). But this contradicts the hypothesis that $n < p$. Therefore $p \neq m$, so $m < p$.

5) By definition of $<$, $m = n$ excludes $m < n$ and $n < m$. And if $m < n$ and $n < m$ were both true, then $m \leq n$ and $n \leq m$, so $m = n$, which is impossible.

6*) The proof is postponed. •

It seems like the proof of part 6) of the theorem shouldn't be difficult. Informally, you simply pair an element of M with an element of N, and keep repeating this process until either a bijection is created between M and N or until one of the sets has no remaining elements. If one of the sets is used up before the other, then it has the smaller cardinal. In fact, this works for finite sets, but for infinite sets it is hard to make precise what "keep repeating this process until ..." means. To make the argument precise, the Axiom of Choice has to be used in some form. We could develop the machinery to complete the proof here, but it would digress too much from the main ideas. So for use in our examples, we'll simply assume part 6*) for now.*

From parts 5) and 6*) of the theorem, we immediately get the following corollary.

Corollary 13.3 For any two cardinals m and n, exactly one of the relations $m < n,\ m = n,$ or $m > n$ is true.

Examples 13.4

1) If $k = |K|$ is a finite cardinal, then K is equivalent to a subset of $\mathbb{N}$ but K is not equivalent to $\mathbb{N}$. And if m is an infinite cardinal number, there is a one-to-one function $f : \mathbb{N} \to M$. These observations show that $k < \aleph_0 \leq m$, so $\aleph_0$ is the smallest infinite cardinal number.

2) $\aleph_0 < c < \; | \, \mathbb{R}^{\mathbb{R}} \, |$ *(Explain)*

14. The Arithmetic of Cardinal Numbers

We want to define exponentiation, addition and multiplication for cardinal numbers. Of course, for finite cardinals, these operations will agree with the usual arithmetic operations in $\mathbb{N}$.

We begin with exponentiation. As in the previous section, $m, n,$ and p will denote cardinals represented by sets $M, N,$ and P.

Definition 14.1 $n^m = \; | \, N^M \, |$

Thus, n^m is the number of functions from the set M into the set N. As with the definition of the order relation $\leq$, we must check that exponentiation is well-defined. (That is, if one person calculates $c^{\aleph_0}$ using $|\mathbb{R}^{\mathbb{N}}|$ and another person uses $|\,(0,1)^{\mathbb{Q}}|$, will they get the same answer?)

Assume $m = |M| = |M'|$ and $n = |N| = |N'|$. We must show that $N^M \sim (N')^{M'}$, that is, we must produce a bijection $\phi : N^M \to (N')^{M'}$.

By hypothesis, we have bijections $f : M \to M'$ and $g : N \to N'$. For $h \in N^M$, define a function $\phi(h) = g \circ h \circ f^{-1} \in (N')^{M'}$.

ϕ is one-to-one: Suppose $h \neq k \in N^M$. Then for some $m \in M$, $h(m) \neq k(m)$.
Let $m' = f(m)$. Then $\phi(h)(m') = g(h(f^{-1}(m'))) = g(h(m))$ and
$\phi(k)(m') = g(k(f^{-1}(m'))) = g(k(m))$.
But $h(m) \neq k(m)$ and g is one-to-one so $\phi(h)(m') = g(h(m)) \neq g(k(m)) = \phi(k)(m')$. Therefore the functions $\phi(h)$ and $\phi(k)$ have different values at m, so $\phi(h) \neq \phi(k)$. Therefore ϕ is one-to-one.

ϕ is onto: Exercise.

Examples 14.2

1) $|\,[0,1]^{(0,1)}\,| = |\,\mathbb{R}^{\mathbb{R}}| = c^c$

2) $2^{\aleph_0} = |2^{\mathbb{N}}| = c$, since $2^{\mathbb{N}} \sim \mathbb{R}$. Thus, there are c different binary sequences.
As remarked earlier, it is not hard to generalize this to show that $k^{\aleph_0} = c$ is true for any integer $k > 1$.

Students sometimes confuse the fact that $2^{\aleph_0} = c$ with the continuum hypothesis. The continuum hypothesis states that there is no cardinal m such that $\aleph_0 < m < 2^{\aleph_0} = c$.

In other words, CH states that $2^{\aleph_0}$ is the immediate successor (in terms of size) of the cardinal number $\aleph_0$. Here is still another way of putting it: let us <u>*define*</u> *$\aleph_1$ to be the immediate successor of $\aleph_0$ (assuming for now that an immediate successor must exist). Then CH simply states that $c = \aleph_1$. To go a little further, the* <u>*generalized*</u> <u>*continuum*</u> <u>*hypothesis*</u> *(*<u>*GCH*</u> *for short) states that for any infinite cardinal m, its "immediate successor" is 2^m.*

(Since ZFC cannot prove CH, certainly ZFC cannot prove the stronger statement GCH. In fact, GCH is undecidable in ZFC.)

3) For any set M, M is equivalent to $\{\{m\} : m \in M\}$, which is a subset of $\mathcal{P}(M)$. Therefore $m \leq 2^m$. But $m \neq 2^m$ because M is not equivalent to 2^M. Therefore $m < 2^m$ for all cardinals m. It follows that $m < 2^m < 2^{2^m} < 2^{2^{2^m}} < \ ... \ < \ ...$ is an infinite increasing sequence of infinite cardinals. In particular, for $m = \aleph_0$, we have $\aleph_0 < 2^{\aleph_0} = c < 2^{2^{\aleph_0}} < \ \ldots$

We now define multiplication and addition of cardinal numbers.

Definition 14.3 Suppose $m = |M|$ and $n = |N|$,

1) $m \cdot n$ (or simply mn) means $|M \times N|$
2) $m + n$ means $|M \cup N|$, where M and N are <u>disjoint</u> sets representing m and n

Part 2) requires that to add m and n, we choose <u>*disjoint*</u> *representing sets M and N. This is* <u>*always possible*</u> *because if $M \cap N \neq \emptyset$, we can replace M and N by the equivalent* <u>*disjoint*</u> *sets $M \times \{0\}$ and $N \times \{1\}$.*

You should check that multiplication and addition are well-defined – that is, the operations are independent of the sets chosen to represent the cardinals m and n.

Theorem 14.4 For cardinal numbers m, n, p, and q :

1) $m + n = n + m$	$1'$) $mn = nm$
2) $(m + n) + p = m + (n + p)$	$2'$) $m(np) = (mn)p$
3) $p(m + n) = pm + pn$	

If $m \leq n$ and $p \leq q$, then

4) $m + p \leq n + q$	$4'$) $mp \leq nq$

Proof The proofs are all simple. For example, we will prove $4'$).

Suppose m, n, p, q are represented by the sets M, N, P, Q. Since $m \leq n$ and $p \leq q$, there are one-to-one functions $f : M \to N$ and $g : P \to Q$. Define $h : M \times P \to N \times Q$ by $h(m, p) = (f(m), g(p))$. Then h is one-to-one (*why?*), so $mp \leq nq$. •

Addition Examples 14.5

1) For any m : $m + 0 = m$ because $M \cup \emptyset \sim M$.

2) For finite n : $n + \aleph_0 = \aleph_0 + \aleph_0 = \aleph_0$ because the union of two countable sets is countable.

3) For finite n : $n + c = \aleph_0 + c = c + c = c$.

$c + c = c$ is true because $(-\infty, 0) \cup [0, \infty) = \mathbb{R}$. Then write the inequalities

$$\begin{array}{lllllll} 0 & \leq & n & \leq \aleph_0 & \leq & c & \text{and} \\ c & \leq & c & \leq c & \leq & c & \end{array}$$

Add these inequalities and apply Theorem 14.4(4) to get

$$c \ \leq \ n + c \ \leq \ \aleph_0 + c \ \leq \ c + c.$$

But $c + c = c$, so we conclude that $c = n + c = \aleph_0 + c = c + c$.

4) If m is infinite, then $m + \aleph_0 = m$.

To see this, pick M so that $m = |M|$ and $M \cap \mathbb{N} = \emptyset$. M is infinite so there is a one-to-one map $f : \mathbb{N} \to M$ and we can write $M = f[\mathbb{N}] \cup (M - f[\mathbb{N}])$. Because $f[\mathbb{N}]$ is countable, so is $f[\mathbb{N}] \cup \mathbb{N}$. Therefore we have a bijection $g : f[\mathbb{N}] \cup \mathbb{N} \to f[\mathbb{N}]$. We can then define a bijection $h : M \cup \mathbb{N} \to M$ by

$$h(x) = \begin{cases} x & \text{if } x \in M - f[\mathbb{N}] \\ g(x) & \text{if } x \in \mathbb{N} \cup f[\mathbb{N}] \end{cases}$$

The following two facts about addition are also true, but the proofs require more complicated arguments that involve AC. We omit the proofs and, for now, simply assume 5) and 6*).*

5*) If m or n is infinite, then $m + n = \max\{m, n\}$: that is, a sum involving an infinite cardinal number equals the larger of the two cardinals.

6*) If $m < n$ and $p < q$, then $m + p < n + q$.

This result deserves a word of caution. For infinite cardinals, it is not true in general that if $m < n$ and $p \leq q$, then $m + p < n + q$! For example, $\aleph_0 < c$ and $c \leq c$ but $\aleph_0 + c = c + c \ \ (= c)$.

7) The World's Longest Song: "$\aleph_0$ Bottles of Beer on the Wall" (*It fits the tune better with the British pronunciation "Aleph-nought."*)

Multiplication Examples 14.6

1) For any m: $m \cdot 0 = 0$ and $m \cdot 1 = m$ because
$M \times \emptyset \sim \emptyset$ and $M \times \{a\} \sim M$

2) $\aleph_0^2 = \aleph_0$.
If we view $\aleph_0^2$ as shorthand for $\aleph_0 \cdot \aleph_0$, the equation is true because $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$. But an alert reader might point out that we could also interpret $\aleph_0^2$ as an exponentiation, that is, as $|\,\mathbb{N}^{\{0,1\}}|$. However, this gives to the same result because $\mathbb{N}^{\{0,1\}} \sim \mathbb{N} \times \mathbb{N}$. (*Mapping each function $f \in \mathbb{N}^{\{0,1\}}$ to the pair $(f(0), f(1)) \in \mathbb{N} \times \mathbb{N}$ is a bijection.*)

3) If n is finite and $n > 0$, then $n \cdot \aleph_0 = \aleph_0$.
To see this, begin with the inequalities

$$\begin{array}{lllll} 1 & \leq & n & \leq & \aleph_0 \\ \aleph_0 & \leq & \aleph_0 & \leq & \aleph_0 \end{array}$$

Multiplying and applying Theorem 14.4$(4')$ gives

$$\aleph_0 \ \leq \ n \cdot \aleph_0 \ \leq \aleph_0^2$$

Since $\aleph_0 = \aleph_0^2$, we get that $\aleph_0 = n \cdot \aleph_0 = \aleph_0^2$.

4) $c^2 = c$ because $(0,1)^2 \sim (0,1)$

5) If n is finite and $n > 0$, then $nc = \aleph_0 \cdot c = c^2 = c$.
To see this begin with the inequalities

$$\begin{array}{lllllll} 1 \leq & n & \leq & \aleph_0 & \leq & c & \\ c \leq & c & \leq & c & \leq & c & \text{and multiply to get} \end{array}$$

$$c \leq \ nc \leq \ \aleph_0 \cdot c \ \leq \ c^2$$

Since $c = c^2$, we conclude that $c = nc = \aleph_0 \cdot c = c^2$

One additional fact about multiplication of cardinals will be assumed, for now, without proof:

6*) If m is infinite and $n \neq 0$, then $mn = \max\{m, n\}$. In particular, for an infinite cardinal m, we have $m^2 = m$.

In algebraic systems (such as $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$) where subtraction is defined, the definition of subtraction is always given in terms of addition: $a - b = c$ is <u>defined</u> to mean $a = b + c$. This shows why <u>subtraction cannot be sensibly defined for infinite cardinal numbers</u>: should we say $\aleph_0 - \aleph_0 = 0$ because $\aleph_0 = \aleph_0 + 0$? or $\aleph_0 - \aleph_0 = 1$ because $\aleph_0 = \aleph_0 + 1$? or $\aleph_0 - \aleph_0 = \aleph_0$ because

$\aleph_0 = \aleph_0 + \aleph_0$? Similarly, division is usually defined in terms of multiplication: $\frac{a}{b} = c$ means $a = bc$. Think about why there is also no sensible definition for division involving infinite cardinal numbers.

The proof of the next theorem is an excellent check on whether you understand "function notation."

Theorem 14.7 For cardinals $m, n,$ and p : $(m^n)^p = m^{np}$

Proof We want to define a bijection $\phi : (M^N)^P \to M^{(N \times P)}$.

If $f \in (M^N)^P$, then f is a function $P \to M^N$. If $p \in P$, then $f(p) \in M^N$, so $f(p)$ is a function $N \to M$. Therefore, for $n \in N$, $f(p)(n)$ makes sense: it is an element of M. So we can define a function $\phi(f) \in M^{N \times P}$ by the rule $\phi(f)(n, p) = f(p)(n)$.

ϕ is onto: if $g \in M^{N \times P}$, let $f : P \to M^N$ be the function defined by $f(p)(n) = g(n, p)$. Then $f \in (M^N)^P$ and $\phi(f) = g$ because for any pair (n, p), we have $\phi(f)(n, p) = f(p)(n)$
$= g(n, p)$.

ϕ is one-to-one: if $f, g \in (M^N)^P$ and $f \neq g$, then for some $p \in P$, $f(p) \neq g(p)$. Because $f(p)$ and $g(p)$ are different functions $N \to M$, there must be some $n \in N$ for which $f(p)(n) \neq g(p)(n)$. But this says that $\phi(f)(n, p) \neq \phi(g)(n, p)$, so $\phi(f) \neq \phi(g)$. •

Examples 14.8

1) $c^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0} = c$

2) $c = 2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq c^{\aleph_0} = c$, so $\aleph_0^{\aleph_0} = c$

3) $|\mathbb{R}^{\mathbb{R}}| = c^c = (2^{\aleph_0})^c = 2^{\aleph_0 \cdot c} = 2^c$

Caution: We now know that $\aleph_0 < \aleph_0^{\aleph_0} = c$
but that $c^{\aleph_0} = c$

So we might be tempted to conjecture that if $m > \aleph_0$, then $m^{\aleph_0} = m$. But this is false.
In fact, for every cardinal number k, it is possible

to find an $m > k$ for which $m^{\aleph_0} > m$ and also
to find a $p > k$ for which $p^{\aleph_0} = p$.

The proof of this is a little too complicated to look at now.

Exercises

E46. Prove or disprove: Let $\mathbb{A}$ be the set of algebraic numbers. Then every open interval in $\mathbb{R}$ contains a point of $\mathbb{R} - \mathbb{A}$.

E47. a) Suppose S is a countable subset of $\mathbb{R}$. Prove that there exists a fixed real number, c such that $s + c$ is transcendental for every $s \in$ S.

b) For any set $B \subseteq \mathbb{R}$ and $\alpha \in \mathbb{R}$, we write $B + \alpha$ for the set $\{b + \alpha : b \in B\}$. Find a set $A \subseteq \mathbb{R}$ for which $|A| = c$ and $(\mathbb{Q} + \alpha) \cap (\mathbb{Q} + \beta) = \emptyset$ for all $\alpha \neq \beta \in A$.

E48. You and I play the following infinite game. We take turns (you go first) picking "0" or "1" and use our choices as the consecutive digits of a binary decimal which, when completed, represents a real number in the interval $[0, 1]$. You win if this number is transcendental; I win if it is algebraic. Explain how to make your choices so that you are guaranteed to win, no matter what choices I make. (*Hint: Consider the binary expansions of the algebraic numbers in* $[0, 1]$. *Look at the "diagonal proof" that shows* $(0, 1)$ *is uncountable.*)

E49. Find the cardinal number of each of the following sets:

a) the set of all convergent sequences of real numbers

b) the set of all straight lines ℓ in the plane for which $|\ell \cap (\mathbb{Q} \times \mathbb{Q})| \geq 2$

c) the set of all sequences $f : \mathbb{N} \to \mathbb{N}$ that are eventually constant (*Note: f is eventually constant if there are natural numbers l and $m \in \mathbb{N}$ such that $f(n) = l$ for all $n \geq m$.*)

d) the set of all differentiable functions $f : \mathbb{R} \to \mathbb{R}$

Hint: if f, g are differentiable and $f \neq g$, what can you say about $f|\mathbb{Q}$ and $g|\mathbb{Q}$?

e) the set of all geometric progressions in $\mathbb{R}$ (*A sequence (x_n) in $\mathbb{R}$ is a geometric progression if there exists a real number $r \neq 0$ such that $a_{n+1} = ra_n$ for every n.*)

f) the set of all strictly increasing sequences $f : \mathbb{N} \to \mathbb{N}$ (*Note: f is strictly increasing if $l, m \in \mathbb{N}$ and $l < m \Rightarrow f(l) < f(m)$.*)

g) the set of all countable subsets of $\mathbb{R}$. (*Hint: part f) could be used.*)

Note: Often one proves $|A| = m$ by using two separate arguments, one to show $|A| \leq m$ and the other to show $|A| \geq m$. Sometimes one of these arguments is much easier than the other.

E50. Find a subset of $\mathbb{Q}$ that is equivalent to the set of all binary sequences (a_n), or explain why no such subset exists.

E51. Explain why the following statement is true:

The set of all real numbers x which have a decimal expansion of the form

$$x = 0.x_1x_2x_3...x_n 0101\overline{01}... \quad (n \textit{ may depend on } x)$$

is countable.

E52. Prove that for any collection of sets $\{A_\lambda : \lambda \in \Lambda\}$, there must exist a set Y such that $|Y| > |A_\lambda|$ for every $\lambda \in \Lambda$. (*Hint: use the fact that* $X \not\sim \mathcal{P}(X)$.)

E53. Is the following statement true or false? Prove the statement, or give a counterexample.

> If $\mathcal{C}$ is an uncountable collection of uncountable subsets of $\mathbb{R}$, then at least two sets in $\mathcal{C}$ must have an uncountable intersection.

E54. Prove that if m, n, and p are cardinals, then

a) $m^{n+p} = m^n \cdot m^p$
b) $(m \cdot n)^p = m^p \cdot n^p$

E55. Prove or disprove:

a) $\aleph_0^c = 2^c$
b) $2^{(2^{\aleph_0})} = (2^2)^{\aleph_0}$
c) if $2 \leq m < 2^{\aleph_0}$, then $m^{\aleph_0} = c$
d) if m is infinite and $2 \leq n \leq 2^m$, then $n^m = 2^m$

E56. a) Prove that $\mathbb{R}$ has c countable subsets.
b) Prove or disprove: If A and B have the same number of countable subsets, then $A \sim B$.

E57. Prove or disprove: there are exactly c sequences of the form $(A_1, A_2, \ldots, A_n, \ldots)$ where each A_n is a subset of $\mathbb{Q}$.

E58. Find all unjustified steps in the following "proof" of the continuum hypothesis:

If CH is false, then $\aleph_0 < m < c$ for some cardinal m. Since $c = \aleph_0^{\aleph_0} \leq m^{\aleph_0} \leq c^{\aleph_0} = c$, we have $m^{\aleph_0} = c = 2^{\aleph_0} < 2^m$, so $m^{\aleph_0} < 2^m$. Therefore $(m^{\aleph_0})^c < (2^m)^c$, so $m^c < 2^c$, which is impossible because $m > 2$. Therefore no such m can exist, so CH is true.

E59. Find all unjustified steps in the following "disproof" of the continuum hypothesis:

We know $c = 2^{\aleph_0} = 2^{\aleph_0^2} = (2^{\aleph_0})^{\aleph_0}$. However, $(2^{\aleph_0})^{\aleph_0} > \aleph_0^{\aleph_0}$ (because $2^{\aleph_0} > \aleph_0$). Since $\aleph_0 > 1$, we have $\aleph_0^{\aleph_0} > \aleph_0^1 = \aleph_0$. Therefore $c > \aleph_0^{\aleph_0} > \aleph_0$, so CH is false.

E60. Since $\mathbb{Q}$ is countable and $|\mathbb{R}| = c$, we know that $\mathbb{P}$ is uncountable. But that does not automatically mean that $|\mathbb{P}| = c$ (*unless you assume CH, you might think that* $\aleph_0 < |\mathbb{P}| < c$).

a) Without using the properties 5*), 6*) of cardinal addition or multiplication and without using CH, prove that:

if $|B| = c$ and A is a countable subset of B, then $|B - A| = c$.

(*Hints: Obviously* $\aleph_0 < |B - A| \leq c$. *One way to show* $|B - A| = c$: *without loss of generality, you can assume* $B = \mathbb{R}^2$. *Then consider vertical lines in* B. *Of course, there are other approaches.*)

b) Using a), deduce that $|\mathbb{P}| = c$.

E61. Prove that for any infinite set E, there is an infinite sequence of disjoint subsets E_1, E_2, E_3,... such that $E = \bigcup_{n=1}^{\infty} E_n$ and $|E_n| = |E|$ for all n.
(*Hint: You can assume the multiplication rule in Example 14.6(6*). It implies that* $m \cdot \aleph_0 = m$ *for any infinite cardinal* m. .)

E62. Call a function $f : X \to Y$ "double-rooted" if $|f^{-1}(y)| = 2$ for every $y \in Y$. Find the number of double-rooted functions $f : \mathbb{Q} \to \mathbb{Q}$.

E63. Assume the generalized continuum hypothesis (see p. 51). Then

<u>True or false</u> (<u>explain</u>): $\mathcal{P}(A) \sim \mathcal{P}(B)$ implies $A \sim B$.

E64. Say that a pair of sets (A, B) has property (*) if all three of the following conditions are true:

i) $A \cup B = \mathbb{N} \times \mathbb{N}$
ii) every horizontal line intersects A in only finitely many points
iii) every vertical line intersects B in only finitely many points.

We saw in Exercise E44 that such pairs (A, B) *exist.*

Prove or disprove: there are exactly c different pairs (A, B) with property (*).

E65. Let $D = \{a_1, a_2, ..., a_n, ...\}$ be a countable subset of $\mathbb{R}$ and choose positive numbers ϵ_n for which $\sum_{n=1}^{\infty} \epsilon_n < \infty$. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \sum_{a_n \leq x} \epsilon_n$. Clearly $f(x) \leq f(y)$ if $x \leq y$.

Prove that f is discontinuous at each point in D and continuous at each point in $\mathbb{R} - D$.

In Theorem 8.5, we saw *that a monotone function* $f : \mathbb{R} \to \mathbb{R}$ *has a countable set of discontinuities; this result is a "sort of" converse. Note that this* f *is continuous from the right at every point.*)

15. A Final Digression

Let S^2 denote the sphere $\{(x,y,z) \in \mathbb{R}^3 \colon x^2+y^2+z^2=1\}$.

We will prove the following surprising (or not so surprising?) result. Informally stated:

> If a countable set D is removed from S^2, it is possible to write the remainder $S^2 - D$ as the union of two subsets A and B which, when rotated, give the whole sphere S^2 back again.

For a point $v = (x,y,z) \in S^2$, we use vector notation and write $-v = (-x, -y, -z)$.

Since S^2 is uncountable (*why?*), we can choose a point $v \in S^2$ for which $v \notin D \cup (-D) = \{\pm d : d \in D\}$. Then neither v nor $-v$ is in D. Change the coordinate axes so that the z-axis goes through v and $-v$ (so the "north and south poles" of S^2 are not in D).

For any $C \subseteq S^2$, write $C(\beta)$ to represent the set obtained by rotating C on the surface S^2 around the z-axis through angle β. In other words, a point in $C(\beta)$ comes from taking a point in C and adding β to its "longitude" on S^2. With this notation, the precise statement of what we want to prove is:

> Suppose D is a countable subset of S^2. Then there are subsets A and B of S^2, and there are real numbers α and γ such that
>
> i) $S^2 - D = A \cup B$ and
> ii) $S^2 = A(\alpha) \cup B(\gamma)$

First, we claim that we can choose a $\beta \in [0, 2\pi)$ that makes the sets $D,\ D(\beta),\ D(2\beta),\ \ldots,\ D(n\beta),\ \ldots$ all pairwise disjoint. For any point v (other than the north and south poles), let arg$(v) \in [0, 2\pi)$ be the longitude of v measured from the great circle through $(1,0,0)$ (= the "Greenwich meridian"). $D(k\beta)$ and $D(n\beta)$ can intersect only if there are points $a, b \in D$ such that

$$\text{arg}(a) + k\beta = \text{arg}(b) + n\beta + 2j\pi \quad (k, n \in \mathbb{N},\ j \in \mathbb{Z}) \qquad (**)$$

This means $D(n\beta)$ and $D(k\beta)$ can intersect <u>only if</u> β satisfies the equation $\beta = \frac{\text{arg}(a) - \text{arg}(b) - 2j\pi}{n-k}$.

But there are only countably to pick a "5-tuple" of values (a, b, j, n, k) to plug into the right side of the equation – because $\aleph_0^5 = \aleph_0$. So there are only countably many values β for which a pair of the sets $D, D(\beta), D(2\beta), \ \ldots\ , D(n\beta), \ \ldots$ could intersect. <u>Choose</u> any $\beta \in [0, 2\pi)$ different from these countably many values; then the sets will be pairwise disjoint.

Now, define $T = D \cup D(\beta) \cup D(2\beta) \cup \ldots \cup D(n\beta) \cup \ldots = \bigcup_{n=0}^{\infty} D(n\beta)$
and $B = T - D = D(\beta) \cup D(2\beta) \cup \ldots \cup D(n\beta) \cup \ldots = \bigcup_{n=1}^{\infty} D(n\beta)$

A rotation through $(-\beta)$ moves each set $D((k+1)\beta)$ onto the set $D(k\beta)$, so $B(-\beta) = T$.

Let $A = S^2 - T$. Then $A \cup B = (S^2 - T) \cup (T - D) = S^2 - D$.

Since $A(0) = A$ and $B(-\beta) = T$, we have $A(0) \cup B(-\beta) = (S^2 - T) \cup T = S^2$. •

Chapter I Review

For each statement, decide whether it is true or false. Then prove it, or provide a counterexample.

1. $\mathbb{R}^{\mathbb{R}} \sim \mathcal{P}(\mathbb{R})$

2. The continuum hypothesis (CH), which states that $2^{\aleph_0} = c$, is independent of the other usual axioms of the set theory.

3. If $f : \mathbb{R} \to \mathbb{R}$ is strictly decreasing, then there are at least 7 points at which f is continuous.

4. There exists a straight line ℓ in the plane such that ℓ contains exactly three points (x, y) where both x and y are rational.

5. In $\mathbb{R}^3$, it is possible to find uncountably many "solid balls" of the form

$$\{(x, y, z) : (x-a)^2 + (y-b)^2 + (z-c)^2 < \epsilon\}$$

such that any two of them are disjoint.

6. The continuum hypothesis is true iff the set of all sequences of 0's and 1's has cardinality c.

7. There are c different infinite sets of prime numbers.

8. Let A be the set of all sequences (a_n) where $a_n \in \{0, 1, 2, 3\}$ and such that $\{n : a_n = k\}$ is infinite for each $k = 0, 1, 2, 3$. Then A is countable.

9. If $\mathcal{C}$ is an uncountable collection of uncountable subsets of $\mathbb{R}$, then at least two sets in $\mathcal{C}$ must have uncountable intersection. (*Hint: recall that* $c^2 = c$)

10. The set of real numbers which are not transcendental is uncountable.

11. $\mathbb{N}^{\mathbb{R}} \sim \mathcal{P}(\mathbb{R})$

12. Let S^2 denote the unit sphere $\{(x, y, z) \in \mathbb{R}^3 : \ x^2 + y^2 + z^2 = 1\}$ and $P = (0, 0, 1)$. There exists a continuous bijection $f : S^2 - \{P\} \to \mathbb{R}$. (*The values of f are in $\mathbb{R}$, not $\mathbb{R}^2$!*)

13. Assume that for infinite cardinals m, there is never a cardinal strictly between m and 2^m (the Generalized Continuum Hypothesis). Then $\mathcal{P}(A) \sim \mathcal{P}(B)$ implies $A \sim B$.

14. There are exactly c sequences of the form $(A_1, A_2, \ldots, A_n, \ldots)$ where each A_n is a subset of $\mathbb{Q}$.

15. There is an algebraic number between any two real numbers.

16. Let $S = \{f \subseteq \mathbb{R}^2 :$ every horizontal line and every vertical line intersects f in exactly one point$\}$. Then $|S| = 2^c$.

17. There are 2^c subsets of $\mathbb{R}$ none of which contains an interval of positive length.

18. Let $\Gamma : C[a, b] \to \mathbb{R}$ by $\Gamma(f) = f(\frac{a+b}{2})$. Then Γ is one-to-one.

19. Suppose a nonempty set X can be "factored" as $X = Y \times Z$. Then Y and Z are unique.

20. Suppose A, B and C are infinite sets and that $A \sim B \cup C$. Then $A \sim B$ or $A \sim C$.

Chapter II
Metric and Pseudometric Spaces

1. Introduction

By itself, a set doesn't have any structure. For two arbitrary sets A and B, we can ask questions like "Is $A \subseteq B$?" or "Is A equivalent to a subset of B?" but not much more. If we add additional structure to a set, it becomes more interesting. For example, if we define a "multiplication operation" $a \cdot b$ in X that satisfies certain axioms (such as $a \cdot (b \cdot c) = (a \cdot b) \cdot c$)), then X becomes an algebraic structure called a group and a whole area of mathematics known as group theory begins.

We are not interested in making a set X into an algebraic system. Rather, we want additional structure on a set X so that we can talk about "nearness" in X. This is what we need to begin topology; "nearness" lets us discuss topics like "convergence" and "continuity." For example, "f is continuous at a" means (roughly) that "if x is near a, then $f(x)$ is near $f(a)$."

The simplest way to talk about "nearness" is to equip the set X with a distance function d to tell us "how far apart" two elements of X are.

Note: As we proceed we may use some ideas taken from elementary analysis, such as the continuity of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ as a source for motivation or examples, although these ideas will not be carefully defined until later in this chapter.

2. Metric and Pseudometric Spaces

Definition 2.1 Suppose $d : X \times X \to \mathbb{R}$ and that for all $x, y, z \in X$:

1) $d(x, y) \geq 0$
2) $d(x, x) = 0$
3) $d(x, y) = d(y, x)$ (symmetry)
4) $d(x, z) \leq d(x, y) + d(y, z)$ (the triangle inequality)

Such a "distance function" d is called a pseudometric on X. The pair (X, d) is called a pseudometric space. If d also satisfies

5) when $x \neq y$, then $d(x, y) > 0$

then d is called a metric on X and (X, d) is called a metric space. Of course, every metric space is automatically a pseudometric space.

If a pseudometric space (X, d) is not a metric space, it is because there are at least two points $x \neq y$ for which $d(x, y) = 0$. In most situations this doesn't happen; metrics come up in mathematics more often than pseudometrics. However pseudometrics do occasionally arise in a natural way. Moreover, many definitions and proofs actually only require using properties 1)-4). Therefore we will state our results in terms of pseudometrics when possible. But, of course, anything we prove about pseudometric spaces is automatically true for metric spaces.

Example 2.2

1) The usual metric on $\mathbb{R}$ is $d(x, y) = |x - y|$. Clearly, properties 1) - 5) are true. In fact, properties 1) -5) are deliberately chosen so that a metric imitates the usual distance function.

2) The usual metric on $\mathbb{R}^n$ is defined as follows: if $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ are in $\mathbb{R}^n$, then $d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$. You should already know that d has properties 1) - 5). But the details to verify the triangle inequality are a little tricky, so we will go through the steps. First, we prove another useful inequality.

Suppose $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$ are points in $\mathbb{R}^n$. Define

$$P(w) = \sum_{i=1}^{n}(a_i + wb_i)^2 = \sum_{i=1}^{n}a_i^2 + (2\sum_{i=1}^{n}a_ib_i)w + (\sum_{i=1}^{n}b_i^2)w^2$$

$P(w)$ is a quadratic function of w, and $P(w) \geq 0$ because $P(w)$ is a sum of squares. Therefore the equation $P(w) = 0$ has at most one real root, so it follows from the quadratic formula that

$$(2\sum_{i=1}^{n}a_ib_i)^2 - 4(\sum_{i=1}^{n}a_i^2)(\sum_{i=1}^{n}b_i^2) \leq 0, \qquad \text{which gives}$$

$$|\sum_{i=1}^{n}a_ib_i| \leq (\sum_{i=1}^{n}a_i^2)^{1/2}(\sum_{i=1}^{n}b_i^2)^{1/2}$$

This last inequality is called the Cauchy-Schwarz inequality. In vector notation it could be written in the form $|a \cdot b| \leq ||a|| \cdot ||b||$.

Then if $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n)$ and $z = (z_1, z_2, ..., z_n)$ are in $\mathbb{R}^n$, we can calculate

$$d(x, z)^2 = \sum_{i=1}^{n}(x_i - z_i)^2 = \sum_{i=1}^{n}((x_i - y_i) + (y_i - z_i))^2$$

$$= \sum_{i=1}^{n}(x_i - y_i)^2 + 2\sum_{i=1}^{n}(x_i - y_i)(y_i - z_i) + \sum_{i=1}^{n}(y_i - z_i)^2$$

$$\leq \sum_{i=1}^{n}(x_i - y_i)^2 + 2\sum_{i=1}^{n}|x_i - y_i||y_i - z_i| + \sum_{i=1}^{n}(y_i - z_i)^2$$

$$\leq \sum_{i=1}^{n}(x_i - y_i)^2 + 2(\sum_{i=1}^{n}(x_i - y_i)^2)^{1/2}(\sum_{i=1}^{n}(x_i - y_i)^2)^{1/2} + \sum_{i=1}^{n}(y_i - z_i)^2$$

$= (d(x, y) + d(y, z))^2$. Taking the square root of both sides gives

$$d(x, z) \leq d(x, y) + d(y, z).$$

Example 2.3 We can also put other "unusual" metrics on the set $\mathbb{R}^n$.

1) Let d be the usual metric on $\mathbb{R}^n$ and define $d'(x,y) = 100d(x,y)$. Then d' is also a metric on $\mathbb{R}^n$. In $(\mathbb{R}^n, d')$, the "usual" distances are stretched by a factor of 100. This is just a rescaling of distances – as if we changed the units of measurement from meters to centimeters, and that change shouldn't matter in any important way. In fact, it's easy to check that if d is any metric (or pseudometric) on a set X and $\alpha > 0$, then $d' = \alpha \cdot d$ is also a metric (or pseudometric) on X.

2) If $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n)$ are points in $\mathbb{R}^n$, define

$$d_t(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

It is easy to check that d_t satisfies properties 1) - 5) so $(\mathbb{R}^n, d_t)$ is a metric space. We call d_t the taxicab metric on $\mathbb{R}^n$. (*For $n = 2$, this means that distances are measured as if you had to move along a rectangular grid of city streets from x to y – the taxi cannot cut diagonally across a city block*).

3) If $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n)$ are points in $\mathbb{R}^n$, define

$$d^*(x,y) = \max \{|x_i - y_i| : i = 1, 2, ..., n\}$$

Then $(\mathbb{R}^n, d^*)$ is also a metric space. We will refer to d^* as the max metric on $\mathbb{R}^n$.

When $n = 1$, of course, d, d_t and d^* are exactly the same metric on $\mathbb{R}$.

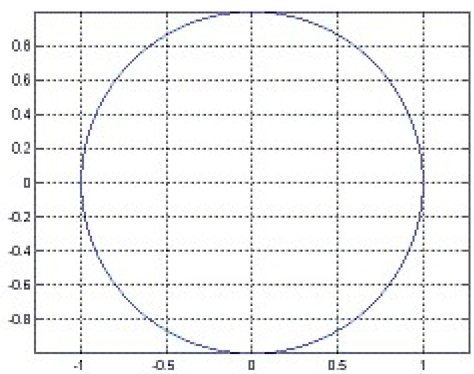
We will see later that "for topological" purposes" d', d_t, d^* are all "equivalent" metrics on $\mathbb{R}^n$. Roughly, this means that whichever of these metrics is used to measure nearness in $\mathbb{R}^n$, exactly the same functions turn out to be continuous and exactly the same sequences converge.

4) The "unit sphere" S^1 is the set of points in $\mathbb{R}^2$ that are at distance 1 from the origin. Sketch the unit sphere in $\mathbb{R}^2$ using the metrics $d,\ d_t,\ d^*$, and $d' = 100d$.

Since there are only two coordinates, we will write a point in $\mathbb{R}^2$ in the usual way as (x, y) rather than (x_1, x_2).

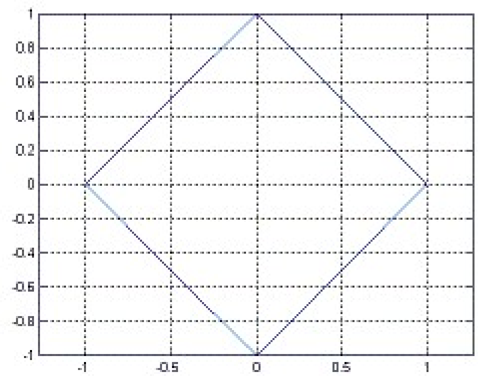
For d, we get

$$S^1 = \{(x,y) : d(\,(x,y),(0,0)) = 1\}$$
$$= \{(x,y) : x^2 + y^2 = 1\}$$



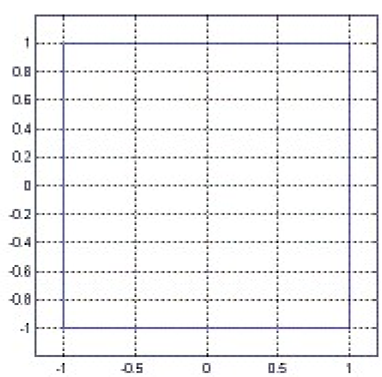
For d_t, we get

$$S^1 = \{(x,y) : d_t(\,(x,y),(0,0)) = 1\}$$
$$= \{(x,y) : |x| + |y| = 1\}$$



For d^*, we get

$$S^1 = \{(x,y) : d^*(\,(x,y),(0,0)) = 1\}$$
$$= \{(x,y) : \max\{|x|,\ |y|\} = 1\}$$



Of course for the metric $d' = 100d$, S^1 has the same shape as for the metric d, but the sphere is reduced in size by a scaling factor of $\frac{1}{100}$.

Switching among the metrics $d,\ d',\ d_t,\ d^*$ produces unit spheres in $\mathbb{R}^n$ with different sizes and shapes. In other words, changing the metric on $\mathbb{R}^n$ may cause dramatic changes in the geometry of the space – for example, "areas" may change and "spheres" may no longer be "round." Changing the metric can also affect smoothness features of the space (spheres may turn out to have sharp corners). But it turns out, as mentioned earlier, that $d,\ d',\ d_t$ and d^* are "equivalent" for "topological purposes." For topology, "size," "geometrical shape," and "smoothness" don't matter.

For working with $\mathbb{R}^n$, the usual metric d is the default – that is, we always assume that $\mathbb{R}^n$, or any subset of $\mathbb{R}^n$, has the usual metric d unless a different metric is explicitly stated.

Example 2.4 For each part, verify that d satisfies the properties of a pseudometric or metric.

1) For a set X, define $d(x, y) = 0$ for all $x, y \in X$. We call d the trivial pseudometric on X: all distances are 0. (*Under what circumstances is this* d *a metric?*)

2) For a set X, define $d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$. We call d the discrete unit metric on X.

To verify the triangle inequality: for points $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$ certainly is true if $x = z$; and if $x \neq z$, then $d(x, z) = 1$ and $d(x, y) + d(y, z) \geq 1$.

Definition 2.5 Suppose (X, d) is a pseudometric space, that $x_0 \in X$ and $\epsilon > 0$. Then $B_\epsilon(x_0) = \{x \in X : d(x, x_0) < \epsilon\}$ is called the ball of radius ϵ with center at x_0.

If there exists an $\epsilon > 0$ such that $B_\epsilon(x_0) = \{x_0\}$, then we say that x_0 is an isolated point in (X, d).

Example 2.6

1) In $\mathbb{R}$, $B_\epsilon(x_0)$ is the interval $(x_0 - \epsilon, x_0 + \epsilon)$. More generally, $B_\epsilon(x_0)$ in $\mathbb{R}^n$ is just the usual spherical ball with radius ϵ and center at x_0 (not including the boundary surface). If the metric d_t is used in $\mathbb{R}^n$, then $B_\epsilon(x_0)$ is the interior of a "diamond-shaped" region centered at x_0. (*See the earlier sketches of* S^1*: in* $(\mathbb{R}^2, d_t)$*,* $B_1((0,0))$ *is the region "inside" the diamond-shaped* S^1*.*)

In $X = [0, 1]$ with the usual metric d, then $B_{\frac{1}{2}}(0) = [0, \frac{1}{2})$, $B_1(0) = [0, 1)$, $B_2(0) = [0, 1]$.

2) If d is the trivial pseudometric on X and $x_0 \in X$, then $B_\epsilon(x_0) = X$ for every $\epsilon > 0$.

3) If d is the discrete unit metric on X, then $B_\epsilon(x_0) = \begin{cases} \{x_0\} & \text{if } \epsilon \leq 1 \\ X & \text{if } \epsilon > 1 \end{cases}$. Therefore every point x_0 in (X, d) is isolated. If we rescale and replace d by the metric αd (where $\alpha > 0$), then it is still true that every point is isolated.

4) Let $C([0, 1]) = \{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous}\}$. For $f, g \in C([0, 1])$, define

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

It is easy to check that d is a pseudometric on $C([0, 1])$. In fact d is a metric: if $f \neq g$, then there must be a point $x_0 \in [0, 1]$ where $|f(x_0) - g(x_0)| > 0$. By continuity, $|f(x) - g(x)| > 0$ for x's near x_0, that is, $|f(x) - g(x)| > 0$ on some interval $[a, b] \subseteq [0, 1]$, where $x_0 \in [a, b]$. (*carefully explain why!*). Let $m = \min_{x \in [a,b]} |f(x) - g(x)|$ (*why does* m *exist?*). Then $m > 0$, so

$$d(f, g) = \int_0^1 |f(x) - g(x)|\, dx \geq \int_a^b |f(x) - g(x)|\, dx \geq \int_a^b m\, dx = m(b - a) > 0.$$

Therefore, d is a metric on $C([0, 1])$.

$C([0,1])$ is a subset of the larger set $Y = \{f \in \mathbb{R}^{[0,1]} : f \text{ is integrable}\}$. We can define a distance function d on Y using the same formula (*). In this case, d is a pseudometric on Y but <u>not</u> a metric. For, example, let

$$f(x) = 0 \text{ for all } x, \text{ and } g(x) = \begin{cases} 0 & \text{if } x \neq \frac{1}{2} \\ 1 & \text{if } x = \frac{1}{2} \end{cases}$$

Then $f \neq g$ but $d(f,g) = \int_0^1 |f(x) - g(x)|\, dx = 0$

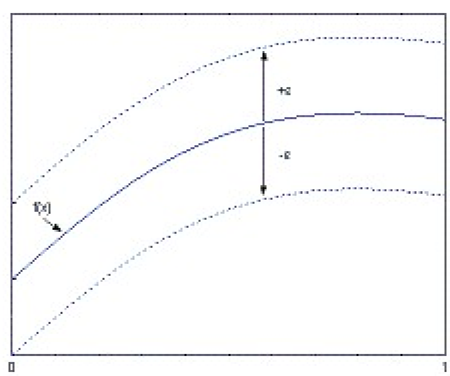
This example shows how a pseudometric that is not a metric can arise naturally in analysis.

5) On $C([0,1])$ we can also define another metric d^* by

$$\begin{aligned} d^*(f,g) &= \sup\ \{|f(x) - g(x)| : x \in [0,1]\} \\ &= \max\ \{|f(x) - g(x)| : x \in [0,1]\} \end{aligned}$$

(*Replacing "sup" with "max" makes sense because a theorem from analysis says that the continuous function* $|f - g|$ *<u>has</u> a maximum value on the closed interval* $[0,1]$.)

Then $d^*(f,g) < \epsilon$ if and only if $|f(x) - g(x)| < \epsilon$ at every point $x \in [0,1]$, so we can picture $B_\epsilon(f)$ in $(C([0,1]), d^*)$ as the set of all functions $g \in C([0,1])$ whose graph lies entirely inside a "tube of width ϵ" containing the graph of f – that is, $g \in B_\epsilon(f)$ iff g is "uniformly within ϵ of f on $[0,1]$." See the following figure.

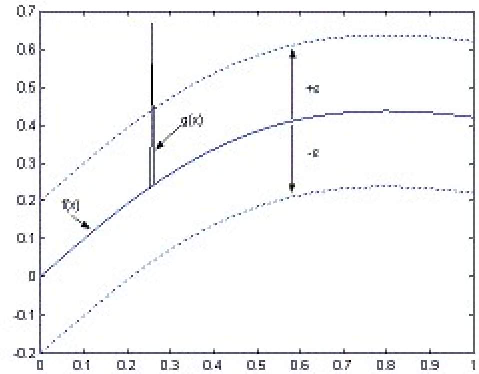


How are the metrics d and d^* from Examples 4) and 5) related? Notice that for all $f, g \in C([0,1])$:

$$d(f,g) = \int_0^1 |f(x) - g(x)|\, dx \leq \int_0^1 \max_{x \in [0,1]} |f(x) - g(x)|\, dx = \int_0^1 d^*(f,g)\, dx = d^*(f,g)\,.$$

We abbreviate this observation by writing $d \leq d^*$. It follows that $B_\epsilon^{d^*}(f) \subseteq B_\epsilon^{d}(f)$: so, <u>for a given $\epsilon > 0$, the larger metric produces the smaller ball.</u> (*Note: the superscript notation on the balls indicates which metric is being used in each case.*)

The following figure shows a function f and the graph of a function $g \in B_\epsilon^d(f) - B_\epsilon^{d^*}(f)$. The graph of g <u>coincides</u> with the graph of f, except for a tall spike: the spike takes the graph of g outside the "ϵ-tube" around the graph of f, but the spike is so thin that the $d(f,g) = \int_0^1 |f(x) - g(x)|\, dx$ $=$ "the total area between the graphs of f and g" $< \epsilon$.



6) Let $\ell_2 = \{f \in \mathbb{R}^{\mathbb{N}} : \sum_{k=1}^{\infty} f^2(k)$ converges$\}$. If we write $f(k) = x_k$ and use the more informal sequence notation, then $\ell_2 = \{(x_k) : x_k \in \mathbb{R}$ and $\sum_{k=1}^{\infty} x_k^2$ converges$\}$. Thus, ℓ_2 is the set of all "square-summable" sequences of real numbers.

Suppose $x = (x_k)$ and $y = (y_k)$ are in ℓ_2 and that $a, b \in \mathbb{R}$. We claim that the sequence $ax + by = (ax_k + by_k)$ is also in ℓ_2. To see this, look at partial sums:

$$\sum_{k=1}^{n}(ax_k + by_k)^2 = a^2\sum_{k=1}^{n}x_k^2 + 2ab\sum_{k=1}^{n}x_k y_k + b^2\sum_{k=1}^{n}y_k^2 \le a^2\sum_{k=1}^{n}x_k^2 + |2a\sum_{k=1}^{n}x_k y_k| + b^2\sum_{k=1}^{n}y_k^2$$

$$\le a^2\sum_{k=1}^{n}x_k^2 + 2|a||b|\,(\sum_{k=1}^{n}x_k^2)^{1/2}\,(\sum_{k=1}^{n}y_k^2)^{1/2} + b^2\sum_{k=1}^{n}y_k^2 \quad \textit{(by the Cauchy-Schwarz inequality)}$$

$$\le a^2\sum_{k=1}^{\infty}x_k^2 + 2|a||b|(\sum_{k=1}^{\infty}x_k^2)^{1/2}(\sum_{k=1}^{\infty}y_k^2)^{1/2} + b^2\sum_{k=1}^{\infty}y_k^2 = M \in \mathbb{R}$$ *(all the series converge because $x, y \in \ell_2$.)*

Therefore the nonnegative series $\sum_{k=1}^{\infty}(ax_k + by_k)^2$ converges because it has bounded partial sums. This means that $ax + by \in \ell_2$.

In particular, if $x, y \in \ell_2$, we now know that $x - y \in \ell_2$ so $\sum_{k=1}^{\infty}(x_k - y_k)^2$ converges. Therefore it makes sense to define $d(x, y) = (\sum_{k=1}^{\infty}(x_k - y_k)^2)^{1/2}$. You should check that d is a metric on ℓ_2.
(For the triangle inequality, notice that $(\sum_{k=1}^{n}(x_k - z_k)^2)^{1/2} \le (\sum_{k=1}^{n}(x_k - y_k)^2)^{1/2} + (\sum_{k=1}^{n}(y_k - z_k)^2)^{1/2}$ *by the triangle inequality in $\mathbb{R}^n$. Letting $n \to \infty$ gives the triangle inequality for ℓ_2.)*

7) Suppose (X_i, d_i) are pseudometric spaces $(i = 1, ..., n)$, and that $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ are points in the product $X = X_1 \times ... \times X_n$. Then each of the following is a pseudometric on X :

$$d(x, y) = (\sum_{i=1}^{n} d_i^2(x_i, y_i))^{1/2} \qquad\qquad d_t(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i)$$

$$d^*(x, y) = \max\{d_i(x_i, y_i) : i = 1, ..., n\}$$

If each d_i is a metric, then so are d, d_t, and d^*. Notice that if each $X_i = \mathbb{R}$ and each d_i is the usual metric on $\mathbb{R}$, then d, d_t, and d^* are just the usual metric, the taxicab metric, and the max metric on $\mathbb{R}^n$. As we shall see, it turns out that these metrics on X are all equivalent for "topological purposes."

Definition 2.7 Suppose (X, d) is a pseudometric space and $O \subseteq X$. We say that <u>O is open in (X, d)</u> if for each $x \in O$ there is an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq O$. (*Of course, ϵ may depend on x.*)

For example,

i) The sets $\emptyset$ and X are open in any space (X, d).

ii) The intervals (a, b), $(-\infty, a)$, (b, ∞), and $(-\infty, \infty) = \mathbb{R}$ are open in $\mathbb{R}$. (*Fortunately this terminology is consistent with the fact that these intervals are called "open intervals" in calculus books.*)

But notice that the interval (a, b), when viewed as a subset of the x-axis in $\mathbb{R}^2$, is <u>not</u> open in $\mathbb{R}^2$. Similarly, $\mathbb{R}$ is an open set in $\mathbb{R}$, but $\mathbb{R}$ (viewed as the x-axis) is <u>not</u> open in $\mathbb{R}^2$.

iii) The intervals $[a, b]$, $[a, b)$ and $(a, b]$ are not open in $\mathbb{R}$. But the sets $[a, b)$ and $[a, b]$ are open in the metric space $([a, b], d)$.

Examples ii) and iii) illustrate that "open" is not a property that depends just on the set A: whether or not a set A is open depends on the larger space in which it "lives" – that is, "open" is a relative term.

The next theorem tells us that the <u>balls in</u> (X, d) <u>are "building blocks" from which all open sets can be constructed</u>.

Theorem 2.8 A set $O \subseteq X$ is open in (X, d) if and only if O is a union of a collection of balls.

Proof If O is open, then for each $x \in O$ there is an $\epsilon_x > 0$ such that $B_{\epsilon_x}(x) \subseteq O$ and therefore we can write $O = \bigcup_{x \in O} B_{\epsilon_x}(x)$.

Conversely, suppose $O = \bigcup_{x \in C} B_{\epsilon_x}(x)$ for some indexing set $C \subseteq O$. We must show that if $y \in O$, then $B_\epsilon(y) \subseteq O$ for some $\epsilon > 0$. Since $y \in O$, we know that $y \in B_{\epsilon_{x_0}}(x_0)$ for some $x_0 \in C$. Then $d(x_0, y) = \delta < \epsilon_{x_0}$. Let $\epsilon = \frac{1}{2}(\epsilon_{x_0} - \delta) > 0$ and consider $B_\epsilon(y)$. If $z \in B_\epsilon(y)$, then
$d(z, x_0) \leq d(z, y) + d(y, x_0) < \epsilon + \delta = \frac{1}{2}(\epsilon_{x_0} - \delta) + \delta = \frac{1}{2}\epsilon_{x_0} + \frac{1}{2}\delta < \frac{1}{2}\epsilon_{x_0} + \frac{1}{2}\epsilon_{x_0} = \epsilon_{x_0}$,
so $z \in B_{\epsilon_{x_0}}(x_0)$. Therefore $B_\epsilon(y) \subseteq B_{\epsilon_{x_0}}(x_0) \subseteq O$. ●

Corollary 2.9 a) Each ball $B_\epsilon(x)$ is open in (X, d).

b) A point x_0 in a pseudometric space (X, d) is isolated iff $\{x_0\}$ is an open set.

Definition 2.10 Suppose (X, d) is a pseudometric space. The topology $\mathcal{T}_d$ generated by d is the collection of all open sets in (X, d). In other words, $\mathcal{T}_d = \{O : O \text{ is open in } (X, d)\} = \{O : O \text{is a union of balls}\}$.

Theorem 2.11 Let $\mathcal{T}_d$ be the topology in (X, d). Then

i) $\emptyset, X \in \mathcal{T}_d$
ii) if $O_\alpha \in \mathcal{T}_d$ for each $\alpha \in A$, then $\bigcup_{a \in A} O_a \in \mathcal{T}_d$
iii) if $O_1, ..., O_n \in \mathcal{T}_d$, then $O_1 \cap ... \cap O_n \in \mathcal{T}_d$.

(Conditions ii) and iii) say that the collection $\mathcal{T}_d$ is "closed under unions" and "closed under finite intersections.")

Proof $\emptyset$ is the union of the empty collection of open balls, and $X = \bigcup_{x \in X} B_1(x)$, so $\emptyset, X \in \mathcal{T}_d$.

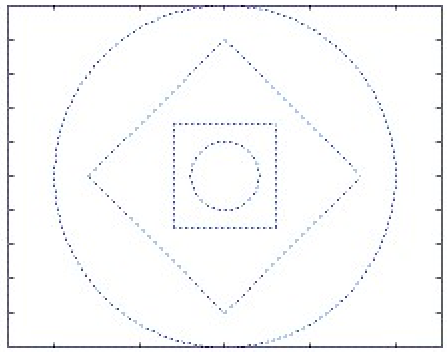
Suppose $x \in O = \bigcup_{\alpha \in A} O_\alpha$ where each $O_\alpha \in \mathcal{T}_d$. Then x is in one of these open sets, say O_{α_0}. So for some $\epsilon > 0$, $x \in B_\epsilon(x) \subseteq O_{\alpha_0} \subseteq O$. Therefore O is open, that is, $O \in \mathcal{T}_d$.

To verify iii), suppose $O_1, O_2, \ldots, O_n \in \mathcal{T}_d$ and that $x \in O_1 \cap O_2 \cap ... \cap O_n$. For each $i = 1, ..., n$, there is an $\epsilon_i > 0$ such that $x \in B_{\epsilon_i}(x) \subseteq O_i$. Let $\epsilon = \min\{\epsilon_1, \epsilon_2, ..., \epsilon_n\} > 0$. Then $B_\epsilon(x) \subseteq O_1 \cap O_2 \cap ... \cap O_n$. Therefore $O_1 \cap O_2 \cap ... \cap O_n \in \mathcal{T}_d$. •

Example 2.12 The set $O_n = (-\frac{1}{n}, \frac{1}{n})$ is open in $\mathbb{R}$ for every $n \in \mathbb{N}$. However, $\bigcap_{n=1}^{\infty} O_n = \{0\}$ is not open in $\mathbb{R}$: so an intersection of infinitely many open sets might not be open. (*Where does the proof for part iii) in Theorem 2.11 break down if we intersect infinitely many open sets?*)

Notice that different pseudometrics can produce the same topology on a set X. For example, if d is a metric on X and we set $d' = 2d$, then d and d' produce the same collection of balls (with radii measured differently): for each $\epsilon > 0$, the ball $B_\epsilon^d(x)$ is the same set as the ball $B_{2\epsilon}^{d'}(x)$. If we get the same balls from each metric, then we must also get the same open sets: $\mathcal{T}_d = \mathcal{T}_{d'}$ (*see Theorem 2.8*).

We can see a less trivial example in $\mathbb{R}^2$. Let d, d_t, and d^* be the usual metric, the taxicab metric, and the max metric on $\mathbb{R}^2$. Clearly any set which is a union of d-balls (or d^* balls) can also be written as a union of d_t-balls, and vice-versa. (*Explain why! See the following picture for* $\mathbb{R}^2$.)



Therefore all three metrics produce the same topology: $\mathcal{T}_d = \mathcal{T}_{d_t} = \mathcal{T}_{d^*}$ even though the balls are different for each metric. It turns out that the open sets in (X, d) are the most important objects from a topological point of view, so in that sense these metrics are all equivalent. (As mentioned earlier, these metrics do change the "shape" and "smoothness" of the balls and therefore these metrics are not equivalent geometrically.)

Definition 2.13 Suppose d and d' are two pseudometrics (or metrics) on a set X. We say that d and d' are equivalent (written $d \sim d'$) if $\mathcal{T}_d = \mathcal{T}_{d'}$, that is, if d and d' generate the same collection of open sets.

Example 2.14

1) If d is the discrete unit metric on X, then each singleton set $\{x\} = B_1(x)$ is a ball, so each $\{x\}$ is open – equivalently, every point x is isolated in (X, d). If $A \subseteq X$, then $A = \bigcup_{a \in A}\{a\}$ is open because A is a union of balls. Therefore $\mathcal{T}_d = \mathcal{P}(X)$, called the discrete topology on X. This is the largest possible topology on X.

If $d'(x,y) = \begin{cases} 17 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$ on a set X, then $d' \sim d$, where d is the discrete unit metric.

More generally, $\alpha d \sim d$ for any $\alpha > 0$., and all of these metrics generate the discrete topology.

2) Let $X = \{\frac{1}{n} : n \in \mathbb{N}\}$. Let d be the usual metric on X and let d' be the discrete unit metric on X. For each n, $B^d_\epsilon(\frac{1}{n}) = \{\frac{1}{n}\}$ if we choose a sufficiently small ϵ. Therefore, just as in part 1), every subset of X is open in (X, d). But every subset in (X, d') also is open, so $d \sim d'$ (even though d and d' are not constant multiples of each other).

3) Let d be the trivial pseudometric on a set X. Are there any other pseudometrics d' on X for which $d' \sim d$?

3. The topology of $\mathbb{R}$

What do the open sets in $\mathbb{R}$ look like? Since ϵ-balls in $\mathbb{R}$ are intervals of the form $(a - \epsilon, a + \epsilon)$, the open sets are precisely the sets which are unions of open intervals. But we can actually say more to make the situation even clearer. We begin by making a precise definition for the term "interval."

Definition 3.1 A subset I of $\mathbb{R}$ is convex if whenever $x \leq y \leq z$ and $x, z \in I$, then $y \in I$. A convex subset of $\mathbb{R}$ is called an interval.

It is easy to give examples of intervals in $\mathbb{R}$. The following theorem states that the obvious examples are the only examples.

Theorem 3.2 $I \subseteq \mathbb{R}$ is an interval iff I has one of the following forms (where $a < b$):

$$(-\infty, \infty),\ (-\infty, a),\ (-\infty, a],\ [a, \infty),\ (a, \infty),\ (a, b),\ [a, b),\ (a, b],\ [a, b],\ \{a\},\ \emptyset \qquad (*)$$

Proof It is clear that each of the sets in the list is convex and therefore each is an interval.

Conversely, we need to show that every interval I has one of these forms. Clearly, if $|I| \leq 1$, then $I = \emptyset$ or $I = \{a\}$. If $|I| \geq 2$ then the definition of interval implies that I must be infinite. The remainder of the proof uses the completeness property (= "least upper bound property") of $\mathbb{R}$, and the argument falls into several cases:

Case I: I is bounded both above and below. Then I has a least upper bound and a greatest lower bound: let $a = \inf I$ and $b = \sup I$. Of course a and b might or might not be in I.

a) if $a, b \in I$, we claim $I = [a, b]$
b) if $a \in I$ but $b \notin I$, we claim $I = [a, b)$
c) if $a \notin I$ but $b \in I$, we claim $I = (a, b]$
d) if $a, b \notin I$, we claim $I = (a, b)$

Case II: I is bounded below but not above. In this case, let $a = \inf I$.

a) if $a \in I$, we claim $I = [a, \infty)$
b) if $a \notin I$, we claim $I = (a, \infty)$

Case III: I is bounded above but not below. In this case, let $b = \sup I$.

a) if $b \in I$, we claim $I = (-\infty, b]$
b) if $b \notin I$, we claim $I = (-\infty, b)$

Case IV: I is not bounded above or below. In this case, we claim $I = (-\infty, \infty)$.

In all cases, the proofs are similar and use properties of sups and infs. To illustrate, we prove case Ic):

If $x \in I$, then $x \leq \sup I = b$. Also, $x \geq \inf I = a$, and because $a \notin I$, we get $x > a$. So $I \subseteq (a, b]$.

We still need to show that $(a, b] \subseteq I$, so suppose $x \in (a, b]$. Then $x > a = \inf I$, so x is not a lower bound for I. This means that there is a point $z \in I$ such that $z < x$. Then $z < x \leq b$ where $z, b \in I$, and I is convex, so $x \in I$. Therefore $(a, b] \subseteq I$, so $I = (a, b]$. ∙

Note: *We used the completeness property to prove Theorem 3.2. In fact, Theorem 3.2 is* equivalent *to the completeness property. To see this* :

> *Assume Theorem 3.2 is true and that A is a nonempty subset of $\mathbb{R}$ that has an upper bound. Let $I = \{x \in \mathbb{R} : x \le c$ for some $c \in A\}$. Then I is an interval (suppose $x \le y \le z$ where $x, z \in I$. Then $z \le c$ for some $c \in A$; therefore $y \le c$ so, by definition of I, $y \in I$.).*
>
> *Since $A \ne \emptyset$, I must be infinite. An upper bound for A must also be an upper bound for I. Since I is an interval with an upper bound, I must have one of the forms $(-\infty, b]$, $(-\infty, b)$, (a,b), $[a, b)$, $(a, b]$, $[a, b]$. Then it's not hard to check that A has a least upper bound, namely $\sup A = b$.*

This is an observation I owe to Professor Robert McDowell.

It is clear that an intersection of intervals in $\mathbb{R}$ is an interval (*why*?). But a union of intervals might not be an interval: for example, $[0, 1] \cup [2, 3]$. However, if every pair of intervals in a collection "overlap," then the union is an interval. The following theorem makes this precise.

Theorem 3.3 Suppose $\mathcal{I}$ is a collection of intervals in $\mathbb{R}$ and that for every pair $I, J \in \mathcal{I}$, we have $I \cap J \ne \emptyset$. Then $\bigcup \mathcal{I}$ is an interval. In particular, if $\bigcap \mathcal{I} \ne \emptyset$, then $\bigcup \mathcal{I}$ is an interval.

Proof Let $a, b \in \bigcup \mathcal{I}$ and suppose that $a \le x \le b$. We need to show that $x \in \bigcup \mathcal{I}$.

Pick intervals I, J in $\mathcal{I}$ with $a \in I$ and $b \in J$, and choose a point $z \in I \cap J$. If $x = z$, then $x \in \bigcup \mathcal{I}$ and we are done. Otherwise, either $z < x \le b$ or $a \le x < z$. Therefore either x is either between two points of J and so $x \in J$; or x is between two points of I, so $x \in I$. Either way, we conclude that $x \in \bigcup \mathcal{I}$. •

We can now give a more careful description of the open sets in $\mathbb{R}$.

Theorem 3.4 Suppose $O \subseteq \mathbb{R}$. O is open in $\mathbb{R}$ if and only if O is the union of a countable collection of pairwise disjoint open intervals.

Proof ($\Leftarrow$) Open intervals in $\mathbb{R}$ are open sets, and a union of any collection of open sets is open.

($\Rightarrow$) Suppose O is open in $\mathbb{R}$. For each $x \in O$, there is an open interval (ball) I such that $x \in I \subseteq O$. Let $G_x = \bigcup \{I : I$ is an open interval and $x \in I \subseteq O\}$. Then $x \in G_x \subseteq O$.
By Theorem 3.3, G_x is also an open interval (*in fact, G_x is the* largest *open interval containing x and inside O; why?*). It is easy to see that there can be distinct points $x, y \in O$ for which $G_x = G_y$. In fact, we claim that if $x, y \in O$, then either $G_x = G_y$ or $G_x \cap G_y = \emptyset$.

> If $G_x \cap G_y \ne \emptyset$, then there is a point $z \in G_x \cap G_y$. By Theorem 3.3, $G_x \cup G_y$ is an open interval, a subset of O, and containing both x and y. Therefore $G_x \cup G_y$ is a set in the collection whose union is G_x. Therefore $G_x \cup G_y \subseteq G_x$. Similarly, $G_x \cup G_y \subseteq G_y$, so $G_x = G_y$.

Removing any repetitions, we let $\mathcal{D}$ be the collection of the distinct intervals G_x that arise in this way. Clearly, $O = \cup \mathcal{D}$ and $\mathcal{D}$ is countable because the members of $\mathcal{D}$ are pairwise disjoint:

> for each $I \in \mathcal{D}$, we can pick a rational number $q_I \in I$, and these q_I's are distinct. So there can be no more $\mathcal{I}$'s in $\mathcal{D}$ than there are rational numbers. (*More formally, the function $f : \mathcal{D} \to \mathbb{Q}$ given by $f(I) = q_I$ is one-to-one.*) $\bullet$

We are now able to count the open sets in $\mathbb{R}$.

Corollary 3.5 There are exactly c open sets in $\mathbb{R}$.

Proof Let $\mathcal{T}_d$ be the usual topology on $\mathbb{R}$. We want to prove that $|\mathcal{T}_d| = c$.

For each $r \in \mathbb{R}$, the interval $(-\infty, r) \in \mathcal{T}_d$, so $|\,\mathcal{T}_d\,| \geq c$.

Let $\mathcal{I}$ be the set of all open intervals in $\mathbb{R}$. Then $|\,\mathcal{I}\,| = c$ (*why?*). For each $O \in \mathcal{T}_d$, pick a sequence $I_1, I_2, ...I_n, ... \in \mathcal{I}$ for which $O = \bigcup_{n=1}^{\infty} I_n$. (*We could also choose the I_n's to be pairwise disjoint, but that is unnecessary in this argument – the important thing here is that there are only countably many I_n's.*) Then we have a function $f : \mathcal{T}_d \to \mathcal{I}^{\mathbb{N}}$ given by $f(O) = (I_1, I_2, ..., I_n, ...)$. The function f is clearly one-to-one, so $|\,\mathcal{T}\,| \leq |\,\mathcal{I}^{\mathbb{N}}\,| = c^{\aleph_0} = c$. $\bullet$

Exercises

E1. The following two statements refer to a metric space (X, d). Either prove or give a counterexample for each statement. (*These statements illustrate the danger of assuming that familiar features of* $\mathbb{R}^n$ *necessarily carry over to arbitrary pseudometric spaces.*)

a) $B_\epsilon(x) = B_\epsilon(y)$ implies $x = y$ (i.e., "a ball can't have two centers")

b) The diameter of $B_\epsilon(x)$ must be bigger than ϵ. (*The diameter of a set A in a metric space is defined to be sup* $\{d(x, y) : x, y \in A\} \leq \infty$.)

E2. The "taxicab" metric on $\mathbb{R}^2$ is defined by $d_t((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$. Draw the set of points in $(\mathbb{R}^2, d_t)$ that are equidistant from $(0, 0)$ and $(3, 4)$.

E3. Suppose (X, d) is a metric space.

a) Define $d^*(x, y) = \min\{1, d(x, y)\}$. Prove that d^* is also a metric on X, and that $\mathcal{T}_d = \mathcal{T}_{d^*}$.

b) Define $d^{**}(x, y) = \frac{d(x,y)}{1+d(x,y)}$. Prove that d^{**} is also a metric on X and that $\mathcal{T}_d = \mathcal{T}_{d^{**}}$

Hint: Let d *be a metric on* X *and suppose* f *is a function from the nonnegative real numbers to the nonnegative real numbers for which:* $f(0) = 0$, $x \leq y \Rightarrow f(x) \leq f(y)$, *and* $f(x + y) \leq f(x) + f(y)$ *for all nonnegative* x, y. *Prove that* $d'(x, y) = f(d(x, y))$ *is also a metric on* X. *Then consider the particular function* $f(x) = \frac{x}{1+x}$.

Note: For all x, y in X, $d^*(x, y) \leq 1$ and $d^{**}(x, y) \leq 1$ – that is, d^* and d^{**} are bounded metrics on X. Thus any metric d on X can be replaced by an equivalent bounded metric – that is, a bounded metric that generates the same topology. "Boundedness" is a property determined by the particular metric, not by the topology.

E4. Suppose $f : \mathbb{R} \to \mathbb{R}$. Let d be the usual metric on $\mathbb{R}$, and d' the usual metric on $\mathbb{R}^2$. Define a new distance function d'' on $\mathbb{R}$ by $d''(x, y) = d'((x, f(x)), (y, f(y)))$. Prove that d'' is a metric on $\mathbb{R}$.

Must d'' be equivalent to d ? If not, can you (*precisely, or informally*) describe conditions which will guarantee that $d'' \sim d$?

E5. Suppose a function $d : X \times X \to \mathbb{R}$ satisfies conditions 1), 2), 3), and 5) in the definition of a metric, but that the triangle inequality is replaced by: $x, y, z \in X$

$$d(x, z) \geq d(x, y) + d(y, z).$$

Prove that $|X| \leq 1$.

E6. Suppose A is a finite open set in a metric space (X, d). Prove that every point of A is isolated in (X, d).

E7. Suppose (X, d) is a metric space and $x \in X$. Prove that the following two statements are equivalent:

i) x is not an isolated point of X

ii) every open set that contains x is infinite.

E8. The definition of an open set in (X,d) reads: O is open if for all $x \in O$, there is an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq O$. In this definition, ϵ may depend on x.

Suppose we define O to be <u>uniformly open</u> if there is an $\epsilon > 0$ such that for all $x \in O$, $B_\epsilon(x) \subseteq O$ – that is, the same ϵ works for every $x \in O$. (*"Uniformly open" is <u>not</u> a standard term.*)

a) What are the uniformly open subsets of $\mathbb{R}^n$?
b) What are the uniformly open sets in (X,d) if d is the trivial pseudometric?
c) What are the uniformly open sets in (X,d) if d is the discrete unit metric?

E9. Let p be a fixed prime number. We define the <u>p-adic absolute value</u> $|\ |_p$ (sometimes called the p-adic norm) on the set of rational numbers $\mathbb{Q}$ as follows:

If $0 \neq x \in \mathbb{Q}$, write $x = \frac{p^k m}{n}$ for integers k,m,n, where p does not divide m or n, and define $|x|_p = p^{-k} = \frac{1}{p^k}$. (*Of course, k may be negative.*). Also, define $|0|_p = 0$.

Prove that $|\ |_p$ "behaves the way an absolute value (norm) should" – that is, for all $x, y \in \mathbb{Q}$

a) $|x|_p \geq 0$ and $|x|_p = 0$ iff $x = 0$
b) $|xy|_p = |x|_p \cdot |y|_p$
c) $|x+y|_p \leq |x|_p + |y|_p$

$|\ |_p$ actually satisfies an inequality <u>stronger</u> than the one in part c). Prove that

d) $|x+y|_p \leq \max\{|x|_p, |y|_p\} \leq |x|_p + |y|_p$

Whenever we have an absolute value (norm), we can use it to define a distance function:

for $x, y \in \mathbb{Q}$, let $d_p(x,y) = |x-y|_p$

e) Prove that d_p is a metric on $\mathbb{Q}$, and show that d_p actually satisfies an inequality <u>stronger</u> than the usual triangle inequality, namely:

for all $x, y, z \in \mathbb{Q}$, $d_p(x,z) \leq \max\{d_p(x,y), d_p(y,z)\}$

f) Give a specific example for x, y, z, p for which

$d_p(x,z) < \max\{d_p(x,y), d_p(y,z\})$

(*Hint: It might be convenient to be able to refer to the exponent "k" associated with a particular x. If $x = \frac{p^k m}{n}$, then k roughly refers to the "number of p's that can be factored out of x" – so we can call $k = \nu(x)$. Prove that $\nu(a-b) \geq \min\{\nu(a), \nu(b)\}$ whenever $a,b \in \mathbb{Q}$ with $a,b \neq 0$ and $a \neq b$. Note that strict inequality can occur here: for example, when $p = 3$, $\nu(8) = \nu(2) = 0$, but $\nu(8-2) = \nu(6) = 1$.*)

g) Suppose $p = 2$. Calculate $d_2(2^n, 0)$. What are $\lim_{n\to\infty} d_2(2^n, 0)$ and $\lim_{n\to\infty} d_2(4^n, 0)$?

4. Closed Sets and Operators on Sets

Definition 4.1 Suppose (X, d) is a pseudometric space and that $F \subseteq X$. We say that F is closed in (X, d) if $X - F$ is open in (X, d).

From the definitions:

F is closed in (X, d)
iff $X - F$ is open in (X, d)
iff for all $x \in X - F$, there is an $\epsilon > 0$ for which $B_\epsilon(x) \subseteq X - F$
iff for all $x \in X - F$, there is an $\epsilon > 0$ for which $B_\epsilon(x) \cap F = \emptyset$.

The open sets in X completely determine the closed sets, and vice-versa. This means that, in some sense, the topology $\mathcal{T}_d$ (the collection of open sets) and the collection of closed sets contain exactly the same "information" about a space (X, d).

The close connection between the closed sets and the open sets is reflected in the following theorem.

Theorem 4.2 In any pseudometric space (X, d),

i) $\emptyset$ and X are closed
ii) if F_α is closed for each $\alpha \in A$, then $\bigcap_{\alpha \in A} F_\alpha$ is closed
("an intersection of closed sets is closed")
iii) if $F_1, ..., F_n$ are closed, then $\bigcup_{i=i}^{n} F_i$ is closed
("a finite union of closed sets is closed")

Proof These statements follow from the corresponding properties of open sets just by taking complements. Since $\emptyset$ and X are open, their complements $X - \emptyset = X$ and $X - X = \emptyset$ are closed.

Suppose F_α is closed for each $\alpha \in A$. Then each set $X - F_\alpha$ is open, so $\bigcup_{\alpha \in A}(X - F_\alpha)$ is open, and therefore its complement $X - (\bigcup_{\alpha \in A}(X - F_\alpha)) = \bigcap_{\alpha \in A} X - (X - F_\alpha) = \bigcap_{\alpha \in A} F_\alpha$ is closed.

The proof of part iii) is similar to that for part ii) and uses the fact that a finite intersection of open sets is open. ●

Exercise: Give an example to show that an infinite union of closed sets might not be closed.

Example 4.3

1) The interval $[0, 1]$ is closed in $\mathbb{R}$ because its complement $\mathbb{R} - [0, 1] = (-\infty, 0) \cup (1, \infty)$ is open. Equivalently, we could say that $[0, 1]$ is closed because for each $x \notin [0, 1]$, there is an $\epsilon > 0$ for which $B_\epsilon(x) \cap [0, 1] = (x - \epsilon, x + \epsilon) \cap [0, 1] = \emptyset$.

2) A set might be neither open nor closed: as examples, consider the following subsets of $\mathbb{R}$:

$$[0, 1) \qquad \mathbb{Q} \qquad \mathbb{P}$$

3) A set can be both open and closed – that is, these terms are not mutually exclusive. Such sets in (X, d) are called clopen sets. For example, $\emptyset$ and X are clopen in every pseudometric space (X, d). Sometimes X contains other clopen sets and sometimes not. For example:

a) in $\mathbb{R}$, for example, $\emptyset$ and $\mathbb{R}$ are the only clopen subsets. (*This fact is not too hard to prove but it is also not obvious – the proof depends on the completeness property* (= *"least upper bound property"*) *in* $\mathbb{R}$. *We will prove this fact later when we need it in Chapter V.*)

b) in the space $X = [0, 1] \cup [3, 4]$ with the usual metric d, the set $[0, 1]$ is clopen.

4) In any pseudometric space (X, d), the set $F = \{x \in X : d(a, x) \leq \epsilon\}$ is a closed set. To see this, suppose $y \notin F$, Then $d(a, y) = \delta > \epsilon$. Let $\epsilon_1 = \delta - \epsilon > 0$. Then $B_{\epsilon_1}(y) \cap F = \emptyset$. (If $z \in B_{\epsilon_1}(y) \cap F$, then we would have $d(a, y) \leq d(a, z) + d(z, y) < \epsilon + \epsilon_1 = \epsilon + (\delta - \epsilon) = \delta$, which is false.)

$\{x \in X : d(a, x) \leq \epsilon\}$ is called the closed ball centered at x with radius ϵ.

For example, $[0, 1] = \{x \in \mathbb{R} : d(\frac{1}{2}, x) = |x - \frac{1}{2}| \leq \frac{1}{2}\}$ is the "closed ball" centered at $\frac{1}{2}$.

5) Let $X = [0, 1] \cup [2, 5)$ with the usual metric d.

Both $[0, 1]$ and $[2, 5)$ are clopen sets in X
but $[3, 4)$ is neither open nor closed in X

Notice again that "open" and "closed" are not absolute terms: whether a set A is open (or closed) is relative to the larger space X that contains A.

6) Let d be the discrete unit metric. Then $\mathcal{T}_d$ is the discrete topology and every subset of X is open. Then every subset of X is clopen.

7) Let d be the trivial pseudometric on X. For every $\epsilon > 0$ and every $x \in X$, the ball $B_\epsilon(x) = X$. In this space, a union of balls must be either X or $\emptyset$ (for the union of an empty collection of balls). Therefore $\mathcal{T}_d = \{\emptyset, X\}$. This is called the trivial topology on X.

Since $\emptyset$ and X must be open in any space (X, d), the trivial topology is the smallest possible topology on X. In (X, d), the only closed sets are $\emptyset$ and X.

Using open and closed sets in (X, d), we can define some useful "operators" on subsets of X. An "operator" creates a new subset of X from an old one.

Definition 4.4 Suppose (X, d) is a pseudometric space and $A \subseteq X$.

the interior of A in X	$= \text{int}_X\, A$	$= \bigcup\{O : O \text{ is open and } O \subseteq A\}$
the closure of A in X	$= \text{cl}_X A$	$= \bigcap\{F : F \text{ is closed and } F \supseteq A\}$
the frontier (or boundary) of A in X	$= \text{Fr}_X A$	$= \text{cl}_X A \cap \text{cl}_X(X - A)$

We will omit the subscript "X" when the context makes clear the space X in which the operations are being performed. Sometimes int A and cl A are denoted A° and $\overline{A}$ respectively. Some books use the notation ∂A for Fr A, but the symbol ∂A has a different meaning in algebraic topology so we will avoid using it here.

Theorem 4.5 Suppose (X, d) is a pseudometric space and that $A \subseteq X$. Then

1) a) int A is the largest open subset of A (that is, if O is open and $O \subseteq A$, then $O \subseteq$ int A).
   b) A is open iff $A =$ int A (since int $A \subseteq A$, the equality is equivalent to $A \subseteq$ int A).
   c) $x \in$ int A iff there is an open set O such that $x \in O \subseteq A$
   iff $\exists\, \epsilon > 0$ such that $B_\epsilon(x) \subseteq A$.

   Informally, we can think of 1c) as saying that the interior of A consist of those points "comfortably inside" A ("surrounded by a small cushion"). These are the points not "on the boundary of A."

2) a) cl A is the smallest closed set that contains A (that is, if F is closed and $F \supseteq A$, then $F \supseteq$ cl A)
   b) A is closed iff $A =$ cl A (since $A \subseteq$ cl A, the equality is equivalent to cl $A \subseteq A$).
   c) $x \in$ cl A iff for every open set O containing x, $O \cap A \neq \emptyset$
   iff for every $\epsilon > 0$, $B_\epsilon(x) \cap A \neq \emptyset$.

   Informally, 2c) states that cl A consists of the points in X that can be approximated arbitrarily closely by points from within the set A.

3) a) Fr A is closed and Fr A = Fr $(X - A)$.
   b) A is clopen iff Fr $A = \emptyset$.
   c) $x \in$ Fr A iff for every open set O containing x, $O \cap A \neq \emptyset$ and $O \cap (X - A$

   Informally, 3c) states that Fr A consists of those points in X that can be approximated arbitrarily closely both by points from within A and by points from outside A.

Proof 1) a) int A is a union of open sets, so it is open and, by definition int $A \subseteq A$.
If O is open and $O \subseteq A$, then O is one of the sets whose union gives int A, so $O \subseteq$ int A.

b) If $A =$ int A, then A is open.
Conversely, if A is open, then A is clearly the largest open subset of A, so $A =$ int A.

c) Since int A is a union of open sets, it is clear that $x \in$ int A iff $x \in O \subseteq A$ for some open set O. Since a open set is open iff it is a union of ϵ-balls, the remainder of the assertion is obviously true.

2) Exercise

3) a) $\text{Fr}\, A$ is closed because it is an intersection of two closed sets, and

$$\text{Fr}\,(X - A) = \text{cl}(X - A) \cap \text{cl}(X - (X - A)) = \text{cl}(X - A) \cap \text{cl}(A) = \text{Fr}\, A.$$

b) If A is clopen, then so is $X - A$. Therefore $\text{Fr}(A) = \text{cl}\,(A) \cap \text{cl}(X - A)$
$= A \cap (X - A) = \emptyset$.
Conversely, if $\text{Fr}\, A = \text{cl}\,(A) \cap \text{cl}(X - A) = \emptyset$, then $\text{cl}\, A \subseteq X - \text{cl}(X - A)$
$\subseteq X - (X - A) = A$, so A is closed.
Similarly we show that $X - A$ is closed, so A is clopen.

c) $x \in \text{Fr}\, A$ iff x in both $\text{cl}\, A$ and $\text{cl}\,(X - A)$. By 2c), this is true iff each open set containing x intersects both A and $X - A$.
Since an open set is a union of ϵ-balls, the remainder of the assertion is clearly true. ∙

Notice that in each part of Theorem 4.5, item c) gives you a criterion that uses only the open sets (not ϵ-balls!) to decide whether or not $x \in \text{int}\, A$, $x \in \text{cl}\, A$, or $x \in \text{Fr}\, A$. *This is important!* It means that if we change the d to an equivalent pseudometric d', then int A, cl A, and Fr A do not change, since d and d' produce the same open sets. In other words, we can say that int, cl, and Fr are topological operators: they depend on only the topology, and not on the particular metric that produces the topology. For example if $A \subseteq \mathbb{R}^n$, then A will have the same interior, same closure, and same frontier whether we measure distances using the usual metric d, the taxicab metric d_t, or the max-metric d^*.

Example 4.6 (*Be sure you understand each statement!*)

1) In $\mathbb{R}$:

$\text{int}\, \mathbb{R} = \mathbb{R}$ $\quad$ $\text{cl}\, \mathbb{R} = \mathbb{R}$ $\quad$ $\text{Fr}\, \mathbb{R} = \emptyset$
$\text{int}\, \mathbb{Q} = \emptyset$ $\quad$ $\text{cl}\, \mathbb{Q} = \mathbb{R}$ $\quad$ $\text{Fr}\, \mathbb{Q} = \mathbb{R}$

$\text{int}\,[0, 1) = (0, 1)$ $\quad$ $\text{cl}\,[0, 1) = [0, 1]$
$\text{Fr}\,[0, 1) = \{0, 1\}$ $\quad$ $\text{Fr}\,(\text{Fr}\, \mathbb{Q}) = \text{Fr}\,(\mathbb{R}) = \emptyset$.

In any space (X, d), it is obviously true that $\text{int}(\text{int}\, A) = \text{int}\, A$ and $\text{cl}(\text{cl}\, A) = \text{cl}\, A$. But this need not be true for Fr, as the last example shows. (*It is true that $Fr(Fr(Fr\, A)) = Fr(Fr\, A)$ in any space X. However, this is not a useful fact, and it is also not very interesting to prove.*)

2) $X = [0, 2)$ (with the usual metric)

$\text{cl}_X\,[0, 1) = [0, 1]$ $\quad$ $\text{int}_X[0, 1) = [0, 1)$
$\text{cl}_X\,[1, 2) = [1, 2)$ $\quad$ $\text{int}_X[1, 2) = (1, 2)$

$\text{Fr}_X\,[0, 1) = [0, 1] \cap [1, 2) = \{1\}$. *Note that $0 \notin Fr_X[0, 1)$ because 0 cannot be "approximated arbitrarily closely" by points from $X - [0, 1)$.*

3) Suppose d is the discrete unit metric on X. If $A \subseteq X$, then A is clopen so we get $\text{cl}\, A = A$, $\text{int}\, A = A$, and $\text{Fr}\, A = \emptyset$.
On the other hand, d is the trivial pseudometric on X. If A is any nonempty, proper subset of X, then $\text{cl}\, A = X$, $\text{int}\, A = \emptyset$, and $\text{Fr}\, A = X$.

4) In (ℓ_2, d), let A be the set of sequences with all terms rational:

$$A = \{x = (x_i) \in \ell_2 : \forall i,\ x_i \in \mathbb{Q}\} = \mathbb{Q}^{\mathbb{N}} \cap \ell_2.$$

We claim that cl $A = \ell_2$ — in other words, that any point $y = (y_i) \in \ell_2$ can be approximated arbitrarily closely by a point from A. So let $\epsilon > 0$. We must show that $B_\epsilon(y) \cap A \neq \emptyset$.

Since $\sum_{i=1}^{\infty} y_i^2$ converges, we can pick an N such that $\sum_{i=N+1}^{\infty} y_i^2 < \frac{\epsilon^2}{2}$. Using this value of N, choose rational numbers a_i so that $|a_i - y_i| < \frac{\epsilon}{\sqrt{2N}}$ for $i = 1, ..., N$.
Define $a = (a_1, ..., a_N, 0, 0, 0, ...)$. Then $a \in A$ and

$$d(a,y) = \sqrt{\sum_{i=1}^{\infty}(a_i - y_i)^2} = \sqrt{\sum_{i=1}^{N}(a_i - y_i)^2 + \sum_{i=N+1}^{\infty}(a_i - y_i)^2}$$

$$= \sqrt{\sum_{i=1}^{N}(a_i - y_i)^2 + \sum_{i=N+1}^{\infty} y_i^2} < \sqrt{N \cdot \frac{\epsilon^2}{2N} + \frac{\epsilon^2}{2}} = \sqrt{\epsilon^2} = \epsilon.$$

Therefore $a \in B_\epsilon(y) \cap A$.

We also claim that int $A = \emptyset$. To prove this, we need to show that if $x = (x_i) \in A$, then no ball centered at x is a subset of A. To see this, pick any $\epsilon > 0$ and choose an irrational y_1 such that $|y_1 - x_1| < \epsilon$. Create a new point y from x by changing x_1 to y_1 so that $y = (y_1, x_2, ..., x_n, ...)$. Clearly, $y \in \ell_2$, $d(x,y) = |x_1 - y_1| < \epsilon$ and $y \notin A$, so $B_\epsilon(x) \not\subseteq A$.

What is Fr A ?

5) In any pseudometric space (X,d), $B_\epsilon(a)$ is a subset of the closed set $\{x \in X : d(a,x) \le \epsilon\}$. Therefore cl $B_\epsilon(a) \subseteq \{x \in X : d(a,x) \le \epsilon\}$.

But these two sets are not necessarily equal: sometimes the closed ball is larger than the closure of the open ball $B_\epsilon(a)$! For example, suppose d is the discrete unit metric on a set X where $|X| > 1$. Then

$$\{a\} = B_1(a) = \text{cl}\, B_1(a) \underset{\neq}{\subset} X = \{x \in X : d(a,x) \le 1\}$$

Definition 4.7 Let (X,d) be a pseudometric space and $D \subseteq X$. We say that D is dense in (X,d) if cl $D = X$. The space (X,d) is called separable if it possible to find a countable dense set D in X. (*Note the spelling: "separable," not "seperable."*)

Notice that "separable" is defined in terms of closure "cl" and the closure operator depends only on the topology, not the particular metric that generates the topology. Therefore separability is a topological property: if (X,d) is separable and $d' \sim d$, then (X,d') is also separable.

More informally, "D is dense in X" means that each point $x \in X$ can be approximated arbitrarily closely by a point from D. If X is uncountable, then a countable dense set D in X (if one exists) is a relatively small set which we can use to approximate any point in X arbitrarily closely.

Example 4.8

1) $\mathbb{R}^n$ is separable because $\mathbb{Q}^n$ is a countable dense set in $\mathbb{R}^n$; in particular, $\mathbb{Q}$ is a countable dense set in $\mathbb{R}$, so $\mathbb{R}$ is separable. $\mathbb{P}$ is an example of an <u>uncountable</u> dense subset of $\mathbb{R}$.

2) If X is countable, then (X, d) is automatically separable, because X is dense in X.

3) Suppose $\mathcal{T}_d$ is the discrete topology on X. Then every subset of X is closed, so a subset D is dense iff $D = X$. Therefore (X, d) is separable iff X is countable.

Suppose $\mathcal{T}_d$ is the trivial topology on X. Then cl $D = X$ for <u>every</u> nonempty subset D since the only closed set containing D is X. So every nonempty subset D is dense and therefore (X, d) is separable – because, for example, any one-point set $\{x\}$ is dense.

4) The set $A = \mathbb{Q}^{\mathbb{N}} \cap \ell_2$ is dense in ℓ_2 (*see Example 4.6(4)*). This set A is <u>uncountable</u> because every sequence of rationals (x_i) with $|x_i| \leq \frac{1}{i}$ is in A (*why?*), and there are c such sequences (*why?*). However, (ℓ_2, d) <u>is</u> separable. Can you find a <u>countable</u> dense set D?
(*The computation in Example 4.6.4 might give you an idea.*)

Definition 4.9 Let (X, d) be a pseudometric space and suppose A, B are nonempty subsets of X. We define the distance between A and B by

$$\text{dist}(A, B) = \inf\{d(a, b) : a \in A \text{ and } b \in B\}.$$

We usually abbreviate dist(A, B) by $d(A, B)$ even though this is an "abuse of notation."

If $A \cap B \neq \emptyset$, then clearly $d(A, B) = 0$. But <u>notice</u>: if $A \cap B = \emptyset$, you <u>cannot</u> conclude that $d(A, B) > 0$, <u>not even</u> if A and B are both closed sets. For example, let A be the y-axis in $\mathbb{R}^2$ and $B = \{(x, y) \in \mathbb{R}^2 : y = \frac{1}{x}\}$. Then A and B are disjoint, <u>closed</u> sets but $d(A, B) = 0$.

We also go one step further and abbreviate $d(\{a\}, B)$ as $d(a, B) =$ "the distance from a to set B."

If $a \in B$, then clearly $d(a, B) = 0$. But the converse may not be true. For example, let $B = \{1, \frac{1}{2}, ..., \frac{1}{n}, ...\} \subseteq \mathbb{R}$. Then $d(0, B) = 0$ even though $0 \notin B$.

We can use "distance from a point to a set" to describe the closure of a set.

Theorem 4.10 Suppose (X, d) is a pseudometric space and that $A \subseteq X$. Then $x \in \text{cl}\, A$ iff $d(x, A) = 0$.

Proof $x \in \text{cl}\, A$ iff for every $\epsilon > 0$, $B_\epsilon(x) \cap A \neq \emptyset$
iff for every $\epsilon > 0$ there is a $y \in A$ with $d(x, y) < \epsilon$
iff $d(x, A) = 0$. •

Exercises

E10. Let (X, d) be a pseudometric space. Prove or disprove each statement.

a) $B_\epsilon(x)$ is never a closed set.
b) If $A \subseteq X$, then Fr $A = \text{cl } A - \text{int } A$.
c) For any $A \subseteq X$, diam $(A) = \text{diam}(\text{cl } A)$.
(*The diameter of a set A in a metric space is defined to be sup* $\{d(x, y) : x, y \in A\} \leq \infty$.)
d) For any $A \subseteq X$, diam $(A) = \text{diam}(\text{int } A)$.
e) For any $A \subseteq X$, $\text{int}(A \cup B) = \text{int } A \cup \text{int } B$.
f) For every $x \in X$ and $\epsilon > 0$, cl $(B_\epsilon(x)) = \{y \in X : \ d(x, y) \leq \epsilon\}$.

E11. a) Give an example of a metric space (X, d) that contains a proper nonempty clopen subset.
b) Give an example of a metric space (X, d) and a subset that is neither open nor closed.
c) Give an example of a metric space (X, d) and a subset A for which every point in A is a limit point of A. (*Note: a point x is called a limit point of a set A if, for every open set O containing x, $O \cap (A - \{x\}) \neq \emptyset$.*)
d) Give an example of a metric space (X, d) and a nonempty subset A such that every point is a limit point of A but $\text{int}(A) = \emptyset$. Can you also arrange that A is closed in X?
e) For each of the following subsets of $\mathbb{R}$, find the interior, closure and frontier ("boundary") in $\mathbb{R}$. Which points of the set are isolated in $\mathbb{R}$? which points of the set are isolated in the set ?

i) $A = \{m + n\pi : m, n \in \mathbb{N}\}$
ii) $B = \{\frac{1}{m} + \frac{1}{n} : m, n \in \mathbb{Z}\}$

E12. An infinite union of closed sets need not be closed. However if infinitely many closed sets are "spread out" enough from each other, then their union is closed. Parts a) and b) illustrate this.

a) Suppose that for each $n \in \mathbb{N}$, F_n is a closed set in $\mathbb{R}$ and that $F_n \subseteq (n, n+1)$.
Prove that $\bigcup_{n=1}^{\infty} F_n$ is closed in $\mathbb{R}$.

b) More generally, suppose F_α is a closed set in (X, d) for each α in some index set A; suppose also that for each point $x \in X$ there is an $\epsilon > 0$ such that $B_\epsilon(x) \cap F_\alpha = \emptyset$ for all but at most finitely many α's.
Prove that $\bigcup_{\alpha \in A} F_\alpha$ is closed in (X, d). (*Notice that b) $\Rightarrow$ a). Why?*)

E13. a) Give an example of 2^c subsets of $\mathbb{R}$ all of which have the same closure. Do the same in $\mathbb{R}^2$.

b) Prove or disprove: there exist 2^c subsets of $\mathbb{R}$ such that any two have different closures.
Is the situation the same in $\mathbb{R}^2$?

E14. The Hilbert cube, H, is a subset of ℓ_2: $H = \{x \in \ell_2 : \ |x_i| \leq \frac{1}{i}\}$. Prove that H is closed in ℓ_2. Prove or disprove: H is open in ℓ_2.

E15. In (X, d), a subset A is called a G_δ set if A can be written as a countable intersection of open sets; and A is called an F_σ set if A can be written as a countable union of closed sets.

> *Note: Open sets are often denoted using letters like O, U, or V (from open, and from the French ouvert), and sometimes by the letter G —from older literature where the German word was "Gebiet." Closed sets often are denoted by the letter F—from the French "ferme." Preferring these particular letters, of course, is just a common tradition – but many topologists follow it and most would wince to read something like "let F be an open set."*
>
> *The names G_δ and F_σ go back to the classic book Mengenlehre by the German mathematician Felix Hausdorff. The σ and the δ in the notation come from the German words used for union and intersection: Summe and Durchschnitt.*

a) Prove that every closed set in a pseudometric space (X, d) is a G_δ set, and every open set is an F_σ set.

b) Prove that the set of irrationals, $\mathbb{P}$, is a G_δ set in $\mathbb{R}$.

c) Find the error in the following argument which "proves" that every subset of $\mathbb{R}$ is a G_δ set:

> *Let $A \subseteq \mathbb{R}$. For $x \in A$, let $J_n = \bigcup\{B_{\frac{1}{n}}(x) : x \in A\}$. J_n is open for each $n \in \mathbb{N}$. Since $\{x\} = \bigcap_{n=1}^{\infty} B_{\frac{1}{n}}(x)$, it follows that $A = \bigcap_{n=1}^{\infty} J_n$, so A is a countable intersection of open sets, that is, A is a G_δ set.*

c) In part b), the truth is that $\bigcap_{n=1}^{\infty} J_n = ?$

d) Suppose we list the members of $\mathbb{Q}$: $x_1, x_2, ..., x_n, ...$. Let $J = \bigcup_{n=1}^{\infty} B_{\frac{1}{n}}(x_n)$ where, of course, $B_{\frac{1}{n}}(x_n)$ is the interval $(x_n - \frac{1}{n}, x_n + \frac{1}{n}) \subseteq \mathbb{R}$. Is $J = \mathbb{R}$?

E16. Let (X, d) be a pseudometric space. Suppose that for every $\epsilon > 0$, there exists a countable subset D_ϵ of X with the following property: $\forall x \in X$, $\exists y \in D_\epsilon$ such that $d(x, y) < \epsilon$. Prove that (X, d) is separable.

E17. Suppose that X is an uncountable set and d is any metric on X which produces the discrete topology. (*Such a metric d might not be a constant multiple of the discrete unit metric: compare Example 2.14.2*). Show that for some $\epsilon > 0$ there is an uncountable subset A of X such that $d(x, y) \geq \epsilon$ for all $x \neq y \in A$.

E18. Let (X, d) be an infinite metric space. Prove that there exists an open set U such that both U and $X - U$ are infinite. *Hint: Consider a non-isolated point, if one exists.*

E19. A metric space (X, d) is called extremally disconnected if the closure of every open set is open. (*Note: "extremally" is the correct spelling; it is not the same as the everyday word "extremely."*)

Prove that if (X, d) is extremally disconnected, then the topology $\mathcal{T}_d$ is the discrete topology.

E20. In Definition 4.9, the function "dist" provides a measure of "distance" between nonempty subsets of (X, d). Is $(\mathcal{P}(X) - \{\emptyset\}, \text{dist})$ a metric (or pseudometric) space?

E21. Suppose (x_n) is a sequence in (X, d). We say that x_0 is a cluster point of (x_n) if for every open set O containing x_0 and for all $n \in \mathbb{N}$, $\exists k > n$ such that $x_k \in O$. (*This is clearly equivalent to saying that* $\forall \epsilon > 0$ *and* $\forall n \in \mathbb{N}$, $\exists k > n$ *such that* $x_k \in B_\epsilon(x_0)$.) Informally, x_0 is a cluster point of (x_n) if the sequence is "frequently in every open set containing x_0."

a) Show that there is a sequence in $\mathbb{R}$ for which every real number r is a cluster point.

b) A neurotic mathematician is walking along $\mathbb{R}$ from 0 toward 1. Halfway to 1, she remembers that she forgot something at 0 and starts back. Halfway back to 0, she decides to go to 1 anyway and turns around, only to change her mind again after traveling half the remaining distance to 1. She continues in this back-and-forth fashion forever. Find the cluster point(s) of the sequence (x_n), where x_n is the point where she reverses direction for the n^{th} time.

5. Continuity

Suppose $a \in A \subseteq \mathbb{R}$ and that $f : A \to \mathbb{R}$. In elementary calculus, the set A is usually an interval, and the idea of continuity at a point a in A is introduced very informally. Roughly, it means that "if x is a point in the domain and near a, then $f(x)$ is near $f(a)$." In advanced calculus or analysis, the idea of "continuity of f at a" is defined carefully. The intuitive version of continuity – stated in terms of "nearness" – is made precise by measuring distances:

f is continuous at a means:

$$\forall \epsilon > 0 \;\; \exists \delta > 0 \text{ such that if } x \in A \text{ and } |x - a| < \delta, \text{ then } |f(x) - f(a)| < \epsilon.$$

The important thing to notice here is that we use the distance function in $\mathbb{R}$ to write the definition: $|x - a| = d(x, a)$ and $|f(x) - f(a)| = d(f(x), f(a))$. Since we have a way to measure distances in pseudometric spaces, we can make an a completely analogous definition of continuity for functions from one pseudometric space to another.

Definition 5.1 Let $f : X \to Y$, where (X, d) and (Y, s) are pseudometric spaces, and $a \in X$. We say that f is continuous at a if:

$$\forall \epsilon > 0 \;\; \exists \delta > 0 \text{ such that if } x \in X \text{ and } d(a, x) < \delta, \text{ then } s(f(x), f(a)) < \epsilon.$$

Notice that the sets X and Y may have completely unrelated metrics d and s: d measures distances in X and s measures distances in Y. But the idea is exactly the same as in calculus: "continuity of f at a" means, roughly, that "if x is near a in the domain X, then the image $f(x)$ is near $f(a)$ in Y."

Theorem 5.2 Suppose (X, d) and (Y, s) are pseudometric spaces. If $a \in X$ and $f : X \to Y$, then the following statements are equivalent:

1) f is continuous at a
2) $\forall \epsilon > 0 \, \exists \delta > 0$ such that $f[B_\delta(a)] \subseteq B_\epsilon(f(a))$
3) $\forall \epsilon > 0 \, \exists \delta > 0 \;\; B_\delta(a) \subseteq f^{-1}[B_\epsilon(f(a))]$
4) $\forall N \subseteq Y$: if $f(a) \in \operatorname{int} N$, then $a \in \operatorname{int} f^{-1}[N]$.

Proof It is clear that conditions 1) - 2) - 3) are just equivalent restatements the definition of continuity at a in terms of images and inverse images of balls. Condition 4), however, seems a bit strange. We will show that 3) and 4) are equivalent.

3) $\Rightarrow$ 4) Suppose $f(a) \in \operatorname{int} N$. By definition of interior, there is an $\epsilon > 0$ such that $B_\epsilon(f(a)) \subseteq N$, so that $f^{-1}[B_\epsilon(f(a))] \subseteq f^{-1}[N]$. By 3), we can pick $\delta > 0$ so that $a \in B_\delta(a) \subseteq f^{-1}[B_\epsilon(f(a))] \subseteq f^{-1}[N]$. Since the δ-ball at a is an open subset of $f^{-1}[N]$, we get that $a \in \operatorname{int} f^{-1}[N]$, as desired.

4) $\Rightarrow$ 3) Suppose $\epsilon > 0$ is given. Let $N = B_\epsilon(f(a))$. Then N is open and $f(a) \in N = \operatorname{int} N$. We conclude from 4) that $a \in \operatorname{int} f^{-1}[N] = \operatorname{int} f^{-1}[B_\epsilon(f(a))]$. Therefore for some $\delta > 0$, $B_\delta(a) \subseteq \operatorname{int} f^{-1}[B_\epsilon(f(a))] \subseteq f^{-1}[B_\epsilon(f(a))]$, so 3) holds. ∙

Definition 5.3 If $N \subseteq X$ and $x \in \text{int}\, N$, then N is called a neighborhood of x in (X, d). Thus, N is a neighborhood of x if there is an open set O such that $x \in O \subseteq N$.

Notice that:

1) The term "neighborhood" goes together with a point $x \in X$. We might say "O is an open set in X," but we would never say "N is a neighborhood in X" – but rather "N is a neighborhood of x in X," where x is some point in int(N).

2) A neighborhood N of x need not be an open set. However, be aware that in some books "a neighborhood of x" means "an open set containing x." It's not really important which way we make the definition of neighborhood (each version has its own technical advantages), but it is important that we all agree in these notes.

So in $\mathbb{R}^n$, for example, we say that the closed ball $N = \{(a, x) : d(a, x) \leq \epsilon\}$ is a neighborhood of a; in fact N is a neighborhood of each point x in its interior. A point x for which $d(a, x) = \epsilon$ is in N, but N is not a neighborhood of such a point x.

The following observation is almost trivial but it is important enough to state and remember.

Theorem 5.4 A subset N in (X, d) is open iff N is a neighborhood of each of its points.

Proof Suppose N is open. Then int $N = N$, so each point x of N is automatically in int N. So N is a neighborhood of each of its points.

Conversely, if N is a neighborhood of each of its points, then for every $x \in N$, we have $x \in \text{int}\, N$. Therefore $N \subseteq \text{int}\, N$, so $N = \text{int}\, N$ and N is open. ●

With this new terminology, we can restate the equivalence of 1) and 4) in Theorem 5.2 as:

f is continuous at a iff

whenever N is a neighborhood of $f(a)$ (*in* Y), then $f^{-1}[N]$ is a neighborhood of a (*in* X).

This tells us something very important. Interiors (and therefore neighborhoods of points) are defined in terms of the open sets in X – without needing to mention the distance function. This means that the neighborhoods of a point x depend only on the topology, not on the specific metric that generates the topology. Therefore whether or not f is continuous at a does not actually depend on the specific metrics but only on the topologies in the domain and range. In other words, "f is continuous at a" is a topological property.

For example, the function sin : $\mathbb{R} \to \mathbb{R}$ is continuous at each point a in $\mathbb{R}$, and this is remains true if we measure distances in the domain with, say, the taxicab metric d_t and distances in the range with the max-metric d^* – since these are both equivalent to the usual metric d on $\mathbb{R}$.

We now define "f is a continuous function" in the usual way.

Definition 5.5 Suppose (X, d) and (Y, s) are pseudometric spaces. We say that $f : X \to Y$ is continuous if f is continuous at each point of X.

Theorem 5.6 Suppose $f : X \to Y$ where (X, d) and (Y, s) are pseudometric spaces. The following are equivalent:

1) f is continuous
2) if O is open in Y, then $f^{-1}[O]$ is open in X
3) if F is closed in Y, then $f^{-1}[F]$ is closed in X.

Proof 1) $\Rightarrow$ 2) Suppose O is open in Y and that $x \in f^{-1}[O]$. Since O is a neighborhood of $f(x)$ and f is continuous at x, we know from Theorem 5.2 that $f^{-1}[O]$ is a neighborhood of x. Therefore $f^{-1}[O]$ is a neighborhood of each of its points, so $f^{-1}[O]$ is open.

2) $\Rightarrow$ 3) If F is closed in Y, then $Y - F$ is open. By 2), $f^{-1}[Y - F] = X - f^{-1}[F]$ is open in X, so $X - (X - f^{-1}[F]) = f^{-1}[F]$ is closed in X.

3) $\Rightarrow$ 2) Exercise

2) $\Rightarrow$ 1) Suppose $a \in X$ and that N is a neighborhood of $f(a)$ in Y, so that $f(a) \in \text{int}\, N \subseteq N$. By 2), $f^{-1}[\text{int}\, N]$ is open in X, and $a \in f^{-1}[\text{int}\, N] \subseteq f^{-1}[N]$. Therefore $f^{-1}[N]$ is a neighborhood of a. Therefore f is continuous at a. Since a was an arbitrary point in X, f is continuous. ●

Notice again: Theorem 5.6 shows that continuity is completely described in terms of the open sets (or equivalently, the closed sets), and the proof of the theorem is phrased entirely in terms of open (closed) sets, without any explicit mention of the pseudometrics on X and Y. Replacing d and s with equivalent pseudometrics would not affect the continuity of f.

Theorem 5.7 Suppose (X, d), (Y, s) and (Z, t) are pseudometric spaces and that $f : X \to Y$ and $g : Y \to Z$. If f is continuous at $a \in X$ and g is continuous at $f(a) \in Y$, then $g \circ f$ is continuous at a. (Therefore, if f and g are continuous, so is $g \circ f$.)

Proof If N is a neighborhood of $g(f(a))$, then $g^{-1}[N]$ is a neighborhood of $f(a)$, because g is continuous at $f(a)$. Since f is continuous at a, $f^{-1}[g^{-1}[N]]$ is a neighborhood of a. But $f^{-1}[g^{-1}[N]] = (g \circ f)^{-1}[N[$, so $g \circ f$ is continuous at a. ●

Example 5.8

1) Suppose $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{if } x \neq 0 \end{cases}$. Then $N = (\frac{1}{2}, \frac{3}{2})$ is a neighborhood of $f(0) = 1$, but $f^{-1}[N] = \{0\}$ is not a neighborhood of 0. Therefore f is not continuous at 0. To see the same thing using slightly different language: f is not continuous at 0 because, choosing $\epsilon = \frac{1}{2}$, there is no choice of $\delta > 0$ such that $f[B_\delta(0)] \subseteq B_\epsilon(f(0)) = B_\epsilon(1)$.

2) If $f:(X,d) \to (Y,s)$ is a constant function, then f is continuous. To see this, suppose $f(x) = c$ for every x. If O is open in Y, then $f^{-1}[O] = \begin{cases} \emptyset & \text{if } c \notin O \\ X & \text{if } c \in O \end{cases}$. In both cases, $f^{-1}[O]$ is open.

3) Suppose $f:\mathbb{R} \to \mathbb{R}$ is given by $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{P} \end{cases}$. Then f is not continuous at any point $a \in \mathbb{R}$ (why?). However $f|\mathbb{Q} = g : \mathbb{Q} \to \mathbb{R}$ is continuous at every point of $\mathbb{Q}$, because g is a constant function.

There is a curious old result called <u>*Blumberg's Theorem*</u> *which states:*

For any $f:\mathbb{R} \to \mathbb{R}$, there exists a dense set $D \subseteq \mathbb{R}$ such that $f|D = g : D \to \mathbb{R}$ is continuous.

Blumberg's Theorem is rather difficult to prove, and not very useful.

4) Suppose $f, g : X \to \mathbb{R}$ where (X,d) is a pseudometric space. Since these functions are real-valued, it makes sense to define functions $f+g$, $f-g$, $f \cdot g$ and $\frac{f}{g}$ in the obvious way. For example, $(f+g)(x) = f(x) + g(x)$, where the "$+$" on the right is ordinary addition in $\mathbb{R}$.

If f and g are continuous at a point $a \in X$, then the functions $f+g$, $f-g$, and $f \cdot g$ are also continuous at a; and $\frac{f}{g}$ is continuous at a if $g(a) \neq 0$. The proofs are just like those given in calculus (where $X = \mathbb{R}$)

For example, consider $f+g$: given $\epsilon > 0$, then (because f, g are continuous at a), we can find $\delta_1 > 0$ and $\delta_2 > 0$ so that

if $d(x,a) < \delta_1$, then $|f(x) - f(a)| < \frac{\epsilon}{2}$ and
if $d(x,a) < \delta_2$, then $|g(x) - g(a)| < \frac{\epsilon}{2}$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then if $d(x,a) < \delta$, we have

$$|(f+g)(x) - (f+g)(a)| = |f(x) - f(a) + g(x) - g(a)|$$
$$\leq |f(x) - f(a)| + |g(x) - g(a)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

You can find at the other proofs in any analysis textbook.

5) For $x = (x_1, x_2, ..., x_n, ...) \in \ell_2$, define $f : \ell_2 \to \mathbb{R}$ by $f(x) = x_1$. (f *is a "projection" of* ℓ_2 *onto* $\mathbb{R}$.) Then f is continuous at x. To see this, suppose $\epsilon > 0$, and let $\delta = \epsilon$. If $y = (y_1, y_2, ..., y_n, ...) \in B_\delta(x)$, then

$$|f(x) - f(y)| = |x_1 - y_1| \leq d(x,y) = (\sum_{i=1}^{\infty}(x_i - y_i)^2)^{1/2} < \delta = \epsilon,$$

so $f[B_\delta(x)] \subseteq B_\epsilon(f(x))$.

A similar argument shows that each projection $g(x) = x_n$ is also continuous, and an argument only slightly more complicated would show, for example, that the projection function $h : \ell_2 \to \mathbb{R}^3$ given by $h(x) = (x_3, x_9, x_{11})$ is continuous.

6) If a is an isolated point in (X, d), then every function $f : (X, d) \to (Y, s)$ is continuous at a. To see this, suppose N is a neighborhood of $f(a)$. Then $a \in \{a\} \subseteq f^{-1}[N]$. But $\{a\}$ is open in X, so $f^{-1}[N]$ is a neighborhood of a.

If $\mathcal{T}_d$ happens to be the discrete topology, then every point in X is isolated, so f is continuous (*In this case, we could argue instead that whenever O is open in Y then $f^{-1}[O]$ must be open in X – because every subset of X is open.*)

7) A function $f : (X, d) \to (Y, s)$ is called an isometry of X into Y if it preserves distances, that is, if $d(a, b) = s(f(a), f(b))$ for all $a, b \in X$. An isometry is clearly continuous (given $\epsilon > 0$, choose $\delta = \epsilon$).

Note that if d is a metric, then f one-to-one (*Why?*). If f happens to be a bijection, we say that (X, d) and (Y, s) are isometric to each other. In that case, it is clear that the inverse function f^{-1} is also an isometry, so f^{-1} is also continuous.

Theorem 5.9 If (X, d) is a metric space and $x \neq y \in X$, then there exist open sets U and V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$. (*More informally: distinct points in a metric space can be separated by disjoint open sets.*)

Proof Since $x \neq y$, $d(x, y) = \delta > 0$. Let $U = B_{\frac{\delta}{2}}(x)$ and $V = B_{\frac{\delta}{2}}(y)$. These sets are open, and if there were a point $z \in U \cap V$, we would have a contradiction:

$$\delta = d(x, y) \leq d(x, z) + d(z, y) < \tfrac{\delta}{2} + \tfrac{\delta}{2} = \delta. \quad \bullet$$

Theorem 5.9 may not be true if d is not a metric. For example, if d is the trivial pseudometric on X, then the only open sets containing x and y are $U = V = X$.)

Example 5.10

1) Suppose $f : (X, d) \to (Y, s)$, where d is the trivial pseudometric on X and Y is any metric space. We already know that if f is constant, then f is continuous.

If f is not constant, then there are points $a, b \in X$ for which $f(a) \neq f(b)$. Since s is a metric, we can pick disjoint open sets $U$and V in Y with $f(a) \in U$ and $f(b) \in V$. Then $f^{-1}[U] \neq \begin{cases} \emptyset & \text{(because } a \in f^{-1}[U]) \\ X & \text{(because } b \notin f^{-1}[U]) \end{cases}$

Since $\mathcal{T}_d = \{\emptyset, X\}$, we see that $f^{-1}[U]$ is not open so f is not continuous.

So, in this situation we conclude that: f is continuous iff f is constant.

2) Suppose d is the usual metric and s is the discrete unit metric on $\mathbb{R}$. Let $i : (\mathbb{R}, d) \to (\mathbb{R}, s)$ is the identity map $i(x) = x$. For every open set O in $(\mathbb{R}, d)$, the image set $i[O]$ is open in $(\mathbb{R}, s)$, but this is not the criterion for continuity: in fact, this function is not continuous at any point. The criterion for continuity is that the inverse image of every open set must be open.

Example 5.10.2 leads us to a definition.

Definition 5.11 A function $f : (X, d) \to (Y, s)$ is called an open function or open mapping if: whenever O is open in X, then the image set $f[O]$ is open in Y. Similarly, we call f a closed mapping if whenever F is closed in X, then the image set $f[F]$ is closed in Y.

The identity mapping i in Example 5.10.2 is both open and closed but i is not continuous. You can convince yourself fairly easily that the projection function $\pi_x : \mathbb{R}^2 \to \mathbb{R}$ given by $\pi_x(x, y) = x$ is open and continuous, but it is not closed – for example, the set $F = \{(x, y) : y = \frac{1}{x}\}$ is a closed set in $\mathbb{R}^2$ but $\pi_x[F] = (0, \infty)$ is not closed in $\mathbb{R}$. The general point is that for a function $f : (X, d) \to (Y, s)$, the properties "open", "closed", and "continuous" are completely independent. You should provide other examples: for instance, a function that is continuous but not open or closed.

With just the basic ideas about continuous functions, we can already prove some rather interesting results.

Theorem 5.12 Suppose $f, g : (X, d) \to (Y, s)$, where d is a pseudometric and s is a metric. Let D be a dense subset of X. If f and g are continuous and $f|D = g|D$, then $f = g$. (*More informally: if two continuous functions with values in a metric space agree on a dense set, then they agree everywhere.*)

Proof Suppose $f \neq g$. Then, for some point $a \in X$, we have $f(a) \neq g(a)$. Since s is a metric, we can find disjoint open sets U and V in Y with $f(a) \in U$ and $g(a) \in V$. Since f and g are continuous at a, there are open sets U_1 and V_1 in X that contain a and that satisfy $f[U_1] \subseteq U$ and $g[V_1] \subseteq V$.

$U_1 \cap V_1$ is an open set containing a. Since $a \in \text{cl}\, D$, there must be a point $d \in (U_1 \cap V_1) \cap D$. Then $f(d) \neq g(d)$ because $f(d) \in U$, $g(d) \in V$ and $U \cap V = \emptyset$. Therefore $f|D \neq g|D$. •

Example 5.13 If $f, g \in C(\mathbb{R})$ and $f|\mathbb{Q} = g|\mathbb{Q}$, then $f = g$ by Theorem 5.12. In other words, the mapping $\Phi : C(\mathbb{R}) \to C(\mathbb{Q})$ given by $\Phi(f) = f|\mathbb{Q} \in C(\mathbb{Q})$ is one-to-one. Therefore $|C(\mathbb{R})| \leq |C(\mathbb{Q})| \leq c^{\aleph_0} = c$.

On the other hand, each constant function $f(x) = r$ is in $C(\mathbb{R})$, so $|C(\mathbb{R})| \geq c$. It follows that $|C(\mathbb{R})| = c$. In other words, there are exactly c continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

Example 5.14 Find all continuous functions $f : \mathbb{R} \to \mathbb{R}$ that satisfy the functional equation

$$f(x + y) = f(x) + f(y) \text{ for all } x, y \in \mathbb{R} \qquad (*)$$

Simple induction shows that for $n \in \mathbb{N}$, $f(x_1 + ... + x_n) = f(x_1) + ... + f(x_n)$.

By $(*)$,

$$f(0) = f(0 + 0) = f(0) + f(0), \text{ so } f(0) = 0.$$

Let $f(1) = c$. Then

$$\begin{aligned} f(2) &= f(1 + 1) = f(1) + f(1) = c \cdot 2 \\ f(3) &= f(2 + 1) = f(2) + f(1) = c \cdot 2 + c = c \cdot 3 \\ &\vdots \end{aligned}$$

Continuing, we see that $f(n) = cn$ for every $x \in \mathbb{N}$. Similarly, for each $m, n \in \mathbb{N}$, we have

$$f(1) = f(\underbrace{\tfrac{1}{n} + ... + \tfrac{1}{n}}_{n \text{ terms}}) = f(\tfrac{1}{n}) + ... f(\tfrac{1}{n}) = n f(\tfrac{1}{n}), \text{ so } f(\tfrac{1}{n}) = c(\tfrac{1}{n})$$

$$f(\tfrac{m}{n}) = f(\tfrac{1}{n} + ... + \tfrac{1}{n}) = mf(\tfrac{1}{n}) = m \cdot c(\tfrac{1}{n}) = c(\tfrac{m}{n})$$
$$\uparrow$$
$$m \text{ terms}$$

So far, we have shown that a function f satisfying $(*)$ must have the formula $f(x) = cx$ for every positive rational $x = \frac{m}{n}$.

Since $\quad 0 = f(0) = f(\frac{m}{n} + (-\frac{m}{n})) = f(-\frac{m}{n}) + f(\frac{m}{n})$, we get that

$$f(-\tfrac{m}{n}) = -f(\tfrac{m}{n}) = -(\tfrac{m}{n})c = c(-\tfrac{m}{n}).$$

Therefore, $f(x) = cx$ for every $x \in \mathbb{Q}$.

So far, we have not used the hypothesis that f is continuous. Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = cx$. Since f and g are continuous and $f|\mathbb{Q} = g|\mathbb{Q}$, Theorem 5.12 tells us that $f = g$, that is, $f(x) = cx$ for <u>all</u> $x \in \mathbb{R}$.

> *The continuous functions satisfying the functional equation $*$) were first described by Cauchy in 1821. It turns out that there are also discontinuous functions f satisfying $(*)$, but they are not easy to find. In fact, they must satisfy a nasty condition called "nonmeasurability" (which makes them "extremely discontinuous").*

In calculus, another important is the idea of a convergent sequence (in $\mathbb{R}$ or $\mathbb{R}^n$). We can generalize the idea of a convergent sequence in $\mathbb{R}^n$ to any pseudometric space.

Definition 5.15 A <u>sequence</u> (x_n) <u>in (X, d)</u> <u>converges to x</u> $\in X$ if any one of the following (clearly equivalent) conditions holds:

1) $\forall \epsilon > 0 \; \exists N \in \mathbb{N}$ such that if $n \geq N$, then $x_n \in B_\epsilon(x)$
(that is, if the sequence of numbers $(d(x_n, x)) \to 0$ in $\mathbb{R}$)

2) if O is open and $x \in O \subseteq X$, then $\exists N \in \mathbb{N}$ such that if $n \geq N$, then $x_n \in O$

3) if W is a neighborhood of x, then $\exists N \in \mathbb{N}$ such that if $n \geq N$, then $x_n \in W$.

If (x_n) converges to X, we write $(x_n) \to x$.

Condition$s$ 2) and 3) describe the convergence of sequences in terms of open sets (or neighborhoods) rather than directly using the distance functions. Therefore replacing d with an equivalent metric d' does not affect which sequences converge to which points: sequential convergence is a topological property.

If a sequence (x_n) has a certain property P for all $n \geq$ some N, we say that "(x_n) <u>eventually</u> has property P." For example, the sequence (0,3,1,7,7,7,7,...) is <u>eventually constant</u>; the sequence $(-1, -3, 5, -2, 5, 6, 7, 8, ..., n, n+1, ...)$ is <u>eventually increasing</u>. Using this terminology, we can give a completely precise definition of convergence by saying: (x_n) converges to x if (x_n) is eventually in each neighborhood of x.

Example 5.16 1) In $\mathbb{R}$, $(\frac{1}{n}) \to 0$.

2) Suppose d is the discrete unit metric on X. Then each set $\{x\}$ is open so $(x_n) \to x$ iff (x_n) is eventually in each neighborhood of X iff (x_n) is eventually in $\{x\}$. In other words, $(x_n) \to x$ iff $x_n = x$ eventually. Every convergent sequence is eventually constant.

At the other extreme, suppose d is the trivial pseudometric on X. Then every sequence (x_n) converges to every point $x \in X$ (since the only neighborhood of x is X).

Example 5.16.2 shows that limits of sequences in a pseudometric space do not have to be unique: the same sequence can have many limits. However if d is a metric, then sequential limits in (X, d) must be unique, as the following theorem shows.

Theorem 5.17 A sequence (x_n) in a metric space (X, d) has at most one limit.

Proof Suppose $x \neq y \in X$ and let U, V be disjoint open sets with $x \in U$ and $y \in V$. If $(x_n) \to x$, then (x_n) must be eventually in U. Since U and V are disjoint, this means that (x_n) cannot be eventually in V (*in fact, the sequence must be eventually outside* V), so $(x_n) \not\to y$. •

In a pseudometric space, sequences can be used to describe the closure of a set.

Theorem 5.18 Suppose $A \subseteq X$, where (X, d) is a pseudometric space. Then $x \in \text{cl}\, A$ iff there is a sequence (a_n) in A for which $(a_n) \to x$.

Proof First, suppose there is a sequence (a_n) in A for which $(a_n) \to x$. If N is any neighborhood of x, then (a_n) is eventually in N. Therefore $N \cap A \neq \emptyset$, so $x \in \text{cl}\, A$.

Conversely, suppose $x \in \text{cl}\, A$. Then $B_{\frac{1}{n}}(x) \cap A \neq \emptyset$ for each $n \in \mathbb{N}$, so we can choose a point $a_n \in B_{\frac{1}{n}}(x) \cap A$. Then $(a_n) \to x$ (because $d(a_n, x) \to 0$). •

Note: the sequence (a_n) is actually a function $f : \mathbb{N} \to X$ with the property that $a_n = f(n) \in B_{\frac{1}{n}}(x) \cap A$. Informally, the existence of such a function is completely clear. But to be precise, this argument actually depends on the Axiom of Choice. The proof as written doesn't describe how to pick specific a_n's: it depends on making "arbitrary choices." But using AC gives us a function f which "chooses" one point from each set in the collection $\{B_{\frac{1}{n}}(x) \cap A : n \in \mathbb{N}\}$.

Theorem 5.18 tells us something very important about the role of sequences in pseudometric spaces. The set A is closed iff $A = \text{cl}\, A$. But $A \subseteq \text{cl}\, A$ is always true, so we can say A is closed iff $\text{cl}\, A \subseteq A$. But that is true iff the limits of convergent sequences (a_n) from A must also be in A. Therefore a complete knowledge about what sequences converge to what points in (X, d) would let you determine which sets are closed (and therefore, by taking complements, which sets are open). In other words, all the information about "which sets in (X, d) are open or closed?" is revealed by the convergent sequences. We summarize this by saying that in a pseudometric space (X, d), sequences are sufficient to describe the topology.

Example 5.19 If d is a pseudometric on X, then d' defined by $d'(x,y) = \min\{1, d(x,y)\}$ is also a pseudometric on X. It is clear that $d'(x_n, x) \to 0$ iff $d(x_n, x) \to 0$. In other words, the metrics d and d' produce exactly the same convergent sequences and limits in (X,d). Since sequences are sufficient to determine the topology in pseudometric spaces, we conclude that d and d' are equivalent pseudometrics on X.

This example also shows that for any given (X,d) there is always an equivalent pseudometric d' on X for which all distances are ≤ 1: <u>every pseudometric is equivalent to a bounded pseudometric</u>.

Another modification of d that accomplishes the same thing is $d''(x,y) = \frac{d(x,y)}{1+d(x,y)}$. This time, it is a little harder to verify that d'' is in fact a pseudometric – the triangle inequality for d'' takes a bit of work. Clearly $d''(x,y) \leq 1$, and $d''(x_n, x) \to 0$ iff $d(x_n, x) \to 0$. So $d'' \sim d \sim d'$.

Definition 5.20 The <u>diameter</u> of a set A in (X,d) is defined by $\text{diam}(A) = \sup\{d(x,y) : x, y \in A\}$ (*we allow the possibility that* $\mathit{diam}(A) = \infty$). If A has finite diameter, we say that A is a <u>bounded set.</u>

It is an easy exercise to show that A is bounded iff $A \subseteq B_k(x_0)$ for some sufficiently large k (where x_0 can be any point in X).

The diameter of a set depends on the particular metric being used. Since we can always replace d by an equivalent metric d' or d'' for which $\text{diam}(X) \leq 1$, <u>boundedness is not a topological property</u>.

The fact that the convergent sequences determine the topology in (X,d) gives us an upper bound on the size of certain metric spaces.

Theorem 5.21 If D is a dense set in a <u>metric</u> space (X,d), then $|X| \leq |D|^{\aleph_0}$. In particular, for a <u>separable</u> metric space (X,d), it must be true that $|X| \leq \aleph_0^{\aleph_0} = c$.

Proof For each $x \in X$, pick a sequence (d_n) in D such that $(d_n) \to x$. This sequence is actually a function $f_x \in D^{\mathbb{N}}$. Since a sequence in a metric space has at most one limit, the mapping $\Phi : X \to D^{\mathbb{N}}$ given by $\Phi(x) = f_x$ is one-to-one, so $|X| \leq |D^{\mathbb{N}}| = |D|^{\aleph_0}$. •

> *Note: If* $|D| = m > \aleph_0$, *do not jump to the (possibly false) conclusion that* $|X| \leq m^{\aleph_0} = m$. *See the note of caution in Chapter I at the end of Example I.14.8.*

Theorem 5.21 is <u>not</u> true if (X,d) is merely a pseudometric space. For example, let X be an uncountable set (with arbitrarily large cardinality) and let d be the trivial pseudometric on X. Then any singleton $\{x\}$ is dense, but $|X| > 1^{\aleph_0} = 1$. In this case, where does the proof of Theorem 5.21 fall apart?

Sequences are sufficient to determine the topology in a pseudometric space, and continuity is characterized in terms of open sets, so it should not be a surprise that sequences can be used to decide whether or not a function f between pseudometric spaces is continuous.

Theorem 5.22 Suppose (X,d) and (Y,s) are pseudometric spaces, and that $f : X \to Y$. Then f is continuous at $a \in X$ iff $(f(x_n)) \to f(a)$ for every sequence $(x_n) \to a$.

Proof Suppose f is continuous at a and consider any sequence $(x_n) \to a$. If W is a neighborhood of $f(a)$, then $f^{-1}[W]$ is a neighborhood of a, so (x_n) is eventually in $f^{-1}[W]$. This implies that $(f(x_n))$ is eventually in W, so $(f(x_n)) \to f(a)$.

Conversely, if f is not continuous at a, then $\sim (\forall \epsilon > 0 \, \exists \delta > 0 \;\; f[B_\delta(x)] \subseteq B_\epsilon(f(a)))$, that is, $\exists \epsilon > 0 \; \forall \delta > 0 \;\; f[B_\delta(x)] \not\subseteq B_\epsilon(f(a))$.

For this ϵ and $\delta = \frac{1}{n}$, we have that $f[B_{\frac{1}{n}}(x)] \not\subseteq B_\epsilon(f(a))$. So for each n we can choose a point $x_n \in B_{\frac{1}{n}}(x)$ for which $f(x_n) \notin B_\epsilon(f(x))$. Then $(x_n) \to x$ (because $d(x_n, x) < \frac{1}{n} \to 0$), but $(f(x_n)) \not\to f(a)$ (because $s(f(x_n), f(a)) \geq \epsilon$ for all n). Therefore f is not continuous at a. •

Note: the first half of the proof is phrased completely in terms of neighborhoods of x_0 and $f(x_0)$; that part of the proof is topological. However the second half makes explicit use of the metric. In fact, as we shall see later, the second half of the proof must involve a little more than just the open sets.

Notice also that the second half of the proof makes uses the Axiom of Choice (the function x "chooses" x_n for each n.

The following theorem and its corollaries are often technically useful. Moreover, they show us that a pseudometric space (X,d) has lots of "built-in" continuous functions – these functions can be defined using pseudometric d.

Theorem 5.23 In a pseudometric space (X,d):

if $(x_n) \to x$ and $(y_n) \to y$, then $d(x_n, y_n) \to d(x,y)$.

Proof Given $\epsilon > 0$, pick N large enough so that $n \geq N$ implies that $d(x_n, x) < \frac{\epsilon}{2}$ and $d(y_n, y) < \frac{\epsilon}{2}$ are both true. Since $d(x_n, y_n) \leq d(x_n, x) + d(x, y_n) \leq d(x_n, x) + d(x,y) + d(y, y_n)$, we get that

$$\text{if } n \geq N, \quad d(x_n, y_n) < \tfrac{\epsilon}{2} + d(x,y) + \tfrac{\epsilon}{2} = d(x,y) + \epsilon \qquad (*)$$

Similarly, $d(x,y) \leq d(x, x_n) + d(x_n, y_n) + d(y_n, y)$, so that

$$\text{if } n \geq N, \quad d(x,y) < \tfrac{\epsilon}{2} + d(x_n, y_n) + \tfrac{\epsilon}{2} = d(x_n, y_n) + \epsilon \quad (**)$$

Combining $(*)$ and $(**)$ gives that $|d(x_n, y_n) - d(x,y)| < \epsilon$ if $n \geq N$, so $d(x_n, y_n) \to d(x,y)$. •

Corollary 5.24 Let (X,d) be a pseudometric space and $a \in X$. If $(x_n) \to x$ in X, then $d(x_n, a) \to d(x, a)$ (in $\mathbb{R}$).

Proof In Theorem 5.23, let (y_n) be the constant sequence where $y_n = a$ for every n. •

Corollary 5.25 Suppose $a \in (X, d)$. Define $f : X \to \mathbb{R}$ by $f(x) = d(x, a)$. Then f is continuous.

Proof Let x_0 be a point in X. If $(x_n) \to x_0$, then by Corollary 5.24, $(d(x_n, a)) \to d(x_0, a)$, that is $(f(x_n)) \to f(x_0)$. So f is continuous at x_0 (by Theorem 5.22). •

Recall that for a nonempty subset A of X, we defined $d(x, A) = \inf\{d(x, a) : a \in A\}$.
The following theorem is also useful.

Theorem 5.26 If A is a nonempty subset of the pseudometric space (X, d), then the function $f : X \to \mathbb{R}$ defined by $f(x) = d(x, A)$ is continuous.

Proof We show that f is continuous at each point $x_0 \in X$. Let $a \in A$. Then the following inequalities are true for every $x \in X$:

$$\begin{aligned} d(a, x_0) &\leq d(a, x) + d(x, x_0) \\ d(a, x) &\leq d(a, x_0) + d(x_0, x) \end{aligned}$$

Apply "$\inf_{a \in A}$" to each inequality to get

$$\begin{aligned} d(x_0, A) &\leq d(x, A) + d(x, x_0) \text{ or } & d(A, x_0) - d(A, x) &\leq d(x, x_0) \\ d(x, A) &\leq d(x_0, A) + d(x_0, x) \text{ or } & d(A, x) - d(A, x_0) &\leq d(x, x_0) \end{aligned}$$

so for all $x \in X$, $\quad |d(A, x) - d(A, x_0)| \leq d(x, x_0)$. In other words,

for all $x \in X$, $\quad |f(x) - f(x_0)| \leq d(x, x_0)$. (*)

So for $\epsilon > 0$, we can choose $\delta = \epsilon$. Then if $d(x, x_0) < \delta$, we have $|f(x) - f(x_0)| < \epsilon$. Therefore f is continuous at x_0. •

Comments on the proof:

i) For a given $\epsilon > 0$: the <u>same choice</u> $\delta = \epsilon$ can be used at every point x_0. A function f that satisfies this condition – stronger than mere continuity – is called <u>uniformly continuous</u>. We will say more about uniform continuity in Chapter IV.)

ii) From the last inequality (*) we could have argued instead: for any sequence $(x_n) \to x_0$, we have $d(x_n, x_0) \to 0$, and this forces $(f(x_n)) \to f(x_0)$. Therefore f is continuous at x_0.
However this argument obscures the observation about "uniform continuity" made in i).

Exercises

E22. Suppose $a \in A \subseteq \mathbb{R}$ and that $f : A \to \mathbb{R}$. We said that f is continuous at a if

$\forall\, \epsilon > 0\ \exists\, \delta > 0$ such that if $x \in A$ and $|x - a| < \delta$, then $|f(x) - f(a) < \epsilon$.

The order of the quantifiers is important. What functions are described by each of the following modifications of the definition?

a) $\forall\, \delta > 0\ \exists\, \epsilon > 0$ such that if $x \in A$ and $|x - a| < \delta$, then $|f(x) - f(a) < \epsilon$
b) $\forall\, \epsilon > 0\ \forall\, \delta > 0$ such that if $x \in A$ and $|x - a| < \delta$, then $|f(x) - f(a) < \epsilon$
c) $\exists\, \epsilon > 0\ \exists\, \delta > 0$ such that if $x \in A$ and $|x - a| < \delta$, then $|f(x) - f(a) < \epsilon$.

In each case, what happens if the restriction " > 0" is dropped on either ϵ or δ ?

E23. Suppose $f : \mathbb{R} \to \mathbb{R}$ and that $f(a) = b$. What does each of the following statements tell us about f? *Throughout this exercise, "interval" always means "a bounded open interval* (c, d).

a) For every interval I containing a and every interval J containing b, $f[I] \subseteq J$.

b) There exists an interval J containing b and there exists an interval I containing a such that $f[I] \subseteq J$.

c) There exists an interval J containing b such that for every interval I containing a, $f[I] \subseteq J$

d) There exists an interval J containing b such for every interval I containing a, $f[I] \not\subseteq J$

. e) For every interval I containing a, there exists an interval J containing b such that $f[I] \subseteq J$

f) There exists an interval I containing a such that for every interval J containing b, $f[I] \subseteq J$

E24. (The Pasting Lemma, easy version) *The two parts of the problem give conditions when a collection of continuous functions defined on subsets of X can be "united" (= "pasted together") to form a new continuous function.* Let sets O_α be open in (X, d), where $\alpha \in A$ (an indexing set of any size); and let $F_1, ..., F_n$ $(n \in \mathbb{N})$ be closed in (X, d).

a) If functions $f_\alpha : O_\alpha \to (Y, s)$ are continuous and $f_\alpha|(O_\alpha \cap O_\beta) = f_\beta\,|(O_\alpha \cap O_\beta)$ for all α and β in A *(that is, f_α and f_β agree where their domains overlap)*, then prove that $\bigcup_{\alpha \in A} f_\alpha = f : \bigcup_{\alpha \in A} O_\alpha \to Y$ is continuous.

b) If functions $f_i : F_i \to (Y, s)$ are continuous for each $i = 1, ..., n$ and $f_i|(F_i \cap F_j) = f_j|(F_i \cap F_j)$ for all i and j *(that is, f_i and f_j agree where their domains overlap)*, then prove that $f = \bigcup_{i=1}^n f_i : \bigcup_{i=1}^n F_i \to Y$ is continuous.

c) Give an example to show that b) may be false for an infinite collection of functions f_i $(i \in \mathbb{N})$ defined on closed subsets of X, even if the domains F_i are pairwise disjoint.

E25. A point x_0 in (X, d) is a cluster point of the sequence (x_n) if for every neighborhood N of x_0 and for all $n \in \mathbb{N}$, $\exists k > n$ such that $x_k \in N$. (*When this condition is true, we say that "(x_n) is frequently in every neighborhood of x_0."*)

Prove that if $f : (X, d) \to (Y, s)$ is continuous and x_0 is a cluster point of (x_n) in X, then $f(x_0)$ is a cluster point of the sequence $(f(x_n))$ in Y.

E26. The characteristic function of a set $A \subseteq X$ is defined by $\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$.
For which sets A is $\chi_A : \mathbb{R} \to \mathbb{R}$ continuous?

E27. Show that a set O is open in (X, d) if and only if there is a continuous function $f : X \to \mathbb{R}$ and an open set W in $\mathbb{R}$ such that $O = f^{-1}[W]$.

E28. Let $x \in X$, where (X, d) is a pesudometric space. Suppose (x_n) is a sequence in X such that $(f(x_n)) \to f(x)$ for every $f \in C(X)$. Either prove that $(x_n) \to x$, or give a example to show that this may be false.

E29. Let d be the usual metric on $\mathbb{R}$. If possible, find a metric d' on $\mathbb{R}$ such that $(x_n) \to 0$ with respect to d iff $(x_n) \to 0$ with respect to d', but d' is not equivalent to d.

E30. a) Suppose A is a closed set in the pseudometric space (X, d) and that $x_0 \notin A$. Prove that there is a continuous function $f : X \to [0, 1]$ such that $f|A = 0$ and $f(x_0) = 1$. (*Hint: Consider the function "distance to the set A."*)

b) Suppose A and B are disjoint closed sets in (X, d). Prove that there exists a continuous function $f : X \to \mathbb{R}$ such that $f|A = 0$ and $f|B = 1$. (*Hint: Consider* $\frac{d(x,A)}{d(x,A)+d(x,B)}$)

c) Using b) (*or by some other method*) prove that if A and B are disjoint closed sets in (X, d), then there exist open sets U and V such that $A \subseteq U$, $B \subseteq V$ and $U \cap V = \emptyset$. Can U and V always be chosen so that $\operatorname{cl} U \cap \operatorname{cl} V = \emptyset$?

E31. *A* function $f: (X, d) \to (Y, s)$ is called an isometry of (X, d) into (Y, s) if f preserves distances: that is, if $d(x, y) = s(f(x), f(y))$ for all $x, y \in X$. Such an f is also called an isometric embedding.

If f is also onto, we say that (X, d) and (Y, s) are isometric to each other. Otherwise, (X, d) is isometric to a subset of (Y, s)

Let $\mathbb{R}$ and $\mathbb{R}^2$ have their usual metrics.

a) Prove that $\mathbb{R}$ and $\mathbb{R}^2$ are not isometric to each other.
b) Let $a \in \mathbb{R}$. Prove that there are exactly two isometries from $\mathbb{R}$ onto $\mathbb{R}$ which hold the point a fixed (that is, for which $f(a) = a$).
c) Give an example of a metric space (X, d) which is isometric to a proper subset of itself.

E32. Use convergent sequences to prove the theorem that two continuous functions f and g from (X,d) into a metric space are identical if they agree on a dense set in X.

E33. Suppose $f:\mathbb{R}\to\mathbb{R}$. Then we can define $F:\mathbb{R}\to\mathbb{R}^2$ by $F(x)=(x,f(x))$, so that ran(F) is the graph of f.

a) Prove that the following statements are equivalent:

i) f is continuous
ii) F is continuous
iii) The sets $\{(x,y):\ y\geq f(x)\}$ and $\{(x,y):y\leq f(x)\}$ are both closed in $\mathbb{R}^2$.

b) Prove that if f is continuous, then its graph is a closed set in $\mathbb{R}^2$. Give a proof or a counterexample for the converse.

E34. Suppose $X=\{x,y,z,w\}$ is a four point set.

a) Show that there is one and only one metric on X that satisfies the following conditions: $s(x,y)=s(y,z)=s(z,x)=2$ and $s(x,w)=\mathrm{s}(y,w)=s(z,w)=1$.

b) Show that (X,s) cannot be isometrically embedded into the plane $\mathbb{R}^2$ (with its usual metric).

c) Prove or disprove: (X,s) can be isometrically embedded in (ℓ_2,d), where d is the usual metric on ℓ_2.

E35. Suppose (X,d) is a metric space for which $|X|>1$ and in which $\emptyset$ and X are the only clopen sets. Prove that $|X|\geq c$.

(*Hint: First prove that there must be a nonconstant continuous function* $f:X\to\mathbb{R}$. *What can you say about the range of* f*?*)

E36. Let X be a finite set and let $C^*(X)$ be the set of all bounded continuous functions from X into $\mathbb{R}$. Let s denote the "uniform metric" on $C^*(X)$ given by $s(f,g)=\sup\{|f(x)-g(x)|:\ x\in X\}$.

a) Prove that $(C^*(X),s)$ is separable.

b) If $X=\mathbb{N}$, is part a) still true?

E37. Find all continuous functions $f:\mathbb{R}\to\mathbb{R}$ is continuous that satisfy the functional equation $f(x+y)=f(x)+f(x)-f(x)f(y)$ for all x,y.

(*Hint: let* $g(x)=f(x)+1$. *What simpler functional equation does* g *satisfy? What is* $g(x)$ *when* x *is rational?*)

Chapter II Review

Explain why each statement is true, or provide a counterexample.

1. A finite set in a metric space must be closed.

2. For $m, n \in \mathbb{N}$, write $m - n = 2^k z$, where z is an integer not divisible by 2.

$$\text{Let } \begin{cases} d(m,m) = 0 \\ d(m,n) = k & \text{if } m \neq n \end{cases}$$

Then d is a pseudometric on $\mathbb{N}$.

3. Consider the set $[1,\infty)$ with the metric $s(x,y) = \frac{3|x-y|}{3+3|x-y|}$. Let $\mathbb{N}$ have its usual metric d and define $f : [1,\infty) \to \mathbb{N}$ by $f(x) =$ "the largest integer $\leq x$". Then f is continuous.

4. If N_1 and N_2 are neighborhoods of x in (X,d), then $N_1 \cap N_2$ is also a neighborhood of x.

5. For any open subset O of a metric space (X,d), $\text{int}(\text{cl}(O)) = O$.

6. The metric $d(n,m) = |\frac{1}{n} - \frac{1}{m}|$ on $\mathbb{N}$ is equivalent to the usual metric on $\mathbb{N}$.

7. Define $f : \mathbb{N} \to \mathbb{R}$ by $f(n) =$ the n^{th} digit of the decimal expansion of π. Then f is continuous.

8. The set of all real numbers with a decimal expansion of the form $x = 0.x_1x_1x_3...x_n010101...$ is dense in $[0,1]$.

9. There are exactly c countable dense subsets of $\mathbb{R}$.

10. Suppose we measure distances in $\mathbb{R}$ using the metric $d(x,y) = \frac{|x-y|}{1+|x-y|}$. Then the function $\cos : \mathbb{R} \to \mathbb{R}$ is continuous at every point $a \in \mathbb{R}$.

11. A subset A in a pseudometric space (X,d) is dense if and only if int $(X - A) = \emptyset$.

12. Let d_t denote the "taxicab" metric $d_t(x,y) = \sum_{i=1}^{n} |x_i - y_i|$ on $\mathbb{R}^n$. $\mathbb{Q}^n$ is dense in $(\mathbb{R}^n, d_t)$.

13. If B is a countable subset of $\mathbb{R}$, then $\mathbb{R} - B$ is dense in $\mathbb{R}$.

14. If $A \subseteq [0,1]$ and cl $A \neq [0,1]$, then int $A \neq \emptyset$.

15. Let $A = \{\frac{1}{n} : n \in \mathbb{N}\}$. The discrete unit metric produces the same topology on A as the usual metric.

16. In a pseudometric space (X,d), cl $A =$ cl $(X - A)$ if and only if A is clopen.

17. If U is an open set in $\mathbb{R}$ and $U \supseteq \mathbb{Q}$, then $U = \mathbb{R}$.

18. There is a sequence of open sets O_n in $\mathbb{R}$ such that $\bigcap_{n=1}^{\infty} O_n = \mathbb{P}$.

19. If $A \subseteq (X, d)$, then $\operatorname{int} A = X - \operatorname{cl}(X - A)$.

20. There are exactly c continuous functions $f : \mathbb{N} \to \mathbb{N}$.

21. Let $\mathbb{R}$ have the metric $d(x, y) = \frac{|x - y|}{1 + |x - y|}$ and let $a_n = \frac{n^2}{n^2+1}$. Then $(a_n) \to 1$ in the space $(\mathbb{R}, d)$.

22. Suppose d_1 is a metric on $\mathbb{R}$ with the following property: for every sequence (r_n) in $\mathbb{R}$

$$(r_n) \to 5 \text{ with respect to } d_1 \text{ if and only if } (r_n) \to 5 \text{ with respect to } d.$$

Then $d_1 \sim d =$ the usual metric on $\mathbb{R}$.

23. Suppose $A \subseteq B \subseteq (X, d)$. If $\operatorname{cl} A = \operatorname{cl} B$, $\operatorname{int} A = \operatorname{int} B$, and $\operatorname{Fr} A = \operatorname{Fr} B$, then $A = B$.

24. Suppose $C([0, 1])$ has the metric $d(f, g) = \int_0^1 |f - g|$ and define $\Phi : (C([0, 1]), d) \to \mathbb{R}$ by $\Phi(f) = \int_0^1 f$. Then Φ is continuous.

25. Let d be the trivial pseudometric on $\mathbb{R}$. In $(\mathbb{R}, d)$, each real number r is the limit of a sequence of irrational numbers.

26. Suppose $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are continuous and that $f \neq g$. Then there must exist a point $p \in \mathbb{Q}$ where $f(p) \neq g(p)$ and a point $q \in \mathbb{P}$ where $f(q) \neq g(q)$.

27. Suppose $f : (X, d) \to (Y, s)$ is continuous <u>at the point</u> $a \in X$, and suppose O is an open set in Y with $f(a) \in O$. Then $f^{-1}[O]$ is an open set containing a.

28. If every convergent sequence in a metric space (X, d) is eventually constant, then $\mathcal{T}_d = \mathcal{P}(X)$.

29. Let $x \in (X, d)$ Suppose (x_n) is a sequence such that $(f(x_n)) \to f(x)$ for every continuous $f : X \to \mathbb{R}$. Then $(x_n) \to x$.

30. For $f, g \in C([0, 1])$, let $d^*(f, g) = \sup\{|f(x) - g(x)| : x \in [0, 1]\}$. Let (f_n) be a sequence such that $(f_n) \to f$ in $(C([0, 1]), d^*)$, where f is the function $f(x) = x + 2$,
Then there is an $N \in \mathbb{N}$ such that whenever $n \geq N$, we have $f_n(x) \geq x$ for all $x \in [0, 1]$.

31. If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then the graph of f is a closed subset of $\mathbb{R}^2$.

32. If the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is closed subset of $\mathbb{R}^2$, then f is continuous.

33. The are exactly c different metrics d on $\mathbb{R}$ for which $\mathcal{T}_d =$ the usual topology on $\mathbb{R}$.

34. In $\mathbb{R}$, the interval $[-2, 1]$ can be written as a countable intersection of open sets.

35. For any $f \in \mathbb{N}^{\mathbb{N}}$, then there is a continuous function $g \in \mathbb{R}^{\mathbb{R}}$ such that $g|\mathbb{N} = f$.

36. Let d be the usual metric on $\mathbb{R}$ and d' the discrete unit metric. Suppose $f : (\mathbb{R}, d) \rightarrow (\mathbb{N}, d')$ is continuous. Then f is constant.

37. If $|X| > 1$, then there are infinitely many different metrics d on X for which $\mathcal{T}_d$ is the discrete topology.

38. The space $\mathbb{P}$ of irrational numbers is separable.

39. Suppose (X, d) and (Y, d') are pseudometric spaces and that $f : X \rightarrow Y$ is continuous at a. If $d(a, b) = 0$, then f is also continuous at b.

40. If A and B are subsets of (X, d), then $\text{int}(A \cap B) = \text{int}(A) \cap \text{int}(B)$.

41. Let d^* be the "max metric" on $\mathbb{R}$ and let d_t be the "taxicab metric" on $\mathbb{R}$. For $x \in \mathbb{R}$, let $f(x) = \cos(x^3)$. Then $f : (\mathbb{R}, d^*) \rightarrow (\mathbb{R}, d_t)$ is continuous.

42. Let d be a pseudometric on the set $X = \{0, 1\}$. Then either $\mathcal{T}_d = \{\emptyset, X\}$ or $\mathcal{T}_d = \mathcal{P}(X)$.

43. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuous and $f(p) = p - \sqrt{2}$ for each <u>irrational</u> p. Then $f(17) \in \mathbb{P}$.

44. A finite set in a metric space must be closed.

45. If $f : \mathbb{R} \rightarrow \mathbb{N}$ is continuous and $f(1) = 1$, then there must exist an irrational number x for which $f(x) = 1$.

46. In a metric space (X, d), it cannot happen that $B_\epsilon(a) = X - B_\epsilon(b)$.

47. The discrete unit metric produces the same topology on $\mathbb{N}$ as the usual metric on $\mathbb{N}$.

48. If D is an uncountable dense subset of $\mathbb{R}$ and C is countable, then $D - C$ is dense in $\mathbb{R}$.

49. Suppose that $f : \ell_2 \rightarrow \mathbb{R}$ is continuous and that $f(x) = 0$ for whenever x is any sequence in ℓ_2 that is eventually 0. Then $f((1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...)) = 0$.

50. If $f : (X, d) \rightarrow (Y, d')$ is continuous and onto, and X is separable, then Y is separable.

51. In a pseudometric space (X, d), $\text{cl}\, A = \text{int}\, A$ if and only if A is clopen.

52. There exists a dense subset, D, of $\mathbb{R}$ such that every infinite countable subset of D is dense in $\mathbb{R}$.

This page is intentionally left blank.

Chapter III
Topological Spaces

1. Introduction

In Chapter I we looked at properties of sets, and in Chapter II we added some additional structure to a set – a distance function d – to create a pseudometric space. We then saw some of the most basic fundamental definitions and properties of pseudometric spaces. There is much more, and some of the most useful and interesting properties of pseudometric spaces will be discussed in Chapter IV. But in Chapter III we look at an important generalization.

We observed, early in Chapter II, that the idea of continuity (in calculus) depends on talking about "nearness," so we used a distance function d to make the idea of "nearness" precise. In that way, we were able to extend the definition of continuity from $\mathbb{R}^n$ to pseudometric spaces. The distance function d also led us to the idea of open sets in a pseudometric space. From there we developed properties of closed sets, closures, interiors, frontiers, dense sets, continuity, and sequential convergence.

An important observation in Chapter II was that open (or closed) sets are really all that we require to talk about many of these ideas. In other words, we can often do what's necessary using the open sets without knowing which specific d generated the open sets: the topology $\mathcal{T}_d$ is what really matters. For example, int A is defined in (X, d) in terms of the open sets, so int A doesn't change if d is replaced with a different but equivalent metric d' – one that generates the same open sets. Similarly, changing d to an equivalent metric d' doesn't affect closures, continuity, or convergent sequences. In summary: for many purposes d is logically unnecessary once d has done the job of creating the topology $\mathcal{T}_d$ (although having d might still be convenient).

This suggests a way to generalize our work. For a particular set X we can simply assign a topology – that is, a collection of "open" sets given without any mention of a pseudometric that might have generated them. Of course when we do this, we want the "open sets" to "behave the way open sets should behave" as described in Theorem II.2.11. This leads us to the definition of a topological space.

2. Topological Spaces

Definition 2.1 A topology $\mathcal{T}$ on a set X is a collection of subsets of X such that

i) $\emptyset, X \in \mathcal{T}$
ii) if $O_\alpha \in \mathcal{T}$ for each $\alpha \in A$, then $\bigcup_{a \in A} O_a \in \mathcal{T}$
iii) if $O_1, ..., O_n \in \mathcal{T}$, then $O_1 \cap ... \cap O_n \in \mathcal{T}$.

A set $O \subseteq X$ is called open if $O \in \mathcal{T}$. The pair $(X, \mathcal{T})$ is called a topological space.

Sometimes we will just refer to "a topological space X." In that case, it is assumed that there is some topology $\mathcal{T}$ on X but, for short, we are just not bothering to write down the "$\mathcal{T}$")

We emphasize that in a topological space there is no distance function d: therefore phrases like "distance between two points" and "ϵ-ball" make no sense in $(X, \mathcal{T})$. There is no preconceived idea about what "open" means: to say "O is open" means nothing more or less than "$O \in \mathcal{T}$."

In a topological space $(X, \mathcal{T})$, we can go on to define closed sets and isolated points just as we did in pseudometric spaces.

Definition 2.2 A subset F in $(X, \mathcal{T})$ is called closed if $X - F$ is open, that is, if $X - F \in \mathcal{T}$.

Definition 2.3 A point $a \in (X, \mathcal{T})$ is called isolated if $\{a\}$ is open, that is, if $\{a\} \in \mathcal{T}$.

The proof of the following theorem is the same as it was for pseudometric spaces; we just take complements and apply properties of open sets.

Theorem 2.4 In any topological space $(X, \mathcal{T})$

i) $\emptyset$ and X are closed
ii) if F_α is closed for each $a \in A$, then $\bigcap_{\alpha \in A} F_\alpha$ is closed
iii) if $F_1, ..., F_n$ are closed, then $\bigcup_{i=i}^{n} F_i$ is closed.

More informally, ii) and iii) state that intersections and finite unions of closed sets are closed.

Proof Read the proof for Theorem II.4.2. •

For a particular topological space $(X, \mathcal{T})$, it might or might not be possible to find a pseudometric d on X that "creates" this topology – that is, one for which $\mathcal{T}_d = \mathcal{T}$.

Definition 2.5 A topological space $(X, \mathcal{T})$ is called pseudometrizable if there exists a pseudometric d on X for which $\mathcal{T}_d = \mathcal{T}$. If d is a metric, then $(X, \mathcal{T})$ is called metrizable.

Examples 2.6

1) Suppose X is a set and $\mathcal{T} = \{\emptyset, X\}$. $\mathcal{T}$ is called the trivial topology on X and it is the smallest possible topology on X. $(X, \mathcal{T})$ is called a trivial topological space. The only open (or closed) sets are $\emptyset$ and X. If we put the trivial pseudometric d on X, then $\mathcal{T}_d = \mathcal{T}$. So a trivial topological space turns out to be pseudometrizable.

At the opposite extreme, suppose $\mathcal{T} = \mathcal{P}(X)$. Then $\mathcal{T}$ is called the discrete topology on X and it is the largest possible topology on X. $(X, \mathcal{T})$ is called a discrete topological space. Every subset is open (and also closed). Every point of X is isolated. If we put the discrete unit metric d (or any equivalent metric) on X, then $\mathcal{T}_d = \mathcal{T}$. So a discrete topological space is metrizable.

2) Suppose $X = \{0, 1\}$ and let $\mathcal{T} = \{\emptyset, \{1\}, X\}$. $(X, \mathcal{T})$ is a topological space called <u>Sierpinski space</u>. In this case it is <u>not</u> possible to find a pseudometric d on X for which $\mathcal{T}_d = \mathcal{T}$, so Sierpinski space is not pseudometrizable. To see this, consider any pseudometric d on X.

If $d(0,1) = 0$, then d is the trivial pseudometric on X and $\{\emptyset, X\} = \mathcal{T}_d \neq \mathcal{T}$.

If $d(0,1) = \delta > 0$, then the open ball $B_\delta(0) = \{0\} \in \mathcal{T}_d$, so $\mathcal{T}_d \neq \mathcal{T}$.
(*In this case* $\mathcal{T}_d$ *is actually the discrete topology:* d *is just a rescaling of the discrete unit metric.*)

Another possible topology on $X = \{0, 1\}$ is $\mathcal{T}' = \{\emptyset, \{0\}, X\}$, although $(X, \mathcal{T})$ and $(X, \mathcal{T}')$ seem very much alike: both are two-point spaces, each with containing exactly one isolated point. One space can be obtained from the other simply renaming "0" and "1" as "1" and "0" respectively. Such "topologically identical" spaces are called "homeomorphic." (We will give a precise definition later in this chapter.)

3) For a set X, let $\mathcal{T} = \{O \subseteq X : O = \emptyset \text{ or } X - O \text{ is finite}\}$. $\mathcal{T}$ is a topology on X :

i) Clearly, $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$.

ii) Suppose $O_\alpha \in \mathcal{T}$ for each $\alpha \in A$. If $\bigcup_{\alpha \in A} O_\alpha = \emptyset$, then $\bigcup_{\alpha \in A} O_\alpha \in \mathcal{T}$. Otherwise there is at least one $O_{\alpha_0} \neq \emptyset$. Then $X - O_{\alpha_0}$ is finite, so $X - \bigcup_{\alpha \in A} O_\alpha = \bigcap_{\alpha \in A} (X - O_\alpha) \subseteq X - O_{\alpha_0}$. Therefore $X - \bigcup_{\alpha \in A} O_\alpha$ is also finite, so $\bigcup_{\alpha \in A} O_\alpha \in \mathcal{T}$.

iii) If $O_1, ..., O_n \in \mathcal{T}$ and some $O_i = \emptyset$, then $\bigcap_{i=1}^n O_i = \emptyset$ so $\bigcap_{i=1}^n O_i \in \mathcal{T}$. Otherwise each O_i is nonempty, so each $X - O_i$ is finite. Then $(X - O_1) \cup ... \cup (X - O_n) = X - \bigcap_{i=1}^n O_i$ is finite, so $\bigcap_{i=1}^n O_i \in \mathcal{T}$.

In $(X, \mathcal{T})$, a set F is closed iff $F = \emptyset$ or F is finite. Because the open sets are $\emptyset$ and the <u>co</u>mplements of <u>finite</u> sets, $\mathcal{T}$ is called the <u>cofinite topology</u> on X.

If X is a finite set, then the cofinite topology is the same as the discrete topology on X. (*Why?*) In X is infinite, then no point in $(X, \mathcal{T})$ is isolated.

Suppose X is an infinite set with the cofinite topology $\mathcal{T}$. If U and V are nonempty open sets, then $X - U$ and $X - V$ must be finite so $(X - U) \cup (X - V) = X - (U \cap V)$ is finite. Since X is infinite, this means that $U \cap V \neq \emptyset$ (*in fact,* $U \cap V$ *must be infinite*). Therefore every pair of nonempty open sets in $(X, \mathcal{T})$ has nonempty intersection! This shows us that an infinite cofinite space $(X, \mathcal{T})$ is not pseudometrizable:

i) if d is the trivial pseudometric on X, then certainly $\mathcal{T}_d \neq \mathcal{T}$, and

ii) if d is not the trivial pseudometric on X, then there exist points $a, b \in X$ for which $d(a,b) = \delta > 0$. In that case, $B_{\delta/2}(a)$ and $B_{\delta/2}(b)$ would be <u>disjoint</u> nonempty open sets in $\mathcal{T}_d$, so $\mathcal{T}_d \neq \mathcal{T}$.

4) On $\mathbb{R}$, let $\mathcal{T} = \{(a, \infty):\ a \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$. It is easy to verify that $\mathcal{T}$ is a topology on $\mathbb{R}$, called the right-ray topology. Is $(\mathbb{R}, \mathcal{T})$ metrizable or pseudometrizable?

If $X = \emptyset$, then $\mathcal{T} = \{\emptyset\}$ is the only possible topology on X, and $\mathcal{T} = \{\emptyset, \{a\}\}$ is the only possible topology on a singleton set $X = \{a\}$. But for $|X| > 1$, there are many possible topologies on X. For example, there are four possible topologies on the set $X = \{a, b\}$. These are the trivial topology, the discrete topology, $\{\emptyset, \{a\}, X\}$, and $\{\emptyset, \{b\}, X\}$ although, as mentioned earlier, the last two can be considered as "topologically identical."

If $\mathcal{T}$ is a topology on X, then $\mathcal{T}$ is a collection of subsets of X, so $\mathcal{T} \subseteq \mathcal{P}(X)$. This means that $\mathcal{T} \in \mathcal{P}(\mathcal{P}(X))$, so $|\mathcal{P}(\mathcal{P}(X))| = 2^{(2^{|X|})}$ is an upper bound for the number of possible topologies on X. For example, there are no more than $2^{2^7} \approx 3.4 \times 10^{38}$ topologies on a set X with 7 elements. But this upper bound is actually very crude, as the following table (given without proof) indicates:

$n = \lvert X \rvert$	Actual number of distinct topologies on X
0	1
1	1
2	4
3	29
4	355
5	6942
6	209527
7	9535241 (*many less than* 10^{38})

Counting topologies on finite sets is really a question about combinatorics and we will not pursue this topic.

Each concept we defined for pseudometric spaces can be carried over directly to topological spaces if the concept was defined in topological terms $-$ that is, in terms of open (or closed) sets. This applies, for example, to the definitions of interior, closure, and frontier in pseudometric spaces, so these definitions can also be carried over verbatim to a topological space $(X, \mathcal{T})$.

Definition 2.7 Suppose $A \subseteq (X, \mathcal{T})$. We define

$\text{int}_X A =$ the interior of A in X $= \bigcup\{O : O \text{ is open and } O \subseteq A\}$
$\text{cl}_X A =$ the closure of A in X $= \bigcap\{F : F \text{ is closed and } F \supseteq A\}$
$\text{Fr}_X A =$ the frontier (or boundary) of A in X $= \text{cl}_X A \cap \text{cl}_X(X - A)$

As before, we will drop the subscript "X" when the context makes it clear.

The properties for the operators cl, int, and Fr (except those that mention a pseudometric d or an ϵ-ball) remain true. The proofs in the preceding chapter were deliberately phrased in topological terms so they would carry over to the more general setting of topological spaces.

Theorem 2.8 Suppose $A \subseteq (X, d)$. Then

1) a) int A is the largest open subset of A (that is, if O is open and $O \subseteq A$, then $O \subseteq$ int A).
   b) A is open iff $A =$ int A (*Note: int $A \subseteq A$ is true for* <u>*every*</u> *set A so we could say: A is open iff $A \subseteq$ int A.*)
   c) $x \in$ int A iff there is an open set O such that $x \in O \subseteq A$

2) a) cl A is the smallest closed set containing A (that is, if F is closed and $F \supseteq A$, then $F \supseteq$ cl A).
   b) A is closed iff $A =$ cl A (*Note: $A \subseteq$ cl A is true for* <u>*every*</u> *set A, so we could say: A is closed iff cl $A \subseteq A$.*)
   c) $x \in$ cl A iff for every open set O containing x, $O \cap A \neq \emptyset$

3) a) Fr A is closed and Fr $A =$ Fr$(X - A)$.
   b) $x \in$ Fr A iff for every open set O containing x, $O \cap A \neq \emptyset$ and $O \cap (X - A) \neq \emptyset$
   c) A is clopen iff Fr $A = \emptyset$.

See the proof of Theorem II.4.5

At this point, we add a few additional facts about these operators. Some of the proofs are left as exercises.

Theorem 2.9 Suppose A, B are subsets of a topological space $(X, \mathcal{T})$. Then

1) cl$(A \cup B) =$ cl$(A) \cup$ cl(B)
2) cl $A = A \cup$ Fr A
3) int $A = A -$ Fr $A = X -$ cl$(X - A)$
4) Fr $A =$ cl $A -$ int A
5) $X =$ int $A \cup$ Fr $A \cup$ int$(X - A)$, and these 3 sets are disjoint.

Proof 1) $A \subseteq A \cup B$ so, from the definition of closure, we have cl $A \subseteq$ cl$(A \cup B)$. Similarly, cl $B \subseteq$ cl$(A \cup B)$ Therefore cl $A \cup$ cl $B \subseteq$ cl$(A \cup B)$.

On the other hand, cl $A \cup$ cl B is the union of two closed sets, so cl $A \cup$ cl B is closed and cl $A \cup$ cl $B \supseteq A \cup B$, so cl $A \cup$ cl $B \supseteq$ cl$(A \cup B)$. (*As an exercise, try proving 1) instead using the characterization of closures given above in Theorem 2.8.2c.*)

Is 1) true if "$\cup$" is replaced by "$\cap$"?

2) Suppose $x \in$ cl A but $x \notin A$. If O is any open set containing x, then $O \cap A \neq \emptyset$ (because $x \in$ cl A) and $O \cap (X - A) \neq \emptyset$ (because the intersection contains x). Therefore $x \in$ Fr A, so $x \in A \cup$ Fr A.

Conversely, suppose $x \in A \cup$ Fr A. If $x \in A$, then $x \in$ cl A. And if $x \notin A$, then $x \in$ Fr $A =$ cl $A \cap$ cl$(X - A)$, so $x \in$ cl A. Therefore cl $A = A \cup$ Fr A.

The proofs of 3) $-$ 5) are left as exercises. $\bullet$

Theorem 2.9 shows us that complements, closures, interiors and frontiers are interrelated and therefore some of these operators are redundant. That is, if we wanted to very "economical," we could discard some of them. For example, we could avoid using "Fr" and "int" and just use "cl" and complement because Fr $A = \text{cl}\, A \cap \text{cl}\,(X - A)$ and int $A = A - \text{Fr}\, A$
$= A - (\text{cl}\, A \cap \text{cl}\,(X - A))$. Of course, the most economical way of doing things is not necessarily the most convenient. (*Could we get by only using complements and "Fr" – that is, can we define "int" and ""cl" in terms of "Fr" and complements? Or could we use just "int" and complements?*)

Here is a famous related problem from the early days of topology: for $A \subseteq (X, \mathcal{T})$, is there an upper bound for the number of different subsets of X which might created from A using only complements and closures, repeated in any order? (*As we just observed, using the interior and frontier operators would not help to create any additional sets.*) For example, one might start with A and then consider such sets as cl A, $X - \text{cl}A$, $\text{cl}(X - \text{cl}(X - A))$, and so on. An old theorem of Kuratowski (1922) says that for any set A in any space $(X, \mathcal{T})$, the upper bound is 14. Moreover, this upper bound is "sharp" – there is a set $A \subseteq \mathbb{R}$ from which 14 different sets can actually be obtained! Can you find such a set?

Definition 2.10 Suppose $(X, \mathcal{T})$ is a topological space and $D \subseteq X$. D is called dense in X if cl $D = X$. The space $(X, \mathcal{T})$ is called separable if there exists a countable dense set D in X.

Example 2.11 Let $\mathcal{T}$ be the cofinite topology on $\mathbb{R}$. Let $\mathbb{E} = \{2, 4, 6, ...\}$.

int $\mathbb{E} = \emptyset$, because each nonempty open set O has finite complement and therefore $O \not\subset E$. In fact, for any $B \subseteq \mathbb{R}$: if $\mathbb{R} - B$ is infinite, then int $B = \emptyset$

cl $\mathbb{E} = \mathbb{R}$ because the only closed set containing $\mathbb{E}$ is $\mathbb{R}$. Therefore $(\mathbb{R}, \mathcal{T})$ is separable. In fact, any infinite set B is dense.

Fr $\mathbb{E} = \text{cl}\,(\mathbb{E}) \cap \text{cl}\,(\mathbb{R} - \mathbb{E}) = \mathbb{R} \cap \mathbb{R} = \mathbb{R}$

$(\mathbb{R}, \mathcal{T})$ is not pseudometrizable (*why?*)

Example 2.12 Let $\mathcal{T}$ be the right-ray topology on $\mathbb{R}$. In $(\mathbb{R}, \mathcal{T})$,

int $\mathbb{Z} = \emptyset$

cl $\mathbb{Z} = \mathbb{R}$, so $(\mathbb{R}, \mathcal{T})$ is separable

Fr $\mathbb{Z} = \mathbb{R}$

Any two nonempty open sets intersect, so $(\mathbb{R}, \mathcal{T})$ is not metrizable. Is it pseudometrizable?

3. Subspaces

There is a natural way to create a new topological space from a subset of a given topological space. The new space is called a subspace (not merely a subset) of the original space.

Definition 3.1 Suppose $(X, \mathcal{T})$ is a topological space and that $A \subseteq X$. The subspace topology on A is defined as $\mathcal{T}_A = \{A \cap O : O \in \mathcal{T}\}$ and $(A, \mathcal{T}_A)$ is called a subspace of $(X, \mathcal{T})$.
We sometimes call $O \cap A$ the "restriction of O to A" or the "trace of O on A"

You should check the conditions in Definition 2.1 are satisfied: in other words, that $\mathcal{T}_A$ really is a topology on A. When we say that A is a subspace of $(X, \mathcal{T})$, we mean that A is a subset of X with the subspace topology $\mathcal{T}_A$. To indicate that A is a subspace, we sometimes write $A \subseteq (X, \mathcal{T})$ rather than $A \subseteq X$.

Example 3.2 Consider $\mathbb{N} \subseteq \mathbb{R}$, where $\mathbb{R}$ has its usual topology. For each $n \in \mathbb{N}$, the interval $(n-1, n+1)$ is open in $\mathbb{R}$. Therefore $(n-1, n+1) \cap \mathbb{N} = \{n\}$ is open in the subspace $\mathbb{N}$, so every point n is isolated in the subspace. The subspace topology is the discrete topology. Notice: the subspace topology on $\mathbb{N}$ is the same as what we get if we use the usual metric on $\mathbb{N}$ to generate open sets in $\mathbb{N}$. Similarly, it is easy to check that in $\mathbb{R}^2$ the subspace topology on the x-axis is the same as the usual metric topology on $\mathbb{R}$.

More generally: suppose $A \subseteq X$, where (X, d) is a pseudometric space. Then we can think of two ways to make A into a topological space.

i) d gives a topology $\mathcal{T}_d$ on X. Take the open sets in $\mathcal{T}_d$ and intersect them with A. This gives us the subspace topology on A, which we call $(\mathcal{T}_d)_A$ and $(A, (\mathcal{T}_d)_A)$ is a subspace of $(X, \mathcal{T})$.

2) Or, we could view A as a pseudometric space by using d to measure distances in A. To be very precise, $d : X \times X \rightarrow \mathbb{R}$, so we make a "new" pseudometric d' defined by $d' = d|A \times A$. Then (A, d') is a pseudometric space and we can use d' to generate open sets in A: the topology $\mathcal{T}_{d'}$.
Usually, we would be less compulsive about notation and continue to use the name "d" also for the pseudometric on A. But for a moment it will be helpful to distinguish carefully between d and $d' = d|A \times A$.

Fortunately, it turns out that 1) and 2) produce the same open sets in A: the open sets in (A, d') are just the open sets from (X, d) restricted to A. That's just what Theorem 3.3 says, in "fancier" notation.

Theorem 3.3 Suppose $A \subseteq X$, where (X, d) is a pseudometric space. Then $(\mathcal{T}_d)_A = \mathcal{T}_{d'}$.

Proof If $U \in (\mathcal{T}_d)_A$, then $U = O \cap A$ where $O \in \mathcal{T}_d$. Let $a \in U$. There is an $\epsilon > 0$ such that $B^d_\epsilon(a) \subseteq O$. Since $d' = d|A \times A$, we get that $B^{d'}_\epsilon(a) = B^d_\epsilon(a) \cap A \subseteq O \cap A = U$, so $U \in \mathcal{T}_{d'}$.
Conversely, suppose $U \in \mathcal{T}_{d'}$. For each $a \in U$ there is an $\epsilon_a > 0$ such that $B^{d'}_{\epsilon_a}(a) \subseteq U$, and $U = \bigcup_{a \in U} B^{d'}_{\epsilon_a}(a)$. Let $O = \bigcup_{a \in U} B^d_{\epsilon_a}(a)$, an open set in $\mathcal{T}_d$. Since $B^{d'}_{\epsilon_a}(a) = B^d_{\epsilon_a}(a) \cap A$, we get $O \cap A = (\bigcup_{a \in O} B^d_{\epsilon_a}(a)) \cap A = (\bigcup_{a \in O} B^d_{\epsilon_a}(a) \cap A)$
$= \bigcup_{a \in U} B^{d'}_{\epsilon_a}(a) = U$. Therefore $U \in (\mathcal{T}_d)_A$. ●

Exercise Verify that in any topological space $(X, \mathcal{T})$

i) If U is open in $(X, \mathcal{T})$ and A is an open set in the subspace U, then A is open in X. (*"An open subset of an open set is open."*)

ii) If F is closed in $(X, \mathcal{T})$ and A is a closed set in the subspace F, then F is closed in X. (*"A closed subset of a closed set is closed."*)

4. Neighborhoods

Definition 4.1 Let $(X, \mathcal{T})$ be a topological space and suppose $N \subseteq X$. If $x \in \operatorname{int} N$, then we say that N is a neighborhood of x.

The collection $\mathcal{N}_x = \{N \subseteq X : N \text{ is a neighborhood of } x\}$ is called the neighborhood system at x.

Note that

1) $\mathcal{N}_x \neq \emptyset$, because every point x has at least one neighborhood – for example, $X \in \mathcal{N}_x$.
2) If N_1 and $N_2 \in \mathcal{N}_x$, then $x \in \operatorname{int} N_1 \cap \operatorname{int} N_2 =$ (*why?*) $\operatorname{int}(N_1 \cap N_2)$. Therefore $N_1 \cap N_2 \in \mathcal{N}_x$.
3) If $N \in \mathcal{N}_x$ and $N \subseteq N'$, then $x \in \operatorname{int} N \subseteq \operatorname{int} N'$, so $N' \in \mathcal{N}_x$ (that is, if N' contains a neighborhood of x, then N' is also a neighborhood of x.)

Just as in pseudometric spaces, it is clear that a set O in $(X, \mathcal{T})$ is open iff O is a neighborhood of each of its points. (*See Theorem II.5.4*)

In a pseudometric space, we use the ϵ-balls centered at x to measure "nearness" to x. For example, if "every ϵ-ball in $\mathbb{R}$ centered at x contains an irrational number," this tells us that "there are irrational numbers arbitrarily near to x." Of course, we could convey the same information in terms of neighborhoods by saying "every neighborhood of x in $\mathbb{R}$ contains an irrational number." Or, instead, we could say it in terms of "open sets": "every open set in $\mathbb{R}$ containing x contains an irrational number." These are all equivalent ways to say "there are irrational numbers arbitrarily near to x." *This isn't that surprising since open sets and neighborhoods in (X, d) were defined in terms of ϵ-balls.*

In a topological space $(X, \mathcal{T})$ we don't have ϵ-balls, but we still have open sets and neighborhoods. We now think of the neighborhoods in $\mathcal{N}_x$ (or, if we prefer, the collection of open sets containing x) as the tool we use to talk about "nearness to x."

For example, suppose X has the trivial topology $\mathcal{T}$. For any $x \in X$, the only neighborhood of x is X: therefore every y in X is in every neighborhood of x : the neighborhoods of x are unable to "separate" x and y, and that's analogous to having $d(x, y) = 0$ (if we had a pseudometric). In that sense, all points in $(X, \mathcal{T})$ are "very close together": so close together, in fact, that they are "indistinguishable." The neighborhoods of x tell us this.

At the opposite extreme, suppose X has the discrete topology $\mathcal{T}$ and that $x \in X$. If $x \in W$ then (*since W is open*), W is a neighborhood of x. The smallest neighborhood of x is $N = \{x\}$. So every point x has a neighborhood N that excludes all other points y: for every $y \neq x$, we could say "y is not within the neighborhood N of x." This is analogous to saying "y is not within ϵ of x" (if we had a pseudometric). Because no point y is "within N of x," we call x isolated. The neighborhoods of x tell us this.

Of course, if we prefer, we could use "U is an open set containing x" instead of "N is a neighborhood of x" to talk about nearness to x.

The complete neighborhood system $\mathcal{N}_x$ of a point x often contains more neighborhoods than we actually need to talk about nearness to x. For example, the open balls $B_{\frac{1}{n}}(x)$ in a pseudometric space (X, d) are enough to let us talk about continuity at x. Therefore, we introduce the idea of a neighborhood base at x to choose a smaller collection of neighborhoods of x that is

i) good enough for all our purposes, and
ii) from which all the other neighborhoods of x can be obtained if we want them.

Definition 4.2 A collection $\mathcal{B}_x \subseteq \mathcal{N}_x$ is called a neighborhood base at x if for every neighborhood N of x, there is a neighborhood $B \in \mathcal{B}_x$ such that $x \in B \subseteq N$. We refer to the sets in $\mathcal{B}_x$ as basic neighborhoods of x.

According to the definition, each set in $\mathcal{B}_x$ must be a neighborhood of x, but the collection $\mathcal{B}_x$ may be much simpler than the whole neighborhood system N_x. The crucial thing is that every neighborhood N of x must contain a basic neighborhood B of x.

Example 4.3 In (X, d), possible ways to choose a neighborhood base at x include:

i) $\mathcal{B}_x$ = the collection of all balls $B_\epsilon(x)$, or
ii) $\mathcal{B}_x$ = the collection of all balls $B_\epsilon(x)$, where ϵ is a positive rational, or
iii) $\mathcal{B}_x$ = the collection of all balls $B_{\frac{1}{n}}(x)$ for $n \in \mathbb{N}$, or
iv) $\mathcal{B}_x = \mathcal{N}_x$ (*The neighborhood system is always a base for itself, but it is not an "efficient" choice; the goal is to get a base $\mathcal{B}_x$ that's much simpler than N_x.*)

Which $\mathcal{B}_x$ to use is our choice: each of i)-iv) gives a neighborhood base at x. But ii) or iii) might be more convenient – because ii) and iii) are countable neighborhood bases $\mathcal{B}_x$. (*If $X = \mathbb{R}$, for example, with the usual metric d, then the collections $\mathcal{B}_x$ in i) and iv) are uncountable.*) Of the four, iii) is probably the simplest choice for $\mathcal{B}_x$.

Suppose we want to check whether some property that involves neighborhoods of x is true. Often all we need to do is to check whether the property holds for neighborhoods in the simpler collection $\mathcal{B}_x$. For example, in (X, d) : O is open iff O contains a neighborhood N of each $x \in O$. But that is true iff O contains a set $B_{\frac{1}{n}}(x)$ around each of its points x.

Similarly, $x \in \text{cl}\, A$ iff $N \cap A \neq \emptyset$ for every $N \in \mathcal{N}_x$ iff $B \cap A \neq \emptyset$ for every $B \in \mathcal{B}_x$. For example, suppose we want to check, in $\mathbb{R}$, that $1 \in \text{cl}\, \mathbb{P}$. It is sufficient just to check that $B_{\frac{1}{n}}(1) \cap \mathbb{P} \neq \emptyset$ for each $n \in \mathbb{N}$, because this implies that $N \cap \mathbb{P} \neq \emptyset$ for every $N \in \mathcal{N}_1$.

Therefore it's often desirable to make an "efficient" choice of neighborhood base $\mathcal{B}_x$ at each point $x \in X$.

Definition 4.4 We say that a space $(X, \mathcal{T})$ satisfies the first axiom of countability (or, more simply, that $(X, \mathcal{T})$ is first countable) if at each point $x \in X$, it is possible to choose a countable neighborhood base $\mathcal{B}_x$.

Example 4.5

1) The preceding Example 4.3 shows that every pseudometric space is first countable.

2) If $\mathcal{T}$ is the discrete topology on X, then $(X, \mathcal{T})$ is first countable. In fact, at each point x, we can choose a neighborhood base that consists of a single set: $\mathcal{B}_x = \{\,\{x\}\,\}$.

3) Let $\mathcal{T}$ be the cofinite topology on an uncountable set X. For any $x \in X$, there cannot be a countable neighborhood base $\mathcal{B}_x$ at x.

We prove this by "contradiction": suppose there were a countable neighborhood base at some point x : call it $\mathcal{B}_x = \{B_1, ..., B_n, ...\}$.
For any $y \neq x$, $\{y\}$ is closed, so $X - \{y\}$ is a neighborhood of x. Then, by the definition of neighborhood base, there is some k for which

$x \in \text{int}\, B_k \subseteq B_k \subseteq X - \{y\}$. Therefore $y \notin \bigcap_{n=1}^{\infty} \text{int}\, B_n$.

But $x \in \bigcap_{n=1}^{\infty} \text{int}\, B_n$ so $\bigcap_{n=1}^{\infty} \text{int}\, B_n = \{x\}$.

Then $X - \{x\} = X - \bigcap_{n=1}^{\infty} \text{int}\, B_n = \bigcup_{n=1}^{\infty} (X - \text{int}\, B_n)$. Since $X - \text{int}\, B_n$ is finite (*why*?), this would mean that $X - \{x\}$ is countable $-$ which is impossible.

Since any pseudometric space is first countable, the example gives us another way to see that this space $(X, \mathcal{T})$ is not pseudometrizable.

In $(X, \mathcal{T})$, the neighborhood system $\mathcal{N}_x$ at each point x is completely determined by the topology $\mathcal{T}$, but $\mathcal{B}_x$ is not. As the preceding examples illustrate, there are usually many possible choices for $\mathcal{B}_x$. (*Can you describe all the spaces $(X, \mathcal{T})$ for which $\mathcal{B}_x$ is uniquely determined at each point x?*)

On the other hand, if we were given $\mathcal{B}_x$ at each point $x \in X$ we could

1) "reconstruct" the whole neighborhood system $\mathcal{N}_x$:

$\mathcal{N}_x = \{N \subseteq X : \exists B \in \mathcal{B}_x$ such that $x \in B \subseteq N\}$, and then we could

2) "reconstruct" the whole topology $\mathcal{T}$:

$\mathcal{T} = \{O : O$ is a neighborhood of each of its points$\}$
$\quad = \{O : \forall x \in O\ \exists B \in \mathcal{B}_x\ \ x \in B \subseteq O\}$, that is,
O is open iff O contains a basic neighborhood of each of its points.

This illustrates one method of describing a topology: by telling the neighborhood basis $\mathcal{B}_x$ at each point. Various effective methods to describe topologies are discussed in the next section.

5. Describing Topologies

How can a topological space be described? If $X = \{0, 1\}$, it is simple to give a topology by just writing $\mathcal{T} = \{\emptyset, \{0\}, X\}$. However, describing all the sets in $\mathcal{T}$ explicitly is often not the easiest way to go.

In this section we look at three important – different but closely related – ways to define a topology on a set. All of them will be used throughout the course. A fourth method – by using a "closure operator" – is not used much nowadays. It is included just as an historical curiosity.

A. Basic neighborhoods at each point

Suppose that at each point $x \in (X, \mathcal{T})$ we have picked a neighborhood base $\mathcal{B}_x$. As mentioned above, the collections $\mathcal{B}_x$ implicitly contain all the information about the topology: a set O is in $\mathcal{T}$ iff O contains a basic neighborhood of each of its points. This suggests that if we start with just a set X, then we could define a topology on X if we begin by saying what the $\mathcal{B}_x$'s should be. Of course, we can't just put "random" sets in $\mathcal{B}_x$: the sets in each $\mathcal{B}_x$ must "act the way basic neighborhoods are supposed to act." And how is that? The next theorem describes the crucial behavior of a collection of basic neighborhoods at x in any topological space.

Theorem 5.1 Suppose $(X, \mathcal{T})$ is a topological space and that for each point $x \in X$, $\mathcal{B}_x$ is a neighborhood base.

1) $\mathcal{B}_x \neq \emptyset$ and $B \in \mathcal{B}_x \Rightarrow x \in B \subseteq X$

2) if B_1 and $B_2 \in \mathcal{B}_x$, then $\exists\, B_3 \in \mathcal{B}_x$ such that $x \in B_3 \subseteq B_1 \cap B_2$

3) if $B \in \mathcal{B}_x$, then $\exists\, I$ such that $x \in I \subseteq B$ and,
$\forall\, y \in I$, $\exists\, B_y \in \mathcal{B}_y$ such that $y \in B_y \subseteq I$

4) $O \in \mathcal{T} \Leftrightarrow \forall x \in O\ \exists\, B \in \mathcal{B}_x$ such that $x \in B \subseteq O$.

Proof 1) Since X is a neighborhood of x, there is a $B \in \mathcal{B}_x$ such that $x \in B \subseteq X$. Therefore $\mathcal{B}_x \neq \emptyset$. If $B \in \mathcal{B}_x \subseteq \mathcal{N}_x$, then B is a neighborhood of x, so $x \in \operatorname{int} B \subseteq B \subseteq X$.

2) The intersection of the two neighborhoods B_1 and B_2 of x is a neighborhood of x. Therefore, by the definition of a neighborhood base, there is a set $B_3 \in \mathcal{B}_x$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

3) Let $I = \operatorname{int} B$. Then $x \in I \subseteq B$ and because I is open, I is a neighborhood of each its points y. Since $\mathcal{B}_y$ is a neighborhood base at y, there is a set $B_y \in \mathcal{B}_y$ such that $y \in B_y \subseteq I$.

4) $\Rightarrow$: If O is open, then O is a neighborhood of each of its points x. Therefore for each $x \in O$ there must be a set $B \in \mathcal{B}_x$ such that $x \in B \subseteq O$.

$\Leftarrow$: The condition implies that O contains a neighborhood of each of its points. Therefore O is a neighborhood of each of its points, so O is open. ●

Theorem 5.1 lists the crucial features of the behavior of a neighborhood base at x. The next theorem tells us that we can put a topology on a set X by assigning a "properly behaved" collection of sets to become the basic neighborhoods at each point x.

Theorem 5.2 (The Neighborhood Base Theorem) Let X be a set. Suppose that for each $x \in X$ we give a collection $\mathcal{B}_x$ of subsets of X in such a way that conditions 1) - 3) of Theorem 5.1 are true. Define $\mathcal{T} = \{O \subseteq X : \forall x \in O \; \exists B \in \mathcal{B}_x$ such that $x \in B \subseteq O\}$. Then $\mathcal{T}$ is a topology on X and $\mathcal{B}_x$ is now a neighborhood base at x in $(X, \mathcal{T})$.

Note: In Theorem 5.2, we do not ask that the $\mathcal{B}_x$'s satisfy condition 4) of Theorem 5.1 – since X is a set with no topology (yet), condition 4) would be meaningless. Rather, Condition 4) becomes the motivation for how to define a topology $\mathcal{T}$ using the B_x's.

Proof We need to prove three things: a) $\mathcal{T}$ is a topology
b) each $B \in \mathcal{B}_x$ is now a neighborhood of x in $(X, \mathcal{T})$, and
c) the collection $\mathcal{B}_x$ is now a neighborhood base at x.

a) Clearly, $\emptyset \in \mathcal{T}$. If $x \in X$ then, by condition 1), we can choose any $B \in \mathcal{B}_x$ and $x \in B \subseteq X$. Therefore $X \in \mathcal{T}$.

Suppose $O_\alpha \in \mathcal{T}$ for all $\alpha \in A$. If $x \in \bigcup\{O_\alpha : \alpha \in A\}$, then $x \in O_{\alpha_0}$ for some $\alpha_0 \in A$. By definition of $\mathcal{T}$, there is a set $B \in \mathcal{B}_x$ such that $x \in B \subseteq O_{\alpha_0} \subseteq \bigcup\{O_\alpha : \alpha \in A\}$, so $\bigcup\{O_\alpha : \alpha \in A\} \in \mathcal{T}$.

To finish a), it is sufficient to show that if O_1 and $O_2 \in \mathcal{T}$, then $O_1 \cap O_2 \in \mathcal{T}$. Suppose $x \in O_1 \cap O_2$. By the definition of $\mathcal{T}$, there are sets B_1 and $B_2 \in \mathcal{B}_x$ such that $x \in B_1 \subseteq O_1$ and $x \in B_2 \subseteq O_2$, so $x \in B_1 \cap B_2 \subseteq O_1 \cap O_2$. By condition 2), there is a set $B_3 \in \mathcal{B}_x$ such that $x \in B_3 \subseteq B_1 \cap B_2 \subseteq O_1 \cap O_2$. Therefore $O_1 \cap O_2 \in \mathcal{T}$.

Therefore $\mathcal{T}$ is a topology on X – so now we have a topological space $(X, \mathcal{T})$ – and we must show that $\mathcal{B}_x$ is a neighborhood base at x in $(X, \mathcal{T})$. Doing so involves the awkward-looking condition 3) – which we have not yet used.

b) If $B \in \mathcal{B}_x$, then $x \in B$ (by condition 1) and (by condition 3), there is a set $I \subseteq X$ such that $x \in I \subseteq B$ and $\underline{\forall\, y \in I,\ \exists\, B_y \in \mathcal{B}_y \text{ such that } y \in B_y \subseteq I}$. The underlined phrase states that I satisfies the condition for $I \in \mathcal{T}$, so I is open. Since I is open and $x \in I \subseteq B$, B is a neighborhood of x, that is, $\mathcal{B}_x \subseteq \mathcal{N}_x$.

c) To complete the proof, we have to check that $\mathcal{B}_x$ is a neighborhood base at x. If N is a neighborhood of x, then $x \in \operatorname{int} N$. Since $\operatorname{int} N$ is open, $\operatorname{int} N$ must satisfy the criterion for membership in $\mathcal{T}$, so there is a set $B \in \mathcal{B}_x$ such that $x \in B \subseteq \operatorname{int} N \subseteq N$. Therefore $\mathcal{B}_x$ forms a neighborhood base at x. ∙

Example 5.3 For each $x \in \mathbb{R}$, let $\mathcal{B}_x = \{[x, b) : b > x\}$. We can easily check the conditions 1) - 3) from Theorem 5.2:

1) For each $x \in \mathbb{R}$, certainly $\mathcal{B}_x \neq \emptyset$ and $x \in [x, b)$ for each set $[x, b) \in \mathcal{B}_x$
2) If $B_1 = [x, b_1)$ and $B_2 = [x, b_2)$ are in $\mathcal{B}_x$, then (in this example) we can choose $B_3 = B_1 \cap B_2 = [x, b_3) \in \mathcal{B}_x$, where $b_3 = \min\{b_1, b_2\}$.
3) If $B = [x, b) \in \mathcal{B}_x$, then (in this example) we can let $I = B$. If $y \in I = [x, b)$, pick c so $y < c < b$. Then $B_y = [y, c) \in \mathcal{B}_y$ and $y \in [y, c) \subseteq I$.

According to Theorem 5.2, $\mathcal{T} = \{O \subseteq \mathbb{R} : \forall x \in O\ \exists\, [x, b) \in \mathcal{B}_x$ such that $x \in [x, b) \subseteq O\}$ is a topology on $\mathbb{R}$ and $\mathcal{B}_x$ is a neighborhood base at x in $(\mathbb{R}, \mathcal{T})$. The space $(\mathbb{R}, \mathcal{T})$ is called the Sorgenfrey line.

Notice that in this example each set $[x, b) \in \mathcal{T}$: that is, the sets in $\mathcal{B}_x$ turn out to be open, not merely neighborhoods of x. (*This does not always happen.*)

It is easy to check that sets of form $(-\infty, x)$ and $[b, \infty)$ are open, so $(-\infty, x) \cup [b, \infty)$ $= \mathbb{R} - [x, b)$ is open. Therefore $[x, b)$ is also closed. So at each point x in the Sorgenfrey line, there is a neighborhood base $\mathcal{B}_x$ consisting of clopen sets.

We can write $(a, c) = \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, c)$, so (a, c) is open in the Sorgenfrey line. Because every usual open set in $\mathbb{R}$ is a union of sets of the form (a, c), we conclude that every usual open set in $\mathbb{R}$ is also open in the Sorgenfrey line. The usual topology on $\mathbb{R}$ is strictly smaller that the Sorgenfrey topology: $\mathcal{T}_d \underset{\neq}{\subseteq} \mathcal{T}$.

$\mathbb{Q}$ is dense in the Sorgenfrey line: if $x \in \mathbb{R}$, then every basic neighborhood $[x, b)$ of x intersects $\mathbb{Q}$, so $x \in \operatorname{cl} \mathbb{Q}$. Therefore the Sorgenfrey line is separable. It is also clear that the Sorgenfrey line is first countable: at each point x the collection $\{[x, x + \frac{1}{n}) : n \in \mathbb{N}\}$ is a countable neighborhood base.

Example 5.4 Similarly, we can define the Sorgenfrey plane by putting a new topology on $\mathbb{R}^2$. At each point $(x, y) \in \mathbb{R}^2$, let $\mathcal{B}_{(x,y)} = \{[x, b) \times [y, c) : b > x,\ c > y\}$. The families $\mathcal{B}_{(x,y)}$ satisfy the conditions in the Neighborhood Base Theorem, so they give a topology $\mathcal{T}$ for which $\mathcal{B}_{(x,y)}$ is a neighborhood base at (x, y). A set $O \subseteq \mathbb{R}^2$ is open iff: for each $(x, y) \in O$, there are $b > x$ and $c > y$ such that $(x, y) \in [x, b) \times [y, c) \subseteq O$. (*Make a sketch!*) You should check that the sets $[x, b) \times [y, c) \in \mathcal{B}_{(x,y)}$ are actually clopen in the Sorgenfrey plane. It is also easy to check the usual topology $\mathcal{T}_d$ on the plane is strictly smaller that the Sorgenfrey topology. It is clear that $\mathbb{Q}^2$ is dense, so $(\mathbb{R}^2, \mathcal{T})$ is separable. Is the Sorgenfrey plane first countable?

Example 5.5 At each point $p \in \mathbb{R}^2$, let $C_\epsilon(p) = \{x \in \mathbb{R}^2 : d(x, p) \leq \epsilon\}$ and define $\mathcal{B}_p = \{C_\epsilon(p) : p \in \mathbb{R}^2\}$. It is easy to check that the conditions 1) - 3) of Theorem 5.2 are satisfied. (*For $C_\epsilon(p)$ in condition 3), let $I = B_\epsilon(p)$.*) The topology generated by the $\mathcal{B}_p$'s is just the usual topology; the sets in $\mathcal{B}_p$ are basic neighborhoods of p in the usual topology – as they should be – but the sets in $\mathcal{B}_p$ did not turn out to be open sets.

Example 5.6 Let $\Gamma = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$ = the "closed upper half-plane."

For a point $p = (x, y) \in \Gamma$ with $y > 0$, let $\mathcal{B}_p = \{B_\epsilon(p) : \epsilon < |\,y\,|\}$
For a point $p = (x, 0) \in \Gamma$, let

$$\mathcal{B}_p = \{\{p\} \cup A : A \text{ is a usual open disc in the upper half-plane, tangent to the } x\text{-axis at } p\}.$$

It is easy to check that the collections $\mathcal{B}_p$ satisfy the conditions 1) - 3) of Theorem 5.2 and therefore give a topology on Γ. In this topology, the sets in $\mathcal{B}_p$ turn out to be open neighborhoods of p.

The space Γ, with this topology, is called the "Moore plane." Notice that Γ is separable and first countable. The subspace topology on the x-axis is the discrete topology. (*Verify these statements!*)

B. Base for the topology

Definition 5.7 A collection of open sets in $(X, \mathcal{T})$ is called a base for the topology $\mathcal{T}$ if each $O \in \mathcal{T}$ is a union of sets from $\mathcal{B}$. More precisely, $\mathcal{B}$ is a base if $\mathcal{B} \subseteq \mathcal{T}$ and for each $O \in \mathcal{T}$, there exists a subfamily $\mathcal{A} \subseteq \mathcal{B}$ such that $O = \bigcup \mathcal{A}$. We also call $\mathcal{B}$ a base for the open sets and we refer to the open sets in $\mathcal{B}$ as basic open sets.

If $\mathcal{B}$ is a base, then it is easy to see that: $O \in \mathcal{T}$ iff $\forall x \in O\ \exists B \in \mathcal{B}$ such that $x \in B \subseteq O$. This means that if we were given $\mathcal{B}$, we could use it to decide which sets are open and thus "reconstruct" $\mathcal{T}$.

Of course, one example of a base is $\mathcal{B} = \mathcal{T}$: every topology $\mathcal{T}$ is a base for itself. But usually there are many ways to choose a base, and the idea is that a simpler base $\mathcal{B}$ would be easier to work with. For example, the set of all balls is a base for the topology $\mathcal{T}_d$ in any pseudometric space (X, d); a different base would be the set containing only the balls with positive rational radii. (*Can you describe those topological spaces for which $\mathcal{T}$ is the only base for $\mathcal{T}$?*)

The following theorem tells us the crucial properties of a base $\mathcal{B}$ in $(X, \mathcal{T})$.

Theorem 5.8 If $(X, \mathcal{T})$ is a topological space with a base $\mathcal{B}$ for $\mathcal{T}$, then

1) $X = \bigcup\{B : B \in \mathcal{B}\}$

2) if B_1 and $B_2 \in \mathcal{B}$ and $x \in B_1 \cap B_2$, then there is a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

Proof 1) Certainly $\bigcup \mathcal{B} \subseteq X$ and – since X is open – the definition of base implies that X is the union of a subfamily of $\mathcal{B}$. Therefore $\bigcup \mathcal{B} = X$.

2) If B_1 and $B_2 \in \mathcal{B}$, then $B_1 \cap B_2$ is open so $B_1 \cap B_2$ must be the union of some sets from $\mathcal{B}$. Therefore, if $x \in B_1 \cap B_2$, there must be a set $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$. ●

The next theorem tells us that if we are given a collection $\mathcal{B}$ of subsets of a set X with properties 1) and 2), we can use it to define a topology.

Theorem 5.9 (The Base Theorem) Suppose X is a set and that $\mathcal{B}$ is a collection of subsets of X that satisfies conditions 1) and 2) in Theorem 5.8. Define $\mathcal{T} = \{O \subseteq X : O$ is a union of sets from $\mathcal{B}\} = \{O \subseteq X : \forall x \in O\ \exists B \in \mathcal{B}$ such that $x \in B \subseteq O\}$.
Then $\mathcal{T}$ is a topology on X and $\mathcal{B}$ is a base for $\mathcal{T}$.

Proof First we show that $\mathcal{T}$ is a topology on X. Since $\emptyset$ is the union of the empty subfamily of $\mathcal{B}$, we get that $\emptyset \in \mathcal{T}$, and condition 1) simply states that $X \in \mathcal{T}$.

If $O_\alpha \in \mathcal{T}$ $(\alpha \in A)$, then $\bigcup\{O_\alpha : \alpha \in A\}$ is a union of sets O_α each of which is itself a union of sets from $\mathcal{B}$. Then clearly $\bigcup\{O_\alpha : \alpha \in A\}$ is a union of sets from $\mathcal{B}$, so $\bigcup\{O_\alpha : \alpha \in A\} \in \mathcal{T}$.

Suppose O_1 and $O_2 \in \mathcal{T}$ and that $x \in O_1 \cap O_2$. For each such x we can use 2) to pick a set $B_x \in \mathcal{B}$ such that $x \in B_x \subseteq O_1 \cap O_2$. Then $O_1 \cap O_2$ is the union of all the B_x's chosen in this way, so $O_1 \cap O_2 \in \mathcal{T}$.

Now we know that we have a topology, $\mathcal{T}$, on X. By definition of $\mathcal{T}$ it is clear that $\mathcal{B} \subseteq \mathcal{T}$ and that each set in $\mathcal{T}$ is a union of sets from $\mathcal{B}$. Therefore $\mathcal{B}$ is a base for $\mathcal{T}$. ●

Example 5.10 The collections

$$\begin{aligned}
\mathcal{B} &= \{B_p(\epsilon) : p \in \mathbb{R}^2, \epsilon > 0\} \\
\mathcal{B}' &= \{B_p(\tfrac{1}{k}) : p \in \mathbb{Q}^2, k \in \mathbb{N}\} \text{ and} \\
\mathcal{B}'' &= \{(a,b) \times (c,d) : a,b,c,d \in \mathbb{R},\ a < b,\ c < d\}
\end{aligned}$$

each satisfy the conditions 1) - 2) in Theorem 5.9, so each collection is the base for a topology on $\mathbb{R}^2$. In fact, all three are bases for the same topology on $\mathbb{R}^2$ – that is, the usual topology (*check this!*) Of the three, $\mathcal{B}'$ is the simplest choice – it is a countable base for the usual topology.

Example 5.11 Suppose $(X, \mathcal{T}')$ and $(Y, \mathcal{T}'')$ are topological spaces. Let $\mathcal{B} =$ the set of "open boxes" in $X \times Y = \{U \times V : U \in \mathcal{T}'$ and $V \in \mathcal{T}''\}$. (*Verify that $\mathcal{B}$ satisfies conditions 1) and 2) of The Base Theorem.*) The product topology on the set $X \times Y$ is the topology for which $\mathcal{B}$ is a base. We always assume that $X \times Y$ has the product topology unless something else is stated.

Therefore a set $A \subseteq X \times Y$ is open (in the product topology) iff : for all $(x, y) \in A$, there are open sets $U \subseteq X$ and $V \subseteq Y$ such that $(x, y) \in U \times V \subseteq A$. (*Note that A itself might not be a "box."*)

Let $\pi_x : X \times Y \to X$ be the "projection" defined by $\pi_x(x, y) = x$. If U is any open set in X, then $\pi_x^{-1}[U] = U \times Y \in \mathcal{B}$. Therefore $\pi_x^{-1}[U]$ is open. Similarly, for $\pi_y : X \times Y \to Y$ defined by $\pi_y(x, y) = y$: if V is open in Y, then $\pi_y^{-1}[V] = X \times V$ is open in $X \times Y$. (*As we see in Section 8, this means that the projection maps are continuous. It is not hard to show that* *the projection maps π_x and π_y are also open maps: that is, the image of open sets in the product is open*).

If D_1 is dense in X and D_2 is dense in Y, we claim that $D_1 \times D_2$ is dense in $X \times Y$. If $(x, y) \in A$, where A is open, then there are nonempty open sets $U \subseteq X$ and $V \subseteq Y$ for which $(x, y) \in U \times V \subseteq A$. Since $x \in \text{cl}\, D_1$, we know that $U \cap D_1 \neq \emptyset$; and similarly $V \cap D_2 \neq \emptyset$. Therefore $(U \times V) \cap (D_1 \times D_2) = (U \cap D_1) \times (V \cap D_2) \neq \emptyset$, so $A \cap (D_1 \times D_2) \neq \emptyset$. Therefore $(x, y) \in \text{cl}\,(D_1 \times D_2)$. So $D_1 \times D_2$ is dense in $X \times Y$. In particular, this shows that the product of two separable spaces is separable.

Example 5.12 The open intervals (a, b) form a base for the usual topology in $\mathbb{R}$, so each set $(a, b) \times (c, d)$ is in the base $\mathcal{B}$ for the product topology on $\mathbb{R} \times \mathbb{R}$. It is easy to see that every "open box" $U \times V$ in $\mathcal{B}$ can be written as a union of "simple open boxes" like $(a, b) \times (c, d)$. Therefore $\mathcal{B}' = \{(a, b) \times (c, d) : a < b,\, c < d\}$ also is a (simpler) base for the product topology on $\mathbb{R} \times \mathbb{R}$. From this, it is clear that the product topology on $\mathbb{R} \times \mathbb{R}$ is the usual topology on the plane $\mathbb{R}^2$ (*see Example 5.10*).

In general, the open sets U and V in the base for the product topology on $X \times Y$ can be replaced by sets "U chosen from a base for X" and "V chosen from a base for Y," as in this example. So in the definition of the product topology, it is sufficient to say that basic open sets are of the form $U \times V$, where U and V are basic open sets from X and from Y.

Definition 5.13 We say that a space $(X, \mathcal{T})$ satisfies the second axiom of countability (or, more simply, that $(X, \mathcal{T})$ is a second countable space) if it is possible to find a countable base $\mathcal{B}$ for the topology $\mathcal{T}$.

For example, $\mathbb{R}$ is second countable because, for example, $\mathcal{B} = \{(a, b) : a, b \in \mathbb{Q}\}$ is a countable base. Is $\mathbb{R}^2$ is second countable (*why or why not*) ?

Example 5.14 The collection $\mathcal{B} = \{[x, x + \frac{1}{n}) : x \in \mathbb{R},\, n \in \mathbb{N}\}$ is a base for the Sorgenfrey topology on $\mathbb{R}$. But the collection $\{[x, x + \frac{1}{n}) : x \in \mathbb{Q},\, n \in \mathbb{N}\}$ is not a countable base for the Sorgenfrey topology. Why not?

Since the sets in a base may be simpler than arbitrary open sets, they are often more convenient to work with, and working with the basic open sets is often all that is necessary – not a surprise since all the information about the open sets in contained in the base $\mathcal{B}$. For example, you should check that

1) If $\mathcal{B}$ is a base for $\mathcal{T}$, then $x \in \text{cl}\, A$ iff each basic open set B containing x satisfies $B \cap A \neq \emptyset$.

2) If $f : (X, d) \to (Y, d')$ and $\mathcal{B}$ is a base for the topology $\mathcal{T}_{d'}$ on Y, then f is continuous iff $f^{-1}[B]$ is open for each $B \in \mathcal{B}$. This means that we needn't check the inverse images of all open sets to verify that f is continuous.

C. Subbase for the topology

Definition 5.15 Suppose $(X, \mathcal{T})$ is a topological space. A family $\mathfrak{S}$ of open sets is called a subbase for the topology $\mathcal{T}$ if the collection $\mathcal{B}$ of all finite intersections of sets from $\mathfrak{S}$ is a base for $\mathcal{T}$. (*Clearly, if* $\mathcal{B}$ *is a base for* $\mathcal{T}$*, then* $\mathcal{B}$ *is automatically a subbase for* $\mathcal{T}$.)

Examples i) The collection $\mathfrak{S} = \{I : I = (-\infty, b) \text{ or } I = (a, \infty),\ a, b \in \mathbb{R}\}$ is a subbase for a topology on $\mathbb{R}$. All intervals of the form $(a, b) = (-\infty, b) \cap (a, \infty)$ are in $\mathcal{B}$, so $\mathcal{B}$ is a base for the usual topology on $\mathbb{R}$.

ii) The collection $\mathfrak{S}$ of all sets $U \times Y$ and $V \times X$ (for U open in X and V open in Y) is a subbase for the product topology on $X \times Y$: these sets are open in $X \times Y$ and the collection $\mathcal{B}$ of all finite intersections of sets in $\mathfrak{S}$ includes all the open boxes $U \times V = (U \times Y) \cap (X \times V)$.

We can define a topology on a set X by giving a collection $\mathfrak{S}$ of subsets as the subbase for a topology. Surprisingly, any collection $\mathfrak{S}$ can be used: no special conditions on $\mathfrak{S}$ are required.

Theorem 5.16 (The Subbase Theorem) Suppose X is a set and $\mathfrak{S}$ is any collection of subsets of X. Let $\mathcal{B}$ be the collection of all finite intersections of sets from $\mathfrak{S}$. Then $\mathcal{B}$ is a base for a topology $\mathcal{T}$, and $\mathfrak{S}$ is a subbase for $\mathcal{T}$.

Proof First we show that $\mathcal{B}$ satisfies conditions 1) and 2) of The Base Theorem.

1) $X \in \mathcal{B}$ since X is the intersection of the empty subcollection of $\mathfrak{S}$ (*this follows the convention that the intersection of an empty family of subsets of* X *is* X *itself. See Example I.4.5.5*). Since $X \in \mathcal{B}$, certainly $X = \bigcup\{B\colon B \in \mathcal{B}\}$.

2) Suppose B_1 and $B_2 \in \mathcal{B}$, and $x \in B_1 \cap B_2$. We know that $B_1 = S_1 \cap ... \cap S_m$ and $B_2 = S_{m+1} \cap ... \cap S_{m+k}$ for some $S_1, ..., S_m, ..., S_{m+k} \in \mathfrak{S}$, so

$$x \in B_3 = B_1 \cap B_2 = S_1 \cap ... \cap S_m \cap S_{m+1} \cap ... \cap S_{m+k} \in \mathcal{B}$$

Therefore $\mathcal{T} = \{O : O \text{ is a union of sets from } \mathcal{B}\}$ is a topology, and $\mathcal{B}$ is a base for $\mathcal{T}$. By definition of $\mathcal{B}$ and $\mathcal{T}$, we have $\mathfrak{S} \subseteq \mathcal{B} \subseteq \mathcal{T}$ and each set in $\mathcal{T}$ is a union of finite intersections of sets from $\mathfrak{S}$. Therefore $\mathfrak{S}$ is a subbase for $\mathcal{T}$. ●

Example 5.17

1) Let $\mathbb{E} = \{2, 4, 6, ...\} \subseteq \mathbb{N}$ and $\mathfrak{S} = \{\{1\}, \{2\}, \mathbb{E}\}$. $\mathfrak{S}$ is a subbase for a topology on $\mathbb{N}$. A base for this topology is the collection $\mathcal{B}$ of all finite intersections of sets in $\mathfrak{S}$:

$$\mathcal{B} = \{\emptyset, \mathbb{N}, \{2\}, \{1\}, \mathbb{E}\}.$$

The not-very-interesting topology $\mathcal{T}$ generated by the base $\mathcal{B}$ is collection of all possible unions of sets from $\mathcal{B}$:

$$\mathcal{T} = \{\emptyset, \mathbb{N}, \{1\}, \{2\}, \{1,2\}, \mathbb{E}, \mathbb{E} \cup \{1\}\}$$

2) For each $n \in \mathbb{N}$, let $S_n = \{n, n+1, n+2, ...\}$. The collection $\mathfrak{S} = \{S_n : n \in \mathbb{N}\}$ is a subbase for a topology on $\mathbb{N}$. Here, $S_n \cap S_m = S_k$ where $k = \max\{m, n\}$, so the collection of all finite intersections from $\mathfrak{S}$ is just $\mathfrak{S}$ itself. So $\mathcal{B} = \mathfrak{S}$ is actually a base for a topology. The topology is $\mathcal{T} = \mathfrak{S} \cup \{\emptyset\}$.

3) Let $\mathfrak{S} = \{\ell : \ell$ is a straight line in $\mathbb{R}^2\}$ in $\mathbb{R}^2$. For every point $p \in \mathbb{R}^2$, $\{p\}$ is the intersection of two sets from $\mathfrak{S}$, so $\{p\} \in \mathcal{B}$. $\mathfrak{S}$ is a subbase for the discrete topology on $\mathbb{R}^2$.

4) Let $\mathfrak{S} = \{\ell : \ell$ is a vertical line in $\mathbb{R}^2\}$. $\mathfrak{S}$ generates a topology on $\mathbb{R}^2$ for which $\mathcal{B} = \{\mathbb{R}^2, \emptyset\} \cup \mathfrak{S}$ is a base and $\mathcal{T} = \{O : O$ is a union of vertical lines$\}$.

5) Let $\mathfrak{S}$ be a collection of subsets of X and suppose that $\mathcal{T}'$ is <u>any</u> topology on X for which $\mathfrak{S} \subseteq \mathcal{T}'$. Since $\mathcal{T}'$ is a topology, it must contain all finite intersections of sets in $\mathfrak{S}$, and therefore must contain all possible unions of such intersections. Therefore $\mathcal{T}'$ contains the topology $\mathcal{T}$ for which $\mathfrak{S}$ is a subbase. To put it another way, <u>the topology for which $\mathfrak{S}$ is a subbase is in the smallest topology on X containing the collection $\mathfrak{S}$</u>. In fact, as an exercise, you can check that

$$\mathcal{T} = \bigcap\{\mathcal{T}' : \mathcal{T}' \supseteq \mathfrak{S} \text{ and } \mathcal{T}' \text{ is a topology on } X\}.$$

Caution We said earlier that for some purposes it is sufficient (and simpler) to work with <u>basic</u> open sets, rather than <u>arbitrary</u> open sets – for example to check whether $x \in \operatorname{cl} A$, it is sufficient to check whether $U \cap A \neq \emptyset$ for every <u>basic</u> open set U that contains x. However, it is <u>not always</u> sufficient to work with <u>subbasic</u> open sets. Some caution is necessary.

For example, $\mathfrak{S} = \{I : I = (-\infty, b)$ or $I = (a, \infty),\ a, b \in \mathbb{R}\}$ is a subbase for the usual topology on $\mathbb{R}$. We have $I \cap \mathbb{Z} \neq \emptyset$ for every <u>subbasic</u> open set I containing $\frac{1}{2}$, but $\frac{1}{2} \notin \operatorname{cl} \mathbb{Z}$.

D. The closure operator

Usually we describe a topology $\mathcal{T}$ by giving a subbase $\mathfrak{S}$, a base $\mathcal{B}$, or by giving collections $\mathcal{B}_x$ to be the basic neighborhoods at each point x. In the early history of general topology, one other method was sometimes used. We will never actually use it, but we include it here as a curiosity.

Let $\operatorname{cl}_{\mathcal{T}}$ be the closure operator in $(X, \mathcal{T})$ (*normally, we would just write "cl" for the closure operator; here we write "$\operatorname{cl}_{\mathcal{T}}$" to emphasize that this closure operator comes from the topology $\mathcal{T}$ on X*). It gives us all the information about $\mathcal{T}$. That is, using $\operatorname{cl}_{\mathcal{T}}$, we can decide whether any set A is closed (by asking "is $\operatorname{cl}_{\mathcal{T}} A = A$?") and therefore can decide whether any set B is open (by asking "is $X - B$ closed ?"). It should be not be a surprise, then, that we can define a topology on a <u>set</u> X if we are start with an "operator" which "behaves like a closure operator." How is that? Our first theorem tells us the crucial properties of a closure operator.

Theorem 5.18 Suppose $(X, \mathcal{T})$ is a topological space and A, B are subsets of X. Then

1) $\text{cl}_{\mathcal{T}}\emptyset = \emptyset$
2) $A \subseteq \text{cl}_{\mathcal{T}} A$
3) A is closed iff $A = \text{cl}_{\mathcal{T}} A$
4) $\text{cl}_{\mathcal{T}} A = \text{cl}_{\mathcal{T}}(\text{cl}_{\mathcal{T}} A)$
5) $\text{cl}_{\mathcal{T}}(A \cup B) = \text{cl}_{\mathcal{T}} A \cup \text{cl}_{\mathcal{T}} B$

Proof 1) Since $\emptyset$ is closed, $\text{cl}_{\mathcal{T}}\emptyset = \emptyset$

2) $A \subseteq \bigcap\{F : F \text{ is closed and } A \subseteq F\} = \text{cl}_{\mathcal{T}} A$

3) $\Rightarrow$: If A is closed, then A itself is one of the closed sets F used in the definition $\text{cl}_{\mathcal{T}} A = \cap\{F : F \text{ is closed and } F \supseteq A\}$, so $A = \text{cl}_{\mathcal{T}} A$.
$\Leftarrow$: If $A = \text{cl}_{\mathcal{T}} A$, then A is closed because $\text{cl}_{\mathcal{T}} A$ is an intersection of closed sets.

4) $\text{cl}_{\mathcal{T}} A$ is closed, so by 3), $\text{cl}_{\mathcal{T}} A = \text{cl}_{\mathcal{T}}(\text{cl}_{\mathcal{T}} A)$

5) $A \cup B \supseteq A$, so $\text{cl}_{\mathcal{T}}(A \cup B) \supseteq \text{cl}_{\mathcal{T}} A$. Similarly, $\text{cl}_{\mathcal{T}}(A \cup B) \supseteq \text{cl}_{\mathcal{T}} B$.
Therefore $\text{cl}_{\mathcal{T}}(A \cup B) \supseteq \text{cl}_{\mathcal{T}}(A) \cup \text{cl}_{\mathcal{T}}(B)$.
On the other hand, $\text{cl}_{\mathcal{T}} A \cup \text{cl}_{\mathcal{T}} B$ is a closed set that contains $A \cup B$, and therefore $\text{cl}_{\mathcal{T}} A \cup \text{cl}_{\mathcal{T}} B \supseteq \text{cl}_{\mathcal{T}}(A \cup B)$. $\bullet$

The next theorem tells us that we can use an operator "cl" to create a topology on a set.

Theorem 5.19 (The Closure Operator Theorem) Suppose X is a <u>set</u> and that for each $A \subseteq X$, a subset $\text{cl}\, A$ is defined (*that is, we have a function cl* : $\mathcal{P}(X) \to \mathcal{P}(X)$) in such a way that conditions 1), 2), 4) and 5) of Theorem 5.18 are satisfied. <u>Define</u> $\mathcal{T} = \{O : \text{cl}(X - O) = X - O\}$. Then $\mathcal{T}$ is a topology on X, and cl is the closure operator for this topology (*that is, cl* $=$ *cl*$_{\mathcal{T}}$).

Note: i) Such a function cl is called a "Kuratowski closure operator," or just "closure operator" for short.
ii) The Closure Operator Theorem does not ask that cl satisfy condition 3): initially, there is no topology on the given set X, so 3) would be meaningless. But 3) motivates the definition of $\mathcal{T}$ as the collection of sets whose complements are unchanged when "cl" is applied..

Proof (*Numbers in parentheses refer to properties of "cl"*)

First note that:

$$(*) \qquad \text{if } A \subseteq B, \text{ then } \text{cl}\, A \subseteq \text{cl}\, A \cup \text{cl}\, B \overset{(5)}{=} \text{cl}(A \cup B) = \text{cl}\, B$$

$\mathcal{T}$ has the properties required for a topology on the set X :

i) $X \overset{(2)}{\subseteq} \text{cl}\, X$, so $X = \text{cl}\, X$. Therefore $\text{cl}\,(X - \emptyset) = \text{cl}\, X = X = X - \emptyset$, so $\emptyset \in \mathcal{T}$.
Also, $\text{cl}(X - X) = \text{cl}\,\emptyset \overset{(1)}{=} \emptyset = X - X$, so therefore $X \in \mathcal{T}$.

ii) Suppose $O_\alpha \in \mathcal{T}$ for each $\alpha \in A$. For each particular $\alpha_0 \in A$, we have

$\text{cl}(X - \bigcup O_\alpha) \overset{(*)}{\subseteq} \text{cl}(X - O_{\alpha_0}) = X - O_{\alpha_0}$ because $O_{\alpha_0} \in \mathcal{T}$. This is true for <u>every</u> $\alpha_0 \in \Lambda$, so $\text{cl}(X - \bigcup O_\alpha) \subseteq \bigcap(X - O_\alpha) = X - \bigcup O_\alpha$.

But by 2), we know that $X - \bigcup O_\alpha \subseteq \text{cl}\,(X - \bigcup O_\alpha)$.

Therefore $\text{cl}(X - \bigcup O_\alpha) = X - \bigcup O_\alpha$, so $\bigcup O_\alpha \in \mathcal{T}$.

iii) If O_1 and O_2 are in $\mathcal{T}$, then $\text{cl}(X - (O_1 \cap O_2)) = \text{cl}((X - O_1) \cup (X - O_2))$
$\overset{(5)}{=} \text{cl}(X - O_1) \cup \text{cl}(X - O_2) = (X - O_1) \cup (X - O_2)$ (since O_1 and $O_2 \in \mathcal{T}$)
$= X - (O_1 \cap O_2)$. Therefore $O_1 \cap O_2 \in \mathcal{T}$. Therefore $\mathcal{T}$ is a topology on X.

Having this topology $\mathcal{T}$ now gives us an associated closure operator $\text{cl}_\mathcal{T}$ and we want to show that $\text{cl}_\mathcal{T} = \text{cl}$. First, we observe that for <u>closed</u> sets in $(X, \mathcal{T})$:

(**) $\text{cl}_\mathcal{T}\, B = B$ iff B is closed in $(X, \mathcal{T})$
iff $X - B \in \mathcal{T}$ iff $\text{cl}\,(X - (X - B)) = X - (X - B)$ iff $\text{cl}\, B = B$.

To finish, we must show that $\text{cl}\, A = \text{cl}_\mathcal{T} A$ for <u>every</u> $A \subseteq X$.

$A \subseteq \text{cl}_\mathcal{T} A$, so (*) gives that $\text{cl}\, A \subseteq \text{cl}(\text{cl}_\mathcal{T} A)$. But $\text{cl}_\mathcal{T} A$ is closed in $(X, \mathcal{T})$, so using $B = \text{cl}_\mathcal{T} A$ in (**) gives $\text{cl}\,\text{cl}_\mathcal{T} A = \text{cl}_\mathcal{T} A$. Therefore $\text{cl}\, A \subseteq \text{cl}_\mathcal{T} A$.

On the other hand, $\text{cl}\,\text{cl}\, A \overset{(4)}{=} \text{cl}\, A$, so using $B = \text{cl}\, A$ in (**) gives that $\text{cl}\, A$ is closed in $(X, \mathcal{T})$.

But $\text{cl}\, A \overset{(2)}{\supseteq} A$, so $\text{cl}\, A$ is one of the closed sets in the intersection that defines $\text{cl}_\mathcal{T} A$. Therefore $\text{cl}\, A \supseteq \text{cl}_\mathcal{T} A$. Therefore $\text{cl}\, A = \text{cl}_\mathcal{T} A$. •

Example 5.20

1) Let X be a set. For $A \subseteq X$, define $\text{cl}\, A = \begin{cases} A & \text{if } A \text{ is finite} \\ X & \text{if } A \text{ is infinite} \end{cases}$.

Then cl satisfies the conditions in the Closure Operator Theorem. Since $\text{cl}\, A = A$ iff A is finite or $A = X$, the closed sets in the topology generated by cl are precisely X and the finite sets $-$ that is, cl generates the cofinite topology on X.

2) For each subset A of $\mathbb{R}$, define

$$\text{cl}\, A = \{x \in \mathbb{R} : \text{ there is a sequence } (a_n) \text{ in } A \text{ with each } a_n \geq x \text{ and } |a_n - x| \to 0\}.$$

It is easy to check that cl satisfies the hypotheses of the Closure Operator Theorem. Moreover, a set A is open in the corresponding topology iff $\forall x \in A\ \exists b > x$ such that $[x, b) \subseteq A$. Therefore the topology generated by cl is the Sorgenfrey topology on $\mathbb{R}$. What happens in this example if " $\geq$ " is replaced by " $>$ " in the definition of cl ?

Since closures, interiors and Frontiers are all related, it shouldn't be surprising that we can also describe a topology by defining an appropriate "int" operator or "Fr" operator on a set X.

Exercises

E1. Let $X = \{0, 1, 2, ., ., .\} = \{0\} \cup \mathbb{N}$. For $O \subseteq X$, let

$$\phi_O(n) = \text{the number of elements in } O \cap [1, n] = |\, O \cap [1, n]\, |. \text{ Then define}$$
$$\mathcal{T} = \{\, O : 0 \notin O \text{ or } (0 \in O \text{ and } \lim_{n \to \infty} \tfrac{\phi_O(n)}{n} = 1)\, \}.$$

a) Prove that $\mathcal{T}$ is a topology on X.

b) In any topological space: a point x is called a <u>limit point of the set A</u> if $N \cap (A - \{x\}) \neq \emptyset$ for every neighborhood N of x. *Informally, x is a limit point of A means that there are points of A, other than x itself, arbitrarily close to x.* Prove that in any topological space, a set B is closed iff B contains all of its limit points.

c) For $(X, \mathcal{T})$ as defined in a): prove that x is a limit point of X if and only if $x = 0$.

E2. Suppose $(X, \mathcal{T})$ is a topological space and that $A_\alpha \subseteq X$ for each $\alpha \in A$.

a) Prove that if $\bigcup \{\text{cl}\, A_\alpha : \alpha \in A\}$ is closed, then

$$\bigcup \{\text{cl}\, A_\alpha : \alpha \in A\} = \text{cl}\,(\bigcup \{A_\alpha : \alpha \in A\})$$

(*Note that " $\subseteq$ " is true for <u>any</u> collection of sets A_α.*)

b) A family $\{B_\alpha : \alpha \in A\}$ of subsets of X is called <u>locally finite</u> if each point $x \in X$ has a neighborhood N such that $N \cap B_\alpha \neq \emptyset$ for only finitely many α's. Prove that if $\{B_\alpha : \alpha \in A\}$ is locally finite, then

$$\bigcup \{\text{cl}\, B_\alpha : \alpha \in A\} = \text{cl}\,(\bigcup \{B_\alpha : \alpha \in A\})$$

c) Prove that in $(X, \mathcal{T})$, the union of a locally finite family of closed sets is closed.

E3. Suppose $f : (X, d) \to (Y, s)$. Let $\mathcal{B}$ be a <u>base</u> for the topology $\mathcal{T}_s$ and let $\mathfrak{S}$ be a <u>subbase</u> for $\mathcal{T}_s$. Prove or disprove: f is continuous iff $f^{-1}[O]$ is open for all $O \in \mathcal{B}$ iff $f^{-1}[O]$ is open for all $O \in \mathfrak{S}$.

E4. A space $(X, \mathcal{T})$ is called a T_1-<u>space</u> if $\{x\}$ is closed for every $x \in X$.

a) Give an example where $(X, \mathcal{T})$ is not a T_1-space and $\mathcal{T}$ is not the trivial topology.
b) Prove that X is a T_1-space if and only if, given any two distinct points $x, y \in X$, each point is contained in an open set not containing the other point.
c) Prove that in a T_1-space, each set $\{x\}$ can be written as an intersection of open sets.
d) Prove that a subspace of a T_1-space is a T_1-space.
e) Prove that if a pseudometric space (X, d) is a T_1-space, then d must in fact be a metric.
f) Prove that if X and Y are T_1-spaces, so is $X \times Y$.

E5. $(X, \mathcal{T})$ is called a T_2-space (or *Hausdorff space*) if whenever $x, y \in X$ and $x \neq y$, then there exist disjoint open sets U and V with $x \in U$ and $y \in V$.

a) Give an example of a space $(X, \mathcal{T})$ which is a T_1-space but not a T_2-space (*see E4.*) .
b) Prove that a subspace of a Hausdorff space is Hausdorff.
c) Prove that if X and Y are Hausdorff, then so is $X \times Y$.

E6. Prove that every infinite T_2-space contains an infinite discrete subspace $-$ that is, a subset which is discrete in the subspace topology (*see E5*).

E7. Suppose that $(X, \mathcal{T})$ and $(Y, \mathcal{T}')$ are topological spaces. Recall that the product topology on $X \times Y$ is the topology for which a base is the collection of "open boxes"

$$\mathcal{B} = \{U \times V : U \in \mathcal{T},\ V \in \mathcal{T}'\}$$

Therefore a set $O \subseteq X \times Y$ is open in the product topology iff for all $(x, y) \in O$, there exist open sets U in X and V in Y such that $(x, y) \in U \times V \subseteq O$. (*Note that the product topology on $\mathbb{R} \times \mathbb{R}$ is the usual topology on $\mathbb{R}^2$. (We always assume that the topology used on a product of two spaces $X \times Y$ is the product topology unless something different is explicitly stated.*)

a) Verify that $\mathcal{B}$ is, in fact, a base for a topology on $X \times Y$.
b) Consider the projection map $\pi_X : X \times Y \to X$. Prove that if O is any open set in $X \times Y$ (*not necessarily a "box"*), then $\pi_X[O]$ is open in X. (*We say π_X is an open map. The same is true for the projection π_Y.)*
c) Prove that if $A \subseteq X$ and $B \subseteq Y$, then $\mathrm{cl}_{X \times Y}(A \times B) = \mathrm{cl}_X A \times \mathrm{cl}_Y B$. Use this to explain why "the product of two closed sets is closed in $X \times Y$."
d) Show that $X \times Y$ has a countable base iff each of X and Y has a countable base.
e) Show that there is a countable neighborhood base at $(x, y) \in X \times Y$ iff there is a countable neighborhood base $\mathcal{B}_x$ at $x \in X$ and a countable neighborhood base $\mathcal{B}_y$ at $y \in Y$.
f) Suppose (X, d_1) and (Y, d_2) are pseudometric spaces. Define a pseudometric d on the set $X \times Y$ by

$$d((x_1, y_1), (x_2, y_2)) = d_1(x_1, x_2) + d_2(y_1, y_2).$$

Prove that the product topology on $X \times Y$ is the same as the topology $\mathcal{T}_d$.

Note: d is the analogue of the taxicab metric in $\mathbb{R}^2$. There are other equivalent pseudometrics that produce the product topology on $X \times Y$, for example

$$d'((x_1, y_1), (x_2, y_2)) = (d_1(x_1, x_2)^2 + d_2(y_1, y_2)^2)^{1/2}, \text{ and}$$

$$d''((x_1, y_1), (x_2, y_2)) = \max\{d_1(x_1, x_2), d_2(y_1, y_2)\}.$$

E8. Suppose $A \subseteq (X, \mathcal{T})$. The set A is called regular open if $A = \text{int}\,(\text{cl}\,A)$ and A is regular closed if $A = \text{cl}\,(\text{int}\,A)$.

a) Show that for any subset B

$$\begin{aligned} \text{i)}\quad X - \text{cl}\,B &= \text{int}\,(X - B) \\ \text{ii)}\quad X - \text{int}\,B &= \text{cl}\,(X - B) \end{aligned}$$

b) Give an example of a closed subset of $\mathbb{R}$ which is not regular closed.

c) Show that the complement of a regular open set in $(X, \mathcal{T})$ is regular closed and vice-versa.

d) Show that the interior of any closed set in $(X, \mathcal{T})$ is regular open.

e) Show that the intersection of two regular open sets in $(X, \mathcal{T})$ is regular open.

f) Give an example of the union of two regular open sets that is not regular open.

E9. In each part, prove the statement or provide a counterexample:

a) For any x in a topological space $(X, \mathcal{T})$, $\{x\}$ is equal to the intersection of all open sets containing x.

b) In a topological space, a finite set must always be closed.

c) Suppose we have topologies $\mathcal{T}_\alpha$ on X, one for each $\alpha \in A$. Then $\bigcap\{\mathcal{T}_\alpha : \alpha \in A\}$ is also a topology on X.

d) If $\mathcal{T}_1$ and $\mathcal{T}_2$ are topologies on X, then there is a unique smallest topology $\mathcal{T}_3$ on X such that $\mathcal{T}_1 \cup \mathcal{T}_2 \subseteq \mathcal{T}_3$.

e) Suppose, for each $\alpha \in A$, that $\mathcal{T}_\alpha$ is a topology on X. Then there is a unique smallest topology $\mathcal{T}$ on X such that for each α, $\mathcal{T} \supseteq \mathcal{T}_\alpha$.

E10. Assume that natural number, except 1, can be factored into primes; you shouldn't need any other information about prime numbers. For $a \in \mathbb{Z}$ and $d \in \mathbb{N}$, let

$$B_{a,d} = \{..., a - 2d,\ a - d,\ a,\ a + d,\ a + 2d,\ ...\} = \{a + kd : k \in \mathbb{Z}\}$$

and let $\mathcal{B} = \{B_{a,d} : a \in \mathbb{Z},\ d \in \mathbb{N}\}$

a) Prove that $\mathcal{B}$ is a base for a topology $\mathcal{T}$ on $\mathbb{Z}$
b) Show that each set $B_{a,d}$ is closed in $(\mathbb{Z}, \mathcal{T})$
c) What is the set $\bigcup\{B_{0,p} :\ p \text{ is a prime number}\}$?
d) Part c) tells you what famous fact about the set of prime numbers?

6. Countability Properties of Spaces

Countable sets are often easier to work with than uncountable sets, so it is not surprising that spaces with certain "countability properties" are viewed as desirable. Most of these properties have already been defined, but the definitions are collected together here for convenience.

Definition 6.1 $(X, \mathcal{T})$ is called

first countable if we can choose a countable neighborhood base $\mathcal{B}_x$ at every point $x \in X$. (*We also say that X satisfies the first axiom of countability.*)

second countable if there is a countable base $\mathcal{B}$ for the topology $\mathcal{T}$. (*We also say that X satisfies the second axiom of countability.*)

separable if there is a countable dense set D in X

Lindelöf if whenever $\mathcal{U}$ is a collection of open sets for which $\bigcup \mathcal{U} = X$, then there is a countable subcollection $\mathcal{U}' = \{U_1, U_2, ..., U_n, ...\} \subseteq \mathcal{U}$ for which $\bigcup \mathcal{U}' = X$.

$\mathcal{U}$ is called an open cover of X, and $\mathcal{U}'$ is called a subcover from $\mathcal{U}$. Thus, X is Lindelöf if "*every open cover has a countable subcover*.")

Example 6.2

1) A countable discrete space $(X, \mathcal{T})$ is second countable because $\mathcal{B} = \{\{x\} : x \in X\}$ is a countable base. Is X first countable? separable? Lindelöf ?

2) $\mathbb{R}$ is second countable because $\mathcal{B} = \{(a, b) : a, b \in \mathbb{Q}\}$ is a countable base. Similarly, $\mathbb{R}^n$ is second countable since the collection $\mathcal{B}$ of boxes $(x_1, y_1) \times (x_2, y_2) \times ... \times (x_n, y_n)$ with rational endpoints is a countable base (*check!*)

3) Let X be a countable set with the cofinite topology $\mathcal{T}$. X has only countably many finite subsets (*see Theorem I.11.1*) so there are only countably many sets in $\mathcal{T}$. $(X, \mathcal{T})$ is second countable because we could choose $\mathcal{B} = \mathcal{T}$ as a countable base.

The following theorem implies that each of the spaces in Example 6.2 is also first countable, separable, and Lindelöf. (*However, it is really worthwhile to try to verify these properties, in each example, directly from the definitions.*)

Theorem 6.3 A second countable topological space $(X, \mathcal{T})$ is also separable, first countable, and Lindelöf.

Proof Let $\mathcal{B} = \{O_1, O_2, ...O_n, ...\}$ be a countable base for $\mathcal{T}$.

i) For each n, pick a point $x_n \in O_n$ and let $D = \{x_n : n \in \mathbb{N}\}$. The countable set D is dense. To see this, notice that if U is any nonempty open set in X, then for some n, $x_n \in O_n \subseteq U$ so $U \cap D \neq \emptyset$. Therefore X is separable.

ii) For each $x \in X$, let $\mathcal{B}_x = \{O \in \mathcal{B} : x \in O\}$. Clearly $\mathcal{B}_x$ is a neighborhood base at x, so X is first countable.

iii) Let $\mathcal{U}$ be any open cover of X. If $x \in X$, then $x \in$ some set $U_x \in \mathcal{U}$. For each x, we can then pick a basic open set $O_x \in \mathcal{B}$ such that $x \in O_x \subseteq U_x$. Let $\mathcal{V} = \{O_x : x \in X\}$. Since each $O_x \in \mathcal{B}$, there can be only countably many different sets O_x: that is, $\mathcal{V}$ may contain "repeats." Eliminate any "repeats" and list only the different sets in $\mathcal{V}$, so $\mathcal{V} = \{O_{x_1}, O_{x_2}, ...O_{x_n}, ...\}$ where $O_{x_n} \subseteq U_{x_n} \in \mathcal{U}$. Every x is in one of the sets O_{x_n}, so $\mathcal{U}' = \{U_{x_1}, U_{x_2}, ...U_{x_n}, ...\}$ is a countable subcover from $\mathcal{U}$. Therefore X is Lindelöf. •

The following examples show that no other implications exist among the countability properties in Theorem 6.3.

Example 6.4

1) Suppose X is uncountable and let $\mathcal{T}$ be the cofinite topology on X.

$(X, \mathcal{T})$ is separable since any infinite set is dense.

$(X, \mathcal{T})$ is Lindelöf. To see this, let $\mathcal{U}$ be an open cover of X. Pick any one nonempty set $U \in \mathcal{U}$. Then $X - U$ is finite, say $X - U = \{x_1, ..., x_n\}$. For each x_i, pick a set $U_i \in \mathcal{U}$ with $x_i \in U_i$. Then $\mathcal{U}' = \{U\} \cup \{U_i : i = 1, ..., n\}$ is a countable (actually, finite) subcover chosen from $\mathcal{U}$.

However X is not first countable (Example 4.5.3), and therefore, by Theorem 6.3, X is also not second countable.

2) Suppose X is uncountable. Define $\mathcal{T} = \{O \subseteq X : O = \emptyset \text{ or } X - O \text{ is countable}\}$.

$\mathcal{T}$ is a topology on X (*check!*) called the cocountable topology. A set $C \subseteq X$ is closed iff $C = X$ or C is countable. (*This is an "upscale" analogue of the cofinite topology.*)

An argument very similar to the one in the preceding example shows that $(X, \mathcal{T})$ is Lindelöf. But $(X, \mathcal{T})$ is not separable $-$ every countable subset is closed and therefore not dense. By Theorem 6.3, $(X, \mathcal{T})$ also cannot be second countable.

3) Suppose X is uncountable set and choose a particular point $p \in X$. Define $\mathcal{T} = \{O \subseteq X : O = \emptyset \text{ or } p \in O\}$. (*Check that $\mathcal{T}$ is a topology.*)

$(X, \mathcal{T})$ is separable because $\{p\}$ is dense.

$(X, \mathcal{T})$ is not Lindelöf $-$ because the cover $\mathcal{U} = \{\{x, p\} : x \in X\}$ has no countable subcover.

Is $(X, \mathcal{T})$ first countable?

4) Suppose X is uncountable and let $\mathcal{T}$ be the discrete topology on X. Then $(X, \mathcal{T})$ is

first countable	because $\mathcal{B}_x = \{\{x\}\}$ is a neighborhood base at x
not second countable	because each open set $\{x\}$ would have to be in a base $\mathcal{B}$
not separable	because cl $D = D \neq X$ for any countable set D
not Lindelöf	because the cover $\mathcal{U} = \{\{x\} : x \in X\}$ has no countable subcover. (*In fact, not a single set can be omitted from $\mathcal{U}$: $\mathcal{U}$ has no proper subcover.*)

For "special" topological spaces – pseudometrizable ones, for example – it turns out that things are better behaved. For example, we noted earlier that every pseudometric space (X, d) is first countable (*Example 4.3*). The following theorem shows that in (X, d) the other three countability properties are equivalent to each other: that is, either all of them are true in (X, d) or none are true.

Theorem 6.5 Any pseudometric space (X, d) is first countable. (X, d) is second countable iff (X, d) is separable iff (X, d) is Lindelöf.

Proof i) Second countable $\Rightarrow$ Lindelöf: by Theorem 6.3, this implication is true in any topological space.

ii) Lindelöf $\Rightarrow$ separable: suppose (X, d) is Lindelöf. For each $n \in \mathbb{N}$, let $\mathcal{U}_n = \{B_{\frac{1}{n}}(x) : n \in \mathbb{N},\, x \in X\}$. For each n, $\mathcal{U}_n$ is an open cover so $\mathcal{U}_n$ has a countable subcover – that is, for each n we can find countably many $\frac{1}{n}$-balls that cover X: say $X = \bigcup_{k=1}^{\infty} B_{\frac{1}{n}}(x_{n,k})$. Let D be the set of centers of all these balls: $D = \{x_{n,k} : n, k \in \mathbb{N}\}$. For any $x \in X$ and every n, we have $x \in B_{\frac{1}{n}}(x_{n,k})$ for some k, so $d(x, x_{n,k}) < \frac{1}{n}$. Therefore, for every n, $B_{\frac{1}{n}}(x) \cap D \neq \emptyset$ – in other words, x can be approximated arbitrarily closely by points from D. Therefore D is dense, so (X, d) is separable.

iii) Separable $\Rightarrow$ second countable: suppose (X, d) is separable and that $D = \{x_1, x_2, ..., x_k, ...\}$ is a countable dense set. Let $\mathcal{B} = \{B_{\frac{1}{n}}(x_k) : n, k \in \mathbb{N}\}$. $\mathcal{B}$ is a countable collection of open balls and we claim $\mathcal{B}$ is a base for the topology $\mathcal{T}_d$.

> Suppose $y \in O \in T_d$. By the definition of "open," there is an $\epsilon > 0$ for which $B_\epsilon(y) \subseteq O$. Pick n so that $\frac{1}{n} < \frac{\epsilon}{2}$, and pick $x_k \in D$ so that $d(x_k, y) < \frac{1}{n}$. Then $y \in B_{\frac{1}{n}}(x_k) \subseteq B_\epsilon(y) \subseteq O$ (because $z \in B_{\frac{1}{n}}(x_k)$ $\Rightarrow d(z, y) \leq d(z, x_k) + d(x_k, y) < \frac{1}{n} + \frac{1}{n} < 2(\frac{\epsilon}{2}) = \epsilon$). •

It's customary to call a metric space that has these three equivalent properties a "separable metric space" rather than a "second countable metric space" or "Lindelöf metric space."

Theorem 6.5 implies that the spaces in parts 1), 2), 3) of Example 6.4 are not pseudometrizable. In general, to show a space $(X, \mathcal{T})$ is not pseudometrizable we can i) show that it fails to have some property shared by all pseudometric spaces (for example, first countability), or ii) show that it has one but not all of the properties "second countable," "Lindelöf," or "separable."

Exercises

E11. Define $\mathcal{T} = \{U \cup V : U$ is open in the usual topology on $\mathbb{R}$ and $V \subseteq \mathbb{P}\}$.

a) Show that $\mathcal{T}$ is a topology on $\mathbb{R}$. If x is irrational, describe an "efficient" neighborhood base at x. Do the same if x is rational.

b) Is $(\mathbb{R}, \mathcal{T})$ first countable? second countable? Lindelöf ? separable?

The space $(\mathbb{R}, \mathcal{T})$ is called the "scattered line." We could change the definition of $\mathcal{T}$ by replacing $\mathbb{P}$ with some other set $A \subseteq \mathbb{R}$. This creates a space in which the set A is "scattered." Hint: See Example I.7.9.6. It is possible to find open intervals I_n such that $\bigcup_{n=1}^{\infty} I_n \supseteq \mathbb{Q}$ and for which $\sum_{n=1}^{\infty} \text{length}(I_n) < 1$.

E12. A point $x \in (X, \mathcal{T})$ is called a <u>condensation point</u> if every neighborhood of x is uncountable.

a) Let C be the set of all condensation points in X. Prove that C is closed.

b) Prove that if X is second countable, then $X - C$ is countable.

E13. Suppose $(X, \mathcal{T})$ is a second countable space and let $\mathcal{B}$ be a countable base for the topology. Suppose $\mathcal{B}'$ is another base (<u>not</u> <u>necessarily countable</u>) for $\mathcal{T}$ containing open sets all of which have some property P. (For example, "P" could be "clopen" or "separable.") Show that there is a <u>countable</u> base $\mathcal{B}''$ consisting of open sets with property P.
Hint: think about the Lindelöf property.

E14. A space $(X, \mathcal{T})$ is called *hereditarily Lindelöf* if every subspace of X is Lindelöf.

a) Prove that a second countable space is hereditarily Lindelöf.

In any space, a point x is called a <u>limit</u> <u>point</u> <u>of</u> <u>the</u> <u>set</u> <u>A</u> if $N \cap (A - \{x\}) \neq \emptyset$ for every neighborhood N of x. Informally, x is a limit point of A if there are points in A different from x and arbitrarily close to x.)

b) Suppose X is hereditarily Lindelöf. Prove that the set $A = \{x \in X : x$ is not a limit point of $X\}$ is countable.

E15. A space $(X, \mathcal{T})$ is said to satisfy the *countable chain condition* ($= CCC$) if every family of disjoint open sets must be countable.

a) Prove that a separable space $(X, \mathcal{T})$ satisfies the CCC.

b) Give an example of a space that satisfies the CCC but that is not separable. (*It is not necessary to do so, but can you find an example which is a metric space?)*

E16. Suppose $(X, \mathcal{T})$ is a topological space and that $\mathcal{P}$ and $\mathcal{B}$ are two bases for the topology $\mathcal{T}$, and that $\mathcal{P}$ and $\mathcal{B}$ are infinite.

a) Prove that there is a subfamily $\mathcal{B}' \subseteq \mathcal{B}$ such that $\mathcal{B}'$ is also a base and $|\mathcal{B}'| \leq |\mathcal{P}|$.
(*Hint: For each pair* $t = (U, V) \in \mathcal{P} \times \mathcal{P}$, *pick, if possible, a set* $W_t \in \mathcal{B}$ *such that* $U \subseteq W \subseteq V$; *otherwise set* $W_t = \emptyset$.)

b) Use part a) to prove that the Sorgenfrey line is not second countable.
(*Hint: Show that otherwise there would be a countable base of sets of the form* $[a, b)$, *but that this is impossible.*)

7. More About Subspaces

Suppose $(X, \mathcal{T})$ is a topological space. In Definition 3.1, we defined the subspace topology $\mathcal{T}_A$ on $A \subseteq X$: $\quad \mathcal{T}_A = \{A \cap O : O \in \mathcal{T}\}$. In this section we explore some simple but important properties of subspaces.

If $A \subseteq B \subseteq X$, there are two ways to put a topology on A :

1) we can give A the subspace topology $\mathcal{T}_A$ from X, or

2) we can give B the subspace topology $\mathcal{T}_B$, and then give A the subspace topology from the space $(B, \mathcal{T}_B)$ – that is, we can give A the topology $(\mathcal{T}_B)_A$.

In other words, we can think of A as a subspace of X or as a subspace of the subspace B. Fortunately, the next theorem says that these two topologies are the same. More informally, Theorem 7.1 says that in a space X "a subspace of a subspace is a subspace."

Theorem 7.1 If $A \subseteq B \subseteq X$, and $\mathcal{T}$ is a topology on X, then $\mathcal{T}_A = (\mathcal{T}_B)_A$.

Proof $U \in \mathcal{T}_A$ iff $U = O \cap A$ for some $O \in \mathcal{T}$ iff $U = O \cap (B \cap A)$ iff $U = (O \cap B) \cap A$. But $O \cap B \in \mathcal{T}_B$, so the last equation holds iff $U \in (\mathcal{T}_B)_A$. •

We always assume a subset A has the subspace topology (unless something else is explicitly stated). The notation $A \subseteq (X, \mathcal{T})$ emphasizes that A is considered a subspace, not merely a subset.

By definition, a set is open in the subspace topology on A iff it is the intersection with A of an open set in X. The same is also true for closed sets.

Theorem 7.2 Suppose $F \subseteq A \subseteq (X, \mathcal{T})$. F is closed in A iff $F = A \cap C$ where C is closed in X.

Proof F is closed in A iff $A - F$ is open in A iff $A - F = O \cap A$ (for some open set O in X) iff $F = A - (O \cap A) = (X - O) \cap A = C \cap A$ (where $C = X - O$ is a closed set in X). •

Theorem 7.3 Suppose $A \subseteq (X, \mathcal{T})$.

1) Let $a \in A$. If $\mathcal{B}_a$ is a neighborhood base at a in X, then $\{B \cap A : B \in \mathcal{B}_a\}$ is a neighborhood base at a in A.

2) If $\mathcal{B}$ is a base for $\mathcal{T}$, then $\{B \cap A : B \in \mathcal{B}\}$ is a base for $\mathcal{T}_A$.

With a slight abuse of notation, we can informally write these collections as $\mathcal{B}_a \cap A$ *and.*$\mathcal{B} \cap A$. *Why is this an "abuse?" What do* $\mathcal{B}_a \cap A$ and $\mathcal{B} \cap A$ *mean if taken literally*?

Proof 1) Suppose $a \in A$ and that N is a neighborhood of a in A. Then $a \in \text{int}_A N \subseteq N$, so there is an open set O in X such that $a \in \text{int}_A N = O \cap A \subseteq O$.

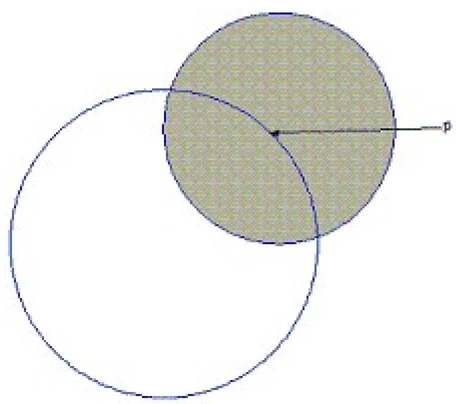
Since $\mathcal{B}_a$ is a neighborhood base at a in X, there is a neighborhood $B \in \mathcal{B}_a$ such that $a \in B \subseteq O$. Then $a \in (\text{int}_X B) \cap A \subseteq B \cap A \in \mathcal{B}_a \cap A$. Since $(\text{int}_X B) \cap A$ is open in A, we see that $B \cap A$ is a neighborhood of a in A. And since $a \in B \cap A \subseteq O \cap A = \text{int}_A N \subseteq N$, we see that $\mathcal{B}_a \cap A$ is a neighborhood base at a in A. •

2) Exercise

Theorem 7.3 tells us that in the subspace A we can get a neighborhood base at a point a by choosing a neighborhood base at a in X are then restricting all its sets to A; and that the same applies to a base for the subspace topology.

Corollary 7.4 Every subspace of a first countable (or second countable) space $(X, \mathcal{T})$ is first countable (or second countable).

Example 7.5 Suppose S^1 is a circle in $\mathbb{R}^2$ and that $p \in S^1 \subseteq \mathbb{R}^2$. $\mathcal{B}_p = \{B_\epsilon(p) : \epsilon > 0\}$ is a neighborhood base at p in $\mathbb{R}^2$, and therefore $\mathcal{B}_p \cap S^1$ is a neighborhood base at p in the subspace S^1. The sets in $\mathcal{B}_p \cap S^1$ are "open arcs on S^1 containing p." (See the figure,)



The following theorem relates closure in a subspace to closure in the larger space. It turns out to be a very useful technical observation.

Theorem 7.6 Suppose $A \subseteq B \subseteq (X, \mathcal{T})$, then $\text{cl}_B A = B \cap \text{cl}_X A$.

Proof $B \cap \text{cl}_X A$ is a closed set in B that contains A, so $B \cap \text{cl}_X A \supseteq \text{cl}_B A$.

On the other hand, suppose $b \in B \cap \text{cl}_X A$. To show that $b \in \text{cl}_B A$, pick an open set U in B that contains b. We need to show $U \cap A \neq \emptyset$. There is an open set O in X such that $O \cap B = U$. Since $b \in \text{cl}_X A$, we have that $\emptyset \neq O \cap A = O \cap (B \cap A) = (O \cap B) \cap A$
$= U \cap A$. •

Example 7.7 1) $\mathbb{Q} = \text{cl}_{\mathbb{Q}}\mathbb{Q} = \text{cl}_{\mathbb{R}}\mathbb{Q} \cap \mathbb{Q}$

2) $\text{cl}_{(0,2]}\,(0,1) = \text{cl}_{\mathbb{R}}(0,1) \cap (0,2] = [0,1] \cap (0,2] = (0,1]$

3) <u>The analogous calculations are not necessarily not true for interiors and boundaries</u>. For example:

$$\mathbb{Q} = \text{int}_{\mathbb{Q}}\mathbb{Q} \neq \text{int}_{\mathbb{R}}\mathbb{Q} \cap \mathbb{Q} = \emptyset \text{ , and}$$

$$\emptyset = \text{Fr}_{\mathbb{Q}}\,\mathbb{Q} \neq \text{Fr}_{\mathbb{R}}\mathbb{Q} \cap \mathbb{Q} = \mathbb{Q}.$$

Why does "cl" have a privileged position here? Is there a "reason" why you would expect a better connection between closures in A and closures in X than you would expect between interiors in A and interiors in X?

Definition 7.8 A property P of topological spaces is called <u>hereditary</u> if whenever a space X has property P, then every subspace A also has property P.

For example, Corollary 7.4 tells us that first and second countability are hereditary properties. Other hereditary properties include "finite cardinality" and "pseudometrizability." On the other hand, "infinite cardinality" is not a hereditary property.

Example 7.9

1) Separability is not a hereditary property. For example, consider the Sorgenfrey plane X (*see Example 5.4*). X is separable because $\mathbb{Q}^2$ is dense.

Consider the subspace $D = \{(x,y) : x + y = 1\}$. The set $U = [a, a+1) \times [b, b+1)$ is open in X so if $(a,b) \in D$, then $U \cap D = \{(a,b)\}$ is open in the subspace D. Therefore D is a discrete subspace of X, and an uncountable discrete space is not separable.

Similarly, the Moore place Γ is separable (*see Example 5.6*); the x-axis in Γ is an uncountable discrete subspace which is not separable.

2) The Lindelöf property is not heredity. Let A be an uncountable set and let $X = A \cup \{p\}$, where p is any additional point not in A. Put a topology on X by giving a neighborhood base at each point.

$$\begin{cases} \mathcal{B}_a = \{\{a\}\} \text{ for } a \in A \\ \mathcal{B}_p = \{B : \; p \in B \text{ and } X - B \text{ is finite}\} \end{cases}$$

(*Check that the collections $\mathcal{B}_x$ satisfy the hypotheses of the Neighborhood Base Theorem 5.2.*)

If $\mathcal{V}$ is an open cover of X, then $p \in V$ for some $V \in \mathcal{V}$. By ii), every neighborhood of p in X has a finite complement, so $X - V$ is finite. For each y in the finite set $X - V$, we can choose a set $V_y \in \mathcal{V}$ with $y \in V_y$. Then $\mathcal{V}' = \{V\} \cup \{V_y : y \in X - V\}$ is a countable (in fact, finite) subcover from $\mathcal{V}$, so X is Lindelöf.

The definition of $\mathcal{B}_a$ implies that each point of A is isolated in A; that is, A is an uncountable discrete subspace. Then $\mathcal{U} = \{\{a\} : a \in A\}$ is an open cover of A that has no countable subcover. Therefore the subspace A is not Lindelöf.

Even when a property is not hereditary, it is sometimes "inherited" by certain subspaces – perhaps, for example, by closed subspaces, or by open subspaces. The next theorem illustrates this.

Theorem 7.10 A closed subspace of a Lindelöf space is Lindelöf (*so we say that the Lindelöf property is "closed hereditary"*).

Proof Suppose K is a closed subspace of the Lindelöf space X. Let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ be a cover of K by sets U_α that are open in K. For each α, there is an open set V_α in X such that $V_\alpha \cap K = U_\alpha$.

Since K is closed, $X - K$ is open and $\mathcal{V} = \{\{X - K\}\} \cup \{V_\alpha : \alpha \in A\}$ is an open cover of X. But X is Lindelöf, so $\mathcal{V}$ has a countable subcover from $\mathcal{V}$, say $\mathcal{V}' = \{X - K, V_1, ..., V_n, ...\}$. (*The set $X - K$ might not be needed in $\mathcal{V}'$, but it can't hurt to include it.*) Clearly then, the collection $\{U_1, ..., U_n, ...\}$ is a countable subcover of K from $\mathcal{U}$. ●

Note:

1. A little reflection on the proof shows that to prove K is Lindelöf, it would be equivalent to show that every cover of K by sets open in X has a countable subcover.)

2. A space X with the property that every open cover has a finite subcover is called compact. See, for example, the space in Example 6.4.1.

An obvious "tweak" to the proof for Theorem 7.10 shows that a closed subspace of a compact space is compact. We will look at the properties of compact spaces in much more detail in Chapter 4 and beyond.

8. Continuity

We first defined continuous functions between pseudometric spaces by using the distance function d to mimic the definition of continuity given in calculus. But then we saw that our definition could be restated in other equivalent ways in terms of open sets, closed sets, or neighborhoods. For topological spaces, we do not have available any distance functions to use to define continuity. But we can still make a definition using neighborhoods (or open set, or closed sets) since the neighborhoods of a describe "nearness" to a, and, of course, the definition parallels the way neighborhoods describe continuity in pseudometric spaces.

Definition 8.1 A function $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is <u>continuous at a</u> $\in X$ if whenever N is a neighborhood of $f(a)$, then $f^{-1}[N]$ is a neighborhood of a. We say f is <u>continuous</u> if f is continuous at each point of X.

The statement that f is continuous at a is clearly equivalent to each of the following statements: :

i) for each neighborhood N of $f(a)$ there is a neighborhood W of a such that $f[W] \subseteq N$
ii) for each open set V containing $f(a)$ there is an open set U containing a such that $f[U] \subseteq V$
iii) for each basic open set V containing $f(a)$ there is a basic open set U containing a such that $f[U] \subseteq V$.

In the following theorem, the conditions i) - iii) for continuity are the same as those in Theorem II.5.6 for pseudometric spaces. Condition iv) was not mentioned in Chapter II, but it is sometimes handy.

Theorem 8.2 Suppose $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$. The following are equivalent.

i) f is continuous
ii) if $O \in \mathcal{T}'$, then $f^{-1}[O] \in \mathcal{T}$ (*the inverse image of an open set is open*)
iii) if F is closed in Y, then $f^{-1}[F]$ is closed in X (*the inverse image of a closed set is closed*)
iv) for every $A \subseteq X$: $f[\mathrm{cl}_X(A)] \subseteq \mathrm{cl}_Y(f[A])$.

Proof The proof that i) - iii) are equivalent is identical to the proof for Theorem II.5.6 for pseudometric spaces. That proof was deliberately worded in terms of open sets, closed sets, and neighborhoods so that it would carry over to this new situation.

iii) $\Rightarrow$ iv) $A \subseteq f^{-1}[f[A]] \subseteq f^{-1}[\mathrm{cl}_Y(f[A])]$. Since $\mathrm{cl}_Y(f[A])$ is closed in Y, iii) tells us that $f^{-1}[\mathrm{cl}_Y(f[A])]$ is a closed set in X that contains A. Therefore $\mathrm{cl}_X(A) \subseteq f^{-1}[\mathrm{cl}_Y(f[A])]$, so $f[\mathrm{cl}_X(A)] \subseteq [\mathrm{cl}_Y(f[A])]$.

iv) $\Rightarrow$ i) Suppose $a \in X$ and that N is a neighborhood of $f(a)$. Let $K = X - f^{-1}[N]$ and $U = X - \mathrm{cl}_X K \subseteq f^{-1}[N]$. U is open, and we claim that $a \in U -$ which will show that $f^{-1}[N]$ is a neighborhood of a, completing the proof. So we need to show that $a \notin \mathrm{cl}_X K$. But this is clear because: if $a \in \mathrm{cl}_X K$, then using iv) would give us that $f(a) \in f[\mathrm{cl}_X K] \subseteq \mathrm{cl}_Y f[K]$, which is impossible since $N \cap f[K] = \emptyset$. •

Example 8.3 Sometimes we want to know whether a certain property is "preserved by continuous functions" – that is, if X has property P and $f : X \to Y$ is continuous and onto, must the image Y also have the property P?

For example, condition iv) in Theorem 8.2 implies that continuous maps preserve separability. Suppose D is a countable dense set in X. Then $f[D]$ is countable and $f[D]$ is dense in Y because $Y = f[X] = f[\mathrm{cl}\, D] \subseteq \mathrm{cl}\, f[D]$.

By contrast, continuous maps do not preserve first countability: for example, let $(Y, \mathcal{T})$ be any topological space. Let $\mathcal{T}'$ be the discrete topology on Y. $(Y, \mathcal{T}')$ is first countable and the identity map $i : (Y, \mathcal{T}') \to (Y, \mathcal{T})$ is continuous and onto. Thus, every space $(Y, \mathcal{T})$ is the continuous image of a first countable space.

Do continuous maps preserve other properties that we have studied – such as Lindelöf, second countable, or metrizable?

The following theorem makes a few simple and useful observations about continuity.

Theorem 8.4 Suppose $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$.

1) Let $B = \mathrm{ran}(f) \subseteq Y$. Then f is continuous iff $f : (X, \mathcal{T}) \to (B, \mathcal{T}'_B)$ is continuous. *In other words, $B =$ the range of f (a subspace of the codomain Y) is what matters for the continuity of f; points of Y not in B (if any) are irrelevant. For example, the function* $\sin : \mathbb{R} \to \mathbb{R}$ *is continuous iff the function* $\sin : \mathbb{R} \to [-1, 1]$ *is continuous.*

2) Let $A \subseteq X$. If f is continuous, then $f|A = g : A \to Y$ is continuous. *That is, the restriction of a continuous function to a subspace is continuous. For example, sin* $: \mathbb{R} \to \mathbb{R}$ *is continuous, so sin* $: \mathbb{Q} \to [-1, 1]$ *is continuous. (The second function is really sin$|\mathbb{Q}$, but it's abbreviated here to just "sin".*

3) If $\mathcal{S}$ is a subbase for $\mathcal{T}'$ (in particular, if $\mathcal{S}$ is a base), then f is continuous iff $f^{-1}[U]$ is open whenever $U \in \mathcal{S}$. In other words, t*o check continuity, it is sufficient to show that the inverse image of every subbasic open set is open.*

Proof 1) Exercise: the crucial observation is that if $O \subseteq Y$, then $f^{-1}[O] = f^{-1}[B \cap O]$.

2) If O is open in Y, then $g^{-1}[O] = f^{-1}[O] \cap A$ – which is an open set in A.

3) Exercise: the proof depends only on the definition of a subbase and set theory:

$$f^{-1}[\bigcup U_\alpha : \alpha \in A] = \bigcup\{f^{-1}[U_\alpha] : \alpha \in A\} \text{ and}$$
$$f^{-1}[\bigcap U_\alpha : \alpha \in A] = \bigcap\{f^{-1}[U_\alpha] : \alpha \in A\} \quad \bullet$$

Example 8.5

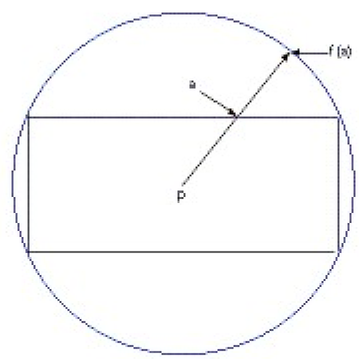
1) For any topological spaces X and Y, every constant function $f : X \to Y$ must be continuous. (*Suppose $f(x) = y_0$ for all x. If O is open in Y, then $f^{-1}[O] =$?*)

If X has the discrete topology and Y is any topological space, then every function $f : X \to Y$ is continuous.

2) Suppose X has the trivial topology and that $f : X \to \mathbb{R}$. If f is not constant, then there are points $a, b \in X$ for which $f(a) \neq f(b)$. Let I be an open set in $\mathbb{R}$ containing $f(a)$ but not $f(b)$. Then $f^{-1}[I]$ is not open in X so f is not continuous. We conclude that f is continuous iff f is constant.

In this example, we could replace $\mathbb{R}$ by any metric space (Y, d); or, for that matter, by any topological space $(Y, \mathcal{T})$ that has what property?

3) Let X be a rectangle inscribed inside a circle Y centered at P. For $a \in X$, let $f(a)$ be the point where the ray from P through a intersects Y. (*The function f is called a "central projection."*). Then both $f : X \to Y$ and $f^{-1} : Y \to X$ are continuous bijections.



Example 8.6 (Weak topologies) Suppose X is a set. Let $\mathcal{F} = \{f_\alpha : \alpha \in A\}$ be a collection of functions where each $f_\alpha : X \to \mathbb{R}$. If we put the discrete topology on X, then all of the functions f_α will be continuous. But a topology on X smaller than the discrete topology might also make all the f_α's continuous. The smallest topology on X that makes all the given f_α's continuous is called the weak topology $\mathcal{T}$ on X generated by the collection $\mathcal{F}$.

How can we describe that topology more directly? $\mathcal{T}$ makes all the f_α's continuous iff for each open $O \subseteq \mathbb{R}$ and each $\alpha \in A$, the set $f_\alpha^{-1}[O]$ is in $\mathcal{T}$. Therefore the weak topology generated by $\mathcal{F}$ is the smallest topology that contains all these sets. According to Example 5.17.5, this means that weak topology $\mathcal{T}$ is the one for which the collection $\mathfrak{S} = \{f_\alpha^{-1}[O] : O \text{ open in } \mathbb{R}, \alpha \in A\}$ is a subbase. (*It is clearly sufficient here to use only basic open sets O from $\mathbb{R}$ – that is, open intervals (a, b) : why ? Would using all open sets O put any additional sets into $\mathcal{T}$?*)

For example, suppose $X = \mathbb{R}^2$ and that $\mathcal{F} = \{\pi_x, \pi_y\}$ contains the two projection maps $\pi_x(x, y) = x$ and $\pi_y(x, y) = y$. For an open interval $U = (a, b) \subseteq \mathbb{R}$, $\pi_x^{-1}[U]$ is the "open vertical strip" $U \times \mathbb{R}$; and $\pi_y^{-1}[V]$ is the "open horizontal strip" $\mathbb{R} \times V$. Therefore a subbase for the weak topology on $\mathbb{R}^2$ generated by $\mathcal{F}$ consists of all such open horizontal or vertical strips. Two such strips intersect in an "open box" $(a, b) \times (c, d)$ in $\mathbb{R}^2$, so it is easy to see that the weak topology is the product topology on $\mathbb{R} \times \mathbb{R}$, that is, the usual topology of $\mathbb{R}^2$.

Suppose $A \subseteq \mathbb{R}$ and that $i : A \to \mathbb{R}$ is the identity function $i(x) = x$. What is the weak topology on the domain A generated by the collection $\mathcal{F} = \{i\}$?

Definition 8.7 A mapping $f: (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is called

open if whenever O is open in X, then $f[O]$ is open in Y, and
closed if whenever F is closed in X, then $f[F]$ is closed in Y.

Suppose $|X| > 1$. Let $\mathcal{T}$ be the discrete topology on X and let $\mathcal{T}'$ be the trivial topology on X. The identity map $i : (X, \mathcal{T}) \to (X, \mathcal{T}')$ is continuous but neither open nor closed, and $i : (X, \mathcal{T}') \to (X, \mathcal{T})$ is both open and closed but not continuous. Open and closed maps are quite different from continuous maps – even when the mapping is a bijection! Here are some examples that are more interesting.

Example 8.8

1) $f : [0, 2\pi) \to S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ given by $f(\theta) = (\cos\theta, \sin\theta)$. It is easy to check that f is continuous, one-to-one, and onto. The set $F = [\pi, 2\pi)$ is closed in $[0, 2\pi)$ but $f[\,[\pi, 2\pi)]$ is not closed in S^1. Also, $[0, \pi)$ is open in $[0, 2\pi)$ but $f[\,[0, \pi)\,]$ is not open in S^1. A continuous, one-to-one, onto mapping does not need to be open or closed!

2) Suppose X and Y are topological spaces and that $f : X \to Y$ is a bijection. Then there is an inverse function $g = f^{-1} : Y \to X$, and f^{-1} is continuous iff f is open. To check this, consider an open set) in X. Then $y \in g^{-1}[O]$ iff $g(y) \in O$ iff $y = f(g(y)) \in f[O]$, so $f[O] = g^{-1}[O]$. So $f[O]$ is open iff $g^{-1}[O]$ is open. So, for a bijection f, f is open iff f^{-1} is continuous.

If O is replaced in the argument by a closed set $F \subseteq X$, then similar reasoning shows that a bijection f is closed iff f^{-1} is continuous.

In part 1), the bijection f is not open and therefore $f^{-1} : S^1 \to [0, 2\pi)$ is not continuous. (*Explain directly, without part 2), why f^{-1} is not continuous.*)

Definition 8.9 A mapping $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is called a homeomorphism if f is a bijection and f and f^{-1} are both continuous. If a homeomorphism f exists, we say that X and Y are homeomorphic and write $X \simeq Y$.

Note: The term is "homeomorphism," not "homomorphism" (a term from algebra). The etymologies are closely related: "-morphism" comes from the Greek word ($\mu o\rho\phi\eta$) for "shape" or "form." The prefixes "homo" and "homeo" come from Greek words meaning same" and "similar" respectively. There was a major dispute in western religious history, mostly during the 4^{th} century AD, that hinged on the distinction between "homeo" and "homo."

As noted in the preceding example, we could also describe a homeomorphism as a "continuous open bijection" or a "continuous closed bijection."

It is obvious that among topological spaces, homeomorphism is an equivalence relation, that is, for topological spaces X, Y, and Z :

i) $X \simeq X$
ii) if $X \simeq Y$, then $Y \simeq X$
iii) if $X \simeq Y$ and $Y \simeq Z$, then $X \simeq Z$.

Example 8.10

1) The function $f : [0, 2\pi) \to S^1$ given by $f(\theta) = (\cos\theta, \sin\theta)$ is not a homeomorphism even though f is continuous, one-to-one, and onto.

2) The "central projection" from the rectangle to the circle (*Example 8.5.3*) is a homeomorphism.

3) It is easy to see that any two open intervals (a, b) in $\mathbb{R}$ are homeomorphic (just use a linear map of one interval onto the other).

The mapping $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ is a homeomorphism, so that each nonempty open interval in $\mathbb{R}$ is actually homeomorphic to $\mathbb{R}$ itself.

3) If $f : (X, d) \to (Y, s)$ is an isometry (onto) between metric spaces, then both f and f^{-1} are continuous, so f is a homeomorphism.

4) If d and d' are equivalent metrics (so $\mathcal{T}_d = \mathcal{T}_{d'}$), then the identity map $i : (X, d) \to (X, d')$ is a homeomorphism. Notice, however, that i doesn't preserve distances (unless $d = d'$).

In general, a homeomorphism between metric spaces need not be an isometry. But, of course, an isometry is automatically a homeomorphism

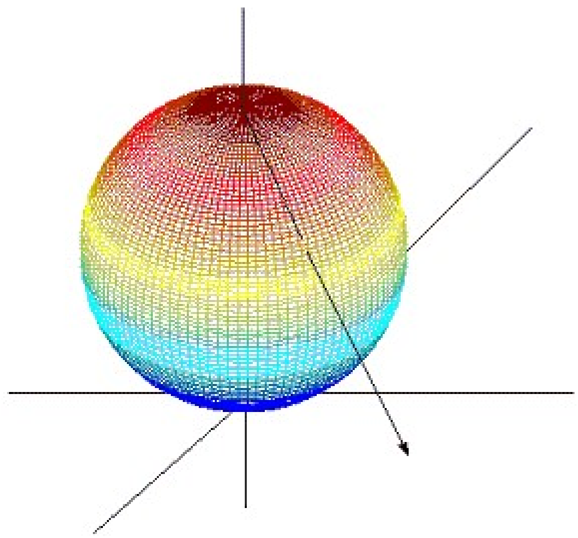
5) The function $f : \{\frac{1}{n} : n \in \mathbb{N}\} \to \mathbb{N}$ given by $f(\frac{1}{n}) = n$ is a homeomorphism (*both spaces have the discrete topology!*) Topologically, these spaces are the identical: both are just countable infinite sets with the discrete topology. f is not an isometry

In general, two discrete spaces X and Y are homeomorphic iff they have the same cardinality:any bijection between them is a homeomorphism. Roughly speaking, "size is the only possible topological difference between two discrete spaces."

6) Let P denote the "north pole" of the sphere

$$S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\} \subseteq \mathbb{R}^3.$$

The function f illustrated below is a "stereographic projection". The arrow starts at P, runs for a while inside the sphere and then exits through the surface of the sphere at a point (x, y, z). Let $f(x, y, z)$ be the point where the tip of the arrow hits the xy-plane $\mathbb{R}^2$. In this way f maps each point in $S^2 - \{P\}$ to a point in $\mathbb{R}^2$. The function f is a homeomorphism. (*See the figure below. Consider the images or inverse images of open sets.*)



In general, what is the significance of a homeomorphism $f : X \to Y$?

i) f is a bijection so it sets up a perfect one-to-one correspondence between the points in X and Y: $x \leftrightarrow y = f(x)$. We can imagine that f just "renames" the points in X. There is also a perfect one-to-one correspondence between the subsets X_α and Y_α of X and Y: $X_\alpha \leftrightarrow Y_\alpha = f[X_\alpha]$. Because f is a bijection, each subset $Y_\alpha \subseteq Y$ corresponds in this way to one and only one subset $X_\alpha \subseteq X$.

ii) f is a bijection, so f treats unions, intersections and complements "nicely":

a) $f^{-1}[\bigcup Y_\alpha : \alpha \in A] = \bigcup\{f^{-1}[Y_\alpha] : \alpha \in A\}$
b) $f^{-1}[\bigcap Y_\alpha : \alpha \in A] = \bigcap\{f^{-1}[Y_\alpha] : \alpha \in A\}$, and
c) $f[\bigcup X_\alpha : \alpha \in A] = \bigcup\{f[X_\alpha] : \alpha \in A\}$
d) $f[\bigcap X_\alpha : \alpha \in A] = \bigcap\{f[X_\alpha] : \alpha \in A\}$, and
e) $f[X - C] = f[X] - f[C] = Y - f[C]$

(*Actually a), b), c) are true for <u>any</u> function ; but d) and e) depend on f being a bijection.*)

These properties say that this correspondence between subsets preserves unions: if each $X_\alpha \leftrightarrow Y_\alpha$, then $\bigcup X_\alpha \leftrightarrow \bigcup Y_\alpha = \bigcup f[X_\alpha] = f[\bigcup X_\alpha]$. Similarly, f preserves intersection and complements.

iii) Finally if f and f^{-1} are continuous, then open (closed) sets in X correspond to open (closed) sets in Y and vice-versa.

The total effect is that all the "topological structure" in X is exactly "duplicated" in Y and vice versa: we can think of points, subsets, open sets and closed sets in Y are just "renamed" copies of their counterparts in X. Moreover f preserves unions, intersections and complements, so f also preserves all properties of X that can defined be using unions, intersections and complements of open sets. For example, we can check that if f is a homeomorphism and $A \subseteq X$, then $f[\text{int}_X A] = \text{int}_Y f[A]$, that $f[\text{cl}_X A] = \text{cl}_Y f[A]$, and that $f[\text{Fr}_X A] = \text{Fr}_Y f[A]$. That is, f takes interiors to interiors, closures to closures, and boundaries to boundaries.

Definition 8.11 A property P of topological spaces is called a topological property if, whenever a space X has property P and $Y \simeq X$, then the space Y also has property P.

If X and Y are homeomorphic, then the very definition of "topological property" says that X and Y have the same topological properties. Conversely, if two topological spaces X and Y have the same topological properties, then X and Y must be homeomorphic. (*Why? Let P be the property "is homeomorphic to X." P is a topological property because if Y has P (that is, if $Y \simeq X$) and $Z \simeq Y$, then Z also has P. Moreover, X has this property P, because $X \simeq X$. So if we assume that Y has the same topological properties as X, then Y has the property P, that is, Y is homeomorphic to X.*)

So we think of two homeomorphic spaces as "topologically identical" $-$ they are homeomorphic iff they have exactly the same topological properties. We can show that two spaces are not homeomorphic by naming a topological property of one space that the other space doesn't possess.

Example 8.12 Let P be the property that "every continuous real-valued function achieves a maximum value." Suppose a space X has property P and that $h : X \to Y$ is a homeomorphism. We claim that Y also has property P.

Let f be any continuous real-valued function defined on Y.
Then $f \circ h : X \to \mathbb{R}$ is continuous :

$$\begin{tikzcd} & \mathbb{R} \\ X \arrow[ur, "g"] \arrow[r, "h"'] & Y \arrow[u, "f"'] \end{tikzcd}$$

By assumption, $g = f \circ h$ achieves a maximum value at some point $a \in X$, and we claim that f must achieve a maximum value at the point $b = h(a) \in Y$. If not, then there is a point $y \in Y$ where $f(y) > f(b)$. Let $x = h^{-1}(y)$. Then

$$g(x) = f(h(x))) = f(h(h^{-1}(y)))) = f(y) > f(b) = f(h(a)) = g(a),$$

which contradicts the fact that g achieves a maximum value at a.

Therefore P is a topological property.

For example, the closed interval $[0,1]$ has property P discussed in Example 8.12 (*this is a well-known fact from elementary analysis – which we will prove later*). But $(0,1)$ and $[0,1)$ do not have this property P (*why?*). So we can conclude that $[0,1]$ is not homeomorphic to either $(0,1)$ or $[0,1)$.

Some other simple examples of topological properties are cardinality, first and second countability, Lindelöf, separability, and (pseudo)metrizability. In the case of metrizability, for example:

> If (X,d) is a metric space and $f:(X,d)\rightarrow(Y,\mathcal{T})$ is a homeomorphism, then we can define a metric d' on Y as $d'(a,b)=d(f^{-1}(a),f^{-1}(b))$ for $a,b\in Y$. You then need to check that $\mathcal{T}_{d'}=\mathcal{T}$ (using the properties of a homeomorphism and the definition of d'). This shows that $(Y,\mathcal{T})$ is metrizable. Be sure you can do this!

9. Sequences

In Chapter II we saw that sequences are a useful tool for working with pseudometric spaces. In fact, sequences are sufficient to describe the topology in a pseudometric space – because the convergent sequences in (X,d) determine the closure of a set.

We can easily define convergent sequences in any topological space $(X,\mathcal{T})$. But, as we will see, sequences need to be used with more care in spaces that are not pseudometrizable. Whether or not a sequence (x_n) converges to a particular point x is a "local" question – it depends on "the x_n's approaching nearer and nearer to x" and, in the absence of a distance function, we use the neighborhoods of x to determine "nearness to x." If the neighborhood system $\mathcal{N}_x$ is "too large" or "too complicated," then it may be impossible for a sequence to "get arbitrarily close" to x. Soon we will see a specific example where such a difficulty actually occurs. But first, we look at some of the things that <u>do</u> work out just as nicely for topological spaces as they do in pseudometric spaces.

Definition 9.1 Suppose (x_n) is a sequence in $(X,\mathcal{T})$. We say that (x_n) <u>converges</u> <u>to</u> <u>x</u> if, for every neighborhood W of x, $\exists k\in\mathbb{N}$ such that $x_n\in W$ when $n\geq k$. In this case we write $(x_n)\rightarrow x$. More informally, we can say that $(x_n)\rightarrow x$ if (x_n) is <u>eventually in every neighborhood</u> W of x.

Clearly, we can replace "every neighborhood W of x" in the definition with "every basic neighborhood B of x" or "every open set O containing x." Be sure you are convinced of this.

In a pseudometric space a sequence can converge to more than one point, but we proved that in a metric space limits of convergent sequences must be unique. A similar distinction holds in topological spaces: the important issue is whether we can "separate points by open sets."

Definition 9.2 $(X, \mathcal{T})$ is a T_1-space if whenever $x \neq y \in X$ there exist open sets U and V such that $x \in U$, $y \notin U$ and $y \in V$, $x \notin V$ *(that is, each point is in an open set that does not contain the other point)*.

$(X, \mathcal{T})$ is a T_2-space (or Hausdorff space) if whenever $x \neq y \in X$ there exist disjoint open sets U and V such that $x \in U$ and $y \in V$.

It is easy to check that i) X is a T_1-space iff for every $x \in X$, $\{x\}$ is closed, that
ii) every T_2-space is a T_1-space
iii) every metric space (X, d) is a T_2 -space (Hausdorff space)

There is a hierarchy of ever stronger "separation axioms" called $T_0, T_1, T_2, T_3,$ and T_4 that a topological space might satisfy. Eventually we will look at all of them. Each condition is stronger than the preceding ones in the list (for example, $T_2 \Rightarrow T_1$). The letter "T" is used here because in the early (German) literature, the word for "separation axioms" was "Trennungsaxiome."

Theorem 9.3 In a Hausdorff space $(X, \mathcal{T})$, a sequence can converge to at most one point.

Proof Suppose $x \neq y \in X$. Choose disjoint open sets U and V with $x \in U$ and $y \in V$. If $(x_n) \to x$, then (x_n) is eventually in U, so (x_n) is not eventually in V. Therefore (x_n) does not also converge to y. ●

When we try to generalize results from pseudometric spaces to topological spaces, we often get a better insight about where the heart of a proof lies. For example, to prove that limits of sequences are unique it is the Hausdorff property that is important, not the presence of a metric. Here is another example: for a pseudometric space (X, d) we proved that $x \in \text{cl } A$ iff x is a limit of a sequence (a_n) in A. That proof (*see Theorem II.5.18*) used the fact that there was a countable neighborhood base $\{B_{\frac{1}{n}}(x) : n \in \mathbb{N}\}$ at each point x. We can see now that the countable neighborhood base was the crucial fact – because we can prove the same result in any first countable topological space $(X, \mathcal{T})$.

But first, two technical lemmas are helpful.

Lemma 9.4 Suppose $\{V_1, V_2, ..., V_k, ...\}$ is a countable neighborhood base at $a \in X$. Define $U_k = \text{int}(V_1 \cap ... \cap V_k)$. Then $\{U_1, U_2, ..., U_k, \; ...\}$ is also a neighborhood base at a.

Proof $V_1 \cap ... \cap V_k$ is a neighborhood of a, so $a \in \text{int}(V_1 \cap ... \cap V_k) = U_k$. Therefore U_k is a open neighborhood of a. If N is any neighborhood of a, then $a \in V_k \subseteq N$ for some k, so then $a \in U_k \subseteq V_k \subseteq N$. Therefore the U_k's are a neighborhood base at a. ●

The exact formula for the U_k's in Lemma 9.4 doesn't matter; the important thing is that we get a "much improved" neighborhood base $\{U_1, U_2, ..., U_k, \; ...\}$ – one in which the U_k's are open and $U_1 \supseteq U_2 \supseteq ... \supseteq U_k \supseteq ...$. This new neighborhood base at a plays a role like the neighborhood base $B_1(a) \supseteq B_{\frac{1}{2}}(a) \supseteq ... \supseteq B_{\frac{1}{k}}(a) \supseteq ...$ in a pseudometric space. We call $\{U_1, U_2, ..., U_k, ...\}$ an open, shrinking neighborhood base at a.

Lemma 9.5 Suppose $\{U_1, U_2, ..., U_k, ...\}$ is a shrinking neighborhood base at x and that $a_n \in U_n$ for each n. Then $(a_n) \to x$.

Proof If W is any neighborhood of x, then there is a k such that $x \in U_k \subseteq W$. Since the U_k's are a shrinking neighborhood base, we have that for any $n \geq k$, $a_n \in U_n \subseteq U_k \subseteq W$. So $(a_n) \to x$. ●

Theorem 9.6 Suppose $(X, \mathcal{T})$ is first countable and $A \subseteq X$. Then $x \in \text{cl}\, A$ iff there is a sequence (a_n) in A such that $(a_n) \to x$. (*More informally, "sequences are sufficient" to describe the topology in a first countable topological space.*)

Proof ($\Leftarrow$) Suppose (a_n) is a sequence in A and that $(a_n) \to x$. For each neighborhood W of x, (a_n) is eventually in W. Therefore so $W \cap A \neq \emptyset$, so $x \in \text{cl}\, A$. (*This half of the proof works in any topological space: it does not depend on first countability.*)

($\Rightarrow$) Suppose $x \in \text{cl}\, A$. Using Lemma 9.4, choose a countable shrinking neighborhood base $\{U_1, ..., U_n, ...\}$ at x. Since $x \in \text{cl}\, A$, we can choose a point $a_n \in U_n \cap A$ for each n. By Lemma 9.5, $(a_n) \to x$. ●

We can use Theorem 9.6 to get an upper bound on the size of certain topological spaces, analogous to what we did for pseudometric spaces. This result is not very important, but it illustrates that in Theorem II.5.21 the properties that are really important are "first countability" and "Hausdorff," not the actual presence of a metric d.

Corollary 9.7 If D is a dense subset in a first countable Hausdorff space $(X, \mathcal{T})$, then $|X| \leq |D|^{\aleph_0}$. In particular, If X is a separable, first countable Hausdorff space, then $|X| \leq \aleph_0^{\aleph_0} = c$.

Proof X is first countable, so for each $x \in X$ we can pick a sequence (d_n) in D such that $(d_n) \to x$; formally, this sequence is a function $f_x : \mathbb{N} \to D$, so $f_x \in D^{\mathbb{N}}$. Since X is Hausdorff, a sequence cannot converge to two different points: so if $x \neq y \in X$, then $f_x \neq f_y$. Therefore the function $\Phi : X \to D^{\mathbb{N}}$ given by $\Phi(x) = f_x$ is one-to-one, so $|X| \leq |D^{\mathbb{N}}| = |D|^{\aleph_0}$. ●

The conclusion in Theorem 9.6 may not be true if X is not first countable: sequences are not always "sufficient to describe the topology" of X – that is, convergent sequences cannot always determine the closure of a set.

Example 9.8 (the space L)

Let $L = \{(m, n) : m, n \in \mathbb{Z},\, m, n \geq 0\}$, and let C_j be "the j^{th} column of L," that is $C_j = \{(j, n) \in L : n = 0, 1, ...\}$. We put a topology on L by giving a neighborhood base at each point p :

$$\mathcal{B}_p = \begin{cases} \{(m,n)\} & \text{if } p = (m,n) \neq (0,0) \\ \{B : (0,0) \in B \text{ and } C_j - B \text{ is finite for all but finitely many } j\} & \text{if } p = (0,0) \end{cases}$$

(*Check that this definition satisfies the conditions in the Neighborhood Base Theorem 5.2 and therefore does describe a topology for* L.)

If $p \neq (0,0)$ then p is isolated in L. A basic neighborhood of $(0,0)$ is a set which contains $(0,0)$ and which, we could say, contains "most of the points from most of the columns." With this topology, L is a Hausdorff space.

Certainly $(0,0) \in \text{cl}\,(L - \{(0,0)\})$, but no sequence from $L - \{(0,0)\}$ converges to $(0,0)$. To see this, consider any sequence (a_n) in $L - \{(0,0)\}$:

i) if there is a column C_{j_0} that contains infinitely many of the terms a_n, then $N = (L - C_{j_0}) \cup \{(0,0)\}$ is a neighborhood of $(0,0)$ and (a_n) is not eventually in N.

ii) if every column C_j contains only finitely many a_n's, then $N = L - \{a_n : n \in \mathbb{N}\}$ is a neighborhood of $(0,0)$ and (a_n) is not eventually in N (in fact, the sequence is never in N).

In L, sequences are not sufficient to describe the topology: convergent sequences can't show us that $(0,0) \in \text{cl}\,(L - \{(0,0)\})$. According to Theorem 9.6, this means that L cannot be first countable $-$ there is a countable neighborhood base at each point $p \neq (0,0)$ but not at $(0,0)$.

The neighborhood system at $(0,0)$ "measures nearness to" $(0,0)$ but the ordering relationship ($\supseteq$) among the basic neighborhoods at $(0,0)$ is very complicated $-$ much more complicated than the neat, simple nested chain of neighborhoods $U_1 \supseteq U_2 \supseteq ... \supseteq U_n \supseteq ...$ that could form a base at x in a first countable space. Roughly, the complexity of the neighborhood system is the reason why the terms of a sequence can't get "arbitrarily close" to $(0,0)$.

Sequences do suffice to describe the topology in a first countable space, so it is not surprising that we can use sequences to determine the continuity of a function defined on a first countable space X.

Theorem 9.9 Suppose $(X, \mathcal{T})$ is first countable and $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$. Then f is continuous at $a \in X$ iff whenever $(x_n) \to a$, then $(f(x_n)) \to f(a)$.

Proof ($\Rightarrow$) If f is continuous at a and W is a neighborhood of $f(a)$, then $f^{-1}[W]$ is a neighborhood of a. Therefore (x_n) is eventually in $f^{-1}[W]$, so $f((x_n))$ is eventually in W. (*This half of the proof is valid for any topological space* X*: continuous functions always "preserve convergent sequences."*)

($\Leftarrow$) Let $\{U_1, ..., U_n, ...\}$ be a shrinking neighborhood base at a. If f is not continuous at a, then there is a neighborhood W of $f(a)$ such that for every n, $f[U_n] \nsubseteq W$. For each n, choose a point $x_n \in U_n - f^{-1}[W]$. Then (since the U_n's are shrinking) we have $(x_n) \to a$ but $(f(x_n))$ fails to converge to $f(a)$ because $f(x_n)$ is never in W. •

(*Compare this to the proof of Theorem II.5.22.*)

10. Subsequences

Definition 10.1 Suppose $f : \mathbb{N} \to X$ is a sequence in X and that $\phi : \mathbb{N} \to \mathbb{N}$ is strictly increasing. The composition $f \circ \phi : \mathbb{N} \to X$ is called a <u>subsequence</u> of f.

$$\begin{array}{ccc} & f & \\ \mathbb{N} & \longrightarrow & X \\ \phi \uparrow & \nearrow & f \circ \phi \\ \mathbb{N} & & \end{array}$$

If we write $f(n) = x_n$ and $\phi(k) = n_k$, then $(f \circ \phi)(k) = f(n_k) = x_{n_k}.$ We write the sequence f informally as (x_n) and the subsequence $f \circ \phi$ as (x_{n_k}). Since ϕ is increasing, we have that $n_k \to \infty$ as $k \to \infty$.

For example, if $\phi(k) = n_k = 2k$, then $f \circ \phi$ is the subsequence written informally as $(x_{n_k}) = (x_{2k})$, that is, the subsequence $(x_2, x_4, x_6, ..., x_{2k}, ...)$. But if $\phi(n) = 1$ for all n, then $f \circ \phi$ is <u>not</u> a subsequence: informally, $(x_1, x_1, x_1, ..., x_1, ...)$ is not a subsequence of $(x_1, x_2, x_3, ..., x_n, ...)$. Every sequence f is a subsequence of itself: just let $\phi(n) = n$.

Theorem 10.2 Suppose $x \in (X, \mathcal{T})$. Then $(x_n) \to x$ iff every subsequence $(x_{n_k}) \to x$.

Proof ($\Leftarrow$) This is clear because (x_n) is a subsequence of itself.

($\Rightarrow$) Suppose $(x_n) \to x$ and that (x_{n_k}) is a subsequence. If W is any neighborhood of x, then $x_n \in W$ for all $n >$ some n_0. Since the n_k's are strictly increasing, $n_k > n_0$ for all $k >$ some k_0. Therefore (x_{n_k}) is eventually in W, so $(x_{n_k}) \to x$. ∙

Definition 10.3 Suppose $x \in (X, \mathcal{T})$. We say that x is a <u>cluster point</u> of the sequence (x_n) if for each neighborhood W of x and <u>for each $k \in \mathbb{N}$</u>, <u>there is an $n > k$</u> for which $x_n \in W$. More informally, we say that x is a cluster point of (x_n) if the sequence is <u>frequently in every neighborhood W of x</u>.

Definition 10.4 Suppose $x \in (X, \mathcal{T})$ and $A \subseteq X$. We say that x is a <u>limit point</u> of A if $N \cap (A - \{x\}) \neq \emptyset$ for every neighborhood N of x $-$ that is, every neighborhood of x contains points of <u>arbitrarily close</u> to x <u>but different</u> from x.

Example 10.5.

1) Suppose $X = [0,1] \cup \{2\}$ and $A = \{2\}$. Then $W \cap A \neq \emptyset$ for every neighborhood W of 2, but 2 is not a limit point of A $-$ because $W = \{2\}$ is a neighborhood of 2 in X and $W \cap (A - \{2\}) = \emptyset$. Each $x \in [0,1]$ is a limit point of $[0,1]$ and also a limit point of X. Since $\{2\}$ is open in X, 2 is also not a limit point of X.

2) Every point r in $\mathbb{R}$ is a limit point of $\mathbb{Q}$. If $A \subseteq \mathbb{N}$, then A has no limit points (in $\mathbb{N}$ or in $\mathbb{R}$).

3) If $(x_n) \to x$, then x is a cluster point of (x_n). More generally, if (x_n) has a subsequence (x_{n_k}) that converges to x, then x is a cluster point of (x_n). (*Why?*)

4) A sequence can have many cluster points. For example, if the sequence (q_n) lists all the elements of $\mathbb{Q}$, then every $r \in \mathbb{R}$ is a cluster point of (q_n).

5) In $\mathbb{R}$, the sequence $(x_n) = ((-1)^n)$ has exactly two cluster points: -1 and 1. But the set $\{x_n : n \in \mathbb{N}\} = \{-1, 1\}$ has no limit points in $\mathbb{R}$. The set of cluster points of a sequence is not always the same as the set of limit points of the set of terms in the sequence! (*Is one of these sets always a subset of the other?*)

Theorem 10.6 Suppose a is a cluster point of (x_n) in a first countable space $(X, \mathcal{T})$. Then there is a subsequence $(x_{n_k}) \to a$.

Proof Let $\{U_1, U_2, ..., U_n, ...\}$ be a countable shrinking neighborhood base at a. Since (x_n) is frequently in U_1, we can pick n_1 so that $x_{n_1} \in U_1$. Since (x_n) is frequently in U_2, we can pick an $n_2 > n_1$ so that $x_{n_2} \in U_2 \subseteq U_1$. Continue inductively: having chosen $n_1 < ... < n_k$ so that $x_{n_k} \in U_k \subseteq ... \subseteq U_1$, we can then choose $n_{k+1} > n_k$ so that $x_{n_{k+1}} \in U_{k+1} \subseteq U_k$. Then (x_{n_k}) is a subsequence of (x_n) and $(x_n) \to a$. •

Example 10.7 (the space L, revisited)

Let L be the space in Example 9.8 and let (x_n) be a sequence which lists all the elements of $L - \{(0,0)\}$.

Every basic neighborhood B of $(0,0)$ is infinite, so B must contain terms x_n for arbitrarily large n. This means that (x_n) is frequently in B, so $(0,0)$ is a cluster point of (x_n).

But no subsequence of (x_n) can converge to $(0,0)$ – because we showed in Example 9.8 that no sequence whatsoever from $L - \{(0,0)\}$ can converge to $(0,0)$. Therefore Theorem 10.6 may not be true if the space X is not first countable.

Consider any sequence $(x_n) \to (0,0)$ in L. If there were infinitely many $x_n \neq (0,0)$, then we could form the subsequence that contains those terms, and that subsequence would be a sequence in $L - \{(0,0)\}$ that converges to $(0,0)$ – which is impossible. Therefore we conclude that eventually $x_n = (0,0)$.

Suppose now that $f : L \to \mathbb{N}$ is any bijection, and let $k = f(0,0)$.

- whenever a sequence $(x_n) \to (0,0)$ in L, then $(f(x_n)) \to f(0,0) = k$ in $\mathbb{N}$ (*because, by the preceding paragraph,* $f(x_n) = f(0,0) = k$ *eventually)*
- the topology on $\mathbb{N}$ is discrete so $\{k\}$ is a neighborhood of $f(0,0)$, but $f^{-1}[\{k\}] = \{(0,0)\}$ is not a neighborhood of $(0,0)$. Therefore f is not continuous at $(0,0)$

Theorem 9.9 does not apply to L: if a space is not first countable, sequences may be inadequate to check whether a function f is continuous at a point.

Exercises

E17. Suppose $f, g : (X, \mathcal{T}_1) \to (Y, \mathcal{T}_2)$ are continuous functions, that D is dense in X, and that $f|D = g|D$. Prove that if Y is Hausdorff, then $f = g$. (*This generalizes the result in Chapter 2, Theorem 5.12.)*

E18. A function $f : (X, \mathcal{T}) \to \mathbb{R}$ is called <u>lower semicontinuous</u> if

$$f^{-1}[(b,\infty)] = \{x : f(x) > b\} \text{ is open for every } b \in \mathbb{R},$$

and f is called <u>upper semicontinuous</u> if

$$f^{-1}[(-\infty, b)] = \{x : f(x) < b\} \text{ is open for each } b \in \mathbb{R}.$$

a) Show that f is continuous iff f is both upper and lower semicontinuous.

b) Give an example of a lower semicontinuous $f : \mathbb{R} \to \mathbb{R}$ which is not continuous. Do the same for upper semicontinuous.

c) Suppose $A \subseteq X$. Prove that the characteristic function χ_A is lower semicontinuous if A is open in X and upper semicontinuous if A is closed in X.

E19. Suppose X is an infinite set with the cofinite topology, and that Y has the property that every singleton $\{y\}$ is a closed set. (*You might want to check: this is equivalent to saying that Y is a T_1-space: see Definition 9.2*). Prove that if $f: X \to Y$ is continuous and onto, then <u>either</u> f is constant <u>or</u> X is homeomorphic to Y.

1) Note: the problem does <u>not</u> say that if f is not constant, then f is a homeomorphism.
2) Hint: Prove first that if f is not constant, then $|X| = |Y|$. Then examine the topology of Y.

E20. Suppose X is a countable set with the cofinite topology. State and prove a theorem that completely answers the question: "what sequences in $(X, \mathcal{T})$ converge to what points?"

E21. Suppose that $(X, \mathcal{T})$ and $(Y, \mathcal{T}')$ are topological spaces. Recall that the <u>product topology</u> on $X \times Y$ is the topology for which the collection of "open boxes"

$$\mathcal{B} = \{U \times V : U \in \mathcal{T},\ V \in \mathcal{T}'\} \text{ is a base.}$$

a) The "projection maps" $\pi_x : X \times Y \to X$ and $\pi_y : X \times Y \to Y$ are defined by

$$\pi_x(x, y) = x \text{ and } \pi_y(x, y) = y.$$

We showed in Example 5.11 that π_x and π_y are continuous. Prove that π_x and π_y are open maps. Give examples to show that π_x and π_y might not be closed.

b) Suppose that $(Z, \mathcal{T}'')$ is a topological space and that $f : Z \to X \times Y$. Prove that f is continuous iff both compositions $\pi_x \circ f : Z \to X$ and $\pi_y \circ f : Z \to Y$ are continuous. (*Informally: a mapping into a product is continuous iff its composition with each projection is continuous.*)

c) Prove that $((x_n, y_n)) \to (x, y) \in X \times Y$ iff $(x_n) \to x$ in X and $(y_n) \to y$ in Y. (*For this reason, the product topology is sometimes called the "topology of coordinatewise convergence."*)

d) Prove that $X \times Y$ is homeomorphic to $Y \times X$.
(*Topological products are commutative.*)

e) Prove that $(X \times Y) \times Z$ is homeomorphic to $X \times (Y \times Z)$
(*Topological products are associative.*)

E22. Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. Suppose $f : X \to Y$. Let

$$\Gamma(f) = \{(x, y) \in X \times Y \colon y = f(x)\} = \text{"the graph of } f.\text{"}$$

Prove that the map $h \colon X \to \Gamma(f)$ defined by $h(x) = (x, f(x))$ is a homeomorphism if and only if f is continuous.

Note: if we think of f as a set of ordered pairs, the "graph of f" is f. More informally, however, the problem states a function is continuous iff its graph is homeomorphic to its domain.

E23. In $(X, \mathcal{T})$, a family of sets $\mathcal{F} = \{B_\alpha : \alpha \in \Lambda\}$ is called locally finite if each point $x \in X$ has a neighborhood N such that $N \cap B_\alpha \neq \emptyset$ for only finitely many α's. (*Part b) was also in Exercise E2.*)

a) Suppose (X, d) is a metric space and that $\mathcal{F}$ is a family of closed sets. Suppose there is an $\epsilon > 0$ such that $d(B_1, B_2) \geq \epsilon$ for all $B_1, B_2 \in \mathcal{F}$. Prove that $\mathcal{F}$ is locally finite.

b) Prove that if $\mathcal{F}$ is a locally finite family of sets in $(X, \mathcal{T})$, then $\mathrm{cl}\,(\bigcup_{\alpha \in A} B_\alpha) = \bigcup_{\alpha \in A} \mathrm{cl}\,(B_\alpha)$. Explain why this implies that if all the B_α's are closed, then $\bigcup_{\alpha \in A} B_\alpha$ is closed. (*This would apply, for example, to the sets in part a*). *See also Exercise E2.*)

c) (*The Pasting Lemmas: compare Exercise II.E24*) Let $(X, \mathcal{T})$ and $(Y, \mathcal{T}')$ be topological spaces. For each $\alpha \in A$, suppose $B_\alpha \subseteq X$, that $f_\alpha : B_\alpha \to Y$ is a continuous function, and that $f_\alpha | (B_\alpha \cap B_\beta) = f_\beta | (B_\alpha \cap B_\beta)$ for all $\alpha, \beta \in A$ Then $\bigcup_{\alpha \in A} f_\alpha = f$ is a function and $f : \bigcup B_\alpha \to Y$. (*Informally: each pair of functions f_α and f_β agree wherever their domains overlap; this allows use to define f by "pasting together" all the "function pieces"*)

i) Show that if all the B_α's are open, then f is continuous.

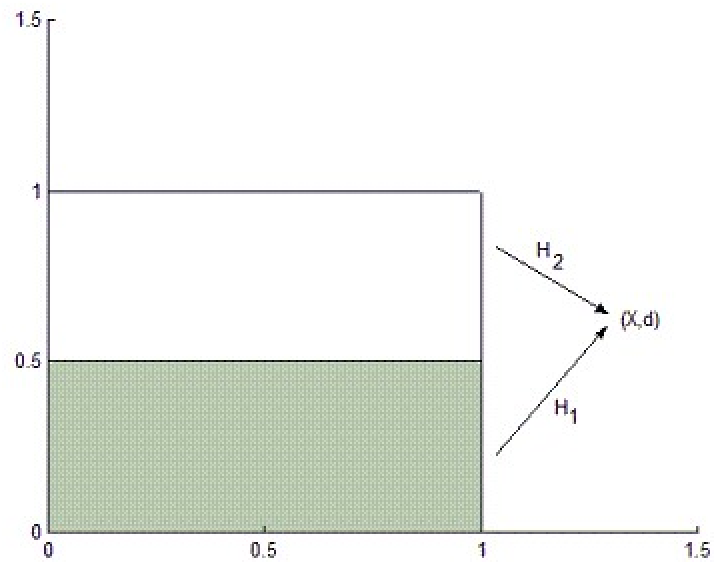
ii) Show that if there are only finitely many B_α's and they are all closed, then f is continuous. (*Hint: use a characterization of continuity in terms of closed sets.*)

iii) Give an example to show that f might not be continuous when there are infinitely many B_α's all of which are closed.

iv) Show that if $\mathcal{F}$ is a locally finite family of closed sets, then f is continuous. (*Of course, iv) $\Rightarrow$ ii)*).

Note: the most common use of the Pasting Lemma is when the index set A is finite. For example, suppose

$H_1 : [0,1] \times [0, \frac{1}{2}] \to (X,d)$ *is continuous, and*
$H_2 : [0,1] \times [\frac{1}{2}, 1] \to (X,d)$ *is continuous, and*
$H_1(t, \frac{1}{2}) = H_2(t, \frac{1}{2})$ *for all* $t \in [0,1]$



H_1 is defined on the lower closed half of the box $[0,1]^2$, H_2 is defined on the upper closed half, and they agree on the "overlap" – that is, on the horizontal line segment $[0,1] \times \{\frac{1}{2}\}$. Part b) (or part c)) says that the two functions can be pieced together into a continuous function $H : [0,1]^2 \to (X,d)$, where $H = H_1 \cup H_2$, that is

$$H(t,y) = \begin{cases} H_1(t,y) & \textit{if } t \leq \frac{1}{2} \\ H_2(t,y) & \textit{if } t \geq \frac{1}{2} \end{cases}$$

Chapter III Review

Explain why each statement is true, or provide a counterexample. If nothing else is mentioned, X and Y are topological spaces with no other properties assumed.

1. For every possible topology $\mathcal{T}$, the space $(\{0,1,2\},\mathcal{T})$ is pseudometrizable.

2. A convergent sequence in a first countable topological space has at most one limit.

3. A one point set $\{x\}$ in a pseudometric space (X,d) is closed.

4. Suppose $\mathcal{T}$ and $\mathcal{T}'$ are topologies on X and that for every subset A of X, $\mathrm{cl}_{\mathcal{T}}(A) = \mathrm{cl}_{\mathcal{T}'}(A)$. Then $\mathcal{T} = \mathcal{T}'$.

5. Suppose $f\colon X \to Y$ and $A \subseteq X$. If $f|A : A \to Y$ is continuous, then f is continuous at each point of A.

6. Suppose $\mathcal{S}$ is a subbase for the topology on X and that $D \subseteq X$. If $U \cap D \neq \emptyset$ for every $U \in \mathcal{S}$, then D is dense in X.

7. If D is dense in $(X,\mathcal{T})$ and $\mathcal{T}^*$ is another topology on X with $\mathcal{T} \subseteq \mathcal{T}^*$, then D is dense in $(X,\mathcal{T}^*)$.

8. If $f\colon X \to Y$ is both continuous and open, then f is also closed.

9. Every space is a continuous image of a first countable space.

10. Let $X = \{0,1\}$ with topology $\mathcal{T} = \{\emptyset, X, \{1\}\}$. There are exactly 3 continuous functions $f : (X,\mathcal{T}) \to (X,\mathcal{T})$.

11. If A and B are subspaces of $(X,\mathcal{T})$ and both A and B are discrete in the subspace topology, then $A \cup B$ is discrete in the subspace topology.

12. If $A \subseteq (X,\mathcal{T})$ and X is separable, then A is separable.

13. If $\mathcal{T}$ is the cofinite topology on X. Every bijection $f: (X, \mathcal{T}) \to (X, \mathcal{T})$ is a homeomorphism.

14. If X has the cofinite topology, then the closure of any open set in X is open.

15. A continuous bijection from $\mathbb{R}$ to $\mathbb{R}$ must be a homeomorphism.

16. For every cardinal m, there is a separable topological space $(X, \mathcal{T})$ with $\mid X \mid = m$.

17. If every family of disjoint open sets in $(X, \mathcal{T})$ is countable, then $(X, \mathcal{T})$ is separable.

18. For $a \in \mathbb{R}^2$ and $\epsilon > 0$, let $C_\epsilon(a) = \{x \in \mathbb{R}^2 \colon d(x, a) = \epsilon\}$. Let $\mathcal{T}$ be the topology on $\mathbb{R}^2$ for which the collection $\{C_\epsilon(a) \colon \epsilon > 0, a \in \mathbb{R}^2\}$ is a subbasis. Let $\mathcal{U}$ be the usual topology on $\mathbb{R}^2$. Then the function $f: (\mathbb{R}^2, \mathcal{T}) \to (\mathbb{R}^2, \mathcal{U})$ given by $f(x, y) = (\sin x, \sin y)$ is continuous.

19. If $D \subseteq \mathbb{R}$ and each point of D is isolated in D, then cl D must be countable.

20. Consider the separation property S : every minimal nonempty closed set F is a singleton. (F *is a* *minimal* *nonempty closed set means: if* A *is a nonempty closed set and* $A \subseteq F$, *then* $A = F$). If X is a T_1-space, then X has property S.

21. A one-to-one, continuous, onto map $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ must be a homeomorphism.

22. An uncountable closed set in $\mathbb{R}$ must contain an interval of positive length.

23. A countable metric space has a base consisting of clopen sets.

24. Suppose $D \subseteq X$ and that D is dense in $(X, \mathcal{T})$. If $\mathcal{T}^*$ is topology on X with $\mathcal{T}^* \subseteq \mathcal{T}$, then D is dense in $(X, \mathcal{T}^*)$.

25. If $D \subseteq \mathbb{R}$ and D is discrete in the subspace topology, then D is countable.

26. The Sorgenfrey plane has a subspace homeomorphic to $\mathbb{R}$ (with its usual topology).

27. In $\mathbb{N}$, let $\mathcal{B}_n = \{B \subseteq \mathbb{N} : n \in B \text{ and } B \subseteq \{1, 2, ..., n\}\}$. At each point n, the $\mathcal{B}_n$'s satisfy the conditions in the Neighborhood Base Theorem and therefore describe a topology on $\mathbb{N}$.

28. At each point $n \in \mathbb{N}$, let $\mathcal{B}_n = \{B : B \supseteq \{n, n+1, n+2, ...\}\}$. At each point n, the $\mathcal{B}_n$'s satisfy the conditions in the Neighborhood Base Theorem and therefore describe a topology on $\mathbb{N}$.

29. Suppose $(X, \mathcal{T})$ has a base $\mathcal{B}$ with $|\mathcal{B}| = c$. Then X has a dense set D with $|D| \leq c$.

30. Suppose $(X, \mathcal{T})$ has a base $\mathcal{B}$ with $|\mathcal{B}| = c$. Then at each point $x \in X$, there is a neighborhood base with $|\mathcal{B}_x| \leq c$.

31. Suppose X is an infinite set. Let $\mathcal{T}_1$ be the cofinite topology on X and let $\mathcal{T}_2$ be the discrete topology on X. If a function $f : \mathbb{R} \to (X, \mathcal{T}_1)$ is continuous, then $f : \mathbb{R} \to (X, \mathcal{T}_2)$ is also continuous.

32. Let $\mathcal{T}$ be the "right-ray topology" on $\mathbb{R}$, that is, $\mathcal{T} = \{(a, \infty) : a \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$. The space $(\mathbb{R}, \mathcal{T})$ is first countable.

Chapter IV
Completeness and Compactness

1. Introduction

In Chapter III we introduced topological spaces as a generalization of pseudometric spaces. This allowed us to see how certain ideas – continuity, for example – can be extended into a setting where there is no "distance" between points. Continuity does not really depend on the pseudometric d but only on the topology.

Looking at topological spaces also highlighted the particular properties of pseudometric spaces that were really important for certain purposes. For example, it turned out that first countability is the crucial ingredient for proving that sequences are sufficient to describe a topology, and that Hausdorff property, not the metric d, is what matters to prove that limits of sequences are unique.

We also looked at some properties that are distinct in topological spaces but equivalent in the special case of pseudometric spaces – for example, second countability and separability.

Most of the earlier definitions and theorems are just basic "tools" for our work. Now we look at some deeper properties of pseudometric spaces and some significant theorems related to them.

2. Complete Pseudometric Spaces

Definition 2.1 A sequence (x_n) in a pseudometric space (X, d) is called a Cauchy sequence if $\forall \epsilon > 0 \ \exists N \in \mathbb{N}$ such that if $m, n > N$, then $d(x_m, x_n) < \epsilon$.

Informally, a sequence (x_n) is Cauchy if its terms "get closer and closer to each other." It should be intuitively clear that this happens if the sequence converges, as the next theorem confirms.

Theorem 2.2 If $(x_n) \to x$ in (X, d), then (x_n) is Cauchy.

Proof Let $\epsilon > 0$. Because $(x_n) \to x$, there is an $N \in \mathbb{N}$ such that $d(x_n, x) < \frac{\epsilon}{2}$ for $n > N$. So if $m, n > N$, we get $d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. Therefore (x_n) is Cauchy. •

However, a Cauchy sequence does not always converge. For example, look at the space $(\mathbb{Q}, d)$ where d is the usual metric. Consider a sequence in $\mathbb{Q}$ that converges to $\sqrt{2}$ in $\mathbb{R}$; for example, we could use $(q_n) = (1, 1.4, 1.41, 1.414, \ldots)$.

- Since (q_n) is convergent in $\mathbb{R}$, (q_n) is a Cauchy sequence in $\mathbb{Q}$ (*"Cauchy" depends only on d and the numbers q_n, not whether we are thinking of these q_n's as elements of* $\mathbb{Q}$ *or* $\mathbb{R}$.
- But (q_n) has no limit in $(\mathbb{Q}, d)$. (*Why? To say that "$\sqrt{2} \notin \mathbb{Q}$" is part of the answer.*)

Definition 2.3 A pseudometric space (X, d) is called complete if every Cauchy sequence in (X, d) has a limit in X.

Example 2.4 In all parts of this example, d is the usual metric on subsets of $\mathbb{R}$.

1) $(\mathbb{R}, d)$ is complete (*for the moment, we assume this as a simple fact from analysis; however we will prove it soon*) but $(\mathbb{Q}, d)$ and $(\mathbb{P}, d)$ are not complete: completeness is not a hereditary property !

2) If (x_n) is a Cauchy sequence in $(\mathbb{N}, d)$, then there is some N such that $|x_m - x_n| < 1$ for $m, n > N$. Because the x_n's are integers, this implies that $x_m = x_n$ for $m, n > N$ – that is, (x_n) is eventually constant and therefore converges. Therefore $(\mathbb{N}, d)$ is complete.

3) Consider $\mathbb{N}$ with the equivalent metric $d'(m, n) = |\frac{1}{m} - \frac{1}{n}|$. (*$d' \sim d$ because both metrics produce the discrete topology on* $\mathbb{N}$.)

In $(\mathbb{N}, d')$, the sequence $(x_n) = (n)$ is Cauchy, and if $k \in \mathbb{N}$, then $d'(x_n, k) = |\frac{1}{n} - \frac{1}{k}| \to \frac{1}{k} \neq 0$ as $n \to \infty$, so $(x_n) \not\to k$. So the space $(\mathbb{N}, d')$ is not complete.

Since $(\mathbb{N}, d')$ has the discrete topology, the map $i(n) = n$ is a homeomorphism between $(\mathbb{N}, d')$ and the complete space $(\mathbb{N}, d)$. So completeness is not a topological property.

A homeomorphism takes convergent sequences to convergent sequences and nonconvergent sequences to nonconvergent sequences. For example, (n) does not converge in $(\mathbb{N}, d')$ and $(i(n)) = (n)$ does not converge in $(\mathbb{N}, d)$. These two homeomorphic metric spaces have exactly the same convergent sequences – but they do not have the same Cauchy sequences. Changing a metric d to a equivalent metric d' does not change the open sets and therefore does not change which sequences converge. But the change might create or destroy Cauchy sequences because the Cauchy property does depend specifically on how distances are measured.

Another similar example: we know that any open interval (a, b) is homeomorphic to $\mathbb{R}$. However, with the usual metric on each space, $\mathbb{R}$ is complete and (a, b) is not complete.

4) In light of the observations in 3) we can ask: if (X, d) is not complete, might there be some equivalent metric d' for which (X, d') is complete? To take a specific example: could we find some metric d' on $\mathbb{Q}$ so $d \sim d'$ but for which $(\mathbb{Q}, d')$ is complete? We could also ask this question about $(\mathbb{P}, d)$. Later in this chapter we will see the answer – and perhaps surprisingly, the answers for $\mathbb{Q}$ and $\mathbb{P}$ are different!

5) $(\mathbb{N}, d')$ and $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$ are countable discrete spaces, so they are homeomorphic. But (even better!) the function $f(n) = \frac{1}{n}$ is actually an isometry between these spaces because $d'(m, n) = |\frac{1}{m} - \frac{1}{n}| = d(f(m), f(n))$. An isometry is a homeomorphism, but since it preserves distances, an isometry also takes Cauchy sequences to Cauchy sequences and non-Cauchy sequences to non-Cauchy sequences. Therefore if there is an isometry between two pseudometric spaces, then one space is complete iff the other space is complete.

For example, the sequence $(x_n) = n$ is Cauchy in $(\mathbb{N}, d')$ and the isometry f carries (x_n) to the Cauchy sequence $(f(x_n)) = (\frac{1}{n})$ in $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$. It is clear that (n) has no limit in the domain and $(\frac{1}{n})$ has no limit in the range.

$(\mathbb{N}, d')$ and $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$ "look exactly alike" – not just topologically but as metric spaces. We can think of f as simply renaming the points in a distance-preserving way.

Theorem 2.5 If x is a cluster point of a Cauchy sequence (x_n) in (X, d), then $(x_n) \to x$.

Proof Let $\epsilon > 0$. Pick N so that $d(x_m, x_n) < \frac{\epsilon}{2}$ when $m, n > N$. Since (x_n) clusters at x, we can pick a $K > N$ so that $x_K \in B_{\frac{\epsilon}{2}}(x)$. Then for this K and for $n > N$,

$$d(x_n, x) \leq d(x_n, x_K) + d(x_K, x) < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon$$

so $(x_n) \to x$. ∙

Corollary 2.6 In a pseudometric space (X, d): a Cauchy sequence (x_n) converges iff (x_n) has a cluster point.

Corollary 2.7 In a metric space (X, d), a Cauchy sequence can have at most one cluster point.

Theorem 2.8 A Cauchy sequence (x_n) in (X, d) is bounded – that is, the set $\{x_1, x_2, ..., x_n, ...\}$ has finite diameter.

Proof Pick N so that $d(x_m, x_n) < 1$ when $m, n \geq N$. Then $x_n \in B_1(x_N)$ for all $n \geq N$. Let $r = \max\{1, d(x_1, x_N), d(x_2, x_N), ..., d(x_{N-1}, x_N)\}$. Therefore, for all m, n we have $d(x_m, x_n) \leq d(x_n, x_N) + d(x_N, x_m) \leq 2r$, so diam $\{x_1, x_2, \dots, x_n, ...\} \leq 2r$. ∙

Now we will prove that $\mathbb{R}$ (with its usual metric d) is complete. The proof depends on the completeness property (also known as the "least upper bound property") of $\mathbb{R}$. We start with two lemmas which might be familiar from analysis. A sequence (x_n) is called weakly increasing if $x_n \leq x_{n+1}$ for all n, and weakly decreasing if $x_n \geq x_{n+1}$ for all n. A monotone sequence is one that is either weakly increasing or weakly decreasing.

Lemma 2.9 If (x_n) is a bounded monotone sequence in $\mathbb{R}$, then (x_n) converges.

Proof Suppose (x_n) is weakly increasing. Since (x_n) is bounded, $\{x_n : n \in \mathbb{N}\}$ has a least upper bound in $\mathbb{R}$. Let $x = \sup\{x_n : n \in \mathbb{N}\}$. We claim $(x_n) \to x$.

Let $\epsilon > 0$. Since $x - \epsilon < x$, we know that $x - \epsilon$ is not an upper bound for $\{x_n : n \in \mathbb{N}\}$. Therefore $x - \epsilon < x_N \leq x$ for some N. Since (x_n) is weakly increasing and x is an upper bound for $\{x_n : n \in \mathbb{N}\}$, it follows that $x - \epsilon < x_N \leq x_n \leq x$ for all $n \geq N$. Therefore $|x_n - x| \leq \epsilon$ for all $n \geq N$, so $(x_n) \to x$.

If (x_n) is weakly decreasing, then the sequence $(-x_n)$ is a weakly increasing, so $(-x_n) \to a$ for some a; then $(x_n) \to -a$. ∙

Lemma 2.10 Every sequence (x_n) in $\mathbb{R}$ has a monotone subsequence.

Proof Call x_k a peak point of (x_n) if $x_k \geq x_n$ for all $n \geq k$. We consider two cases.

i) If (x_n) has only finitely many peak points, then there is a last peak point x_{n_0} in (x_n). Then x_n is not a peak point for $n > n_0$. Pick an $n_1 > n_0$. Since x_{n_1} is not a peak point, there is an $n_2 > n_1$ with $x_{n_1} < x_{n_2}$. Since x_{n_2} is not a peak point, there is an $n_3 > n_2$ with $x_{n_2} < x_{n_3}$. We continue in this way to pick an increasing subsequence (x_{n_k}) of (x_n).

ii) If (x_n) has infinitely many peak points, then list the peak points as a subsequence $x_{n_1}, x_{n_2}, \ldots$ (where $n_1 < n_2 < \ldots$). Since each of these points is a peak point, we have $x_{n_k} \geq x_{n_{k+1}}$ for each k, so (x_{n_k}) is a weakly decreasing subsequence of (x_n). •

Theorem 2.11 $(\mathbb{R}, d)$ is complete.

Proof Let (x_n) be a Cauchy sequence in $(\mathbb{R}, d)$. By Theorem 2.8, (x_n) is bounded and by Lemma 2.10, (x_n) has a monotone subsequence (x_{n_k}) which is, of course, also bounded. By Lemma 2.9, (x_{n_k}) converges to some point $x \in \mathbb{R}$. Therefore x is a cluster point of the Cauchy sequence (x_n), so $(x_n) \to x$. •

Corollary 2.12 $(\mathbb{R}^n, d)$ is complete.

Proof Exercise
(*Hint: For* $n = 2$, *the sequence* $((x_n, y_n))$ *in* $\mathbb{R}^2$ *is Cauchy iff both sequences* (x_n) *and* (y_n) *are Cauchy sequences in* $\mathbb{R}$. *If* $(x_n) \to x$ *and* $(y_n) \to y$, *then* $((x_n, y_n) \to (x, y)$ *in* $\mathbb{R}^2$.)

3. Subspaces of Complete Spaces

Theorem 3.1 Let (X, d) be a metric space suppose $A \subseteq X$.

1) If (X, d) is complete and A is closed, then (A, d) is complete.
2) If (A, d) is complete, then A is closed in (X, d).

Proof 1) Let (a_n) be a Cauchy sequence in A. Then (a_n) is also Cauchy in the complete space (X, d), so $(a_n) \to x$ for some $x \in X$. But A is closed, so this limit x must be in A, that is, $(a_n) \to x \in A$. Therefore (A, d) is complete.

2) If $x \in \operatorname{cl} A$, we can pick a sequence in A such that $(a_n) \to x$. Since (a_n) converges, it is a Cauchy sequence in A. But (A, d) is complete, so we know $(a_n) \to a$ for some $a \in A \subseteq X$. Since limits of sequences are unique in metric spaces, we conclude that $x = a$. Therefore if $x \in \operatorname{cl} A$, then $x \in A$. So A is closed. •

Part 1) of the proof is valid when d is a pseudometric, but the proof of part 2) requires that d be a metric. Can you give an example of a pseudometric space (X, d) *for which part 2) of the theorem is false?*

The definition of completeness is stated in terms of the existence of limits for Cauchy sequences. The following theorem gives a different characterization – in terms of the existence of points in certain intersections. This illustrates an important idea: the completeness property for (X, d) can be expressed in different ways, but each characterization somehow asserts the existence of certain points.

Theorem 3.2 (The Cantor Intersection Theorem) The following are equivalent for a metric space (X, d)

1) (X, d) is complete
2) Whenever $F_1 \supseteq F_2 \supseteq ... \supseteq F_n \supseteq ...$ is a decreasing sequence of nonempty closed sets with diam$(F_n) \to 0$, then $\bigcap_{n=1}^{\infty} F_n = \{x\}$ for some $x \in X$.

Proof 1) $\Rightarrow$ 2) For each k, pick a point $x_k \in F_k$. Since diam$(F_k) \to 0$, the sequence (x_k) is Cauchy so $(x_k) \to x$ for some $x \in X$.

For each n, the subsequence $(x_{n+k}) = (x_{n+1}, x_{n+2}, ...) \to x$. Since the F_n's are decreasing, this subsequence is inside the closed set F_n, so $x \in F_n$. Therefore $x \in \bigcap_{n=1}^{\infty} F_n$.

If $x, y \in \bigcap_{n=1}^{\infty} F_n$, then $d(x, y) \leq$ diam(F_n) for every n. Since diam$(F_n) \to 0$, this means that $d(x, y) = 0$. Therefore $x = y$, and so $\bigcap_{n=1}^{\infty} F_n = \{x\}$.

2) $\Rightarrow$ 1) Suppose 1) is false, and let (x_n) be a nonconvergent Cauchy sequence (x_n). We will construct sets F_n which violate condition 2). Without loss of generality, we may assume that all the x_n's are distinct.

(*Why? If any value $x_k = y$ were repeated infinitely often, then y would be a cluster point of (x_n) and we would have $(x_n) \to y$. Therefore each term of (x_n) can occur only finitely often, and we can pick a subsequence from (x_n) whose terms are distinct. This subsequence is also a nonconvergent Cauchy sequence. We could now construct the F_n's using the terms of the subsequence. But to keep the notation a bit simpler, we might as well assume all the terms in the original sequence are distinct.*)

Let $F_n = \{x_n, x_{n+1}, ..., x_{n+k}, ...\} =$ the "n^{th} tail" of the sequence $(x_n) \neq \emptyset$. Then each $F_n \supseteq F_{n+1}$, and $\bigcap_{n=1}^{\infty} F_n = \emptyset$. For $\epsilon > 0$, we can pick N so that $d(x_m, x_n) < \frac{\epsilon}{2}$ if $m, n \geq N$. Then diam$(F_N) < \epsilon$ and for $n \geq N$, diam$(F_n) \leq$ diam$(F_N) < \epsilon$. So diam$(F_n) \to 0$.

We claim that these F_n's are closed – which will contradict 2) and complete the proof. If some particular F_{n_0} were not closed, then there would be a point $x \in \text{cl}\, F_{n_0} - F_{n_0}$ and we could find a sequence of distinct terms in F_{n_0} that converges to x. But such a sequence is automatically a subsequence of the original sequence (x_n), so x would be a cluster point of (x_n) – which is impossible since a nonconvergent Cauchy sequence (x_n) can't have a cluster point. •

How would Theorem 3.2 be different if d were only a pseudometric? How would the proof change?

Example 3.3

1) In the complete space $(\mathbb{R}, d)$ consider:

a) $F_n = [n, \infty)$
b) $F_n = (0, \frac{1}{n})$
c) $F_n = [n,\, n + \frac{1}{n}]$

In each case, $\bigcap_{n=1}^{\infty} F_n = \emptyset$. Do these examples contradict Theorem 3.2? Why not?

2) Let $F_n = \{x \in \mathbb{Q} : [\sqrt{2} - \frac{1}{n} \leq x \leq \sqrt{2} + \frac{1}{n}]\}$. The F_n's satisfy all the hypotheses in Theorem 3.2), but $\bigcap_{n=1}^{\infty} F_n = \emptyset$. Does this contradict Theorem 3.2?

Example 3.4 Here is a proof, using the Cantor Intersection Theorem, that the closed interval $[0,1]$ is uncountable.

> If $[0,1]$ were countable, we could list its elements in a sequence (x_n). Pick a subinterval $[a_1, b_1]$ of $[0,1]$ with length less than $\frac{1}{2}$ and excluding x_1. Then pick an interval $[a_2, b_2] \subseteq [a_1, b_1]$ of length less than $\frac{1}{4}$ and excluding x_2. Continuing inductively, choose an interval $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ of length less than $\frac{1}{2^{n+1}}$ and excluding x_{n+1}. Then $[a_1, b_1] \supseteq [a_2, b_2] \supseteq ... \supseteq [a_n, b_n] \supseteq ...$ and these sets clearly satisfy the conditions in part 2) of the Cantor Intersection Theorem. But $\bigcap_{n=1}^{\infty}[a_n, b_n] = \emptyset$. This is impossible since $[0,1]$ is complete.

The following theorem has some interesting consequences. In addition, the proof is a very nice application of Cantor's Intersection Theorem – in it, completeness is used to prove the existence of "very many" points in X.

Theorem 3.6 Suppose (X, d) is a nonempty complete metric space with no isolated points. Then $|X| \geq c$.

To prove Theorem 3.5, we will use the following lemma.

Lemma 3.5 Suppose (X, d) is a metric space that $A \subseteq X$. If $a, b \in \text{int}\, A$ and $a \neq b$, then we can find disjoint closed balls F_a and F_b (centered at a and b) with $F_a \subseteq \text{int}\, A$ and $F_b \subseteq \text{int}\, A$.

Proof Let $\epsilon = d(a,b)$. Since $a, b \in \text{int}\, A$, we can choose positive radii ϵ_a and ϵ_b so that $B_{\epsilon_a}(a) \subseteq \text{int}\, A$ and $B_{\epsilon_b}(b) \subseteq \text{int}\, A$. In addition, we can choose both ϵ_a and ϵ_b less than $\frac{\epsilon}{2}$.
Let $\epsilon' = \min\{\epsilon_a, \epsilon_b\}$. Then we can use the closed balls $F_a = \{x : d(x,a) \leq \epsilon'\}$ and $F_b = \{x : d(x,b) \leq \epsilon'\}$. •

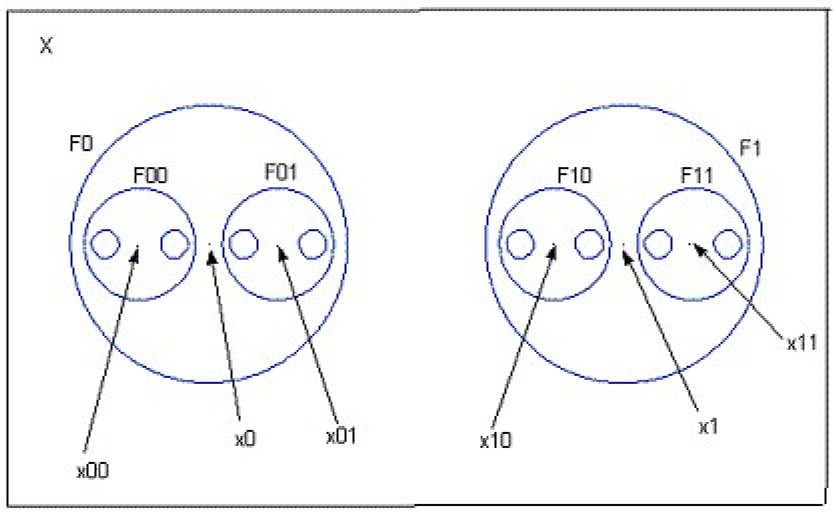
Proof of Theorem 3.6 The idea of the proof is to construct, inductively, c different descending sequences of closed sets, each of which satisfies condition 2) in the Cantor Intersection Theorem, and to do this in a way that the intersection of each sequence gives a <u>different</u> point in X. It then follows that $|X| \geq c$. The idea is simple but the notation gets a bit complicated. First we will give the idea of how the construction is done. The actual details of the induction step are relegated to the end of the proof.

<u>Stage 1</u> X is nonempty and has no isolated points, so there must exist two points $x_0 \neq x_1 \in X$. Pick disjoint closed balls F_0 and F_1, centered at x_0 and x_1, each with diameter $< \frac{1}{2}$.

<u>Stage 2</u> Since $x_0 \in \text{int}\, F_0$ and x_0 is not isolated, we can pick distinct points x_{00} and x_{01} (both $\neq x_0$) in $\text{int}\, F_0$ and use Lemma 3.5 to pick disjoint closed balls F_{00} and F_{01} (centered at x_{00} and x_{01}) and both $\subseteq \text{int}\, F_0$. We can then shrink the balls, if necessary, so that each has diameter $< \frac{1}{2^2}$.

We can repeat similar steps inside F_1 : since $x_1 \in \text{int}\, F_1$ and x_1 is not isolated, we can pick distinct points x_{10} and x_{11} (both $\neq x_1$) in $\text{int}\, F_1$ and use the Lemma 3.5 to pick disjoint closed balls F_{10} and F_{11} (centered at x_{10} and x_{11}) and both $\subseteq \text{int}\, F_1$. We can then shrink the balls, if necessary, so that each has diameter $< \frac{1}{2^2}$. At the end of Stage 2, we have 4 disjoint closed balls: F_{00}, F_{01}, F_{10}, F_{11}.

<u>Stage 3</u> We now repeat the same construction inside <u>each</u> of the 4 sets $F_{00}, F_{01}, F_{10}, F_{11}$. For example, we can pick distinct points x_{000} and x_{001} (both $\neq x_{00}$) in $\text{int}\, F_{00}$ and use the Lemma 3.5 pick disjoint closed balls F_{000} and F_{001} (centered at x_{000} and x_{001}) and both $\subseteq$ int F_{00}. Then we can shrink the balls, if necessary, so that each has diameter $< \frac{1}{2^3}$. *See the figure below.*



After Stage 3, we have 8 disjoint closed balls: $F_{000}, F_{001}, F_{010}$, … . We have the beginnings of 8 descending sequences of closed balls at this stage. In each sequence, at the n^{th} "stage," the sets have diameter $< \frac{1}{2^n}$.

$$\begin{tikzcd}[column sep=small, row sep=small]
 & & \supseteq F_{000} & \\
 & \supseteq F_{00} \arrow[ur] \arrow[dr] & & \\
 & & \supseteq F_{001} & \\
F_0 \arrow[uur] \arrow[ddr] & & & \text{.... etc.} \\
 & & \supseteq F_{010} & \\
 & \supseteq F_{01} \arrow[ur] \arrow[dr] & & \\
 & & \supseteq F_{011} &
\end{tikzcd}$$

and

$$\begin{tikzcd}[column sep=small, row sep=small]
 & & \supseteq F_{100} & \\
 & \supseteq F_{10} \arrow[ur] \arrow[dr] & & \\
 & & \supseteq F_{101} & \\
F_1 \arrow[uur] \arrow[ddr] & & & \text{.... etc.} \\
 & & \supseteq F_{110} & \\
 & \supseteq F_{11} \arrow[ur] \arrow[dr] & & \\
 & & \supseteq F_{111} &
\end{tikzcd}$$

We continue inductively (*see the details below*) in this way – at a given stage, each descending sequence of closed balls splits into two "disjoint branches." The "split" happens when we choose two new nonempty disjoint closed balls inside the current one, making sure that their diameters keep shrink toward 0.

In the end, <u>for each binary sequence</u> $s = (n_1, n_2, ..., n_k, ...) \in \{0,1\}^{\mathbb{N}}$ we will have a corresponding descending sequence of nonempty closed sets whose diameters $\rightarrow 0$:

$$F_{n_1} \supseteq F_{n_2 n_2} \supseteq F_{n_1 n_2 n_3} \supseteq ... \supseteq F_{n_1 n_2 ... n_k} \supseteq ...$$

For example, the binary sequence $s = (0, 1, 1, 0, 0, 1...) \in \{0,1\}^{\mathbb{N}}$ corresponds to the descending sequence of closed sets

$$F_0 \supseteq F_{01} \supseteq F_{011} \supseteq F_{0110} \supseteq F_{01100} \supseteq F_{011001} \supseteq ...$$

The Cantor Intersection theorem tells us <u>for each s</u>, there is an $x_s \in X$ such that

$$F_{n_1} \cap F_{n_2 n_2} \cap F_{n_1 n_2 n_3} \cap ... \cap F_{n_1 n_2 ... n_k} \cap ... = \bigcap_{k=1}^{\infty} F_{n_1 n_2 ... n_k} = \{x_s\}$$

For two different binary sequences, say $t = (m_1, m_2, ..., m_k, ...) \neq (n_1, n_2, ..., n_k, ...) = s$, there is a <u>smallest</u> k for which $m_k \neq n_k$. Then $x_s \in F_{n_1 n_2 ... n_{k-1} n_k}$ and $x_t \in F_{n_1 n_2 ... n_{k-1} m_k}$. Since these sets are disjoint, we have $x_s \neq x_t$. Thus, mapping $s \mapsto x_s$ gives a one-to-one function from $\{0,1\}^{\mathbb{N}}$ into X. We conclude that $c = 2^{\aleph_0} = |\{0,1\}|^{\mathbb{N}} \leq |X|$.

<u>Here are the details of the formal induction step in the proof.</u>.

Induction Hypothesis: Suppose we have completed k stages – that is, for each $i = 1, ..., k$ and for each i-tuple $(n_1, ..., n_i) \in \{0,1\}^i$ we have defined points $x_{n_1 ... n_i}$ and closed balls $F_{n_1 ... n_i}$ centered at $x_{n_1 ... n_i}$, with $\text{diam}(F_{n_1 ... n_i}) < \frac{1}{2^i}$ and so that

for each $(n_1, ..., n_i)$, $F_{n_1} \supseteq F_{n_1 n_2} \supseteq ... \supseteq F_{n_1 n_2 ... n_i}$

Induction step: We must construct the sets for stage $k+1$. For each $(k+1)$-tuple $(n_1, ..., n_k, n_{k+1}) \in \{0,1\}^{k+1}$, we need to define a point $x_{n_1 ... n_k n_{k+1}}$ and a closed ball $F_{n_1 ... n_k n_{k+1}}$ in such a way that the conditions in the induction hypothesis remain true with $k+1$ replacing k.

For any $(n_1, ..., n_k)$: we have $x_{n_1,,n_k} \in \text{int}\, F_{n_1 ... n_k}$. Since $x_{n_1,,n_k}$ is not isolated, we can pick distinct points $x_{n_1 ... n_k 0}$ and $x_{n_1 ... n_k 1}$ (both $\neq x_{n_1 ... n_k}$) in int $F_{n_1 ... n_k}$ and use the Lemma 3.5 to pick disjoint closed balls $F_{n_1 ... n_k 0}$ and $F_{n_1 ... n_k 1}$ (centered at $x_{n_1,,n_k 0}$ and $x_{n_1 ... n_k 1}$) and both $\subseteq$ int $F_{n_1 ... n_k}$. We can then shrink the balls, if necessary, so that each has diameter $< \frac{1}{2^{k+1}}$. •

Corollary 3.7 If (X, d) is a nonempty complete separable metric space with no isolated points, then $|X| = c$.

Proof Theorem II.5.21 (using separability) tells us that $|X| \leq c$; Theorem 3.6 gives us $|X| \geq c$. •

The following corollary gives us another variation on the basic result.

Corollary 3.8 If (X, d) is an uncountable complete separable metric space, then $|X| = c$.
(*So we might say that "the Continuum Hypothesis holds among complete separable metric spaces."*)

Proof Since (X, d) is separable, Theorem II.5.21 gives us $|X| \leq c$. If we knew that X had no isolated points, then Theorem 3.6 (or Corollary 3.7) would complete the proof. But X might have isolated points, so a little more work is needed.

Call a point $x \in X$ a condensation point if every neighborhood of x is uncountable. Let C be the set of all condensation points in X. Each point $a \in X - C$ has a countable open neighborhood O. Each point of O is also a non-condensation point, so $a \in O \subseteq X - C$. Therefore $X - C$ is open so C is closed, and therefore (C, d) is a complete metric space.

Since (X, d) is separable (and therefore second countable), $X - C$ is also second countable and therefore Lindelöf. Since $X - C$ can be covered by countable open sets, it can be covered by countably many of them, so $X - C$ is countable: therefore $C \neq \emptyset$ (in fact, C must be uncountable).

Finally, (C, d) has no isolated points in C: if $b \in C$ were isolated in C, then there would be an open set O in X with $O \cap C = \{b\}$. Since $O - \{b\} \subseteq X - C$, O would be countable $-$ which is impossible since b is a condensation point.

Therefore Corollary 3.7 therefore applies to (C, d), and therefore $|X| \geq |C| \geq c$. ●

Why was the idea of a "condensation point" introduced in the example? Will the argument work if, throughout, we replace "condensation point" with "non-isolated point?" If not, precisely where would the proof break down?

The next corollary answers a question we raised earlier: is there a metric d' on $\mathbb{Q}$ which is equivalent to the usual metric d but for which $(\mathbb{Q}, d')$ is complete – that is, is $\mathbb{Q}$ "completely metrizable?"

Corollary 3.9 $\mathbb{Q}$ is not completely metrizable.

Proof Suppose d' is a metric on $\mathbb{Q}$ equivalent to the usual metric d, so that $\mathcal{T}_d = \mathcal{T}_{d'}$. Then $(\mathbb{Q}, d')$ is a nonempty metric space and, since $d \sim d'$, the space $(\mathbb{Q}, d')$ has no isolated points (no set $\{q\}$ is in $\mathcal{T}_d = \mathcal{T}_{d'}$: "isolated point" is a topological notion.). If $(\mathbb{Q}, d')$ were complete, Theorem 3.6 would imply that $\mathbb{Q}$ is uncountable. ●

Exercises

E1. Prove that in a metric space (X, d), the following statements are equivalent:

a) every Cauchy sequence is eventually constant
b) (X, d) is complete and $\mathcal{T}_d$ is the discrete topology
c) every subspace of (X, d) is complete

E2. Suppose that X is a dense subspace of the pseudometric space (Y, d) and that every Cauchy sequence in X converges to some point in Y. Prove that (Y, d) is complete.

E3. Suppose that (X, d) is a metric space and that (x_n) is a Cauchy sequence with only finitely many distinct terms. Prove that (x_n) is eventually constant, i.e., that for some $n \in \mathbb{N}$, $x_n = x_{n+1} = \ldots$

E4. Let p be a fixed prime number. The <u>p-adic norm</u> $|\ |_p$ is defined on $\mathbb{Q}$ by:

For $0 \neq x \in \mathbb{Q}$, write $x = \frac{p^k m}{n}$ for integers k, m, n, where p does not divide m or n. (*Of course, k may be negative.*) Define $|x|_p = p^{-k} = \frac{1}{p^k}$. . We define $|0|_p = 0$.

For all $x, y \in \mathbb{Q}$,

a) $|x|_p \geq 0$ and $|x|_p = 0$ iff $x = 0$
b) $|xy|_p = |x|_p \cdot |y|_p$
c) $|x + y|_p \leq |x|_p + |y|_p$
d) $|x + y|_p \leq \max\{|x|_p, |y|_p\}$

$|\ |_p$ is shares properties a), b) and c) with the ordinary norm (or absolute value), $|\ |$ on $\mathbb{Q}$. But d) is a condition stronger than the usual triangle inequality: that is, d) $\Rightarrow$ c)

The <u>p-adic metric</u> d_p is defined on $\mathbb{Q}$ by $d_p(x, y) = |x - y|_p$

Because of d), the metric d_p satisfies a strengthened version of the triangle inequality for metric spaces: what is this stronger inequality ? For $x, y, z \in \mathbb{Q}$, $d_p(x, z)$

<u>Prove or disprove</u> : $(\mathbb{Q}, d_p)$ is complete.

E5. Suppose (X, d) is complete and that $F_1 \supseteq F_2 \supseteq ... \supseteq F_n \supseteq ...$ is a sequence of closed sets in X for which diam$(F_n) \to 0$. Prove that if Y is a Hausdorff space and $f : X \to Y$ is continuous, $\bigcap_{n=1}^{\infty} f[F_n] = f[\bigcap_{n=1}^{\infty} F_n]$.

E6. A metric space (X, d) is called <u>locally complete</u> if every point x has a neighborhood N_x (necessarily closed) which is complete.

a) Give an example of a metric space (X, d) that is locally complete but not complete.

b) Prove that if (D, d) is a locally complete dense subspace the complete metric space (X, d), then D is open in X.

Hint: it may be helpful to notice if O is open and D is dense, then cl (O ∩ D) = cl(O). This is true in any topological space $(X, \mathcal{T})$.

E7. Suppose A is an uncountable subset of $\mathbb{R}$ with $|A| = m < c$. (*Of course, there is no such set A if the Continuum Hypothesis is assumed.*) Is it possible for A to be closed? Explain.

4. The Contraction Mapping Theorem

Definition 4.1 A point $x \in X$ is called a fixed point of the function $f : X \to X$ if $f(x) = x$. We say that a topological space X has the fixed point property if every continuous $f : X \to X$ has a fixed point.

The fixed point property is a topological property (*this is easy to check; do it!*)

Example 4.2

1) A function $f : \mathbb{R} \to \mathbb{R}$ has a fixed point if and only if the graph of f intersects the line $y = x$.

If $A \subseteq \mathbb{R}$, then A is the set of fixed points for some function: for example $f(x) = x \cdot \chi_A(x)$, where χ_A is the characteristic function of A. *Is every set $A \subseteq \mathbb{R}$ the set of fixed points for some continuous function $f : \mathbb{R} \to \mathbb{R}$? Are there any restrictions on the cardinality of A?*

2) The interval $[a, b]$ has the fixed point property: a continuous $f : [a, b] \to [a, b]$ must have a fixed point. Certainly this is true if $f(a) = a$ or $f(b) = b$. So assume that $f(a) > a$ and $f(b) < b$ and define $g(x) = f(x) - x$. Then g is continuous, $g(a) = f(a) - a > 0$ and $g(b) = f(b) - b < 0$. By the Intermediate Value Theorem (from analysis) there must be a point $c \in (a, b)$ where $g(c) = f(c) - c = 0$, that is, $f(c) = c$.

3) The Brouwer Fixed Point Theorem states that $D^n = \{x \in \mathbb{R}^n : d(x, 0) \leq 1\}$ has the fixed point property. For $n > 1$, the proof of Brouwer's theorem is rather difficult; the usual proofs use techniques from algebraic topology. For $n = 1$, D^n is homeomorphic to $[a, b]$, so Brouwer's Theorem generalizes part 2) of this example.

In fact, Brouwer's Theorem can be generalized as the Schauder Fixed Point Theorem: every nonempty compact convex subset K of a Banach space has the fixed point property.

Definition 4.3 $f : (X, d) \to (X, d)$ is a contraction mapping (for short, a contraction) if there is a constant $\alpha \in (0, 1)$ such that $d(f(x), f(y)) \leq \alpha d(x, y)$ for all $x, y \in X$.

Notice that a contraction f is automatically continuous: for $\epsilon > 0$, choose $\delta = \epsilon$. Then, for all $x \in X$, $d(x, y) < \delta$ implies $d(f(x), f(y)) < \alpha\delta < \epsilon$. In fact, this choice of δ depends only on ϵ and not the point x. When that is true, we say that f is uniformly continuous.

The definition of uniform continuity for f on X reads:

$$\forall \epsilon \ \boldsymbol{\exists \delta} \ \boldsymbol{\forall x} \ \forall y \ \ (d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon)$$

Compare this to the weaker requirement for f to be continuous on X:

$$\forall \epsilon \ \boldsymbol{\forall x} \, \boldsymbol{\exists \delta} \ \forall y \ \ (d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \epsilon)$$

Example 4.4

1) The function $f:[0,\frac{1}{4}] \to [0,\frac{1}{4}]$ given by $f(x)=x^2$ is a contraction because for any $x,y \in [0,\frac{1}{4}]$, we have $|f(x)-f(y)| = |x^2-y^2| = |x+y||x-y| \le \frac{1}{2}|x-y|$.

2) However, the function $f:\mathbb{R} \to \mathbb{R}$ given by $f(x)=x^2$ is not a contraction. For f to be a contraction, there would have to be some $\alpha \in (0,1)$ for which

$$\begin{array}{lll} |x^2-y^2| \le \alpha|x-y| & \text{for all } x,y \text{ in } \mathbb{R}\text{, and this would imply that} \\ |x+y| \quad \le \alpha & \text{for all } x \ne y \text{ (which is false).} \end{array}$$

We could also argue that f is not a contraction by noticing that $f:\mathbb{R}\to\mathbb{R}$ is not uniformly continuous. For example, if we choose $\epsilon = 2$, then no $\delta > 0$ will satisfy the condition $\forall x\ \forall y\ \ (d(x,y)<\delta \Rightarrow d(f(x),f(y)) < 2)$. For example, given $\delta > 0$, we can let $x=\frac{2}{\delta}$ and $y=\frac{2}{\delta}+\frac{\delta}{2}$. Then $|f(x)-f(y)| = 2+\frac{\delta^2}{4} > 2$.

Theorem 4.5 (The Contraction Mapping Theorem) If (X,d) is a nonempty complete metric space and $f: X \to X$ is a contraction, then f has a unique fixed point. (*"Nonempty" is included in the hypothesis because the empty function $\emptyset$ from the complete space $\emptyset$ to $\emptyset$ is a contraction with no fixed point.*)

A fixed point provided by the Contraction Mapping Theorem can be useful, as we will soon see in Theorem 4.9 (Picard's Theorem). The proof of the Contraction Mapping Theorem is also useful because it shows how to make some handy numerical estimates (*see the example following Picard's Theorem.*)

Proof Suppose $0<\alpha<1$ and that $d(f(x),f(y)) \le \alpha d(x,y)$ for all $x,y\in X$. Pick any point $x_0 \in X$ and apply the function f repeatedly to define

$$\begin{aligned} x_1 &= f(x_0), \\ x_2 &= f(x_1) = f(f(x_0)) = f^2(x_0) \\ &\vdots \\ x_n &= f(x_{n-1}) = \ldots = f^n(x_0) \\ &\vdots \end{aligned}$$

We claim that this sequence (x_n) is Cauchy and that its limit in X is a fixed point for f. (*Here you can imagine a "control system" where the initial input is x_0 and each output becomes the new input in a "feedback loop." This system approaches a "steady state" where "input = output."*)

Suppose $\epsilon > 0$. We want to show that $d(x_n,x_m) < \epsilon$ for large enough m,n. Assume $m>n$. Applying the "contraction property" n times gives:

$$\begin{aligned} d(x_n,x_m) &= d(f^n(x_0), f^m(x_0)) = d(f(f^{n-1}(x_0)), f(f^{m-1}(x_0))) \\ &\le \alpha\, d(f^{n-1}(x_0), f^{m-1}(x_0)) \le \alpha^2\, d(f^{n-2}(x_0), f^{m-2}(x_0)) \\ &\le \ \ldots\ \le\ \alpha^n d(x_0, f^{m-n}(x_0)) = \alpha^n\, d(x_0, x_{m-n}) \end{aligned}$$

Using the triangle inequality and the contraction property again, we get

$$d(x_n, x_m) \leq \alpha^n \, d(x_0, x_{m-n})$$

$$\leq \alpha^n \left(d(x_0, x_1) + d(x_1, x_2) + d(x_2, x_3) + \ldots + d(x_{m-n-1}, x_{m-n})\right)$$

$$\leq \alpha^n \left(d(x_0, x_1) + \alpha d(x_0, x_1) + \alpha^2 d(x_0, x_1) + \ldots + \alpha^{m-n-1} d(x_0, x_1)\right)$$

$$= \alpha^n \, d(x_0, x_1) \left(1 + \alpha + \alpha^2 + \ldots + \alpha^{m-n-1}\right) < \alpha^n \, d(x_0, x_1) \textstyle\sum_{k=0}^{\infty} \alpha^k$$

$$= \frac{\alpha^n d(x_0, x_1)}{1-\alpha} \to 0 \text{ as } n \to \infty.$$

So $\frac{\alpha^n d(x_0,x_1)}{1-\alpha} < \epsilon$ if $n >$ some N. Then $d(x_n, x_m) < \epsilon$ if $m > n > N$. Therefore (x_n) is Cauchy. Since (X, d) is complete $(x_n) \to x$ for some $x \in X$.

By continuity, $(f(x_n)) \to f(x)$. But also $(f(x_n)) \to x$ because $f(x_n) = x_{n+1}$. Because a sequence in a metric space can have at most one limit. we get that $f(x) = x$. (*If d were just a pseudometric, x might not actually be a fixed point: we could only say that* $d(x, f(x)) = 0$.)

If y is also a fixed point for f, then $d(x, y) = d(f(x), f(y)) \leq \alpha \, d(x, y)$. Since $0 < \alpha < 1$, this implies that $d(x, y) = 0$, so $x = y$ – that is, the fixed point x is unique. (*If d were just a pseudometric, f could have several fixed points at distance* 0 *from each other.*) •

Notes about the proof

1) The proof gives us that if $m > n$, then $d(x_n, x_m) < \frac{\alpha^n d(x_0,x_1)}{1-\alpha}$ for each $n \in \mathbb{N}$. If we fix n and let $m \to \infty$, Then $(x_m) \to x$, so $d(x_n, x) \leq \frac{\alpha^n d(x_0,x_1)}{1-\alpha}$. Thus we have a computable bound on how well x_n approximates the fixed point x.

2) For large n we can think of x_n as an "approximate fixed point" – in the sense that f doesn't move the point x_n very much. To be more precise,

$$d(x_n, f(x_n)) \leq d(x_n, x) + d(x, f(x_n)) = d(x_n, x) + d(f(x), f(x_n))$$

$$\leq d(x_n, x) + \alpha \, d(x, x_n) = (1+\alpha) \, d(x_n, x)$$

$$\leq (1+\alpha) \frac{\alpha^n d(x_0,x_1)}{1-\alpha} \to 0 \text{ as } n \to \infty.$$

Example 4.6 Consider the following functions that map the complete space $\mathbb{R}$ into itself:

1) $f(x) = 2x + 1$: f is (uniformly) continuous, f is not a contraction, and f has a unique fixed point: $x = -1$
2) $g(x) = x^2 + 1$ is not a contraction and has no fixed point.
3) $h(x) = x + \sin x$ has infinitely many fixed points.

We are going to use the Contraction Mapping Theorem to prove a fundamental theorem about the existence and uniqueness of a solution to a certain kind of initial value problem in differential equations. We will use the Contraction Mapping Theorem applied to the complete space in the following example.

Example 4.7 Let X be a topological space and $C^*(X)$ be the set of bounded continuous real-valued functions $f\colon X \to \mathbb{R}$, with the metric $\rho(f,g) = \sup\{|f(x) - g(x)|\colon x \in X\}$. (*Since f, g are bounded, so is $|f - g|$ and therefore this sup is always a real number. It is easy to check that ρ is a metric.*) The metric ρ is called the "metric of uniform convergence" because $\rho(f,g) < \epsilon$ implies that f is "uniformly close" to g :

$$\rho(f,g) < \epsilon \qquad \Rightarrow \quad |f(x) - g(x)| < \epsilon \text{ at every point } x \in X \quad (*)$$

In the specific case of $(C^*(\mathbb{R}), \rho)$, the statement that $(f_n) \xrightarrow{\rho} f$ is equivalent the statement (in analysis) that "(f_n) converges to f uniformly on $\mathbb{R}$."

Theorem 4.8 $(C^*(X), \rho)$ is complete.

Proof Suppose that (f_n) is a Cauchy sequence in $(C^*(X), \rho)$: given $\epsilon > 0$, there is some N so that $\rho(f_n, f_m) < \epsilon$ whenever $n, m > N$. This implies (*see* (*), *above*) that for each $x \in X$, $|f_n(x) - f_m(x)| < \epsilon$ whenever $n, m > N$. For any fixed x, then, $(f_n(x))$ is a Cauchy sequence of real numbers, so $(f_n(x)) \to$ some $r_x \in \mathbb{R}$. Define $f : X \to \mathbb{R}$ by $f(x) = r_x$. To complete the proof, we claim that $f \in C^*(X)$ and that $(f_n) \xrightarrow{\rho} f$

First we argue that f is bounded. Pick N so that $\rho(f_n, f_m) < 1$ whenever $n, m \geq N$.
Using $m = N$, this gives $\rho(f_n, f_N) < 1$ for $n \geq N$ and therefore $|f_n(x) - f_N(x)| < 1$ for every $x \in X$. Now let $n \to \infty$, to get that

for every x $\qquad |f(x) - f_N(x)| \leq 1,$

so, for every x $\quad -1 \leq f(x) - f_N(x) \leq 1.$

Since f_N is bounded, there is a constant M such that

$$-M \leq f_N(x) \leq M \qquad \text{for every } x.$$

Adding the last two inequalities shows that f is bounded because, for every x

$$|f(x)| \leq M + 1.$$

We now claim that $(f_n) \xrightarrow{\rho} f$. (*Technically, this is an abuse of notation – because ρ is defined on $C^*(X)$ and we don't yet know that f is in $C^*(X)$! However, the same definition for ρ makes sense on any collection of bounded real-valued functions, continuous or not. We are using this "extended definition" of ρ here.*) Let $\epsilon > 0$ and pick N so that $\rho(f_n, f_m) < \frac{\epsilon}{2}$ whenever $m, n \geq N$. Then $|f_n(x) - f_m(x)| < \frac{\epsilon}{2}$ for every x whenever $n, m \geq N$. Letting $m \to \infty$, we get that $|f_n(x) - f(x)| \leq \frac{\epsilon}{2}$ for all x and all $n \geq N$. Therefore $\rho(f_n, f) < \epsilon$ if $n \geq N$.

Finally, we show that f is continuous at every point $a \in X$. Let $\epsilon > 0$. For $x \in X$ and $N \in \mathbb{N}$ we have

$$|f(x) - f(a)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(a)| + |f_N(a) - f(a)|$$

Since $(f_n) \xrightarrow{\rho} f$, we can choose N large enough to make $\rho(f_N, f) < \frac{\epsilon}{3}$. This implies that $|f(x) - f_N(x)| < \frac{\epsilon}{3}$ and $|f_N(a) - f(a)| < \frac{\epsilon}{3}$. Since f_N is continuous at a, we can choose a neighborhood W of a so that $|f_N(x) - f_N(a)| < \frac{\epsilon}{3}$ for all $x \in W$. Then for $x \in W$ we have

$|f(x) - f(a)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$, so f is continuous at a. •

Theorem 4.9 (Picard's Theorem) Let (x_0, y_0) be an interior point of a closed box D in $\mathbb{R}^2$, and suppose that $f : D \to \mathbb{R}$ is continuous. In addition, suppose that there is a constant M such that for all (x_1, y_1) and (x_1, y_2) in D:

$$|f(x_1, y_1) - f(x_1, y_2)| \leq M|y_1 - y_2| \quad \text{(L)}$$

Then

i) there exists an interval $I = [x_0 - a, x_0 + a] = \{x : |x - x_0| \leq a\}$, and
ii) there exists a unique differentiable function $g : I \to \mathbb{R}$

such that
$$\begin{cases} y_0 = g(x_0) \\ g'(x) = f(x, g(x)) \end{cases} \quad \text{for } x \in I$$

In other words, on some interval I centered at x_0, there is a unique solution $y = g(x)$ to the initial value problem

$$\begin{cases} y' = f(x, y) \\ y(x_0) = y_0 \end{cases} \qquad (*)$$

Before beginning the proof, we want to make some comments about the hypotheses.

1) The "strange" condition (L) says that "f satisfies a Lipschitz condition in the variable y on D." For our purposes, it's enough to notice that (L) will be true when the partial derivative f_y exists and is continuous on the closed box D. In that case, we know (from analysis) that $|f_y| \leq$ some constant M on D. Then, if we think of $f(x_1, y)$ as a single-variable function of y and use the ordinary Mean Value Theorem, we get a point z between y_1 and y_2 for which

$$|f(x_1, y_1) - f(x_1, y_2)| = |f_y(x_1, z)| \cdot |y_1 - y_2| \leq M|y_1 - y_2|$$

2) Consider a specific example: $f(x, y) = y - x$, $D = [-1, 1] \times [-1, 1]$, and $(x_0, y_0) = (0, 0)$. Since $f_y = 1$, the preceding comment tells us that condition (L) is true.

Consider the initial value problem $\begin{cases} y' = y - x \\ y(0) = 0 \end{cases}$. Picard's Theorem states that there is a unique differentiable function $y = g(x)$, defined on some interval $[-a, a]$ that solves this system: $g(0) = 0$ and $g'(x) = f(x, g(x))$ (that is, $y' = y - x$). As we will see, the proofs of Picard's Theorem and the Contraction Mapping Theorem can actually help us to find this solution. Of course, once a solution is actually found, a direct check might show that the solution is actually valid on an interval larger than the interval I that comes up in the proof.

Proof (Picard's Theorem) We begin by changing the initial value problem $(*)$ into an equivalent problem that involves an integral equation $(**)$ rather than a differential equation. Let $y = g(x)$ be a continuous function defined on an interval $I = [x_0 - a, x_0 + a]$:

Suppose g satisfies (*): f is continuous, so $g'(x) = f(x, g(x))$ is continuous. (*Why?*) Therefore the Fundamental Theorem of Calculus tells us for all $x \in I$,

$$g(x) - g(x_0) = \int_{x_0}^{x} g'(t)\,dt = \int_{x_0}^{x} f(t, g(t))\,dt, \text{ so}$$

$$g(x) = y_0 + \int_{x_0}^{x} f(t, g(t))\,dt \qquad (**)$$

Suppose g satisfies (**): then $g(x_0) = y_0 + \int_{x_0}^{x_0} f(t, g(t))\,dt = y_0$. Since $f(t, g(t))$ is continuous, the Fundamental Theorem gives that for all $x \in I$, $g'(x) = f(x, g(x))$. Therefore the function $g(x)$ satisfies (*).

Therefore a function continuous g on I satisfies (*) iff g satisfies (**). (*Note that the initial condition* $y_0 = g(x_0)$ *is "built into" the single equation* $(**)$.) Now we show that a unique solution g to (**) exists on some interval I centered at x_0.

First, we establish some notation:

We are given a constant M in the Lipschitz condition (L).

Pick a constant K so that $|f(x, y)| \le K$ for all $(x, y) \in D$. (*We will prove later in this chapter that a continuous real-valued function on a closed box in* $\mathbb{R}^2$ *must be bounded. For now, we assume that fact from analysis.)*

Because $(x_0, y_0) \in \operatorname{int} D$, we can pick a constant $a > 0$ so that

i) $\{x : |x - x_0| \le a\} \times \{y : |y - y_0| \le Ka\} \subseteq D$, and
ii) $aM < 1$

Let $I = \{x : |x - x_0| \le a\} = [x_0 - a, x_0 + a]$.

We consider $(C^*(I), \rho)$, where ρ is the metric of uniform convergence. By Theorem 4.8, this space is complete. Let $B = \{g \in C^*(I) : \forall t \in I,\ |g(t) - y_0| \le Ka\}$.

For an arbitrary $g \in C^*(I)$, $g(t)$ might be so large that $(t, g(t)) \notin D$. But if we look only at the functions in B, we avoid this problem: if $g \in B$ and $t \in I$, then $(t, g(t)) \in D$ because of how we chose a. So for $g \in B$, we know that $f(t, g(t))$ is defined for all $t \in I$.)

Notice that $B \neq \emptyset$ since the constant function $g(x) = y_0$ is certainly in B.

$\underline{B \text{ is closed}}$ in $C^*(I)$: if $h \in \text{cl}\, B$, then there is sequence of functions (g_n) in B such that $(g_n) \to h$ in the metric ρ. Therefore $g_n(x) \to h(x)$ for all $x \in I$, and subtracting y_0 gives

$$|g_n(x) - y_0| \to |h(x) - y_0|.$$

Since $|g_n(x) - y_0| \le Ka$ for each $x \in I$, then $|h(x) - y_0| \le Ka$ for each $x \in I$. Therefore $h \in B$, so B is closed. So (B, ρ) is a <u>nonempty complete metric space</u>.

For $g \in B$, define on I a function $h = T(g)$ using the formula

$$h(x) = T(g)(x) = y_0 + \int_{x_0}^{x} f(t, g(t))\, dt.$$

Notice that h is continuous (in fact, differentiable) and that, for every $x \in I$,

$$|h(x)| \le |y_0| + |\int_{x_0}^{x} f(t, g(t))\, dt\,| \le |y_0| + \int_{x_0}^{x} |f(t, g(t))|\, dt \le |y_0| + \int_{x_0}^{x} K\, dt \le |y_0| + Ka,$$

so h is bounded. Therefore $T : B \to C^*(I)$. But in fact, even <u>more</u> is true: $T\colon B \to B$, because

$$\forall x \in I,\ |T(g)(x) - y_0| = |\,h(x) - y_0| = |\int_{x_0}^{x} f(t, g(t))\, dt\,| \le Ka$$

Now we claim that $T : (B, \rho) \to (B, \rho)$ is a contraction. To see this, we simply compute distances: if g_1 and $g_2 \in B$, then

$$\begin{aligned}
\rho(T(g_1), T(g_2)| &= \text{sup}\{\,|\,T(g_1)(x) - T(g_2)(x)|\ : x \in I\} \\
&= \text{sup}\{\,|\int_{x_0}^{x} f(t, g_1(t)) - f(t, g_2(t))\, dt|\ : x \in I\} \\
&\le \text{sup}\{\,|\int_{x_0}^{x} |f(t, g_1(t)) - f(t, g_2(t))|\ dt| : x \in I\} \\
&\le \text{sup}\{\,|\int_{x_0}^{x} M\, |g_1(t) - g_2(t)|\, dt\,| : x \in I\} \qquad \text{(from Condition L)} \\
&\le \text{sup}\{\,|\int_{x_0}^{x} M\rho(g_1, g_2)\, dt\,| : x \in I\} \\
&= \text{sup}\{\, M\rho(g_1, g_2)\ |\ x - x_0| : x \in I\} \\
&= aM\rho(g_1, g_2) \\
&= \alpha\rho(g_1, g_2), \text{ where } \alpha = aM < 1.
\end{aligned}$$

Then the Contraction Mapping Theorem gives us a unique function $g \in B$ for which $g = T(g)$. From the definition of T, that simply means:

$$\forall x \in I,\ \ g(x) = T(g)(x) = y_0 + \int_{x_0}^{x} f(t, g(t))dt$$

which is precisely condition (**). •

Example 4.10 If we combine the method in the proof of Picard's Theorem with the numerical estimates in the proof of the Contraction Mapping Theorem, we can get useful information about a specific initial value problem. To illustrate, we consider

$$\begin{cases} y' &= y - x \\ y(0) = 0 \end{cases}$$

and find a solution that is valid on some interval containing 0.

We begin by choosing a box D with $(x_0, y_0) = (0,0)$ in its interior: we select (rather arbitrarily) the box $D = [-1,1] \times [-1,1]$. Since $|f(x,y)| = |y - x| \leq |x| + |y| \leq 2$ on D, we can use $K = 2$ in the proof. Because $|f_y(x,y)| = 1$ throughout D, the Lipschitz condition (L) is satisfied with $M = 1$.

Following the proof of Picard's Theorem, we now choose a constant a so that

i) $\{x : |x| \leq a\} \times \{y : |y| \leq Ka = 2a\} \subseteq D$ and
ii) $aM = a \cdot 1 < 1$

Again rather arbitrarily, we choose $a = \frac{1}{2}$, so that $I = [x_0 - a, x_0 + a] = [-\frac{1}{2}, \frac{1}{2}]$.
Then $B = \{g \in C^*(I) : |g(x) - 0| \leq Ka\} = \{g \in C^*(I) : |g(x)| \leq 1$ for all $x \in [-\frac{1}{2}, \frac{1}{2}]\}$.

Finally, we choose <u>any</u> function $g_0 \in B$; to make things as simple as possible, we might as well choose g_0 to be the constant function $g_0(x) = 0$ on $[-\frac{1}{2}, \frac{1}{2}]$. (*If for simplicity we use a constant* g_0, *then* $g_0 = 0$ *is the "best possible" choice since its graph goes through the point* $(x_0, y_0) = (0,0)$.)

According to the proof of the Contraction Mapping Theorem, the sequence of functions $(g_n) = (T^n(g_0))$ will converge (with respect to the metric ρ) to g, where g is a fixed point for T; g will be the solution to our initial value problem. We calculate:

$$g_1(x) = y_0 + \int_{x_0}^{x} f(t, g_0(t))\, dt = 0 + \int_0^x f(t,0)\, dt = \int_0^x -t\, dt = -\frac{x^2}{2}$$

$$g_2(x) = T(g_1(x)) = 0 + \int_0^x f(t, -\tfrac{t^2}{2})\, dt = \int_0^x -\tfrac{t^2}{2} - t\, dt = -\frac{x^3}{6} - \frac{x^2}{2}$$

$$g_3(x) = T(g_2(x)) = \int_0^x f(t, -\tfrac{t^3}{6} - \tfrac{t^2}{2})\, dt = \int_0^x -\tfrac{t^3}{6} - \tfrac{t^2}{2} - t\, dt = -\frac{x^4}{24} - \frac{x^3}{6} - \frac{x^2}{2}$$

and, in general,

$$g_n(x) = T(g_{n-1}(x)) = \dots = -\frac{x^2}{2!} - \frac{x^3}{3!} - \dots - \frac{x^{n+1}}{(n+1)!}$$

The functions g_n converge uniformly to the solution g.

In this particular problem, moreover, we are lucky enough to recognize that the functions $g_n(x)$ are just the partial sums of the series $\sum_{n=2}^{\infty} -\frac{x^n}{n!} = 1 + x - \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x - e^x$. Therefore $g(x) = 1 + x - e^x$ is a solution of our initial value problem, and we know it is valid for all $x \in I = [-\frac{1}{2}, \frac{1}{2}]$. (*You can check by substitution that the solution is correct – <u>and</u>, after the fact, that the solution actually is valid for x's in the much larger interval* $\mathbb{R}$.)

Even if we couldn't recognize a neat formula for the limit $g(x)$, we could still make some useful approximations. From the proof of the Contraction Mapping Theorem, we know that $\rho(g_n, g) \leq \frac{\alpha^n \rho(g_0, g_1)}{1 - \alpha}$. In this example $\alpha = aM = \frac{1}{2}$, so that $\rho(g_n, g) \leq \frac{\frac{1}{2^n}\rho(g_0, g_1)}{1 - \frac{1}{2}}$
$= \frac{1}{2^{n-1}} \sup\{|0 - (-\frac{x^2}{2})| : x \in I\} = \frac{1}{2^{n-1}} \cdot \frac{1}{2^3} = \frac{1}{2^{n+2}}$. Therefore, on the interval $[-\frac{1}{2}, \frac{1}{2}]$, $g_n(x)$ is <u>uniformly</u> within distance $\frac{1}{2^{n+2}}$ of the exact solution $g(x)$

Finally, recall that our initial choice of $g_0 \in B$ was arbitrary. Since $|\sin x| \leq 1$, we know that $\sin \in B$, and we could just as well have chosen $g_0(x) = \sin x$. Then the functions $g_n(x)$ computed as $T(g_0) = g_1$, $T(g_2) = g_2$, ... would be quite different (try $computing$ g_1 *and* g_2) but it would still be true that $g_n(x) \to g(x) = 1 + x - e^x$ uniformly on I : this must be same limit g because the solution g is unique. •

The Contraction Mapping Theorem can be used to prove other results $-$ for example, the Implicit Function Theorem. (*You can see details in Topology, by James Dugundji.*)

Exercises

E8. $a)$ Suppose $f : \mathbb{R} \to \mathbb{R}$ is differentiable and that there is a constant $K < 1$ such that $|f'(x)| \leq K$ for all x. Prove that f is a contraction (and therefore has a unique fixed point.)

b) Give an example of a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that

$$|f(x) - f(y)| < |x - y| \text{ for all } x \neq y \quad (**)$$

for all $x \neq y \in \mathbb{R}$ but such that f has no fixed point.

Note: The function f is not a contraction mapping. If we allow $\alpha = 1$ in the definition of contraction, your example shows that Contraction Mapping Theorem may not be true – not even if (as above) we "compensate" by using " $<$ " instead of " $\leq$ " in the definition.

c) Find an example satisfying (**) where $f : [0,1] \to [0,1]$ and f is not a contraction (*so compactness + (**) doesn't force f to be a contraction*)

d) (Edelstein's Theorem) Show that if $f : (X,d) \to (X,d)$ satisfies $(**)$ and X is compact, then f has a unique fixed point a and that $a = \lim\limits_{n\to\infty} x_n$ where x_0 is any point of X and $x_n = f^n(x_0)$ *Hints: $g(x) = d(x, f(x))$ has a minimum value at some point $x = a$. Show that a is the unique fixed point of f. Without loss of generality, assume all $x_n \neq a$; then $d(x_n, a) \to \ell \geq 0$. Prove that $\ell = 0$. At some point, use the fact that each sequence in X has a convergent subsequence.*

E9. Let $f\colon (X,d) \to (X,d)$, where (X,d) is a nonempty complete metric space. Let f^k denote the "k^{th} iteration of f" – that is, f composed with itself k times.

a) Suppose that $\exists k \in \mathbb{N}$ for which f^k is a contraction. Then, by the Contraction Mapping Theorem, f^k has a unique fixed point p. Prove that p is also the unique fixed point for f.

b) Prove that the function $\cos : \mathbb{R} \to \mathbb{R}$ is not a contraction.

c) Prove that for some $k \in \mathbb{N}$, $\cos^k$ is a contraction . (*Hint: the Mean Value Theorem may be helpful.*)

d) Pick $k \in \mathbb{N}$ so that $g = \cos^k$ is a contraction and let p be the unique fixed point of g. By a), p is also the unique solution of the equation $\cos x = x$. Start with 0 as a "first approximation" for p and use the technique in the proof of the Contraction Mapping Theorem to find an $n \in \mathbb{N}$ so that $|g^n(0) - p| < 0.00001$.

e) For this n, use a computer or calculator to evaluate $g^n(0)$. (*This "solves" the equation $\cos x = x$ with $|Error| < 0.00001$.*)

E10. Consider the differential equation $y' = x + y$ with the initial condition $y(0) = 1$. Choose a suitable rectangle D and suitable constants K, M and a as in the proof of Picard's Theorem. Use the technique in the proof of the contraction mapping theorem to find a solution for the initial value problem. Identify the interval I in the proof. Is the solution you found actually valid on an interval larger than I ?

5. Completions

The set of rationals $\mathbb{Q}$ (with the usual metric d) is not complete. However, $\mathbb{Q}$ is a dense subspace of the complete space $(\mathbb{R}, d)$. This is a model for the definition of a completion for a metric space. We will focus on the important case of metric spaces, but make a few additional comments along the way about pseudometric spaces where slight adjustments are necessary.

Definition 5.1 $(\widetilde{X}, \widetilde{d})$ is a completion of (X, d) if: $(\widetilde{X}, \widetilde{d})$ is complete
$\widetilde{d} \mid (X \times X) = d$, and
X is a dense subspace of $\widetilde{X}$.

Since $\widetilde{d} = d$ on X, we will simplify – and slightly abuse – the notation by also using d to refer to the metric on $\widetilde{X}$. Occasionally, if it's helpful to distinguish between d and the "extended" metric on $\widetilde{X}$, we may revert to more precise notation and use a different name, such as $\widetilde{d}$, for the extended metric on $\widetilde{X}$.

Loosely speaking, a completion of (X, d) contains the additional points necessary (and no others) to provide limits for Cauchy sequences that fail to converge in (X, d). In the case of $(\mathbb{Q}, d)$, the additional points are the irrational numbers, and the resulting completion is $(\mathbb{R}, d)$.

If a metric space (X, d) is already complete, what would its completion look like? In that case, (X, d) is a complete subspace of $(\widetilde{X}, d)$, so X must be closed in $\widetilde{X}$. But X must also be dense in $\widetilde{X}$, so $X = \widetilde{X}$ – that is, a complete metric space is its own completion.

> *If (X, d) is a complete pseudometric space, then X might not be closed in $\widetilde{X}$ and $\widetilde{X}$ might contain additional points y. But each $y \in \widetilde{X} - X$ is the limit of a (Cauchy) sequence (x_n) from X, and (x_n) already has a limit $x \in X$, so $d(x, y) = 0$. Any new points y in $\widetilde{X} - X$ are "unnecessary additions" because every Cauchy sequence in X already has a limit in X, and each of these "unnecessary additions" y is at distance 0 from a point in X.*

Of course different spaces might have the same completion – for example, $(\mathbb{R}, d)$ is a completion for both $(\mathbb{Q}, d)$ and $(\mathbb{P}, d)$.

Notice that a completion of (X, d) depends on d, not just the topology $\mathcal{T}_d$. For example, $(\mathbb{N}, d)$ and $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$ are homeomorphic topological spaces (countable sets, each with the discrete topology). But the completion of $(\mathbb{N}, d)$ ($=$ itself!) is not homeomorphic to the completion $(\{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}, d)$ of $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$.

One way to create a completion: recall that if f is an (onto) isometry between two metric spaces (X, d) and (Y, s), then (X, d) and (Y, s) can be regarded as "the same" metric space. We can think of f as just "assigning new names" to the points of X. If $f : (X, d) \to (Y, s)$ is an isometry from X into Y, we can identify (X, d) with $(f[X], s) =$ an "exact metric copy" of (X, d) inside (Y, s). If (Y, s) happens to be complete, then $(f[X], s)$ is dense in the complete space $(\text{cl}_Y f[X], s)$. We agree to identify $(f[X], s)$ with (X, d) and call $(\text{cl}_Y f[X], s)$ a completion of (X, d) even though X is not literally a subset of $\text{cl}_Y f[X]$. To find a completion of (X, d), then, it is sufficient to find an isometry f from (X, d) into any complete space (Y, s).

Example 5.2 Consider $\mathbb{N}$ with the metric $d'(n,m) = |\frac{1}{n} - \frac{1}{m}|$. $(\mathbb{N}, d')$ is isometric to $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$ where d is the usual metric on $\mathbb{N}$ – in fact, the mapping $f(n) = \frac{1}{n}$ is an isometry. So these spaces are "exact metric copies" of the other.

Because $(\{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}, d)$ is a completion of its dense subspace $(\{\frac{1}{n} : n \in \mathbb{N}\}, d)$, we can also think of $(\{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}, d)$ as a completion of $(\mathbb{N}, d')$ – even though $\mathbb{N}$ is not literally a subspace of $\{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$.

The next theorem tells us every metric space has a completion and, as we will see later, it is essentially unique (so the method for creating it usually doesn't matter).

Theorem 5.3 Every metric space (X, d) has a completion.

Proof By our earlier comments, it is sufficient to find an isometry of (X, d) into some complete metric space. We will use $(C^*(X), \rho)$ where, as usual, ρ is the metric of uniform convergence (*see Example 4.7 and Theorem 4.8*)

The theorem is trivial if $X = \emptyset$, so we assume that we can pick a point $p \in X$. For each point $a \in X$, define a function $\phi_a : X \to \mathbb{R}$ using the formula $\phi_a(x) = d(x,a) - d(x,p)$. Each map ϕ_a is continuous because it is a difference of continuous functions and, for each $x \in X$, $|\phi_a(x)| = |d(x,a) - d(x,p)| \leq d(a,p)$, so ϕ_a is bounded. Therefore $\phi_a \in C^*(X)$.

Define $\Phi : (X,d) \to (C^*(X), \rho)$ by $\Phi(a) = \phi_a$. We complete the proof by showing that Φ is an isometry. This just involves computing some distances: for any $a, b \in X$,

$$\begin{aligned}\rho(\phi_a, \phi_b) &= \sup\{|\phi_a(x) - \phi_b(x)| : x \in X\} \\ &= \sup\{|(d(x,a) - d(x,p)) - (d(x,b) - d(x,p))| : x \in X\} \\ &= \sup\{|d(x,a) - d(x,b)| : x \in X\}.\end{aligned}$$

For every x, $|d(x,a) - d(x,b)| \leq d(a,b)$, and letting $x = b$ gives $|d(x,a) - d(x,b)| = d(a,b)$. Therefore

$$\rho(\phi_a, \phi_b) = \sup\{|d(x,a) - d(x,b)| : x \in X\} = d(a,b). \quad \bullet$$

The completion of (X, d) given by this proof is $(\mathrm{cl}_{C^*(X)}\Phi[X], \rho)$. The proof is "slick" but it has very little intuitive content. For example, if we apply the proof to $(\mathbb{Q}, d)$, it is not at all clear that the resulting completion is isometric to $(\mathbb{R}, d)$ (as we would expect). In fact, there is another more intuitive way to construct a completion of (X, d), but verifying all the details is much more tedious. (*To reiterate: In Theorem 5.4, we will see that the method doesn't matter: the completion of (X, d) – no matter what method is used to construct it – always comes out "the same."*
We will simply sketch this alternate construction here, and in the discussion it's probably clearer if we use $\widetilde{d}$ to refer to the metric d extended to $\widetilde{X}$.

> We call two Cauchy sequences (x_n) and (y_n) in (X, d) equivalent if $d(x_n, y_n) \to 0$. It is easy to check that $\sim$ is an equivalence relation among Cauchy sequences in (X, d). Clearly, if $(x_n) \to z \in X$, then any equivalent Cauchy sequence also converges to z. And if two nonconvergent Cauchy sequences are equivalent, then they are "trying to converge to the same point" but that point is "missing" in X.

We denote the equivalence class of a Cauchy sequence (x_n) by $[(x_n)]$ and let $\widetilde{X}$ be the set of equivalence classes. Define the distance $\widetilde{d}$ between two equivalence classes by $\widetilde{d}\ ([(x_n)], [(y_n)]) = \lim_{n\to\infty} d(x_n, y_n)$. (*Why must this limit exist?*) It is easy to check that $\widetilde{d}$ does not depend on the choice of representative sequences from the equivalence classes, and that $\widetilde{d}$ is a metric on $\widetilde{X}$. One then checks (this is most tedious part) that $(\widetilde{X}, \widetilde{d}\ )$ is complete: any $\widetilde{d}$ -Cauchy sequence of equivalence classes $(\,[(x_n)]\,)$ must converge to an equivalence class in $(\widetilde{X}\ ,\ \widetilde{d}\ )$.

For each $x \in X$, the sequence $(x, x, x, ...)$ is Cauchy, and so $[(x, x, x, ...)] \in \widetilde{X}$.
The mapping $f : (X, d) \to (\widetilde{X}, \widetilde{d}\ )$ given by $f(x) = [(x, x, x, ...)]$ is an isometry, and it is easy to check that $f[X]$ is dense in $(\widetilde{X}, \widetilde{d}\ )$ – so that $(\widetilde{X}, \widetilde{d}\ )$ is a completion for (X, d).

This method is one of the standard ways to construct the real numbers from the rationals: $\mathbb{R}$ is defined as the set of these equivalence classes of Cauchy sequences of rational numbers. *Note: $\mathbb{R}$ can also be constructed as a completion of $\mathbb{Q}$ by using a method called Dedekind cuts in $\mathbb{Q}$. However that approach makes use of the ordering $<$ in $\mathbb{Q}$, so we cannot imitate this construction in the general setting of metric spaces where no ordering of elements exists.*

The next theorem gives us the good news that it doesn't really matter, in the end, how we construct a completion – because the completion of (X, d) is "essentially" unique: all completions are isometric, and in a "special" way! We state the theorem for metric spaces and in the proof it is clearer to give the metrics on the completions new names – not referring to them as d. (*Are there modifications of the statement and proof to handle the case where d is merely a pseudometric?*)

Theorem 5.4 The completion of (X, d) is unique in the following sense: if (Y, s) and (Z, t) are both complete metric spaces containing X as a dense subspace, then there is an (onto) isometry $f : (Y, s) \to (Z, t)$ such that $f|X$ is the identity map on X. (*In other words, not only are the completions Y and Z isometric, but there is an isometry between them that holds X fixed. The isometry merely "renames" the new points in the "outgrowth" $Y - X$.*)

Proof For each $y \in Y$, we can pick a sequence (x_n) in X which converges to y. Since (x_n) is convergent, (x_n) is Cauchy in (Y, s) – and therefore also in (Z, t) since both s and t agree with d on X. Because (Z, t) is complete, $(x_n) \to$ some point $z \in Z$, and we can define $f(y) = z$. (*It is easy to check that $f : (Y, s) \to (Z, t)$ is well-defined – that is, we get the same z no matter which sequence (x_n) we first chose converging to y.*)

For $x \in X$, what is $f(x)$? We can choose (x_n) to be the constant sequence $x_n = x$, and therefore $f(x) = x$. So $f|X$ is the identity map on X.

We need to verify that f is an isometry. Suppose $y' \in Y$ and we choose $(x'_n) \to y'$. Then

$$d(x_n, x'_n) = s(x_n, x'_n) \to s(y, y') \text{ and}$$
$$d(x_n, x'_n) = t(x_n, x'_n) \to t(z, z') = t(f(y), f(y')), \text{ so}$$
$$s(y, y') = t(f(y), f(y')) \quad \bullet$$

According to Theorem 5.4, the space $(\mathbb{R}, d)$ – no matter how we construct it – is the completion of the space $(\mathbb{Q}, d)$.

6. Category

There are many different mathematical ways to compare the "size" of sets, and these methods are used for different purposes. One of the simplest ways is to say that one set is "bigger" if it has "more points" than another set $-$ that is, by comparing their cardinal numbers.

In a totally different spirit, we might call one subset of $\mathbb{R}^2$ "bigger" than another if it has a larger area, and in analysis there is a further a generalization of area. A certain collection $\mathcal{M}$ of subsets of $\mathbb{R}^2$ contains sets that are called measurable, and for each set $S \in \mathcal{M}$ a nonnegative real number $\mu(S)$ is assigned. $\mu(S)$ is called the "measure of S" and we can think of $\mu(S)$ as a kind of "generalized area." A set with a larger measure is "bigger."

In this section, we will look at a third completely different and topologically useful idea for comparing the size of certain sets. The most interesting results in this section will be about complete metric spaces, but the basic definitions make sense in any topological space $(X, \mathcal{T})$.

Definition 6.1 A subset A of a topological space $(X, \mathcal{T})$ is called nowhere dense in X if $\text{int}_X(\text{cl}_X A) = \emptyset$. (*In some books, a nowhere dense set is called rare.*)

A set has empty interior iff its complement is dense. Therefore we also can say that A is nowhere dense in X iff $X - \text{cl}_X A$ is dense.

Intuitively, we can think of an open set O as including some "elbow room" around each of its points $-$ if $x \in O$, then all sufficiently near points are also in O. Then we think of a nowhere dense set as being "skinny" $-$ so skinny that not only does it contain no "elbow room" around any of its points, but even its closure contains no "elbow room" around any of its points.

Example 6.2

1) A closed set F is nowhere dense in X iff $\text{int}\, F = \emptyset$ iff $X - F$ is dense in X. In particular, if a singleton set $\{p\}$ is a closed set in X, then $\{p\}$ is nowhere dense unless p is isolated in X.

2) Suppose $A \subseteq \mathbb{R}$. A is nowhere dense iff $\text{cl}\, A$ contains no interval (a, b).

For example, each singleton set $\{r\}$ is nowhere dense in $\mathbb{R}$. In particular, for $n \in \mathbb{N}$, $\{n\}$ is nowhere dense in $\mathbb{R}$. But note that $\{n\}$ is not nowhere dense in $\mathbb{N}$, because n is isolated in $\mathbb{N}$. Whether a set is nowhere dense is relative to the space in which a set "lives."

If $B \subseteq A \subseteq X$, then $\text{int}_X(\text{cl}_X B)) \subseteq \text{int}_X(\text{cl}_X A))$. Therefore if A is nowhere dense in X, then B is also nowhere dense in X. However, the set B might not be nowhere dense in A. For example, consider $\{1\} \subseteq \mathbb{N} \subseteq \mathbb{R}$.

$\{\frac{1}{n} : n \in \mathbb{N}\}$ is nowhere dense in $\mathbb{R}$.

$\mathbb{Q}$ and $\mathbb{P}$ are not nowhere dense in $\mathbb{R}$ (*the awkward "double negative" in English is one reason why some authors prefer to use the term "rare" for "nowhere dense."*

3) Since $\text{cl}_X A = \text{cl}_X(\text{cl}_X A)$, the set on the left side has empty interior iff the set on the right side has empty interior – that is, A is nowhere dense in X iff $\text{cl}_X A$ is nowhere dense in X.

Theorem 6.3 Let $(X, \mathcal{T})$ be a topological space and $B \subseteq A \subseteq X$. If B is nowhere dense in A, then B is nowhere dense in X.

Proof Suppose not. Then there is a point $x \in \text{int}_X(\text{cl}_X B) \subseteq \text{cl}_X B$. Since $\text{int}_X(\text{cl}_X B)$ is an open set containing x and $x \in \text{cl}_X B$, then $\text{int}_X(\text{cl}_X B) \cap B \neq \emptyset$.

So $\emptyset \neq \text{int}_X(\text{cl}_X B) \cap B \subseteq \text{int}_X(\text{cl}_X B) \cap A \subseteq \text{cl}_X B \cap A = \text{cl}_A B$.

But $\text{int}_X(\text{cl}_X B) \cap A$ is a nonempty open set in A, so $\text{int}_A(\text{cl}_A B) \neq \emptyset$ – which contradicts the hypothesis that B is nowhere dense in A. •

The following technical results are sometimes useful for handling manipulations involving open sets and dense sets.

Lemma 6.4 In a topological space $(X, \mathcal{T})$:

1) If D is dense in X and O is open in X, then $\text{cl}\,(O \cap D) = \text{cl}\, O$.
2) If D is dense in X and O is open in X, then $O \cap D$ is dense in O.
3) If D is dense in X and O is open and dense in X, then $O \cap D$ is dense in X.

In particular, the intersection of two – and therefore finitely many – dense open sets is dense. (*But this is* <u>*not*</u> *true for countable intersections; can you provide an example?*)

Proof 1) To show $\text{cl}\, O \subseteq \text{cl}\,(O \cap D)$, suppose $x \in \text{cl}\, O$. If U is any open set containing x, then $U \cap O \neq \emptyset$. Since D is dense, this implies that $\emptyset \neq (U \cap O) \cap D = U \cap (O \cap D)$. Therefore $x \in \text{cl}\,(O \cap D)$.

2) Using part 1), we have $\text{cl}_O(O \cap D) = \text{cl}_X\,(O \cap D) \cap O = (\text{cl}_X\, O) \cap O = \text{cl}_O O = O$.

3) If O is dense, then part 1) gives $\text{cl}(O \cap D) = \text{cl}\, O = X$. •

Theorem 6.5 A <u>finite</u> union of nowhere dense sets in $(X, \mathcal{T})$ is nowhere dense in X.

Proof We will prove this for the union of two nowhere dense sets. The general case follows using a simple induction. If A_1 and A_2 are nowhere dense in X, then

$$X - \text{cl}\,(A_1 \cup A_2) = X - (\text{cl}\, A_1 \cup \text{cl}\, A_2) = (X - \text{cl}\, A_1) \cap (X - \text{cl}\,_2).$$

The last two sets are open and dense, so Lemma 6.4(3) gives that $X - \text{cl}\,(A_1 \cup A_2)$ is dense. Therefore $A_1 \cup A_2$ is nowhere dense. •

Notice that Theorem 6.5 is <u>false for infinite unions</u>: for each $q \in \mathbb{Q}$, $\{q\}$ is nowhere dense in $\mathbb{R}$, but $\bigcup_{q \in \mathbb{Q}} \{q\} = \mathbb{Q}$ is not nowhere dense in $\mathbb{R}$.

Definition 6.6 In $(X, \mathcal{T})$, a subset A is called a <u>first category set in X</u> if A can be written as a countable union of sets that are nowhere dense in X. If A is not first category in X, we say that A is <u>second</u> <u>category</u> <u>in</u> <u>X</u>. (*Books that use the terminology* <u>*rare*</u> *for nowhere dense sets usually use the word* <u>*meager*</u> *for first category sets.*)

If we think of a nowhere dense set in X as "skinny," then a first category set is a bit larger – merely "thin."

Example 6.7

1) If A is nowhere dense in X, then A is first category in X.

2) $\mathbb{Q}$ is a first category set in $\mathbb{R}$.

3) If $B \subseteq A \subseteq X$ and A is first category in X, then B is first category in X. However B may not be first category in A, as the example $\{0\} \subseteq \{0, 1\} \subseteq \mathbb{R}$ shows. However, using Theorem 6.3, we can easily prove that if B is first category in A, then B is also first category in X.

4) A countable union of first category sets in X is first category in X.

5) $A = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$. A is nowhere dense (and therefore first category) in $\mathbb{R}$; but A is second category <u>in A</u> – because any subset of A that contains 1 is not nowhere dense in A.

6) Cardinality and category are totally independent ways to talk about the "size" of a set.

a) Consider the right-ray topology $\mathcal{T} = \{\emptyset, \mathbb{R}\} \cup \{(a, \infty) : a \in \mathbb{R}\}$ on $\mathbb{R}$. $\mathbb{R}$ is uncountable, but $\mathbb{R} = \bigcup_{n=1}^{\infty} F_n$, where $F_n = (-\infty, n]$. Each F_n is nowhere dense in $(\mathbb{R}, \mathcal{T})$, so $\mathbb{R}$ is first category in $(\mathbb{R}, \mathcal{T})$.

b) On the other hand, $\mathbb{N}$ is countable but $\mathbb{N}$ (with the usual topology) is second category in $\mathbb{N}$.

7) Let $n > 1$. A straight line is nowhere dense in $\mathbb{R}^n$, so a countable union of straight lines in $\mathbb{R}^n$ is first category in $\mathbb{R}^n$. (*Can $\mathbb{R}^n$ be written as a countable union of straight lines? The answer follows immediately from the Baire Category Theorem, proved below. But, in fact, the answer is clear from a simple argument using countable and uncountable sets.*)

More generally, when $k < n$, a countable union of k-dimensional linear subspaces of $\mathbb{R}^n$ is first category in $\mathbb{R}^n$. (*Can a countable union of k-dimensional linear subspaces $= \mathbb{R}^n$?*)

It is not always easy to say what the "category" of a set is. Is $\mathbb{P}$ a first or second category set in $\mathbb{R}$? For that matter, is $\mathbb{R}$ first or second category in $\mathbb{R}$? (*If you know the answer to either of these questions, you also know the answer to the other: why are the questions equivalent?*)

As we observed in Lemma 6.4, the intersection of two (and therefore, finitely many) dense open sets is dense. <u>Sometimes</u>, however, the intersection of a countable collection of dense open sets is dense. The following theorem discusses this condition and leads us to a definition.

Theorem 6.8 In any topological space $(X, \mathcal{T})$, the following two statements are <u>equivalent</u>:

1) If A is first category in X, then $X - A$ is dense in X
2) For each sequence $G_1, G_2, ..., G_k, ...$ of dense open sets, the intersection $\bigcap_{k=1}^{\infty} G_k$ is also dense.

Proof (1 $\Rightarrow$ 2) Let $G_1, G_2, ..., G_k, ...$ be a sequence of dense open sets. Each $X - G_k$ is closed and nowhere dense, so $\bigcup_{k=1}^{\infty}(X - G_k)$ is first category. By 1), $X - \bigcup_{k=1}^{\infty}(X - G_k) = \bigcap_{k=1}^{\infty} G_k$ is dense.

(2 $\Rightarrow$ 1) Let A be a first category set, say $A = \bigcup_{n=1}^{\infty} N_k$, where each N_k is nowhere dense in X. Then each $X - \text{cl}\, N_k = G_k$ is dense so, by 2), $\bigcap_{k=1}^{\infty}(X - \text{cl}\, N_k)$ is dense. Since $\bigcap_{k=1}^{\infty}(X - \text{cl}\, N_k) = X - \bigcup_{k=1}^{\infty} \text{cl}\, N_k \subseteq X - \bigcup_{k=1}^{\infty} N_k = X - A$, we see that $X - A$ is dense. ●

Definition 6.9 A space $(X, \mathcal{T})$ is called a Baire space if one (therefore both) of the conditions in Theorem 6.8 holds.

Intuitively we can think of a nonempty Baire space as one which is "thick," at least in some places. A "thin" first category set A can't "fill up" X : in fact, its complement $X - A$ is dense (and therefore nonempty). This fact is the basis for how a Baire space is often used in "applications": if you want to prove that there is an element in a Baire space X with a certain property P, you can consider $A = \{x \in X : x$ does <u>not</u> have property $P\}$. If you can show that A is first category in X, it follows that $X - A \neq \emptyset$.

Clearly, a space homeomorphic to a Baire space is a Baire space: being a Baire space is a topological property.

The following theorem tells us a couple of important facts.

Theorem 6.10 Suppose O is an open set in a Baire Space $(X, \mathcal{T})$.

1) If $O \neq \emptyset$, then O is second category in X. In particular, a <u>nonempty</u> Baire space X is second category in itself.

2) O is a Baire space.

Proof 1) If O were first category in X, then (by definition of a Baire space) $X - O$ would be dense in X; since $X - O$ is closed, it would follow that $X - O = X$, and therefore $O = \emptyset$.

2) Suppose O is open in the Baire space X, and let A be a first category set <u>in O</u>. We must show that $O - A$ is dense in O. Suppose $A = \bigcup_{k=1}^{\infty} N_k$, where N_k is nowhere dense in O. By Theorem 6.3, N_k is also nowhere dense in X. Therefore A is first category in X so $X - A$ is dense in X. Since O is open, $O \cap (X - A) = O - A$ is dense in O by part 2) of Lemma 6.4. ●

Example 6.11 By Theorem 6.10.1, a nonempty Baire space is second category in itself. The converse is false. To see this, let $X = (\mathbb{Q} \cap [0, 1]) \cup \{2\}$, with its usual topology. If we write $X = \bigcup_{k=1}^{\infty} N_k$, then the isolated point "2" must be in some N_k, so N_k is not nowhere dense. Therefore X is second category in itself. If X were Baire, then, by Theorem 6.10, the open subspace $\mathbb{Q} \cap [0, 1]$ would also be Baire and therefore second category in itself. However this is false since $\mathbb{Q} \cap [0, 1]$ is a countable union of (nowhere dense) singleton sets.

To make use of properties of Baire spaces, it would be helpful to have a "large supply" of Baire spaces. The next theorem provides us with many Baire spaces.

Theorem 6.12 (The Baire Category Theorem) A complete metric space (X, d) is a Baire space. (*and by 6.10, therefore, a nonempty complete metric space must be second category in itself.*)

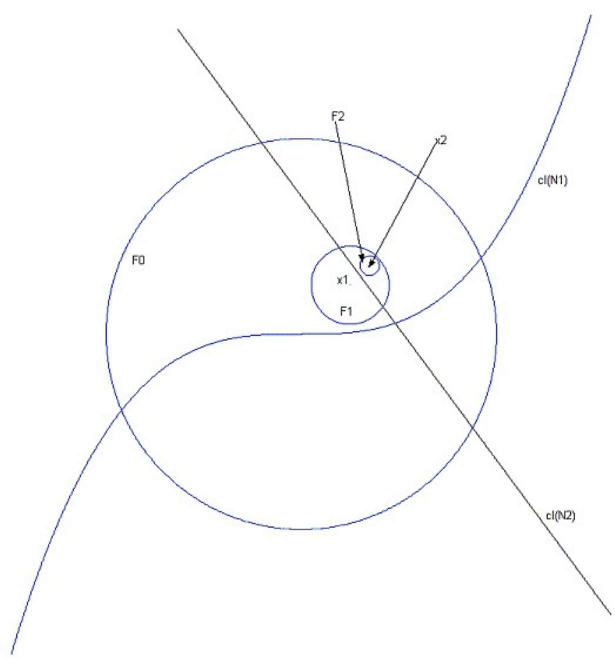
Proof If $X = \emptyset$, then X is Baire, so we assume $X \neq \emptyset$.

Let $A = \bigcup_{k=1}^{\infty} N_k$, where N_k is nowhere dense. We must show that $X - A$ is dense in X. Suppose $x_0 \in X$. The closed balls of the form $\{x : d(x_0, x) \leq \epsilon\}$ form a neighborhood base at x_0. Let F_0 be such a closed ball, centered at x_0 with radius $\epsilon > 0$. We will be done if we can show $F_0 \cap (X - A) \neq \emptyset$. We do that by using the Cantor Intersection Theorem.

Since int $F_0 \neq \emptyset$ (it contains x_0) and since N_1 is nowhere dense, we know that int F_0 is not a subset of cl N_1. Therefore we can choose a point x_1 in the open set int F_0 $-$ cl N_1. Pick a closed ball F_1, centered at x_1, so that $x_1 \in F_1 \subseteq$ int $F_0 -$ cl $N_1 \subseteq F_0$. If necessary, choose F_1 even smaller so that diam$(F_1) < \frac{1}{2}$.

(*This is the first step of an inductive construction. We could now move to the induction step, but actually include "step two" to be sure the process is clear.*)

Since int $F_1 \neq \emptyset$ (it contains x_1) and since N_2 is nowhere dense, we know that int F_1 is not a subset of cl N_2. Therefore we can choose a point x_2 in the open set int F_1 $-$ cl N_2. Pick a closed ball F_2, centered at x_2, so that $x_2 \in F_2 \subseteq$ int $F_1 -$ cl $N_2 \subseteq F_1$. If necessary, choose F_2 even smaller so that diam$(F_2) < \frac{1}{3}$.



For the induction step: suppose we have defined closed balls $F_1, F_2, ..., F_k$ centered at points $x_1, x_2, ..., x_k$, with diam$(F_i) < \frac{1}{i+1}$ and so that $x_i \in F_i \subseteq$ int $F_{i-1} -$ cl $N_i \subseteq F_{i-1}$.

Since int $F_k \neq \emptyset$ (it contains x_k) and since N_{k+1} is nowhere dense, we know that int F_k is not a subset of cl N_{k+1}. Therefore we can choose a point x_{k+1} in the open set int F_k $-$ cl N_{k+1}. Pick a closed ball F_{k+1}, centered at x_{k+1}, so that $x_{k+1} \in F_{k+1} \subseteq$ int $F_k -$ cl $N_{k+1} \subseteq F_k$. If necessary, choose F_{k+1} even smaller so that $\operatorname{diam}(F_{k+1}) < \frac{1}{k+2}$.

By induction, the closed sets F_k are defined for all k, $\operatorname{diam}(F_k) \to 0$ and $F_0 \supseteq F_1 \supseteq ... \supseteq F_k \supseteq ...$. Since (X, d) is complete, the Cantor Intersection Theorem says that there is a point $x \in \bigcap_{k=1}^{\infty} F_k \subseteq F_0$. For each $k \geq 1$, $x \in F_k \subseteq$ int $F_{k-1} -$ cl N_k, so $x \notin$ cl N_k, so $x \notin N_k$. Therefore $x \in X - \bigcup_{k=1}^{\infty} N_k = X - A$, so $x \in F_0 \cap (X - A)$ and we are done. •

Example 6.13

1) Now we can see another reason why $\mathbb{Q}$ is not completely metrizable. If $(\mathbb{Q}, d')$ is complete, then $(\mathbb{Q}, d')$ is a Baire space. But $\mathbb{Q}$, with the usual metric d is not a Baire space. Therefore $\mathcal{T}_d \neq \mathcal{T}_{d'}$ so $d \not\sim d'$.

2) Suppose (X, d) is a nonempty complete metric space without isolated points. We proved in Theorem 3.6 that $|X| \geq c$. We can now see even more: that every point in X must be a condensation point.

Otherwise, there would be a non-condensation point $x_0 \in X$, and there would be some countable (possibly finite) open set $O = \{x_0, x_1, x_2, ...x_n, ...\}$. Since each x_n is non-isolated, the singleton sets $\{x_n\}$ are nowhere dense in (X, d), so O is first category in (X, d). Since (X, d) is Baire, this would mean that the closed set $X - O$ is dense $-$ which is impossible.

Notice that this example illustrates again that $\mathbb{Q}$ cannot be completely metrizable: if its topology were produced by a some complete metric, then each point in $\mathbb{Q}$ would have to be a condensation point.

3) The set $\mathbb{P}$ is a second category set in $\mathbb{R}$: if not, we could write $\mathbb{R} = \mathbb{P} \cup \mathbb{Q}$, so that $\mathbb{R}$ would be a first category set in $\mathbb{R}$ $-$ which contradicts the Baire Category Theorem. (*Is $\mathbb{P}$ second category in $\mathbb{P}$?*)

4) Recall that a subset A of $(X, \mathcal{T})$ is called an F_σ set if A can be written as a countable union of closed sets, and that A is a G_δ set if it can be written countable intersection of open sets. The complement of a G_δ set is an F_σ set and vice-versa.

$\mathbb{P}$ is not an F_σ set in $\mathbb{R}$. To see this, suppose that $\mathbb{P} = \bigcup_{n=1}^{\infty} F_n$, where the F_n's are closed in $\mathbb{R}$. Since $\mathbb{P}$ is second category in $\mathbb{R}$, one of these F_n's must be not nowhere dense in $\mathbb{R}$. This F_n must therefore contain an open interval (a, b). But then $(a, b) \subseteq F_n \subseteq \mathbb{P}$, which is impossible because there are rational numbers in any interval (a, b).

Taking complements, we see that $\mathbb{Q}$ is not a G_δ-set in $\mathbb{R}$.

Example 6.14 Suppose $f:[0,1]\to\mathbb{R}$ is continuous. Then f has an antiderivative f_1, by the Fundamental Theorem of Calculus. Since f_1 is differentiable (therefore continuous), it has an antiderivative f_2. Continuing in this way, let f_n denote an antiderivative of f_{n-1}. It is a trivial observation that:

$$(\exists k\,\forall x\, f_k(x)=0) \;\Rightarrow\; f(x)=0 \text{ for all } x\in[0,1]$$

We will use the Baire Category Theorem to prove a much less obvious fact:

$$(\forall x\;\exists k\; f_k(x)=0) \;\Rightarrow f(x)=0 \text{ for all } x \text{ in } [0,1] \quad (*)$$

In other words, if $f(x)$ is not identically 0 on $[0,1]$, then there must exist a point $x_0\in[0,1]$ such that every antiderivative $f_k(x_0)\neq 0$.

The hypothesis in (*) lets us write $[0,1]=\bigcup_{k=1}^{\infty}F_k$, where $F_k=\{x\in[0,1]: f_k(x)=0\}$ $=f_k^{-1}[\{0\}]$; these are closed sets in $[0,1]$ because each f_k is continuous. Since $[0,1]$ is second category in itself, one of the sets F_{k_0} is not nowhere dense and therefore contains an interval (a,b). Since f_{k_0} is identically 0 on (a,b), f is identically 0 on (a,b).

We can repeat the same argument on any closed subinterval $J=[c,d]\subseteq[0,1]$: by letting $g=f|J$ and $g_k=f_k|J$, we can conclude that J contains an open interval on which f is identically 0. In particular, each closed subinterval $J=[c,d]$ contains a point x at which $f(x)=0$. Therefore $\{x: f(x)=0\}$ is dense in $[0,1]$. Since f is continuous and f is 0 on a dense set, f must be 0 everywhere in $[0,1]$. (*See Theorem II.5.12 and Exercise III.E.16.*)

Example 6.15 (The Banach-Mazur Game) The equipment for the game consists of two disjoint sets A, B where $A\cup B=[0,1]$. Set A belongs to Andy and set B belongs to Beth. Andy gets the first "move" and selects a closed interval $I_1\subseteq[0,1]$. Beth then chooses a closed interval $I_2\subseteq I_1$, where I_2 has length $\leq\frac{1}{2}$. Andy then selects a closed interval $I_3\subseteq I_2$, where I_3 has length $\leq\frac{1}{3}$. They continue back-and-forth in this way forever. (*Of course, to finish in a finite time they must make their choices faster and faster.*) When all is done, they look at $\bigcap_{n=1}^{\infty}I_n=\{x\}$. If $x\in A$, Andy wins; if $x\in B$, Beth wins. We claim that if A is first category, then Beth can always win.

Suppose $A=\bigcup_{n=1}^{\infty}N_k$, where A_k is nowhere dense in $[0,1]$. Let Andy's k^{th} choice be J $(=I_{2k-1})$. Of course, $J\cap N_k$ is nowhere dense in $[0,1]$; but actually $J\cap N_k$ is also nowhere dense in the interval J:

> If $\text{int}_J\text{cl}_J(J\cap N_k)\neq\emptyset$, then $\text{cl}_J(J\cap N_k)$ must contain an nonempty open interval (a,b). But then
> $$(a,b)\subseteq \text{cl}_J(J\cap N_k)\subseteq \text{cl}_J N_k=\text{cl}_{[0,1]}N_k\cap J\subseteq \text{cl}_{[0,1]}N_k,$$
> which contradicts the fact that N_k is nowhere dense in $[0,1]$.

Then the open set $\text{int}\,J-\text{cl}_J(J\cap N_k)$ is nonempty and Beth can make her k^{th} choice I_{2k} to be a closed interval $[a,b]\subseteq \text{int}\,J-\text{cl}_J(J\cap N_k)$. This implies that $I_{2k}\cap N_k=\emptyset$, since

$$I_{2k}\subseteq \text{int}\,J-\text{cl}_J(J\cap N_k)\subseteq \text{int}\,J-(J\cap N_k)\subseteq J-(J\cap N_k)=J-N_k$$

Therefore $\bigcap_{k=1}^{\infty}I_k\cap\bigcup_{k=1}^{\infty}N_k=\emptyset$ and Beth wins!

Example 6.16 The Baire Category Theorem can also be used to prove the existence of a continuous function on $[0,1]$ that is nowhere differentiable. The details can be found, for example, in *General Topology* (S. Willard). Roughly, one looks at the space $C^*([0,1])$ with the uniform metric ρ, and argues that the set N of functions which have a derivative at one or more points is a ("thin") first category set in $C^*([0,1])$. But the complete space $(C^*([0,1]),\rho)$ is a ("thick") Baire space. Therefore it is second category in itself so $C^*([0,1]) - N \neq \emptyset$. In fact, $C^*([0,1]) - N$ is dense in $C^*([0,1])$. Any function in $C^*([0,1]) - N$ does the trick.

7. Complete Metrizability

Which metric spaces (X,d) are completely metrizable – that is, when does there exist a metric d' on X for which (X,d') is complete and $d \sim d'$? We have already seen, using the Baire Category Theorem, that $\mathbb{Q}$ is not completely metrizable, and that certain familiar spaces like $\{\frac{1}{n} : n \in \mathbb{N}\}$ are completely metrizable. In this section, we will answer to this question.

Lemma 7.1 Let (X,d) be a pseudometric space. If $g : X \to \mathbb{R}$ continuous and $g(a) \neq 0$, then $\frac{1}{g} : X \to \mathbb{R}$ is continuous at a.

Proof The proof is a perfect "mimic" of the proof in analysis (where $X = \mathbb{R}$ or $\mathbb{R}^n$).

Our first theorem tells us that certain subspaces of complete spaces are completely metrizable.

Theorem 7.2 If (X,d) is complete and O is open in X, then there is a metric $d' \sim d$ on O such that (O,d') is complete – that is, O is completely metrizable.

If (O,d) is not already complete, it is because there are some nonconvergent Cauchy sequences in O. But those Cauchy sequences do have limits in (X,d) – they converge to points outside O but in Fr O. The idea of the proof is to create a new metric d' on O that is equivalent to d, and which "blows up distances" near the boundary of O: in other words, the new metric destroys the "Cauchyness" of the nonconvergent Cauchy sequences.

Here is a concrete (but slightly simpler) example to illustrate the idea.

> Let d be the usual metric on the interval $J = (-\frac{\pi}{2}, \frac{\pi}{2})$. The sequence $(x_n) = (\frac{\pi}{2} - \frac{1}{n})$ is a nonconvergent Cauchy sequence in J. The function $\tan : J \to \mathbb{R}$ is a homeomorphism for which $(f(x_n)) \to \infty$.
>
> $d'(x,y) = |\tan x - \tan y|$ defines a new metric on J, and $\tan : (J,d') \to (\mathbb{R},d)$ is an isometry because $d'(x,y) = \ = |\tan x - \tan y| = d(\tan x, \tan y)$. Since $(\mathbb{R},d)$ is complete, so in (J,d'). (*The sequence* (x_n) *is still nonconvergent in* (J,d') *because* $d \sim d'$; *but in* (J,d'), *it is no longer a Cauchy sequence.*)
>
> In the proof of Theorem 7.2, we will not have a homeomorphism like "tan" to use. Instead, we define a continuous $f : (O,d) \to \mathbb{R}$ and use f to create a new metric d' on O that destroys Cauchy sequences (x_n) when they approach Fr O.

Proof of 7.2 Define $f(x) = \frac{1}{d(x, X-O)}$. Since O is open, the denominator is never 0 for $x \in O$ so $f : O \to \mathbb{R}$ is continuous, by Lemma 7.1. On O, define $d'(x,y) = d(x,y) + |f(x) - f(y)|$. It is easy to check that d' is a metric on O.

Since sequences are sufficient to determine the topology in pseudometric spaces, we can show that $d' \sim d$ on O by showing that they produce the same convergent sequences in O. Suppose (x_n) is a sequence in O and $x \in O$:

If $d'(x_n, x) \to 0$, then since $d' \geq d$ we get $d\,(x_n, x) \to 0$.

If $d\,(x_n, x) \to 0$, then $|f(x_n) - f(x)| \to 0$ (since f is continuous), so $d'\,(x_n, x) \to 0$.

(O, d') is complete : suppose (x_n) is a d'-Cauchy sequence in O. Since $d' \geq d$, the sequence (x_n) is also d-Cauchy, so $(x_n) \overset{d}{\to}$ some $x \in X$. But, in fact, this x must be in O.

If $x \in X - O$, then $d(x_n, X - O) \to 0$ and so $f(x_n) \to \infty$. In particular, this means that for every n we can find $m > n$ for which $f(x_m) > f(x_n) + 1$. Then $|f(x_m) - f(x_n)| > 1$, so $d'(x_m, x_n) > 1$. This contradicts the assumption that (x_n) is d'-Cauchy.

Therefore $(x_n) \overset{d}{\to} x \in O$. Since $d \sim d'$ on O, $(x_n) \overset{d'}{\to} x \in O$ and so (O, d') is complete. •

The next theorem generalizes this result, but the really interesting idea is actually in Theorem 7.2. The proof of Theorem 7.3 just uses Theorem 7.2 repeatedly to "patch together" a more general result.

Theorem 7.3 (Alexandroff) If (X, d) is complete metric space and A is a G_δ in X, then there is a metric ρ on A, with $\rho \sim d$ on A, for which (A, ρ) is complete. In other words: a G_δ subset of a complete space is completely metrizable.

Proof Let $A = \bigcap_{i=1}^{\infty} O_i$, where O_i is open in X. Let d_i' be a metric on O_i, equivalent to d on O_i, such that $(O_i,\, d_i')$ is complete. Let $d_i(x,y) = \min\{d_i'(x,y), 1\}$. Then $d_i \sim d_i' \sim d$ on O, and (O_i, d_i) is complete. (*Note that d_i and d_i' are identical for distances smaller than one: this is why they produce the same convergent sequences and the same Cauchy sequences.*)

For $x, y \in A$, define $\rho(x,y) = \sum_{i=1}^{\infty} \frac{d_i(x,y)}{2^i}$ (the series converges since all $d_i(x,y) \leq 1$).

ρ is a metric on A:

Exercise

$\rho \sim d$ on A:

We show that ρ and d produce the same convergent sequences in A.

Suppose $(a_n) \xrightarrow{d} a \in A \subseteq O_i$. Let $\epsilon > 0$. For every i, $d \sim d_i$ on O_i, so $(a_n) \xrightarrow{d_i} a \in O_i$. Pick k so that $\sum_{j=k+1}^{\infty} \frac{1}{2^j} < \frac{\epsilon}{2}$. For each $i = 1, \dots, k$, we can pick N_i so if $n > N_i$ then $\frac{d_i(a_n,a)}{2^i} < \frac{\epsilon}{2k}$. Then if $n > N = \max\{N_1, \dots, N_k\}$, we get $\rho(a_n,a) = \sum_{i=1}^{k} \frac{d_i(a_n,a)}{2^i} + \sum_{i=k+1}^{\infty} \frac{d_i(a_n,a)}{2^i} \leq k \cdot \frac{\epsilon}{2k} + \frac{\epsilon}{2} = \epsilon$, so $(a_n) \xrightarrow{\rho} a \in A$.

Conversely, if $(a_n) \xrightarrow{\rho} a \in A$, then for $\epsilon > 0$, we can pick N so that $n > N \Rightarrow \rho(a_n, a) = \sum_{i=1}^{\infty} \frac{d_i(a_n,a)}{2^i} < \frac{\epsilon}{2} \quad \Rightarrow d_1(a_n, a) < \epsilon$, so $(a_n) \xrightarrow{d_1} a \in O_1$. But $d_1 \sim d$ on O_1, so $(a_n) \xrightarrow{d} a$.

(A, ρ) is complete:

Let (a_n) be ρ-Cauchy in A. Let $\epsilon > 0$. For any $i \in \mathbb{N}$, we can choose N_i so that

if $m, n > N_i$, then $\rho(a_n, a_m) < \frac{\epsilon}{2^i}$, and therefore $d_i(a_n, a_m) < \epsilon$.

Therefore (a_n) is Cauchy in the complete space (O_i, d_i) and there is a point $x_i \in O_i$ such that $(a_n) \xrightarrow{d_i} x_i \in O_i$. But $d_i \sim d$ on O_i, so $(a_n) \xrightarrow{d} x_i \in O_i \subseteq X$.

This is true for each i. But (a_n) can have only one limit $a \in (X, d)$, so we conclude that $x_1 = x_2 = \dots = x_i = \dots = a$. Since $a \in O_i$ for each i, we have $a \in A$, so $(a_n) \xrightarrow{d} a \in A$. But $d \sim \rho$ on A, so $(a_n) \xrightarrow{\rho} a \in A$. Therefore (A, ρ) is complete. $\bullet$

Corollary 7.4 $\mathbb{P}$ is completely metrizable (and therefore $\mathbb{P}$ is a Baire space, so $\mathbb{P}$ is second category <u>in $\mathbb{P}$</u>).

Proof $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}$, so $\mathbb{Q}$ is an F_σ set in $\mathbb{R}$. So, taking complements, we get that $\mathbb{P} = \bigcap_{q \in \mathbb{Q}} (\mathbb{R} - \{q\})$ is a G_δ set in $\mathbb{R}$. $\bullet$

In fact, a sort of "converse" to Alexandroff's Theorem is also true.

Theorem 7.5 For any metric space (X, d), the following are <u>equivalent</u>:

i) (X, d) is completely metrizable
ii) (X, d) is homeomorphic to a G_δ set in some complete metric space (Y, d')
iii) If (Z, d'') is any metric space and $f : (X, d) \to (f[X], d'') \subseteq (Z, d'')$ is a homeomorphism, then $f[X]$ is a G_δ set in (Z, d'') (*therefore we say that "X is an absolute G_δ set among metric spaces"*)
iv) X is a G_δ set in the completion $(\widetilde{X}, \widetilde{d})$.

Various parts of Theorem 7.5 are due to Mazurkiewicz (1916) & Alexandroff (1924).

The proof of one of the implications in Theorem 7.5 requires a technical result whose proof we will omit. (See, for example Willard, *General Topology*)

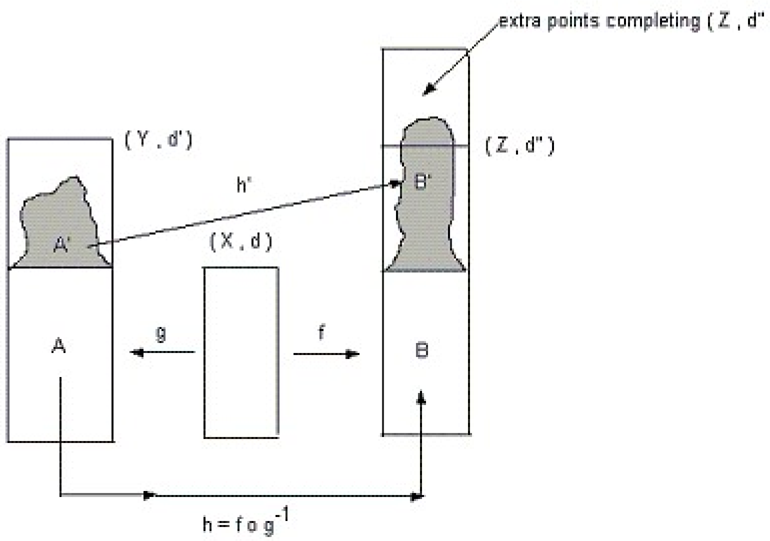
Theorem 7.6 (Lavrentiev) Suppose (X, d) and (Y, d') are complete metric spaces with $A \subseteq X$ and $B \subseteq Y$. Let h be a homeomorphism from A onto B. Then there exist G_δ sets A' in X and B' in Y with $A \subseteq A' \subseteq \text{cl}\, A$ and $B \subseteq B' \subseteq \text{cl}\, B$ and there exists a homeomorphism h' from A' onto B' such that $h'|A = h$. (*Loosely stated: a homeomorphism between subsets of two complete metric spaces can always be "extended" to a homeomorphism between G_δ sets.*)

Assuming Lavrentiev's Theorem, we now prove Theorem 7.5.

Proof i) $\Rightarrow$ ii) Suppose (X, d) is completely metrizable and let d' be a complete metric on X equivalent to d. Then the identity map $i : (X, d) \to (X, d')$ is a homeomorphism and $i[X] = X$ is certainly a G_δ set in (X, d'). (*In other words, we can use $(X, d) = (Y, d')$ in Part 2).*

ii) $\Rightarrow$ iii) Suppose we have a homeomorphism $g : (X, d) \to g[X] = A$, where A is a G_δ set in a complete space (Y, d'). Let $f : (X, d) \to (Z, d'') \subseteq$ the completion $(\widetilde{Z}, \widetilde{d}'')$ be a homeomorphism (into Z). We want to prove that $B = f[X]$ must be a G_δ set in Z.

We have that $h = fg^{-1} : A \to B$ is a homeomorphism. Using Lavrentiev's Theorem, we get an extension of h to a homeomorphism $h' : A' \to B'$ where A', B' are G_δ sets with $A \subseteq A' \subseteq Y$ and $B \subseteq B' \subseteq \widetilde{Z}$ (*see the figure*).



Since A is a G_δ in Y, there are open sets O_n in Y such that $A = \bigcap_{n=1}^{\infty} O_n = A' \cap \bigcap_{n=1}^{\infty} O_n = \bigcap_{n=1}^{\infty} (O_n \cap A')$. Therefore A is also a G_δ in A'. But $h' : A' \to B'$ is a homeomorphism, so $h'[A] = h[A] = B$ is a G_δ set in B'. Therefore $B = \bigcap_{n=1}^{\infty} V_n$, where each V_n is open in B' and, in turn, $V_n = B' \cap W_n$ where W_n is open in $\widetilde{Z}$.

Also, B' is a G_δ in $\widetilde{Z}$, so $B' = \bigcap_{n=1}^{\infty} U_n$, where the U_n's are open in $\widetilde{Z}$.

Putting all this together, $B = \bigcap_{n=1}^{\infty} V_n = \bigcap_{n=1}^{\infty} (W_n \cap B') = (\bigcap_{n=1}^{\infty} W_n) \cap B'$
$= (\bigcap_{n=1}^{\infty} W_n) \cap (\bigcap_{n=1}^{\infty} U_n)$. The last expression shows that B is a G_δ set in $\widetilde{Z}$. But then $B = Z \cap B = Z \cap (\bigcap_{n=1}^{\infty} W_n) \cap (\bigcap_{n=1}^{\infty} U_n) = (\bigcap_{n=1}^{\infty} W_n \cap Z) \cap (\bigcap_{n=1}^{\infty} U_n \cap Z)$
is a G_δ set in Z.

iii) $\Rightarrow$ iv) iii) states that any "homeomorphic copy" of X in a metric space (Z, d'') must be a G_δ in Z. Letting $(Z, d'') = (\widetilde{X}, \widetilde{d})$, it follows that X must be a G_δ set in the completion $(\widetilde{X}, \widetilde{d})$.

iv) $\Rightarrow$ i) This follows from Alexandroff's Theorem 7.3: a G_δ set in a complete space is completely metrizable. •

Theorem 7.5 characterizes, for metric spaces, the "absolute G_δ sets. This suggests other questions: what spaces are "absolutely open"? what spaces are "absolutely closed?" Of course in each case a satisfactory answer might involve some kind of qualification. For example, "among Hausdorff spaces, a space X is absolutely closed iff ... "

Example 7.7 Since $\mathbb{Q}$ is not a G_δ in $\mathbb{R}$ (an earlier consequence of the Baire Category Theorem), Theorem 7.5 gives us yet another reason why $\mathbb{Q}$ is not completely metrizable.

Example 7.8 Since $\mathbb{P}$ is completely metrizable, it follows that $\mathbb{P}$ is a Baire space. If d is the usual metric on $\mathbb{P}$, then $(\mathbb{P}, d)$ is an example of a Baire metric space that isn't complete.

To finish this section on category we mention, just as a curiosity and without proof, a generalization of an earlier result (*Blumberg's Theorem: see Example II.5.8*). I have never seen it used anywhere.

Theorem 7.9 Suppose (X, d) is a Baire metric space. For every $f : X \to \mathbb{R}$, there exists a dense subset D of X such that $f|D : D \to \mathbb{R}$ is continuous,

The original theorem of this type was proved for $X = \mathbb{R}$ or $\mathbb{R}^2$. (Blumberg, *New properties of all real-valued functions*, Transactions of the American Mathematical Society, 24(1922) 113-128).

The more general result stated in Theorem 7.9 is (essentially) due to Bradford and Goffman (*Metric spaces in which Blumberg's Theorem holds*, Proceedings of the American Mathematical Society 11(1960), 667-670)

Exercises

E11. In a pseudometric space (X, d), every closed set is a G_δ set and every open set is an F_σ set (*see Exercise II.E.15*).

a) Find a topological space $(X, \mathcal{T})$ containing a closed set F that is not a G_δ set.

b) Recall that the "scattered line" is the space $(\mathbb{R}, \mathcal{T})$ where

$$\mathcal{T} = \{ U \cup V : U \text{ is a usual open set in } \mathbb{R} \text{ and } V \subseteq \mathbb{P} \}$$

Prove that the scattered line is not metrizable. (*Hint: G_δ or F_σ sets are relevant.*)

E12. a) Prove that if O is open in $(X, \mathcal{T})$, then $\text{Fr}\,(O)$ is nowhere dense.

b) Suppose that D is a discrete subspace of a Hausdorff space $(X, \mathcal{T})$ and that X has no isolated points. Prove that D is nowhere dense in X.

E13. Let $\mathcal{T}$ be the cofinite topology. Prove that $(X, \mathcal{T})$ is a Baire space if and only if X is either finite or uncountable.

E14. Suppose d is any metric on $\mathbb{Q}$ which is equivalent to the usual metric on $\mathbb{Q}$. Prove that the completion $(\widetilde{\mathbb{Q}}, \widetilde{d})$ is uncountable.

E15. Suppose that (X, d) is a nonempty complete metric space and that $\mathcal{F}$ is a family of continuous functions from X to $\mathbb{R}$ with the following property:

$$\forall x \in X, \ \exists \text{ a constant } M_x \text{ such that } |f(x)| \le M_x \text{ for all } f \in \mathcal{F}$$

Prove that there exists a nonempty open set U and a constant M (independent of x) such that $|f(x)| \le M$ for all $x \in U$ and all $f \in \mathcal{F}$.

This result is called the Uniform Boundedness Principle.

Hint: Let $E_k = \{x \in X : |f(x)| \le k$ for all $f \in \mathcal{F}\}$. Use the Baire Category Theorem.

E16. Suppose $f\colon (X, \mathcal{T}) \to (Y, d)$, where (Y, d) is a metric space where d is a metric. For each $x \in X$, define

$$\omega_f(x) \;=\; \text{"the oscillation of } f \text{ at } x\text{"} = \inf\{\text{diam}(f[N])\colon N \text{ is a neighborhood of } x\}$$

a) Prove that f is continuous at a if and only if $\omega_f(a) = 0$.

b) Prove that for $n \in \mathbb{N}$, $\{x \in X : \omega_f(x) < \frac{1}{n}\}$ is open in X.

c) Prove that $\{x \in X : f$ is continuous at $x\}$ is a G_δ-set in X.

Note: Since $\mathbb{Q}$ is not a G_δ-set in $\mathbb{R}$, a continuous function $f : \mathbb{R} \to \mathbb{R}$ cannot have $\mathbb{Q}$ as its set of points of continuity. On the other hand, $\mathbb{P}$ is a G_δ-set in $\mathbb{R}$ and, from analysis, you should know an example of a continuous function $f : \mathbb{R} \to \mathbb{R}$ that is continuous at each $p \in \mathbb{P}$ and discontinuous at each point $q \in \mathbb{Q}$.

d) Prove that there cannot exist a function $f : \mathbb{R} \to (0, \infty)$ such that for all $x \in \mathbb{Q}$ and all $y \in \mathbb{P}$, $f(x)f(y) \leq \mid x - y \mid$. *Hint for d): Suppose f exists.*

i) First prove that if (x_n) is a sequence of rationals converging to an irrational, then $(f(x_n)) \to 0$ and, likewise, if (y_n) is a sequence of irrationals converging to a rational, then $(f(y_n)) \to 0$.

ii) Define a new function $g : \mathbb{R} \to \mathbb{R}$ by $g \mid \mathbb{Q} = 0$ and $g(y) = f(y)$ for y irrational. Examine the set of points where g is continuous.)

E17. There is no continuous function $f : \mathbb{R} \to \mathbb{R}$ for which the set of points of continuity is $\mathbb{Q}$. The reason (*see problem E16*) ultimately depends on the Baire Category Theorem. Find the error in the following "more elementary proof."

Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuous and that f is continuous at x iff $x \in \mathbb{Q}$.
Then $\mathbb{Q} = \{x \in \mathbb{R} : \omega_f(x) = 0\} = \bigcap_{n=1}^{\infty} U_n$, where $U_n = \{x \in \mathbb{R} :\ \omega_f(x) < \frac{1}{n}\}$.

Each $U_n \supseteq \mathbb{Q}$, and U_n can be written as a countable union of disjoint open intervals. If (a, b) and (c, d) are consecutive intervals in U_n, then $b = c$ – or else there would be a rational in (b, c). Therefore $\mathbb{R} - U_n$ consists only of the endpoints of some disjoint open intervals. Therefore $\mathbb{R} - U_n$ is countable.

It follows that $\{x \in \mathbb{R} : f$ is not continuous at $x\} = \mathbb{R} - \bigcap_{n=1}^{\infty} U_n = \bigcup_{n=1}^{\infty} \mathbb{R} - U_n$ is countable. Therefore f is continuous on more than just the points in $\mathbb{Q}$.

E18. For each irrational p, construct an equilateral triangle T_p (including its interior) in $\mathbb{R}^2$ with one vertex at $(p, 0)$ and its opposite side above and parallel to the x-axis. Prove that $\bigcup\{T_p : p \in \mathbb{P}\}$ must contain an "open box" of the form $(a, b) \times (0, \frac{1}{k})$ for some $a, b \in \mathbb{R}$ and some $k \in \mathbb{N}$. In the hypothesis, we could weaken "equilateral" to read " ... " ?
(*Hint: Consider $N_k = \{p \in \mathbb{P} : T_p$ has height $\geq \frac{1}{k}\}$.*)

E19. Suppose $f : \mathbb{R} \to \mathbb{R}$. Prove that if f is discontinuous at every irrational $p \in \mathbb{P}$, then there must exist an interval $I = (a, b)$ such that f is discontinuous at every point in I.

E20. $(X, \mathcal{T})$ is "absolutely open" if whenever $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is a homeomorphism (into), then $f[X]$ must be open in Y. Find all absolutely open spaces.

8. Compactness

Compactness is one of the most powerful topological properties. Formally, it is a stronger version of the Lindelöf property (although it preceded the Lindelöf property historically).

Compact spaces are relatively simple to work with because of the "rule of thumb" that "compact spaces often act like finite spaces."

Definition 8.1 A topological space $(X, \mathcal{T})$ is called compact if every open cover of X has a finite subcover.

Example 8.2

1) A finite space is compact.

2) Any set X with the cofinite topology is compact – because any one nonempty open set covers X except for perhaps finitely many points.

3) An infinite discrete space is not compact – because the cover consisting of all singleton sets has no finite subcover. In particular, $\{\frac{1}{n} : n \in \mathbb{N}\}$ is not compact.

4) The space $X = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$ is compact: if $\mathcal{U}$ is any open cover, and $0 \in U \in \mathcal{U}$, then the set U covers X except for perhaps finitely many points.

5) $\mathbb{R}$ is not compact since $\mathcal{U} = \{(-n, n) : n \in \mathbb{N}\}$ has no finite subcover.

Definition 8.3 A family $\mathcal{F}$ of sets has the finite intersection property (FIP) if every finite subfamily of $\mathcal{F}$ has nonempty intersection.

It is sometimes useful to have a characterization of compact spaces stated in terms of closed sets, and we can get one using FIP. As you read the proof of Theorem 8.4, you should see that the characterization is nothing more a restatement of the definition of compactness that uses complements to "convert" to closed sets.

Theorem 8.4 The following are equivalent for any topological space $(X, \mathcal{T})$

1) X is compact
2) every family $\mathcal{F}$ of closed sets in X with the finite intersection property also has $\bigcap \mathcal{F} \neq \emptyset$.

Proof 1) $\Rightarrow$ 2) Suppose X is compact and that $\mathcal{F}$ is a family of closed sets with FIP. Let $\mathcal{U} = \{X - F : F \in \mathcal{F}\}$. For any $F_1, F_2, ..., F_n \in \mathcal{F}$, the FIP tells us that $X \neq X - \bigcap_{i=1}^{n} F_n$ $= (X - F_1) \cup ... \cup (X - F_n)$. In other words, no finite subcollection of $\mathcal{U}$ covers X. Since X is compact, $\mathcal{U}$ cannot be a cover of cover X, that is, $X - \bigcup \mathcal{U} = \bigcap \{X - U : U \in \mathcal{U}\}$ $= \bigcap \mathcal{F} \neq \emptyset$.

2) $\Rightarrow$ 1) The proof of the converse is similar and left as an exercise. ∙

Theorem 8.5 For any space $(X, \mathcal{T})$

a) If X is compact and F is closed in X, then F is compact.
b) If X is a Hausdorff space and K is a compact subset, then K is closed in X.

Proof a) *The proof is like the proof that a closed subspace of a Lindelöf space is Lindelöf. (Theorem III.3.10)* Suppose F is a closed set in a compact space X, and let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ be a cover of F by sets open in F. For each α, pick an open set V_α in X such that $U_\alpha = V_\alpha \cap F$. Since F is closed, the collection $\mathcal{V} = \{X - F\} \cup \{V_\alpha : \alpha \in A\}$ is an open cover of X. Therefore we can find $V_{\alpha_1}, ..., V_{\alpha_n}$ so that $\{X - F, V_{\alpha_1}, ..., V_{\alpha_n}\}$ covers X. Then $\{U_{\alpha_1}, ...U_{\alpha_n}\}$ covers F, so F is compact.

Note: the definition of compactness requires that we look at a cover of F using sets open in F, but it should be clear that it is equivalent to look at a cover of F by sets open in X.

b) Suppose K is a compact set in a Hausdorff space X. Let $y \in X - K$. For each $x \in K$, we can choose disjoint open sets U_x and V_x in X where $x \in U_x$ and $y \in V_x$. Then $\mathcal{U} = \{U_x : x \in K\}$ covers K, so there are finitely many points $x_1, ..., x_n \in K$ such that $U_{x_1}, ..., U_{x_n}$ covers K. Then $y \in V = V_{x_1} \cap ... \cap V_{x_n} \subseteq X - K$, so $X - K$ is open and K is closed. •

Notes 1) Reread the proof of part b) assuming K is a finite set. This highlights how "compactness" has been used in place of "finiteness" and illustrates the rule of thumb that compact spaces often behave like finite spaces.

2) The "Hausdorff" hypothesis cannot be omitted in part b). For example, suppose $|X| > 1$ and that d is the trivial pseudometric on X. Then every singleton set $\{x\}$ is compact but not closed.

3) Compactness is a clearly a topological property, so part b) implies that if a compact space X is homeomorphic to a subspace K of a Hausdorff space Y, then K is closed in Y. So we can say that a compact space is "absolutely closed" – among Hausdorff spaces, "it's closed wherever you put it." In fact, the converse is also true among Hausdorff spaces – a space which is absolutely closed is compact – but we do not have the machinery to prove that now.

Corollary 8.6 A compact metric space (X, d) is complete.

Proof (X, d) is a dense subspace of its completion $(\widetilde{X}, \widetilde{d})$. But (X, d) is compact, so X must be closed in $\widetilde{X}$. Therefore $(X, d) = (\widetilde{X}, \widetilde{d})$, so (X, d) is complete. •
(*Note: this proof doesn't work for pseudometric spaces (why?). But is Corollary 8.6 still true for pseudometric spaces? Would a proof using the Cantor Intersection Theorem work?*)

You may have seen some different definition of compactness, perhaps in an analysis course. In fact there are several different "kinds of compactness" and, in general they are not equivalent. But we will see that they are all equivalent in certain spaces – for example, in $\mathbb{R}^n$.

Definition 8.7 A topological space $(X, \mathcal{T})$ is called

sequentially compact	if every sequence in X has a convergent subsequence in X
countably compact	if every countable open cover of X has a finite subcover (*therefore "Lindelöf + countably compact = compact"*)
pseudocompact	if every continuous $f : X \to \mathbb{R}$ is bounded (*Check that this is equivalent to saying that every continuous real-valued function on X assumes both a maximum and a minimum value*).

We want to look at the relations between these different varieties of compactness.

Lemma 8.8 If every sequence in $(X, \mathcal{T})$ has a cluster point, then every infinite set in X has a limit point. The converse is true if $(X, \mathcal{T})$ is a T_1-space – that is, if all singleton sets $\{x\}$ are closed in X.

Proof Suppose A is an infinite set in X. Choose a sequence (a_n) of distinct terms in A and let x be a cluster point of (a_n). Then every neighborhood N of x contains infinitely many a_n's, so $N \cap (A - \{x\}) \neq \emptyset$. Therefore x is a limit point of A.

Suppose X is a T_1-space in which every infinite set has a limit point. Let (x_n) be a sequence in X. We want to show that (x_n) has a cluster point. Without loss of generality, we may assume that the terms of the sequence are distinct (why?), so that $A = \{x_n : n \in \mathbb{N}\}$ is infinite. Let x be a limit point of the set A. We claim x is a cluster point for (x_n).

Suppose U is an open set containing x and $n \in \mathbb{N}$. Let $V = U - \{x_1, ..., x_n\} \cup \{x\}$. Since $\{x_1, ..., x_n\}$ is closed, V is open and $x \in V \subseteq U$. Then $V \cap (A - \{x\}) \neq \emptyset$, so $U \cap (A - \{x\}) \neq \emptyset$. Therefore U contains a term x_k for some $k > n$, so x is a cluster point of (x_n). ●

Example 8.9 The "T_1" hypothesis in the second part of Lemma 8.8 cannot be dropped. Let $X = \{\frac{1}{n} : n \in \mathbb{N}\} \cup \{p_n : n \in \mathbb{N}\}$ where the p_n's are distinct points and $p_n \neq \frac{1}{k}$ for all $n, k \in \mathbb{N}$. The idea is to make $d(p_n, x_n) = 0$ for each n so that p_n is a sort of "double" for $\frac{1}{n}$. So we define

$$\begin{cases} d(p_n, x_n) = 0 \\ d(\frac{1}{n}, \frac{1}{m}) = |\frac{1}{n} - \frac{1}{m}| = d(p_n, p_m) \text{ if } m \neq n \\ d(p_n, \frac{1}{m}) = |\frac{1}{n} - \frac{1}{m}| \text{ if } m \neq n \end{cases}$$

It is easy to check that d is a pseudometric on X. In (X, d), every nonempty set A (finite or infinite) has a limit point – because if $x \in A$, then its "double" is a limit point of A. But the sequence $(\frac{1}{n})$ has no cluster point in (X, d).

Theorem 8.10 $(X, \mathcal{T})$ is countably compact iff every sequence in X has a cluster point (So, by Lemma 8.8, a T_1-space $(X, \mathcal{T})$ is countably compact iff every infinite set in X has a limit point).

Proof Suppose $(X, \mathcal{T})$ is countably compact and consider any sequence (x_n). Let T_n = the "n^{th} tail" of $(x_n) = \{x_k : k \geq n\}$. We claim that $\bigcap_{n=1}^{\infty} \text{cl}\, T_n \neq \emptyset$.

> Assume not. Then every x is in $(X - \text{cl}\, T_n)$ for some n, so $\{X - \text{cl}\, T_n : n \in \mathbb{N}\}$ is a countable open cover of X. Since X is countably compact, there exists an N such that $X = (X - \text{cl}\, T_1) \cup ... \cup (X - \text{cl}\, T_N)$. Taking complements gives $\emptyset = \bigcap_{n=1}^{N} \text{cl}\, T_n = \text{cl}\, T_N$, which is impossible.

Let $x \in \bigcap_{n=1}^{\infty} \text{cl}\, T_n$ and let N be a neighborhood of x. Then $N \cap T_n \neq \emptyset$ for every n, so N contains an x_k for arbitrarily large values of n. Therefore x is a cluster point of (x_n).

Conversely, suppose X is not countably compact. Then X has a countable open cover $\{U_1, ..., U_n, ...\}$ with no finite subcover. For each n, pick a point $x_n \in X - \bigcup_{i=1}^{n} U_i$. Then the sequence (x_n) has no cluster point. To see this, pick any $x \in X$; we know x is in some U_n. This U_n is a neighborhood of x and $x_m \notin U_n$ for $m > n$. ∙

The next theorem tells us some connections between the different types of compactness.

Theorem 8.11 In any space $(X, \mathcal{T})$, the following implications are true:

$$\left\{\begin{array}{l} X \text{ is compact} \\ \qquad \text{or} \\ X \text{ is sequentially compact} \end{array}\right. \Rightarrow X \text{ is countably compact} \Rightarrow X \text{ is pseudocompact}$$

Proof It is clear from the definitions that a compact space is countably compact.

Suppose X is sequentially compact. Then each sequence (x_n) in X has a subsequence that converges to some point $x \in X$. Then x is a cluster point of (x_n). From Lemma 8.8 we conclude that X is countably compact.

Suppose X is countably compact and that $f : X \to \mathbb{R}$ is continuous. The sets $U_n = f^{-1}[(-n, n)] = \{x \in X : -n < f(x) < n\}$ form a countable open cover of X, so for some N, the sets $U_1, ..., U_N$ cover X. Since $U_1 \subseteq U_2 \subseteq ... \subseteq U_N$, this implies that $U_N = X$. So $-N < f(x) < N$ for all $x \in X$ – in other words, f is bounded, so X is pseudocompact. ∙

In general, no other implications hold among these four types of compactness, but we do not have the machinery to provide counterexamples now. (*See Corollary 8.5 in Chapter VIII and Example 6.5 in Chapter X*). We will prove, however, that they are all equivalent in any pseudometric space (X, d). Much of the proof is developed in the following sequence of lemmas – some of which have intrinsic interest of their own.

Lemma 8.12 If X is first countable, then X is sequentially compact iff X is countably compact (so, in particular, sequential and countable compactness are equivalent in pseudometric spaces.)

Proof We know from Theorem 8.11 that if X is sequentially compact, then X is countably compact. Conversely, assume X is countably compact and that (x_n) is a sequence in X. By Theorem 8.10, we know (x_n) has a cluster point, x. Since X is first countable, there is a subsequence$(x_{n_k}) \to x$ (*see Theorem III.10.6*). Therefore X is sequentially compact. ∙

Definition 8.13 A pseudometric space (X, d) is called totally bounded if, for each $\epsilon > 0$, X can be covered by a finite number of ϵ-balls. If $X \neq \emptyset$, this is the same as saying that for $\epsilon > 0$, there exist $x_1, x_2, ..., x_n \in X$ such that $X = B_\epsilon(x_1) \cup ... \cup B_\epsilon(x_n)$.

More informally: a totally bounded pseudometric space is one that can be kept under full surveillance using a finite number of policemen with an arbitrary degree of nearsightedness.

Example 8.14

1) Neither $\mathbb{R}$ nor $\mathbb{N}$ (with the usual metric) is totally bounded since neither can be covered by a finite number of 1-balls.

2) A compact pseudometric space (X, d) is totally bounded: for any $\epsilon > 0$, we can pick a finite subcover from the open cover $\mathcal{U} = \{B_\epsilon(x) : x \in X\}$.

Lemma 8.15 If (X, d) is countably compact, then (X, d) is totally bounded.

Proof If $X = \emptyset$, then X is totally bounded, so assume $X \neq \emptyset$. If (X, d) is not totally bounded, then for some $\epsilon > 0$, no finite collection of ϵ-balls can cover X. Choose any point $x_1 \in X$. Then we can choose a point x_2 so that $d(x_1, x_2) \geq \epsilon$ (or else $B_\epsilon(x_1)$ would cover X).

We continue inductively. Suppose we have chosen points $x_1, x_2, ..., x_n$ in X such that $d(x_i, x_j) \geq \epsilon$ for each $i \neq j$, $i, j = 1, ..., n$. Then we can choose a point x_{n+1} at distance $\geq \epsilon$ from each of $x_1, ..., x_n$ $-$ because otherwise the ϵ-balls centered at $x_1, ..., x_n$ would cover X.

The sequence (x_n) chosen in this way cannot have a cluster point because, for any x, $B_{\frac{\epsilon}{2}}(x)$ contains at most one x_n. Therefore X is not countably compact. •

Lemma 8.16 A totally bounded pseudometric space (X, d) is separable.

Proof For each n, choose a finite number of $\frac{1}{n}$-balls that cover X and let D_n be the (finite) set that contains the centers of these balls. For any $y \in X$, $d(x, y) < \frac{1}{n}$ for some $x \in D_n$. This means that $D = \bigcup_{n=1}^\infty D_n$ is dense. Since D is countable, X is separable. •

Theorem 8.17 In a pseudometric space (X, d), the properties of compactness, countable compactness, sequential compactness and pseudocompactness are equivalent.

Proof We already have the implications from Theorem 8.11.

If (X, d) is countably compact, then (X, d) is totally bounded (*see Lemma 8.15*) and therefore separable (*see Lemma 8.16*). But a separable pseudometric space is Lindelöf *(see Theorem III.6.5*) and a countably compact Lindelöf space is compact. Therefore compactness and countable compactness are equivalent in (X, d).

We observed in Lemma 8.12 that countable compactness and sequential compactness are equivalent in (X, d).

Since a countably compact space is pseudocompact, we complete the proof by showing that if (X, d) is not countably compact, then (X, d) is not pseudocompact. This is the only part of the proof that takes some work. (*A bit of the maneuvering in the proof is necessary because d is a pseudometric. If d is actually a metric, some minor simplifications are possible.*)

If (X, d) is not countably compact, we can choose a sequence (x_n) with no cluster point (*see Theorem 8.10*). In fact, we can choose (x_n) so that all the x_n's are distinct, and $d(x_m, x_n) > 0$ if $m \neq n$ (*why?*). Then we can find open sets U_n such that i) $x_n \in U_n$, ii) $U_n \cap U_m = \emptyset$ for $m \neq n$, and so that iii) $\operatorname{diam} U_n \to 0$ as $n \to \infty$.

Here is a sketch for finding the U_n's; the details are left to check as an exercise. First check that for any $a, b \in X$ and positive reals r_1, r_2 : if $r_1 + r_2 < d(a, b)$ then $B_{r_1}(a)$ and $B_{r_2}(b)$ are disjoint – in fact, they have disjoint closures.

Since (x_n) has no cluster point we can find, for each n, a ball $B_{\delta_n'}(x_n)$ that contains no other x_m – that is, $d(x_n, x_m) \geq \delta_n'$ for all $m \neq n$. Let $\delta_n = \delta_n'/2$. Then for $m \neq n$, we have $x_m \notin B_{\delta_n}(x_n)$ and $\delta_n < d(x_n, x_m)$. Of course, two of these balls might overlap. To get the U_n's we want, we shrink these balls (choose smaller radii ϵ_n to replace the δ_n) to eliminate any overlap. We define the ϵ_n's inductively:

Let $\epsilon_1 = \delta_1 < d(x_1, x_2)$.
Pick $\epsilon_2 > 0$ so that

$$\begin{aligned} \epsilon_1 + \epsilon_2 &< d(x_1, x_2) \text{ and} \\ \epsilon_2 &< \delta_2 \end{aligned}$$

Since

$$\begin{aligned} &\epsilon_1 = \delta_1 < d(x_1, x_3) \text{ and} \\ &\epsilon_2 < \delta_2 < d(x_2, x_3)\text{, we can} \end{aligned}$$

pick $\epsilon_3 > 0$ so that

$$\begin{aligned} \epsilon_1 + \epsilon_3 &< d(x_1, x_3) \text{ and} \\ \epsilon_2 + \epsilon_3 &< d(x_2, x_3) \text{ and} \\ \epsilon_3 &< \delta_3 \end{aligned}$$

Continue in this way. At the n^{th} step, we get a new ball $B_{\epsilon_n}(x_n)$ for which

if $m \neq n$, $x_m \notin B_{\epsilon_n}(x_n)$ (because $\epsilon_n < \delta_n$) and
$B_{\epsilon_n}(x_n) \cap B_{\epsilon_j}(x_j) = \emptyset$ for all $j < n$ (because $\epsilon_j + \epsilon_n < d(x_j, x_n)$

Finally, when we choose ϵ_n at each step, we can also add the condition that $\epsilon_n < \frac{1}{n}$, so that $\text{diam}(B_{\epsilon_n}(x_n)) \to 0$.

Then we can let $U_n = B_{\epsilon_n}(x_n)$.

For each n, define $f_n : X \to \mathbb{R}$ by $f_n(x) = \frac{nd(x, X - U_n)}{d(x_n, X - U_n)}$. Since $x_n \in U_n$ and $X - U_n$ is closed, the denominator of $f_n(x)$ is not 0. Therefore f_n is continuous, and $f_n(x_n) = n$. Define $f(x) = \sum_{n=1}^{\infty} f_n(x)$. *(For any $x \in X$, this series converges because x is in at most one U_n and therefore at most one term $f_n(x) \neq 0$.)* Then f is unbounded on X because $f(x_n) = f_n(x_n)$ $= n$. So we are done if we can show that f is continuous at each in X.

Let $a \in X$. We claim that there is an open set V_a containing a such that $V_a \cap U_n \neq \emptyset$ for at most finitely many n's.

If $d(a, x_n) = 0$ for some x_n, we can simply let $V_a = U_n$.

Suppose $d(a, x_n) > 0$ for all n. Since a is not a cluster point of (x_n), we can choose $\epsilon > 0$ so that $B_\epsilon(a)$ contains none of the x_n's. For this ϵ, choose N so that if $n > N$, then diam $U_n < \frac{\epsilon}{2}$. Then $B_{\frac{\epsilon}{2}}(a) \cap U_n = \emptyset$ for $n > N$ (*if* $n > N$ *and* $z \in B_{\frac{\epsilon}{2}}(a) \cap U_n$, *then* $d(a, x_n) \leq d(a, z) + d(z, x_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$). Then let $V_a = B_{\frac{\epsilon}{2}}(a)$.

Then $f_n|V_a$ is identically 0 for all but finitely many n. Therefore $f|V_a$ is really only a finite sum of continuous functions, so $f|V_a : V_a \to \mathbb{R}$ is continuous. Suppose W is any neighborhood of $f(a)$. Then there is an open set U in V_a containing a for which $(f|V_a)[U] \subseteq W$. But V_a is open in X, so U is also open in X, and $f[U] = (f|V_a)[U] \subseteq W$. So f is continuous at a. •

We now explore some properties of compact metric spaces.

Theorem 8.18 Suppose A is a compact subset of a metric space (X, d). Then A is closed and bounded (that is, A has finite diameter).

Proof A compact subset of the Hausdorff space must be closed (*see Theorem 8.5b*).

If $A = \emptyset$, then A is certainly bounded. If $A \neq \emptyset$, choose a point $a_0 \in A$ and let $f : A \to \mathbb{R}$ by $f(a) = d(a, a_0)$. Then f is continuous and f is bounded (since X is pseudocompact): say $|f(x)| < M$ for all $x \in A$. But that means that $A \subseteq B_M(a_0)$, so A is a bounded set . •

Caution: the converse of Theorem 8.18 is false. For example, suppose d is the discrete unit metric on an infinite set X. Every subset of (X, d) is both closed and bounded, but an infinite subspace of X is not compact. However, the converse to Theorem 8.18 is true in $\mathbb{R}^n$ (*with the usual metric*), as you may remember from analysis.

Theorem 8.19 Suppose $A \subseteq \mathbb{R}^n$. The A is compact iff A is closed and bounded.

Proof Because of Theorem 8.18, we only need to prove that if A is closed and bounded, then A is compact. First consider the case $n = 1$. Then A is a closed subspace of some interval $I = [a, b]$ and it is sufficient to show that I is compact. To do this, consider any sequence (x_n) in I and choose a monotone subsequence (x_{n_k}) using Lemma 2.10. Since (x_{n_k}) is bounded, we know that $(x_{n_k}) \to$ some point $r \in I$ (*see Lemma 2.9*). Since (x_n) has a convergent subsequence, I is sequentially compact and therefore compact.

When $n > 1$, the proof is similar. We illustrate for $n = 2$. If A is a closed bounded set in $\mathbb{R}^2$, then A is a closed subspace of some closed box of the form $I = [a, b] \times [c, d]$. Therefore it is sufficient to prove that I is compact. To do this, consider any sequence (x_n, y_n) in I. Since $[a, b]$ is compact, the sequence (x_n) has a subsequence (x_{n_k}) that converges to some point $x \in [a, b]$. Now consider the subsequence (x_{n_k}, y_{n_k}) in I. Since $[c, d]$ is compact, the sequence (y_{n_k}) has a subsequence $(y_{n_{k_l}})$ that converges to some point $y \in [c, d]$. But in $\mathbb{R}^2$, $(u_n, v_n) \to (u, v)$ iff $(u_n) \to u$ and $(v_n) \to v$ in $\mathbb{R}$. So $(x_{n_{k_l}}, y_{n_{k_l}}) \to (x, y) \in I$. This shows that I is sequentially compact and therefore compact.

For $n > 2$, a similar argument clearly works – the argument just involves "taking subsequences" n times. •

9. Compactness and Completeness

We already know that a compact metric space (X, d) is complete. Therefore all the "big" theorems that we proved for complete spaces are true, in particular, for compact metric spaces $-$ for example Cantor's Intersection Theorem, the Contraction Mapping Theorem and the Baire Category Theorem (which implies that a nonempty compact metric space is second category in itself).

Of course, a complete space (X, d) need not be compact: for example, $\mathbb{R}$. We want to see the exact relationship between compactness and completeness.

We begin with a couple of preliminary results.

Theorem 9.1 If (X, d) is totally bounded, then (X, d) is bounded.

Proof Let $\epsilon = 1$. Pick points $x_1, ..., x_n$ so that $X = B_1(x_1) \cup ... \cup B_1(x_n)$ and let $M = \max\{d(x_i, x_j) : i, j = 1, ..., n\}$. For any pair of points $x, y \in X$, we have $x \in B_1(x_i)$ and $y \in B_1(x_j)$ for some i, j, so that

$$d(x, y) \leq d(x, x_i) + d(x_i, y_j) + d(y_j, y) < 1 + M + 1 = M + 2.$$

Therefore (X, d) has finite diameter $-$ that is, (X, d) is bounded. $\bullet$

Notice that the converse to the theorem is false: "total boundedness" is a stronger condition than "boundedness." For example, let $d'(x, y) = \min\{1, |x - y|\}$ on $\mathbb{R}$. Then diam $(\mathbb{R}, d') = 1$ so $(\mathbb{R}, d')$ is bounded. Because $d'(x, y) = |x - y|$ when $|x - y| < 1$, we have that $B_1^{d'}(x) = (x - 1, x + 1)$. Since $\mathbb{R}$ cannot be covered by a finite number of these 1-balls, $(\mathbb{R}, d')$ is not totally bounded.

Theorem 9.2 If (X, d) is totally bounded and $A \subseteq X$, then (A, d) totally bounded: that is, a subspace of a totally bounded space is totally bounded.

Proof Let $\epsilon > 0$ and choose $x_1, ..., x_n$ so that $X = B_{\frac{\epsilon}{2}}(x_1) \cup ... \cup B_{\frac{\epsilon}{2}}(x_n)$. Of course these balls also cover A. But to show that (A, d) is totally bounded, the definition requires us to show that we can cover A with a finite number of ϵ-balls <u>centered at points in A</u>.

Let $J = \{j : B_{\frac{\epsilon}{2}}(x_j) \cap A \neq \emptyset\}$ and, for each $j \in J$, pick $a_j \in B_{\frac{\epsilon}{2}}(x_j) \cap A$. We claim that the balls $B_\epsilon^A(a_j)$ cover A. In fact, if $a \in A$, then $a \in B_{\frac{\epsilon}{2}}(x_j)$ for some j; and of course a_j is also in $B_{\frac{\epsilon}{2}}(x_j)$. Then $d(a, a_j) \leq d(a, x_j) + d(x_j, a_j) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$, so $a \in B_\epsilon^A(a_j)$. $\bullet$

The next theorem gives the exact connection between compactness and completeness. It is curious because it states that "compactness" (a topological property) is the "sum" of two nontopological properties.

Theorem 9.3 (X,d) is compact iff (X,d) is complete <u>and</u> totally bounded.

Proof $(\Rightarrow)$ We have already seen that the compact space (X,d) is totally bounded and complete. (*For completeness, here is a fresh argument: let* (x_n) *be a Cauchy sequence in* (X,d)*. Since* X *is countably compact,* (x_n) *has a cluster point* x *in* X*. But a Cauchy sequence must converge to a cluster point. So* (X,d) *is complete.*)

$(\Leftarrow)$ Let (x_n) be a sequence in (X,d). We will show that (x_n) has a Cauchy subsequence which (by completeness) must converge. That means (X,d) is sequentially compact and therefore compact.

Since (X,d) is totally bounded, we can cover X with a finite number of 1-balls, and one of them – call it B_1 – must contain x_n for infinitely many n. Since (B_1,d) is totally bounded (*see Theorem 9.2*) we can cover B_1 with a finite number of $\frac{1}{2}$-balls and one of them – call it $B_2 \subseteq B_1$ – must contain x_n for infinitely many n.

Continue inductively in this way. At the induction step, suppose we have already defined $\frac{1}{k}$-balls $B_1 \supseteq B_2 \supseteq ... \supseteq B_k$ where each ball contains x_n for infinitely many n. Since (B_k,d) is totally bounded, we can cover it with a finite number of $\frac{1}{k+1}$-balls, one of which – call it $B_{k+1} \subseteq B_k$ – must contain x_n for indefinitely many n. This inductively defines an infinite descending sequence of balls $B_1 \supseteq B_2 \supseteq ... \supseteq B_k \supseteq ...$ where each B_k contains x_n for infinitely many values of n.

Then we can choose

$$\begin{aligned} &x_{n_1} \in B_1 \\ &x_{n_2} \in B_2, \text{ with } n_2 > n_1 \\ &\vdots \\ &x_{n_k} \in B_k, \text{ with } n_k > n_{k-1} > ... > n_1 \\ &\vdots \end{aligned}$$

The subsequence (x_{n_k}) is Cauchy since $x_{n_l} \in B_k$ for $l \geq k$ and diam $B_k \to 0$. •

Example 9.4 With the usual metric d:

i) $\mathbb{N}$ is complete and not compact (and therefore not totally bounded).

ii) $\{\frac{1}{n} : n \in \mathbb{N}\}$ is totally bounded, because it is a subspace of the totally bounded space $[0,1]$. But it is not compact (and therefore not complete).

iii) $\mathbb{N}$ and $\{\frac{1}{n} : n \in \mathbb{N}\}$ are homeomorphic because both are countable discrete spaces. Total boundedness is not a topological property. *We remark here, without proof, that for every* <u>*separable*</u> *metric space* (X,d)*, there is an* <u>*equivalent*</u> *metric* d' *such that* (X,d') *is totally bounded.*

iv) $\{\frac{1}{n} : n \in \mathbb{N}\} \cup \{0\}$ is compact and is therefore both complete and totally bounded.

During the earlier discussion of the Contraction Mapping Theorem, we defined uniform continuity. For convenience, the definition is repeated here.

Definition 9.5 A function $f:(X,d)\to(Y,d')$ is uniformly continuous if

$$\forall\epsilon\,\exists\delta\,\forall x\,\forall y\in X\ \ (d(x,y)<\delta\Rightarrow d'(f(x),f(y))<\epsilon)$$

Clearly, if $X=\{x_1,...,x_n\}$ is finite, then every continuous function $f:(X,d)\to(Y,s)$ is uniformly continuous. (*For* $\epsilon>0$ *and each* $i=1,...,n$, *we pick the the* δ_i *that works at* x_i *in the definition of continuity. Then* $\delta=min\,\{\delta_1,...\delta_n\}$ *works "uniformly" across the space* X.)

The next theorem generalizes this observation to compact metric spaces (X,d) – and illustrates once more "rule of thumb" that "compact spaces act like finite spaces."

Theorem 9.6 If (X,d) is compact and $f:(X,d)\to(Y,d')$ is continuous, then f is uniformly continuous.

Proof If f is not uniformly continuous, then for some $\epsilon>0$, "no δ works." In particular, for that ϵ, $\delta=\frac{1}{n}$ "doesn't work" – so we can find, for each n, a pair of points u_n and v_n with $d(u_n,v_n)<\frac{1}{n}$ but $d'(f(u_n),f(v_n))\geq\epsilon$. By compactness, there is a subsequence $(u_{n_k})\to x\in X$. The corresponding subsequence $(v_{n_k})\to x$ also because $d(v_{n_k},x)\leq d(v_{n_k},u_{n_k})+d(u_{n_k},x)<\frac{1}{n_k}+d(x_{n_k},x)\to 0$. Therefore, by continuity, we should have $(f(u_{n_k}))\to f(x)$ and also $(f(v_{n_k}))\to f(x)$. But this is impossible since
$d'(f(u_{n_k}),f(v_{n_k}))\geq\epsilon$ for all k. •

Uniform continuity is a strong and useful condition. For example, Theorem 9.6 implies that a continuous function $f:[a,b]\to\mathbb{R}$ must be uniformly continuous. This is one of the reasons why "continuous functions on closed intervals" are so nice to work with in calculus – the following example is a simple illustration:

Example 9.7 Suppose $f:[a,b]\to\mathbb{R}$ be bounded. Choose any partition of $[a,b]$

$$x_0=a<x_1<...<x_{i-1}<x_i<...<x_n=b$$

and let m_i and M_i denote the inf and sup of the set $f[\,[x_{i-1},x_i]\,]$. Let $\Delta_i x=x_i-x_{i-1}$.

The sums $$\sum_{i=1}^{n}m_i\Delta_i x\ \leq\ \sum_{i=1}^{n}M_i\Delta_i x$$

are called the lower and upper sums for f associated with this partition.

t is easy to see that the sup of the lower sums (over all possible partitions) is finite: it is called the lower integral of f on $[a,b]$ and denoted $\underline{\int_a^b} f(x)\,dx$. Similarly, the inf of all the upper sums is finite: it is called the upper integral of f on $[a,b]$ and denoted $\overline{\int_a^b} f(x)\,dx$. It is easy to verify that $\underline{\int_a^b} f(x)\,dx\leq\overline{\int_a^b} f(x)\,dx$. If the two are actually equal, we say f is (Riemann) integrable on $[a,b]$ and we write $\int_a^b f(x)\,dx=\underline{\int_a^b} f(x)\,dx=\overline{\int_a^b} f(x)\,dx$.

Uniform continuity is just what we need to prove that a continuous f on $[a,b]$ is integrable. If f is continuous on $[a,b]$, then f is uniformly continuous. Therefore, for $\epsilon > 0$, we can choose $\delta > 0$ such that $|f(x) - f(y)| < \frac{\epsilon}{b-a}$ whenever $|x - y| < \delta$. Pick a partition of $[a,b]$ for which each $\Delta_i x < \delta$. Then for each i, $M_i - m_i \leq \frac{\epsilon}{b-a}$. Therefore

$$\overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx \leq \sum_{i=1}^{n} M_i \Delta_i x - \sum_{i=1}^{n} m_i \Delta_i x$$
$$= \sum_{i=1}^{n} (M_i - m_i)\,\Delta_i x < \epsilon \sum_{i=1}^{n} \Delta_i x = \frac{\epsilon}{b-a}(b-a) = \epsilon$$

Since $\epsilon > 0$ was arbitrary, we conclude that $\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx$.

10. The Cantor Set

The Cantor set is an example of a compact subspace of $\mathbb{R}$ with a surprising combination of properties. Informally, we can construct the Cantor set C as follows. Begin with the closed interval $[0,1]$ and delete the open "middle third." The remainder is the union of 2 disjoint closed intervals: $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. At the second stage of the construction, we then delete the open middle thirds of each interval – leaving a remainder which is the union of 2^2 disjoint closed intervals: $[0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. We repeat this process of deleting the middle thirds "forever." The Cantor set C is the set of survivors – the points that are never discarded.

Clearly there <u>are</u> some survivors: for example, the endpoint of a deleted middle third – for example, $\frac{1}{9}$ – clearly survives forever and ends up in C.

To make this whole process precise, we name the closed subintervals that remain after each stage. Each interval contains the 2 new remaining intervals below it:

$$\begin{array}{cccccccc}
\multicolumn{4}{c}{[0, \frac{1}{3}]} & \multicolumn{4}{c}{[\frac{2}{3}, 1]} \\
\multicolumn{4}{c}{(F_0)} & \multicolumn{4}{c}{(F_2)} \\
\multicolumn{2}{c}{[0, \frac{1}{9}]} & \multicolumn{2}{c}{[\frac{2}{9}, \frac{1}{3}]} & \multicolumn{2}{c}{[\frac{2}{3}, \frac{7}{9}]} & \multicolumn{2}{c}{[\frac{8}{9}, 1]} \\
\multicolumn{2}{c}{(F_{00})} & \multicolumn{2}{c}{(F_{02})} & \multicolumn{2}{c}{(F_{20})} & \multicolumn{2}{c}{(F_{22})} \\
[0, \frac{1}{27}] & [\frac{2}{27}, \frac{1}{9}] & [\frac{2}{9}, \frac{7}{27}] & [\frac{8}{27}, \frac{1}{3}] & [\frac{2}{3}, \frac{19}{27}] & [\frac{20}{27}, \frac{7}{9}] & [\frac{8}{9}, \frac{25}{27}] & [\frac{26}{27}, 1] \\
(F_{000}) & (F_{002}) & (F_{020}) & (F_{022}) & (F_{200}) & (F_{202}) & (F_{220}) & (F_{222})
\end{array}$$

etc. etc.

At the k^{th} stage the set remaining is the union 2^k closed subintervals, each with length $\frac{1}{3^k}$. We label them $F_{n_1 n_2 \ldots n_k}$ where $(n_1, n_2, \ldots n_k) \in \{0,2\}^k$.

Notice that, for example, $F_0 \supseteq F_{02} \supseteq F_{020} \supseteq F_{0200} \supseteq F_{02002} \supseteq \ldots$. As we go down the chain "toward" the Cantor set, each new 0 or 2 in the subscript indicates whether the next set down is "the remaining left interval" or "the remaining right subinterval." So for every sequence $s = (n_1, n_2, \ldots, n_k, \ldots) \in \{0,2\}^{\mathbb{N}}$, we have

$$F_{n_1} \supseteq F_{n_1 n_2} \supseteq \ldots \supseteq F_{n_1 n_2 \ldots n_k} \supseteq \ldots$$

Since $[0,1]$ is complete, the Cantor Intersection Theorem gives that $\bigcap_{k=1}^{\infty} F_{n_1 n_2,...n_k}$ = a one-point set $\{x_s\}$. The Cantor set C is then defined as

$$C = \{x_s : s \in \{0,2\}^{\mathbb{N}}\} \subseteq [0,1].$$

Comments:

1) If $s \neq s' = (n'_1, n'_2, ..., n'_k, ...) \in \{0,2\}^{\mathbb{N}}$, we can look at the first k for which $n_k \neq n'_k$. Then $x_s \in F_{n_1...n_{k-1}n_k}$ and $x_{s'} \in F_{n_1...n_{k-1},n'_k}$. Since these closed intervals are disjoint, we conclude $x_s \neq x_{s'}$. Therefore each sequence $s \in \{0,2\}^{\mathbb{N}}$ corresponds to a different point in C, so $|C| \geq |\{0,2\}^{\mathbb{N}}| = 2^{\aleph_0} = c$. Since $C \subseteq [0,1]$, we conclude $|C| = c$.

2) An observation we don't really need: each $x \in [0,1]$ can be written using a "ternary decimal" expansion $x = \sum_{n=1}^{\infty} \frac{t_n}{3^n} = 0.t_1 t_2 ... t_k ..._{\text{three}}$, where each $t_k \in \{0,1,2\}$. Just as with base 10 decimals, the expansion for a particular x is not always unique: for example, $\frac{1}{3} = 0.1000..._{\text{three}}$ $= 0.022222..._{\text{three}}$. It's easy to check x has two different ternary expansions iff x is the endpoint of one of the deleted "open middle thirds" in the construction. It is also easy to see that a point is in C iff it has a ternary expansion involving only 0's and 2's. For example, $\frac{1}{3} \in C$ and $\frac{1}{2} \notin C$.

What are some of the properties of C?

i) $[0,1] - C$ is the union of the deleted open "middle third" intervals, so $[0,1] - C$ is open in $[0,1]$. Therefore C is closed in $[0,1]$, so C is a compact (therefore also complete) metric space.

ii) Every point in C is a limit point of C – that is, C has no isolated points. (*Note: such a space is sometimes called dense-in-itself. This is an awkward but well-established term; it is awkward because it means something different from the obvious fact that every space X is a dense subset of itself.*) To see this, suppose $x_s \in C$, where $s = (n_1, n_2, ..., n_k, n_{k+1}, n_{k+2} ...) \in \{0,2\}^{\mathbb{N}}$. Given $\epsilon > 0$, pick k so $\frac{1}{3^k} < \epsilon$. Define n'_{k+1} ($= 0$ or 2) so that $n'_{k+1} \neq n_{k+1}$, and let

$$s' = (n_1, n_2, ..., n_k, n'_{k+1}, n_{k+2} ...) \in \{0,2\}^{\aleph_0}.$$

Then x_s and $x_{s'}$ are distinct points in C, but both are in $F_{n_1,...n_k}$ and $\operatorname{diam} F_{n_1,...n_k} = \frac{1}{3^k} < \epsilon$. Therefore $|x_s - x_{s'}| < \epsilon$, so x_s is not isolated in C.

iii) With the usual metric d, C is a nonempty complete metric space with no isolated points. It follows from Theorem 3.6 that $|C| \geq c$ and, since $C \subseteq [0,1]$, then $|C| = c$. However, we can also see this from the discussion in ii): there are $2^{\aleph_0} = c$ different ways we could redefine the infinite "tail" $(n_{k+1}, ...n_{k+l}, ...)$ of s – and each version produces a different point in C at distance $< \epsilon$ from x_s. From Example 6.13(2), every point in C is a condensation point.

iv) In some ways, C is "big." For example, the Cantor set has as many points as $[0,1]$. Also, since C is complete, it is second category in itself. But C is "small" in other ways. For one thing, C is nowhere dense in $[0,1]$ (and therefore in $\mathbb{R}$). Since C is closed, this simply means that C contains no nonempty open interval.

To see this, suppose $I = (a, b) \subseteq C$. At the k^{th} stage of the construction, we must have $I \subseteq$ exactly one of the $F_{n_1, \ldots n_k}$. (This follows immediately from the definition of an interval.) Therefore $b - a < \frac{1}{3^k}$ for every k, so $a = b$ and $I = \emptyset$.

v) C is also "small" in another sense. In the construction of C, the open intervals that are deleted have total length $\frac{1}{3} + 2\left(\frac{1}{9}\right) + 4\left(\frac{1}{27}\right) + \ldots = \frac{\frac{1}{3}}{1 - \frac{2}{3}} = 1$. In some sense, the "length" of what remains is 0! This is made precise in measure theory, where we say that $[0, 1] - C$ has measure 1 and that C has measure 0. Measure is yet another way, differing from both cardinality and category, to measure the "size" of a set.

The following theorem states that the topological properties of C which we have discussed actually characterize the Cantor set. The theorem can be used to prove that a "Cantor set" obtained by some other construction is actually the same topologically as C.

Theorem 10.1 Suppose A is a nonempty compact subset of $\mathbb{R}$ which is dense-in-itself and contains no nonempty open interval. Then A is homeomorphic to the Cantor set C.

Proof The proof is omitted. (See Willard, *General Topology*.)

It is possible to modify the construction of C by changing "middle thirds" to, say, "middle fifths," or even by deleting "middle thirds" at stage one, "middle fifths" at stage two, etc. Theorem 10.1 can be used to show that the resulting "generalized Cantor set" is homeomorphic to C, but the total lengths of the deleted open intervals may be different: that is, the result is can be a topological Cantor set but with positive measure. "Measure" is not a topological property.

Here's one additional observation that we don't need, but which you might see in an analysis course. As we noted earlier, C consists of the points $x \in [0, 1]$ for which we can write

$$x = \sum_{n=1}^{\infty} \frac{t_n}{3^n} = 0.t_1 t_2 \ldots t_k \ldots_{\text{three}}, \text{ with each } t_n \in \{0, 2\}.$$

We can obviously rewrite this as $x = \sum_{n=1}^{\infty} \frac{2b_n}{3^n}$, where each $b_n \in \{0, 1\}$ and therefore we can define a function $g : C \to [0, 1]$ as follows:

$$\text{for } x \in C, \;\; g(x) = g\left(\sum_{n=1}^{\infty} \frac{2b_n}{3^n}\right) = \sum_{n=1}^{\infty} \frac{b_n}{2^n}.$$

More informally,:

$$g(0.\,2b_1\,2b_2\,2b_3\,\ldots_{\text{three}}) = 0.b_1 b_2 b_3 \ldots_{\text{two}}$$

This mapping g is not one-to-one . For example,

$$\begin{aligned} g\left(\tfrac{7}{9}\right) &= g(0.202222\ldots._{\text{three}}) &&= 0.101111\ldots_{\text{two}} &&= \tfrac{3}{4} \quad \text{and} \\ g\left(\tfrac{8}{9}\right) &= g(0.220000\ldots_{\text{three}}) &&= 0.11000\ldots_{\text{two}} &&= \tfrac{3}{4} \end{aligned}$$

In fact, for $a, b \in C$, we have $g(a) = g(b)$ iff the interval (a, b) is one of the "deleted middle thirds." It should be clear that g is continuous and that if $a < b$ in C, then $g(a) \leq g(b)$ – that is, g is weakly increasing. (*Is g onto?*)

The function g is not defined on $[0, 1] - C$. But $[0, 1] - C$ is, by construction, a union of <u>disjoint</u> open intervals. We can therefore <u>extend</u> the definition of g to a mapping $G : [0, 1] \rightarrow [0, 1]$ in the following simple-minded way:

$$G(x) = \begin{cases} g(x) & \text{if } x \in C \\ g(a) & \text{if } x \in (a, b), \text{ where } (a, b) \text{ is a "deleted middle third"} \end{cases}$$

Since we know $g(a) = g(b)$, this amounts to extending the graph of g horizontally over each "deleted middle third" (a, b) from $(a, g(a))$ to $(b, g(b))$. The result is the graph of the continuous function G.

G is sometimes called the Cantor-Lebesgue function. It satisfies:

i) $G(0) = 0,\ G(1) = 1$
ii) G is continuous
iii) G is (weakly) increasing
iv) at any point x in a "deleted middle third" (a, b), G is differentiable and $G'(x) = 0$.

Recall that C has "measure 0." Therefore we could say "$G'(x) = 0$ almost everywhere" – even though G rises monotonically from 0 to 1!

Note: The technique that we used to extend g works in a similar way for any closed set $F \subseteq \mathbb{R}$ and any continuous function $g : F \rightarrow \mathbb{R}$. Since $\mathbb{R} - F$ is open in $\mathbb{R}$, we know that we can write $\mathbb{R} - F$ as the union of a countable collection of disjoint open intervals I_n which have form (a, b), $(-\infty, b)$ or (a, ∞) We can extend f to a continuous function $F : \mathbb{R} \rightarrow \mathbb{R}$ simply by extending the graph of f over linearly over each I_n :

if $I_n = (a, b)$, then let the graph F over I_n be the straight line segment joining $(a, f(a)$, to $(b, f(b))$.
if $I_n = (-\infty, b)$, then let F have the constant value $f(b)$ on I_n
if $I_n = (a, \infty)$, then let F have the constant value $f(a)$ on I_n,

There is a much more general theorem that implies that whenever F is a closed subset of (X, d), then each $f \in C(F)$ and be extended to a function $F \in C(X)$. In the particular case $X = \mathbb{R}$, proving this was easy because $\mathbb{R}$ is ordered and we completely understand the structure of the open sets in $\mathbb{R}$..

Another curious property of C, mentioned without proof, is that its "difference set" $\{x - y : x, y \in C\} = [-1, 1]$. Although C has measure 0, the difference set in this case has measure 2!

Exercises

E21. Suppose that $f : X \to Y$ is continuous and onto, where X and Y are any topological spaces. Prove If X is compact, then Y is compact. ("*A continuous image of a compact space is compact.*")

E22. a) Suppose that X is compact and Y is Hausdorff. Let $f : X \to Y$ be a continuous bijection. Prove that f is a homeomorphism.

b) Let $\mathcal{T}$ be the usual topology on $[0,1]$ and suppose $\mathcal{T}_1$ and $\mathcal{T}_2$ are two other topologies on $[0,1]$ such that $\mathcal{T}_1 \underset{\neq}{\subseteq} \mathcal{T} \underset{\neq}{\subseteq} \mathcal{T}_2$ Prove that $([0,1], \mathcal{T}_1)$ is not Hausdorff and that $([0,1], \mathcal{T}_2)$ is not compact. (*Hint: Consider the identity map* $i : [0,1] \to [0,1]$.)

c) Is part b) true if $[0,1]$ is replaced by an arbitrary compact Hausdorff space $(X, \mathcal{T})$?

E23. Prove that a nonempty topological space X is pseudocompact iff every continuous $f : X \to \mathbb{R}$ achieves both a maximum and a minimum value.

E24. Suppose $f : (X, d) \to (Y, d')$. Prove that f is continuous iff $f|K$ is continuous for every compact set $K \subseteq X$.

E25. Suppose that A and B are nonempty disjoint closed sets in (X, d) and that A is compact. Prove that $d(A, B) > 0$. Is this necessarily true if both A and B are not compact?

E26. Let X and Y be topological spaces.

a) Prove that X is compact iff every open cover by basic open sets has a finite subcover.

b) Suppose $X \times Y$ is compact. Prove that if $X, Y \neq \emptyset$, then X and Y are compact. (*By induction, a similar statement applies to any finite product.*)

c) Prove that if X and Y are compact, then $X \times Y$ is compact. (*By induction, a similar statement applies to any finite product.*)
(*Hint: for any* $x \in X$, $\{x\} \times Y$ *is homeomorphic to* Y. *Part a) is relevant.*)

d) Point out explicitly why the proof in c) cannot be altered to prove that a product of two countably compact spaces is countably compact. (*An example of a countably compact space* X *for which* $X \times X$ *is not even pseudocompact is given in Chapter X, Example 6.8.*)

Note: In fact, an arbitrary product of compact spaces is compact. This is the "Tychonoff Product Theorem" which we will prove later.

It is true that the product of a countably compact space and a compact space is countably compact. You might trying proving this fact. A very similar proof shows that a product of a compact space and a Lindelöf space is Lindelöf. That proof does not generalize, however, to show that a product of two Lindelöf spaces is Lindelöf. Can you see why?

E27. Prove that if (X, d) is a compact metric space and $|X| = \aleph_0$, then X has infinitely many isolated points.

E28. Suppose X is any topological space and that Y is compact. Prove that the projection map $\pi_X : X \times Y \to X$ is closed (*that is: a projection "parallel to a compact factor" is closed*).

E29. Let X be an uncountable set with the discrete topology $\mathcal{T}$. Prove that there does not exist a totally bounded metric d on X such that $\mathcal{T}_d = \mathcal{T}$.

E30. a) Give an example of a metric space (X, d) which is totally bounded and an isometry from (X, d) into (X, d) which is not onto. *Hint: on the circle S^1, start with a point p and repeatedly rotate it by some fixed angle βon S^1 a*

b) Prove that if (X, d) is totally bounded and f is an isometry from (X, d) into (X, d), then $f[X]$ must be dense in (X, d). *Hint: Given x, $\epsilon > 0$, cover X with finitely many $\frac{\epsilon}{2}$-balls; one of these balls must contain $f^n(x)$ for infinitely many values of n.*

c) Show that a compact metric space cannot be isometric to a proper subspace of itself. *Hint: you might use part b).*

d) Prove that if each of two compact metric spaces is isometric to a subspace of the other, then the two spaces are isometric to each other. *Note: Part d) is an analogue of the Cantor-Schroeder-Bernstein Theorem for compact metric spaces.*

E31. Suppose (X, d) is a metric space and that (X, d') is totally bounded for every metric $d' \sim d$. Must (X, d) be compact?

E32. Let $\mathcal{U}$ be an open cover of the compact metric space (X, d). Prove that there exists a constant $\delta > 0$ such that for all x: $B_\delta(x) \subseteq U$ for some $U \in \mathcal{U}$. (*The number δ is called a Lebesgue number for $\mathcal{U}$.*)

E33. Suppose (X, d) is a metric space with no isolated points. Prove that X is compact iff $d(A, B) > 0$ for every pair of disjoint closed sets $A, B \subseteq X$.

E34. Suppose $A \subseteq \mathbb{R}^n$ (with the usual metric). Prove that A is totally bounded if and only if A is bounded.

Chapter IV Review

Explain why each statement is true, or provide a counterexample.

1. Let A be the set of fixed points of a continuous function $f : X \to X$, where X is a Hausdorff space. A is closed in X.

2. Let S denote the set of all Cauchy sequences in $\mathbb{Q}$ that converge to a point in $\mathbb{R}$. Then $| S | = c$.

3. For a metric space (X, d): if every continuous function $f : X \to \mathbb{R}$ assumes a minimum value, then every infinite set in X has a limit point.

4. There exists a continuous function $f : \mathbb{R} \to \mathbb{R}$ for which the Cantor set is the set of fixed points.

5. If X has the cofinite topology, then every subspace of X is pseudocompact.

6. Let $[0, 1]$ have the subspace topology from the Sorgenfrey line. Then $[0, 1]$ is countably compact.

7. Let $\mathfrak{S} = \{A \subseteq \mathbb{R} : A \text{ is first category in } \mathbb{R}\}$, and let $\mathcal{T}$ be the topology on $\mathbb{R}$ for which $\mathfrak{S}$ is a subbase. Then $(\mathbb{R}, \mathcal{T})$ is a Baire space.

8. Every sequence in $[0, 1]^2$ has a Cauchy subsequence.

9. If A is a dense subspace of (X, d) and every Cauchy sequence in A converges to some point in X, then (X, d) is complete.

10. If $\mathcal{T}$ is the cofinite topology on $\mathbb{N}$, then $(\mathbb{N}, \mathcal{T})$ has the fixed point property.

11. Every sequence in $[0, 1)$ has a Cauchy subsequence.

12. Let $Z = (0, \frac{1}{3}) \cap C$, where C is the Cantor set. Then Z is completely metrizable.

13. The Sorgenfrey plane is countably compact.

14. If every sequence in the metric space (X, d) has a convergent subsequence, then every continuous real valued function on X must have a minimum value.

15. Suppose (X, d) is complete and $f : X \to X$. The set $C = \{x \in \mathbb{R} : f \text{ is continuous at } x\}$ is second category in itself.

16. In the space $C([0, 1])$, with the metric ρ of uniform convergence, the subset of all polynomials is first category.

17. If O is nonempty and open in $(X, \mathcal{T})$, then O is <u>not</u> nowhere dense.

18. A nonempty nowhere dense subset of $\mathbb{R}$ must contain an isolated point.

19. There are exactly c nowhere dense subsets of $\mathbb{R}$.

20. Suppose B is nonempty subset of $\mathbb{Q}$ with no isolated points. B cannot be completely metrizable.

21. Suppose (X, d) is a countable complete metric space. If $A \subseteq X$, then (A, d) may not be complete, but A is completely metrizable.

22. If A is first category in X and $B \subseteq A$, then B is first category in A.

23. There are c different metrics d on $\mathbb{Q}$, each equivalent to the usual metric, for which $(\mathbb{Q}, d)$ is complete.

24. If $\mathbb{N}$ has the cofinite topology, then every infinite subset is sequentially compact.

25. Suppose $f\colon [0,1]^2 \to \mathbb{R}$ and $A = \{p \in (0,1)^2\colon\ \frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ both exist at $p\}$. Then A, with its usual metric, is totally bounded.

26. A space $(X, \mathcal{T})$ is a Baire space iff X is second category in itself.

27. Suppose $f\colon [0,1] \to \mathbb{R}$ is continuous. Let E be the subset of the Cantor set C that consists of the endpoints of the open intervals deleted from $[0,1]$ in the construction of C.
If $f[E] \subseteq [0, 0.1]$. Then $f[C] \subseteq [0, 0.1]$.

28. Let d be a metric on the irrationals $\mathbb{P}$ which is equivalent to the usual metric and such that $(\mathbb{P}, d)$ is complete. Then $(\mathbb{P}, d)$ must be totally bounded.

29. Suppose (X, d) is a nonempty complete metric space and A is a closed subspace such that (A, d) is totally bounded. Then every sequence in A has a cluster point.

30. There is a metric d on $[0,1]$, equivalent to the usual metric, such that $([0,1], d)$ is not complete.

31. Let $Z = [0,1]^2$ have the usual topology. Every nonempty closed subset of Z is second category in itself.

32. Suppose $f : [0,1]^n \to \mathbb{R}^m$ and $C = \{x \in [0,1]^n : f$ is continuous at $x\}$. There is a metric d equivalent to the usual metric on C for which (C, d) is complete.

33. If the continuum hypothesis (CH) is true, then $\mathbb{R}$ cannot be written as the union of fewer than c nowhere dense sets.

34. Suppose A is a complete subspace of $[0,1]$. Then A is a G_δ set in $\mathbb{R}$.

35. There are 2^c different subsets of $\mathbb{R}$ none of which contains an interval of positive length.

36. Suppose (X, d) is a nonempty complete metric space and A is a closed subspace such that (A, d) is totally bounded. Then every sequence in A has a cluster point.

37. Let d' be a metric on the irrationals $\mathbb{P}$ which is equivalent to the usual metric and for which $(\mathbb{P}, d')$ is totally bounded. Then $(\mathbb{P}, d')$ cannot be complete.

38. The closure of a discrete subspace of $\mathbb{R}$ may be uncountable.

39. Suppose $f : [0,1]^n \to \mathbb{R}^m$ and $C = \{x \in [0,1]^n : f \text{ is continuous at } x\}$. Then there is a metric d on C, equivalent to the usual metric, such that (C, d) is totally bounded.

40. Suppose $(X, \mathcal{T})$ is compact, that $f : X \to \mathbb{R}$ is a continuous function, and that $f(x) > 0$ for all $x \in X$. Then there is an $\epsilon > 0$ such that $f(x) > \epsilon$ for all $x \in X$.

41. If every point of (X, d) is isolated, then (X, d) is complete.

42. If (X, d) has no isolated points, then its completion also has no isolated points.

43. Suppose that for each point $x \in (X, d)$, there is an open set U_x such that (U_x, d) is complete (*that is, X is "locally complete"*). Then there is a metric $d' \sim d$ on X such that (X, d') is complete.

44. For a metric space (X, d) with completion $(\widetilde{X}, \widetilde{d})$, it can happen that $|\widetilde{X} - X| = \aleph_0$.

45. Let d be the usual metric on $\mathbb{Q}$. There is a completion $(\widetilde{\mathbb{Q}}, \widetilde{d})$ of $(\mathbb{Q}, d)$ for which $|\widetilde{\mathbb{Q}} - \mathbb{Q}| = \aleph_0$.

46. A subspace of $\mathbb{R}$ is bounded iff it is totally bounded.

47. A countable metric space must have at least one isolated point.

48. The intersection of a sequence of dense open subsets in $\mathbb{P}$ must be dense in $\mathbb{P}$.

49. If C is the Cantor set, then $\mathbb{R} - C$ is an F_σ set in $\mathbb{R}$.

50. A subspace A of a metric space (X, d) is compact iff A is closed and bounded.

51. Let $\mathcal{B} = \{ [a, b) \cap [0,1] : a, b \in \mathbb{R},\ a < b \}$ be the base for a topology $\mathcal{T}$ on $[0,1]$. The space $([0,1], \mathcal{T})$ is compact.

52. Let d be the usual metric on ℓ_2. Then (ℓ_2, d) is sequentially compact.

53. Let d be the usual metric on ℓ_2. Then (ℓ_2, d) is countably compact.

54. A subspace of a pseudocompact space is pseudocompact.

55. An uncountable closed set in $\mathbb{R}$ must contain an interval of positive length.

56. Suppose C is the Cantor set and $f : C \to C$ is continuous. Then the graph of f (a subspace of $C \times C$, with its usual metric) is totally bounded.

57. Let C denote the Cantor set $\subseteq [0, 1]$, with the metric $d(x, y) = \frac{|x-y|}{1+|x-y|}$. Then (C, d) is totally bounded.

58. A discrete subspace of $\mathbb{R}$ must be closed in $\mathbb{R}$

59. A subspace of $\mathbb{R}$ which is discrete in its relative topology must be countable.

60. If A and B are subspaces of $(X, \mathcal{T})$ and each is discrete in its subspace topology, then $A \cup B$ is discrete in the subspace topology.

61. Let $\cos^n$ denote the composition of cos with itself n times. Then for each $x \in [-1, 1]$, there exists an n (perhaps depending on x) such that $\cos^n x = x$.

62. Suppose d' is any metric on $0, 1]$ equivalent to the usual metric. Then $([0, 1], d')$ is totally bounded.

63. Let S^1 denote the unit circle in $\mathbb{R}^2$. S^1 is homeomorphic to a subspace of the Cantor set C.

64. There is a metric space (X, d) satisfying the following condition: for every metric $d' \sim d$, (X, d') is totally bounded.

65. $[0, 1]^2$, with the subspace topology from the Sorgenfrey plane, is compact.

Chapter V
Connected Spaces

1. Introduction

In this chapter we introduce the idea of connectedness. Connectedness is a topological property quite different from any property we considered in Chapters 1-4. A connected space X need not have any of the other topological properties we have discussed earlier. Conversely, the only topological properties that imply "X is connected" are very extreme – such as "$|X| \leq 1$" or "X has the trivial topology."

2. Connectedness

Intuitively, a space is connected if it is all in one piece; equivalently a space is disconnected if it can be written as the union of two nonempty "separated" pieces. To make this precise, we need to decide what "separated" should mean. For example, we think of $\mathbb{R}$ as connected even though $\mathbb{R}$ can be written as the union of two disjoint pieces: for example, $\mathbb{R} = A \cup B$, where $A = (-\infty, 0]$ and $B = (0, \infty)$. Evidently, "separated" should mean something more than "disjoint."

On the other hand, if we remove the point 0 to "cut" $\mathbb{R}$, then we probably think of the remaining space $X = \mathbb{R} - \{0\}$ as "disconnected." We can write $X = A \cup B$, where $A = (-\infty, 0)$ and $B = (0, \infty)$. Here, A and B are disjoint, nonempty sets and (*unlike A and B in the preceding paragraph*) they satisfy the following (equivalent) conditions:

i) A and B are open in X
ii) A and B are closed in X
iii) $(B \cap \mathrm{cl}_X A) \cup (A \cap \mathrm{cl}_X B) = \emptyset$ – that is, each of A and B is disjoint from the closure of the other. This is true even if we use $\mathrm{cl}_{\mathbb{R}}$ instead of cl_X.)

Condition iii) is important enough to deserve a name.

Definition 2.1 Suppose A and B are subspaces of $(X, \mathcal{T})$. A and B are called separated if each is disjoint from the closure of the other – that is, if $(B \cap \mathrm{cl}_X A) \cup (A \cap \mathrm{cl}_X B) = \emptyset$.

It follows immediately from the definition that

i) separated sets are disjoint, and

ii) if A and B are separated, and $C \subseteq A$, $D \subseteq B$, then C and D are also separated: that is, subsets of separated sets are separated.

Example 2.2

1) In $\mathbb{R}$, the sets $A = (-\infty, 0]$ and $B = (0, \infty)$ are disjoint but not separated. Likewise in $\mathbb{R}^2$, the sets $A = \{(x, y) : x^2 + y^2 \leq 1\}$ and $B = \{(x, y) : (x-2)^2 + y^2 < 1\}$ are disjoint but not separated.

2) The intervals $A = (-\infty, 0)$ and $B = (0, \infty)$ are separated in $\mathbb{R}$ but $\mathrm{cl}_{\mathbb{R}} A \cap \mathrm{cl}_{\mathbb{R}} B \neq \emptyset$. The same is true for the open balls $A = \{(x,y) : x^2 + y^2 < 1\}$ and $B = \{(x,y) : (x-2)^2 + y^2 < 1\}$ in $\mathbb{R}^2$.

So, the condition that two sets are separated is stronger than saying they are disjoint, but weaker than saying that the sets have disjoint closures.

Theorem 2.3 In any topological space $(X, \mathcal{T})$, the following are equivalent:

1) $\emptyset$ and X are the only clopen sets in X
2) if $A \subseteq X$ and $\mathrm{Fr}\, A = \emptyset$, then $A = \emptyset$ or $A = X$
3) X is not the union of two disjoint nonempty open sets
4) X is not the union of two disjoint nonempty closed sets
5) X is not the union of two nonempty separated sets.

Note: Condition 2) is not frequently used. However it is fairly expressive: to say that $\mathrm{Fr}\, A = \emptyset$ says that no point x in X can be "approximated arbitrarily closely" from both inside and outside A – so, in that sense, A and $B = X - A$ are pieces of X that are "separated" from each other.

Proof 1) $\Leftrightarrow$ 2) This follows because A is clopen iff $\mathrm{Fr}\, A = \emptyset$ (*see Theorems III.2.8, II.4.5.3*).

1) $\Rightarrow$ 3) Suppose 3) is false and that $X = A \cup B$ where A, B are disjoint, nonempty and open. Then $B = X - A$ is a nonempty proper clopen subset of X, which shows that 1) is false.

3) $\Leftrightarrow$ 4) This is clear.

4) $\Rightarrow$ 5) If 5) is false, then $X = A \cup B$, where A, B are nonempty and separated. Since $\mathrm{cl}\, B \cap A = \emptyset$, we conclude that $\mathrm{cl}\, B \subseteq B$, so B is closed. Similarly, A must be closed. Therefore 4) is false.

5) $\Rightarrow$ 1) Suppose 1) is false and that A is a nonempty proper clopen subset of X. Then $B = X - A$ is nonempty and clopen, so A and B are separated. Since $X = A \cup B$, 5) is false. ∙

Definition 2.4 A topological space $(X, \mathcal{T})$ is connected if any (and therefore all) of the conditions in Theorem 2.3 are true. If $C \subseteq X$, we say that C is connected if C is connected in the subspace topology.

According to the definition, a subspace $C \subseteq X$ is disconnected if we can write $C = A \cup B$, where the following (equivalent) statements are true:

1) A and B are disjoint, nonempty and open in C
2) A and B are disjoint, nonempty and closed in C
3) A and B are nonempty and separated in C.

If C is disconnected, such a pair of sets A, B will be called a disconnection or separation of C.

The following technical theorem and its corollary are very useful in working with connectedness in subspaces.

Theorem 2.5 Suppose $A, B \subseteq C \subseteq X$. Then A and B are separated in C iff A and B are separated in X.

Proof $\text{cl}_C B = C \cap \text{cl}_X B$ (*see Theorem III.7.6*), so $A \cap \text{cl}_C B = \emptyset$ iff $A \cap (\text{cl}_X B \cap C) = \emptyset$ iff $(A \cap C) \cap \text{cl}_X B = \emptyset$ iff $A \cap \text{cl}_X B = \emptyset$. Similarly, $B \cap \text{cl}_C A = \emptyset$ iff $B \cap \text{cl}_X A = \emptyset$. $\bullet$

Caution: According to Theorem 2.5, C is disconnected iff $C = A \cup B$ where A and B are nonempty separated sets in C iff $C = A \cup B$ where A and B are nonempty separated sets in X. Theorem 2.5 is very useful because it means that we don't have to distinguish here between "separated in C" and "separated in X" $-$ because these are equivalent. In contrast, if we say that C is disconnected when C is the union of two disjoint, nonempty open (*or closed*) sets A, B in C, then phrase "in C" cannot be omitted: the sets A, B might not be open (*or closed*) in X.

For example, suppose $X = [0,1]$ and $C = [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$. The sets $A = [0, \frac{1}{2})$ and $B = (\frac{1}{2}, 1]$ are open, closed and separated in C. By Theorem 2.5, they are also separated in X, but they are neither open nor closed in X.

Example 2.6

1) Clearly, connectedness is a topological property. More generally, suppose $f : X \to Y$ is continuous and onto. If B is proper nonempty clopen set in Y, then $f^{-1}[B]$ is a proper nonempty clopen set in X. Therefore a continuous image of a connected space is connected.

2) A discrete space X is connected iff $|X| \leq 1$. For example, $\mathbb{N}$ and $\mathbb{Z}$ are not connected.

3) $\mathbb{Q}$ is not connected since we can write $\mathbb{Q}$ as the union of two nonempty separated sets: $\mathbb{Q} = \{q \in \mathbb{Q} : q^2 < 2\} \cup \{q \in \mathbb{Q} : q^2 > 2\}$. Similarly, we can show $\mathbb{P}$ is not connected.

More generally, a connected subset C of $\mathbb{R}$ must be an interval: otherwise, there would be points $a < z < b$ where $a, b \in C$ but $z \notin C$. Then $\{x \in C : x < z\} = \{x \in C : x \leq z\}$ would be a nonempty proper clopen set in C.

In fact, a subset C of $\mathbb{R}$ is connected iff C is an interval. It is not very hard, using the least upper bound property of $\mathbb{R}$, to prove that each interval in $\mathbb{R}$ is connected. (*Try it as an exercise!*) However, we will give a short proof soon using a different argument (Corollary 2.12).

4) The Intermediate Value Theorem If X is connected and $f : X \to \mathbb{R}$ is continuous, then $\text{ran}(f)$ is connected (*by part 1*) so $\text{ran}(f)$ is an interval (*by part 3*). Therefore if $a, b \in X$ and $f(a) < z < f(b)$, there must be a point $c \in X$ for which $f(c) = z$.

5) The Cantor set C is not connected (since it is not an interval). But much more is true. Suppose $x, y \in A \subseteq C$ and that $x < y$. Since C is nowhere dense (*see IV.10*), the interval $(x, y) \nsubseteq C$, so we can choose $z \notin C$ with $x < z < y$. Then $B = (-\infty, z) \cap A$ $= (-\infty, z] \cap A$ is clopen in A, and B contains x but not y. So A is not connected. It follows that every connected subset of C contains at most one point.

A space $(X, \mathcal{T})$ is called totally disconnected if its only nonempty connected subspaces are the one-point sets.

6) X is connected iff every continuous $f : X \to \{0,1\}$ is constant: certainly, if f is continuous and not constant, then $f^{-1}[\{0\}]$ is a proper nonempty clopen set in X so X is not connected. Conversely, if X is not connected and A is a proper nonempty clopen set, then the characteristic function $\chi_A : X \to \{0,1\}$ is continuous but not constant.

Theorem 2.7 Suppose $f : X \to Y$. Let $\Gamma = \{(x, y) \in X \times Y : y = f(x)\}$ = "the graph of f." If f is continuous, then graph of f is homeomorphic to the domain of f; in particular, the graph of a continuous function is connected iff its domain is connected.

Proof We want to show that X is homeomorphic to Γ. Let $h : X \to \Gamma$ be defined by $h(x) = (x, f(x))$. Clearly h is a one-to-one map from X onto Γ.

Let $a \in X$ and suppose $(U \times V) \cap \Gamma$ is a basic open set containing $h(a) = (a, f(a))$. Since f is continuous and $f(a) \in V$, there exists an open set O in X containing a and such that $f[O] \subseteq V$. Then $a \in U \cap O$, and $h[U \cap O] \subseteq (U \times V) \cap \Gamma$, so h is continuous at a.

If U is open in X, then $h[U] = (U \times Y) \cap \Gamma$ is open in Γ, so h is open. Therefore h is a homeomorphism. ●

Note: It is not true that a function f with a connected graph must be continuous. See Example 2.22.

The following lemma makes a simple but very useful observation.

Lemma 2.8 Suppose M, N are separated subsets of X. If $C \subseteq M \cup N$ and C is connected, then $C \subseteq M$ or $C \subseteq N$.

Proof $C = (C \cap M) \cup (C \cap N)$. The sets $C \cap M$ and $C \cap N$ are separated (because $C \cap M \subseteq M$ and $C \cap N \subseteq N$). Since C is connected, either $C \cap M = \emptyset$ (so $C \subseteq N$) or $C \cap N = \emptyset$ (so $C \subseteq M$). ●

The next theorem and its corollaries are simple but powerful tools for proving that certain sets are connected. Roughly, the theorem states that if we have one "central " connected set C and other connected sets none of which is separated from C, then the union of all the sets is connected. It is as if "C links all the connected pieces together."

Theorem 2.9 Suppose C and C_α ($\alpha \in I$) are connected subsets of X and that for each α, C_α and C are not separated. Then $S = C \cup \bigcup C_\alpha$ is connected.

Proof Suppose that $S = M \cup N$ where M and N are separated. By Lemma 2.8, either $C \subseteq M$ or $C \subseteq N$. Suppose $C \subseteq M$. By the same reasoning we conclude that for each α, either $C_\alpha \subseteq M$ or $C_\alpha \subseteq N$. But if some $C_\alpha \subseteq N$, then C and C_α would be separated. Hence every $C_\alpha \subseteq M$. Therefore $N = \emptyset$ and the pair M, N is not a disconnection of S. ●

Corollary 2.10 Suppose that for each $\alpha \in I$, C_α is a connected subset of X and that $C_\alpha \cap C_\beta \neq \emptyset$ for all $\alpha \neq \beta \in I$. Then $\bigcup\{C_\alpha : \alpha \in I\}$ is connected.

Proof If $I = \emptyset$, then $\bigcup\{C_\alpha : \alpha \in I\} = \emptyset$ is connected. If $I \neq \emptyset$, pick an $\alpha_0 \in I$ and let C_{α_0} be the "central set" C in Theorem 2.9. For all $\alpha \in I$, $C_\alpha \cap C_{\alpha_0} \neq \emptyset$, so C_α and C_{α_0} are not separated. By Theorem 2.9, $\bigcup\{C_\alpha : \alpha \in I\}$ is connected. •

Corollary 2.11 For each $n \in \mathbb{N}$, suppose C_n is a connected subset of X and that $C_n \cap C_{n+1} \neq \emptyset$. Then $\bigcup_{n=1}^{\infty} C_n$ is connected.

Proof Let $A_n = \bigcup_{k=1}^{n} C_k$. Corollary 2.10 (and simple induction) shows that the A_n's are connected. Then $\emptyset \neq A_1 \subseteq A_2 \subseteq ... \subseteq A_n \subseteq ...$ Another application of Corollary 2.10 gives that $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} C_n$ is connected. •

Corollary 2.12 Let $I \subseteq \mathbb{R}$. Then I is connected iff I is an interval. In particular, $\mathbb{R}$ is connected (so $\mathbb{R}$ and $\emptyset$ are the only clopen sets in $\mathbb{R}$).

Proof We have already shown that if I is not an interval, then I is not connected (*Example 2.6.3*), so we need to show each interval I is connected. This is obvious if $I = \emptyset$ or $I = \{r\}$.

If $I = [0, 1]$ and A, B are nonempty disjoint closed sets in I, then there are points $a_0 \in A$ and $b_0 \in B$ for which $|a_0 - b_0| = d(a_0, b_0) = d(A, B)$.

> To see this, define $f : A \times B \to [0, 1]$ by $f(x, y) = |x - y|$. $A \times B$ is a closed bounded set in $\mathbb{R}^2$ so $A \times B$ is compact. Therefore f <u>has</u> a <u>minimum</u> value, occurring at some point $(a_0, b_0) \in A \times B$ (*see Exercise IV.E23.*)

Let $z = \frac{a_0 + b_0}{2} \in [0, 1]$. Since $|z - b_0| = |\frac{a_0 + b_0}{2} - b_0| = |\frac{a_0 - b_0}{2}| < |a_0 - b_0|$, we conclude $z \notin A$. Similarly, $z \notin B$. Therefore $[0, 1] \neq A \cup B$, so $[0, 1]$ is connected.

Suppose $a < b$. Since $[a, b]$ is homeomorphic to $[0, 1]$, each interval $[a, b]$ is connected. Since $(a, b) = \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, b - \frac{1}{n}]$, Corollary 2.10 implies that (a, b) is connected. We can use Corollary 2.10 in a similar way to show that all other possible intervals are connected:

$$
\begin{aligned}
\bigcup_{n=1}^{\infty} [a, b - \tfrac{1}{n}] &= [a, b) \\
\bigcup_{n=1}^{\infty} [a + \tfrac{1}{n}, b] &= (a, b] \\
\bigcup_{n=1}^{\infty} [a, a + n] &= [a, \infty) \\
\bigcup_{n=1}^{\infty} (a, a + n) &= (a, \infty) \\
\bigcup_{n=1}^{\infty} [a - n, a] &= (-\infty, a] \\
\bigcup_{n=1}^{\infty} [a - n, a) &= (-\infty, a) \\
\bigcup_{n=1}^{\infty} [-n, n] &= \mathbb{R} \bullet
\end{aligned}
$$

Corollary 2.13 For every $n \in \mathbb{N}$, $\mathbb{R}^n$ is connected.

Proof By Corollary 2.12, $\mathbb{R}^1$ is connected. $\mathbb{R}^n$ can be written as a union of straight lines (each homeomorphic to $\mathbb{R}$) through the origin and Corollary 2.10 implies that $\mathbb{R}^n$ is connected. $\bullet$

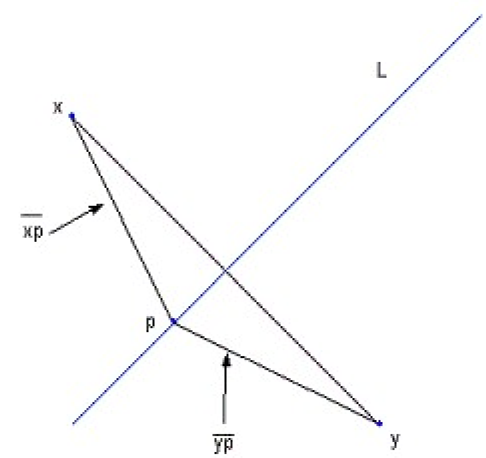
Corollary 2.14 Suppose that for all $x, y \in X$ there exists a connected set $C_{xy} \subseteq X$ with $x, y \in C_{xy}$. Then X is connected.

Proof Certainly $X = \emptyset$ is connected.

If $X \neq \emptyset$, pick a point $a \in X$. For each y in X there is, by hypothesis, a connected set C_{ay} that contains both a and y. By Corollary 2.10, $X = \bigcup\{C_{ay} : y \in X\}$ is connected. $\bullet$

Example 2.15 Suppose C is a countable subset of $\mathbb{R}^n$, where $n \geq 2$. Then $\mathbb{R}^n - C$ is connected. In particular, $\mathbb{R}^n - \mathbb{Q}^n$ is connected. To see this, suppose x, y are any two points in $\mathbb{R}^n - C$.

Choose a straight line L which is perpendicular to the line segment $\overline{xy}$ joining x and y. For each $p \in L$, let C_p be the union of the two line segments $\overline{xp} \cup \overline{py}$. C_p is the union of two intervals with a point in common, so C_p is connected.



If $p' \neq p$, then $C_{p'} \cap C_p = \{x, y\}$, so each $z \in C$ is in at most one C_p. Since C is countable, there must be a point $p^* \in L$ for which $C_{p^*} \cap C = \emptyset$. Then $x, y \in C_{p^*} \subseteq \mathbb{R}^n - C$ so $\mathbb{R}^n - C$ is connected by Corollary 2.14 (with $C_{xy} = C_{p^*}$).

The definition of connectedness agrees with our intuition in the sense that every set that you think (intuitively) should be connected <u>is</u> actually connected according to Definition 2.4. But according to Definition 2.4, certain strange sets also turn out "unexpectedly" to be connected. $\mathbb{R}^n - \mathbb{Q}^n$ might fall into that category. So the official definition forces us to try to expand our intuition about what "connected" means. *For example, is* $\mathbb{R}^n - \mathbb{P}^n$ *connected?*

This situation is analogous to what happens with the "ϵ-δ definition" of continuity. It turns out, using that definition, that every function that you expect (intuitively) should be continuous is actually continuous. If there is problem with the official definition of continuity, it would be that it seems too generous – it allows some "unexpected" functions also to be continuous. An example is the well-known function from elementary analysis: $f : \mathbb{R} \to \mathbb{R}$, where

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } x = \frac{p}{q} \text{ in lowest terms} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

f is continuous at x iff x is irrational (*check!*).

Definition 2.16 If X is connected and $X - \{p\}$ is not connected, then p is called a cut point in X.

Suppose $f : X \to Y$ is a homeomorphism. It is easy to check that p is a cut point in X iff $f(p)$ is a cut point in Y. Therefore homeomorphic spaces have the same number of cut points.

Example 2.17 $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}$ if $n \geq 2$.

Proof Every point in $\mathbb{R}$ is a cut point, but $\mathbb{R}^n$ clearly has no cut points when $n \geq 2$ (or look at Example 2.15).

It is true – but much harder to prove – that $\mathbb{R}^m$ and $\mathbb{R}^n$ are not homeomorphic whenever $m \neq n$. One way to prove this is to develop theorems about a topological property called dimension. Then it turns out (thankfully) that dim $\mathbb{R}^m = m \neq n =$ dim $\mathbb{R}^n$ so these spaces are not homeomorphic. But usually, the proof is done as a part of algebraic topology.

Example 2.18 How is S^1 topologically different from $[0,1]$? They are both compact connected metric spaces with cardinality c, and there is no topological property from Chapters 1-4 that can distinguish between them. The difference has to do with connectivity. Each point in $[0,1]$, except the endpoints, is a cut point; but S^1 has no cut points since $S^1 - \{p\}$ is homeomorphic to $(0,1)$ for every $p \in S^1$.

Corollary 2.19 Suppose $(X, \mathcal{T})$ and $(Y, \mathcal{T}')$ are nonempty topological spaces. Then $X \times Y$ is connected iff X and Y are connected. (*It follows by induction that the same result holds for any finite product of spaces. When infinite products are defined in Chapter 6, it will turn out that the product of any collection of connected spaces is connected.*)

Proof $\Rightarrow$ Suppose $X \times Y$ is connected. Since $X \times Y \neq \emptyset$, we have $X = \pi_X[X \times Y]$. Therefore X is the continuous image of a connected space, so X is connected. Similarly, Y is connected.

$\Leftarrow$ Let X and Y be nonempty connected spaces, and consider any two points (a,b) and (c,d) in $X \times Y$. Then $X \times \{b\}$ and $\{c\} \times Y$ are homeomorphic to X and Y, so these "slices" of the product are connected and both contain the point (c,b). By Corollary 2.10, the "cross" $C = (X \times \{b\}) \cup (\{c\} \times Y)$ is a connected set that contains both (a,b) and (c,d). By Corollary 2.14, we conclude that $X \times Y$ is connected. •

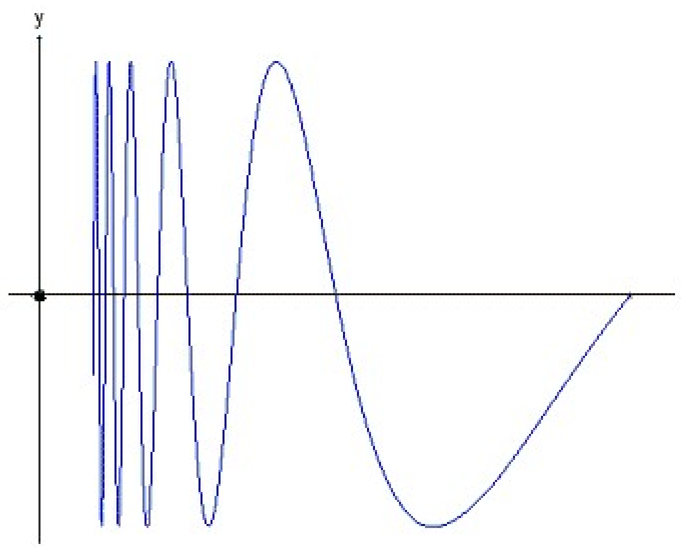
(This corollary gives an another reason why $\mathbb{R}^n$ is connected for $n > 1$.)

Corollary 2.20 Suppose C is a connected subset of X. If $C \subseteq A \subseteq \operatorname{cl} C$, then A is connected. In particular, the closure of a connected set is connected.

Proof For each $a \in A$, $\{a\}$ and C are not separated. By Theorem 2.9, $A = C \cup \bigcup\{\{a\} : a \in A\}$ is connected. •

Example 2.21 By Corollary 2.20, the completion of a connected pseudometric space (X, d) must be connected.

Example 2.22 Let $f(x) = \begin{cases} \sin \frac{\pi}{x} & 0 < x \leq 1 \\ 0 & x = 0 \end{cases}$. Then $f(\frac{1}{n}) = 0$ for every $n \in \mathbb{N}$ and the graph oscillates more and more rapidly between ± 1 as $x \to 0^+$. Part of the graph is pictured below. Of course, f is not continuous at $x = 0$. Let Γ be the graph of the restricted function $g = f|(0,1]$. Since g is continuous, Theorem 2.7 shows that Γ is homeomorphic to $(0,1]$ so Γ is connected.



Γ *is sometimes called the "topologist's sine curve."*

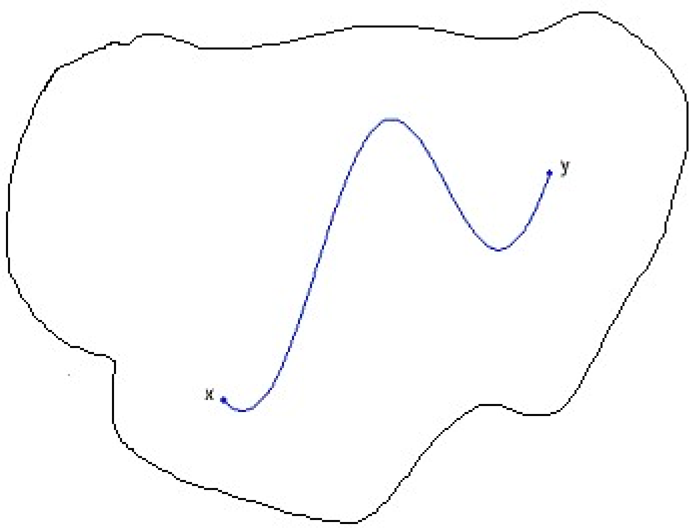
Because $\operatorname{cl}\Gamma = \Gamma \cup (\{0\} \times [-1,1])$, Corollary 2.20 gives that $\Gamma \cup A$ is connected for any set $A \subseteq \{0\} \times [-1,1]$. In particular, $\Gamma_f = \Gamma \cup \{(0,0)\}$ (the graph of f) is connected.

Therefore, a function f with a connected graph need not be continuous. However, it is true that if the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is a closed connected subset of $\mathbb{R}^2$, then f is continuous. (The proof is easy enough to read: see C.E. Burgess, *Continuous Functions and Connected Graphs*, The American Mathematical Monthly, April 1990, 337-339.)

3. Path Connectedness and Local Path Connectedness

In some spaces X, every pair of points can be joined by a path in X. This seems like a very intuitive way to describe "connectedness". However, this condition is actually stronger than our definition for a connected space. .

Definition 3.1 A path in X is a continuous map $f : [0,1] \to X$. The path starts at its initial point $f(0)$ and ends at its terminal point $f(1)$. We say f is a path from $f(0)$ to $f(1)$. We can picture this in $\mathbb{R}^2$:



Sometimes it helps to visualize a path by thinking of a point moving in X from $f(0)$ to $f(1)$ with $f(t)$ representing its position at "time" $t \in [0,1]$. Remember, however, that the path, by definition, is the function f, not the set $\text{ran}(f) \subseteq X$. To illustrate the distinction: suppose f is a path from x to y. Then the function $g : [0,1] \to X$ defined by $g(t) = f(1-t)$ is a different path (in the "opposite direction," from y to x), even though $\text{ran}(f) = \text{ran}(g)$.

Definition 3.2 A topological space X is called path connected if, for every pair of points $x, y \in X$, there is a path from x to y in X.

Note: X is called arcwise connected if, for every pair of points $x, y \in X$, there exists a homeomorphism $f : [0,1] \to X$ with $f(0) = x$ and $f(1) = y$. Such a path f is called an arc from x to y. If a path f in a Hausdorff space X is not an arc, the reason must be that the path intersects itself, that is, that f is not one-to-one (why?). It can be proved that a Hausdorff space is path connected iff it is arcwise connected. Therefore some books use "arcwise connected" to mean the same thing as "path connected."

Theorem 3.3 A path connected space X is connected.

Proof $\emptyset$ is connected, so assume $X \neq \emptyset$ and choose a point $a \in X$. For each $x \in X$, there is a path f_x from a to x. Let $C_x = \text{ran}(f_x)$. Each C_x is connected and contains a. By Corollary 2.10, $X = \bigcup\{C_x : x \in X\}$ is connected. •

Sometimes path connectedness and connectedness are equivalent. For example, a subset $I \subseteq \mathbb{R}$ is connected iff I is an interval iff I is path connected. But in general, the converse to Theorem 3.3 is false, as the next example shows.

Example 3.4 Consider $f(x) = \begin{cases} \sin\frac{\pi}{x} & 0 < x \leq 1 \\ 0 & x = 0 \end{cases}$. In Example 2.22, we showed that the graph Γ_f is connected. However, we claim that there is no path in Γ_f from $(0,0)$ to $(1,0)$ and therefore Γ_f is not path connected.

Suppose, on the contrary, that $h : [0,1] \rightarrow \Gamma_f$ is a path from $(0,0)$ to $(1,0)$. For $t \in [0,1]$, write $h(t) = (h_1(t), h_2(t)) \in \Gamma_f$. The coordinate functions h_1 and h_2 are continuous *(why?)*. Since $[0,1]$ is compact, h is uniformly continuous (*Theorem IV.9.6*) so we can choose $\delta_1 > 0$ for which $|u - v| < \delta_1 \Rightarrow d(h(u), h(v)) < 1 \Rightarrow |h_2(u) - h_2(v)| < 1$.

We have $0 \in h^{-1}((0,0))$. Let $t^* = \sup h^{-1}(0,0)$. Then $0 \leq t^* < 1$. Since $h^{-1}(0,0)$ is a closed set, $t^* \in h^{-1}(0,0)$ so $h(t^*) = (0,0)$. (*We can think of t^* as the last "time" that the path h goes through the origin*).

Choose a positive $\delta < \delta_1$ so that $0 \leq t^* < t^* + \delta < 1$. Since $h_1(t^*) = 0$ and $h_1(t^* + \delta) > 0$, we can choose a positive integer N for which

$$0 = h_1(t^*) \leq \tfrac{2}{N+1} < \tfrac{2}{N} < h_1(t^* + \delta).$$

By the Intermediate Value Theorem, there exist points $u, v \in (t^*, t^* + \delta)$ where $h_1(u) = \frac{2}{N+1}$ and $h_1(v) = \frac{2}{N}$. Then $h_2(u) = \sin\frac{(N+1)\pi}{2}$ and $h_2(v) = \sin\frac{N\pi}{2}$, so $|h_2(u) - h_2(v)| = 1$. But this is impossible since $|u - v| < \delta < \delta_1$ and therefore $|h_2(u) - h_2(v)| < 1$. •

Note: Let Γ be the graph of the restriction $g = f|(0\,1]$. For any nonempty set $A \subseteq \{0\} \times [-1,1]$, a similar argument shows that $\Gamma \cup A$ is not path connected. In particular, $cl\,\Gamma = \Gamma \cup (\{0\} \times [-1,1])$ is not path connected. But Γ is homeomorphic to $(0,1]$, so Γ is path connected. So the closure of a path connected space need not be path connected.

Definition 3.5 A space $(X, \mathcal{T})$ is called

a) locally connected if for each point $x \in X$ and for each neighborhood N of x, there is a connected open set U such that $x \in U \subseteq N$.
b) locally path connected if for each point $x \in X$ and for each neighborhood N of x, there is a path connected open set U such that $x \in U \subseteq N$.

Note: to say U is path connected means that any two points in U can be joined by a path in U. Roughly, "locally path connected" means that "nearby points can be joined by short paths."

Example 3.6

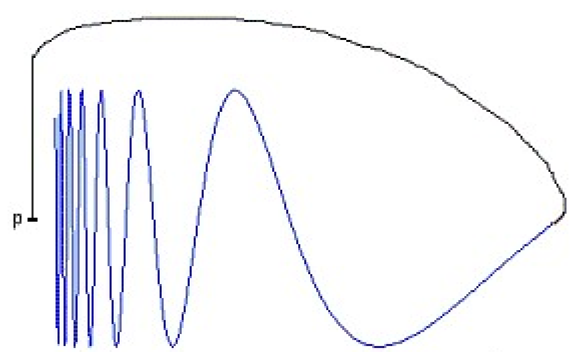
1) $\mathbb{R}^n$ is connected, locally connected, path connected and locally path connected.

2) A locally path connected space is locally connected.

3) Connectedness and path connected are "global" properties of a space X: they are statements about X "as a whole." Local connectedness and local path connectedness are statement about what happens "locally" (in arbitrarily small neighborhoods of points) in X. In general, global properties do not imply local properties, nor vice-versa.

a) Let $X = (0,1) \cup (1,2]$. X is not connected (and therefore not path connected) but X is locally path connected (and therefore locally connected). The same is true if X is a discrete space with more than one point.

b) Let X be the subset of $\mathbb{R}^2$ pictured below. Note that X contains the "topologist's sine curve" as a subspace $-$ you need to imagine it continuing to oscillate faster and faster as it approaches the vertical line segment in the picture:

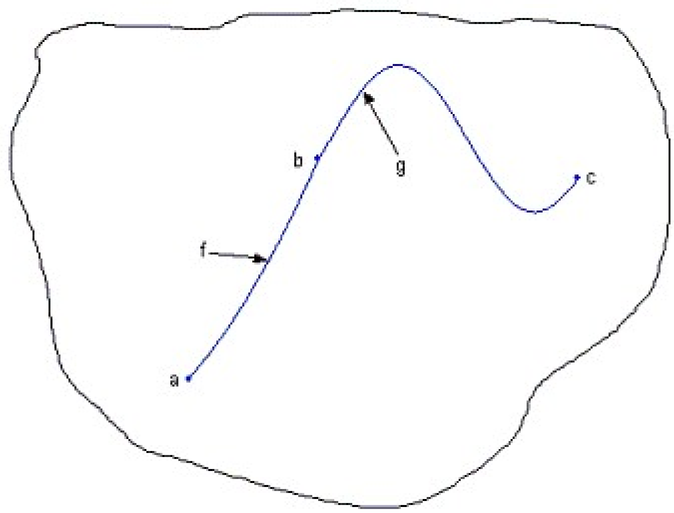


X is path connected (therefore connected), but X is not locally connected: if $p = (0,0)$, there is no open connected set containing p inside the neighborhood $N = B_{\frac{1}{2}}(p) \cap X$. Therefore X is also not locally path connected.

Notice that Examples a) and b) also show that neither "(path) connected" nor "locally (path) connected" implies the other.

Lemma 3.7 Suppose that f is a path in X from a to b and g is a path from b to c. Then there exists a path h in X from a to c.

Proof f ends where g begins, so we feel intuitively that we can "join" the two paths "end-to-end" to get a path h from a to c. The only technical detail we need to provide to provide is that, by definition, a path h must be a function with domain $[0,1]$. To get h we simply "join and reparametrize:"



Define $h:[0,1]\to X$ by $h(t)=\begin{cases} f(2t) & 0\le t\le \frac{1}{2} \\ g(2t-1) & \frac{1}{2}\le t\le 1\end{cases}$. (*You can imagine a point moving twice as fast as before: first along the path f and then continuing along the path g.*)
The function h is continuous by the Pasting Lemma (*see Exercise III.E22*). •

Theorem 3.8 If X is connected and locally path connected, then X is path connected.

Proof If $X=\emptyset$, then X is path connected. So assume $X\neq\emptyset$. For any $a\in X$, let $C=\{x\in X:$ there exists a path in X from a to $x\}$. Then $C\neq\emptyset$ since $a\in C$ (*why?*). We want to show that $C=X$.

Suppose $x\in C$. Let f be a path in X from a to x. Choose a path connected open set U containing x. For any point $y\in U$, there is a path g in U from x to y. By Lemma 3.7, there is a path h in X from a to y, so $y\in C$. Therefore $x\in U\subseteq C$, so C is open.

Suppose $x\notin C$ and choose a path connected open set U containing x. If $y\in U$, there is a path g in U from y to x. Therefore there cannot exist a path in X from a to y – or else, by Lemma 3.7, there would be a path h from a to x and x would be in C. Therefore $y\notin C$, so $x\in U\subseteq X-C$, so C is closed – and therefore clopen.

X is connected and C is a nonempty clopen set, so $C=X$. Therefore X is path connected. •

Here is another situation (particularly useful in complex analysis) where connectedness and path connected coincide:

Corollary 3.9 An <u>open</u> connected set O <u>in $\mathbb{R}^n$</u> is path connected.

Proof Suppose $x \in O$. If N is any neighborhood of x in O, then $x \in \text{int}_O N = U \subseteq O$. Since O is open in $\mathbb{R}^n$, and U is open in O, U is also open in $\mathbb{R}^n$. Therefore there is an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq U \subseteq N$. Since $B_\epsilon(x)$ is an ordinary ball in $\mathbb{R}^n$, $B_\epsilon(x)$ is path connected. (*Of course, that might not be true for a ball in an arbitrary metric space.*) We conclude that O is locally path connected so, by Theorem 3.8, O is path connected. ∙

4. Components

The components of a space X are its largest connected subspaces. A connected space X has exactly one component – X itself. In a totally disconnected space, for example $\mathbb{N}$, the components are the singletons $\{x\}$. In very simple examples, the components "look like" just what you imagine. In more complicated situations, some mild surprises can occur.

Definition 4.1 A component C of a space X is a maximal connected subspace. (Here, "maximal connected" means: C is connected and if $C \subseteq D \subseteq X$ where D is connected, then $C = D$)

For any $p \in X$, let $C_p = \bigcup\{A : p \in A \subseteq X \text{ and } A \text{ is connected}\}$. Then $p \in C_p$ and C_p is connected since C_p is a union of connected sets each containing p (*Corollary 2.10*). If $C_p \subseteq D$ and D is connected, then D was one of the sets A in the collection whose union defines C_p – so $D \subseteq C_p$ and therefore $C_p = D$. Therefore C_p is a component of X that contains p, and X can be written as the union of components: $X = \bigcup_{p \in X} C_p$.

Of course it can happen that $C_p = C_q$ when $p \neq q$: for example, in a connected space X, $C_p = X$ for every $p \in X$. But if $C_p \neq C_q$, then $C_p \cap C_q = \emptyset$: for if $x \in C_p \cap C_q$, then $C_p \cup C_q$ would be a connected set strictly larger than C_p.

The preceding paragraphs show that the distinct components form a partition of X : a pairwise disjoint collection whose union is X. If we define $p \sim q$ to mean that p and q are in the same component of X, then it is easy to see that $\sim$ is an equivalence relation on X and that C_p is the equivalence class of p.

Theorem 4.2 X is the union of its components, and any two components of X are disjoint. Each component is a closed connected set.

Proof In light of the preceding comments, it only remains to show that each component C_p is closed. But this is clear: $C_p \subseteq \text{cl}\, C_p$ and $\text{cl}\, C_p$ is connected (*Corollary 2.20*). By maximality, we conclude that $C = \text{cl}\, C_p$. ∙

It should be clear that a homeomorphism maps components to components. Therefore homeomorphic spaces have the same number of components.

Example 4.3

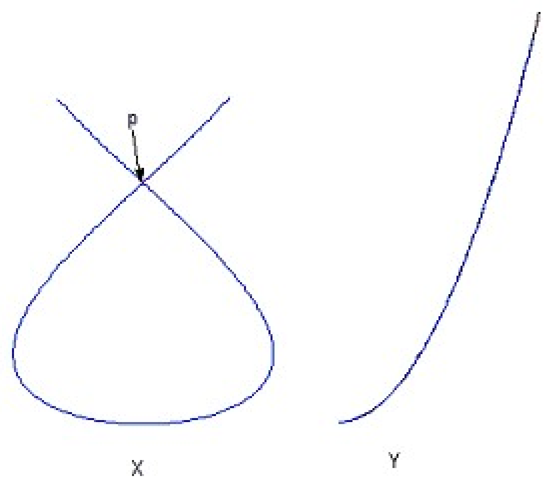
1) Let $X = [1,2] \cup [3,4] \cup [5,6] \subseteq \mathbb{R}$. X has three components: $[1,2]$, $[3,4]$, and $[5,6]$. For each $0 \leq p \leq 1$, we have $C_p = [0,1]$. If a space has only a finite number of components, then each component is also open, because its complement is closed – it is the union of the other finitely many, closed components.

However, a space can have infinitely many components and in general they need not be open. For example, if $X = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\} \subseteq \mathbb{R}$, then the components are the singleton sets $\{x\}$ (*why?*). The component $\{0\}$ is not open in X.

2) In $\mathbb{R}^2$, $X = \bigcup_{n=1}^{2} \mathrm{B}_{1/4}((n,0))$ is not homeomorphic to $Y = \bigcup_{n=1}^{3} \mathrm{B}_{1/4}((n,0))$ because X has two components but Y has three.

3) Suppose C is a nonempty, connected, clopen subset of X. If $C \subsetneq D$, then C is clopen in D, so D is not connected unless $C = D$. Therefore C is a component of X.

4) The sets X and Y in $\mathbb{R}^2$ pictured below are not homeomorphic since X contains a cut point p for which $X - \{p\}$ has three components. Y contains no such cut point.



Example 4.4 The following examples are meant to help "fine-tune" your intuition about components by pointing out some false impressions that you need to avoid. (*Take a look back at Definition 2.4 to be sure you understand what is meant by a "disconnection."*)

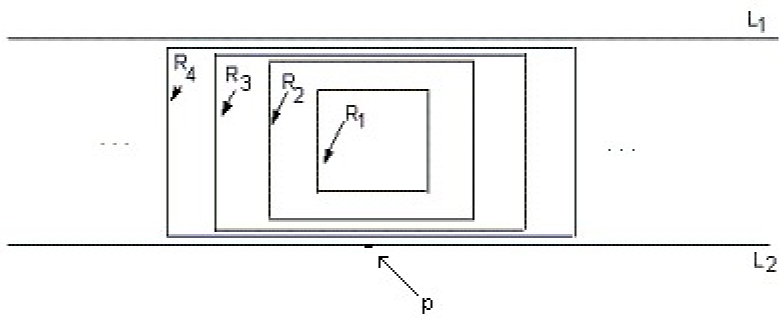
1) Let $X = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$. One of the components of X is $\{0\}$, but $\{0\}$ is <u>not</u> clopen in X. Therefore the sets $A = \{0\}$ and $B = \{\frac{1}{n} : n \in \mathbb{N}\}$ do <u>not</u> form a disconnection of X. <u>A component and its complement might not form a disconnection of X</u>.

2) <u>If X is the union of disjoint closed connected sets, these sets need not be components</u>. For example, $[0,1] = \bigcup_{r \in [0,1]} \{r\}$.

3) If x and y are in different components of X, then there might not exist a disconnection A, B of X for which $x \in A$ and $y \in B$. For example, consider $X = L_1 \cup L_2 \cup \bigcup_{n=1}^{\infty} R_n \subseteq \mathbb{R}^2$, where:

L_1 and L_2 are the straight lines with equations $y = 2$ and $y = -2$.

For each $n \in \mathbb{N}$, R_n is the rectangle $\{(x, y) : |x| = n \text{ and } |y| = 2 - \frac{1}{n}\}$. The top and bottom edges of the R_n's approach the lines L_1 and L_2. The first four R_n's are pictured:



Each R_n is connected and clopen in X. Therefore each R_n is a component of X. (*See Example 4.3.3.*)

L_1 and L_2 are the other components of X. For example:

L_1 is connected. Let C be the component that contains L_1. C must be disjoint from each component R_n, so if $L_1 \neq C$, then the additional points in C are from L_2 – that is, $C = L_1 \cup D$ where $D \subseteq L_2$. But in that case, L_1 would be a nontrivial clopen set in C and C would not be connected. Therefore $L_1 = C$.

Suppose A, B is a disconnection of X – that is, A and B are disjoint clopen sets in X for which $X = A \cup B$. A and B are separated so, by Lemma 2.8, L_2 is either a subset of A or a subset of B: without loss of generality, assume that $L_2 \subseteq A$ and let $p \in L_2$. Since A is open, p has a neighborhood $N \subseteq A$. But N intersects infinitely many connected R_n's, each of which, therefore, must also be a subset of A. Since the top edges of those R_n's approach L_1, there are points on L_1 in $\text{cl}\, A = A$. Therefore L_1 intersects A, so $L_1 \subseteq A$. So $L_1 \cup L_2 \subseteq A$.

In particular: $(0, 2) \in L_1$ and $(0, -2) \in L_2$ are in different components of X, but both are in the same piece A of a disconnection.

Conversely, however: suppose $A \cup B$ is a disconnection of some space Y, with $x \in A$ and $y \in B$. Then x and y must be in different components of Y. (*Why?*)

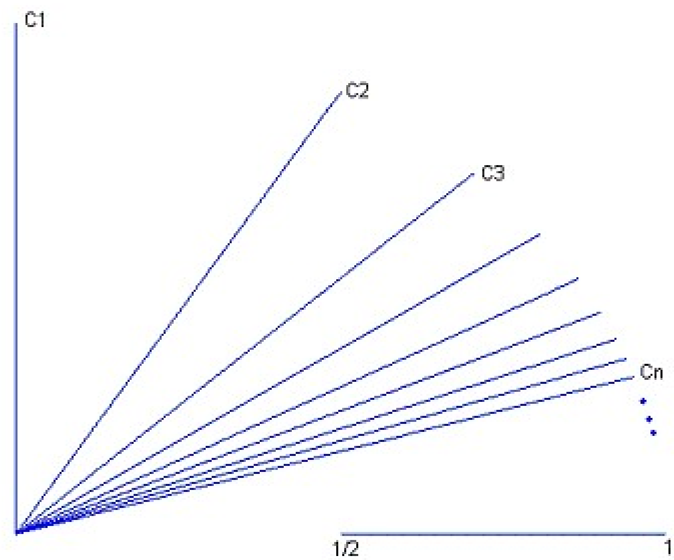
4) Suppose X is a connected space with a cut point v, and let C be a component of $X - \{v\}$. (*Illustrate with a few simple pictures before reading on.*)

It can happen that $v \notin \text{cl}_X\, C$. (*Did your pictures suggest otherwise?*)

For example, consider the following subspace $X = C \cup B$ of $\mathbb{R}^2$ where

C is the interval $[\frac{1}{2}, 1]$ on the x-axis and

$B = \{(0,0)\} \cup \bigcup_{n=1}^{\infty} C_n$ is a "broom" made up of disjoint "straws" C_n (each a copy of $(0,1]$) extending out from the origin and arranged so that slope$(C_n) \to 0$.



The broom B is connected (*because it's path connected*), so $\text{cl}_X B = X$ is connected.

Let $v = (0,0)$. Each straw C_n is connected and clopen in $X - \{v\}$. Therefore each C_n is a component of $X - \{v\}$, and the remaining connected subset, C is the remaining component of $X - \{v\}$.

So v is a cut point of X and $v \notin \text{cl}_X C$.

Note: for this example, $v \notin \text{cl}_X C$, but v is in the closure of the each of the other components C_n of $X - \{v\}$.

In a much more complicated example, due to Knaster and Kuratowski (*Fundamenta Mathematicae, v. 2, 1921*) X is a connected set in $\mathbb{R}^2$ with a cut point v such that $X - \{v\}$ is totally disconnected. Intuitively, all the singleton sets $\{x\}$ are "tied together" at the point v to create the connected space X; removing v causes X to "explode" into "one-element fragments." In contrast to the "broom space," all components in $X - \{v\}$ are singletons, so v is not in the closure (in X) of any of them.

Here is a description of the Knaster-Kuratowski space X (sometimes called "Cantor's teepee"). The proof that is has the properties mentioned is omitted. (*You can find it on p. 145 of the Steen & Seebach's book* Counterexamples in Topology.) Let

$C =$ the Cantor set ($\subseteq [0,1]$) on the x-axis in $\mathbb{R}^2$, and let $v = (\frac{1}{2}, 1)$.

$D = \{p \in C :$ p is the endpoint of one of the "deleted middle thirds" in the construction of $C\}$
$= \{p \in C :$ $p = 0.a_1 a_2 a_3 ... a_n ..._{\text{base three}}$, where the a_n's are eventually equal to 0 or eventually equal to $2\}$. Of course, D is countable.

$E = C - D$ ($=$ "the points in C that are not isolated on either side in C")

Then $C = D \cup E$, $D \cap E = \emptyset$ and both D and E are dense in C.

For each $p \in C$, let $\overline{vp}$ $=$ the line segment from v to p and define a subset of $\overline{vp}$ by

$$C_p = \begin{cases} \{(x,y) \in \overline{vp} : y \text{ is rational}\} & \text{if } p \in D \\ \{(x,y) \in \overline{vp} : y \text{ is irrational}\} & \text{if } p \in E. \end{cases}$$

Cantor's teepee is the space $X = \bigcup_{p \in C} C_p$. X is connected and $X - \{v\}$ is totally disconnected.

5. Sierpinski's Theorem

Let $\mathcal{T}$ be the cofinite topology on $\mathbb{N}$. Clearly, $(\mathbb{N}, \mathcal{T})$ is connected. But is it path connected? (*Try to prove or disprove!*) This innocent-sounding question turns out to be harder than you might expect.

Suppose $f : [0,1] \to (\mathbb{N}, \mathcal{T})$ is a path from (say) 1 to 2 in $(\mathbb{N}, \mathcal{T})$. Then ran$(f)$ is a connected set containing at least two points. But every finite subspace of $(\mathbb{N}, \mathcal{T})$ is discrete, so ran(f) must be infinite. Therefore $[0,1] = \bigcup_{n=1}^{\infty} f^{-1}(n)$, where infinitely many of the sets $f^{-1}(n)$ are nonempty. Is this possible? The question is not particularly easy. In fact, the question of whether $(\mathbb{N}, \mathcal{T})$ is path connected is equivalent to the question of whether $[0,1]$ can be written as a countable union of pairwise disjoint nonempty closed sets.

The answer lies in a famous old theorem of Sierpinski which states that a compact connected Hausdorff space X cannot be written as a countable union of two or more nonempty pairwise disjoint closed sets. (*The "finite union" case is trivial: X cannot be the union of n nonempty disjoint closed sets $(n \geq 2)$ since each set would be clopen $-$ an impossibility since X is connected. The "infinite union" case is the interesting part of the theorem.*)

We will prove Sierpinski's result after a series of several lemmas. The line of argument used is due to R. Engelking. (*It is possible to prove Sierpinski's theorem just for the special case $X = [0,1]$. That proof is a little easier but still nontrivial.*)

Lemma 5.1 If A and B are disjoint closed sets in a compact Hausdorff space X, then there exist disjoint open sets U and V with $A \subseteq U$ and $B \subseteq V$.

Proof Consider first the case where $A = \{x\}$, a singleton set. For each $y \in B$, choose disjoint open sets U_y and V_y with $x \in U_y$ and $y \in V_y$. The open sets V_y cover the compact set B so a finite number of them cover B, say $B \subseteq V_{y_1} \cup ... \cup V_{y_n} = V$. Let $U = U_{y_1} \cap ... \cap U_{y_n}$. Then $A \subseteq U$, $B \subseteq V$ and U, V are disjoint open sets.

Suppose now that A and B are any pair of disjoint closed sets in X. For each $x \in A$, pick disjoint open sets U_x and V_x such that $\{x\} \subseteq U_x$ and $B \subseteq V_x$. The open U_x's cover the compact set A, so a finite number of them cover A, say $A \subseteq U_{x_1} \cup ... \cup U_{x_n} = U$. Let $V = V_{x_1} \cap ... \cap V_{x_n}$. Then $A \subseteq U$, $B \subseteq V$ and U, V are disjoint open sets. •

Note: If A and B were both finite, an argument analogous to the proof given above would work in <u>*any*</u> *Hausdorff space X. The proof of Lemma 5.1 illustrates the rule of thumb that "compact sets act like finite sets."*

Lemma 5.2 Suppose O is an open set in the compact space $(X, \mathcal{T})$. If $\mathcal{F} = \{F_\alpha : \alpha \in I\}$ is a family of closed sets in X for which $\bigcap \mathcal{F} \subseteq O$, then there exist $\alpha_1, ..., \alpha_n \in I$ such that $F_{\alpha_1} \cap F_{\alpha_2} \cap ... \cap F_{\alpha_n} \subseteq O$.

Proof For each $y \in X - O$, there is an α such that $y \notin F_\alpha$. Therefore $\{X - F_\alpha : \alpha \in I\}$ is an open cover of the compact set $X - O$. There exist $\alpha_1, ...\alpha_n \in I$ such that $(X - F_{\alpha_1}) \cup (X - F_{\alpha_2}) \cup ... \cup (X - F_{\alpha_n}) \supseteq X - O$. Taking complements gives that $F_{\alpha_1} \cap F_{\alpha_2} \cap ... \cap F_{\alpha_n} \subseteq O$. •

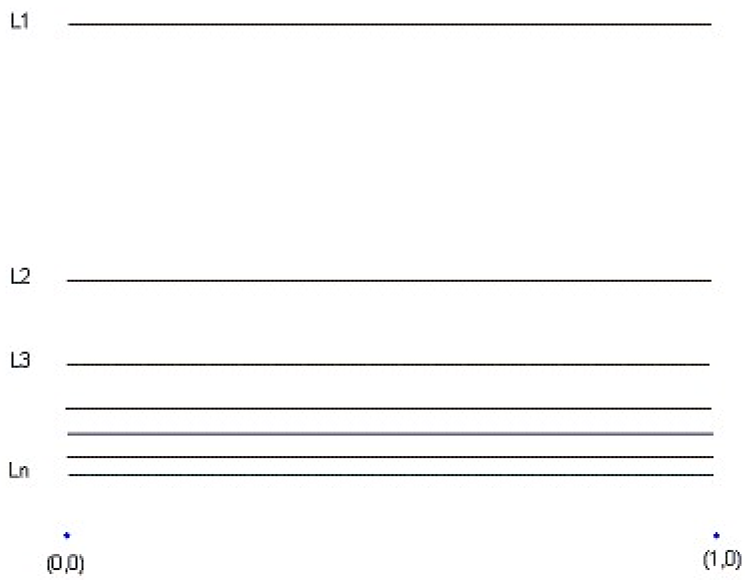
Definition 5.3 Suppose $p \in X$. The set $Q_p = \bigcap\{C \subseteq X : p \in C \text{ and } C \text{ is clopen in } X\}$ is called the <u>quasicomponent</u> of X containing p.

Q_p is always a closed set in X. The next two lemmas give some relationships between the component C_p that contains p and the quasicomponent Q_p.

Lemma 5.4 If $p \in X$, then $C_p \subseteq Q_p$.

Proof If C is any clopen set containing p, then $C_p \subseteq C$ – for otherwise $C_p \cap C$ and $C_p \cap (X - C)$ disconnect C_p. Therefore $C_p \subseteq \bigcap\{C : p \in C \text{ and } C \text{ is clopen}\} = Q_p$. •

Example 5.5 An example where $C_p \neq Q_p$. In $\mathbb{R}^2$, let L_n be the horizontal line segment $[0,1] \times \{\frac{1}{n}\}$ and define $X = \bigcup_{n=1}^{\infty} L_n \cup \{(0,0),(1,0)\}$.



The components of X are the sets L_n and the two singleton sets $\{(0,0)\}$ and $\{(1,0)\}$.

If C is any clopen set in X that contains $(0,0)$, then C intersects infinitely many L_n's so (since the L_n's are connected) C contains those L_n's. Hence the closed set C contains points arbitrarily close to $(1,0)$ $-$ so $(1,0)$ is also in C. Therefore $(0,0)$ and $(1,0)$ are both in $Q_{(0,0)}$, so $C_{(0,0)} \neq Q_{(0,0)}$. (*In fact, it is easy to check that* $Q_{(0,0)} = \{(0,0),\ (1,0)\}$.)

Lemma 5.6 If X is a compact Hausdorff space and $p \in X$, then $C_p = Q_p$.

Proof C_p is a maximal connected set and $C_p \subseteq Q_p$, so we can show that $C_p = Q_p$ by showing that Q_p is connected. Suppose $Q_p = A \cup B$, where A, B are disjoint closed sets in Q_p. We can assume that $p \in A$. We will show that $B = \emptyset$ $-$ in other words, that there is no disconnection of Q_p.

Q_p is closed in X, so A and B are also closed in X. By Lemma 5.1, we can choose disjoint open sets U and V in X with $p \in A \subseteq U$ and $B \subseteq V$. Then $Q_p = A \cup B \subseteq U \cup V$. Since X is compact and Q_p is an intersection of clopen sets in X, Lemma 5.2 lets us pick finitely many clopen sets $C_1, ..., C_n$ such that $Q_p \subseteq C_1 \cap ... \cap C_n \subseteq U \cup V$. Let $C = C_1 \cap ... \cap C_n$. C is clopen in X and $Q_p \subseteq C \subseteq U \cup V$.

$U \cap C$ is open in X and, in fact, $U \cap C$ is also closed in X: since $\text{cl}\, U \subseteq X - V$, we have that $U \cap C = \text{cl}\, U \cap C = \text{cl}\, U \cap \text{cl}\, C \supseteq \text{cl}\,(U \cap C)$. Therefore $U \cap C$ is one of the clopen sets containing p whose intersection defines Q_p, so $Q_p \subseteq U \cap C \subseteq U$. Therefore $Q_p \cap B = \emptyset$, so $B = \emptyset$. $\bullet$

Definition 5.7 A continuum is a compact connected Hausdorff space.

Lemma 5.8 Suppose A is a closed subspace of a continuum X and that $\emptyset \neq A \neq X$. If C is a component of A, then $C \cap \mathrm{Fr}\, A \neq \emptyset$.

Proof Let C be a component of A and let $p \in C$. Since $C \subseteq A = \mathrm{cl}\, A$, we have that $C \cap \mathrm{Fr}\, A = C \cap \mathrm{cl}\, A \cap \mathrm{cl}\,(X - A)) = C \cap \mathrm{cl}\,(X - A)$, so we need to show that $C \cap \mathrm{cl}\,(X - A) \neq \emptyset$. We do this by contraposition: assuming $C \cap \mathrm{cl}\,(X - A) = \emptyset$, we will prove $A = X$.

A is compact so Lemma 5.6 gives $C = C_p = Q_p = \bigcap\{C_\alpha \subseteq A : p \in C_\alpha$ and C_α is clopen in $A\}$. By assumption, $C \subseteq X - \mathrm{cl}(X - A)$, so by Lemma 5.2 there exist indices α_1, α_2,..., α_n for which $C \subseteq C_{\alpha_0} = C_{\alpha_1} \cap C_{\alpha_2} ... \cap C_{\alpha_n} \subseteq X - \mathrm{cl}(X - A)$. Since $C_{\alpha_0} \cap \mathrm{cl}\,(X - A) = \emptyset$, we have $C_{\alpha_0} \cap \mathrm{Fr}\, A = \emptyset$.

C_{α_0} is clopen in A and $C_{\alpha_0} \subseteq A - \mathrm{Fr}\, A = \mathrm{int}\, A$. Since C_{α_0} is open in $\mathrm{int}\, A$ which is open in X, C_{α_0} is open in X. But C_{α_0} is also closed in the closed set A, so C_{α_0} is closed in X. Since X is connected, we conclude that $A = X$. $\bullet$

Lemma 5.9 Suppose X is a continuum and that $X = \bigcup\{F_n : n \in \mathbb{N}\}$ where the F_n's are pairwise disjoint closed sets and $F_n \neq \emptyset$ for <u>at least</u> two values of n. Then for <u>each</u> n there exists a continuum $C_n \subseteq X$ such that $C_n \cap F_n = \emptyset$ and $C_n \cap F_i \neq \emptyset$ for <u>at least two</u> $i \in \mathbb{N}$.

Before proving Lemma 5.9, consider the formal statement of Sierpinski's Theorem..

Theorem 5.10 (Sierpinski) Let X be a continuum. If $X = \bigcup\{F_n : n \in \mathbb{N}\}$ where the F_n's are pairwise disjoint closed sets, then <u>at most one</u> F_n is nonempty.

(*Of course, the statement of the theorem includes the easy "finite union" case.*) *In proving Sierpinski's theorem we will assume that* $X = \bigcup\{F_n : n \in \mathbb{N}\}$ *where the* F_n*'s are pairwise disjoint closed sets and* $F_n \neq \emptyset$ *for <u>at least</u> two values of* n. *Then we will apply Lemma 5.9 to arrive at a contradiction. When all the smoke clears we see that, in fact, <u>there are no continua which satisfy the hypotheses of Lemma 5.9</u>. Lemma 5.9 is really the first part of the proof (by contradiction) of Sierpinski's theorem – set off as a preliminary lemma to break the argument into more manageable pieces.*

Proof of Lemma 5.9

If $F_n = \emptyset$, let $C_n = X$.

Assume $F_n \neq \emptyset$. Choose $m \neq n$ with $F_m \neq \emptyset$ and pick a point $p \in F_m$. By Lemma 5.1, we can choose disjoint open sets U, V in X with $F_n \subseteq U$ and $p \in F_m \subseteq V$. Let C_n be the component of p in $\mathrm{cl}\, V$. Certainly C_n is a continuum, and we prove that this choice of C_n works. We have that $C_n \cap F_n = \emptyset$ (since $C_n \subseteq \mathrm{cl}\, V \subseteq X - U$) and that $C_n \cap F_m \neq \emptyset$ (since

$p \in C_n \cap F_m$). Therefore, to complete proof, we need only show that for some $i \neq m, n$, $C_n \cap F_i \neq \emptyset$.

Since $p \in \mathrm{cl}\, V \subseteq X - U$, we have that $\mathrm{cl}\, V \neq \emptyset$; and $\mathrm{cl}\, V \neq X$ because $\emptyset \neq F_n \subseteq U$. Therefore, by Lemma 5.8, there is a point $q \in C_n \cap \mathrm{Fr}\,(\mathrm{cl}\, V)$. Since $q \in \mathrm{Fr}\,(\mathrm{cl}\, V)$ and $F_m \subseteq V \subseteq \mathrm{int}(\mathrm{cl}\, V)$, we have $q \notin F_m$. And since $q \in \mathrm{Fr}\,(\mathrm{cl}\, V) \subseteq \mathrm{cl}\, V \subseteq X - U$, we have that $q \notin F_n$. But X is covered by the F_i's, so $q \in F_i$ for some $i \neq m, n$. Therefore $C_n \cap F_i \neq \emptyset$. •

Proof of Theorem 5.10 We want to show that if $X = \bigcup_{n=1}^{\infty} F_n$ where the F_n's are disjoint closed sets, then <u>at most one</u> $F_n \neq \emptyset$. Looking for a contradiction, we suppose at least two F_n's are nonempty.

By Lemma 5.9, there is a continuum C_1 in X with $C_1 \cap F_1 = \emptyset$ and such that C_1 has nonempty intersection with a least two F_n's. We can write $C_1 = C_1 \cap X = C_1 \cap \bigcup_{n=1}^{\infty} F_n$ $= \bigcup_{n=1}^{\infty}(C_1 \cap F_n) = \bigcup_{n=2}^{\infty}(C_1 \cap F_n)$, where at least two of the sets $C_1 \cap F_n$ are nonempty.

Applying Lemma 5.9 again (to the continuum C_1) we find a continuum $C_2 \subseteq C_1$ such that $C_2 \cap (C_1 \cap F_2) = C_2 \cap F_2 = \emptyset$ and C_2 intersects at least two of the sets $C_1 \cap F_n$.
Then $C_2 = C_2 \cap C_1 = \bigcup_{n=2}^{\infty}(C_2 \cap (C_1 \cap F_n)) = \bigcup_{n=2}^{\infty}(C_2 \cap F_n) = \bigcup_{n=3}^{\infty}(C_2 \cap F_n)$, where at least two of the sets $C_2 \cap F_n$ are nonempty.

We continue this process inductively, repeatedly applying Lemma 5.9, to and generate a decreasing sequence of nonempty continua $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_n \supseteq \ldots$ such that for each n, $C_n \cap F_n = \emptyset$. This gives $\emptyset = \bigcap_{n=1}^{\infty} C_n \cap \bigcup_{n=1}^{\infty} F_n = \bigcap_{n=1}^{\infty} C_n \cap X = \bigcap_{n=1}^{\infty} C_n$. But this is impossible: the C_n's have the finite intersection property and X is compact, so $\bigcap_{n=1}^{\infty} C_n \neq \emptyset$. •

Example 5.11 By Theorem 5.10, we know that $[0,1]$ cannot be written as the union of m pairwise disjoint nonempty closed sets if $1 < m \leq \aleph_0$. And, of course, $[0,1]$ can easily be written as the union of $m = c$ such sets: for example, $[0,1] = \bigcup_{x \in [0,1]} \{x\}$. What if $\aleph_0 < m < c$?

There are other related questions you could ask yourself. For example, can $[0,1]$ be written as the union of c disjoint closed sets each of which is uncountable? The answer is "yes." For example, take a continuous onto function $f : [0,1] \to [0,1]^2$ (*a space-filling curve, whose existence you should have seen in an advanced calculus course*). For each $x \in [0,1]$, let $L_x = \{x\} \times [0,1] =$ "the vertical line segment at x in $[0,1]^2$. Then the sets $f^{-1}[L_x]$ do the job.

We could also ask: is it possible to write $[0,1]$ as the union of uncountably many pairwise disjoint closed sets each of which is countably infinite? (See Exercise VIII.E27).

Exercises

E1. Suppose $X = A \cup B$ where $A - B$ and $B - A$ are separated.

a) Prove that if $C \subseteq X$, then $\mathrm{cl}_X C = \mathrm{cl}_A(A \cap C) \cup \mathrm{cl}_B(B \cap C)$.
b) Conclude that C is closed if $C \cap A$ is closed in A and $C \cap B$ is closed in B.
c) Conclude that C is open if $C \cap A$ is open in A and $C \cap B$ is open in B.
d) Suppose $X = A \cup B$ where $A - B$ and $B - A$ are separated. Prove that if $f : X \to Y$ and both $f|A$ and $f|B$ are continuous, then f is continuous.

E2. Prove that $[0, 1]$ is connected directly from the definition of connected.
(*Use the least upper bound property of* $\mathbb{R}$.)

E3. Suppose both A, B are closed subsets of $(X, \mathcal{T})$. Prove that $A - B$ is separated from $B - A$. Do the same assuming instead that A and B are both open.

E4. Suppose S is a connected subset of $(X, \mathcal{T})$. Prove that if $S \cap E \neq \emptyset$ and $S \cap (X - E) \neq \emptyset$, then $S \cap \mathrm{Fr}\, E \neq \emptyset$.

E5. Let (X, d) and (Y, d') be two connected metric spaces. Suppose $k > 0$ and that $(a, b) \in X \times Y$. Let $K = \{(x, y) \in X \times Y : d(x, a) \leq k \text{ and } d'(y, b) \leq k\}$.

a) Give an example to show that the complement of K in $X \times Y$ might not be connected.
b) Prove that the complement of K in $X \times Y$ is connected if (X, d) and $Y, d')$ are unbounded.

E6. Prove that there does not exist a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that $f[\mathbb{P}] \subseteq \mathbb{Q}$ and $f[\mathbb{Q}] \subseteq \mathbb{P}$.

Hint: One method: What do you know about ran f? What else do you know?)
Another method: if such an f exists, let $g = \frac{1}{1+|f|}$ and let $h = g|[0, 1]$. What do you know about h?

E7. a) Find the cardinality of the collection of all compact connected subsets of $\mathbb{R}^2$.
b) Find the cardinality of the collection of all connected subsets of $\mathbb{R}^2$.

E8. Suppose (X, d) is a connected metric space with $|X| > 1$. Prove that $|X| \geq c$.

E9. Suppose each point in a metric space (X, d) has a neighborhood base consisting of clopen sets (such a metric space is called *zero-dimensional*). Prove that (X, d) is totally disconnected.

E10. A metric space (X, d) satisfies the ϵ-chain condition if for all $\epsilon > 0$ and all $x, y \in X$, there exists a finite set of points $x_1, x_2, \ldots, x_{n-1}, x_n$ where $x_1 = x$, $x_n = y$, and $d(x_i, x_{i+1}) < \epsilon$ for all $i = 1, \ldots, n-1$.

a) Give an example of a metric space which satisfies the ϵ-chain condition but which is not connected.

b) Prove that if (X, d) is connected, then (X, d) satisfies the ϵ-chain condition.

c) Prove that if (X, d) is compact and satisfies the ϵ-chain condition, then X is connected.

d) Prove that $(\{0\} \times [-1, 1]) \cup \{(x, \sin \frac{1}{x}) : 0 < x \leq 1\} \subseteq \mathbb{R}^2$ is connected.
(*Use c) to give a different proof than the one given in Example 3.4.*)

e) In any space $(X, \mathcal{T})$, a simple chain from a to b is a finite collection of sets $\{A_1, \ldots, A_n\}$ such that:

$a \in A_1$ and $a \notin A_i$ if $i \neq 1$
$b \in A_n$ and $b \notin A_i$ if $i \neq n$
for all $i = 1, \ldots, n-1$, $A_i \cap A_{i+1} \neq \emptyset$
$A_i \cap A_j = \emptyset$ if $|j - i| \neq 1$

Prove X is connected iff for every open cover $\mathcal{U}$ and every pair of points $a, b \in X$, there is a simple chain from a to b consisting of sets taken from $\mathcal{U}$.

E11. a) Prove that X is locally connected iff the components of every open set O are also open in X.

b) The path components of a space are the maximal path connected subsets. Show that X is locally path connected iff the path components of every open set O are also open in X.

E12. Let $\mathcal{T}$ be the cofinite topology on $\mathbb{N}$. We know that $(\mathbb{N}, \mathcal{T})$ is not path connected (because of Sierpinski's Theorem applied to the closed interval $[0, 1]$). Prove that the statement "$(\mathbb{N}, \mathcal{T})$ is not path connected" is equivalent to "Sierpinski's Theorem for the case $X = [0, 1]$."

E13. Let $n > 1$ and suppose $f : [0, 1] \rightarrow \mathbb{R}^n$ is a homeomorphism (into); then ran(f) is called an arc in $\mathbb{R}^n$. Use a connectedness argument to prove that an arc is nowhere dense in $\mathbb{R}^n$. Is the same true if $[0, 1]$ is replaced by the circle S^1?

E14. a) Prove that for any space X and $n \geq 2$,

> if X has $\geq n$ components, then there are nonempty pairwise separated sets $H_1, ..., H_n$ for which $X = H_1 \cup ... \cup H_n$ (**)

Hints. For a given n, do not start with the components and try to group them to form the H_n's. Start with the fact that X is not connected. Use induction. When X has infinitely many components, then X has $\geq n$ components for every n.

b) Recall that a <u>disconnection</u> of X means a pair of nonempty separated sets A, B for which $X = A \cup B$. Remember also that if C is a component of X, C is not necessarily "one piece in a disconnection of X" *(see Example 4.4)*.

Prove that X has only finitely many components $n \, (n \geq 2)$ iff X has only finitely many disconnections.

E15. A metric space (X, d) is called locally separable if, for each $x \in X$, there is an open set U containing x such that (U, d) is separable. Prove that a connected, locally separable metric space is separable.

E16. In $(X, \mathcal{T})$, define $x \sim y$ if there does <u>not</u> exist a disconnection $X = A \cup B$ with $x \in A$ and $y \in B$, i.e., if "X can't be split between x and y." Prove that $\sim$ is an equivalence relation and that the equivalence class of a point p is the quasicomponent Q_p. (*It follows that X is the disjoint union of its quasicomponents.*)

E17. For the following alphabet (capital Arial font), decide which pairs of letters are homeomorphic:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

E18. Suppose $f : \mathbb{R} \to X = \{(x, y) \in \mathbb{R}^2 : x = 0 \text{ or } y = 0\}$ is continuous and onto. Prove that $f^{-1}[\{(0, 0)\}]$ contains at least 3 points.

E19. Show how to write $\mathbb{R}^2 = A \cup B$ where A and B are nonempty, disjoint, connected, dense and congruent by translation (i.e., $\exists \, (u, v) \in \mathbb{R}^2$ such that $B = \{(x + u, y + v) : (x, y) \in A\}$).

E20. Suppose X is connected and $|X| \geq 2$. Show that X can be written as $A \cup B$ where A and B are connected proper subsets of X.

E21. Prove or disprove: a nonempty product $X \times Y$ is totally disconnected iff both X and Y are totally disconnected.

Chapter V Review

Explain why each statement is true, or provide a counterexample.

1. There exists a continuous function $f : \mathbb{R} \to \mathbb{R}$ which is onto and for which $f[\mathbb{Q}] \subseteq \mathbb{N}$.

2. Let $\mathcal{T}$ be the right ray topology on $\mathbb{R}$. Then $(\mathbb{R}, \mathcal{T})$ is path connected.

3. There exists a countably infinite compact connected metric space.

4. The letter T is homeomorphic to the letter F.

5. If A and B are connected and not separated, then $A \cup B$ is connected.

6. If A and B are nonempty and $A \cup B$ is connected, then $\text{cl}\, A \cap \text{cl}\, B \neq \emptyset$.

7. If the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is connected, then f is continuous.

8. $\mathbb{N}$, with the cofinite topology, is connected.

9. In a space $(X, \mathcal{T})$, the component containing the point p is a subset of the intersection of all clopen sets containing a point p.

10. In a topological space: if $A, B \subseteq X$ and A is clopen in $A \cup B$, then X is not connected.

11. If $f : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is continuous and onto and $(X, \mathcal{T})$ is path connected, then $(Y, \mathcal{T}')$ is path connected.

12. Suppose $A \subseteq \mathbb{R}$ and that $|\, A \,| > 1$. If A is nowhere dense, then A is not connected.

13. Let C be the Cantor set. There exists a nonempty space X for which $X \times C$ is connected.

14. Let C be the Cantor set. Then $\mathbb{R}^2 - C^2$ is connected.

15. A component in a complete metric space is complete.

16. Let S^1 be the unit circle in $\mathbb{R}^2$. S^1 is homeomorphic to a subspace of the Cantor set.

17. If X is the union of an uncountable collection of pairwise disjoint nonempty, connected closed sets C_α, then the C_α's are components of X.

18. If X is the union of a finite collection of disjoint nonempty connected closed sets C_α, then the C_α's are components of X.

19. If X is the union of a countable collection of disjoint connected closed sets C_α, then the C_α's are components of X.

20. If (X, d) is connected, then its completion is also connected.

21. Suppose A is connected and B is clopen. If $A \cap B \neq \emptyset$, then $A \subseteq B$.

22. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, that $f((0, 0, ..., 0)) = (0, 0, ..., 0)$ and $f((1, 0, ..., 0)) = (2, 0, ..., 0)$. Let A denote the set of all fixed points of f. It is possible that A is open in $\mathbb{R}^n$ (*Note: we are not assuming f is a contraction, so f may have more than one fixed point.)*

23. Suppose (X, d) is a connected separable metric space with $\mid X \mid > 1$. Then $\mid X \mid = c$.

24. If a subset A of $\mathbb{R}$ contains an open interval around each of its points, then A must be connected.

25. There exists a connected metric space (X, d) with $|X| = \aleph_0$.

26. If $\mathbb{R}^2 \supseteq A_1 \supseteq A_2 \supseteq ... \supseteq A_n \supseteq ...$ is a nested sequence of connected sets in the plane, then $\bigcap_{n=1}^{\infty} A_n$ is connected.

27. $(\mathbb{P} \times \mathbb{R}) \cup (\mathbb{R} \times \mathbb{P})$ is a connected set in $\mathbb{R}^2$.

28. Let $A = \{(x, \sin \frac{1}{x}) : x > 0\} \subseteq \mathbb{R}^2$ and suppose $f : \mathrm{cl}\,(A) \to \{0, 1\}$ is continuous. Then f must be constant.

29. Suppose $X \neq \emptyset$. X is connected if and only if there are exactly two functions $f \in C(X)$ such that $f^2 = f$.

30. If X is nonempty, countable and connected, then every $f \in C(X)$ is constant.

31. Every path connected set in $\mathbb{R}^2$ is locally path connected.

32. If B is a dense connected set in $\mathbb{R}$, then $B = \mathbb{R}$.

33. In a metric space (X, d) the sets A and B are separated iff $d(A, B) > 0$.

34. A nonempty clopen subset of a space X must be a component of X.

35. Suppose $\mathcal{T}$ and $\mathcal{T}'$ are topologies on X and that $\mathcal{T} \subseteq \mathcal{T}'$. If $(X, \mathcal{T})$ is connected, then $(X, \mathcal{T}')$ is connected.

36. The letter X is homeomorphic to the letter Y.

Chapter VI
Products and Quotients

1. Introduction

In Chapter III we defined the product $X \times Y$ of two topological spaces and considered some of the simple properties of products. (*See Examples III.5.10-5.12 and Exercise IIIE20.*) The properties we explored hold equally well for products of any finite number of spaces $X_1 \times ... \times X_n$. For example, the product of two compact spaces is compact, so a simple induction argument shows that the product of any finite number of compact spaces is compact. Now we turn our attention to infinite products which will lead us to some very nice theorems. For example, infinite products will eventually help us decide which topological spaces are metrizable.

2. Infinite Products and the Product Topology

The set $X \times Y$ was defined as $\{(x, y) : x \in X, y \in Y\}$. How can we define an "infinite product" set $X = \prod\{X_\alpha : \alpha \in A\}$? Informally, we want to say something like

$$X = \prod\{X_\alpha : \alpha \in A\} = \{(x_\alpha) : x_\alpha \in X_\alpha\}$$

so that a point x in the product consists of "coordinates" x_α chosen from the X_α's. But what exactly does a symbol like $x = (x_\alpha)$ mean if there are "uncountably many coordinates?"

We can get an idea by first thinking about a countable product. For sets $X_1, X_2, ..., X_n$,... we can informally define the product set as a certain set of sequences: $X = \prod_{n=1}^{\infty} X_n = \{(x_n) : x_n \in X_n\}$. But if we want to be careful about set theory, then a legal definition of X should have the form $X = \{(x_n) \in U : x_n \in X_n\}$. From what "pre-existing set" U will the sequences in X be chosen? The answer is easy: given the sets X_n, the ZF axioms guarantee that the set $(\bigcup_{n=1}^{\infty} X_n)^{\mathbb{N}}$ exists. Then

$$X = \prod_{n=1}^{\infty} X_n = \{x \in (\bigcup_{n=1}^{\infty} X_n)^{\mathbb{N}} : x(n) = x_n \in X_n\}.$$

Thus the elements of $\prod_{n=1}^{\infty} X_n$ are certain functions (sequences) defined on the index set $\mathbb{N}$. This idea generalizes naturally to any product.

Definition 2.1 Let $\{X_\alpha : \alpha \in A\}$ be a collection of sets. We define the product set $X = \prod\{X_\alpha : \alpha \in A\} = \{x \in (\bigcup X_\alpha)^A : x(\alpha) \in X_\alpha\}$. The X_α's are called the factors of X. For each α, the function $\pi_\alpha : \prod\{X_\alpha : \alpha \in A\} \to X_\alpha$ defined by $\pi_\alpha(x) = x_\alpha$ is called the α^{th}-projection map. For $x \in X$, we write more informally $x_\alpha = x(\alpha) =$ the α^{th}-coordinate of x and write $x = (x_\alpha)$.

Caution: the index set A might not be ordered. So even though we use the informal notation $x = (x_\alpha)$, such phrases as "the first coordinate of x," "the next coordinate in x after x_α," and "the coordinate in x preceding x_α" may not make sense. The notation (x_α) is handy but can lead you into errors if you're not careful.

By definition, a point x in $\prod\{X_\alpha : \alpha \in A\}$ is a function that "chooses" a coordinate x_α from each set in the collection $\{X_\alpha : \alpha \in A\}$. To say that such a "choice function" x must exist if all the X_α's are nonempty is precisely the Axiom of Choice. (*See the discussion following Theorem I.6.8.*)

Theorem 2.2 The Axiom of Choice (AC) is equivalent to the statement that every product of nonempty sets is nonempty.

Note: In ZF set theory, certain special products can be shown to be nonempty without using AC. For example, if $X_n = \mathbb{N}$, then $\prod_{n=1}^{\infty} X_n = \{x \in \mathbb{N}^{\mathbb{N}} : x(n) \in \mathbb{N}\} = \mathbb{N}^{\mathbb{N}}$. Without using AC, we can precisely describe a point (= function) in the product − for example, the identity function $i = \{(m,n) \in \mathbb{N} \times \mathbb{N} : m = n\}$ − so $\mathbb{N}^{\mathbb{N}} = \prod_{n=1}^{\infty} X_n \neq \emptyset$. Can you give other similar examples?

We will often write $\prod\{X_\alpha : \alpha \in A\}$ as $\prod_{\alpha \in A} X_\alpha$. If the indexing set A is clearly understood, we may simply write $\prod X_\alpha$.

Example 2.3

1) If $A = \emptyset$, then $\prod\{X_\alpha : \alpha \in A\} = \{(\bigcup_{\alpha \in A} X_\alpha)^A : x(\alpha) \in X_\alpha\} = \{\emptyset\}$.

2) Suppose $X_{\alpha_0} = \emptyset$ for some $\alpha_0 \in A$. Then $\prod_{\alpha \in A} X_\alpha = \{(\bigcup_{\alpha \in A} X_\alpha)^A : x(\alpha) \in X_\alpha\}$. But $x(\alpha_0) \in X_{\alpha_0}$ is impossible so $\prod_{\alpha\alpha \in A} X_\alpha = \emptyset$.

3) Strictly speaking, we now have two different definitions for a finite product $X_1 \times X_2$:

i) $X_1 \times X_2 = \{(x_1, x_2) : x_1 \in X_1,\ x_2 \in X_2\}$ (a set of ordered pairs)

ii) $X_1 \times X_2 = \{x \in (X_1 \cup X_2)^{\{1,2\}} : x(i) \in X_i\}$ (a set of functions)

But there is an obvious way to regard these two sets as "the same": the ordered pair (x_1, x_2) corresponds to the function $x = \{(1, x_1), (2, x_2)\} \in (X_1 \cup X_2)^{\{1,2\}}$.

4) If $A = \mathbb{N}$, then $\prod\{X_n : n \in \mathbb{N}\} = \prod_{n=1}^{\infty} X_n = \{x \in (\bigcup_{n=1}^{\infty} X_n)^{\mathbb{N}} : x_n \in X_n\}$ $= \{(x_1, x_2, ..., x_n, ...,) : x_n \in X_n\} =$ the set of all sequences (x_n) where $x_n \in X_n$.

5) Suppose the X_α's are identical, say $X_\alpha = Y$ for all $\alpha \in A$. Then $\prod\{X_\alpha : \alpha \in A\}$ $= \{x \in (\bigcup_{\alpha \in A} X_\alpha)^A : x(\alpha) \in X_\alpha\} = \{x \in Y^A : x_\alpha \in Y\} = Y^A$. If $|A| = m$, we will sometimes write this product simply as Y^m = "the product of m copies of Y" because the <u>number</u> of factors m is often more important than the specific index set A.

6) Discuss: is the equation $\prod_{i \in I} A_i \cap \prod_{j \in J} B_j = \prod_{(i,j) \in I \times J} (A_i \cap B_j)$ always true? sometimes true? never true?

Now that we have a definition of the <u>set</u> $\prod\{X_\alpha : \alpha \in A\}$, we can think about a product <u>topology</u>. We begin by recalling the definition and a few basic facts about the "weak topology." (*See Example III.8.6.*)

Definition 2.4 Let X be a set. For each $\alpha \in A$, suppose $(X_\alpha, \mathcal{T}_\alpha)$ is a topological space and that $f_\alpha : X \to X_\alpha$. The weak topology on X generated by the collection $\mathcal{F} = \{f_\alpha : \alpha \in A\}$ is the smallest topology on X that makes all the f_α's continuous.

Certainly, there is at least one topology on X that makes all the f_α's continuous: the discrete topology. Since the intersection of a collection of topologies on X is a topology (*why?*), the weak topology exists – we can describe it "from the top down" as $\bigcap\{\mathcal{T}: \mathcal{T}$ is a topology on X that makes all the f_α's continuous$\}$.

However, this efficient description of the weak topology doesn't give us a useful idea about what sets are open. Usually it is more useful to describe the weak topology on X "from the bottom up." To make all the f_α's continuous it is necessary and sufficient that

for each $\alpha \in A$ and for each open set $U_\alpha \subseteq X_\alpha$, the set $f_\alpha^{-1}[U_\alpha]$ must be open.

Therefore the weak topology $\mathcal{T}$ is the smallest topology that contains all such sets $f_\alpha^{-1}[U_\alpha]$, and that is the topology for which $\mathfrak{S} = \{f_\alpha^{-1}[U_\alpha] : \alpha \in A, U_\alpha$ open in $X_\alpha\}$ is a subbase. (*See Example III.8.6.*)

Therefore a base for the weak topology consists of all finite intersections of sets from $\mathfrak{S}$. A typical basic open set has form $f_{\alpha_1}^{-1}[U_{\alpha_1}] \cap f_{\alpha_2}^{-1}[U_{\alpha_2}] \cap ... \cap f_{\alpha_n}^{-1}[U_{\alpha_n}]$ where each $\alpha_i \in A$ and each U_{α_i} is open in X_{α_i}. To cut down on symbols, we will use a special notation for these subbasic and basic open sets: we will write

$< U_\alpha > \ = f_\alpha^{-1}[U_\alpha] \ =$ a typical subbasic open set, and then
$< U_{\alpha_1} > \cap \ < U_{\alpha_2} > \cap ... \cap < U_{\alpha_n} > \ = U$ for a typical basic open set.
We then abbreviate further as $U = \ < U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} >$.
So $x \in U = \ < U_{\alpha_1}, U_{\alpha_2}, ...U_{\alpha_n} >$ iff $f_{\alpha_i}(x) \in U_{\alpha_i}$ for each $i = 1, ..., n$.

This notation is not standard—but it should be because it's very handy. You should verify that to get a base for the weak topology $\mathcal{T}$ on X, it is sufficient to use only the sets $< U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} >$ where each U_{α_i} is a basic (or even subbasic) open set in X_{α_i}.

Example 2.5 Suppose $A \subseteq (X, \mathcal{T})$ and let $i : A \to X$ be the inclusion map $i(a) = a$. Then the subspace topology on A is the same as the weak topology generated by $\mathcal{F} = \{i\}$. To see this, just notice that a base for the weak topology is $\{i^{-1}[U] : U$ open in $X\}$, and the sets $i^{-1}[U] = U \cap A$ are exactly the open sets in the subspace topology.

The following theorem tells us that a map f into a space X with the weak topology is continuous iff each composition $f_\alpha \circ f$ is continuous.

Theorem 2.6 Suppose $f : Z \to X$, where Z is a topological space and X has the weak topology generated by maps $f_\alpha : X \to (X_\alpha, \mathcal{T}_\alpha)$ $(\alpha \in A)$. Then f is continuous if and only if $f_\alpha \circ f : Z \to X_\alpha$ is continuous for every α.

Proof If f is continuous, then each composition $f_\alpha \circ f$ is continuous. Conversely, suppose each $f_\alpha \circ f$ is continuous. To show that f is continuous, it is sufficient to show that $f^{-1}[V]$ is open in Z whenever V is a subbasic open set in X (*why?*). So let $V = \ < U_\alpha >$ with U_α open in X_α. Then $f^{-1}[V] = f^{-1}[f_\alpha^{-1}[U_\alpha]] = (f_\alpha \circ f)^{-1}[U_\alpha]$, which is open because $f_\alpha \circ f$ is continuous. •

Definition 2.7 For each $\alpha \in A$, let $(X_\alpha, \mathcal{T}_\alpha)$ be a topological space. The product topology $\mathcal{T}$ on the set $\prod X_\alpha$ is the weak topology generated by the collection of projection maps $\mathcal{F} = \{\pi_\alpha : \alpha \in A\}$.

The product topology is sometimes called the "Tychonoff topology." *We always assume that a product $\prod X_\alpha$ of topological spaces has the product topology unless some other topology is explicitly stated.*

Because the product topology is a weak topology, a subbase consists of all sets of form $< U_\alpha > = \pi_\alpha^{-1}[U_\alpha]$, where U_α is open in X_α. A base, then, consists of all possible finite intersections of these sets :

$$< U_{\alpha_1} > \cap ... \cap < U_{\alpha_n} > = \pi_{\alpha_1}^{-1}[U_{\alpha_1}] \cap ... \cap \pi_{\alpha_n}^{-1}[U_{\alpha_n}] \qquad (n \in \mathbb{N})$$

$$= \{x \in \prod X_\alpha : x_{\alpha_i} \in U_{\alpha_i} \text{ for each } i = 1, ..., n\}$$

$$= \prod_{\alpha \in A} U_\alpha \ \underline{\text{where}}\ U_\alpha = X_\alpha \text{ for } \alpha \neq \alpha_1, \alpha_2, ..., \alpha_n \quad (*)$$

It is sufficient to use only U_α's which are basic (or even subbasic) open sets in X_α. Why?

A basic open set in $\prod X_\alpha$ "depends on only finitely many coordinates" in the following sense:

> $x \in < U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} >$ iff x satisfies the finitely many restrictions $x_{\alpha_i} \in U_{\alpha_i}$. The (basic) open sets containing x are what we use to describe "closeness" to x, so we can say informally that in the product topology "closeness depends on only finitely many coordinates."

If the index set A is finite, then the condition in (*) is satisfied automatically, and a base for the product topology is the set of all "open boxes" $\prod_{\alpha \in A} U_\alpha$:

$$\prod_{\alpha \in A} U_\alpha = \prod_{i=1}^n U_i = < U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} > = U_1 \times U_2 \times ... \times U_n$$

Thus, when A is finite, Definition 2.7 agrees with the earlier definition for product topologies in Chapter III (Example 5.11).

You might not have expected Definition 2.7. A "first guess" to define a topology on products might have been to use all boxes $\prod_{\alpha \in A} U_\alpha$ (U_α open in X_α) rather than the more restricted collection in (*). As just noted, that would be equivalent to Definition 2.7 for finite products, but not for infinite products. One can define a topology on the set $\prod_{\alpha \in A} X_\alpha$ using all boxes of the form $\prod_{\alpha \in A} U_\alpha$ as a base – an alternate topology called the box topology that contains, in general, many more open sets than the product topology because of the omission of the restriction on the U_α's in (*). We will try to indicate, below, why our definition of the product topology is the "right" one to use.

Theorem 2.8 Each projection $\pi_\alpha : \prod X_\alpha \to X_\alpha$ is continuous and open, and π_α is onto if $\prod X_\alpha \neq \emptyset$. A function $f : Z \to \prod X_\alpha$ is continuous if and only if $\pi_\alpha \circ f : Z \to X_\alpha$ is continuous for every α.

Proof Each π_α is onto if the product is nonempty (*why?*). By definition, the product topology makes all the π_α's continuous. To show that π_α is open, it is sufficient to show that the image of a basic open set $U = < U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} >$ is open.

This is clear if $U = \emptyset$, and

for $U \neq \emptyset$, we get $\pi_\alpha[U] = \pi_\alpha[\ < U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} > \ ] = \begin{cases} U_{\alpha_i} & \text{if } \alpha = \alpha_i \\ X_\alpha & \text{if } \alpha \neq \alpha_1, ..., \alpha_n \end{cases}$.

Finally, f is continuous iff each composition $\pi_\alpha \circ f$ is continuous because the product has the weak topology generated by the projections (Theorem 2.6). ●

Generally, projection maps are <u>*not*</u> *closed. For example,* $F = \{(x, y) \in \mathbb{R}^2 : y = \frac{1}{x},\ x > 0\}$ *is closed in* $\mathbb{R} \times \mathbb{R}$*, but its projection* $\pi_1[F]$ *is not closed in* $\mathbb{R}$.)

Example 2.9

1) A subbasic open set in $\mathbb{R} \times \mathbb{R}$ has form $\pi_1^{-1}[U] = U \times \mathbb{R}$ or $\pi_2^{-1}[V] = \mathbb{R} \times V$, where U and V are open in $\mathbb{R}$. (*We still get a subbase if we only use open* <u>*intervals*</u> U, V *in* $\mathbb{R}$.). Then basic open sets have form $(U \times \mathbb{R}) \cap (\mathbb{R} \times V) = U \times V$. Therefore the product topology on $\mathbb{R} \times \mathbb{R}$ is the usual topology on $\mathbb{R}^2$.

The function $f : \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ given by $f(t) = (t^2, \sin t^2)$ is continuous because the compositions $(\pi_1 \circ f)(t) = t^2$ and $(\pi_2 \circ f)(t) = \sin t^2$ are both continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

2) Let $X = \mathbb{N}^{\aleph_0} =$ "the product of countably many copies of $\mathbb{N}$." The singleton sets $\{n\}$ are a basis for $\mathbb{N}$. A base for the product topology consists of all sets $U = \prod U_n$ where finitely many U_n's are singletons and all the others are equal to $\mathbb{N}$. Each basic open set U is infinite (*in fact* $|U| = c$ *– why?*), so X has no isolated points and therefore X is not discrete. (*In fact,* $X \simeq \mathbb{P}$). By similar reasoning, an infinite product of discrete spaces, each with more than 1 point, is <u>not discrete</u>.

For each $x = (k_1, k_2, ..., k_n, ...) \in X$, the set $\{x\} = \prod_{n=1}^{\infty} \{k_n\}$ is open <u>in the box topology</u> so, in contrast, the box topology on X <u>is</u> the discrete topology. (*For a finite product, the box and product topologies care the same: a* <u>*finite*</u> *product of discrete spaces* <u>*is*</u> *discrete.*)

3) Consider $X = \prod \{X_r : r \in \mathbb{R}\}$ where each $X_r = \mathbb{R}$. By definition of product, each point in X is a function $f : \mathbb{R} \to \mathbb{R}$. In other words, $X = \mathbb{R}^{\mathbb{R}}$ and for each r, $\pi_r(f) = f(r)$. As basic open sets in X, we can use sets $U = \ < U_{r_1}, U_{r_2}, ..., U_{r_n} >$, where the U_{r_i}'s are open intervals in $\mathbb{R}$. Then $f \in U$ if and only if $f(r_i) \in U_{r_i}$ for each $i = 1, ..., n$. If g is <u>also</u> in U, then f and g are "close" at the finitely many points r_i – in the sense that $f(r_i)$ and $g(r_i)$ are both in the interval U_{r_i}. If, for example, the U_{r_i}'s each have diameter less than ϵ, then $|f(r_i) - g(r_i)| < \epsilon$ for each $i = 1, ..., n$. *Of course, this is much weaker than saying* f *and* g *are "uniformly" close.*

Why is the product topology the "correct" topology for set $\prod X_\alpha$? Of course there is no "right" answer, but a few observations should make it seem a good choice.

Example 2.10

1) For finite products, the box and product topologies are exactly the same. When it comes to <u>infinite</u> products, there's no obvious reason to favor the box topology or the product topology. Moreover, if one of them seems more natural, then at least we should be cautious: our intuition, after all, is only comfortable with finite sets, and we always run risks when we apply naive intuition to infinite collections.

2) Consider $(0,1)$: for two points $x = 0.x_1 \ldots x_n \ldots$ and $a = 0.a_1 \ldots a_n \ldots$, it will be true that "$x$ is close to a" when x and a agree in the first n decimal places (for a sufficiently large n). Roughly speaking, "closeness" depends on only finitely many decimal places ("coordinates").

Now consider the Hilbert cube $H = \prod_{n=1}^{\infty}[0, \frac{1}{n}] = [0,1] \times [0, \frac{1}{2}] \times \ldots \times [0, \frac{1}{n}] \times \ldots \subseteq \ell_2$, where ℓ_2 has its usual metric, d (*see Example II.2.6.6 and Exercise II.E10*). Suppose $x, a \in H$ and let $\epsilon > 0$. What condition on x will guarantee that $d(x,a) < \epsilon$?

Pick N so that $\sum_{i=N+1}^{\infty} \frac{1}{i^2} < \frac{\epsilon^2}{2}$. If $(x_i - a_i)^2 < \frac{\epsilon^2}{2N}$ for each $i = 1, \ldots, N$, then we have $d(x,a) = (\sum_{i=1}^{N}(x_i - a_i)^2 + \sum_{i=N+1}^{\infty}(x_i - a_i)^2)^{1/2} < (N \cdot \frac{\epsilon^2}{2N} + \frac{\epsilon^2}{2})^{1/2} = \epsilon$. Here, in the natural metric topology on the product H, we see that we can achieve "x close to a" by requiring "closeness" in just finitely many coordinates $1, \ldots, N$. This is just what the product topology does. In fact, the product topology on H turns out to be the topology $\mathcal{T}_d$.

A handy "rule of thumb" that has proved true every time I've used it is that if a topology on a product set is such that "closeness depends on only finitely many coordinates," then that topology is the product topology.

3) From a very pragmatic point of view, the product topology appears much more manageable. To "get your mind around" a basic open set $U = < U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n} >$ in the product topology, you only need to think about finitely many sets $U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_n}$; but in the box topology, thinking about U requires taking into account all the U_α's in $U = \prod_{\alpha \in \in} U_\alpha$ – and there may be uncountably many different U_α's.

4) The bottom line, however, is this: a mathematical definition justifies itself by the fruit it bears. The definition of the product topology will lead to some beautiful theorems. Using the product topology, for example, we will see that compact Hausdorff spaces are topologically nothing other than the closed subspaces of cubes $[0,1]^m$ (where m might be infinite). For the time being, you will need to accept that things work out nicely down the road, and that by contrast, the box topology turns out to be rather ill-behaved. (*See Exercise E11.*)

As a simple example of such nice behavior, the following theorem is exactly what one would hope for – and the proof depends on having the "correct" topology on the product. The theorem says that convergence of sequences in a product is "coordinatewise convergence" : that is, in a product, $(x_n) \to x$ iff for all α , the α^{th} coordinate of x_n converges (in X_α) to the α^{th} coordinate of x. For that reason, the product topology is sometimes called the "topology of coordinatewise convergence."

Theorem 2.11 Suppose (x_n) is a sequence in $X = \prod\{X_\alpha : \alpha \in A\}$. Then $(x_n) \to x \in X$ iff $(\pi_\alpha(x_n)) \to \pi_\alpha(x)$ in X_α for all $\alpha \in A$.

Proof If $(x_n) \to x$, then $(\pi_\alpha(x_n)) \to \pi_\alpha(x)$ because each π_α is continuous.

Conversely, suppose $(\pi_\alpha(x_n)) \to \pi_\alpha(x)$ in X_α for each α and consider any basic open set $U = < U_{\alpha_1}, U_{\alpha_2}, \ldots, U_{\alpha_k} >$ that contains $x = (x_\alpha)$. For each $i = 1, \ldots, k$, we have $x_{\alpha_i} \in U_{\alpha_i}$. Since $(\pi_{\alpha_i}(x_n)) \to \pi_{\alpha i}(x)$, we have $\pi_{\alpha_i}(x_n) \in U_{\alpha_i}$ for $n \geq$ some N_i. Let $N = \max\{N_1, \ldots, N_k\}$. Then for $n \geq N$ we have $\pi_{\alpha_i}(x_n) \in U_{\alpha_i}$ for every $i = 1, \ldots, k$. This means that $x_n \in U$ for $n \geq N$, so $(x_n) \to x$. ∙

In the proof, N is the max of a finite set. If X has the box topology, the basic open set $U = \prod U_\alpha$ might involve infinitely many open sets $U_\alpha \neq X_\alpha$. For each such α, we could pick an $N_\alpha \in \mathbb{N}$, just as in the proof. But the set of all N_α's might not have a max, N, and the proof would collapse. Can you create a specific example with the box topology where this happens?

Example 2.12 Consider $\mathbb{R}^{\mathbb{R}} = \prod\{X_r : r \in \mathbb{R}\}$, where $X_r = \mathbb{R}$. Each point f in the product is a function $f : \mathbb{R} \to \mathbb{R}$. Suppose that (f_n) is a sequence of points in $\mathbb{R}^{\mathbb{R}}$. By Theorem 2.11, $(f_n) \to f$ iff $(f_n(r)) \to f(r)$ for each $r \in \mathbb{R}$. With the product topology, convergence of a sequence of functions in $\mathbb{R}^{\mathbb{R}}$ is called (in analysis) pointwise convergence. *Question: if $\mathbb{R}^{\mathbb{R}}$ is given the box topology, is convergence of a sequence (f_n) simply uniform convergence (as defined in analysis)?*

The following "theorem" is stated loosely. You can easily create variations. Any reasonable version of the statement is probably true.

Theorem 2.13 Topological products are associative in any "reasonable" sense: for example, if the index set A is written as $A = B \cup C$ where B and C are disjoint, then

$$\begin{array}{ccccc} \prod\{X_\alpha : \alpha \in A\} & \simeq & \prod\{X_\beta : \beta \in B\} & \times & \prod\{X_\gamma : \gamma \in C\} \\ \| & & \| & & \| \\ X & & Y & \times & Z \end{array}$$

Proof A point $x \in \prod\{X_\alpha : \alpha \in A\}$ is a function $x : A \to \bigcup_{\alpha \in A} X_\alpha$. Define $f : X \to Y \times Z$ by $f(x) = (x|B, x|C)$. Clearly f is one-to-one and onto.

f is a mapping of X into a product $Y \times Z$, so f is continuous iff $\pi_1 \circ f$ and $\pi_2 \circ f$ are both continuous. But $\pi_1 \circ f$ is also a map into a product: $\pi_1 \circ f : X \to Y = \prod\{X_\beta : \in B\}$. So $\pi_1 \circ f$ is continuous if and only if $\pi_\beta \circ \pi_1 \circ f : \prod\{X_\alpha : \alpha \in A\} \to X_\beta$ is continuous for all $\beta \in B$. This is true because $\pi_\beta \circ \pi_1 \circ f = \pi_\beta : \prod\{X_\alpha : \alpha \in A\} \to X_\beta$. The proof that $\pi_2 \circ f$ is continuous is completely similar.

$f^{-1} : Y \times Z \to X = \prod\{X_\alpha : \alpha \in A\}$ is given by $x = f^{-1}(y, z) = y \cup z$ ($=$ *the union of two functions*), and f^{-1} is continuous iff $\pi_\alpha \circ f^{-1} : Y \times Z \to X_\alpha$ is continuous for each $\alpha \in A = B \cup C$. To check this, first suppose $\alpha \in B$: then $\pi_\alpha \circ f^{-1}(y, z) = \pi_\alpha(x)$, where $x = y \cup z$. Since $\alpha \in B$, $x_\alpha = (\pi_\alpha \circ \pi_1)(y, z)$, so $\pi_\alpha \circ f^{-1} = \pi_\alpha \circ \pi_1$, which is continuous. The case where $\alpha \in C$ is completely similar.

Therefore f is a homeomorphism. •

The question of topological commutativity for products only makes sense when the index set A is ordered in some way. But even then: if we view a product as a collection of functions, the question of commutativity is trivial – the question reduces to the fact that set theoretic unions are commutative.
For example, $X_1 \times X_2 = \{x \in (X_1 \cup X_2)^{\{1,2\}} : x(i) \in X_i \text{ for } i = 1, 2\}$
$= \{x \in (X_2 \cup X_1)^{\{1,2\}} : x(i) \in X_i \text{ for } i = 1, 2\} = X_2 \times X_1$.
So viewed as sets of functions, $X_1 \times X_2$ and $X_2 \times X_1$ are exactly the same set! The same observation applies to any product viewed as a collection of functions.

But we might look at an ordered product in another way: for example, thinking of $X_1 \times X_2$ and $X_2 \times X_1$ as sets of ordered pairs. Then generally $X_1 \times X_2 \neq X_2 \times X_1$. From that point of view, the topological spaces $X_1 \times X_2$ and $X_2 \times X_1$ are not literally identical, but there is a homeomorphism between them: $f(x_1, x_2) = (x_2, x_1)$. So the products are topologically identical. We can make a similar argument whenever the order of factors is "commuted" by permuting the index set.

The general rule of thumb is that "whenever it makes sense, topological products are commutative."

Exercise 2.14 X^m denotes the product of m copies of the space X. Prove that $(X^m)^n$ is homeomorphic to X^{mn}. (*Hint: The bijection ϕ in the proof of Theorem I.14.7 is a homeomorphism.*)

Notice that "cancellation properties" may not be true. For example, $\mathbb{N} \times \{0\}$ and $\mathbb{N} \times \{0, 1\}$ are homeomorphic (both are countable discrete spaces) but topologically you can't "cancel the $\mathbb{N}$" : $\{0\}$ is not homeomorphic to $\{0,1\}$!

Here are a few results which are quite simple but very handy to remember**.** The first states that singleton factors are topologically irrelevant in a product.

Lemma 2.15 $\prod_{\alpha \in A} X_\alpha \times \prod_{\beta \in B} \{p_\beta\} \simeq \prod_{\alpha \in A} X_\alpha$

Proof $\prod_{\beta \in B} \{p_\beta\}$ is itself a one-point space $\{p\}$, so we only need to prove that $\prod_{\alpha \in A} X_\alpha \times \{p\} \simeq \prod_{\alpha \in A} X_\alpha$. The map $f(x, p) = x$ is clearly a homeomorphism. •

Lemma 2.16 Suppose $\prod_{\alpha \in A} X_\alpha \neq \emptyset$. For any $B \subseteq A$, $\prod_{\alpha \in B} X_\alpha$ is homeomorphic to a subspace Z of $\prod_{\alpha \in A} X_\alpha$ – that is, $\prod_{\alpha \in B} X_\alpha$ can be embedded in $\prod_{\alpha \in A} X_\alpha$. In fact, if all the X_α's are T_1-spaces, then $\prod_{\alpha \in B} X_\alpha$ is homeomorphic to a closed subspace Z of $\prod_{\alpha \in A} X_\alpha$.

Proof Pick a point $p = (p_\alpha) \in \prod_{\alpha \in A} X_\alpha$. Then by Lemma 2.15,

$$\prod_{\alpha \in B} X_\alpha \simeq \prod_{\alpha \in B} X_\alpha \times \prod_{\alpha \in A - B} \{p_\alpha\} = Z \subseteq \prod_{\alpha \in A} X_\alpha$$

Now suppose all the X_α's are T_1. If $y \in (\prod_{\alpha \in A} X_\alpha) - Z$, then for some $\gamma \in A - B$, $y_\gamma \neq p_\gamma$. Since X_γ is a T_1 space, there is an open set U_γ in X_γ that contains y_γ but not p_γ. Then $y \in \; < U_\gamma >$ and

$$< U_\gamma > \; \cap \, (\prod_{\alpha \in B} X_\alpha \times \prod_{\alpha \in A - B} \{p_\alpha\}) = \emptyset.$$

Therefore Z is closed in $\prod_{\alpha \in A} X_\alpha$. •

Note: 1) Assume each $X_\alpha = \mathbb{R}$. Go through the preceding proof step-by-step when $A = \{1, ..., k\}$ and when $A = \mathbb{N}$ and

2) In the case $B = \{\alpha_0\}$, Lemma 2.16 says that each factor X_{α_0} is homeomorphic to subspace of $\prod_{\alpha \in A} X_\alpha$ (a closed subspace if all the X_α's are T_1).

3) Caution: Lemma 2.16 does not say that if all the X_α's are T_1, then every copy of X_{α_0} embedded in $\prod_{\alpha \in A} X_\alpha$ is closed: only that there exists a closed homeomorphic copy. (It is very easy to show a copy of $\mathbb{R}$ embedded in $\mathbb{R}^2$ that is not closed in $\mathbb{R}^2$, for example ... ?)

Lemma 2.17 Suppose $X = \prod\{X_\alpha : \alpha \in A\} \neq \emptyset$. Then X is a Hausdorff space (or, T_1-space) if and only if every factor X_α is a Hausdorff space (or, T_1-space).

Proof, for T_2 Suppose all the X_α's are Hausdorff. If $x \neq y \in X$, then $x_{\alpha_0} \neq y_{\alpha_0}$ for some α_0. Pick disjoint open sets U_{α_0} and V_{α_0} in X_{α_0} containing x_{α_0} and y_{α_0}. Then $< U_{\alpha_0} >$ and $< V_{\alpha_0} >$ are disjoint (basic) open sets in $\prod X_\alpha$ that contain x and y, so X is Hausdorff.

Conversely, suppose $X \neq \emptyset$. By Lemma 2.16, each factor X_α is homeomorphic to a subspace of X. Since a subspace of a Hausdorff space is Hausdorff (*why?*), each X_α is a Hausdorff space. •

Exercise 2.18 Prove Lemma 2.17 if "Hausdorff" is replaced by "T_1." (*The proof is similar but easier.*)

Theorem 2.19 The product of countably many two-point discrete spaces is homeomorphic to the Cantor set C.

Proof We will show that $\prod_{n=1}^\infty X_n \simeq C$, where each $X_n = \{0, 2\}$.

To construct C we defined, for each sequence $x = (x_1, x_2, ..., x_n, ...) \in \{0, 2\}^{\mathbb{N}}$, a descending sequence of closed sets $F_{x_1} \supseteq F_{x_1 x_2} \supseteq ... \supseteq F_{x_1 x_2 ... x_n} \supseteq ...$ in $[0, 1]$ whose intersection gave a unique point $p \in C$: $\{p\} = \bigcap_{n=1}^\infty F_{x_1 x_2 ... x_n}$ (*see Section IV.10*). For each n, we can write C as a union of 2^n disjoint clopen sets: $C = \bigcup_{(x_1, ..., x_n) \in \{0,2\}^n} (C \cap F_{x_1 x_2 ... x_n})$.

Define $f : C \to \prod_{n=1}^\infty X_n$ by $f(p) = x = (x_1, x_2, ..., x_n, ...)$. Clearly f is one-to-one and onto. To show that f is continuous at $p \in C$, it is sufficient to show that for each n, the function $\pi_n \circ f : C \to \{0, 2\}$ is continuous at p. Pick any $n \in \mathbb{N}$. For this n, there is a clopen set $U = C \cap F_{x_1 x_2 ... x_n}$ that contains p, and $(\pi_n \circ f)|U = x_n$. Thus, $\pi_n \circ f$ is <u>constant</u> on a neighborhood of p in C, so $\pi_n \circ f$ is continuous at p.

By Lemma 2.17, $\prod_{n=1}^\infty X_n$ is Hausdorff. Since f is a continuous bijection of a compact space onto a Hausdorff space, f is a homeomorphism (*why?*). •

Corollary 2.20 $\{0, 2\}^{\aleph_0}$ is compact.

This corollary is a very special case of the Tychonoff Product Theorem which states that any product of compact spaces is compact. The Tychonoff Product Theorem is much harder and will be proved in Chapter IX.

Corollary 2.21 The Cantor set is homeomorphic to a product of countably many copies of itself.

Proof By Exercise 2.14 above, $C^{\aleph_0} \simeq (\{0, 2\}^{\aleph_0})^{\aleph_0} \simeq \{0, 2\}^{\aleph_0 \cdot \aleph_0} \simeq \{0, 2\}^{\aleph_0} \simeq C$. The case of $C^n \simeq C$ for $n \in \mathbb{N}$ is similar. •

Example 2.22 Convince yourself that each assertion is true:

1) If X is the Sorgenfrey line, then $X \times X$ is the Sorgenfrey plane (*see Examples III.5.3 and III.5.4*).

2) Let S^1 be the unit circle in $\mathbb{R}^2$. Then $S^1 \times [0, 1]$ is homeomorphic to the cylinder $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, z \in [0, 1]\}$.

3) $S^1 \times S^1$ is homeomorphic to a torus ($=$ "the surface of a doughnut").

Exercises

E1. Does it ever happen that $X \times \{0\}$ open in $X \times \mathbb{R}$? If so, what is a necessary and sufficient condition on X for this to happen?

E2. a) Suppose X and Y are topological spaces and that $A \subseteq X$, $B \subseteq Y$. Prove that $\text{int}_{X\times Y}(A \times B) = \text{int}_X A \times \text{int}_Y B$: that is, "the interior of the product is the product of the interiors." (*By induction, the same result holds for any* <u>*finite*</u> *product*.) Give an example to show that the statement may be false for <u>infinite</u> products.

b) Suppose $A_\alpha \subseteq X_\alpha$ for all $\alpha \in A$. Prove that in the product $X = \prod X_\alpha$,

$$\text{cl}\,(\prod A_\alpha) = \prod \text{cl}\, A_\alpha.$$

Note: When the A_α's are closed, this shows that $\prod A_\alpha$ is closed: so "any product of closed sets is closed." Can you see any plausible reason why products of closures are better behaved than products of interiors?

c) Suppose $X = \prod X_\alpha \neq \emptyset$ and that $A_\alpha \subseteq X_\alpha$. Prove that $\prod A_\alpha$ is dense in X iff A_α is dense in X_α for each α. *Note: Part c) implies that a finite product of separable spaces is separable, but it doesn't tell us whether or not an infinite product of separable spaces is separable: why not?*

d) For each α, let $q_\alpha \in X_\alpha$. Prove that $B = \{x \in \prod X_\alpha : x_\alpha = q_\alpha$ for all but at most finitely many $\alpha\}$ is dense in $\prod X_\alpha$. *Note: Suppose $X = \mathbb{R}^{\mathbb{N}} = \prod_{n\in\mathbb{N}} X_n$ where each $X_n = \mathbb{R}$. Suppose each q_n is chosen to be a rational – say $q_n = 0$. Then what does d) imply about $\mathbb{R}^{\mathbb{N}}$?*

e) Let $\alpha_0 \in A$. Prove that $Y_{\alpha_0} = \{x \in \prod X_\alpha :\ x_\alpha = p_\alpha$ for all $\alpha \neq \alpha_0\}$ is homeomorphic to X_{α_0}. *Note: So any factor of a product has a "copy" of itself inside the product in a "natural" way. For example, in $\mathbb{R}^n$, the set of points where all coordinates except the first are 0 is homeomorphic to the first factor, $\mathbb{R}$..*

f) Give an example of infinite spaces X, Y, Z such that $X \times Y$ is homeomorphic to $X \times Z$ but Y is not homeomorphic to Z.

E3. Let X be a topological space and consider the "diagonal" Δ of $X \times X$:

$$\Delta = \{(x, x) : x \in X\} \subseteq X \times X.$$

a) Prove that Δ is closed in $X \times X$ if and only if X is Hausdorff.
b) Prove that Δ is open in $X \times X$ if and only if X is discrete.

E4. Suppose X is a Hausdorff space and that $X_\alpha \subseteq X$ for each $\alpha \in A$. Show that $Y = \bigcap\{X_\alpha :\ \alpha \in A\}$ is homeomorphic to a closed subspace of the product $\prod\{X_\alpha :\ \alpha \in A\}$.

E5. For each $n \in \mathbb{N}$, suppose D_n is a countable dense set in $(X_n, \mathcal{T}_n)$.

a) Prove that $D = \prod D_n$ is dense in $\prod X_n$.

b) Prove that $\prod X_n$ is separable. *Hint: Note that D might not be countable ! But closeness in a product depends on only finitely many coordinates.*

E6. a) Suppose $a_{i,j} \in \{0, 2\}$ for all $i, j \in \mathbb{N}$. Prove that there exists a sequence (j_k) in $\mathbb{N}$ such that, for each i, $\lim_{k \to \infty} a_{i,j_k}$ exists. *Hint: "picture" the $a_{i,j}$ in an infinite matrix. For each fixed j, the "j-th column" of the matrix is a point in the Cantor set $C = \{0, 2\}^{\aleph_0}$.*

b) In $\mathbb{R}$, $\sum_{n=1}^{\infty} a'_n$ is called a <u>subseries</u> of $\sum_{n=1}^{\infty} a_n$ if for every n $\begin{cases} a'_n = a_n & \text{or} \\ a'_n = 0 \end{cases}$

Prove that if $\sum_{n=1}^{\infty} a_n$ is absolutely convergent, then $S = \{s \in \mathbb{R} : s \text{ is the sum of a subseries of } \sum_{n=1}^{\infty} a_n\}$ is closed in $\mathbb{R}$. *Hint: "Absolute convergence" guarantees that every subseries converges. Each subseries $\sum_{n=1}^{\infty} a'_n$ can be associated in a natural way with a point $x \in \{0, 1\}^{\aleph_0}$. Consider the mapping $f : \{0, 1\}^{\aleph_0} \to \mathbb{R}$ given by $f(x) = \sum_{n=1}^{\infty} a'_n \in \mathbb{R}$. Must f be a homeomorphism?*

c) Suppose G is an open dense subset of the Cantor set C. Must Fr G be countable? *Hint: Consider $\{(a, b) \in C \times C : b \neq 0\}$.*

E7. Let $\mathbb{N}$ have the cofinite topology.

a) Does the product $\mathbb{N}^m$ have the cofinite topology? Does the answer depend on m?

b) Prove $\mathbb{N}^m$ is separable *Hint: When m is infinite, consider the simplest possible points in the product. Note: part b) implies that an arbitrarily large product of T_1 spaces with more than one point can be separable. However, that is false for Hausdorff spaces – see Theorem 3.8 later in this chapter.*

E8. We can define a topology on any set X by choosing a nonempty family of subsets $\mathcal{F}$ and defining closed sets to be all sets which can be written as an intersection of finite unions of sets from $\mathcal{F}$. $\mathcal{F}$ is called a subbase for the closed sets of X. (*This construction is "complementary" to generating a topology on X by using a collection of sets as a subbase for the open sets.*)

a) Verify that this procedure does define a topology on X.

b) Suppose X_α is a topological space. Give $\prod X_\alpha$ the topology for which collection of "closed boxes" $\mathcal{F} = \{\prod F_\alpha : F_\alpha$ closed in $X_\alpha\}$ is a subbase for the closed sets. Is this topology the product topology?

E9. Prove or disprove:

There exists a bijection $f : X = \{0,1\}^{\aleph_0} \to \mathbb{N}^{\aleph_0} = Y$ such that for all $x \in X$ and for all n, the first n coordinates of $y = f(x)$ are determined by the first m coordinates of x.

Here, m depends on n and x. More formally, we are asking whether there exists a bijection f such that:

$\forall x \in X \ \ \forall n \in \mathbb{N} \ \ \exists m \in \mathbb{N}$ *such that changing x_j for any $j > m$ does not change y_k for $k \leq n$*

Hint: Think about continuity and the definition of the product topology.

3. Productive Properties

We want to consider how some familiar topological properties behave with respect to products.

Definition 3.1 Suppose that, for each $\alpha \in A$, the space X_α has a certain property P. We say that

$$\text{the property } P \text{ is } \begin{cases} \underline{\text{productive}} & \text{if } \prod X_\alpha \text{ must have property } P \\ \underline{\text{countably productive}} & \text{if } \prod X_\alpha \text{ must have property } P \text{ when } A \text{ is countable} \\ \underline{\text{finitely productive}} & \text{if } \prod X_\alpha \text{ must have property } P \text{ when } A \text{ is finite} \end{cases}$$

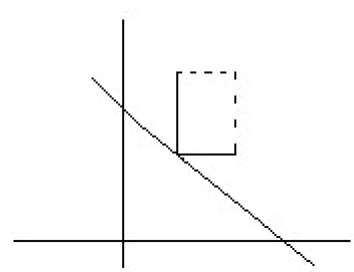
For example, Lemma 2.17 shows that the T_1 and T_2 properties are productive.

Some topological properties behave very badly with respect to products. For example, the Lindelöf property is a "countability property" of spaces, and we might expect the Lindelöf property to be countably productive. Unfortunately, this is not the case.

Example 3.2 The Lindelöf property is not finitely productive; in fact if X is a Lindelöf space, then $X \times X$ may not be Lindelöf. Let X be the set of real numbers with the topology for which a neighborhood base at a is $\mathcal{B}_a = \{[a, b) : a, b \in S,\ b > a\}$. (*Recall that X is called the Sorgenfrey Line: see Example III.5.3.*) We begin by showing that S is Lindelöf.

> It is sufficient to show that a collection $\mathcal{U}$ of <u>basic</u> open sets covering X has a countable subcover. Given such a cover $\mathcal{U}$, Let $\mathcal{V} = \{(a, b) : [a, b) \in \mathcal{U}\}$ and define $A = \bigcup \mathcal{V}$. For a moment, think of A as a subspace of $\mathbb{R}$ with its <u>usual</u> topology. Then A is Lindelöf (*why?*), and $\mathcal{V}$ is a covering of A by usual open sets, so there is a countable subfamily $\mathcal{V}'$ with $\bigcup \mathcal{V}' = A$.
>
> Now replace the left endpoints of the intervals in $\mathcal{V}'$ to get $\mathcal{U}' = \{[a, b) : (a, b) \in \mathcal{V}'\} \subseteq \mathcal{U}$. If $\mathcal{U}'$ covers X we are done, so suppose $X - \bigcup \mathcal{U}' \neq \emptyset$. For each $x \in X - \bigcup \mathcal{U}'$, pick a set $[a, b)$ in $\mathcal{U}$ that contains x. In fact, x <u>must</u> be the left endpoint of $[a, b)$ – because if $x \in [a, b) \in \mathcal{U}$ and $x \neq a$, then $x \in \bigcup \mathcal{V} = \bigcup \mathcal{V}' \subseteq \bigcup \mathcal{U}'$. So, for each $x \notin \bigcup \mathcal{U}'$ we can pick a set $[x, b_x) \in \mathcal{U}$.
>
> If x and y are distinct points not in $\bigcup \mathcal{U}'$, then $[x, b_x)$ and $[y, b_y)$ must be disjoint (*why?*) and there can be at most countably many disjoint intervals $[x, b_x)$. So $\mathcal{U}' \cup \{[x, b_x) : x \notin \bigcup \mathcal{U}'\}$ is a countable subcollection of $\mathcal{U}$ that covers S.

However the Sorgenfrey plane $S \times S$ is <u>not</u> Lindelöf. If it were, then its closed subspace $D = \{(x, y) : x + y = 1\}$ would also be Lindelöf (Theorem III.7.10). But that is impossible since D is uncountable and discrete in the subspace topology. (*See the figure on the following page.*)



Fortunately, many other topological properties do play more nicely with products. Here are several topological properties P to which the "same" theorem applies. We combine these into one large theorem for efficiency.

Theorem 3.3 Suppose $X = \prod\{X_\alpha : \alpha \in A\} \neq \emptyset$. Let P be one of the properties "first countable," "second countable," "metrizable," or "completely metrizable." Then X has property P iff

1) all the X_α's have property P, and
2) at most countably many X_α's are <u>non</u>trivial (i.e., do not have the trivial topology)

For all practical purposes, this theorem is a statement about countable products because:

1) The nontrivial X_α's are the "interesting" factors, and 2) says there are only countably many of them. In practice, one hardly ever works with trivial spaces, and if we totally exclude trivial spaces from the discussion, then the theorem just states that X has property P iff X is a countable product of spaces with property P.

2) A nonempty T_1-space X_α has the trivial topology iff $|X_\alpha| = 1$. So, if we are concerned only with T_1-spaces (as is most often the case) the theorem says that X has property P iff all the X_α's have property P and all but countably many of the X_α's are "topologically irrelevant" singletons. Of course, in the cases that involve metrizability, the T_1 condition is automatically satisfied.

Proof <u>Throughout the proof</u>, let $B =$ the set of "interesting indices" $= \{\alpha \in A : X_\alpha$ is a nontrivial space$\}$. We begin with the case $P =$ "first countable."

<u>Suppose 1) and 2) hold</u> and $x \in X$. We need to produce a countable neighborhood base at x. For each $\alpha \in B$, let $\{U_\alpha^n : n \in \mathbb{N}\}$ be a countable open neighborhood base at $x_\alpha \in X_\alpha$. Let

$$\mathcal{B}_x = \{U : U \text{ is a finite intersection of sets of the form } < U_\alpha^n > ,\ \alpha \in B,\ n \in \mathbb{N}\}$$

Since B is countable, $\mathcal{B}_x$ is a countable collection of open sets containing x and we claim that $\mathcal{B}_x$ is a neighborhood base at $x \in X$. To see this, suppose $V = \ < V_{\alpha_1}, ..., V_{\alpha_k} >$ is a basic open set containing x. (*We may assume that all α_i's are in B : why?*) For each $i = 1, ..., k$, pick $U_{\alpha_i}^{n_i}$ so that $x_{\alpha_i} \in U_{\alpha_i}^{n_i} \subseteq V_{\alpha_i}$. Then $U = \ < U_{\alpha_1}^{n_1}, U_{\alpha_2}^{n_2}, ..., U_{\alpha_k}^{n_k} > \ \in \mathcal{B}_x$ and $x \in U \subseteq V$.

Conversely, suppose X is first countable. We need to prove that 1) and 2) hold.

$X \neq \emptyset$, so by Lemma 2.16 each X_α is homeomorphic to a subspace of X. Therefore each X_α is first countable, so 1) holds.

For 2) we prove the contrapositive: assuming B is uncountable, we find a point $x \in X$ at which there cannot be a countable neighborhood base. Pick any point $p = (p_\alpha) \in X$

For each $\beta \in B$, pick an open set $O_\beta \subseteq X_\beta$ for which $\emptyset \neq O_\beta \neq X_\beta$. Choose $x_\beta \in O_\beta$ and $y_\beta \notin O_\beta$. Define $x = (x_\alpha) \in X$ using the coordinates

$$x_\alpha = \begin{cases} x_\beta & \text{if } \alpha = \beta \in B \\ p_\alpha \in X_\alpha & \text{if } \alpha \notin B \end{cases}$$

Suppose $\mathcal{B}_x$ is a countable collection of neighborhoods of x and for each $N \in \mathcal{B}_x$ pick a basic open set U with $x \in U = \, < U_{\alpha_1}, ..., U_{\alpha_k} > \, \subseteq N$. There are only finitely many α_i's involved in the expression for each chosen U, and there are only countably many N's in $\mathcal{B}_x$. So, since B is uncountable, we can pick a $\beta \in B$ that is not one of the α_i's involved in the expression for any of the sets U that were picked. Then

i) $< O_\beta >$ is an open set that contains x because $x_\beta \in O_\beta$

ii) for all $N \in \mathcal{B}_x$, $x \in N \nsubseteq \, < O_\beta >$, so $\mathcal{B}_x$ cannot be a neighborhood base at x. To see this, define a point $w = (w_\alpha)$ by

$$w_\alpha = \begin{cases} x_\alpha & \text{if } \alpha \neq \beta \\ y_\beta & \text{if } \alpha = \beta \end{cases}$$

Thus, w and x have the same coordinates except the β coordinate. For each $N \in B_x$, we picked $U = \, < U_{\alpha_1}, ..., U_{\alpha_k} > \, \subseteq N$. Since $x \in U$ and w has the same α_1,..., α_n coordinates as x, we also have $w \in U$. But $w \notin \, < O_\beta >$ because $w_\beta = y_\beta \notin O_\beta$. Therefore $U \nsubseteq \, < O_\beta >$, so $N \nsubseteq \, < O_\beta >$.

It's now easy to see that if the product $X = \prod X_\alpha$ has any of the other properties P, then conditions 1) and 2) must hold.

If X is second countable, metrizable or completely metrizable, then X is first countable so, by the first part of the proof, condition 2) must hold.

If X is second countable or metrizable then every subspace has these same properties – so each X_α is second countable or metrizable respectively. If X is completely metrizable, then X and all the subspaces X_α are T_1. By Lemma 2.16, X_α is homeomorphic to a closed – therefore complete – subspace of X. Therefore X_α is completely metrizable.

It remains to show that if $P =$ "second countable," "metrizable," or "completely metrizable" and conditions 1) and 2) hold, then $X = \prod X_\alpha$ also has property P.

Suppose $P =$ "<u>second countable</u>."

For each α in the countable set B, let $\mathcal{B}_\alpha = \{O_\alpha^1, O_\alpha^2, ..., O_\alpha^n, ...\}$ be a countable base for X_α and let

$$\mathcal{B} = \{O : O \text{ is a finite intersection of sets of form } < O_\alpha^n > \text{, where } \alpha \in B \text{ and } n \in \mathbb{N}\}$$

$\mathcal{B}$ is countable and we claim $\mathcal{B}$ is a base for the product topology on X.

Suppose $x \in V = \; < V_{\alpha_1}, ..., V_{\alpha_k} >$, a basic open set in X. For each $i = 1, ..., k$, $x_{\alpha_i} \in V_{\alpha_i}$ so we can choose a basic open set in X_{α_i} such that $x_{\alpha_i} \in O_{\alpha_i}^{n_i} \subseteq V_{\alpha_i}$. Then $x \in O = \; < O_{\alpha_1}^{n_1}, O_{\alpha_2}^{n_2}, ..., O_{\alpha_k}^{n_k} > \; \subseteq V$ and $O \in \mathcal{B}$. Therefore V can be written as a union of sets from $\mathcal{B}$, so $\mathcal{B}$ is a base for X.

Suppose $P =$ "<u>metrizable.</u>"

Since all the X_α's are T_1, condition 2) implies that all but countably many X_α's are singletons, which we can omit without changing X topologically. Therefore it is sufficient to prove that if each space $X_1, X_2, ..., X_n, ...$ is metrizable, then $X = \prod_{n=1}^\infty X_n$ is metrizable.

Let d_n be a metric for X_n, where without loss of generality, we can assume each $d_n \leq 1$ (*why?*). For points $x = (x_n),\; y = (y_n) \in X$, define

$$d(x, y) = \sum_{n=1}^{\infty} \tfrac{d_n(x_n, y_n)}{2^n}$$

Then d is a metric on X (*check*!) and <u>we claim</u> that $\mathcal{T}_d$ is the product topology $\mathcal{T}$. Because X is a countable product of first countable spaces, X is first countable, so $\mathcal{T}$ can be described using sequences: <u>it is sufficient to show</u> that $(z_k) \to z$ in $(X, \mathcal{T})$ iff $(z_k) \to z$ in $(X, \mathcal{T}_d)$. But $(z_k) \to z$ in $(X, \mathcal{T})$ iff the (z_k) converges "coordinatewise." Therefore <u>it is sufficient to show</u> that:

$(z_k) \to z$ in $(X, \mathcal{T}_d)$ iff $\forall n \; (z_k(n)) \to z(n)$ in (X_n, d_n) or equivalently,
$\forall n \;\; d_n(z_k(n), z(n)) \to 0$

i) Suppose $d(z_k, z) \to 0$. Let $\epsilon > 0$ and consider any particular n_0. We can choose K so that $k \geq K$ implies $d(z_k, z) = \sum_{n=1}^{\infty} \frac{d_n(z_k(n),\, z(n))}{2^n} < \frac{\epsilon}{2^{n_0}}$. Then for $k \geq K$, $\frac{d_{n_0}(z_k(n_0),\, z(n_0))}{2^{n_0}} < \frac{\epsilon}{2^{n_0}}$, so $d_{n_0}(z_k(n_0),\, z(n_0)) < \epsilon$.
Therefore $d_{n_0}(z_k(n_0),\, z(n_0)) \to 0$.

ii) On the other hand, suppose $d_n(z_k(n), z(n)) \to 0$ for every n and let $\epsilon > 0$. Choose N so that $\sum_{n=N+1}^{\infty} \frac{1}{2^n} < \frac{\epsilon}{2}$ and then choose K so that if $k \geq K$

$$\begin{aligned} &d_1(z_k(1), z(1))/2^1 < \tfrac{\epsilon}{2N} \\ &d_2(z_k(2), z(2))/2^2 < \tfrac{\epsilon}{2N} \\ &\quad\vdots \\ &d_N(z_k(N), z(N))/2^N < \tfrac{\epsilon}{2N}\;. \end{aligned}$$

Then for $k \geq K$ we have $d(z_k, z) = \sum_{n=1}^{\infty} \frac{d_n(z_k(n), z(n))}{2^n}$

$= \sum_{n=1}^{N} \frac{d_n(z_k(n), z(n))}{2^n} + \sum_{n=N+1}^{\infty} \frac{d_n(z_k(n), z(n))}{2^n} < N \cdot \frac{\epsilon}{2N} + \frac{\epsilon}{2} = \epsilon.$

Therefore $d(z_k, z) \to 0$.

Suppose $P =$ "completely metrizable."

Just as for $P =$ "metrizable," we can assume $X = \prod_{n=1}^{\infty} X_n$ and that d_n is a complete metric on X_n with $d_n \leq 1$. Using these d_n's, we define d in the same way. Then $\mathcal{T}_d$ is the product topology on X. We only need to show that (X, d) is complete.

Suppose (z_k) is a Cauchy sequence in (X, d). From the definition of d it is easy to see that $(z_k(n))$ is Cauchy in (X_n, d_n) for each n, so that $(z_k(n)) \to$ some point $a_n \in X_n$.
Let $a = (a_n) \in X$. Since $(z_k(n)) \to a_n$ for each n, we have $(z_k) \to a$ in the product topology $= \mathcal{T}_d$. Therefore (X, d) is complete. •

What is the correct formulation and proof of the theorem for $P =$ "pseudometrizable" ?

We might wonder why $P =$ "separable" is not included in Theorem 3.3. Since "separable" is a "countability property," we might hope that separability is preserved in countable products – although our experience Lindelöf spaces could make us hesitate. The explanation for the omission is that separability is actually better behaved for products than the other properties. Surprisingly, the product of as many as c separable spaces is separable, and the product of more than c nontrivial separable spaces can sometimes be separable. (*You should try to prove directly that a countable product of separable spaces is separable – remembering that in the product topology, "closeness depends on finitely many coordinates." If necessary, first look at finite products.*)

We begin the treatment of separability and products with a simple lemma which is merely set theory.

Lemma 3.4 Suppose $|A| \leq c$. There exists a countable collection $\mathcal{R}$ of subsets of A with the following property: given distinct $\alpha_1, \alpha_2, ..., \alpha_n \in A$, there are disjoint sets $A_1, A_2, ..., A_n \in \mathcal{R}$ such that $\alpha_i \in A_i$ for each i.

Proof (*Think about how you would prove the theorem if $A = \mathbb{R}$. If you do that, then you'll see that the general case is just a "carry over" of that proof.*)

Since $|A| \leq c$, there is a one-to-one map $\phi : A \to \mathbb{R}$. Let $\mathcal{R} = \{\phi^{-1}[\,(a, b)\,] : a, b \in \mathbb{Q}\}$. For distinct $\alpha_1, \alpha_2, ..., \alpha_n \in A$, we know that $\phi(\alpha_1), \phi(\alpha_2), ..., \phi(\alpha_n)$ are distinct real numbers. Then we can choose $a_i, b_i \in \mathbb{Q}$ so that $\phi(\alpha_i) \in (a_i, b_i)$ and so that the intervals (a_i, b_i) are pairwise disjoint. Then the sets $A_i = \phi^{-1}[(a_i, b_i)] \in \mathcal{R}$ are the ones we need. •

Theorem 3.5

1) Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$. If X is separable, then each X_α is separable.

2) If each X_α is separable and $|A| \leq c$, then $X = \prod_{\alpha \in A} X_\alpha$ is separable.

Part 2) of Theorem 3.5 is attributed (independently) to several people. In a slightly more general version, it is sometimes called the "Hewitt-Marczewski-Pondiczery Theorem." Here is an amusing sidelight, written by topologist Melvin Henriksen online in the Topology Atlas. A few words have been modified to conform with our notation:

> *Most topologists are familiar with the Hewitt-Marczewski-Pondiczery theorem. It states that if* m *is an infinite cardinal, then a product of* 2^m *topological spaces, each of which has a dense set of cardinality* $\leq m$, *also has a dense set with* $\leq m$ *points. In particular, the product of* c *separable spaces is separable (where* c *is the cardinal number of the continuum). Hewitt's proof appeared in [Bull. Amer. Math. Soc. 52 (1946), 641-643], Marczewski's proof in [Fund. Math. 34 (1947), 127-143], and Pondiczery's in [Duke Math. 11 (1944), 835-837]. A proof and a few historical remarks appear in Chapter 2 of Engelking's General Topology. The spread in the publication dates is due to dislocations caused by the Second World War; there is no doubt that these discoveries were made independently.*
>
> *Hewitt and Marczewski are well-known as contributors to general topology, but who was (or is) Pondiczery? The answer may be found in Lion Hunting & Other Mathematical Pursuits, edited by G. Alexanderson and D. Mugler, Mathematical Association of America, 1995. It is a collection of memorabilia about Ralph P. Boas Jr. (1912-1992), whose accomplishments included writing many papers in mathematical analysis as well as several books, making a lot of expository contributions to the American Mathematical Monthly, being an accomplished administrator (e.g., he was the first editor of Mathematical Reviews (MR) who set the tone for this vitally important publication, and was the chairman for the Mathematics Department at Northwestern University for many years and helped to improve its already high quality), and helping us all to see that there is a lot of humor in what we do. He wrote many humorous articles under pseudonyms, sometimes jointly with others. The most famous is "A Contribution to the Mathematical Theory of Big Game Hunting" by H. Petard that appeared in the Monthly in 1938. This book is a delight to read.*
>
> *In this book, Ralph Boas confesses that he concocted the name from Pondicheree (a place in India fought over by the Dutch, English and French), changed the spelling to make it sound Slavic, and added the initials E.S. because he contemplated writing spoofs on extra-sensory perception under the name E.S. Pondiczery. Instead, Pondiczery wrote notes in the Monthly, reviews for MR, and the paper that is the subject of this article. It is the only one reviewed in MR credited to this pseudonymous author.*
>
> *One mystery remains. Did Ralph Boas have a collaborator in writing this paper? He certainly had the talent to write it himself, but facts cannot be established by deduction alone. His son Harold (also a mathematician) does not know the answer to this question...*

Proof 1) Let D be a countable dense set in X. For each α, $\pi_\alpha[D]$ is countable and dense in X_α (because $X_\alpha = \pi_\alpha[X] = \pi_\alpha[\text{cl}\, D] \subseteq \text{cl}\, \pi_\alpha[D]$). Therefore each X_α is separable.

2) Choose a family $\mathcal{R}$ as in Lemma 3.4 and for each α, let $D_\alpha = \{x_\alpha^1, x_\alpha^2, ..., x_\alpha^n, ...\}$ be a countable dense set in X_α. Define a countable set $\mathcal{S}$ by

$$\mathcal{S} = \{(A_1, ..., A_n,\, l_1, ..., l_n) : n \in \mathbb{N},\, l_i \in \mathbb{N},\, A_i \in \mathcal{R} \text{ with the } A_i\text{'s pairwise disjoint}\}.$$

For each α pick a point $p_\alpha \in X_\alpha$, and for each $2n$-tuple $s = (A_1, ..., A_n, l_1, ..., l_n) \in \mathcal{S}$, define a point $x_s \in X$ with coordinates

$$x_s(\alpha) = \begin{cases} x_\alpha^{l_i} & \text{if } \alpha \in A_i \\ p_\alpha & \text{if } \alpha \notin \bigcup A_i \end{cases}$$

Let $D = \{x_s : s \in \mathcal{S}\}$. D is countable and we claim that D is dense in X. To see this, consider any nonempty basic open set $U = \ < U_{\alpha_1}, ..., U_{\alpha_k} >$: we will show that $U \cap D \neq \emptyset$.

For $U = \ < U_{\alpha_1}, ..., U_{\alpha_k} > \ \neq \emptyset$,

i) Choose disjoint sets $A_1, ..., A_k$ in $\mathcal{R}$ so that $\alpha_1 \in A_1, ..., \alpha_k \in A_k$.

ii) For each $i = 1, ..., k$, D_{α_i} is dense in X_{α_i} and we can pick a point $x_{\alpha_i}^{n_i}$ in $D_{\alpha_i} \cap U_{\alpha_i}$ $= \{x_{\alpha_i}^1, x_{\alpha_i}^2, ..., x_{\alpha_i}^n, ...\} \cap U_{\alpha_i}$.

Then, let $s = (A_1, ..., A_i, ..., A_k, n_1, ...n_i, ..., n_k) \in \mathcal{S}$. Because $\alpha_i \in A_i$, we have $x_s(\alpha_i) = x_{\alpha_i}^{n_i} \in U_{\alpha_i}$. Therefore $x_s \in U \cap D$. ●

Example 3.6 The rather abstract construction of a dense set D in the proof of Theorem 3.5 can be nicely illustrated with a concrete example. Consider $\mathbb{R}^{\mathbb{R}}$ ($= \prod_{r \in \mathbb{R}} X_r$, where each $X_r = \mathbb{R}$). Choose $\mathcal{R}$ to be the collection of all open intervals (a, b) with rational endpoints, and make a list these intervals as $A_1, ..., A_n, ...$. In each X_r, choose $D_r = \mathbb{Q} = \{q^1, q^2, ...q^n, ...\}$. (*Since all the D_r's are identical, we can omit the subscript "r" on the points; but just to stay consistent with the notation in the proof, we still use superscripts to index the q's.*) For each r, (arbitrarily) pick $p_r = 0 \in X_r$.

One example of a 6-tuple in the collection $\mathcal{S}$ is $s = (A_6, A_2, A_5, 2, 7, 4)$, where A_1, A_2, A_3 are disjoint open intervals with rational endpoints. The corresponding point $x_s \in \mathbb{R}^{\mathbb{R}}$ is the function $x_s : \mathbb{R} \to \mathbb{R}$ with

$$x_s(r) = \begin{cases} q^2 & \text{for } r \in A_6 \\ q^7 & \text{for } r \in A_2 \\ q^4 & \text{for } r \in A_5 \\ 0 & \text{for } r \notin A_6 \cup A_5 \cup A_2 \end{cases}$$

The dense set D consists of all step functions (such as x_s) that are 0 <u>outside</u> a finite union $A_{n_1} \cup A_{n_2} \cup ... \cup A_{n_k}$ of disjoint open intervals with rational endpoints and which have a constant rational value on each A_i.

<u>Caution</u>: In Example 2.12 we saw that the product topology on $\mathbb{R}^{\mathbb{R}}$ is the topology of pointwise convergence – that is, $(f_n) \to f$ in $\mathbb{R}^{\mathbb{R}}$ iff $(f_n(r)) \to f(r)$ for each $r \in \mathbb{R}$. But $\mathbb{R}^{\mathbb{R}}$ is not first countable (*why?*) so we cannot say that sequences are sufficient to describe the topology. In particular, if $f \in \mathbb{R}^{\mathbb{R}}$, then $f \in \text{cl}\, D$ but we <u>cannot</u> say that there must be sequence of step functions from D that converges pointwise to f.

Since $\mathbb{R}^{\mathbb{R}}$ is not first countable, $\mathbb{R}^{\mathbb{R}}$ is an example of a separable space that is neither second countable nor metrizable.

In contrast to the properties discussed in Theorem 3.3, an arbitrarily large product of nontrivial separable spaces can sometimes turn out to be separable, as the next example shows. However, Theorem 3.8 shows that for Hausdorff spaces, a nonempty product with more than c factors cannot be separable.

Example 3.7 For each $\alpha \in A$, let X_α be a set with $|X_\alpha| > 1$. Choose $p_\alpha \in X_\alpha$ and let $\mathcal{T}_\alpha = \{O \subseteq X_\alpha : p_\alpha \in O\} \cup \{\emptyset\}$. The singleton set $\{p_\alpha\}$ is dense in $(X_\alpha, \mathcal{T}_\alpha)$, so X_α is separable.
If $p = (p_\alpha) \in X = \prod_{\alpha \in A} X_\alpha$, then singleton set $\{p\}$ is dense (*why? look at a nonempty basic open set* U.) So X is separable, and this does not depend on $|A|$..

In this example, the X_α's are not T_1. But Exercise E7 shows that an arbitrarily large product of separable T_1-spaces can turn out to be separable.

Theorem 3.8 Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$ where each X_α is a T_2-space with more than one point.
If X is separable, then $|A| \leq c$.

Proof For each α, we can pick a pair of disjoint, nonempty open sets U_α and V_α in X_α. Let D be a countable dense set in X and let $D_\alpha = \, < U_\alpha > \cap D$ for each α. If $\alpha \neq \beta \in A$, there is a point $p \in \, < U_\alpha, V_\beta > \cap D$ because D is dense. Then $p \in D_\alpha$ but $p \notin D_\beta = \, < U_\beta > \cap D$ since $x_\beta \notin U_\beta$: therefore $D_\alpha \neq D_\beta$. Therefore the map $\phi : A \to \mathcal{P}(D)$ given by $\phi(\alpha) = D_\alpha$ is one-to-one, so $|A| \leq |\mathcal{P}(D)| = 2^{\aleph_0} = c$. ∙

We saw in Corollary V.2.19 that a finite product of connected spaces is connected. The following theorem shows that connectedness actually behaves very nicely with respect to all products. The proof of the theorem is interesting because, unlike previous proofs about products, this proof uses the theorem about finite products to prove the general case.

Theorem 3.9 Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$. X is connected if and only if each X_α is connected.

Proof Suppose X is connected. Since $X \neq \emptyset$, we have $\pi_\alpha[X] = X_\alpha$ for each α. A continuous image of a connected space is connected (Example V.2.6), so each X_α is connected.

Conversely, suppose each X_α is connected. For each α, pick a point $p_\alpha \in X_\alpha$. For each finite set $F \subseteq A$, let $X_F = \prod_{\alpha \in F} X_\alpha \times \prod_{\alpha \in A - F} \{p_\alpha\}$. X_F is homeomorphic to the finite product $\prod_{\alpha \in F} X_\alpha$, so each X_F is connected. Let $D = \bigcup \{X_F : F \text{ is a finite subset of } A\}$. Each X_F contains the point $p = (p_\alpha)$, so Corollary V.2.10 tells us that D is connected. Unfortunately, $D \neq X$ (*except in trivial cases; why?*).

But we claim that D is dense in X. We need to show that $D \cap U \neq \emptyset$ for every nonempty basic open set in X. If $U = \, < U_{\alpha_1}, ..., U_{\alpha_n} >$, choose a point $x_{\alpha_i} \in U_{\alpha_i}$ for each $i = 1, ..., n$, and define a point $x \in X$ with coordinates

$$x(\alpha) = \begin{cases} x_{\alpha_i} & \text{if } \alpha = \alpha_i \\ p_\alpha & \text{if } \alpha \neq \alpha_1, \alpha_2, ..., \alpha_n \end{cases}$$

Let $F = \{a_1, \alpha_2, ..., \alpha_n\}$. Then $x \in X_F \cap U \subseteq D \cap U$, so $D \cap U \neq \emptyset$.

Therefore $X = \operatorname{cl} D$ is connected (Corollary V.2.20). ∙

Question: is the analogue of Theorem 3.9 true for path connected spaces?

Just for reference, we state one more theorem here. We will not prove the theorem until Chapter IX, but we may use it in examples. (*Of course, the proof in Chapter IX will not depend on any of these examples!*)

Theorem 3.10 (Tychonoff Product Theorem) Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$. Then X is compact if and only if each X_α is compact.

One half ($\Rightarrow$) of the proof of Tychonoff's Theorem is very easy (*why?)*, and the easy proof that a finite product of compact spaces is compact was in Exercise IV.E.26.

Exercises

E10. Let $X = [0,1]^{[0,1]}$ have the product topology.

a) Prove that the set of all functions in X with finite range is dense in X. (*Here, we will call such functions step functions. In other settings, the definition of "step function" is more restrictive.*)

b) By Theorem 3.5, X is separable. Describe a countable set of step functions which is dense in X.

c) Let $A = \{x \in X : x$ is the characteristic function of a singleton set $\{r\}\}$. Prove that A, with the subspace topology, is discrete and not separable. Is A closed?

d) Prove that A has exactly one limit point, z, in X and that if N is a neighborhood of z, then $A - N$ is finite.

E11. "Boxes" of the form $\prod\{U_\alpha : \alpha \in A\}$, where U_α is open in X_α, are a base for the box topology on $\prod_{\alpha \in A} X_\alpha$. Throughout this problem, we assume that products have the box topology rather than the usual product topology.

a) Show that the "diagonal map" $f: \mathbb{R} \to \mathbb{R}^{\aleph_0}$ given by $f(x) = (x, x, x, ...)$ is not continuous, but that its composition with each projection map is continuous.

b) Show that $[0,1]^{\aleph_0}$ is not compact. *Hint: let $A_0 = [0,1)$ and $A_1 = (0,1]$. Consider the collection $\mathcal{U}$ of all sets of the form $A_{\epsilon_1} \times A_{\epsilon_2} \times ... \times A_{\epsilon_n} \times ...$, where $(\epsilon_1, \epsilon_2, ..., \epsilon_n, ...) \in \{0,1\}^{\aleph_0}$. By contrast, the Tychonoff Product Theorem (3.10) implies that $[0,1]^m$ (with the product topology) is compact for any cardinal m.*

c) Show that $\mathbb{R}^{\aleph_0}$ is not connected by showing that the set $A = \{x \in \mathbb{R}^{\aleph_0} : x$ is an unbounded sequence in $\mathbb{R}\}$ is clopen.

d) Suppose (X, d) and (X_α, d_α) $(\alpha \in A)$ are metric spaces. Prove that a function $f : X \to \prod X_\alpha$ (with the box topology) is continuous iff each coordinate function $f_\alpha = \pi_\alpha \circ f$ is continuous and each $x \in X$ has a neighborhood on which all but a finite number of the f_α's are constant.

E12. State and prove a theorem that gives a necessary and sufficient condition for a product of spaces to be path connected.

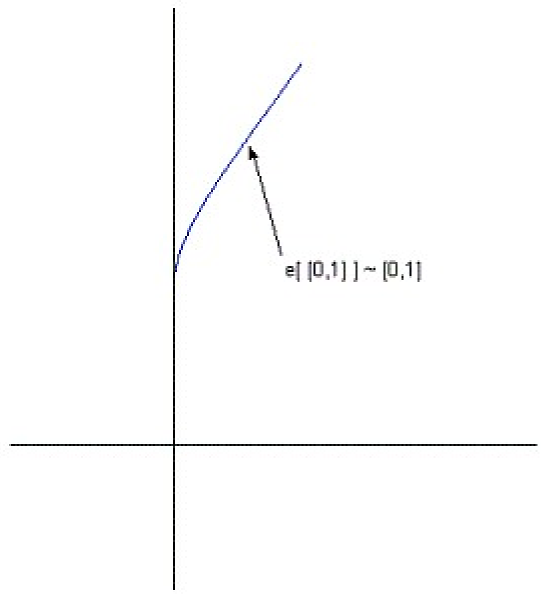
E13. Prove the following more general version of Theorem 3.8:

Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$, and that, for each $\alpha \in A$, there exist disjoint nonempty open sets U_α and V_α in X_α. If X is separable, then $|A| \leq c$.

4. Embedding Spaces in Products

If there is a homeomorphism $h : X \overset{\text{into}}{\to} Y$, then $X \simeq h[X] \subseteq Y$. We say then that X is embedded in and call h an embedding. Phrased differently, h shows that X is homeomorphic to a subspace of Y and, speaking topologically, we might say X "is" a subspace of Y. It is often possible to embed a space X in a product $Y = \prod X_\alpha$. Such embeddings will give us some nice theorems – for example, we will see that there is a separable metric space Y that contains (topologically) all other separable metric spaces – Y is a "universal" separable metric space.

To illustrate the embedding technique that we use, consider two functions $f_1 : [0,1] \to \mathbb{R}$ and $f_2 : [0,1] \to \mathbb{R}$ given by $f_1(x) = x^2$ and $f_2(x) = e^x$. Using f_1 and f_2, we can define $e : [0,1] \to \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ by using f_1 and f_2 as "coordinate functions": $e(x) = (f_1(x), f_2(x))$ $= (x^2, e^x)$. This map e is called the evaluation map defined by the set of functions $\{f_1, f_2\}$. In this example, e is an embedding – that is, e is a homeomorphism of $[0,1]$ into $\mathbb{R}^2$, so that $[0,1] \simeq \text{ran}(e) \subseteq \mathbb{R}^2$ (*see the figure*).



An evaluation map does not always give an embedding: for example, the evaluation map $e : [0,1] \to \mathbb{R}^2$ defined by the family $\{\cos 2\pi x, \sin 2\pi x\}$ is not a homeomorphism between $[0,1]$ and $\text{ran}(e) \subseteq \mathbb{R}^2$ (*why? what is ran*(e)*?*)

We want to generalize the idea of an evaluation map e into a product and to find conditions under which e will be an embedding.

Definition 4.1 Suppose X and X_α $(\alpha \in A)$ are topological spaces and that $f_\alpha : X \to X_\alpha$ for each α. The evaluation map defined by the family $\{f_\alpha : \alpha \in A\}$ is the function $e : X \to \prod_{\alpha \in A} X_\alpha$ given by

$$e(x)(\alpha) = f_\alpha(x)$$

Thus, $e(x)$ is the point in the product $\prod X_\alpha$ whose α^{th} coordinate is $f_\alpha(x)$:. In more informal coordinate notation, $e(x) = (f_\alpha(x))$.

Exercise 4.2 Suppose $X = \prod_{\alpha \in A} X_\alpha$. For each α, there is a projection map $\pi_\alpha : X \to X_\alpha$. What is the evaluation map defined by the family $\{\pi_\alpha : \alpha \in A\}$?

Definition 4.3 Suppose X and X_α $(\alpha \in A)$ are topological spaces and that $f_\alpha : X \to X_\alpha$. We say that the family $\{f_\alpha : \alpha \in A\}$ separates points if, for each pair of points $x \neq y \in X$, there exists an $\alpha \in A$ for which $f_\alpha(x) \neq f_\alpha(y)$.

Clearly, the evaluation map $e : X \to \prod_{\alpha \in A} X_\alpha$ is one-to-one $\Leftrightarrow$ for all $x \neq y \in X$, $e(x) \neq e(y)$
$\Leftrightarrow$ for all $x \neq y \in X$ there is an $\alpha \in A$ for which $e(x)(\alpha) = f_\alpha(x) \neq f_\alpha(y) = e(y)(\alpha)$
$\Leftrightarrow$ the family $\{f_\alpha : \alpha \in A\}$ separates points.

Theorem 4.4 Suppose X has the weak topology generated by the maps $f_\alpha : X \to (X_\alpha, \mathcal{T}_\alpha)$ and that the family $\{f_\alpha : \alpha \in A\}$ separates points. Then e is an embedding – that is, e is a homeomorphism between X and $e[X] \subseteq \prod X_\alpha$.

Proof Since $\{f_\alpha\}$ separates points, e is one-to-one and e is continuous because each composition $\pi_\alpha \circ e = f_\alpha$ is continuous.

e preserves unions and also (since e is one-to-one) intersections. Therefore. to check that e is an open map from X to $e[X]$, it is sufficient to show that e maps subbasic open sets in X to open sets in $e[X]$. Because X has the weak topology, a subbasic open set has the form $U = f_\alpha^{-1}[V]$, where V is open in X_α. But then $e[U] = e[f_\alpha^{-1}[V]] = \pi_\alpha^{-1}[V] \cap e[X]$ is an open set in $e[X]$. $\bullet$

Note: $e[U]$ might not be open in $\prod X_\alpha$, but that is irrelevant. See the earlier example where $e(x) = (x^2, e^x)$.

The converse of Theorem 4.4 is also true: if e is an embedding, then the f_α's separate points and X has the weak topology generated by the f_α's. However, we do not need this fact and will omit the proof (which is not very hard).

Example 4.5

Let (X, d) be a separable metric space. We can assume, without loss of generality, that $d \leq 1$. X is second countable so there is a countable base $\mathcal{B} = \{U_1, ..., U_n, ...\}$ for the open sets. For each n, let $f_n(x) = d(x, X - U_n)$. Then $f_n : X \to [0,1]$ is continuous and, since $X - U_n$ is closed, we have $f_n(x) > 0$ iff $x \in U_n$. If $x \neq y \in X$, there is an n such that $x \in U_n$ and $y \notin U_n$. Then $f_n(y) = 0 \neq f_n(x)$, so f_n's separate points.

We claim that the topology $\mathcal{T}_d$ on X is the same as the weak topology $\mathcal{T}_w$ generated by the f_n's. Because the functions f_n are continuous if X has the topology $\mathcal{T}_d$, we know that $\mathcal{T}_d \supseteq \mathcal{T}_w$. To show $\mathcal{T}_d \subseteq \mathcal{T}_w$, suppose $x \in U \in \mathcal{T}_d$. For some n, we have $x \in U_n \subseteq U$ and therefore $f_n(x) = c > 0$. But $V_n = f_n^{-1}[(\frac{c}{2}, 1]]$ is a (subbasic) open set in the weak topology and $x \in V_n \subseteq U_n \subseteq U$. Therefore $U \in \mathcal{T}_w$.

By Theorem 4.4, $e : X \to [0,1]^{\aleph_0}$ is an embedding, so $X \simeq e[\,X\,] \subseteq [0,1]^{\aleph_0}$. We sometimes write this as $X \overset{\text{top}}{\subseteq} [0,1]^{\aleph_0}$. From Theorems 3.3 and 3.5, we know that $[0,1]^{\aleph_0}$ is itself a separable metrizable space – and therefore all its subspaces are separable and metrizable. Putting all this together, we get that topologically, separable metrizable spaces are nothing more and nothing less than the subspaces of $[0,1]^{\aleph_0} = H$ ("the Hilbert cube").

We can view this fact with a "half-full" or "half-empty" attitude:

> i) separable metric spaces must not be very complicated since topologically they are nothing more than the subspaces of a single very nice space: the "cube" $[0,1]^{\aleph_0}$
>
> ii) separable metric spaces can get quite complicated, so the subspaces of a cube $[0,1]^{\aleph_0}$ are more complicated than we imagined.

Since $[0,1]^{\aleph_0} \simeq \prod_{n=1}^{\infty}[0, \frac{1}{n}] = H$ (the "Hilbert cube") $\subseteq \ell_2$, we can also say that topologically *the separable metrizable spaces X are* <u>*precisely*</u> *the subspaces of* ℓ_2. This is particularly amusing because of the metric d on ℓ_2 is very much like the usual metric on $\mathbb{R}^n$:

$$d(x,y) = (\sum_{n=1}^{\infty}(x_n - y_n)^2\,)^{\frac{1}{2}}$$

In some sense, this elegant "Euclidean-like" metric is adequate to describe the topology of any separable metric space. (*Note: X is topologically a subspace of H with the product topology. If we identify H with a subspace of ℓ_2, as above, how to we know that the metric topology induced on H from ℓ_2 is the same as the product topology on H?*)

We can summarize by saying that each of H and ℓ_2 is a "universal separable metric space." Notice, though, that these two "universal" spaces are not homeomorphic: one is compact and the other is not.)

Example 4.6

Suppose (X,d) is a metric space (not necessarily separable) and that $\{U_\alpha : \alpha \in A\}$ is a base for the topology $\mathcal{T}_d$, where $|A| = m$. Then an argument exactly like the one in Example 4.5 (just replace "n" everywhere with "α") shows that $X \overset{\text{top}}{\subseteq} [0,1]^m$. Therefore *every metric space, topologically, is a subspace of some sufficiently large "cube."* Of course when $m > \aleph_0$, the cube $[0,1]^m$ is <u>not</u> itself metrizable (*why?*); in general this cube will have many subspaces that are nonmetrizable. So the result is not quite as dramatic as in the separable case.

The <u>weight</u> $w(X)$ of a topological space $(X,\mathcal{T})$ is defined as min $\{|\mathcal{B}| : \mathcal{B}$ is a base for $\mathcal{T}\} + \aleph_0$.

> *We are assuming here that the "min" in the definition exists: see Example 5.22 in Chapter VIII. For some very simple spaces, the "min" could be finite – in which case the "$+\aleph_0$" guarantees that $w(X) \geq \aleph_0$ (convenient for purely technical reasons that don't matter in these notes).*

The <u>density</u> $\delta(X)$ is defined as $\min\{|D| : D$ is dense in $(X,\mathcal{T})\} + \aleph_0$. For a <u>metrizable</u> space, it is not hard to prove that $w(X) = \delta(X)$. The proof is just like our earlier proof (in Theorem III.6.5) that separability and second countability are equivalent in metrizable spaces. Therefore we have that for any metric space (X,d),

$$X \overset{\text{top}}{\subseteq} [0,1]^{w(X)} = [0,1]^{\delta(X)} \qquad (*)$$

Notice that, for a given space (X,d), the exponents in this statement are not necessarily the smallest possible. For example, (*) says that $\mathbb{R} \overset{\text{top}}{\subseteq} [0,1]^{w(\mathbb{R})} = [0,1]^{\aleph_0}$, but in fact we can do much better than the exponent $\aleph_0$: $\mathbb{R} \simeq (0,1) \subseteq [0,1] = [0,1]^1$!

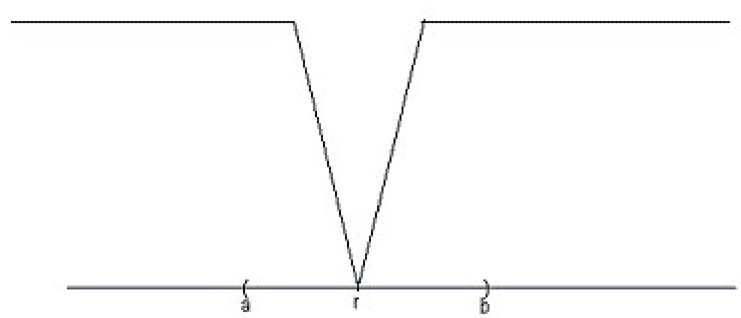
We add one additional comment, without proof: For a given infinite cardinal m, it is possible to define a metrizable space H_m with weight m such that every metric space X with weight m can be embedded in $H_m^{\aleph_0}$. In other words, $H_m^{\aleph_0}$ is a metrizable space which is "universal for all metric spaces of weight m." The price of metrizability, here, is that we need to replace $[0,1]$ by a more complicated space H_m.

> *Without going into all the details, you can think of H_m as a "star" with m different copies of $[0,1]$ ("rays") all placed with 0 at the center of the star. For two points x, y on the same "ray" of the star, $d(x,y) = |x - y|$; if x, y are on different rays, the distance between them is measured "via the origin" : $d(x,y) = |x| + |y|$.*

The condition in Theorem 4.4 that "X has the weak topology generated by a collection of maps $f_\alpha : X \to X_\alpha$" is not always easy to check. The following definition and theorem can sometimes help.

Definition 4.7 Suppose X and X_α $(\alpha \in A)$ are topological spaces and that $f_\alpha : X \to X_\alpha$. We say that the collection $\mathcal{F} = \{f_\alpha : \alpha \in A\}$ separates points from closed sets if whenever F is a closed set in X and $x \notin F$, there is an α such that $f_\alpha(x) \notin \text{cl}\, f_\alpha[F]$.

Example 4.8 Let $X = \mathbb{R}$ and $\mathcal{F} = C(\mathbb{R})$. Suppose F is a closed set in $\mathbb{R}$ and $r \notin F$. There is an open interval (a,b) for which $r \in (a,b) \subseteq \mathbb{R} - F$. Define $f \in C(\mathbb{R})$ with a graph like the one shown in the figure:



Then $0 = f(r) \notin \text{cl}\, f[F]$. Therefore $C(\mathbb{R})$ separates points and closed sets.

The same notation continues in the following lemma.

Lemma 4.9 The family $\mathcal{B} = \{f_\alpha^{-1}[V] :\ \alpha \in A,\ V \text{ open in } X_\alpha\}$ is a base for the topology in X iff

$$\begin{cases} \text{i) the } f_\alpha\text{'s are continuous, and} \\ \text{ii) } \{f_\alpha : \alpha \in A\} \text{ separates points and closed sets.} \end{cases}$$

In particular, i) + ii) imply that $\mathcal{B}$ is a subbase for the topology on X – so that X has the weak topology generated by the f_α's.

Note: the more open (closed) sets there are in X, the harder it is for a given family $\{f_\alpha : \alpha \in A\}$ to succeed in separating points and closed sets. In fact, the lemma shows that a family of continuous functions $\{f_\alpha : \alpha \in A\}$ succeeds in separating points and closed sets <u>only</u> if $\mathcal{T}$ is the <u>smallest</u> topology that makes the f_α's continuous.

Proof Suppose $\mathcal{B}$ is a base. Then the sets $f_\alpha^{-1}[V]$ in $\mathcal{B}$ are open, so the f_α's are continuous and i) holds.

To prove ii), suppose F is closed in X and $x \notin F$. For some α and some V open in X_α, we have $x \in f_\alpha^{-1}[V] \subseteq X - F$. Then $f_\alpha(x) \in V$ and $V \cap f_\alpha[F] = \emptyset$ (since $f_\alpha^{-1}[V] \cap F = \emptyset$). Therefore $f_\alpha(x) \notin \text{cl}\, f_\alpha[F]$ so ii) also holds.

Conversely, suppose i) and ii) hold. If $x \in O$ and O is open in X, we need to find a set $f_\alpha^{-1}[V] \in \mathcal{B}$ such that $x \in f_\alpha^{-1}[V] \subseteq O$. Since $x \notin F = X - O$, condition ii) gives us an α for which $f_\alpha(x) \notin \text{cl}\, f_\alpha[F]$. Then $f_\alpha(x) \in V = X_\alpha - \text{cl}\, f_\alpha[F]$, so $x \in f_\alpha^{-1}[V]$ and we claim $f_\alpha^{-1}[V] \subseteq O$:

If $w \notin O$, then $w \in F$, so $f_\alpha(w) \in f_\alpha[F] \subseteq \text{cl}\, f_\alpha[F]$.
Then $f_\alpha(w) \notin V$, so $w \notin f_\alpha^{-1}[V]$. •

Theorem 4.10 Suppose $f_\alpha : X \to X_\alpha$ is continuous for each $\alpha \in A$. If the collection $\{f_\alpha : \alpha \in A\}$

$$\begin{cases} \text{i) separates points and closed sets, and} \\ \text{ii) separates points} \end{cases}$$

then the evaluation map $e : X \to \prod X_\alpha$ is an embedding.

Proof Since the f_α's are continuous, Lemma 4.9 gives us that X has the weak topology. Then Theorem 4.4 implies that e is an embedding. •

If the space X (that we are trying to embed in a product) is a T_1-space (*as is most often the case*), then i) $\Rightarrow$ ii) in Theorem 4.10 , so we have the simpler statement given in the following corollary.

Corollary 4.11 Suppose $f_\alpha : X \to X_\alpha$ is continuous for each $\alpha \in A$. If X is a T_1-space and $\{f_\alpha : \alpha \in A\}$ separates points and closed sets, then evaluation map $e : X \to \prod X_\alpha$ is an embedding.

Proof By Theorem 4.10, it is sufficient to show that the f_α's separate points, so suppose $x \neq y \in X$. Since x is <u>not</u> in the closed set $\{y\}$, there is an α for which $f_\alpha(x) \notin \text{cl}\, f_\alpha[\{y\}]$. Therefore $f_\alpha(x) \neq f_\alpha(y)$, so $\{f_\alpha : \alpha \in A\}$ separates points. •

Exercises

E14. A space $(X, \mathcal{T})$ is called a T_0 space if whenever $x \neq y \in X$, then $\mathcal{N}_x \neq \mathcal{N}_y$ (equivalently, either there is an open set U containing x but not y, or vice-versa). Notice that the T_0 condition is weaker than T_1 (*see example III.2.6.4).* Clearly, a subspace of a T_0-space is T_0.

a) Prove that a nonempty product $X = \prod\{X_\alpha\colon \alpha \in A\}$ is T_0 iff each X_α is T_0.

b) Let S be "Sierpinski space" – that is, $S = \{0, 1\}$ with the topology $\mathcal{T} = \{\emptyset, \{1\}, \{0, 1\}\}$. Use the embedding theorems to prove that a space X is T_0 iff X is homeomorphic to a subspace of S^m for some cardinal m. *Hint $\Rightarrow$: for each open set U in X, let χ_U be the characteristic function of U. Use an embedding theorem. Nearly all interesting spaces are T_0, and those spaces, topologically, are all just subspaces of S^m for some m.*

E15. a) Let (X, d) be a metric space. Prove that $C(X) = \{f \in \mathbb{R}^X : f \text{ is continuous}\}$ separates points and closed sets. (*Since X is T_1, $C(X)$ therefore also separates points*).

b) Suppose X is any T_0 topological space for which $C(X)$ separates points and closed sets. Prove that X can be embedded in a product of copies of $\mathbb{R}$.

E16. A space X satisfies the countable chain condition (CCC) if every collection of nonempty pairwise disjoint open sets must be countable. (*For example, every separable space satisfies CCC.*)

Suppose that X_α is separable for each $\alpha \in A$. Prove that $X = \prod\{X_\alpha\colon \alpha \in A\}$ satisfies CCC. (*There isn't much to prove when $|A| \leq c$; why?*)

Hint: Let $\{U_t : t \in T\}$ be any such collection. We can assume all the U_t's are basic open sets (why?). Prove that if $S \subseteq T$ and $|S| \leq c$, then $|S| \leq \aleph_0$ and hence T must be countable.)

E17. There are several ways to define dimension for topological spaces. One classical method is the following inductive definition.

Define dim $\emptyset = -1$.

For $p \in X$, we say that X has dimension ≤ 0 at p if there is a neighborhood base at p consisting of sets with -1 dimensional (that is, empty) frontiers. Since a set has empty frontier iff it is clopen, X has dimension ≤ 0 at p iff p has a neighborhood base consisting of clopen sets.

We say X has dimension $\leq n$ at p if there exists a neighborhood base at p in which the frontier of every basic neighborhood has dimension $\leq n-1$.

*We say X has dimension $\leq n$, and write dim$(X) \leq n$, if X has dimension $\leq n$ at p for each $p \in X$ and that **dim$(X) = n$** if dim$(X) \leq n$ but dim$(X) \nleq n-1$. We say **dim$(X) = \infty$** if dim$(X) \leq n$ is false for every $n \in \mathbb{N}$.*

*While this definition of dim(X) makes sense for any topological space X, it turns out that "dim" produces a nicely behaved dimension theory only for separable metric spaces. The dimension function "dim" is sometimes called **small inductive dimension** to distinguish it from other more general definitions of dimension. The classic discussion of small inductive dimension is in Dimension Theory (Hurewicz and Wallman).*

It is clear that "dim$(X) = n$" is a topological property. There is a theorem stating that dim ($\mathbb{R}^n$) $= n$, from which it follows that $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$ if $m \neq n$. In proving the theorem, showing dim $(\mathbb{R}^n) \leq n$ is easy; the hard part is showing that dim $(\mathbb{R}^n) \not\leq n - 1$.

a) Prove that dim $(\mathbb{R}) = 1$.

b) Let C be the Cantor set. Prove that dim$(C) = 0$.

c) Suppose (X, d) is a 0-dimensional separable metric space. Prove that X is homeomorphic to a subspace of C. (*Hint: Show that X has a countable base of clopen sets. View C as $\{0, 2\}^{\aleph_0}$ and apply the embedding theorems.*)

E18. Suppose X and Y are T_0-spaces. Part a) outlines a sufficient condition that Y can be embedded in a product of X's, i.e., that $Y \overset{\text{top}}{\subseteq} X^m$ for some cardinal m. Parts b) and c) look at some corollaries.

a) **Theorem** Let X and Y be T_0 spaces. Then Y can be topologically embedded in X^m for some cardinal m if for every closed set $F \subseteq Y$ and every point $y \notin F$, there exists a $n \in \mathbb{N}$ and a continuous function $f : Y \to X^n$ such that $f(y) \notin \text{cl}\, f[F]$.

Proof Let $T = \{t : t = (y, F),\ F$ is closed in Y and $y \notin F\}$. For each such pair $t = (y, F)$, let f_t be the function given in the hypothesis. Let $X_t = X^n$ (the space containing the range of f_t). Then clearly $\prod\{X_t : t \in T\}$ is homeomorphic to X^m for some m, so it suffices to show Y can be embedded in $\prod\{X_t : t \in T\}$. Let $h : Y \to \prod\{X_t : t \in T\}$ be defined as follows: for $y \in Y$, $h(t)$ has for its t-th coordinate $f_t(y)$, i.e., $h(y)(t) = f_t(y)$.

i) Show h is continuous.

ii) Show h is one-to-one.

iii) Show that h is a closed mapping onto its range $h[Y]$ to complete the proof that h is a homeomorphism between Y and $h[Y] \subseteq \prod\{X_t : t \in T\}$. (*Note: the converse of the theorem is also true. Both the theorem and converse are due to S. Mrowka.*)

b) Let F denote Sierpinski space $\{\{0,1\}, \{\emptyset,\{0\},\{0,1\}\}$. Use the theorem to show that every T_0 space Y can be embedded in F^m for some m.

c) Let D denote the discrete space $\{0,1\}$. Use the theorem to show that every T_1-space Y satisfying dim$(Y) = 0$ (see Exercise E14) can be embedded in D^m for some m.
Parts b) and c) are due to Alexandroff.

Mrowka also proved that there is no T_1-space X such that every T_1-space Y can be embedded in X^m for some m.

E19. Let $X = \{0, 1\}$ with the discrete topology, and let m be an infinite cardinal.

a) Show that X^m contains a discrete subspace of cardinality m.
b) Show that $w(X^m) = m$ (*see Example 4.6*).

E20. X is called <u>totally disconnected</u> if every connected subset A satisfies $|A| \leq 1$. Prove that a totally disconnected compact Hausdorff space is homeomorphic to a closed subspace of $\{0, 1\}^m$ for some m. (*Hint: see Lemma V.5.6*).

E21. Suppose X is a countable space that does not have a countable neighborhood base at the point $a \in X$. (*For instance, let* $a = (0, 0)$ *in the space* L*, Example III.9.8.*)

Let $A_j = \{x \in X^{\aleph_0} : x_i = a \text{ for } i > j\}$ and $A = \bigcup_{j=0}^{\infty} A_j \subseteq X^{\aleph_0}$. Prove that <u>no</u> point in the (countable) space A has a countable neighborhood base. (*Note: it is not necessary that* X *be countable. That condition simply forces* A *to be countable and makes the example more dramatic.*)

5. The Quotient Topology

Suppose that for each $\alpha \in A$ we have a map $g_\alpha : X_\alpha \to Y$, where X_α is a topological space and Y is a set. Certainly there is a topology for Y that will make all the g_α's continuous: for example, the trivial topology on Y. But what is the largest topology on Y that will do this? Let

$$\mathcal{T} = \{O \subseteq Y : \text{for all } \alpha \in A,\ g_\alpha^{-1}[O] \text{ is open in } X_\alpha\}$$

It is easy to check that $\mathcal{T}$ is a topology on Y. For each α, by definition, each set $g_\alpha^{-1}[O]$ is open in X_α, so $\mathcal{T}$ makes all the g_α's continuous. Moreover, if $B \subseteq Y$ and $B \notin \mathcal{T}$, then for at least one α, $g_\alpha^{-1}[B]$ is not open in X_α – so adding B to $\mathcal{T}$ would "destroy" the continuity of at least one g_α. Therefore $\mathcal{T}$ is the largest possible topology on Y making all the g_α's continuous.

Definition 5.1 Suppose $g_\alpha : X_\alpha \to Y$ for each $\alpha \in A$. The strong topology $\mathcal{T}$ on Y generated by the maps g_α is the largest topology on Y making all the maps g_α continuous, and $O \in \mathcal{T}$ iff $g_\alpha^{-1}[O]$ is open in X_α for every α.

The strong topology generated by a collection of maps $\{g_\alpha : \alpha \in A\}$ is "dual" to the weak topology in the sense that it involves essentially the same notation but with "all the arrows pointing in the opposite direction." For example, the following theorem states that a map f out of a space with the strong topology is continuous iff each map $f \circ g_\alpha$ is continuous; but a map f into a space with the weak topology generated by mappings f_α is continuous iff all the compositions $f_\alpha \circ f$ are continuous (*see Theorem 2.6*)

Theorem 5.2 Suppose Y has the strong topology generated by a collection of maps $\{g_\alpha : \alpha \in A\}$. If Z is a topological space and $f : Y \to Z$, then f is continuous if and only if $f \circ g_\alpha : X_\alpha \to Z$ is continuous for each $\alpha \in A$.

Proof For each $\alpha \in A$, we have $X_\alpha \xrightarrow{g_\alpha} Y \xrightarrow{f} Z$, and the g_α's are continuous since Y has the strong topology.

If f is continuous, so is each composition $f \circ g_\alpha$.

Conversely, suppose each $f \circ g_\alpha$ is continuous and that U is open in Z. We want to show that $f^{-1}[U]$ is open in Y. But Y has the strong topology, so $f^{-1}[U]$ is open in Y iff each $g_\alpha^{-1}[f^{-1}[U]]$ is open in X_α. But $g_\alpha^{-1}[f^{-1}[U]] = (f \circ g_\alpha)^{-1}[U]$ which is open because $f \circ g_\alpha$ is continuous. ●

We introduced the idea of the strong topology as a parallel to the definition of weak topology. However, we are going to use the strong topology only in a special case: when there is just one map $g_\alpha = g$ and g is onto.

Definition 5.3 Suppose $(X, \mathcal{T})$ is a topological space and $g : X \to Y$ is onto. The strong topology on Y generated by g is also called the quotient topology on Y. If Y has the quotient topology from g, we say that $g : X \to Y$ is a quotient mapping and we say Y is a quotient of X. We also say the Y is a quotient space of X and sometimes as $Y = X/g$.

From the discussion of strong topologies, we know that if X is a topological space and $g : X \to Y$, then the quotient topology on Y is $\mathcal{T} = \{U : g^{-1}[U] \text{ is open in } X\}$. Thus $U \in \mathcal{T}$ if and only if $g^{-1}[U]$ is open in X: notice in this description that

"only if"	guarantees that g is continuous and
"if"	guarantees that $\mathcal{T}$ is the largest topology on Y making g continuous.

Quotients of X are used to create new spaces Y by "pasting together" ("identifying") several points of X to become a single new point. Here are two intuitive examples:

i) Begin with $X = [0,1]$ and identify 0 with 1 (that is, "paste" 0 and 1 together to become a single point). The result is a circle, S^1. This identification is exactly what the map $g : [0,1] \to S^1$ given by $g(x) = (\cos 2\pi x, \sin 2\pi x)$ accomplishes. It turns out (*see below*) that the usual topology on S^1 is the same as quotient topology generated by the map g. Therefore we can say that g is a quotient map and that S^1 is a quotient of $[0,1]$

ii) If we take the space $X = S^1$ and use a mapping g to "identify" the north and south poles together, the result is a "figure-eight" space Y. The usual topology on Y (from $\mathbb{R}^2$) turns out to be the same as quotient topology generated by g (*see below*). Therefore we can say that g is a quotient mapping and the "figure-eight" is a quotient of S^1.

Suppose we are given an onto map $g : (X, \mathcal{T}) \to (Y, \mathcal{T}')$. How can we tell whether g is a quotient map – that is, how can we tell whether $\mathcal{T}'$ is the quotient topology? By definition, we must check that $U \in \mathcal{T}'$ iff $g^{-1}[U] \in \mathcal{T}$. Sometimes it is fairly straightforward to do this. But the following theorem will sometimes make things much easier.

Theorem 5.4 Suppose $g : (X, \mathcal{T}) \to (Y, \mathcal{T}')$ is continuous and onto. If g is open (or closed), then $\mathcal{T}'$ is the quotient topology, so g is a quotient map. In particular, if X is compact and Y is Hausdorff, g is a quotient mapping.

Note: Whether $g : X \to Y$ is continuous depends, of course, on the topology $\mathcal{T}'$, but if $\mathcal{T}'$ makes g continuous, then so would any smaller topology on Y. The theorem tells us that if g is both continuous and open (or closed), then $\mathcal{T}'$ is completely determined by g: it is the largest topology that makes g continuous.

Proof Suppose $U \subseteq Y$. We must show $U \in \mathcal{T}'$ iff $g^{-1}[U] \in \mathcal{T}$. If $U \in \mathcal{T}'$, then $g^{-1}[U] \in \mathcal{T}$ because g is continuous. On the other hand, suppose $g^{-1}[U] \in \mathcal{T}$. Since g is onto and open, $g[g^{-1}[U]] = U \in \mathcal{T}'$.

For $F \subseteq Y$, $g^{-1}[Y - F] = X - g^{-1}[F]$. It follows easily that F is closed in the quotient space if and only if $g^{-1}[F]$ is closed in X. With that observation, the proof that a continuous, closed, onto map g is a quotient map is exactly parallel to the case when g is open.

If X is compact and Y is T_2, then g must be closed, so g is a quotient mapping. •

Note: Theorem 5.4 implies that the map $g(x) = (\cos 2\pi x, \sin 2\pi x)$ from $[0,1]$ to S^1 is a quotient map, but g is not open. The same formula g defines a quotient map $g : \mathbb{R} \to S^1$ which is not closed (why?). Exercise E25 gives examples of quotient maps g that are neither open nor closed.

Suppose $\sim$ is an equivalence relation on a set X. For each $x \in X$, the equivalence class of x is $[x] = \{z \in X : z \sim x\}$. The equivalence classes partition X into a collection of nonempty pairwise disjoint sets. Conversely, it is easy to see that any partition of X is the collection of equivalence classes for some equivalence relation – namely, $x \sim z$ iff x and z are in the same set of the partition.

The set of equivalence classes, $Y = \{[x] : x \in X\}$, is sometimes written as $X/\sim$. There is a natural onto map $g : X \to X/\sim\; = Y$ given by $g(x) = [x]$. We can think of the elements of Y as "new points" which are created by "identifying together as one" all the members of each equivalence class in X. Conversely, whenever $g : X \to Y$ is any onto mapping, we can think of Y as the set of equivalence classes for some equivalence relation on X – namely $x \sim y \;\Leftrightarrow\; y \in g^{-1}(x) \;\Leftrightarrow\; g(y) = g(x)$. If X is a topological space, we can give the set of equivalence classes Y the quotient topology.

Example 5.5 For $a,\ b \in \mathbb{Z}$, define $a \sim b$ iff $b - a$ is even. There are two equivalence classes $[0] = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$ and $[1] = \{\ldots -3, -1, 1, 3, \ldots\}$ so $\mathbb{Z}/\sim\; = \{[0],\ [1]\}$.

Define $g : \mathbb{Z} \to \mathbb{Z}/\sim$ by $g(a) = [a]$ and give $\mathbb{Z}/\sim$ the quotient topology. A set U is open in $\mathbb{Z}/\sim$ iff $g^{-1}[U]$ is open in $\mathbb{Z}$. But that is true for every $U \subseteq \mathbb{Z}/\sim$ because $\mathbb{Z}$ is discrete. Therefore the quotient $\mathbb{Z}/\sim$ is a two point discrete space.

Example 5.6 Let (X, d) be a pseudometric space. Define an equivalence relation $\sim$ in X by $x \sim z$ iff $d(x, z) = 0$. Let $Y = X/\sim$ and define $g : X \to Y$ by $g(x) = [x]$. Give Y the quotient topology. Then points at distance 0 in X have been "identified with each other" to become one point (an equivalence class) in Y.

For $[x], [z] \in Y$, define $d'([x], [z]) = d(x, z)$. In order to see that d' is well-defined, we need to check that the definition is independent of the representatives chosen from the equivalence classes:

If $[x'] = [x]$ and $[z'] = [z]$, then $d(x, x') = 0$ and $d(z, z') = 0$. Therefore
$d(x, z) \leq d(x, x') + d(x', z') + d(z', z) = d(x', z')$, and similarly
$d(x', z') \leq d(x, z)$. Thus $d(x', z') = d(x, z)$ so $d'([x'], [z']) = d'([x], [z])$.

It is easy to check that d' is a pseudometric on Y. In fact, d' is a metric: if $d'([x], [z]) = 0$, then $d(x, z) = 0$, which means that $x \sim z$ and $[x] = [z]$.

We now have two definitions for topologies on Y: the quotient topology $\mathcal{T}$ and the metric topology $\mathcal{T}_{d'}$. We claim that $\mathcal{T} = \mathcal{T}_{d'}$. To see this, first notice that

$$[y] \in B_\epsilon^{d'}([x]) \text{ iff } d'([y], [x]) < \epsilon \text{ iff } d(y, x) < \epsilon \text{ iff } y \in B_\epsilon^{d}(x)$$

Therefore $g^{-1}[B_\epsilon^{d'}([x])] = B_\epsilon^{d}(x)$ and $g\,[B_\epsilon^{d}(x)] = B_\epsilon^{d'}([x])$. But then we have

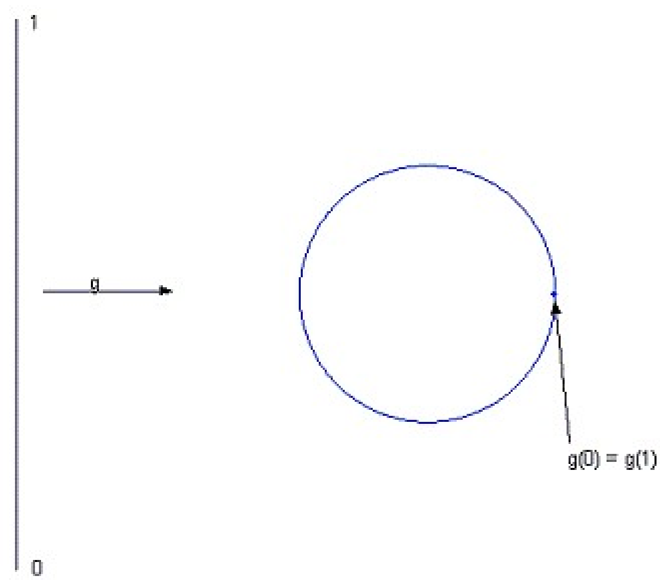
$$\begin{aligned} U \in \mathcal{T}_{d'} \quad & \text{iff } U \text{ is a union of } d'\text{-balls} \\ & \text{iff } g^{-1}[U] \text{ is a union of } d\text{-balls} \\ & \text{iff } g^{-1}[U] \text{ is open in } X \\ & \text{iff } U \in \mathcal{T}. \end{aligned}$$

The metric space (Y, d') is called the metric identification of the pseudometric space (X, d). In effect, we turn the pseudometric space into a metric space by agreeing that points in X at distance 0 are "lumped together" into a single point.

Note: In this particular example, it is easy to verify that the quotient mapping $g : X \to Y$ is open, so g would be a homeomorphism if only g were one-to-one. If the original pseudometric d is actually a metric, then g is one-to-one and a homeomorphism: the metric identification of a metric space (X, d) is itself.

Example 5.7 What does it mean if we say "identify together the endpoints of $[0, 1]$ and get a circle"? Of course, one could simply take this to be the definition of a (topological) "circle." Or, it could mean that we already know what a circle is and are claiming that a certain quotient space is homeomorphic to a circle. We take the latter point of view.

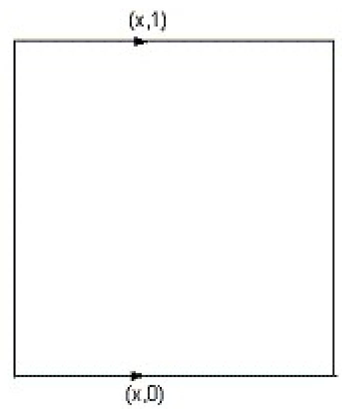
Define $g : [0, 1] \to S^1$ by $g(x) = (\cos 2\pi x, \sin 2\pi x)$. This map is onto would be one-to-one except that $g(0) = g(1)$, so g corresponds to the equivalence relation on X for which $0 \sim 1$ (and there are no other equivalences except that $x \sim x$ for every x). We can think of the equivalence classes $X/\sim$ as corresponding in a natural way to the points of S^1.



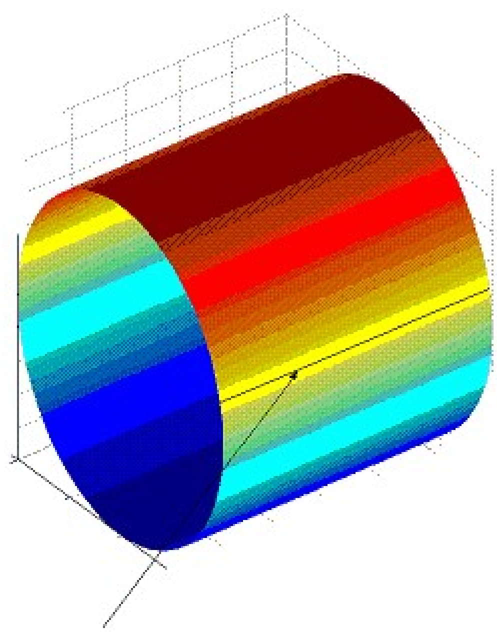
Here S^1 has its usual topology and g is continuous. Since X is compact and S^1 is Hausdorff, Theorem 5.4 gives that the usual topology on S^1 is the quotient topology and g is a quotient map.

When it "seems apparent" that the result of making certain identifications produces some familiar space Y, we need to check that the familiar topology on Y is actually the quotient topology. Example 5.7 is reassuring: if we believed, intuitively, that the result of identifying the endpoints of $[0, 1]$ should be S^1 but then found that the quotient topology on the set $X/\sim$ differed from the usual topology, we would be inclined to think that we had made the "wrong" definition for a quotient.

Example 5.8 Suppose we take a square $[0,1]^2$ and identify points on the top and bottom edges using the equivalence relation $(x,0) \sim (x,1)$. We can schematically picture this identification as



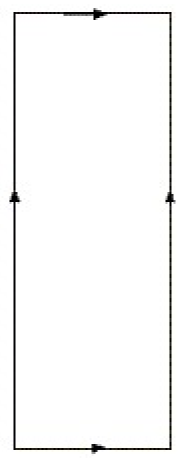
The arrows indicate that the edges are to be identified as we move along the top and bottom edges in the same direction. We have an obvious map g from $[0,1]^2$ to a cylinder in $\mathbb{R}^3$ which identifies points in just this way, and we can think of the equivalence classes as corresponding in a natural way to the points of the cylinder.



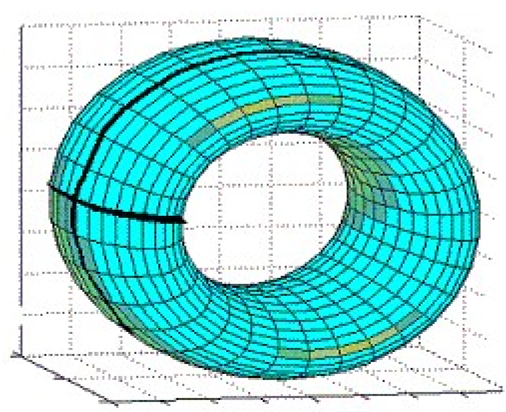
g identifies (x,0) with (x,1)

The cylinder has its usual topology from $\mathbb{R}^3$ and the map g is (clearly) continuous and onto. Again, Theorem 5.4 gives that the usual topology on the cylinder is, in fact, the quotient topology.

Example 5.9 Similarly, we can show that a torus (the "surface of a doughnut") is the result of the following identifications in $[0,1]^2$: $(x,0) \sim (x,1)$ and $(0,y) \sim (1,y)$



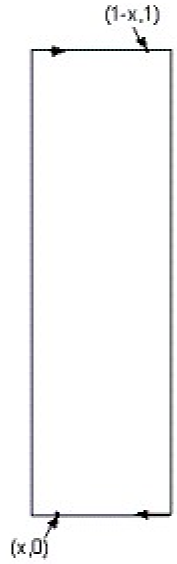
Thinking in two steps, we see that the identification of the two vertical edges produces a cylinder; the circular ends of the cylinder are then identified (in the same direction) to produce the torus.



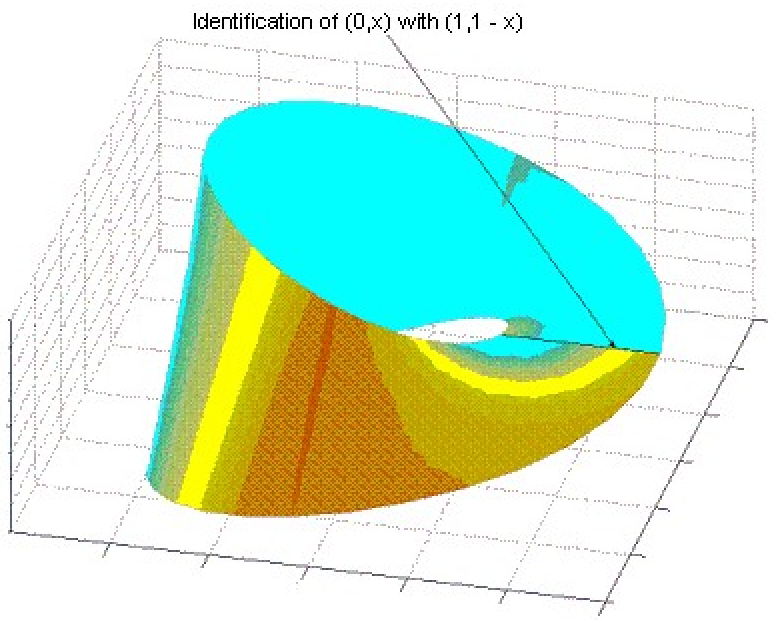
The two circles darkly shaded on the surface represent the identified edges.

We can identify the equivalence classes naturally with the points of this torus in $\mathbb{R}^3$ and just as before we see that the usual topology on the torus is in fact the quotient topology.

Example 5.10 Define an equivalence relation $\sim$ in $[0,1]^2$ by setting $(x,0) \sim (1-x,1)$. Intuitively, the idea is to identify the points on the top and bottom edges with each other as we move along the edges in opposite directions. We can picture this schematically as

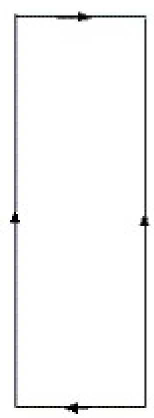


Physically, we can think of a strip of paper and glue the top and bottom edges together after making a "half-twist." The quotient space $X/\sim$ is called a Möbius strip.

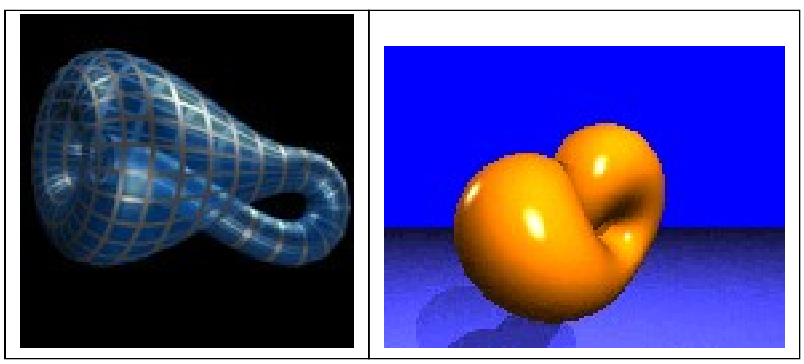


We can take the quotient $X/\sim$ as the definition of a Möbius strip, or we can consider a "real" Möbius strip M in $\mathbb{R}^3$ and define a map $g : [0,1]^2 \to M$ that accomplishes the identification we have in mind. In that case there is a natural way to associate the equivalence classes to the points of the torus in $\mathbb{R}^3$ and again Theorem 5.4 guarantees that the usual topology on the Möbius strip is the quotient topology.

Example 5.11 If we identify the vertical edges of $[0,1]^2$ (to get a cylinder) and then identify its circular ends with a half-twist (reversing orientation): $(0,y) \sim (1,y)$ and $(x,0) \sim (1-x,1)$. We get a quotient space which is called a Klein bottle.



It turns out that a Klein bottle cannot be embedded in $\mathbb{R}^3$ – the physical construction would require a "self-intersection" (that is, additional points identified) which is not allowed. A pseudo-picture looks like



In these pictures, the thin "neck" of the bottle actually intersects the main body in order to re-emerge "from the inside" – in a "real" Klein bottle (in $\mathbb{R}^4$, say), the self-intersection would not happen.

In fact, you can imagine the Klein bottle as a subset of $\mathbb{R}^4$ using color as a 4^{th} dimension. To each point (x,y,z) on the "bottle" pictured above, add a 4th coordinate to get (x,y,z,r). Now color the points on the bottle in varying shades of red and let r be a number measuring the "intensity of red coloration at a point." Do the coloring in such a way that the surface "blushes" as it intersects itself – so that the points of "intersection" seen above in $\mathbb{R}^3$ will be different (in their 4^{th} coordinates).

Alternately, you can think of the Klein bottle as a parametrized surface traced out by a moving point $P = (x,y,z,t)$ where x, y, z depend on time t and t is recorded as a 4^{th}-coordinate. A point on the surface then has coordinates of form (x, y, z, t). At "a point" where we see a self-intersection in $\mathbb{R}^3$, there are really two different points (with different time coordinates t).

Example 5.12

1) In S^1, identify antipodal points – that is, in vector notation, $P \sim -P$ for each $P \in S^1$. Convince yourself that the quotient $S^1/\sim$ is S^1.

2) Let D^2 be the unit disk $\{P \in \mathbb{R}^2 : |P| \leq 1\}$. Identify antipodal points on the boundary of D^2: that is, $P \sim -P$ if $P \in S^1 \subseteq D^2$. The quotient $D/\sim$ is called the projective plane, a space which, like the Klein bottle, cannot be embedded in $\mathbb{R}^3$.

3) For any space X, we can form the product $X \times [0,1]$ and let $(x,1) \sim (y,1)$ for all $x, y \in X$. The quotient $(X \times [0,1])/\sim$ is called the cone over X. (*Why?*)

4) For any space X, we can form the product $X \times [-1,1]$ and define $(x,1) \sim (y,1)$ and $(x,-1) \sim (y,-1)$ for all $x, y \in X$. The quotient $(X \times [-1,1])/\sim$ is called the suspension of X. (*Why?*)

There is one other very simple construction for combining topological spaces. It is often used in conjunction with quotients.

Definition 5.13 For each $\alpha \in A$, let $(X_\alpha, \mathcal{T}_\alpha)$ be a topological space, and assume that the sets X_α are pairwise disjoint. The topological sum (or "free sum") of the X_α's is the space $(\bigcup_{\alpha \in A} X_\alpha, \mathcal{T})$ where $\mathcal{T} = \{O \subseteq \bigcup_{\alpha \in A} X_\alpha : O \cap X_\alpha \text{ is open in } X_\alpha \text{ for every } \alpha \in A\}$. We denote the topological sum by $\sum_{\alpha \in A} X_\alpha$. In the case $|A| = 2$, we use the simpler notation $X_1 + X_2$.

In $\sum_{\alpha \in A} X_\alpha$, each X_α is a clopen subspace. Any set open (or closed) in X_α is open (or closed) in the sum. The topological sum $\sum_{\alpha \in A} X_\alpha$ can be pictured as a union of the disjoint pieces X_α, all "far apart" from each other – so that there is no topological "interaction" between the pieces.

Example 5.14 In $\mathbb{R}^2$, let A_1 and A_2 be open disks with radius 1 and centers at $(0,0)$ and $(3,0)$. Then topological sum $A_1 + A_2$ is the same as $A_1 \cup A_2$ with subspace topology. By contrast, let B_1 be an open disk with radius 1 centered at $(0,0)$ and let B_2 be a closed disk with radius 1 centered at $(2,0)$. Then $B_1 + B_2$ is not the same as $B_1 \cup B_2$ with the subspace topology (*why?*). Are the topologies on $C_1 + C_2$ and $C_1 \cup C_2$ the same if C_1 and C_2 are separated subsets of $\mathbb{R}^2$?

Exercise 5.15 Usually it is very easy to see whether properties of the X_α's do or do not carry over to $\sum_{\alpha \in A} X_\alpha$. For example, you should convince yourself that:

1) If the X_α's are nonempty and separable, then $\sum_{\alpha \in A} X_\alpha$ is separable iff $|A| \leq \aleph_0$.

2) If the X_α's are nonempty and second countable, then $\sum_{\alpha \in A} X_\alpha$ is second countable iff $|A| \leq \aleph_0$.

3) A function $f: \sum_{\alpha \in A} X_\alpha \to Z$ is continuous iff each $f|X_\alpha$ is continuous.

4) If $d_\alpha \leq 1$ is a metric for X_α, then the topology on $\sum_{\alpha \in A} X_\alpha$ is the same as $\mathcal{T}_d$ where

$$d(x, y) = \begin{cases} d_\alpha(x, y) & \text{if } x, y \in X_\alpha \\ 1 & \text{otherwise} \end{cases}$$

so $\sum_{\alpha \in A} X_\alpha$ is metrizable if all the X_α's are metrizable. You should be able to find similar statements for other topological properties such as first countable, second countable, Lindelöf, compact, connected, path connected, completely metrizable, ...

Definition 5.16 Suppose X and Y are disjoint topological spaces and $f: A \to Y$, where $A \subseteq X$. In the sum $X + Y$, define $x \sim y$ iff $y = f(x)$. If we form $(X + Y)/\sim$, we say that we have attached X to Y with f and write this space as $X +_f Y$.

For each $p \in f[A]$, its equivalence class under $\sim$ is $\{p\} \cup f^{-1}[\{p\}]$. You may think of the function "attaching" the two spaces by repeatedly selecting a group of points in X, identifying them together, and "sewing" them all onto a single point in Y – just as you might run a needle and thread through several points in the fabric X and then through a point in Y and pull everything tight.

Example 5.17

1) Consider disjoint cylinders X and Y. Let A be the circle forming one end of X and B the circle forming one end of Y. Let $f: A \to f[A] = B \subseteq Y$ be a homeomorphism. Then f "sews together" X and Y by identifying these two circles. The result is a new cylinder.

2) Consider a sphere $S^2 \subseteq \mathbb{R}^3$. Excise from the surface S^2 two disjoint open disks D_1 and D_2 and let $A_1 \cup A_2$ be union of the two circles that bounded those disks. Let Y be a cylinder whose ends are bounded by the union of two circles, $B_1 \cup B_2$. Let f be a homeomorphism carrying the points of A_1 and A_1 clockwise onto the points of B_1 and B_2 respectively..

Then $S^2 +_f Y$ is a "sphere with a handle."

3) Consider a sphere $S^2 \subseteq \mathbb{R}^3$. Excise from the surface S^2 an open disk and let A be the circular boundary of the hole in the surface S^2. Let M be a Möbius strip and let B be the curve that bounds it. Of course, $B \simeq S^1$. Let $f: A \to B$ be a homeomorphism. The we can use f to join the spaces by "sewing" the edge of the Möbius strip to the edge of the hole in S^2. The result is a "sphere with a crosscap."

There is a very nice theorem, which we will not prove here, which uses all these ideas. It is a "classification" theorem for certain surfaces.

Definition 5.18 A Hausdorff space X is a 2-manifold if each $x \in X$ has an open neighborhood U that is homeomorphic to $\mathbb{R}^2$. Thus, a 2-manifold looks "locally" just like the Euclidean plane. A surface is a Hausdorff 2-manifold.

Theorem 5.19 Let X be a compact, connected surface. The X is homeomorphic to a sphere S^2 or to S^2 with a finite number of handles and crosscaps attached.

You can read more about this theorem and its proof in *Algebraic Topology: An Introduction* (William Massey).

Exercises

E22. a) Let $\sim$ be the equivalence relation on $\mathbb{R}^2$ given by $(x_1, y_1) \sim (x_2, y_2)$ iff $y_1 = y_2$. Prove that $\mathbb{R}^2/\sim$ is homeomorphic to $\mathbb{R}$.

b) Find a counterexample to the following assertion: if $\sim$ is an equivalence relation on a space X and each equivalence class is homeomorphic to the same space Y, then $(X/\sim) \times Y$ is homeomorphic to X.

Why might someone conjecture that this assertion might be true? In part a), we have $X = \mathbb{R} \times \mathbb{R}$, each equivalence class is homeomorphic to $\mathbb{R}$ and $(X/\sim) \times \mathbb{R} \simeq \mathbb{R} \times \mathbb{R} \simeq X$. In this example, you "divide out" equivalence classes that all look like $\mathbb{R}$, then "multiply" by $\mathbb{R}$, and you're back where you started.

c) Let $g : \mathbb{R}^2 \to \mathbb{R}$ be given by $g(x, y) = x^2 + y^2$. Then the quotient space $\mathbb{R}^2/g$ is homeomorphic to what familiar space?

d) On $\mathbb{R}^2$, define an equivalence relation $(x_1, y_1) \sim (x_2, y_2)$ iff $x_1 + y_1^2 = x_2 + y_2^2$. Prove that $\mathbb{R}^2/\sim$ is homeomorphic to some familiar space.

e) Define an equivalence relation on $\mathbb{R}$ by $x \sim y$ if an only if $x - y \in \mathbb{Z}$. What is the quotient space $\mathbb{R}/\sim$? Explain.

E23. For $x, y \in [0, 1]$, define $x \sim y$ iff $x - y$ is rational. Prove that the corresponding quotient space $[0, 1]/\sim$ is trivial.

E24. Prove that a 1-1 quotient map is a homeomorphism.

E25. a) Let $Y_1 = [0, 1]$ with its usual topology and $Y_2 = [2, 3]$ with the discrete topology. Define $g : Y_1 + Y_2 \to Y_1$ by letting $g(x) = \begin{cases} x & \text{if } x \in Y_1 \\ x - 2 & \text{if } x \in Y_2 \end{cases}$. Prove that g is a quotient map that is neither open nor closed.

b) Let $\mathbb{R}^2$ have the topology $\mathcal{T}$ for which a subbase consists of all the usual open sets together with the singleton set $\{(0, 0)\}$. Let $\mathbb{R}$ have the usual topology and let $f : \mathbb{R}^2 \to \mathbb{R}$ be the projection $f(x, y) = x$. Prove that f is a quotient map which is neither open nor closed.

E26. State and prove a theorem of the form:

"for two disjoint subsets A and B of $\mathbb{R}^2$, $A + B$ is homeomorphic to $A \cup B$ iff ... "

E27. Let $\mathbb{R}^2$ have the topology $\mathcal{T}$ for which a subbasis consists of all the usual open sets together with the singleton set $\{(0,0)\}$. Let $\mathbb{R}$ have the usual topology and define $f: \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = x$. Prove that f is a quotient map which is neither open nor closed.

E28. Let $Y_1 = \mathbb{N} + (\mathbb{N} \times (0,1))$ and let $Y_2 = Y_1 + [0,1)$. Prove Y_1 is not homeomorphic to Y_2 but that each is a continuous one-to-one image of the other

E29. Show that no continuous image of $\mathbb{R}$ can be represented as a topological sum $X + Y$, where $X, Y \neq \emptyset$. How can this be result be strengthened?

E30. Suppose X_s (s $\in S$) and Y_t ($t \in T$) are pairwise disjoint spaces. Prove that $\sum_{s \in S} X_s \times \sum_{t \in T} Y_t$ is homeomorphic to $\sum_{s \in S, t \in T} (X_s \times Y_t)$.

E31. This problem outlines a proof (due to Ira Rosenholtz) that every nonempty compact metric space X is a continuous image of the Cantor set C. From Example 4.5, we know that X is homeomorphic to a subspace of $[0,1]^{\aleph_0}$.

a) Prove that the Cantor set $C \subseteq \mathbb{R}$ consists of all reals of the form $\sum_{j=0}^{\infty} \frac{a_j}{3^j}$ where each $a_j = 0$ or 2.

b) Prove that $[0,1]$ is a continuous image of C. *Hint: Define* $g(\sum_{j=0}^{\infty} \frac{a_j}{3^j}) = \sum_{j=0}^{\infty} \frac{a_j}{2^j}$.

c) Prove that the cube $[0,1]^{\aleph_0}$ is a continuous image of C. *Hint: By Corollary 2.21,* $C \simeq \{0,2\}^{\aleph_0} \simeq C^{\aleph_0}$. *Use g from part b) to define* $f : C^{\aleph_0} \to [0,1]^{\aleph_0}$ *by* $f(x_1, x_2, ..., ...) = (g(x_1), g(x_2), ..., ...)$

d) Prove that a closed set $K \subseteq C$ is a continuous image of C. *Hint: C is homeomorphic to the "middle two-thirds" set C' consisting of all reals of the form* $\sum_{j=0}^{\infty} \frac{b_j}{6^j}$. *$C'$ has the property that if $x, y \in C'$, then $\frac{x+y}{2} \notin C'$. If K' is closed in C', we can map $C' \to K$ by sending each point x to the point in K' nearest to x.*)

e) Prove that every nonempty compact metric space X is a continuous image of C.

.

Chapter VI Review

Explain why each statement is true, or provide a counterexample.

1. $(0,1)^{\aleph_0}$ is open in $[0,1]^{\aleph_0}$.

2. Suppose F is a closed set in $[0,1] \times \mathbb{R}$. Then $\pi_2[F]$ is a closed set in $\mathbb{R}$.

3. $\mathbb{N}^{\aleph_0}$ is discrete.

4. If C is the Cantor set, then there is a complete metric on $C^{\aleph_0}$ which produces the product topology.

5. Let $\mathbb{R}^{\mathbb{R}}$ have the box topology. A sequence $(f_n) \to f \in \mathbb{R}^{\mathbb{R}}$ iff $(f_n) \to f$ uniformly.

6. Let $f_n : [0,1] \to [0,1]$ be given by $f_n(x) = x^n$. The sequence (f_n) has a limit in $[0,1]^{[0,1]}$.

7. Let $g \in \mathbb{R}^{\mathbb{R}}$ be defined by $g(x) = x^2$ for all $x \in \mathbb{R}$. Give an example of a sequence (f_n) of distinct functions in $\mathbb{R}^{\mathbb{R}}$ that converges to g.

8. Let C be the Cantor set. Then C is homeomorphic to the topological sum $C + C$.

9. The projection maps π_x and π_y from $\mathbb{R}^2 \to \mathbb{R}$ separate points from closed sets.

10. The letter N is a quotient of the letter M .

11. Suppose $\sim$ is an equivalence relation on X and that for $x \in X$, $[\,x\,]$ represents its equivalence class. If x is a cut point of X, then $[\,x\,]$ is a cut point of the quotient space $X/\sim$.

12. If g: $X \to Y$ is a quotient map and Y is compact T_2, then Y is compact T_2.

13. Suppose A is infinite and that in each space X_α ($\alpha \in A$) there is a nonempty proper open subset O_α. Then $\prod O_\alpha$ is not a basic open set in the product topology on $X = \prod X_\alpha$. Moreover, $\prod O_\alpha$ cannot even be open in the product.

14. If $X = \bigcup_{n=1}^{\infty} A_n$, where the A_n's are disjoint clopen sets in X, then $X \cong \sum_{n=1}^{\infty} A_n$ ($=$ the topological sum of the A_n's).

15. Let $X_n = \{0,1\}$ and $Y_n = \mathbb{N}$ with their usual topologies. Then $\sum\limits_{n=1}^{\infty} X_n$ is homeomorphic to $\sum\limits_{n=1}^{\infty} Y_n$.

16. Let $X_n = \{0,1\}$ and $Y_n = \mathbb{N}$ with their usual topologies. Then $\prod_{n=1}^{\infty} X_n$ is homeomorphic to $\prod_{n=1}^{\infty} Y_n$.

17. Let $A = \{(x,y,z) \in \mathbb{R}^3\colon\; x - y^2 - 2yz - z^2 > |\sin(xyz)|\,\}$, and let $f\colon \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x,y,z) = x + 2$. Then $f\,[A]$ is open but not closed in $\mathbb{R}$.

18. Suppose A_n is a connected subset of $X_n (\neq \emptyset)$ and that $\prod_{n=1}^{\infty} A_n$ is dense in $\prod_{n=1}^{\infty} X_n$. Then each X_n is connected.

19. In $\mathbb{R}^{\mathbb{R}}$, every neighborhood of the function sin contains a step function (that is, a function with finite range).

20. Let X be an uncountable set with the cocountable topology. Then $\{(x,x) : x \in X\}$ is a closed subset of the product $X \times X$..

21. Let $A = \{\frac{1}{n} : n \in \mathbb{N}\} \subseteq \mathbb{R}$, and let m be any cardinal number. A^k is a closed set in $\mathbb{R}^k$.

22. If C is the Cantor set, then $C \times C \times C \times C$ is homeomorphic to $C \times C$.

33. $S^1 \times S^1$ is homeomorphic to the "infinity symbol": ∞

34. 31. Let P be the set of all real polynomials in one variable, with domain restricted to $[0,1]$, for which ran$(P) \subseteq [0,1]$. Then P is dense in $[0,1]^{[0,1]}$.

35. Every metric space is a quotient of a pseudometric space.

37. A separable metric space with a basis of clopen sets is homeomorphic to a subspace of the Cantor set.

38. Let $\prod_{\alpha \in A} X_\alpha$ be a nonempty product space. Then each factor X_α is a quotient of $\prod_{\alpha \in A} X_\alpha$.

39. Suppose X does not have the trivial topology. Then X^{2^c} cannot be separable.

40. Every countable space X is a quotient of $\mathbb{N}$.

41. $\mathbb{N} \times \mathbb{N}$ is homeomorphic to the sum of $\aleph_0$ disjoint copies of $\mathbb{N}$.

42. Suppose $x = (x_\alpha) \in$ int A, where $A \subseteq \prod X_\alpha$. Then for every α, $x_\alpha \in$ int $\pi_\alpha[A]$.

43. The unit circle, S^1, is homeomorphic to a product $\prod_{\alpha \in A} X_\alpha$, where each $X_\alpha \subseteq [0,1]$ (i.e., S^1 can be "factored" into a product of subspaces of $[0,1]$).

44. $\mathbb{N}^{\aleph_0}$ is homeomorphic to $\mathbb{R}$.

Chapter VII
Separation Properties

1. Introduction

"Separation" refers here to whether objects such as points or disjoint closed sets can be enclosed in disjoint open sets. In spite of the similarity of terminology, "separation properties" have no direct connection to the idea of "separated sets" that appeared in Chapter 5 in the context of connected spaces.

We have already met some simple separation properties of spaces: the T_0, T_1 and T_2 (Hausdorff) properties. In this chapter, we look at these and others in more depth. As hypotheses for "more separation" are added, spaces generally become nicer and nicer – especially when "separation" is combined with other properties. For example, we will see that "enough separation" and "a nice base" guarantees that a space is metrizable.

"Separation axioms" translates the German term *Trennungsaxiome* used in the original literature. Therefore the standard separation axioms were historically named T_0, T_1,T_2, T_3, and T_4, each one stronger than its predecessor in the list. Once these had become common terminology, another separation axiom was discovered to be useful and "interpolated" into the list: $T_{3\frac{1}{2}}$. It turns out that the $T_{3\frac{1}{2}}$ spaces (also called Tychonoff spaces) are an extremely well-behaved class of spaces with some very nice properties.

2. The Basic Ideas

Definition 2.1 A topological space X is called a

1) T_0-space if, whenever $x \neq y \in X$, there either exists an open set U with $x \in U$, $y \notin U$ or there exists an open set V with $y \in V$, $x \notin V$

2) T_1-space if, whenever $x \neq y \in X$, there exists an open set U with $x \in U, y \notin V$ and there exists an open set V with $x \notin U, y \in V$

3) T_2-space (or, Hausdorff space) if, whenever $x \neq y \in X$, there exist disjoint open sets U and V in X such that $x \in U$ and $y \in V$.

It is immediately clear from the definitions that $T_2 \Rightarrow T_1 \Rightarrow T_0$.

Example 2.2

1) X is a T_0-space if and only if: whenever $x \neq y$, then $\mathcal{N}_x \neq \mathcal{N}_y$ – that is, different points in X have different neighborhood systems.

2) If X has the trivial topology and $|X| > 1$, then X is not a T_0-space.

3) A pseudometric space (X, d) is a metric space in and only if (X, d) is a T_0-space.

Clearly, a metric space is T_0. On the other hand, suppose (X, d) is T_0 and that $x \neq y$. Then for some $\epsilon > 0$ either $x \notin B_\epsilon(y)$ or $y \notin B_\epsilon(x)$. Either way, $d(x, y) \geq \epsilon$, so d is a metric.

4) In any topological space X we can define an equivalence relation $x \sim y$ iff $\mathcal{N}_x = \mathcal{N}_y$. Let $g : X \to X/\sim\ = Y$ by $g(x) = [x]$. Give Y the quotient topology. Then g is continuous, onto, open (not automatic for a quotient map!) and the quotient is a T_0 space:

If O is open in X, we want to show that $g[O]$ is open in Y, and because Y has the quotient topology this is true iff $g^{-1}[g[O]]$ is open in X. But $g^{-1}[g[O]]$
$= \{x \in X : g(x) \in g[O]\} = \{x \in X :$ for some $y \in O, g(x) = g(y)\}$
$= \{x \in X :\ x$ is equivalent to some point y in $O\} = O$.

If $[x] \neq [y] \in Y$, then x is not equivalent to y, so there is an open set $O \subseteq X$ with (say) $x \in O$ and $y \notin O$. Since g is open, $g[O]$ is open in Y and $[x] \in g[O]$. Moreover, $[y] \notin g[O]$ or else y would be equivalent to some point of O – implying $y \in O$.

Y is called the T_0-identification of X. This identification turns any space into a T_0-space by identifying points that have identical neighborhoods. If X is a T_0-space to begin with, then g is one-to-one and g is a homeomorphism. Applied to a T_0 space, the T_0-identification accomplishes nothing. If (X, d) is a pseudometric space, the T_0-identification is the same as the metric identification discussed in Example VI.5.6 because, in that case, $\mathcal{N}_x = \mathcal{N}_y$ if and only if $d(x, y) = 0$.

5) For $i = 0, 1, 2$: if $(X, \mathcal{T})$ is a T_i space and $\mathcal{T}' \supseteq \mathcal{T}$ is a new topology on X, then $(X, \mathcal{T}')$ is also a T_i space.

Example 2.3

1) (Exercise) It is easy to check that a space X is a T_1 space

iff for each $x \in X$, $\{x\}$ is closed
iff for each $x \in X$, $\{x\} = \bigcap\{O :\ O$ is open and $x \in O\}$

2) A finite T_1 space is discrete.

3) Sierpinski space $X = \{0, 1\}$ with topology $\mathcal{T} = \{\emptyset, \{1\}, \{0, 1\}\})$ is T_0 but not T_1: $\{1\}$ is an open set that contains 1 and not 0; but there is no open set containing 0 and not 1.

4) $\mathbb{R}$, with the right-ray topology, is T_0 but not T_1: if $x < y \in \mathbb{R}$, then $O = (x, \infty)$ is an open set that contains y and not x; but there is no open set that contains x and not y.

5) With the cofinite topology, $\mathbb{N}$ is T_1 but not T_2 because, in an infinite cofinite space, any two nonempty open sets have nonempty intersection.

These separation properties are very well-behaved with respect to subspaces and products.

Theorem 2.4 For $i = 0, 1, 2$:

a) A subspace of a T_i-space is a T_i-space
b) If $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$, then X is a T_i-space iff each X_α is a T_i-space.

Proof All of the proofs are easy. We consider here only the case $i = 1$, leaving the other cases as an exercise.

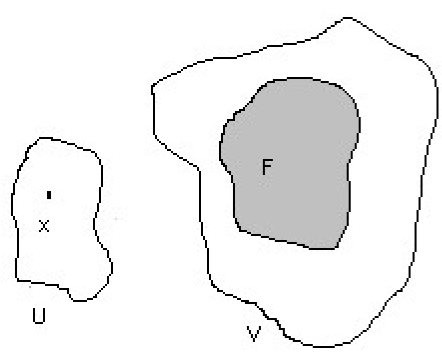
a) Suppose $a \neq b \in A \subseteq X$, where X is a T_1 space. If U' is an open set in X containing x but not y, then $U = U' \cap A$ is an open set in A containing x but not y. Similarly we can find an open set V in A containing y but not x. Therefore A is a T_1-space.

b) Suppose $X = \prod_{\alpha \in A} X_\alpha$ is a nonempty T_1-space. Each X_α is homeomorphic to a subspace of X, so, by part a), each X_α is T_1. Conversely, suppose each X_α is T_1 and that $x \neq y \in X$. Then $x_\alpha \neq y_\alpha$ for some α. Pick an open set U_α in X_α containing x_α but not y_α. Then $U = \, < U_\alpha >$ is an open set in X containing x but not y. Similarly, we find an open set V in X containing y but not x. Therefore X is a T_1-space. ∙

Exercise 2.5 Is a continuous image of a T_i-space necessarily a T_i-space? How about a quotient? A continuous open image?

We now consider a slightly different kind of separation axiom for a space X : formally, the definition is "just like" the definition of T_2, but with a closed set replacing one of the points.

Definition 2.6 A topological space X is called regular if whenever F is a closed set and $x \notin F$, there exist disjoint open sets U and V such that $x \in U$ and $F \subseteq V$.



There are some easy equivalents of the definition of "regular" that are useful to recognize.

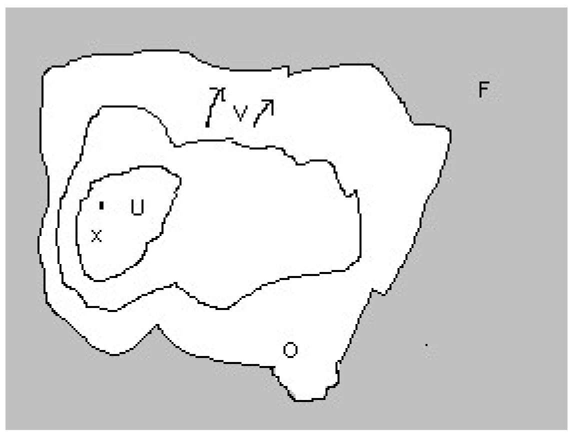
Theorem 2.7 The following are equivalent for any space X:

i) X is regular

ii) if O is an open set containing x, then there exists an open set $U \subseteq X$ such that $x \in U \subseteq \mathrm{cl}\, U \subseteq O$

iii) at each point $x \in X$ there exists a neighborhood base consisting of closed neighborhoods.

Proof i) $\Rightarrow$ ii) Suppose X is regular and O is an open set with $x \in O$. Letting $F = X - O$, we use regularity to get disjoint open sets U, V with $x \in U$ and $F \subseteq V$ as illustrated below:



Then $x \in U \subseteq \mathrm{cl}\, U \subseteq O$ (since $\mathrm{cl}\, U \subseteq X - V$).

ii) $\Rightarrow$ iii) If $N \in \mathcal{N}_x$, then $x \in O = \mathrm{int}\, N$. By ii), we can find an open set U so that $x \in U \subseteq \mathrm{cl}\, U \subseteq O \subseteq N$. Since $\mathrm{cl}\, U$ is a neighborhood of x, the closed neighborhoods of x form a neighborhood base at x.

iii) $\Rightarrow$ i) Suppose F is closed and $x \notin F$. By ii), there is a closed neighborhood K of x such that $x \in K \subseteq X - F$. We can choose $U = \mathrm{int}\, K$ and $V = X - K$ to complete the proof that X is regular. $\bullet$

Example 2.8 Every pseudometric space (X, d) is regular. Suppose $a \notin F$ and F is closed. We have a continuous function $f(x) = d(x, F)$ for which $f(a) = c > 0$ and $f|F = 0$. This gives us disjoint open sets with $a \in U = f^{-1}[(\frac{c}{2}, \infty)]$ and $F \subseteq V = f^{-1}[(-\infty, \frac{c}{2})]$. Therefore X is regular.

At first glance, one might think that regularity is a stronger condition than T_2. But this is false: if (X, d) is a pseudometric space but not a metric space, then X is regular but not even T_0.

To bring things into line, we make the following definition.

Definition A topological space X is called a T_3-space if X is regular <u>and</u> T_1.

It is easy to show that $T_3 \Rightarrow T_2$ ($\Rightarrow T_1 \Rightarrow T_0$): suppose X is T_3 and $x \neq y \in X$. Then $F = \{y\}$ is closed so, by regularity, there are disjoint open sets U, V with $x \in U$ and $y \in \{y\} \subseteq V$.

<u>Caution</u> Terminology varies from book to book. For some authors, the definition of "regular" includes T_1 : for them, "regular" means what we have called "T_3." Check the definitions when reading other books.

Exercise 2.10 Show that a regular T_0 space must be T_3 (*so it would have been equivalent to use "T_0" instead of "T_1" in the definition of "T_3"*).

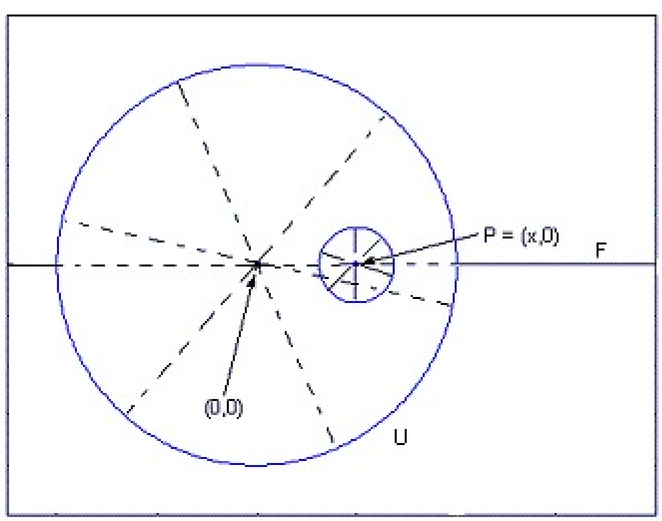
Example 2.11 $T_2 \not\Rightarrow T_3$. We will put a new topology on the set $X = \mathbb{R}^2$. At each point $p \in X$, let a neighborhood base $\mathcal{B}_p$ consist of all sets N of the form

$$N = B_\epsilon(p) - (\text{a finite number of straight lines through } p) \cup \{p\} \text{ for some } \epsilon > 0.$$

(*Check that the conditions in the Neighborhood Base Theorem III.5.2 are satisfied.*) With the resulting topology, X is called the <u>slotted plane</u>. Note that $B_\epsilon(p) \in \mathcal{B}_p$ (*because "0" is a finite number*), so each $B_\epsilon(p)$ is among the basic neighborhoods in $\mathcal{B}_p$ – so the slotted plane topology on $\mathbb{R}^2$ contains the usual Euclidean topology. It follows that X is T_2.

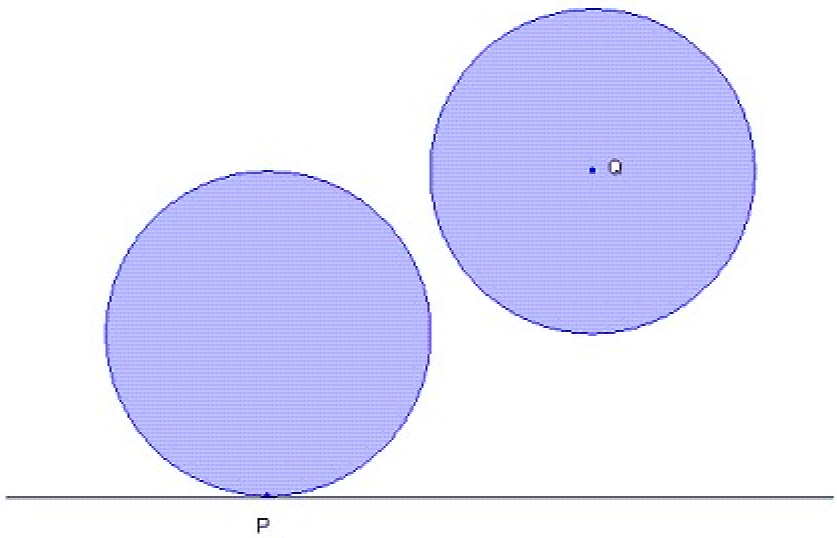
The set $F = \{(x, 0) : x \neq 0\}$ = "the x-axis with the origin deleted" is a closed set in X (*why?*).

If U is any open set containing the origin $(0, 0)$, then there is a basic neighborhood N with $(0, 0) \in N \subseteq U$. Using the ϵ in the definition of N, we can choose a point $p = (x, 0) \in F$ with $0 < x < \epsilon$. <u>Every</u> basic neighborhood set of p must intersect N (*why?) and therefore must intersect* U. It follows that $(0, 0)$ and F cannot be separated by disjoint open sets, so the slotted plane is not regular (and therefore not T_3).



Note: The usual topology in $\mathbb{R}^2$ is regular. This example shows that an "enlargement" of a regular (or T_3) topology may not be regular (or T_3). Although the enlarged topology has more open sets to work with, there are also more "point/closed set pairs x, F" that need to be separated. By contrast, it is easy to see that an "enlargement" of a T_i topology $(i = 0, 1, 2)$ is still T_i.

Example 2.12 The Moore plane Γ (Example III.5.6) is clearly T_2. In fact, at each point, there is a neighborhood base of closed neighborhoods. The figure illustrates this for a point P on the x-axis and a point Q above the x-axis. Therefore Γ is T_3.



Theorem 2.13 a) A subspace of a regular (T_3) space is regular (T_3).
b) Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$. X is regular (T_3) iff each X_α is regular (T_3).

Proof a) Let $A \subseteq X$ where X is regular. Suppose $a \in A$ and that F is a closed set in A that does not contain a. There exists a closed set F' in X such that $F' \cap A = F$. Choose disjoint open sets U' and V' in X with $a \in U'$ and $F' \subseteq V'$. Then $U = U' \cap A$ and $V = V' \cap A$ are open in A, disjoint, $a \in U$, and $F \subseteq V$. Therefore A is regular.

b) If $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$ is regular, then part a) implies that each X_α is regular, because each X_α is homeomorphic to a subspace of X. Conversely, suppose each X_α is regular and that $U = \, < U_{\alpha_1}, ..., U_{\alpha_n} >$ is a basic open set containing x. For each α_i, we can pick an open set V_{α_i} in X_{α_i} such that $x_{\alpha_i} \in V_{\alpha_i} \subseteq \text{cl}\, V_{\alpha_i} \subseteq U_{\alpha_i}$. Then $x \in V = \, < V_{\alpha_1}, ..., V_{\alpha_n} > \, \subseteq \text{cl}\, V \subseteq \, < \text{cl}\, V_{\alpha_1}, ..., \text{cl}\, V_{\alpha_n} > \, \subseteq U$. (*Why is the last inclusion true?*) Therefore X is regular.

Since the T_1 property is hereditary and productive, a) and b) also hold for T_3-spaces •

The obvious "next step up" in separation is the following:

Definition 2.14 A topological space X is called normal if, whenever A, B are disjoint closed sets in X, there exist disjoint open sets U, V in X with $A \subseteq U$ and $B \subseteq V$. X is called a T_4-space if X is normal and T_1.

Example 2.15 a) Every pseudometric space (X, d) is normal (so every metric space is T_4).
In fact, if A and B are disjoint closed sets, we can define $f(x) = \frac{d(x,A)}{d(x,A)+d(x,B)}$. Since the denominator cannot be 0, f is continuous and $f|A = 0$, $f|B = 1$. The open sets $U = \{x : f(x) < \frac{1}{2}\}$ and $V = \{x : f(x) > \frac{1}{2}\}$ are disjoint that contain A and B respectively. Therefore X is normal.
Note: the argument given is slick and clean. Can you show (X, d) is normal by directly constructing a pair of disjoint open sets that contain A and B ?

b) Let $\mathbb{R}$ have the right ray topology $\mathcal{T} = \{(x, \infty); x \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$. $(\mathbb{R}, \mathcal{T})$ is normal because the only possible pair of disjoint closed sets is $\emptyset$ and X and we can separate these using the disjoint open sets $U = \emptyset$ and $V = X$. Also, $(\mathbb{R}, \mathcal{T})$ is not regular: for example 1 is not in the closed set $F = (-\infty, 0]$, but every open set that contains F also contains 1. So normal $\not\Rightarrow$ regular. But $(\mathbb{R}, \mathcal{T})$ is not T_1 and therefore not T_4.

When we combine "normal $+ T_1$" into T_4, we have a property that fits perfectly into the separation hierarchy.

Theorem 2.16 $T_4 \Rightarrow T_3\ (\Rightarrow T_2 \Rightarrow T_1 \Rightarrow T_0)$

Proof Suppose X is T_4. If F is a closed set not containing x, then $\{x\}$ and F are disjoint closed sets. By normality, we can find disjoint open sets separating $\{x\}$ and F. It follows that X is regular and therefore T_3. ●

Exercises

E1. X is called a door space if every subset is either open or closed. Prove that if a T_2-space X contains two points that are not isolated, then X is not a door space, and that otherwise X is a door space.

E2. A base for the closed sets in a space X is a collection of $\mathcal{F}$ of closed subsets such that every closed set F is an intersection of sets from $\mathcal{F}$. Clearly, $\mathcal{F}$ is a base for the closed sets in X iff $\mathcal{B} = \{O : O = X - F,\ F \in \mathcal{F}\}$ is a base for the open sets in X.

For a polynomial P in n real variables, define the zero set of P as

$$Z(P) = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \ P\ (x_1, x_2, \ldots, x_n) \ = 0\}$$

a) Prove that $\{Z(P) : \ P$ a polynomial in n real variables$\}$ is the base for the closed sets of a topology (called the Zariski topology) on $\mathbb{R}^n$.

b) Prove that the Zariski topology on $\mathbb{R}^n$ is T_1 but not T_2.

c) Prove that the Zariski topology on $\mathbb{R}$ is the cofinite topology, but that if $n > 1$, the Zariski topology on $\mathbb{R}^n$ is not the cofinite topology.

Note: The Zariski topology arises in studying algebraic geometry. After all, the sets $Z(P)$ are rather special geometric objects—those "surfaces" in $\mathbb{R}^n$ which can be described by polynomial equations $P(x_1, x_2, \ldots, x_n) \ = 0$.

E3. A space X is a $T_{5/2}$ space if, whenever $x \neq y \in X$, there exist open sets U and V such that $x \in U$, $y \in V$ and $\mathrm{cl}\, U \cap \mathrm{cl}\, V = \emptyset$. (*Clearly,* $T_3 \Rightarrow T_{5/2} \Rightarrow T_2$.)

a) Prove that a subspace of a $T_{5/2}$ space is a $T_{5/2}$ space.

b) Suppose $X = \prod X_\alpha \neq \emptyset$. Prove that X is $T_{5/2}$ iff each X_α is $T_{5/2}$.

c) Let $S = \{(x, y) \in \mathbb{R}^2 : y \geq 0\}$ and $L = \{(x, y) \in S : y = 0\}$. Define a topology on S with the following neighborhood bases:

if $p \in S - L$:	$\mathcal{B}_p \ = \{B_\epsilon(p) \cap S : \epsilon > 0\}$
if $p \in L$:	$\mathcal{B}_p \ = \{B_\epsilon(p) \cap (S - L) \cup \{p\} : \epsilon > 0\}$

You may assume that these $\mathcal{B}_p$'s satisfy the axioms for a neighborhood base.

Prove that S is $T_{5/2}$ but not T_3.

E4. Suppose $A \subseteq X$. Define a topology on X by

$$\mathcal{T} = \{O \subseteq X : O \supseteq A\} \cup \{\emptyset\}$$

Decide whether or not $(X, \mathcal{T})$ is normal.

E5. A function $f : X \to Y$ is called perfect if f is continuous, closed, onto, and, for each $y \in Y$, $f^{-1}(y)$ is compact. Prove that if X is regular and f is perfect, then Y is regular; and that if X is T_3, the Y is also T_3.

E6. a) Suppose X is finite. Prove that $(X, \mathcal{T})$ is regular iff there is a partition $\mathcal{B}$ of X that is a base for the topology.

b) Give an example to show that a compact subset K of a regular space X need not be closed. However, show that if X is regular then $\operatorname{cl} K$ is compact.

c) Suppose F is closed in a T_3-space X. Prove that

i) Prove that $F = \bigcap\{O : O \text{ is open and } F \subseteq O\}$.

ii) Define $x \sim y$ iff $x = y$ or $x, y \in F$. Prove that the quotient space $X/\sim$ is Hausdorff.

d) Suppose B is an infinite subset of a T_3-space X. Prove that there exists a sequence of open sets U_n such that each $U_n \cap B \neq \emptyset$ and that $\operatorname{cl} U_n \cap \operatorname{cl} U_m = \emptyset$ whenever $n \neq m$.

e) Suppose each point y in a space Y has a neighborhood V such that $\operatorname{cl} V$ is regular. Prove that Y is regular.

3. Completely Regular Spaces and Tychonoff Spaces

The T_3 property is well-behaved. For example, we saw in Theorem 2.13 that the T_3 property is hereditary and productive. However, the T_3 property is not sufficiently strong to give us really nice theorems.

For example, it's very useful if a space has many (nonconstant) continuous real-valued functions available to use. Remember how many times we have used the fact that continuous real-valued functions f can be defined on a metric space (X, d) using formulas like $f(x) = d(x, a)$ or $f(x) = d(x, F)$; when $|X| > 1$, we get many nonconstant real functions defined on (X, d). But a T_3-space can sometimes be very deficient in continuous real-valued functions – in 1946, Hewett gave an example of a infinite T_3-space H on which the only continuous real-valued functions are the constant functions.

In contrast, we will see that the T_4 property is strong enough to guarantee the existence of lots of continuous real-valued functions and, therefore, to prove some really nice theorems (for example, see Theorems 5.2 and 5.6 later in this chapter). The downside is that T_4-spaces turn out also to exhibit some very bad behavior: the T_4 property is not hereditary (*explain why a proof analogous to the one given for Theorem 2.13b) doesn't work*) and it is not even finitely productive. Examples of such bad behavior are a little hard to find right now, but later they will appear rather naturally.

These observations lead us to look first at a class of spaces with separation somewhere "between T_3 and T_4." We want a group of spaces that is well-behaved, but also with enough separation to give us some very nice theorems. We begin with some notation and a lemma.

Recall that $\quad C(X) = \{f \in \mathbb{R}^X : f \text{ is continuous}\} =$ the collection of continuous real-valued functions on X

$\quad C^*(X) = \{f \in C(X) : f \text{ is bounded}\} =$ the collection of continuous bounded real-valued functions on X

Lemma 3.1 Suppose $f, g \in C(X)$. Define real-valued functions $f \vee g$ and $f \wedge g$ by

$$(f \vee g)(x) = \max\{f(x), g(x)\}$$
$$(f \wedge g)(x) = \min\{f(x), g(x)\}$$

Then $f \vee g$ and $f \wedge g$ are in $C(X)$.

Proof We want to prove that the max or min of two continuous real-valued functions is continuous. But this follows immediately from the formulas

$$(f \vee g)(x) = \frac{f(x) + g(x)}{2} + \frac{|f(x) - g(x)|}{2}$$

$$(f \wedge g)(x) = \frac{f(x) + g(x)}{2} - \frac{|f(x) - g(x)|}{2} \quad \bullet$$

Definition 3.2 A space X is called <u>completely regular</u> if whenever F is a closed set and $a \notin F$, there exists a function $f \in C(X)$ such that $f(a) = 0$ and $f|F = 1$.

Informally, "completely regular" means that "a and F can be separated by a continuous real-valued function."

<u>Note</u> i) The definition requires that $f|F = 1$, in other words, that $F \subseteq f^{-1}[\{1\}]$. However, these two sets <u>might not</u> be equal.

ii) If there is such a function f, there is also a continuous $g : X \to [0,1]$ such that $g(a) = 0$ and $g|F = 1$. For example, we could use $g = (f \vee 0) \wedge 1$ which, by Lemma 3.1, is continuous.

iii) Suppose $g : X \to [0,1]$ is continuous and $g(a) = 0$, $g|F = 1$. The particular values $0, 1$ in the definition are not important.: they could be any real numbers $r < s$. (*If we choose a homeomorphism $\phi : [0,1] \to [r,s]$, then it must be true that either $\phi(0) = r$, $\phi(1) = s$ or $\phi(0) = s$, $\phi(1) = r$ – why?. Then $h = \phi \circ g : X \to [r,s]$ and, depending on how you chose ϕ, $h(a) = r$ and $h|F = b$ or vice-versa.*)

Putting these observations together, we see Definition 3.2 is equivalent to:

Definition 3.2′ A space X is called <u>completely regular</u> if whenever F is a closed set, $a \notin F$, and r, s are real numbers with $r < s$, then there exists a continuous function $f : X \to [r,s]$ for which $f(a) = r$ and $f|F = s$.

In one way, the definition of "completely regular space" is very different the definitions for the other separation properties: the definition isn't "internal"because an "external" space, $\mathbb{R}$ is an integral part of the definition. While it <u>is</u> possible to contrive a purely internal definition for "completely regular," the definition is complicated and seems completely unnatural: it simply imposes some very unintuitive conditions to force the existence of enough functions in $C(X)$.

Example 3.3 Suppose (X, d) is a pseudometric space with a closed subset F and $a \notin F$. Then $f(x) = \frac{d(x,F)}{d(a,F)}$ is continuous, $f(a) = 1$ and $f|F = 0$. So (X, d) is completely regular, but if d is not a metric, then this space is not even T_0.

Definition 3.4 A completely regular T_1-space X is called a <u>Tychonoff space</u> (or $T_{3\frac{1}{2}}$-space).

Theorem 3.5 $T_{3\frac{1}{2}} \Rightarrow T_3$ ($\Rightarrow T_2 \Rightarrow T_1 \Rightarrow T_0$)

Proof Suppose F is a closed set in X not containing a. If X is $T_{3\frac{1}{2}}$, we can choose $f \in C(X)$ with $f(a) = 0$ and $f|F = 1$. Then $U = f^{-1}[(-\infty, \frac{1}{2})]$ and $V = f^{-1}[(\frac{1}{2}, \infty)]$ are disjoint open sets with $a \in U$, $F \subseteq V$. Therefore X is regular. Since X is T_1, X is T_3. •

Hewitt's example of a T_3 space on which every continuous real-valued function is constant is more than enough to show that a T_3 space may not be $T_{3\frac{1}{2}}$ (the example, in *Ann. Math., 47(1946) 503-509*, is rather complicated.). For that purpose, it is a little easier – but still nontrivial – to find a T_3 space X containing two points p, q such that for all $f \in C(X)$, $f(p) = f(q)$. Then p and $\{q\}$ cannot be separated by a function from $C(X)$ so X is not $T_{3\frac{1}{2}}$. (See D.J. Thomas, *A regular space, not completely regular,* American Mathematical Monthly, 76(1969), 181-182). The space X can then be used to construct an infinite T_3 space H (simpler than Hewitt's example) on which every continuous real-valued function is constant (see Gantner, *A regular space on which every continuous real-valued function is constant,* American Mathematical Monthly, 78(1971), 52.) Although we will not present these constructions here, we will occasionally refer to H in comments later in this section.

Note: We have not yet shown that $T_4 \Rightarrow T_{3\frac{1}{2}}$: this is true (as the notation suggests), but it is not at all easy to prove: try it! This result is in Corollary 5.3.

Tychonoff spaces continue the pattern of good behavior that we saw in preceding separation axioms, and they will also turn out to be a rich class of spaces to study.

Theorem 3.7 a) A subspace of a completely regular ($T_{3\frac{1}{2}}$) space is completely regular ($T_{3\frac{1}{2}}$).
b) Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$. X is completely regular ($T_{3\frac{1}{2}}$) iff each X_α is completely regular ($T_{3\frac{1}{2}}$) .

Proof Suppose $a \notin F \subseteq A \subseteq X$, where X is completely regular and F is a closed set in A. Pick a closed set K in X such that $K \cap A = F$ and an $f \in C(X)$ such that $f(a) = 0$ and $f|K = 1$. Then $g = f|A \in C(A)$, $g(a) = 0$ and $g|F = 1$. Therefore A is completely regular.

If $\emptyset \neq X = \prod_{\alpha \in A} X_\alpha$ is completely regular, then each X_α is homeomorphic to a subspace of X so each X_α is completely regular. Conversely, suppose each X_α is completely regular and that F is a closed set in X not containing a. There is a basic open set U such that

$$a \in U = \ < U_{\alpha_1}, ... U_{\alpha_n} > \ \subseteq X - F$$

For each $i = 1, ..., n$ we can pick a continuous function $f_{\alpha_i} : X_{\alpha_i} \rightarrow [0,1]$ with $f_{\alpha_i}(a_{\alpha_i}) = 0$ and $f_{\alpha_i}|(X_{\alpha_i} - U_{\alpha_i}) = 1$. Define $f : X \rightarrow [0,1]$ by

$$f(x) = \max \{(f_{\alpha_i} \circ \pi_{\alpha_i})(x) : i = 1, ..., n\} = \max \{f_{\alpha_i}(x_{\alpha_i}) : i = 1, ..., n\}$$

Then f is continuous and $f(a) = \max \{f_{\alpha_i}(a_{\alpha_i}) : i = 1, ..., n\} = 0$. If $x \in F$, then for some i, $x_{\alpha_i} \notin U_{\alpha_i}$ and $f_{\alpha_i}(x_{\alpha_i}) = 1$, so $f(x) = 1$. Therefore $f|F = 1$ and X is completely regular.
.

Since the T_1 property is both hereditary and productive, the statements in a) and b) also hold for $T_{3\frac{1}{2}}$. •

Corollary 3.8 For any cardinal m, the "cube" $[0,1]^m$ and all its subspaces are $T_{3\frac{1}{2}}$.

Since a Tychonoff space X is defined using functions in $C(X)$, we expect that these functions will have a close relationship to the topology on X. We want to explore that connection.

Definition 3.9 Suppose $f \in C(X)$. Then $Z(f) = f^{-1}[\{0\}] = \{x \in X : f(x) = 0\}$ is called the zero set of f. If $A = Z(f)$ for some $f \in C(X)$, we call A a zero set in X. The complement of a zero set in X is called a cozero set: $\text{coz}(f) = X - Z(f) = \{x \in X : f(x) \neq 0\}$.

A zero set $Z(f)$ in X is closed because f is continuous. In addition, $Z(f) = \bigcap_{n=1}^{\infty} O_n$, where $O_n = \{x \in X : |f(x)| < \frac{1}{n}\}$. Each O_n is open. Therefore a zero set is always a closed G_δ-set. Taking complements shows that $\text{coz}\,(f)$ is always an open F_σ-set in X.

For $f \in C(X)$, let $g = (-1 \vee f) \wedge 1 \in C^*(X)$. Then $Z(f) = Z(g)$. Therefore $C(X)$ and $C^*(X)$ produce the same zero sets in X (and therefore also the same cozero sets).

Example 3.10

1) A closed set F in a pseudometric space (X, d) is a zero set: $F = Z(f)$, where $f(x) = d(x, F)$.

2) In general, a closed set in X might not be a zero set − in fact, a closed set in X might not even be a G_δ set.

Suppose X is uncountable and $p \in X$. Define a topology on X by letting $\mathcal{B}_x = \{\{x\}\}$ be a neighborhood at each point $x \neq p$ and letting $\mathcal{B}_p = \{B : p \in B \text{ and } X - B \text{ is countable}\}$ be the neighborhood base at p. (*Check that the conditions of the Neighborhood Base Theorem III.5.2 are satisfied.*)

All points in $X - \{p\}$ are isolated and X is clearly T_1. In fact, X is T_4.

If A and B are disjoint closed sets in X, then one of them (say A) satisfies $A \subseteq X - \{p\}$, so A is clopen. We then have open sets $U = A$ and $V = X - A \supseteq B$, so X is normal.

We do not know yet that $T_4 \Rightarrow T_{3\frac{1}{2}}$ in general, but it's easy to see that this space X is also $T_{3\frac{1}{2}}$.

If F is a closed set not containing x, then either $F \subseteq X - \{p\}$ or $\{x\} \subseteq X - \{p\}$. So one of the sets F or $\{x\}$ is clopen and he characteristic function of that clopen set is continuous and works to show that X is completely regular.

The set $\{p\}$ is closed but $\{p\}$ is not a G_δ set in X, so $\{p\}$ <u>is not a zero set</u> in X.

Suppose $p \in \bigcap_{n=1}^{\infty} O_n$ where O_n is open. For each n, there is a basic neighborhood B_n of p such that $p \in B_n \subseteq O_n$, so $X - O_n \subseteq X - B_n$ is countable. Therefore $X - \bigcap_{n=1}^{\infty} O_n = \bigcup_{n=1}^{\infty} (X - O_n)$ is countable. Since X is uncountable, we conclude that $\{p\} \neq \bigcap_{n=1}^{\infty} O_n$.

Even when F is both closed <u>and</u> a G_δ set, F might not be a zero set. We will see examples later.

For purely technical purposes, it is convenient to notice that zero sets and cozero sets can be described in a many different forms. For example, if $f \in C(X)$, then we can see that <u>each set in the left column is a zero set</u> by choosing a suitable $g \in C(X)$:

$$
\begin{array}{lll}
Z = \{x : f(x) = r\} & = Z(g), & \text{where } g(x) = f(x) - r \\
Z = \{x : f(x) \geq 0\} & = Z(g), & \text{where } g(x) = f(x) - |f(x)| \\
Z = \{x : f(x) \leq 0\} & = Z(g), & \text{where } g(x) = f(x) + |f(x)| \\
Z = \{x : f(x) \geq r\} & = Z(g), & \text{where } g(x) = (f(x) - r) - |f(x) - r| \\
Z = \{x : f(x) \leq r\} & = Z(g), & \text{where } g(x) = (f(x) - r) + |f(x) - r|
\end{array}
$$

On the other hand, if $g \in C(X)$, we can write $Z(g)$ in any of the forms listed above by choosing an appropriate function $f \in C(X)$:

$$\begin{array}{ll} Z(g) = \{x : f(x) = r\} & \text{where } \ f(x) = g(x) + r \\ Z(g) = \{x : f(x) \geq 0\} & \text{where } \ f(x) = -|g(x)| \\ Z(g) = \{x : f(x) \leq 0\} & \text{where } \ f(x) = |g(x)| \\ Z(g) = \{x : f(x) \geq r\} & \text{where } \ f(x) = r - |g(x)| \\ Z(g) = \{x : f(x) \leq r\} & \text{where } \ f(x) = r + |g(x)| \end{array}$$

Taking complements, we get the corresponding results for cozero sets: if $f \in C(X)$

i) the sets $\{x : f(x) \neq r\}$, $\{x : f(x) < 0\}$, $\{x : f(x) > 0\}$, $\{x : f(x) < r\}$, $\{x : f(x) > r\}$ are cozero sets, and

ii) any given cozero set can be written in any one of these forms.

Using the terminology of cozero sets, we can see a nice comparison/contrast between regularity and complete regularity. Suppose $x \notin F$, where F is closed in X. If X is regular, we can find disjoint open sets U and V with $x \in U$ and $F \subseteq V$. But if X is completely regular, we can separate x and F with "special" open sets U and V – cozero sets! Just choose $f \in C(X)$ with $f(x) = 0$ and $f|F = 1$, then

$$x \in U = \{x : f(x) < \tfrac{1}{2}\} \qquad \text{and} \qquad F \subseteq V = \{x : f(x) > \tfrac{1}{2}\}$$

In fact this observation characterizes completely regular spaces – that is, if a regular space fails to be completely regular, it is because there is a "shortage" of cozero sets – because there is a "shortage" of functions in $C(X)$ (*see Theorem 3.12, below)*. For the extreme case of a T_3 space H on which the only continuous real valued functions are constant, the only cozero sets are $\emptyset$ and H !

The next theorem reveals the connections between cozero sets, $C(X)$ and the weak topology on X.

Theorem 3.11 For any space $(X, \mathcal{T})$, $C(X)$ and $C^*(X)$ induce the same weak topology $\mathcal{T}_w$ on X, and a base for $\mathcal{T}_w$ is the collection of all cozero sets in X.

Proof A subbase for $\mathcal{T}_w$ consists of all sets of the form $f^{-1}[U]$, where U is open in $\mathbb{R}$ and $f \in C(X)$. Without loss of generality, we can assume the sets U are subbasic open sets of the form (a, ∞) and $(-\infty, b)$, so that the sets $f^{-1}[U]$ have form $\{x \in X : f(x) > a\}$ or $\{x \in X : f(x) < b\}$. But these are cozero sets of X, and every cozero set in X has this form. So the cozero sets are a subbase for $\mathcal{T}_w$ In fact, the cozero sets are actually a base because $\operatorname{coz}(f) \cap \operatorname{coz}(g) = \operatorname{coz}(fg)$: the intersection of two cozero sets is a cozero set.

The same argument, with $C^*(X)$ replacing $C(X)$, shows that the cozero sets of $C^*(X)$ are a base for the weak topology on X generated by $C^*(X)$. But $C(X)$ and $C^*(X)$ produce the same cozero sets in X, and therefore generate the same weak topology $\mathcal{T}_w$ on X. •

Now we can now see the close connection between X and $C(X)$ in completely regular spaces. For any space $(X, \mathcal{T})$ the functions in $C(X)$ certainly are continuous with respect to $\mathcal{T}$ (by definition of $C(X)$). But is $\mathcal{T}$ the smallest topology making this collection of functions continuous? In other words, is $\mathcal{T}$

the weak topology on X generated by $C(X)$? The next theorem says that is true precisely when X is completely regular.

Theorem 3.12 For any space $(X, \mathcal{T})$, the following are equivalent:

a) X is completely regular
b) The cozero sets of X are a base for the topology on X (equivalently, the zero sets of X are a base for the closed sets—meaning that every closed set is an intersection of zero sets)
c) X has the weak topology from $C(X)$ (equivalently, from $C^*(X)$)
d) $C(X)$ (equivalently, $C^*(X)$) separates points from closed sets.

Proof The preceding theorem shows that b) and c) are equivalent.

a) $\Rightarrow$ b) Suppose $a \in O$ where O is open. Let $F = X - O$. Then we can choose $f \in C(X)$ with $f(a) = 0$ and $f|F = 1$. Then $U = \{x : f(x) < \frac{1}{2}\}$ is a cozero set for which $a \in U \subseteq O$. Therefore the cozero sets are a base for X.

b) $\Rightarrow$ d) Suppose F is a closed set not containing a. By b), we can choose $f \in C(X)$ so that $a \in \operatorname{coz}(f) \subseteq X - F$. Then $f(a) = r \neq 0$, so $f(a) \notin \operatorname{cl} f[F] = \{0\}$. Therefore $C(X)$ separates points and closed sets.

d) $\Rightarrow$ a) Suppose F is a closed set not containing a. There is some $f \in C(X)$ for which $f(a) \notin \operatorname{cl} f[F]$. Without loss of generality (*why?*), we can assume $f(a) = 0$. Then, for some $\epsilon > 0$, $(-\epsilon, \epsilon) \cap f[F] = \emptyset$, so that for $x \in F$, $|f(x)| \geq \epsilon$. Define $g \in C^*(X)$ by $g(x) = \min\{|f(x)|, \epsilon\}$. Then $g(a) = 0$ and $g|F = \epsilon$, so X is completely regular.

At each step of the proof, $C(X)$ can be replaced by $C^*(X)$ (*check!*) •

The following corollary is curious and the proof is a good test of whether one understands the idea of "weak topology."

Corollary 3.13 Suppose X is a set and let $\mathcal{T}_{\mathcal{F}}$ be the weak topology on X generated by <u>any</u> family of functions $\mathcal{F} \subseteq \mathbb{R}^X$. Then $(X, \mathcal{T}_{\mathcal{F}})$ is completely regular. $(X, \mathcal{T}_{\mathcal{F}})$ is Tychonoff is $\mathcal{F}$ separates points.

Proof Give X the topology the weak topology $\mathcal{T}_{\mathcal{F}}$ generated by $\mathcal{F}$. Now X has a topology, so the collection $C(X)$ makes sense. Let T_w be the weak topology on X generated by $C(X)$.

The topology $\mathcal{T}_{\mathcal{F}}$ <u>does</u> make all the functions in $C(X)$ continuous, so $\mathcal{T}_w \subseteq \mathcal{T}_{\mathcal{F}}$.

On the other hand: $\mathcal{F} \subseteq C(X)$ by definition of $\mathcal{T}_{\mathcal{F}}$, and the larger collection of functions $C(X)$ generates a (potentially) larger weak topology. Therefore $\mathcal{T}_{\mathcal{F}} \subseteq \mathcal{T}_w$.

Therefore $\mathcal{T}_{\mathcal{F}} = \mathcal{T}_w$. By Theorem 3.12, $(X, \mathcal{T}_{\mathcal{F}})$ is completely regular. •

Example 3.14

1) If $\mathcal{F} = \{f \in \mathbb{R}^{\mathbb{R}} : f \text{ is nowhere differentiable}\}$, then the weak topology $\mathcal{T}_{\mathcal{F}}$ on $\mathbb{R}$ generated by $\mathcal{F}$ is completely regular.

2) If H is an infinite T_3 space on which every continuous real-valued function is constant (*see the comments at the beginning of this Section 3*), then the weak topology generated by $C(X)$ has for a base the collection of cozero sets $\{\emptyset, H\}$. So the weak topology generated by $C(X)$ is the trivial topology, not the original topology on X.

Theorem 3.12 leads to a lovely characterization of Tychonoff spaces.

Corollary 3.15 Suppose X is a Tychonoff space. For each $f \in C^*(X)$, we have $\text{ran}(f) \subseteq [a_f, b_f] = I_f$ for some $a_f < b_f \in \mathbb{R}$. The evaluation map $e : X \to \prod \{I_f : f \in C^*(X)\}$ is an embedding.

Proof X is T_1, the f's are continuous and the collection of f's $(= C^*(X))$ separates points and closed sets. By Corollary VI.4.11, e is an embedding. •

Since each I_f is homeomorphic to $[0,1]$, $\prod\{I_f : f \in C^*(X)\}$ is homeomorphic to $[0,1]^m$, where $m = |C^*(X)|$. Therefore any Tychonoff space can be embedded in a "cube." On the other hand (Corollary 3.8) $[0,1]^m$ and all its subspaces are Tychonoff. So we have:

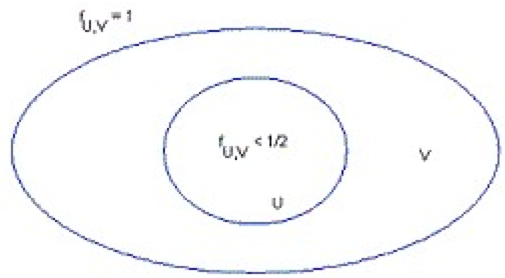
Corollary 3.16 A space X is Tychonoff iff X is homeomorphic to a subspace of a "cube" $[0,1]^m$ for some cardinal number m.

The exponent $m = |C^*(X)|$ in the corollary may not be the smallest exponent possible. As an <u>extreme</u> case, for example, we have $c = |C^*(\mathbb{R})|$, even though we can embed $\mathbb{R}$ in $[0,1] = [0,1]^1$. The following theorem improves the value for m in certain cases (*and we proved a similar result for metric spaces* (X,d)*: see Example VI.4.5.*)

Theorem 3.17 Suppose X is Tychonoff with a base $\mathcal{B}$ of cardinality m. Then X can be embedded in $[0,1]^m$. In particular, X can be embedded in $[0,1]^{w(X)}$.

Proof Suppose m is finite. Since X is T_1, $\{x\} = \bigcap\{B : B \text{ is a basic open set containing } x\}$. Only finitely many such intersections are possible, so X is finite and therefore discrete. Hence $X \overset{\text{top}}{\subseteq} [0,1] \overset{\text{top}}{\subseteq} [0,1]^m$.

Suppose $\mathcal{B}$ is a base of cardinal m where m is infinite. Call a pair $(U,V) \in \mathcal{B} \times \mathcal{B}$ <u>distinguished</u> if there exists a continuous $f_{U,V} : X \to [0,1]$ with $f_{U,V}(x) < \frac{1}{2}$ for all $x \in U$ and $f_{U,V}(x) = 1$ for all $x \in X - V$. Clearly, $U \subseteq V$ for a distinguished pair (U,V). For each distinguished pair, pick such a function $f_{U,V}$ and let $\mathcal{F} = \{f_{U,V} : (U,V) \in \mathcal{B} \times \mathcal{B} \text{ is distinguished}\}$.



We note that if $a \in V \in \mathcal{B}$, then there must exist $U \in \mathcal{B}$ such that $a \in U$ and (U, V) is distinguished. To see this, pick an $f : X \to [0,1]$ so that $f(a) = 0$ and $f|X - V = 1$. Then choose $U \in \mathcal{B}$ so that $a \in U \subseteq f^{-1}[[0, \frac{1}{2})] \subseteq V$.

We claim that $\mathcal{F}$ separates points and closed sets:

> Suppose F is a closed set not containing a. Choose a basic set $V \in \mathcal{B}$ with $a \in V \subseteq X - F$. There is a distinguished pair (U, V) with $a \in U \subseteq V \subseteq X - F$. Then $f_{U,V}(a) = r < \frac{1}{2}$ and $f_{U,V}|F = 1$, so $f_{U,V}(a) \notin \text{cl}\, f_{U,V}[F] = \{1\}$.

By Corollary VI.4.11, $e : X \to [0,1]^{|\mathcal{F}|}$ is an embedding. Since m is infinite, $|\mathcal{F}| \leq |\mathcal{B} \times \mathcal{B}| = m^2 = m$. ∙

A theorem that states that certain topological properties of a space X imply that X is metrizable is called a "metrization theorem." Typically the hypotheses of a metrization theorem involve that
1) X has "enough separation" and 2) X has a "sufficiently nice base." The following theorem is a simple example.

Corollary 3.18 ("Baby Metrization Theorem") A second countable Tychonoff space X is metrizable.

Proof By Theorem 3.17, $X \overset{\text{top}}{\subseteq} [0,1]^{\aleph_0}$. Since $[0,1]^{\aleph_0}$ is metrizable, so is X. ∙

In Corollary 3.18, X turns out to be metrizable and separable (since X is second countable). On the other hand $[0,1]^{\aleph_0}$ and all its subspaces are separable metrizable spaces. Thus, the corollary tells us that the separable metrizable spaces (topologically) are precisely the second countable Tychonoff spaces.

Exercises

E7. Prove that if X is a countable Tychonoff space, then there is a neighborhood base of clopen sets at each point. (Such a space X is sometimes called zero-dimensional.)

E8. Prove that in any space X, a countable union of cozero sets is a cozero set – or, equivalently, that a countable intersection of zero sets is a zero set.

E9. Prove that the following are equivalent in any Tychonoff space X:

a) every zero set is open
b) every G_δ set is open
c) for each $f \in C(X)$: if $f(p) = 0$ then there is a neighborhood N of p such that $f \mid N \equiv 0$

E10. Let $i : \mathbb{R} \to \mathbb{R}$ be the identity map and let

$$(i) = \{f \in C(\mathbb{R}) : f = gi \text{ for some } g \in C(\mathbb{R})\}.$$

(i) is called the ideal in $C(\mathbb{R})$ generated by the element i.

For those who know a bit of algebra: if we definite addition and multiplication of functions pointwise, then $C(\mathbb{R})$ (or, more generally, $C(X)$) is a commutative ring. The constant function $\mathbf{0}$ is the zero element in the ring; there is also a unit element, namely the constant function "1."

a) Prove that $(i) = \{f \in C(\mathbb{R}) : f(0) = 0 \text{ and the derivative } f'(0) \text{ exists}\}$.

b) Exhibit two functions f, g in $C(\mathbb{R})$ for which $fg \in (i)$ yet $f \notin (i)$ and $g \notin (i)$.

c) Let X be a Tychonoff space with more than one point. Prove that there are two functions $f, g \in C(X)$ such that $fg = \mathbf{0}$ on X yet neither f nor g is identically 0 on X.

Thus, there are functions $f, g \in C(X)$ for which $fg = \mathbf{0}$ although $f \neq \mathbf{0}$ and $g \neq \mathbf{0}$. In an algebra course, such elements f and g in the ring $C(X)$ are called "zero divisors."

d) Prove that there are exactly two functions $f \in C(\mathbb{R})$ for which $f^2 = f$. (*In* $C(X)$, the notation $f^2(x)$ *means* $f(x) \cdot f(x)$, *not* $f(f(x))$.)

e) Prove that there are exactly *c* functions f in $C(\mathbb{Q})$ for which $f^2 = f$.

An element in $C(X)$ that equals its own square is called an idempotent. Part d) shows that $C(\mathbb{R})$ and $C(\mathbb{Q})$ are not isomorphic rings since they have different numbers of idempotents. Is either $C(\mathbb{R})$ or $C(\mathbb{Q})$ isomorphic to $C(\mathbb{N})$?

One classic part of general topology is to explore the relationship between the space X and the rings $C(X)$ and $C^*(X)$. For example, if X is homeomorphic to Y, then $C(X)$ is isomorphic to $C(Y)$. This necessarily implies (why?) that $C^*(X)$ is isomorphic to $C^*(Y)$. The question "when does isomorphism *imply homeomorphism?" is more difficult. Another important area of study is how the maximal ideals of the ring $C(X)$ are related to the topology of X. The best introduction to this material is the classic book Rings of Continuous Functions (Gillman-Jerison).*

f) Let $D(\mathbb{R})$ be the set of differentiable functions $f : \mathbb{R} \to \mathbb{R}$. Are the rings $C(\mathbb{R})$ and $D(\mathbb{R})$ isomorphic? *Hint: An isomorphism between $C(\mathbb{R})$ and $D(\mathbb{R})$ preserves cube roots.*

E11. Suppose X is a connected Tychonoff space with more than one point. Prove $|X| \geq c$..

E12. Let X be a topological space. Suppose $f, g \in C(X)$ and that $Z(f)$ is a neighborhood of $Z(g)$ (that is, $Z(g) \subseteq \text{int}\, Z(f)$.)

a) Prove that f is a multiple of g in $C(X)$, that is, prove there is a function $h \in C(X)$ such that $f(x) = g(x)h(x)$ for all $x \in X$.

b) Give an example where $Z(f) \supseteq Z(g)$ but f is not a multiple of g in $C(X)$.

E13. Let X be a Tychonoff space with subspaces F and A, where F is closed and A is countable. Prove that if $F \cap A = \emptyset$, then A is disjoint from some zero set that contains F.

E14. A space X is called pseudocompact if every continuous $f : X \to \mathbb{R}$ is bounded, that is, if $C(X) = C^*(X)$ (s*ee Definition IV.8.7*). Consider the following condition (*) on a space X:

(*) Whenever $V_1 \supseteq V_2 \supseteq ... \supseteq V_n \supseteq ...$ is a decreasing sequence of nonempty open sets,

then $\bigcap_{n=1}^{\infty} \text{cl}\, V_n \neq \emptyset$.

a) Prove that if X satisfies (*), then X is pseudocompact.

b) Prove that if X is Tychonoff and pseudocompact, then X satisfies (*).

Note: For Tychonoff spaces, part b) gives an "internal" characterization of pseudocompactness – that is, a characterization that makes no explicit reference to $\mathbb{R}$.

.

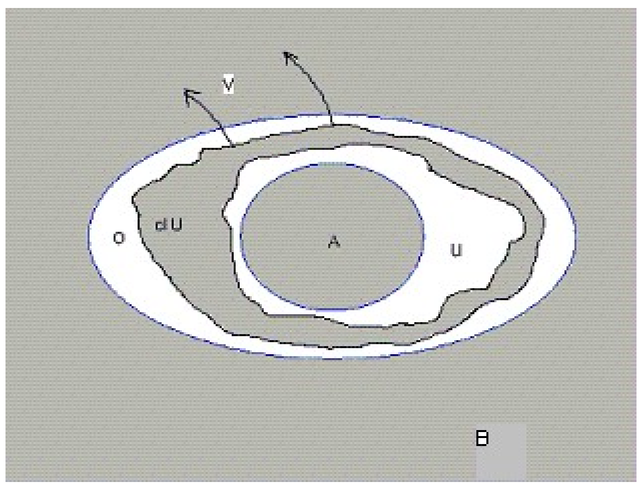
4. Normal and T_4-Spaces

We now return to a topic in progress: normal spaces (and T_4-spaces). Even though normal spaces are badly behaved in some ways, there are still some very important (and nontrivial) theorems that we can prove. One of these will give "$T_4 \Rightarrow T_{3\frac{1}{2}}$" as an immediate corollary.

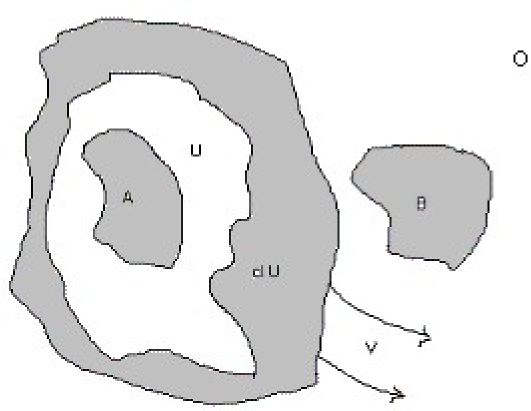
To begin, the following technical variation on the definition of normality is very useful.

Lemma 4.1 A space X is normal iff whenever A is closed, O is open and $A \subseteq O$, there exists an open set U with $A \subseteq U \subseteq \text{cl}\, U \subseteq O$.

Proof Suppose X is normal and that O is an open set containing the closed set A. Then A and $B = X - O$ are disjoint closed sets. By normality, there are disjoint open sets U and V with $A \subseteq U$ and $B \subseteq V$. Then $A \subseteq U \subseteq \text{cl}\, U \subseteq X - V \subseteq O$.



Conversely, suppose X satisfies the stated condition and that A, B are disjoint closed sets.



Then $A \subseteq O = X - B$, so there is an open set U with $A \subseteq U \subseteq \text{cl}\, U \subseteq X - B$. Let $V = X - \text{cl}\, U$. U and V are disjoint closed sets containing A and B respectively, so X is normal. $\bullet$

Theorem 4.2 a) A <u>closed</u> subspace of a normal (T_4) space is normal (T_4).
b) A continuous closed image of a normal (T_4) space normal (T_4).

Proof a) Suppose F is a closed subspace of a normal space X and let A and B be disjoint closed sets in F. Then A, B are also closed in X so we can find disjoint open sets U' and V' in X containing A and B respectively. Then $U = U' \cap F$ and $V = V' \cap F$ are disjoint open sets in F that contain A and B, so F is normal.

b) Suppose X is normal and that $f : X \to Y$ is continuous, closed and onto. If A and B are disjoint closed sets in Y, then $f^{-1}[A]$ and $f^{-1}[B]$ are disjoint closed sets in X. Pick U' and V' disjoint open sets in X with $f^{-1}[A] \subseteq U'$ and $f^{-1}[B] \subseteq V'$. Then $U = Y - f[X - U']$ and $V = Y - f[X - V']$ are open sets in Y.

If $y \in U$, then $y \notin f[X - U']$. Since f is onto, $y = f(x)$ for some $x \in U' \subseteq X - V'$. Therefore $y \in f[X - V']$ so $y \notin V$. Hence $U \cap V = \emptyset$.

If $y \in A$, then $f^{-1}[\{y\}] \subseteq f^{-1}[A] \subseteq U'$, so $f^{-1}[\{y\}] \cap (X - U') = \emptyset$. Therefore $y \notin f[X - U']$ so $y \in Y - f[X - U'] = U$. Therefore $A \subseteq U$ and, similarly, $B \subseteq V$ so Y is normal.

Since the T_1 property is hereditary and is preserved by closed onto maps, the statements in a) and b) hold for T_4 as well as normality. •

The next theorem gives us more examples of normal (and T_4) spaces.

Theorem 4.3 Every regular Lindelöf space X is normal (and therefore every Lindelöf T_3-space is T_4).

Proof Suppose A and B are disjoint closed sets in X. For each $x \in A$, use regularity to pick an open set U_x such that $x \in U_x \subseteq \operatorname{cl} U_x \subseteq X - B$. Since the Lindelöf property is hereditary on <u>closed</u> subsets, a countable number of the U_x's cover A: relabel these as $U_1, U_2, ..., U_n, ...$. For each n, we have $\operatorname{cl} U_n \cap B = \emptyset$. Similarly, choose a sequence of open sets $V_1, V_2, ..., V_n, ...$ covering B such that $\operatorname{cl} V_n \cap A = \emptyset$ for each n.

We have that $\bigcup_{n=1}^{\infty} U_n \supseteq A$ and $\bigcup_{n=1}^{\infty} V_n \supseteq B$, but these unions may not be disjoint. So we define

$$\begin{array}{ll}
U_1^* = U_1 - \operatorname{cl} V_1 & V_1^* = V_1 - \operatorname{cl} U_1 \\
U_2^* = U_2 - (\operatorname{cl} V_1 \cup \operatorname{cl} V_2) & V_2^* = V_2 - (\operatorname{cl} U_1 \cup \operatorname{cl} U_2) \\
\quad\vdots & \quad\vdots \\
U_n^* = U_n - (\operatorname{cl} V_1 \cup \operatorname{cl} V_2 \cup ... \cup \operatorname{cl} V_n) & V_n^* = V_n - (\operatorname{cl} U_1 \cup \operatorname{cl} U_2 \cup ... \cup \operatorname{cl} U_n) \\
\quad\vdots & \quad\vdots
\end{array}$$

Let $U = \bigcup_{n=1}^{\infty} U_n^*$ and $V = \bigcup_{n=1}^{\infty} V_n^*$.

If $x \in A$, then $x \notin \operatorname{cl} V_n$ for all n. But $x \in U_k$ for some k, so $x \in U_k^* \subseteq U$. Therefore $A \subseteq U$ and, similarly, $B \subseteq V$.

To complete the proof, we show that $U \cap V = \emptyset$. Suppose $x \in U$.

Then $x \in U_k^*$ for some k, so $x \notin \text{cl}\, V_1 \cup \text{cl}\, V_2 \cup ... \cup \text{cl}\, V_k$,
so $x \notin V_1 \cup V_2 \cup ... \cup V_k$
so $x \notin V_1^* \cup V_2^* \cup ... \cup V_k^*$
so $x \notin V_n^*$ for any $n \leq k$.

Since $x \in U_k^*$, then $x \in U_k$. So, if $n > k$, then $x \notin V_n^* = V_n - (\text{cl}\, U_1 \cup ... \cup \text{cl}\, U_k \cup ... \cup \text{cl}\, U_n)$

So $x \notin V_n^*$ for all n, so $x \notin V$ and therefore $U \cap V = \emptyset$. •

Example 4.4 The Sorgenfrey line S is regular because the sets $[a, b)$ form a base of closed neighborhoods at each point a. We proved in Example VI.3.2 that S is Lindelöf, so S is normal. Since S is T_1, we have that S is T_4.

5. Urysohn's Lemma and Tietze's Extension Theorem

We now turn our attention to the issue of "$T_4 \Rightarrow T_{3\frac{1}{2}}$". This is hard to prove because to show that a space X is $T_{3\frac{1}{2}}$, we need to prove that certain continuous functions exist; but the hypothesis "T_4" gives us no continuous functions to work with. As far as we know at this point, there could even be T_4 spaces on which every continuous real-valued function is constant! If T_4-spaces are going to have a rich supply of continuous real-valued functions, we will have to show that these functions can be "built from scratch" in a T_4-space. This will lead us to two of the most well-known classical theorems of general topology.

We begin with the following technical lemma. It gives a way to use a certain collection of open sets $\{U_r : r \in Q\}$ to construct a function $f \in C(X)$. The idea in the proof is quite straightforward, but I attribute its elegant presentation (and that of Urysohn's Lemma which follows) primarily to Leonard Gillman and Meyer Jerison.

Lemma 5.1 Suppose X is any topological space and let Q be any dense subset of $\mathbb{R}$. Suppose open sets $U_r \subseteq X$ have been defined, one for each $r \in Q$, in such a way that:

i) $X = \bigcup_{r \in Q} U_r$ and $\bigcap_{r \in Q} U_r = \emptyset$
ii) if $r, s \in Q$ and $r < s$, then $\text{cl}\, U_r \subseteq U_s$.

For $x \in X$, define $f(x) = \inf\{r \in Q : x \in U_r\}$. Then $f : X \to \mathbb{R}$ is continuous.

We will use this Lemma only once, with $Q = \mathbb{Q}$. So if you like, there is no harm in assuming that $Q = \mathbb{Q}$ in the proof.

Proof Suppose $x \in X$. By i) we know that $x \in U_r$ for some r, so $\{r \in Q : x \in U_r\} \neq \emptyset$. And by ii), we know that $x \notin U_s$ for some s. For that s : if $x \in U_r$, then (by ii) $s \leq r$, so s is a lower bound for $\{r \in Q : x \in U_r\}$. Therefore $\{r \in Q : x \in U_r\}$ has a greatest lower bound, so the definition of f makes sense: $f(x) = \inf\{r \in Q : x \in U_r\} \in \mathbb{R}$.

From the definition of f, we get that for $r, s \in Q$,

a) if $x \in \operatorname{cl} U_r$, then $x \in U_s$ for all $s > r$ so $f(x) \leq r$
b) if $f(x) < s$, then $x \in U_s$.

We want to prove f is continuous at each point $a \in X$. Since Q is dense in $\mathbb{R}$,

$$\{[r, s] : r, s \in Q \text{ and } r < f(a) < s\}$$

is a neighborhood base at $f(a)$ in $\mathbb{R}$. Therefore it is sufficient to show that whenever $r < f(a) < s$, then there is a neighborhood U of a such that $f[U] \subseteq [r, s]$.

Since $f(a) < s$, we have $a \in U_s$, and $f(a) > r$ gives us that $a \notin \operatorname{cl} U_r$. Therefore $U = U_s - \operatorname{cl} U_r$ is an open neighborhood of a. If $z \in U$, then $z \in U_s \subseteq \operatorname{cl} U_s$, so $f(z) \leq s$; and $z \notin \operatorname{cl} U_r$, so $z \notin U_r$ and $f(z) \geq r$. Therefore $f[U] \subseteq [r, s]$. ●

Our first major theorem about normal spaces is still traditionally referred to as a "lemma" because it was a lemma in the paper where it originally appeared. Its author, Paul Urysohn, died at age 26, on the morning of 17 August 1924, while swimming off the coast of Brittany.

Theorem 5.2 (Urysohn's Lemma) A space X is normal iff whenever A, B are disjoint closed sets in X, there exists a function $f \in C(X)$ with $f|A = 0$ and $f|B = 1$. (*When such an f exists, we say that A and B are completely separated.*)

Note: Notice that if A and B happen to be disjoint zero sets, say $A = Z(g)$ and $B = Z(h)$, then the conclusion of the theorem is true in any space, without assuming normality: just let $f(x) = \frac{g^2(x)}{g^2(x) + h^2(x)}$. Then f is continuous, $f|A = 0$ and $f|B = 1$.

The conclusion of Urysohn's Lemma only says that $A \subseteq f^{-1}(0)$ and $B \subseteq f^{-1}(1)$: equality might not be true. In fact, if $A = f^{-1}(0)$ and $B = f^{-1}(1)$, then A and B were zero sets in the beginning, and the hypothesis of normality would have been unnecessary.

This shows again that zero sets are very special closed sets: in any space, disjoint zero sets are completely separated. Put another way: given Urysohn's Lemma, we can conclude that every nonnormal space must contain a closed set that is not a zero set.

Proof The proof of Urysohn's Lemma in one direction is almost trivial. If such a function f exists, then $U = \{x : f(x) < \frac{1}{2}\}$ and $V = \{x : f(x) > \frac{1}{2}\}$ are disjoint open sets (in fact, cozero sets) containing A and B respectively. It is the other half of Urysohn's Lemma for which Urysohn deserves credit.

Let A and B be disjoint closed sets in a normal space X. We will define sets open sets U_r ($r \in \mathbb{Q}$) in X in such a way that Lemma 5.1 applies. To start, let $U_r = \emptyset$ for $r < 0$ and $U_r = X$ for $r > 1$.

Enumerate the remaining rationals in $\mathbb{Q} \cap [0,1]$ as $r_1, r_2, ..., r_n, ...,$ beginning the list with $r_1 = 1$ and $r_2 = 0$. We begin by defining $U_{r_1} = U_1 = X - B$. Then use normality to define $U_{r_2}(= U_0)$: since $A \subseteq U_{r_1} = X - B$, we can pick U_{r_2} so that

$$A \subseteq U_{r_2} \subseteq \mathrm{cl}\, U_{r_2} \subseteq U_{r_1} = X - B$$

Then $0 = r_2 < r_3 < r_1 = 1$, and we use normality to pick an open set U_{r_3} so that

$$A \subseteq U_{r_2} \subseteq \mathrm{cl}\, U_{r_2} \subseteq U_{r_3} \subseteq \mathrm{cl}\, U_{r_3} \subseteq U_{r_1} = X - B$$

We continue by induction. Suppose $n \geq 3$ and that we have already defined open sets $U_{r_1}, U_{r_2}, ..., U_{r_n}$ in such a way that whenever $r_i < r_j < r_k$ $(i, j, k \leq n)$, then

$$\mathrm{cl}\, U_{r_i} \subseteq U_{r_j} \subseteq \mathrm{cl}\, U_{r_j} \subseteq U_{r_k} \quad (*)$$

We need to define $U_{r_{n+1}}$ so that $(*)$ holds for $i, j, k \leq n + 1$.

Since $r_1 = 1$ and $r_2 = 0$, and $r_{n+1} \in (0,1)$, it makes sense to define

$r_k =$ the largest among $r_1, r_2, ..., r_n$ that is smaller than r_{n+1}, and
$r_l =$ the smallest among $r_1, r_2, ..., r_n$ that is larger than r_{n+1}.

By the induction hypothesis, we already have $\mathrm{cl}\, U_{r_k} \subseteq U_{r_l}$. Then use normality to pick an open set $U_{r_{n+1}}$ so that

$$\mathrm{cl}\, U_{r_k} \subseteq U_{r_{n+1}} \subseteq \mathrm{cl}\, U_{r_{n+1}} \subseteq U_{r_l}.$$

The U_r's defined in this way satisfy the conditions of Lemma 5.1, so the function $f : X \to \mathbb{R}$ defined by $f(x) = \inf\{r \in \mathbb{Q} : x \in U_r\}$ is continuous. If $x \in A$, then $x \in U_{r_2} = U_0$ and $x \notin U_r$ if $r < 0$, so $f(x) = 0$. If $x \in B$ then $x \notin U_1$, but $x \in U_r = X$ for $r > 1$, so $f(x) = 1$. ∙

Once we have the function f we can replace it, if we like, by $g = (0 \vee f) \wedge 1$ so that A and B are completely separated by a function $g \in C^*(X)$. It is also clear that we can modify g further to get an $h \in C^*(X)$ for which $h|A = a$ and $h|B = b$ where a and b are any two real numbers.

With Urysohn's Lemma, the proof of the following corollary is obvious.

Corollary 5.3 $T_4 \Rightarrow T_{3\frac{1}{2}}$.

There is another famous characterization of normal spaces in terms of $C(X)$. It is a result about "extending" continuous real-valued functions defined on closed subspaces.

We begin with the following two lemmas. Lemma 5.4, called the "Weierstrass M-Test" is a slight generalization of a theorem with the same name in advanced calculus. It can be useful in "piecing together" infinitely many real-valued continuous functions to get a new one. Lemma 5.5 will be used in the proof of Tietze's Extension Theorem (Theorem 5.6).

Lemma 5.4 (Weierstrass M-Test) Let X be a topological space. Suppose $f_n : X \to \mathbb{R}$ is continuous for each $n \in \mathbb{N}$ and that $|f_n(x)| \le M_n$ for all $x \in X$. If $\sum_{n=1}^{\infty} M_n < \infty$, then $f(x) = \sum_{n=1}^{\infty} f_n(x)$ converges (absolutely) for all x and $f : X \to \mathbb{R}$ is continuous.

Proof For each x, $\sum_{n=1}^{\infty} |f_n(x)| \le \sum_{n=1}^{\infty} M_n < \infty$, so $\sum_{n=1}^{\infty} f_n(x)$ converges (absolutely) by the Comparison Test.

Suppose $a \in X$ and $\epsilon > 0$. Choose N so that $\sum_{n=N+1}^{\infty} M_n < \frac{\epsilon}{4}$. Each f_n is continuous, so for $n = 1, ..., N$ we can pick a neighborhood U_n of a such that for $x \in U_n$, $|f_n(x) - f_n(a)| < \frac{\epsilon}{2N}$. Then $U = \bigcap_{n=1}^{N} U_n$ is a neighborhood of a, and for $x \in U$ we get $|f(x) - f(a)|$

$= |\sum_{n=1}^{N} (f_n(x) - f_n(a)) + \sum_{n=N+1}^{\infty} (f_n(x) - f_n(a))| \le \sum_{n=1}^{N} |f_n(x) - f_n(a)| + \sum_{n=N+1}^{\infty} |f_n(x) - f_n(a)|$

$\le \sum_{n=1}^{N} |f_n(x) - f_n(a)| + \sum_{n=N+1}^{\infty} |f_n(x)| + |f_n(a)| < N \cdot \frac{\epsilon}{2N} + \sum_{n=N+1}^{\infty} 2M_n < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$

Therefore f is continuous at a. •

Lemma 5.5 Let A be a closed set in a normal space X and let a be a positive real number. Suppose $h : A \to [-r, r]$ is continuous. Then there exists a continuous $\phi : X \to [-\frac{r}{3}, \frac{r}{3}]$ such that $|h(x) - \phi(x)| \le \frac{2r}{3}$ for each $x \in A$.

Proof Let $A_1 = \{x \in A : h(x) \le -\frac{r}{3}\}$ and $B_1 = \{x \in A : h(x) \ge \frac{r}{3}\}$. A is closed, and A_1 and B_1 are disjoint closed sets in A, so A_1 and B_1 are closed in X. By Urysohn's Lemma, there exists a continuous function $\phi : X \to [-\frac{r}{3}, \frac{r}{3}]$ such that $\phi|A_1 = -\frac{r}{3}$ and $\phi|B_1 = \frac{r}{3}$.

If $x \in A_1$, then $-r \le h(x) \le -\frac{r}{3}$ and $\phi(x) = -\frac{r}{3}$, so $|h(x) - \phi(x)| \le |-r - (-\frac{r}{3})| = \frac{2r}{3}$; and similarly if $x \in B_1$, $|h(x) - \phi(x)| \le \frac{2r}{3}$. If $x \in A - (A_1 \cup B_1)$, then $h(x)$ and $\phi(x)$ are both in $[-\frac{r}{3}, \frac{r}{3}]$ so $|h(x) - \phi(x)| \le \frac{2r}{3}$. •

Theorem 5.6 (Tietze's Extension Theorem) A space X is normal iff whenever A is a closed set in X and $f \in C(A)$, then there exists a function $g \in C(X)$ such that $g|A = f$.

> Note: if A is a closed subset of $X = \mathbb{R}$, then it is quite easy to prove the theorem. In that case, $\mathbb{R} - A$ is open and can be written as a countable union of disjoint open intervals I, where each $I = (a, b)$ or $(-\infty, b)$ or (a, ∞) (*see Theorem II.3.4*). For each of these intervals I, the endpoints are in A, where f is already defined. If $I = (a, b)$ then extend the definition of f over I by using a straight line segment to join $(a, f(a))$ and $(b, f(b))$ on the graph of f. If $I = (a, \infty)$. then extend the graph of f over I using a horizontal right ray at height $f(a)$; if $I = (-\infty, b)$, then extend the graph of f over I using a horizontal left ray at height $f(b)$.

As with Urysohn's Lemma, half of the proof is easy. The significant part of theorem is proving the existence of the extension g when X is normal.

Proof $(\Leftarrow)$ Suppose A and B are disjoint closed sets in X. A and B are clopen in the subspace $A \cup B$ so the function $f : A \cup B \to [0,1]$ defined by $f|A = 0$ and $f|B = 1$ is continuous. Since $A \cup B$ is closed in X, there is a function $g \in C(X)$ such that $g|(A \cup B) = f$. Then $U = \{x : g(x) < \frac{1}{2}\}$ and $V = \{x : g(x) > \frac{1}{2}\}$ are disjoint open sets (cozero sets, in fact) that contain A and B respectively. Therefore X is normal.

$(\Rightarrow)$ The idea is to find a sequence of functions $g_i \in C(X)$ such that $|f(x) - \sum_{i=1}^{n} g_i(x)| \to 0$ as $n \to \infty$ for each $x \in A$ (where f is defined). The sums $\sum_{i=1}^{n} g_i(x)$ are defined on all of X and as $n \to \infty$ we can think of them as giving better and better approximations to the extension g that we want. Then we can let $g(x) = \lim_{n\to\infty} \sum_{i=1}^{n} g_i(x) = \sum_{i=1}^{\infty} g_i(x)$. The details follow. We proceed in three steps, but the heart of the argument is in Case I.

Case I Suppose f is continuous and that $f : A \to [-1,1]$. We claim there is a continuous function $g : X \to [-1,1]$ with $g|A = f$.

Using Lemma 5.5 (with $h = f$, $r = 1$) we get a function $g_1 = \phi : X \to [-\frac{1}{3}, \frac{1}{3}]$ such that for $x \in A$, $|f(x) - g_1(x)| \leq \frac{2}{3}$. Therefore $f - g_1 : A \to [-\frac{2}{3}, \frac{2}{3}]$.

Using Lemma 5.5 again (with $h = f - g_1, r = \frac{2}{3}$), we get a function $g_2 : X \to [-\frac{2}{9}, \frac{2}{9}]$ such that for $x \in A$, $|f(x) - g_1(x) - g_2(x)| \leq \frac{4}{9} = (\frac{2}{3})^2$. So $f - (g_1 + g_2) : A \to [-\frac{4}{9}, \frac{4}{9}]$

Using Lemma 5.5 again (with $h = f - g_1 - g_2, r = \frac{4}{9}$), we get a function $g_3 : X \to [-\frac{4}{27}, \frac{4}{27}]$ such that for $x \in A$, $|f(x) - g_1(x) - g_2(x) - g_3(x)| \leq \frac{8}{27} = (\frac{2}{3})^3$.
So $f - (g_1 + g_2 + g_3) : A \to [-\frac{8}{27}, \frac{8}{27}]$.

We continue, using induction, to find for each i a continuous function $g_i : X \to [-\frac{2^{i-1}}{3^i}, \frac{2^{i-1}}{3^i}]$ such that $|f(x) - \sum_{i=1}^{n} g_i(x)| \leq (2/3)^n$ for $x \in A$.

Since $\sum_{i=1}^{\infty} |g_i(x)| \leq \sum_{i=1}^{\infty} \frac{2^{i-1}}{3^i} < \infty$, the series $g(x) = \sum_{i=1}^{\infty} f_i(x)$ converges (absolutely) for every $x \in X$, and g is continuous by the Weierstrass M-Test. Since $|g(x)| = |\sum_{i=1}^{\infty} g_i(x)|$ $\leq \sum_{i=1}^{\infty} |g_i(x)| \leq \sum_{i=1}^{\infty} \frac{2^{i-1}}{3^i} = 1$, we have $g : X \to [-1,1]$.
Finally, for $x \in A$, $|f(x) - g(x)| = \lim_{n\to\infty} |f(x) - \sum_{i=1}^{n} g_i(x)| \leq \lim_{n\to\infty} (\frac{2}{3})^n = 0$, so $g|A = f$ and the proof for Step I is complete.

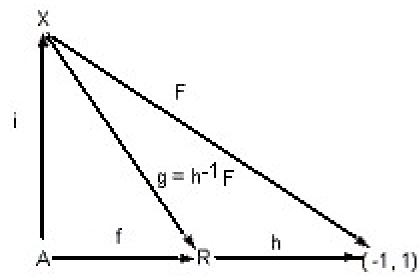
Case II Suppose $f : A \to (-1,1)$ is continuous. We claim there is a continuous function $g : X \to (-1,1)$ with $g|A = f$.

Since $f : A \to (-1,1) \subseteq [-1,1]$, we can apply Case I to find a continuous function $F : X \to [-1,1]$ with $F|A = f$. To get g, we merely make a slight modification to F to get a g that still extends f but where g has all its values in $(-1,1)$.

Let $B = \{x \in X : F(x) = \pm 1\}$. A and B are disjoint closed sets in X, so by Urysohn's Lemma there is a continuous $h : X \to [-1, 1]$ such that $h|B = 0$ and $h|A = 1$. If we let $g(x) = F(x)h(x)$, then $g : X \to (-1, 1)$ and $g|A = f$, completing the proof of Case II.

Case III (the full theorem) Suppose $f : A \to \mathbb{R}$ is continuous. We claim there is a continuous function $g : X \to \mathbb{R}$ with $g|A = f$.

Let $h : \mathbb{R} \to (-1, 1)$ be a homeomorphism. Then $h \circ f : A \to (-1, 1)$ and, by Step II, there is a continuous $F : X \to (-1, 1)$ with $F|A = h \circ f$.



Let $g = h^{-1} \circ F : X \to \mathbb{R}$. Then for $x \in A$ we have $g(x) = h^{-1}(F(x))$
$= h^{-1}((h \circ f))(x) = f(x)$. $\bullet$

It is easy to see that $C^*(X)$ can replace $C(X)$ in the statement of Tietze's Extension Theorem.

Example 5.7 We now know enough about normality to see some of its bad behavior. The Sorgenfrey line S is normal (*Example 4.4*) but the Sorgenfrey plane $S \times S$ is not normal.

To see this, let $D = \mathbb{Q} \times \mathbb{Q}$, a countable dense set in $S \times S$. Every continuous real-valued function on $S \times S$ is completely determined by its values on D. (*See Theorem II.5.12. The theorem is stated for the case of functions defined on a pseudometric space, but the proof is written in a way that applies just as well to functions with any space X as domain.*) Therefore the mapping $C(S \times S) \to C(D)$ given by $f \mapsto f|D$ is one-to-one, so $|C(S \times S)| \leq |C(D) \leq |\mathbb{R}^D| = c^{\aleph_0} = c$.

$A = \{(x, y) \in S \times S : x + y = 1\}$ is closed and discrete in the subspace topology, so every function defined on A is continuous, that is, $\mathbb{R}^A = C(A)$ and so $|C(A)| = c^c = 2^c$. If $S \times S$ were normal, then each $f \in C(A)$ could be extended (by Tietze's Theorem) to a continuous function in $C(S \times S)$. This would mean that $|C(S \times S)| \geq |\mathbb{R}^A| = c^c = 2^c > c$. which is false. Therefore normality is not even finitely productive.

The comments following the statement of Urysohn's Lemma imply that $S \times S$ must contain closed sets that are not zero sets.

A completely similar argument "counting continuous real-valued functions" shows that the Moore plane Γ *(Example III.5.6) is not normal: use that* Γ *is separable and the* x*-axis in* Γ *is an uncountable closed discrete subspace.*

Questions about the normality of products are difficult. For example, it was an open question for a long time whether the product of a normal space X with such a nice, well-behaved space as $[0,1]$ must be normal. In the 1950's, Dowker proved that $X \times [0,1]$ is normal iff X is normal and "countably paracompact."

However, this result was unsatisfying – because no one knew whether a normal space was automatically "countably paracompact." In the 1960's, Mary Ellen Rudin constructed a normal space X which was not countably paracompact. But this example was still unsatisfying because the construction assumed the existence of a space called a "Souslin line" – and whether a Souslin line exists cannot be decided in the ZFC set theory! In other words, the space X she constructed required adding a new axiom to ZFC.

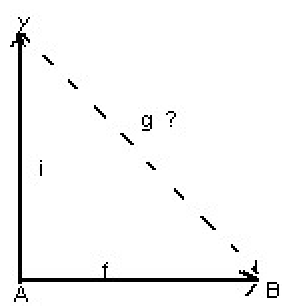
Things were finally settled in 1971 when Mary Ellen Rudin constructed a "real" example of a normal space X whose product with $[0,1]$ is not normal. By "real," we mean that X was constructed in ZFC, with no additional set theoretic assumptions. Among other things, this complicated example made use of the box topology on a product.

Example 5.8 The Sorgenfrey line S is T_4, so S is $T_{3\frac{1}{2}}$ and therefore the Sorgenfrey plane $S \times S$ is also $T_{3\frac{1}{2}}$. So $S \times S$ is an example showing that $T_{3\frac{1}{2}}$ <u>does not imply</u> T_4.

Extension theorems such as Tietze's are an important topic in mathematics. In general, an "extension theorem" has the following form:

$A \subseteq X$ and $f : A \to B$, then there is a function $g : X \to B$ such that $g|A = f$.

For example, in algebra one might ask: if A *is a subgroup of* X *and* $f : A \to B$ *is an isomorphism, can* f *be extended to a homomorphism* $g : X \to B$ *?*



If we let $i : A \to X$ *be the injection* $i(a) = a$*, then the condition "*$g|A = f$*" can be rewritten as* $g \circ i = f$*. In the language of algebra, we are asking whether there is a suitable function* g *which "makes the diagram commute."*

Specific extension theorems impose conditions on A and X, and usually we want g to share some property of f such as continuity. Here are some illustrations, without further details.

1) <u>Extension theorems that generalize of Tietze's Theorem:</u> by putting stronger hypotheses on X, we can relax the hypotheses on B.

Suppose A is closed in X and $f : A \to B$ is continuous.

$$\text{If} \quad \begin{cases} X \text{ is normal} & \text{and } B = \mathbb{R} \quad \text{(Tietze's Theorem)} \\ X \text{ is normal} & \text{and } B = \mathbb{R}^n \\ X \text{ is collectionwise normal**} & \text{and } B \text{ is a separable Banach space*} \\ X \text{ is paracompact**} & \text{and } B \text{ is a Banach space*} \end{cases}$$

then f has a continuous extension $g : X \to B$.

The statement that $\mathbb{R}^n$ can replace $\mathbb{R}$ in Tietze's Theorem is easy to prove:

If X is normal and $f : A \to \mathbb{R}^n$ is continuous, write $f(x) = (f_1(x), f_2(x), ..., f_n(x))$ where each $f_i : A \to \mathbb{R}$. By Tietze's Theorem, there exists for each i a continuous extension $g_i : X \to \mathbb{R}$ with $g_i|A = f_i$. If we let $g(x) = (g_1(x), ..., g_n(x))$, then $g : X \to \mathbb{R}^n$ and $g|A = f$. In other words, we separately extend the coordinate functions in order to extend f. And in this example, n could even be an infinite cardinal.

* *A normed linear space is a vector space V with a norm $|v|$ (= "absolute value") that defines the "length" of each vector. Of course, a norm must satisfy certain axioms – for example, $|v_1 + v_2| \le |v_1| + |v_2|$. These properties guarantee that a norm can be used to define a metric: $d(v_1, v_2) = |v_1 - v_2|$. A <u>Banach space</u> is a normed linear space which is complete in this metric d.*

For example, $\mathbb{R}^n$: the usual norm $|(x_1, x_2, ..., x_n)| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$ produces the usual metric, which is complete. So $\mathbb{R}^n$ is a separable Banach space.

** *Roughly, a "collectionwise normal" space is one in which certain <u>infinite</u> collections of disjoint closed sets can be enclosed in disjoint open sets. We will not give definitions for "collectionwise normal" (or the stronger condition, "paracompactness") here, but is true that*

$$\begin{cases} \textit{metric} \\ \quad \textit{or} \\ \textit{compact } T_2 \end{cases} \Rightarrow \quad \textit{paracompact} \quad \Rightarrow \textit{collectionwise normal} \quad \Rightarrow \quad \textit{normal}$$

Therefore, in the theorems cited above, a continuous map f defined on a closed subset of a metric space (or, compact T_2 space) and valued in a Banach space B can be continuously extended a function $g : X \to B$.

2) The Hahn-Banach Theorem is another example, taken from functional analysis, of an extension theorem. A special case states:

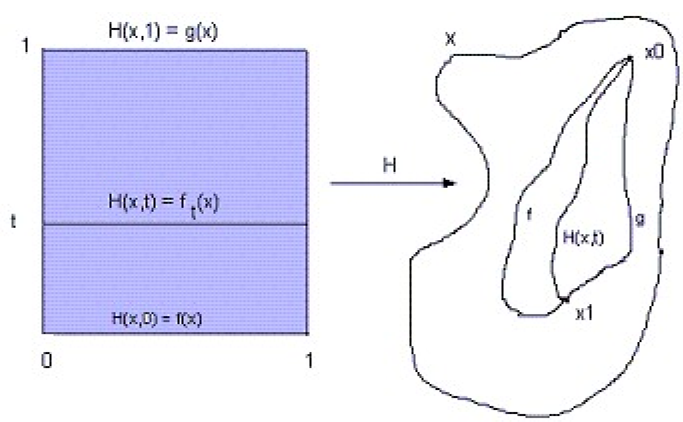
> Suppose M is a linear subspace of a real normed linear space X and that $f : M \to \mathbb{R}$ is linear and satisfies $f(x) \leq ||x||$ for all $x \in M$. There is a linear $F : X \to \mathbb{R}$ such that $F|M = f$ and $F(x) \leq ||x||$ for all $x \in X$.

3) Homotopy is usually not discussed in terms of extension theorems, but extensions are really at the heart of the idea.

> Let $f, g : [0,1] \to X$ be continuous and suppose that $f(0) = g(0) = x_0$ and $f(1) = g(1) = x_1$. Then f and g are paths in X that start at x_0 and end at x_1. Let B be the boundary of the square $[0,1]^2 \subseteq \mathbb{R}^2$ and define $F : B \to X$ by
>
> $$F(x,0) = f(x) \qquad F(x,1) = g(x)$$
> $$F(0,t) = x_0 \qquad F(1,t) = x_1$$

Thus F agrees with f on the bottom edge of B and with g on the top edge. F is constant $(= x_0)$ on the left edge of B and constant $(= x_1)$on the right edge of B. We ask whether F can be extended to a continuous map defined on the whole square, $H : [0,1]^2 \to X$.

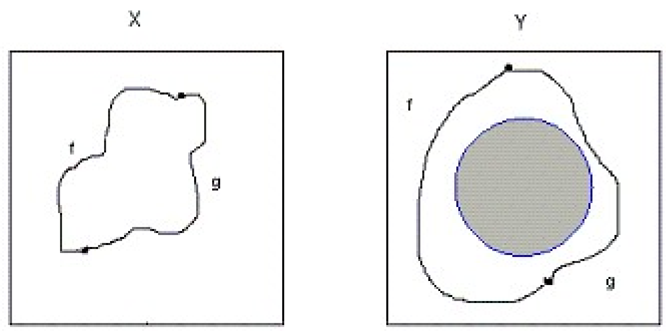
If such an extension H does exist, then we have



For each $t \in [0,1]$, restrict H to the line segment at height t to define $f_t(x) = H(x,t)$. Then for each $t \in [0,1]$, f_t is also a path in X from x_0 to x_1. As t varies from 0 to 1, we can think of the f_t's as a family of paths in X that continuously deform $f_0 = f$ into $f_1 = g$.

The continuous extension H (if it exists) is called a homotopy between f and g with fixed endpoints, and we say that the paths f and g are homotopic with fixed endpoints.

In the space X on the left, below, it seems intuitively clear that f can be continuously deformed (with endpoints held fixed) into g – in other words, that H exists.



However in the space Y pictured on the right, f and g together form a loop that surrounds a "hole" in Y, and it seems intuitively clear that the path f cannot be continuously deformed into the path g inside the space Y – that is, the extension H does not exist.

In some sense, homotopy can be used to detect the presence of certain "holes" in a space, and is one important part of algebraic topology.

The next theorem shows us where compact Hausdorff spaces stand in the discussion of separation properties.

Theorem 5.9 A compact T_2 space X is T_4.

Proof X is Lindelöf, and a regular Lindelöf space is normal (Theorem 4.3). Therefore it is sufficient to show that X is regular. Suppose F is a closed set in X and $x \notin F$. For each $y \in F$ we can pick disjoint open sets U_y and V_y with $x \in U_y$ and $y \in V_y$. F is compact so a finite number of the V_y's cover F – say $V_{y_1}, V_{y_2}, \ldots, V_{y_n}$. Then $x \in \bigcap_{i=1}^n U_{y_i} = U$, $F \subseteq \bigcup_{i=1}^n V_{y_i} = V$, and U, V are disjoint open sets. ●

Therefore, our results line up as:

$$(*) \quad \text{compact metric} \Rightarrow \begin{cases} \text{compact } T_2 \\ \underline{\qquad} \\ \text{metric} \end{cases} \Rightarrow T_4 \Rightarrow T_{3\frac{1}{2}} \Rightarrow T_3 \Rightarrow T_2 \Rightarrow T_1 \Rightarrow T_0$$

In particular, <u>Urysohn's Lemma and Tietze's Extension Theorem hold in metric spaces and in compact T_2 spaces</u>.

Notice that

i) the space $(0,1)$ is T_4 but not compact T_2

ii) the Sorgenfrey line S is T_4 (see example 5.4) but not metrizable. (*If S were metrizable, then $S \times S$ would be metrizable and therefore T_4 – which is false : see Example 5.7*).

iii) $[0,1]^c$ is compact T_2 (*assuming, for now, the Tychonoff Product Theorem VI.3.10*) but not metrizable (*why?*)

iv) $(0,1)$ is metrizable but not compact.

Combining these observations with earlier examples, we see that none of the implications in (*) is reversible.

Example 5.10 (*See Example 5.7*) The Sorgenfrey plane $S \times S$ is $T_{3\frac{1}{2}}$, so $S \times S$ can be embedded in a cube $[0,1]^m$ and $[0,1]^m$ is compact T_2 (*assuming the Tychonoff Product Theorem*). Since $S \times S$ is not normal, we see now that a normal space can have nonnormal subspaces. This example, admittedly, is not terribly satisfying since we can't visualize how $S \times S$ "sits" inside $[0,1]^m$. In Chapter VIII (Example 8.10), we will look at an example of a T_4 space in which it's easy to "see" why a certain subspace isn't normal.

6. Some Metrization Results

Now we have enough information to completely characterize separable metric spaces topologically.

Theorem 6.1 (Urysohn's Metrization Theorem) A second countable T_3-space is metrizable.
Note: We proved a similar metrization theorem in Corollary 3.18, but there the separation hypothesis was $T_{3\frac{1}{2}}$ rather than T_3.

Proof X is second countable so X is Lindelöf, and Theorem 4.3 tells us that a Lindelöf T_3-space is T_4. Therefore X is $T_{3\frac{1}{2}}$. So by Corollary 3.18, X is metrizable. ●

Because a separable metrizable space <u>is</u> second countable and T_3, we have a complete characterization: <u>X is a separable metrizable space iff X is a second countable T_3-space</u>. So, with hindsight, we now see that the hypothesis "$T_{3\frac{1}{2}}$" in Corollary 3.18 was unnecessarily strong. In fact, we see that T_3 and $T_{3\frac{1}{2}}$ are equivalent in a space that is second countable.

Further developments in metrization theory hinged on work of Arthur H. Stone in the late 1940's – in particular, his result that metric spaces have a property called "paracompactness." This led quickly to a complete characterization of metrizable spaces that came roughly a quarter century after Urysohn's work. We state this characterization here without a proof.

A family of sets $\mathcal{B}$ in $(X, \mathcal{T})$ is called locally finite if each point $x \in X$ has a neighborhood N that has nonempty intersection with only finitely many sets in $\mathcal{B}$. The family $\mathcal{B}$ is called σ-locally finite if we can write $\mathcal{B} = \bigcup_{n \in \mathbb{N}} \mathcal{B}_n$ where each subfamily $\mathcal{B}_n$ is locally finite.

Theorem 6.2 (The Bing-Smirnov-Nagata Metrization Theorem) $(X, \mathcal{T})$ is metrizable iff X is T_3 and has a σ-locally finite base $\mathcal{B}$.

Note: If X is second countable, a countable base $\mathcal{B} = \{O_1, O_2, ..., O_n, ...\}$ is σ-locally finite – because we can write $\mathcal{B} = \bigcup \mathcal{B}_n$, where $\mathcal{B}_n = \{O_n\}$. Therefore this Metrization Theorem includes Urysohn's Metrization Theorem as a special case.

The Bing-Smirnov-Nagata Theorem has the typical form of most metrization theorems: X is metrizable iff "X has enough separation" and "X has a nice enough base."

Exercises

E15. Let (X, d) be a metric space and $S \subseteq X$. Prove that if each continuous $f\colon S \to \mathbb{R}$ extends to a continuous $g\colon X \to \mathbb{R}$, then S is closed. (*The converse, of course, follows from Tietze's Extension Theorem.)*

E16. Urysohn's Lemma says that in a T_4-space disjoint closed sets are completely separated. Part a) shows that this is also true in a Tychonoff space if one of the closed sets is compact.

a) Suppose X is Tychonoff and $F, K \subseteq X$ where F is closed, K is compact. X and $F \cap K = \emptyset$. Prove that there is an $f \in C(X)$ such that $f \mid K = 0$ and $f \mid F = 1$. (*This is another example of the rule of thumb that "compact spaces act like finite spaces." If necessary, try proving the result first for a finite set K.*)

b) Suppose X is Tychonoff and that $p \in U$, where U is open in X. Prove $\{p\}$ is a G_δ set in X iff there exists a continuous function $f : X \to [0, 1]$ such that $f^{-1}(1) = \{p\}$ and $f|X - U = 0$.

E17. Suppose Y is a Hausdorff space. Define $x \sim y$ in Y iff there does not exist a continuous function $f : X \to [0, 1]$ such that $f(x) \neq f(y)$. Prove or disprove: $Y/\sim$ is a Tychonoff space.

E18. Prove that a Hausdorff space X is normal iff for each finite open cover $\mathcal{U} = \{U_1, \ldots, U_n\}$ of X, there exist continuous functions $f_i\colon X \to [0, 1]$ $(i = 1, \ldots, n)$ such that $\sum_{i=1}^{n} f_i(x) = 1$ for each $x \in X$ and such that, for each i, $f_i \mid X - U_i \equiv 0$. (*Such a set of functions is called a <u>partition</u> <u>of</u> <u>unity</u> <u>subordinate</u> <u>to</u> <u>the</u> <u>finite</u> <u>cover</u> $\mathcal{U}$.*)

Hint ($\Rightarrow$) First build a new open cover $\mathcal{V} = \{V_1, \ldots, V_n\}$ that "shrinks" $\mathcal{U}$ in the sense that, $V_i \subseteq cl\, V_i \subseteq U_i$ for each i. To begin the construction, let $F_1 = X - \bigcup_{i>1} U_i$. Pick an open V_1 so that $F_1 \subseteq V_1 \subseteq cl\, V_1 \subseteq U_1$. Then $\{V_1, U_2, \ldots, U_n\}$ still covers X. Continue by looking at $F_2 = X - (V_1 \cup \bigcup_{i>2} U_i)$ and defining V_2 so that $\{V_1, V_2, U_3, \ldots, U_n\}$ is still a cover and $V_2 \subseteq cl\, V_2 \subseteq U_2$. Continue in this way to replace the U_i's one by one. Then use Urysohn's lemma to get functions g_i which can then be used to define the f_i's. .

E19. Suppose X is a compact, countable Hausdorff space. Prove that X is completely metrizable.

Hint: 1) For each pair of points $x_n \neq x_m$ in X pick disjoint open sets $U_{m,n}$ and $V_{m,n}$ containing these points. Consider the collection of all finite intersections of such sets.

2) Or: Since X is, countable, every singleton $\{p\}$ is a G_δ set. Use regularity to find a descending sequence of open sets V_n containing p such that $\bigcap_{n=1}^{\infty} cl\, V_n = \{p\}$. Prove that the V_n's are a neighborhood base at p.

E20. A space X is called completely normal if every subspace of X is normal. (*For example, every metric space is completely normal).*

a) Prove that X is completely normal if and only if the following condition holds:

whenever $A, B \subseteq X$ and (cl $A \cap B) \cup (A \cap \text{cl}\, B) = \emptyset$ (that is, each of A, B is disjoint from the closure of the other), then there exist disjoint open sets U and V with $A \subseteq U$ and $B \subseteq V$.

b) Recall that the "scattered line" (*Exercise IIIE.10*) consist of the set $X = \mathbb{R}$ with the topology $\mathcal{T} = \{U \cup V : U$ is open in the usual topology on $\mathbb{R}$ and $V \subseteq \mathbb{P}\}$. Prove that the scattered line is completely normal and therefore T_4.

E21. A T_1-space X is called perfectly normal if whenever A and B are disjoint nonempty closed sets in X, there is an $f \in C(X)$ with $f^{-1}(0) = A$ and $f^{-1}(1) = B$.

a) Prove that every metric space (X, d) is perfectly normal.

b) Prove that X is perfectly normal iff X is T_4 and every closed set in X is a G_δ-set.
Note: Example 3.10 shows a T_4. space X that is not perfectly normal.

c) Show that the scattered line (see *Exercise E20*) is not perfectly normal, even though every singleton set $\{p\}$ is a G_δ-set.

d) Show that the scattered line is T_4.
Hint: Use the fact that $\mathbb{R}$, with the usual topology, is normal. Nothing deeper than Urysohn's Lemma is required but the problem is a bit tricky.

E22. Prove that a T_3 space $(X, \mathcal{T})$ has a locally finite base $\mathcal{B}$ iff $\mathcal{T}$ is the discrete topology. (*Compare to Theorem 6.2.*)

Chapter VII Review

Explain why each statement is true, or provide a counterexample.

1. Suppose $(X, \mathcal{T})$ is a topological space and let $\mathcal{T}_w$ be the weak topology on X generated by $C(X)$. Then $\mathcal{T} \subseteq \mathcal{T}_w$.

2. If X is regular and $x \in \text{cl}\,\{y\}$, then $y \in \text{cl}\,\{x\}$.

3. Every separable Tychonoff space can be embedded in $[0,1]^{\aleph_0}$.

4. If $f \in C(\mathbb{Q})$, we say g is a square root of f if $g \in C(\mathbb{Q})$ and $g^2 = f$. If a function f in $C(\mathbb{Q})$ has more than one square root, then it has c square roots.

5. In a Tychonoff space, every closed set is an intersection of zero sets.

6. A subspace of a separable space need not be separable, but every subspace of the Sorgenfrey line is separable.

7. Suppose $\mathbb{N}$ has the cofinite topology. If A is closed in $\mathbb{N}$, then every $f \in C(A)$ can be extended to a function $g \in C(\mathbb{N})$.

8. For $n = 1, 2, ...$, let $f_n : \mathbb{R} \to \mathbb{R}$ be given by $f_n(x) = x + n$ and let $\mathcal{T}$ be the weak topology on $\mathbb{R}$ generated by the f_n's. Then the evaluation map $e : \mathbb{R} \to \mathbb{R}^{\aleph_0}$ given by $e(x)(n) = f_n(x)$ is an embedding.

9. Let C be the set of points in the Cantor set with the subspace topology from the Sorgenfrey line S. Every continuous function $f : C \to \mathbb{R}$ can be extended to a continuous function $g : S \to \mathbb{R}$.

10. If (X, d) is a metric space, then X is homeomorphic to a dense subspace of some compact Hausdorff space.

11. Suppose $F \subseteq \mathbb{R}^2$. F is closed iff F is a zero set.

12. Suppose K is a compact subset of the Hausdorff space $X \times Y$. Let $A = \pi_X[K]$. Then A is T_4.

13. Let $\mathcal{F} = \{f \in \mathbb{R}^{\mathbb{R}} : f \text{ is not continuous}\}$. The weak topology on $\mathbb{R}$ generated by $\mathcal{F}$ is the discrete topology.

14. A compact T_2 space is metrizable if and only if it is second countable.

15. Suppose F and K are disjoint subsets of a Tychonoff space X, where F is closed and K is compact. There are disjoint cozero sets U and V with $F \subseteq U$ and $K \subseteq V$.

16. Every T_4 space is homeomorphic to a subspace of some cube $[0,1]^m$.

17. Suppose X is a T_4-space with nonempty pairwise disjoint closed subspaces $F_1, ..., F_n$. There is an $f \in C(X)$ such that $f|F_i = i$ for all $i = 1, ..., n$.

Chapter VIII
Ordered Sets, Ordinals and Transfinite Methods

1. Introduction

In this chapter, we will look at certain kinds of ordered sets. If X is an ordered set with a few reasonable properties, then there is a natural way to define an "order topology" on X. For our purposes, we will be primarily interested in ordered sets that satisfy some very strong ordering conditions, including a requirement that every nonempty subset contains a smallest element. These very special ordered sets are called well-ordered. The most familiar example of a well-ordered set is $\mathbb{N}$ and it is the well-ordering property in $\mathbb{N}$ that lets us do mathematical induction. In this chapter we will see "longer" well-ordered sets, and these will give us a new method of proof called "transfinite induction." But we begin with ordered sets that are much simpler.

2. Partially Ordered Sets

Recall that a relation R on a set X is a subset of $X \times X$ (*see Definition I.5.2*). If $(x, y) \in R$, we write xRy. Order relations are usually denoted by symbols such as $\leq$, $<$, $\prec$, or $\preceq$. There are different kinds of order relations on X, and they usually satisfy some of the following conditions:

Definition 2.1 A relation R on X is called:

<u>transitive</u> if :	$\forall\, a, b, c \in X$	$(aRb$ and $bRc) \Rightarrow aRc.$
<u>reflexive</u> if :	$\forall\, a \in X$	aRa
<u>antisymmetric</u> if :	$\forall\, a, b \in X$	$(aRb$ and $bRa) \Rightarrow (a = b)$
<u>symmetric</u> if :	$\forall\, a, b \in X$	$aRb \Leftrightarrow bRa$ (that is, the set R is symmetric with respect to the diagonal $\Delta = \{(x, x) : x \in X\} \subseteq X \times X$).

Example 2.2

1) The relation " $=$ " on a set X is transitive, reflexive, symmetric, and antisymmetric. Viewed as a subset of $X \times X$, the relation " $=$ " is the diagonal set $\Delta = \{(x, x) : x \in X\}$.

2) In $\mathbb{R}$, the usual order relation $<$ is transitive and antisymmetric, but not reflexive or symmetric.

3) In $\mathbb{R}$, the usual order $\leq$ is transitive, reflexive and antisymmetric. It is not symmetric.

4) In Chapter I, we defined an order relation $\leq$ that applies to cardinal numbers. On any set $\mathcal{C}$ of cardinal numbers, the order $\leq$ is transitive, reflexive and antisymmetric (*by the Cantor-Schroeder-Bernstein Theorem I.10.2*). It is not symmetric (unless $|\mathcal{C}| \leq 1$).

Definition 2.3 A relation $\leq$ on a set X is called a partial order if $\leq$ is transitive, reflexive and antisymmetric. The pair $(X, \leq)$ is called a partially ordered set (or, for short, poset).

A relation $\leq$ on a set X is called a linear order if $\leq$ is a partial order and, further, every two elements in X are comparable: that is, $\forall\, a, b \in X$ either $a \leq b$ or $b \leq a$. In this case, the pair $(X, \leq)$ is called a linearly ordered set. For short, a linearly ordered set is also called a chain.

In any ordered set, we write $a < b$ to mean that $a \leq b$ and $a \neq b$; and we can always "invert" the order notation and write $b \geq a$ $(b > a)$ to mean the same thing as $a \leq b$ $(a < b)$.

In some books, a partial order is defined as a "strict" relation $<$ which is transitive and irreflexive $(\forall a \in X,\ a \not< a)$. In that case, we can define $a \leq b$ to mean "$a < b$ or $a = b$" to get a partial order in the sense defined above. This variation in terminology creates no real mathematical problems: it is completely analogous to worrying about whether "$\leq$" or "$<$" should be called the "usual order" on $\mathbb{R}$.

Example 2.4

1) Suppose $A \subseteq X$. For any kind of order $\leq$ on X, we can get an order $\leq_A$ on $A \subseteq X$ by restricting the order $\leq$ to A. More formally, $\leq_A\, = \,\leq \cap (A \times A)$. We always assume that a subset of an ordered set has this natural "inherited" ordering unless something else is explicitly stated. With that understanding, we usually omit the subscript and also write $\leq$ for the order $\leq_A$ on A.

If $(X, \leq)$ is a poset (or chain), and $A \subseteq X$, then $(A, \leq)$ is also a poset (or chain). For example, every subset of $(\mathbb{R}, \leq)$ is a chain.

2) For $z, w \in \mathbb{C}$ ($=$ the set of complex numbers), define $z \preceq w$ iff $|z| \leq |w|$, where $\leq$ is the usual order in $\mathbb{R}$. $(\mathbb{C}, \preceq)$ is not a poset. (*Why?*)

3) Let $(X, \mathcal{T})$ be a topological space. For $f, g \in C(X)$, define

$$f \leq g \text{ iff } \forall x \in X,\ f(x) \leq g(x).$$

As a set,

$$\leq\ = \{(f, g) \in C(X) \times C(X) :\ \forall x \in \mathbb{R}\ \ f(x) \leq g(x)\}.$$

Notice that here we are allowing an ambiguity in notation. We are using the symbol "$\leq$" with two different meanings: we are defining an order "$\leq$" in $C(X)$, but the comparison "$f(x) \leq g(x)$" refers to the usual order in different set, $\mathbb{R}$. Of course, we could be more careful (as in 2)) and write $f \preceq g$ for the new order on $C(X)$, but usually we won't be that fussy when the context makes clear which meaning of "$\leq$" we have in mind.

$(C(X), \leq)$ is a poset but usually not a chain: for example, if $X = \mathbb{R}$ and f, g are given by $f(x) = x$ and $g(x) = x^2$, then $f \not\leq g$ and $g \not\leq f$. When is $(C(X), \leq)$ a chain? (*The answer is not "iff $|X| \leq 1$."*)

4) The following two diagrams represent posets, each with 5 elements. Line segments upward from x to y indicate that $x \leq y$. In Figure (i), for example, $d \leq c$ and $c \leq a$ (so $d \leq a$); in Figure (i), a and b are not comparable.

Figure (ii) shows a chain: $a \leq b \leq c \leq d \leq e$.

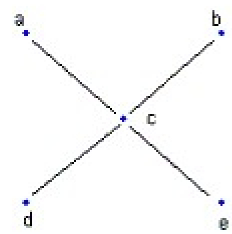


Figure (i)

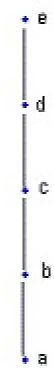


Figure (ii)

5) Suppose $\mathcal{C}$ is a collection of sets. We can define $\leq$ on $\mathcal{C}$ by $A \leq B$ iff $A \subseteq B$. Then $(\mathcal{C}, \leq)$ is a poset. In this case, we say that $\mathcal{C}$ has been <u>ordered by inclusion</u>. In particular, for any set X, we can order $\mathcal{P}(X)$ by inclusion. What conditions on X will guarantee $(\mathcal{P}(X), \leq)$ is a chain ?

6) Suppose $\mathcal{C}$ is a collection of sets. We can define $\leq$ on $\mathcal{C}$ by $A \leq B$ iff $A \supseteq B$. Then $(\mathcal{C}, \leq)$ is a poset. In this case, we say that $\mathcal{C}$ has been <u>ordered by reverse inclusion</u>. In particular, for any set X, we can order $\mathcal{P}(X)$ by reverse inclusion. What conditions on X will guarantee that $(\mathcal{P}(X), \leq)$ is a chain ?

For a given collection $\mathcal{C}$, Examples 5) and 6) are quite similar: one is a "mirror image" of the other. The identity map $i : \mathcal{C} \to \mathcal{C}$ is an "order-reversing isomorphism" between the posets.

For our purposes, the "reverse inclusion" ordering on a collection of sets will turn out to be more useful. For a point x in a topological space X, we can order the neighborhood system $\mathcal{N}_x$ by <u>reverse inclusion</u> $\leq$. The simplicity or complexity of the poset $(\mathcal{N}_x, \leq)$ reflects some topological properties of X and is a measure of how complicated the neighborhood system $\mathcal{N}_x$ is. For example,

i) If x is isolated, then $\{x\} \in N_x$ and $N \leq \{x\}$ for every $N \in N_x$. So the poset $(\mathcal{N}_x, \leq)$ has a largest element. (*Is the converse true?*)

ii) Suppose X is first countable and that $\mathcal{B}_x = \{N_1, N_2, ..., N_k, ...\}$ is a countable neighborhood base at x. Then for every $N \in \mathcal{N}_x$ there is a k such that $N \leq N_k$, and we say that $\mathcal{B}_x$ is <u>cofinal</u> in $\mathcal{N}_x$. More informally, the poset $(\mathcal{N}_x, \leq)$ has a countable subset $\{N_1, N_2, ..., N_k, ...\}$ that contains "arbitrarily large" members of $\mathcal{N}_x$.

iii) The poset $(\mathcal{N}_x, \leq)$ is usually not a chain. But it does have another interesting order property: if $N_1, N_2 \in \mathcal{N}_x$, then $\{N_1, N_2\}$ has an upper bound in $\mathcal{N}_x$. To see this notice that if $N_3 = N_1 \cap N_2$ then $N_1 \leq N_3$ and $N_2 \leq N_3$.

This poset $(\mathcal{N}_x, \leq)$ will turn out (in Chapter 9) to be the inspiration for the definition of "convergent net," a kind of convergence that is more powerful than "convergent sequence." Nets resemble sequences in some ways but are more powerful – unlike sequences, nets can be used to determine closures (and therefore the topology) in any topological space.

Definition 2.5 Suppose $(X, \leq)$ is a poset and let $A \subseteq X$. An element $x \in X$ is called

i) the largest (or last) element in X if $y \leq x$ for all $y \in X$
ii) the smallest (or first) element in X if $x \leq y$ for all $y \in X$
iii) a maximal element in X if $(y \geq x \Rightarrow y = x)$ for all $y \in X$
iv) a minimal element in X if $(y \leq x \Rightarrow y = x)$ for all $y \in X$.

Clearly, a largest (or smallest) element in X, if it exists, is unique. For example: if z_1 and z_2 were both largest in X, then $z_1 \leq z_2$ and $z_2 \leq z_1$ so $z_1 = z_2$.

In Figure (i): both a, b are maximal elements and d, e are minimal elements. This poset has no largest or smallest element. *Suppose a poset $(X, \leq)$ has a unique maximal element z. Must z also be the largest element in $(X, \leq)$?*

In Figure (ii): a is the smallest (and also a minimal element); e is the largest (and also a maximal) element.

v) Suppose $A \subseteq X$ and $x \in X$. We say that $x \in X$ is an upper bound for A if $a \leq x$ for all $a \in A$; x is called a least upper bound (sup) for A if x is the smallest upper bound for A. A set A might have many upper bounds, one upper bound, or no upper bounds in X. If A has more than one upper bound, A might or might not have a least upper bound in X. But if X has a least upper bound x, then the least upper bound is unique (*why?*).

vi) An element $x \in X$ is called a lower bound for A if $x \leq a$ for all $a \in A$; x is called a greatest lower bound (inf) for A if x is the largest lower bound for A. A set A might have many lower bounds, one lower bound, or no lower bounds in X. If A has more than one lower bound, A might or might not have a greatest lower bound in X. But if X has a greatest lower bound x, then the greatest lower bound is unique (*why?*).

vii) If $x < y \in X$ and if $\nexists\, z \in X$ with $x < z < y$, then x is called an immediate predecessor of y and y is called an immediate successor of x. In a poset, an immediate predecessor or successor might not be unique; but if X is a chain, then an immediate predecessor or successor, if it exists, must be unique. (*Why?*)

In Figure (i), the upper bounds on $\{d, e\}$ are a, b, c, and $c = \sup\{d, e\}$; the set $\{d, e\}$ has no lower bounds. Both a, b are immediate successors of c. The elements d, e have no immediate predecessor (in fact, no predecessors at all). In Figure (ii), the immediate predecessor of c is b and d is the immediate successor of c.

Example 2.6 If $\leq_1$ is an order on X, then $\leq_1 \subseteq X \times X$. So if $\leq_1$ and $\leq_2$ are orders on X, it makes sense to ask whether $\leq_1 \subseteq \leq_2$, or vice-versa. If we look at the set $\mathcal{P} = \{ \leq : \leq$ is a partial order on $X\}$, then $\mathcal{P}$ is partially ordered by inclusion. A linear order $\leq$ is a maximal element in $(\mathcal{P}, \subseteq)$ (*Why? Is the converse true?*)

3. Chains

Definition 3.1 Let $(X, \leq)$ be a chain. The order topology on X is the topology for which all sets of the form $\{x \in X : x < a\}$ or $\{x \in X : b < x\}$ $(a, b \in X)$ are a subbase. (*As usual, we write $x < y$ as shorthand for "$x \leq y$ and $x \neq y$."*)

It's handy to use standard "interval notation" when working with chains: for example, if $a, b \in X$:

$$\{x \in X : a < x < b\} = (a, b)$$
$$\{x \in X : a \leq x < b\} = [a, b)$$
$$\{x \in X : x \leq a\} = (-\infty, a]$$

But we need to be careful not to read too much into the notation. For example, the chain in Figure (ii), shows how interval notation can be misleading if not used thoughtfully: $(a, b) = \emptyset$, $(d, \infty) = \{e\}$, $(-\infty, b) = \{a\}$, $(a, c) = \{b\}$, and $(a, e) = [b, d]$.

Example 3.2

1) The order topology on the chain in Figure (ii) is the discrete topology.

2) The order topology on $\mathbb{N}$ is the usual (discrete) topology: $\{1\} = \{k \in \mathbb{N} : k < 2\}$ $= (-\infty, 2)$; and for $n > 1$, $\{n\} = (n-1, n+1)$.

Example 3.3 $\mathbb{P}$ and $\mathbb{Q}$ each have an order inherited from $\mathbb{R}$, and their order topologies are the same as the usual subspace topologies. But, in general, we have to be careful about the topology on $A \subseteq X$ when X is a chain with the order topology. There are two possible topologies on A:

a) The order $\leq$ gives an order topology $\mathcal{T}_{\leq}$ on X and we can give A the subspace topology $(\mathcal{T}_{\leq})_A$.

b) A has an ordering $\leq_A$ (inherited from the order $\leq$ on X) and we can use it to give A an order topology. More formally, we could write this topology as $\mathcal{T}_{\leq_A}$.

Unfortunately, these two topologies might not be the same. Let $A = (0, 1) \cup \{2\} \subseteq \mathbb{R}$. The order topology $\mathcal{T}_{\leq}$ on $\mathbb{R}$ is the usual topology on $\mathbb{R}$, and this topology produces a subspace topology for which 2 is isolated in A.

But in the order topology on A, each basic open set containing 2 must have the form $\{x \in A : x > a\} = (a, 2]$ where $a < 1$. So 2 is not isolated in $(A, \mathcal{T}_{\leq_A})$. In fact, the space $(A, \mathcal{T}_{\leq_A})$ is homeomorphic to $(0, 1]$ (*why?*).

Is there any necessary inclusion $\subseteq$ or $\supseteq$ between $\mathcal{T}_{\leq_A}$ and $(\mathcal{T}_{\leq})|_A$? Can you state any hypotheses on X or A that will guarantee that $\mathcal{T}_{\leq_A} = (\mathcal{T}_{\leq})|_A$?

Example 3.4 We defined the order topology only for chains, but the same definition could be used in any ordered set $(X, \leq)$. We usually restrict our attention to chains because otherwise the order topology may not be very nice. For example, let $X = \{a, b, c\}$ with the partial order represented by the following diagram:

X is the only open set containing a (*why?*), so the order topology is not T_1. (*Can you find a poset for which the order topology is not even T_0?*) By contrast, the order topology for any chain $(X, \leq)$ has good separation properties. For example, it is easy to show that the order topology on a chain must be T_2 :

> Suppose $a \neq b \in X$ where, say, $a < b$. If a is the immediate predecessor of b, we can let $U = \{x \in X : x < b\}$ and $V = \{x \in X : x > a\}$. But if there exists a point c satisfying $a < c < b$, then we can define $U = \{x \in X : x < c\}$ and $V = \{x \in X : x > c\}$. Either way, we have a pair of disjoint open sets U and V with $x \in U$ and $y \in V$.

In fact, the order topology on a chain is always T_4, but the proof is much messier so we will not include it here. *Most of our interest in this Chapter will be with ordered sets that are much more special than chains – the well-ordered sets). For these sets we will prove that the order topology is T_4.*

From now on, when the context is clear, we will simply write X, rather than $(X, \leq)$, for an ordered set. We will denote the orders in many different sets all with the same symbol "$\leq$", letting the context determine which order is being referred to. If it becomes necessary to distinguish carefully between different orders, then we will occasionally add subscripts such as $\leq_1$, $\leq_2$, ...

Definition 3.5 A function $f : X \to Y$ between ordered sets is called an order isomorphism if f is bijection and both f and f^{-1} are order preserving – that is, $a \leq b$ if and only if $f(a) \leq f(b)$. (*Since f is one-to-one, it follows that $x < y$ if and only if $f(x) < f(y)$.*) If such an f exists, we say that X and Y are order isomorphic and write $X \simeq Y$. If f is not onto, then $X \simeq f[X] \subseteq Y$ and we say that f is an order isomorphism of X into Y. Between ordered sets, "$\simeq$" will always refer to order isomorphism.

Theorem 3.6 Let $X, Y,$ and Z be chains (*Parts i-iv), however, are also true for posets*).

i) $X \simeq X$
ii) $X \simeq Y$ iff $Y \simeq X$
iii) if $X \simeq Y$ and $Y \simeq Z$, then $X \simeq Z$
iv) $X \simeq Y$ implies $|X| = |Y|$.
v) if X and Y are finite chains, then $X \simeq Y$ iff $|X| = |Y|$.

Proof The proof is very easy and is left as an exercise. •

Even though the proof is easy, there are some interesting observations to make.

1) To show that $X \simeq X$, the identity map i might not be the only possible order isomorphism. For example, when $X = \mathbb{R}$, the functions $f(x) = x$ and $f(x) = x^3$ are both order isomorphisms between $\mathbb{R}$ and $\mathbb{R}$. *Check: how many order isomorphisms exist between* $\mathbb{N}$ *and* $\mathbb{N}$*?*

2) A chain can be order isomorphic to a proper subset of itself. For example, $f(n) = 2n$ is an order isomorphism between $\mathbb{N}$ and $\mathbb{E}$ (the set of even natural numbers). Both $\mathbb{N}$ and $\mathbb{E}$ are order isomorphic to the set of all prime numbers. *Check: must every two countable infinite chains be order isomorphic?*

3) An order isomorphism between X and Y preserves largest, smallest, maximal and minimal elements (if they exist). Therefore $(0,1)$ and $[0,1]$ are not order isomorphic: for example, $[0,1]$ has a smallest element and $(0,1)$ doesn't. Similarly, $\mathbb{N}$ is not order isomorphic to the set of integers $\mathbb{Z}$.

An order isomorphism preserves "betweenness," so $\mathbb{Z}$ is not order isomorphic to $\mathbb{Q}$: in $\mathbb{Q}$, there is a third element between any two elements, but this is false in $\mathbb{Z}$.

4) Let $\mathbb{C}$ be the set of complex numbers. If $f: \mathbb{C} \to \mathbb{R}$ is any bijection, then we can use f to create a chain $(\mathbb{C}, \leq)$: simply define $z_1 \leq z_2$ iff $f(z_1) \leq f(z_2)$. Then $(\mathbb{C}, \leq) \simeq (\mathbb{R}, \leq)$.

Of course, this chain $(\mathbb{C}, \leq)$ is not very interesting from the point of view of algebra or analysis: we imposed an arbitrary ordering on $\mathbb{C}$ that has nothing to do with the algebraic structure of $\mathbb{C}$. For example, there is no reason to think that $z_1 \leq z_2$ and $z_3 \leq z_4$, then $z_1 + z_3 \leq z_2 + z_4$.

Similarly, a bijection $f: \mathbb{N} \to \mathbb{Q}$ can be used to give $\mathbb{Q}$ a new order $\leq$ so that the two chains are order isomorphic. In this case, f is just a sequence that enumerates $\mathbb{Q}$: $q_1, q_2, \ldots, q_n, \ldots$ and a new order on $\mathbb{Q}$ could simply defined as $q_1 < q_2 < \ldots < q_n < \ldots$.

In general, if $f: X \to Y$ is a bijection and one of X or Y is ordered, we can use f to "transfer" the order to the other set in such a way that $(X, \leq) \simeq (Y, \leq)$.

5) Clearly, order isomorphic chains are homeomorphic in their order topologies. But the converse is false. Suppose $X = \{-\frac{1}{n} : n \in \mathbb{N}\} \cup \{0\}$ and $Y = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$. The order topologies on X and Y are the usual topologies, and the reflection $f(x) = -x$ is a homeomorphism (in fact, an isometry) between them.

The largest element in Y is 1, and it has an immediate predecessor, $\frac{1}{2}$. But 0, the largest element in X, has no immediate predecessor in X. Since an order isomorphism preserves largest elements and immediate predecessors, there is no order isomorphism between X and Y.

The next theorem tells when an ordered set is order isomorphic to the set of rational numbers.

Theorem 3.7 (Cantor) Suppose that $(L, \leq)$ is a nonempty countable chain such that

a) $\forall a \in L, \exists b \in L$ with $a < b$ (L has "no last element")
b) $\forall a \in L, \exists b \in L$ with $b < a$ (L has no first element")
c) $\forall a, b \in L$, if $a < b$ then $\exists c \in L$ such that $a < c < b$.

Then $(L, \leq)$ is order isomorphic to $\mathbb{Q}$.

When a chain that satisfies the third condition – that between any two elements there must exist a third element – we say that X is order-dense. Then Theorem 3.7 can be restated as: a nonempty countable order-dense chain with no first or last element is order isomorphic to $\mathbb{Q}$.

We might try the following: enumerate $L = \{l_1, l_2, .., l_n, ...\}$ and $\mathbb{Q} = \{q_1, q_2, ..., q_n, ...\}$ and let $f(l_1) = q_1$. Then inductively define $f(l_n) =$ some $q \in \mathbb{Q}$ chosen so that $f(l_n) \neq f(l_1), ..., f(l_{n-1})$ and so that $f(l_n)$ has the same order relations (in $\mathbb{Q}$) to $f(l_1), ..., f(l_{n-1})$ as l_n has to $l_1, ..., l_{n-1}$ (in L). For example, if $l_2 < l_3 < l_4 < l_1$ and $f(l_1), f(l_2), f(l_3)$ have been defined, then define $f(l_4) \in \mathbb{Q}$ so that $f(l_2) < f(l_3) < f(l_4) < f(l_1)$. This argument shows that for any countable chain L, there is a one-to-one order preserving map $g : L \to \mathbb{Q}$, so any countable chain is order isomorphic to a subset of $\mathbb{Q}$. However, the function g constructed in this way might not be onto. The "back-and-forth" construction used in the following proof is designed to be sure we end up with an onto order preserving map $g : \mathbb{Q} \to L$ if L has properties a)-c).

First we prove a lemma: it states that an order isomorphism between finite subsets of $\mathbb{Q}$ and L can always be extended to include one more point in its domain (or, one more point in its range).

Lemma: Suppose $A \subseteq \mathbb{Q}$, $B \subseteq L$ where A, B are finite. Let g be an order isomorphism from A onto B.

a) If $q \in \mathbb{Q} - A$, there exists an order isomorphism h that extends g and for which dom$(h) = A \cup \{q\}$ and ran$(h) \subseteq L$.

b) If $l \in L - B$, there exists an order isomorphism h that extends g and for which dom$(h) \subseteq \mathbb{Q}$ and ran $(h) = B \cup \{l\}$.

Proof of a) Suppose $A = \{q_1, ..., q_n\}$ and that $g(q_i) = l_i$. Pick an $l \in L - B$ that has the same order relations to $l_1, ..., l_n$ as q does to $q_1, ..., q_n$. (*For example: if q is greater that all of* $q_1, ..., q_n$, *pick l greater than all of* $l_1, ..., l_n$; *if i and j are the largest and smallest subscripts for which* $q_i < q < q_j$, *then choose l between* l_i *and* l_j.) This choice is always possible since L satisfies a), b), and c). Define h by $h(q) = l$ and $h(x) = g(x)$ for $x \in A$.

Proof for b) The proof is almost identical. •

Proof of Theorem 3.7 L is a countable chain. Since $L \neq \emptyset$ and L has no last element, L must be infinite. Without loss of generality, we may assume that $L \cap \mathbb{Q} = \emptyset$.

We define an order isomorphism g between $\mathbb{Q}$ and L in stages. At each stage, we enlarge the function we have by adding a new point to its domain or range.

Let $M = \mathbb{Q} \cup L = \{m_1, m_2, ..., m_n, ...\}$. Each element of $\mathbb{Q}$ and L appears exactly once in this list. Define $g_0 = \emptyset$, and continue by induction. Suppose $n > 0$ and that an order isomorphism $g_{n-1} : A_{n-1} \to B_{n-1}$ has been defined, where $A_{n-1} \subseteq \mathbb{Q}$ and $B_{n-1} \subseteq L$.

If $m_n \in \mathbb{Q}$ and $m_n \notin A_{n-1}$, use the Lemma to get an order isomorphism g_n that extends g_{n-1} and for which $\text{dom}(g_n) = A_{n-1} \cup \{m_n\} = A_n$. Let $B_n = \text{ran}(g_n) \subseteq L$.

If $m_n \in L$ and $m_n \notin B_{n-1}$, use the Lemma to get an order isomorphism g_n that extends g_{n-1} and for which $\text{ran}(g_n) = B_{n-1} \cup \{m_n\} = B_n$. Let $A_n = \text{dom}(g_n) \subseteq \mathbb{Q}$.

If $m_n \in \text{dom}(g_{n-1}) \cup \text{ran}(g_{n-1})$, let $g_n = g_{n-1}$, $A_n = A_{n-1}$ and $B_n = B_{n-1}$.

By induction, g_n is defined for all n, and since g_n extends g_{n-1} and we can define an order isomorphism $g = \bigcup_{n=0}^{\infty} g_n$. The construction guarantees that $\text{dom}(g) = \mathbb{Q}$ and $\text{ran}(g) = L$. •

"Being order isomorphic" is an equivalence relation among ordered sets so any two chains having the properties in Cantor's theorem are order isomorphic to each other. Since order isomorphic chains have homeomorphic order topologies, we have a topological characterization of $\mathbb{Q}$ in terms of order.

Corollary 3.8 A nonempty countable order-dense chain with no largest or smallest element is homeomorphic to $\mathbb{Q}$.

The following corollary gives a characterization of <u>all</u> order-dense countable chains.

Corollary 3.9 If X is a countable order-dense and $|X| > 1$, then X is order isomorphic to exactly one of the following chains :

a) $\mathbb{Q} \cap (0,1) \simeq \mathbb{Q}$
b) $\mathbb{Q} \cap [0,1)$
c) $\mathbb{Q} \cap (0,1]$
d) $\mathbb{Q} \cap [0,1]$

Proof By looking at largest and smallest elements, we see that no two of these chains are order isomorphic. Therefore no chain is order isomorphic to more than one of them.

If X has no largest or smallest element then, Cantor's theorem gives $X \simeq \mathbb{Q}$ and a second application of Cantor's Theorem gives that $\mathbb{Q} \simeq \mathbb{Q} \cap (0,1)$.

If X has a smallest element, a, but no largest element, then $X - \{a\}$ is nonempty and has no smallest element (*why*?). $X - \{a\}$ clearly satisfies the other hypotheses of the Cantor's theorem so there is an order isomorphism $h : X - \{a\} \to \mathbb{Q} \cap (0,1)$. Then define an order isomorphism $g : X \to \mathbb{Q} \cap [0,1)$ by setting $g(a) = 0$ and $g(x) = h(x)$ for $x \neq a$..

The proofs of the other cases are similar. •

No two of the chains a)-d) mentioned in Corollary 3.9 are order isomorphic, but they are, in fact, all <u>homeomorphic</u> topological spaces. We can see that b) and c) are homeomorphic by using the map

$f(x) = 1 - x$, but the other homeomorphisms are not so obvious. Here is a sketch of a proof, contributed by Edward N. Wilson, that $(0,1) \cap \mathbb{Q}$ is homeomorphic to $(0,1] \cap \mathbb{Q}$.

For short, write $(0,1) \cap \mathbb{Q} = (0,1)_{\mathbb{Q}}$ and $(0,1] \cap \mathbb{Q} = (0,1]_{\mathbb{Q}}$.

In $(0,1)$, choose a strictly increasing sequence of irrationals $(p_n) \to 1 \in \mathbb{R}$. Let $p_0 = 0$.
For $n \geq 0$: let $O_n = (p_n, p_{n+1}) \cap \mathbb{Q}$, $U_n = (\frac{1}{2}p_n, \frac{1}{2}p_{n+1}) \cap \mathbb{Q}$,
and $V_n = (1 - \frac{1}{2}p_{n+1}, 1 - \frac{1}{2}p_n) \cap \mathbb{Q}$. Each of $O_n, U_n,$ and V_n is clopen in $(0,1)_{\mathbb{Q}}$ and in $(0,1]_{\mathbb{Q}}$.

We have $(0,1]_{\mathbb{Q}} = \bigcup_{n=0}^{\infty} O_n \cup \{1\}$ and $(0,1)_{\mathbb{Q}} = \bigcup_{n=0}^{\infty} U_n \cup \{\frac{1}{2}\} \cup \bigcup_{n=0}^{\infty} V_n$
Define a map ϕ: $(0,1]_{\mathbb{Q}} \to \bigcup_{n=0}^{\infty} U_n \cup \{\frac{1}{2}\} \cup \bigcup_{n=0}^{\infty} V_n$ by

$\phi \mid_{O_{2n}}$ = an increasing linear map from O_{2n} onto U_n

$\phi \mid_{O_{2n+1}}$ = a decreasing linear map from O_{2n+1} onto V_n

$\phi(1) = \frac{1}{2}$

Then ϕ is a homeomorphism. (Since the sets O_n, U_n, V_n are clopen, it is clear that ϕ is continuous at all points except perhaps 1 and that ϕ^{-1} is continuous at all points except perhaps $\frac{1}{2}$. These special cases are easy to check separately.)

There is, in fact, a more general theorem that states that <u>*every infinite countable metric space with no isolated points is homeomorphic to* $\mathbb{Q}$</u>*. This theorem is due to Sierpinski (1920).*

We can also characterize the real numbers as an ordered set.

Theorem 3.10 Suppose X is a nonempty chain which

i) has no largest or smallest element
ii) is order-dense
iii) is separable in the order topology
iv) is Dedekind complete (that is, every nonempty subset of X which has an upper bound in X has a least upper bound in X).

Then X is order isomorphic to $\mathbb{R}$ (and therefore X, with its order topology, is homeomorphic to $\mathbb{R}$).

Proof We will not give all the details of a proof, However, the ideas are completely straightforward and the details are easy to fill in.

Let D be a countable dense set in the order topology on X. Then D satisfies the hypotheses of Cantor's Theorem (*why?*) so there exists an (onto) order isomorphism $f : \mathbb{Q} \to D$. For each irrational $p \in \mathbb{R}$, let $\mathbb{Q}_p = \{q \in \mathbb{Q} : q < p\}$ and extend f to an order isomorphism $g : \mathbb{R} \to X$ by defining $f(p) = \sup f[\mathbb{Q}_p]$. ∙

Remark In a separable space, any family of disjoint open sets must be countable (*why?*). Therefore we could ask whether condition iii) in Theorem 3.10 can be replaced by

iii)$'$ every collection of disjoint open intervals in X is countable.

In other words, can we say that

(**) a nonempty chain satisfying i), ii), iii)$'$, and iv), must X be order isomorphic to $\mathbb{R}$?

The Souslin Hypothesis (SH) states that the answer to (**) is "yes." The status of SH was famously unknown for many years. Work of Jech, Tennenbaum and Solovay in the 1970's showed that SH is consistent with and independent of the axioms ZFC for set theory – that is, SH is undecidable in ZFC. We could add either SH or its negation as an additional axiom in ZFC without introducing an inconsistency. If one assumes that SH is false, then there is a nonempty chain satisfying i), ii), iii)$'$, and iv) but not order isomorphic to $\mathbb{R}$: such a chain is called a Souslin line.

SH was of special interest for a while in connection with the question "if X is a T_4-space, is $X \times [0,1]$ necessarily T_4?" In the 1960's, Mary Ellen Rudin showed that if she had a Souslin line to work with – that is, if SH is false – then the answer to the question was "no." (*See the remarks following Example III.5.7*)

There are lots of equivalent ways of formulating SH—for example, in terms of graph theory. There is a very nice expository article by Mary Ellen Rudin on the Souslin problem in the American Mathematical Monthly, 76(1969), 1113-1119. The article was written before the consistency and independence results of the 1970's and deals with aspects of SH in a "naive" way.

Exercises

E1. Prove or disprove: if R is a relation on X which is both symmetric and antisymmetric, then R must be the equality relation " $=$ ".

E2. State and prove a theorem of the form:

A power set $\mathcal{P}(X)$, ordered by inclusion, is a chain iff ...

E3. Suppose $(X, \leq)$ is a poset in which every nonempty subset contains a largest and smallest element. Prove that $(X, \leq)$ is a finite chain.

E4. Prove that any countable chain $(L, \leq)$ is order isomorphic to a subset of $(\mathbb{Q}, \leq)$.
(*Hint: See the "Caution" in the proof of Cantor's Theorem 3.7.*)

E5. Let $(X, \leq)$ be an infinite poset. A subset C of X is called <u>totally unordered</u> if no two distinct elements of C are comparable, that is:

$$\forall a \in C \ \forall b \in C \quad (a \leq b) \ \Leftrightarrow \ (a = b)$$

Prove that <u>either</u> X has a subset C which is an infinite chain <u>or</u> X has a totally unordered infinite subset C.

E6. Let $(X, \leq)$ be a poset in which the <u>longest</u> chain has length n ($n \in \mathbb{N}$). Prove that X can be written as the union of n totally unordered subsets (*see E5*) and the n is the smallest natural number for which this is true.

.

4. Order Types

In Chapter I, we assumed that we can somehow assign a cardinal number $|X|$ to each set X, in such a way that $|X| = |Y|$ iff there exists a bijection $f : X \to Y$. Similarly, we now assume that we can assign to each chain an "object" called its order type and that this is done in such a way that two chains have the same order type iff the chains are order isomorphic. Just as with cardinal numbers, an exact description of how this can be done is not important here. In axiomatic set theory, all the details can be made precise. Of course, then, the order type of a chain turns out itself to be a certain set (since "everything is a set" in ZFC). For our purposes, it is enough just to take the naive view that "order-isomorphic" is an equivalence relation among chains, and that each equivalence class is an order type.

We will usually denote order types by lower case Greek letters such as μ, ν, τ, ω – with a few traditional exceptions mentioned in the next example. If μ is the order type of a chain M, we say that M represents μ.

Example 4.1

1) Two chains with the same order type are order isomorphic. Since the order isomorphism is a bijection, chains with the same order type also have the same cardinal number. But the converse is false: $\mathbb{Q}$ and $\mathbb{N}$ have the same cardinal number but they have different order types because the sets are not order isomorphic.

However, two finite chains have the same cardinality iff they are order isomorphic. Therefore, for finite chains, we will use the same symbol for both the cardinal number and the order type. (*In the precise definitions of axiomatic set theory, the cardinal number and the order type of a finite chain do turn out to be the same set!*)

2) 0 is the order type of $\emptyset$
1 is the order type of $\{0\}$
2 is the order type of $\{0, 1\}$
.
.
.
n is the order type of $\{0, 1, \ldots, n-1\}$
.
.
.
ω_0 is the order type of $\{0, 1, 2, \ldots\}$. ω_0 is also the order type of $\mathbb{N}$. *The subscript "0" hints at bigger things to come.*
ω_0 is also the order type of $\mathbb{E} = \{2, 4, 6, \ldots\}$ since this chain is order isomorphic to $\mathbb{N}$.

Notice that each order type in this example is represented by "the set of all preceding order types."

Definition 4.2 Let μ and ν be order types represented by chains M and N. We say that $\mu \leq \nu$ if there exists an order isomorphism f of M into N. We write $\mu < \nu$ if $\mu \leq \nu$ but $\mu \neq \nu$ (that is, $M \not\simeq N$). (*Check that the definition is independent of the chains M and N chosen to represent μ and ν.*)

Example 4.3 Let μ be the order type of a chain M. Since $\emptyset$ is order isomorphic to a subset of M we have $0 \leq \mu$. More generally, $0 < 1 < 2 < \ldots < \omega_0$.

Suppose $(X, \leq)$ has order type μ. With a little reflection, we can create a new chain $(X, \leq^*)$ by defining $x \leq^* y$ iff $y \leq x$. We write μ^* for the order type of $(X, \leq^*)$. For example, ω_0^* is the order type of the chain $\{\ldots, -2, -1\}$ of negative integers. Since $\omega_0 \not\leq \omega_0^*$ and $\omega_0^* \not\leq \omega_0$ (*why?*), we see that two order types may not be comparable.

The relation $\leq$ between order types is reflexive and transitive but it is not antisymmetric – for example, let μ and ν be order types of the intervals $(0, 1)$ and $[0, 1]$: then $\mu \leq \nu$ and $\nu \leq \mu$ but $\mu \neq \nu$. Therefore $\leq$ is not even a partial ordering among order types.

Definition 4.4 For $\alpha \in A$, let $(M_\alpha, \leq_\alpha)$ be pairwise disjoint chains and suppose that the index set A is also a chain. We define the ordered sum $\sum_{\alpha \in A} M_\alpha$ as the chain $(\bigcup_{\alpha \in A} M_\alpha, \leq)$, where we define

$$x \leq y \quad \text{if} \quad \begin{cases} x, y \in M_\alpha \text{ and } x \leq_\alpha y, & \text{or} \\ \alpha < \beta \in A,\ x \in M_\alpha \text{ and } y \in M_\beta \end{cases}$$

We can "picture" the ordered sum as laying the chains M_α "end-to-end" with larger α's further to the right. In particular, for two disjoint chains $(M, \leq_M)$ and $(N, \leq_N)$, the ordered sum $(M + N, \leq)$ is formed by putting N to the right of ("larger than") M and using the old orders inside each of M and N.

Definition 4.5 Suppose A is a chain and that for each $\alpha \in A$, we have an order type μ_α. Let the μ_α's be represented by pairwise disjoint chains M_α. Then $\sum_{\alpha \in A} \mu_\alpha$ as order type of the chain $(\bigcup_{\alpha \in A} M_\alpha, \leq)$. In particular, if μ and ν are order types represented by disjoint chains M and N, then $\mu + \nu$ is the order type of the ordered sum $(M + N, \leq)$. (*Check that sum of order types is independent of the disjoint chains used to represent the order types.*)

Example 4.6 Addition of order types is clearly associative: $(\mu + \nu) + \tau = \mu + (\nu + \tau)$. It is not commutative. For example $\omega_0 + 1 \neq 1 + \omega_0$, since a chain representing the left side has a largest element but a chain representing the right side does not. In general, for $n \in \mathbb{N}$, $n + \omega_0 = \omega_0$ while, if $m \neq n \in \mathbb{N}$, $\omega_0 \neq \omega_0 + n \neq \omega_0 + m$. Of course, chains representing these different order types all have cardinality $\aleph_0$.

The order type $\omega_0 + \omega_0$ is represented by the ordered set $\{0, 1, 2, 3, \ldots;\ a_1, a_2, a_3, \ldots\}$ where ";" indicates that each a_i is larger than every n. Less abstracting, $\omega_0 + \omega_0$ could also be represented by the chain $\{1 - \frac{1}{n} : n \in \mathbb{N}\} \cup \{2 - \frac{1}{n} : n \in \mathbb{N}\} \subseteq \mathbb{Q}$.

$\omega_0^* + \omega_0$ is the order type of the set of integers $\mathbb{Z}$. Is $\omega_0^* + \omega_0 = \omega_0 + \omega_0^*$? (*Why or why not? Give an example of a subset of $\mathbb{Q}$ that represents $\omega_0 + \omega_0^*$.*)

Example 4.7 It is easy to prove that every countable chain C is order isomorphic to a subset of $\mathbb{Q}$: just list the elements of C and inductively define a one-to-one mapping into $\mathbb{Q}$ that preserves order at each step (*see the "attempted" argument that precedes that actual proof of Cantor's Theorem 3.7*) *Theorem 3.7*). But if every countable order type can be represented by some subset of $\mathbb{Q}$, then there can be at most c different countable order types.

As a matter of fact, we can prove that there are exactly c different countable order types. A sketch of the argument follows: the details are easy to fill in (or see W. Sierpinski, *Cardinal and Ordinal Numbers* – an old "Bible" on the subject with much more information than anybody would want to know.)

$\mathbb{Z}$ has order type $\xi = \omega_0^* + \omega_0$. For each sequence $a = (a_1, a_2, a_3 ...) \in \{0,1\}^{\mathbb{N}}$, we can define an order type

$$\xi_a = \xi + 1 + a_1 + \xi + 1 + a_2 + \xi + 1 + a_3 + \ldots$$

It is not hard to show that the map $a \mapsto \xi_a$ is one-to-one. Here is a sketch of the argument:

We say that two elements of a chain to be in the same component if there are only finitely many elements between them. (*This use of the word "component" has nothing to do with connectedness.*) It is clear all elements in the same component are smaller (or larger) than all elements in a different component; this observation lets us order the components of the chain.

Suppose C is a chain representing ξ_a and let the finite components of C (listed in order of increasing size) be $F_1, F_2, ..., F_n, ...$ where F_n has order type $1 + a_n$.

Let $a \neq b \in \{0,1\}^{\mathbb{N}}$ and suppose C' is a chain that represents ξ_b. Call the finite components of C' (listed in order of increasing size) $F_1', F_2', ..., F_n', ...$ where F_n' has order type $1 + b_n$.

An order isomorphism between C and C' would necessarily carry F_n to F_n' for every n. But this is impossible since, for some n, one of $1 + a_n$ and $1 + b_n$ is 1 and the other is 2.

Thus, there are at least as many different countable order types as there are sequences $a \in \{0,1\}^{\mathbb{N}}$, namely c.

Definition 4.8 Let $(M, \leq_M)$ and $(N, \leq_N)$ be chains representing the order types μ and ν. We define the product $\mu\nu$ to be the order type of $(N \times M, \leq)$, where $(n_1, m_1) \leq (n_2, m_2)$ iff $n_1 <_N n_2$ or ($n_1 = n_2$ and $m_1 \leq_M m_2$).

This ordering $\leq$ on $N \times M$ is called the lexicographic (or dictionary) order since the pairs are ordered "alphabetically."

Example 4.9

a) $2 \cdot \omega_0 = \omega_0$. To see this, we can represent ω_0 by $\mathbb{N}$ and 2 by $\{0,1\}$. Then the chain representing $2 \cdot \omega_0$, listed in increasing (lexicographic) order, is $\mathbb{N} \times \{0,1\} = \{(1,0),\ (1,1),\ (2,0),\ (2,1),\ (3,0),\ (3,1),\ ...,\ (n,0),\ (n,1),\ ...\}$, and this chain is order isomorphic to $\mathbb{N}$. More generally, $n \cdot \omega_0 = \omega_0$ for each $n \in \mathbb{N}$.

However, $\omega_0 \cdot 2 \neq \omega_0$. The product $\omega_0 \cdot 2$ is represented by $\{0,1\} \times \mathbb{N}$ ordered as:

$$\{(0,1),\ (0,2), ..., (0,n), ...\ \ ;\ \ (1,1), (1,2), ..., (1,n), ...\}$$

This chain is not order isomorphic to $\mathbb{N}$: $\mathbb{N}$ has only one element with no "immediate predecessor," while this set has two such elements. In fact, $\omega_0 \cdot 2 = \omega_0 + \omega_0$. Thus, multiplication of order types is <u>not commutative</u>.

b) $\omega_0 = 2 \cdot \omega_0 = (1+1) \cdot \omega_0 \neq 1 \cdot \omega_0 + 1 \cdot \omega_0 = \omega_0 + \omega_0$. Thus the "right distributive" law fails.

c) ω_0^2 can be represented by the lexicographically ordered chain $\mathbb{N} \times \mathbb{N}$. In order of increasing size, this chain is:

$$\{(1,1),\ (1,2), ...\ ;\ (1,n), ...\ ;\ (2,1),\ (2,2),\ ...,\ (2,n), ...\ ;\ \ ...\ \ ;\ ...\ ;\ (n,1),\ (n,2), ...\ ,\ \ ;\ \ ...\}$$

The chain looks like countably many copies of $\mathbb{N}$ placed end-to-end, so we can also write: $\omega_0^2 = \omega_0 + \omega_0 + ... + \omega_0 + ...\ = \sum_{n \in \mathbb{N}} \mu_n$ where each $\mu_n = \omega_0$. A subset of $\mathbb{Q}$ that represents ω_0^2 is $\{k - \frac{1}{n} :\ k, n = 1, 2, 3, ...\}$.

Exercise Prove that

1) multiplication of order types is associative

2) the left distributive law holds for order types: $\mu(\nu + t) = \mu\nu + \mu\tau$.

Note: Other books may define $\mu\nu$ "in reverse" as the order type of the lexicographically ordered set $M \times N$. Under that definition, $\omega_0 \cdot 2 = \omega_0$ and $2\omega_0 = \omega_0 + \omega_0 \neq \omega_0$. Also, under the "reversed" definition, the right distributive law holds but the left distributive law fails.

Which way the definition is made is not important mathematically. The arithmetic of order types under one definition is just a "mirror image" of the arithmetic under the other definition. You just need to be aware of which convention a writer is using.

Exercises

E7. Give $[0,1]^2$ the lexicographic order $\leq$, and let (a,b) represent an open interval in $[0,1]^2$. Describe what a "small" open interval around each of the following <u>points</u> looks like: $(0,\frac{1}{2})$, $(0,1)$, $(\frac{1}{2},0)$, $(1,0)$.

E8. For each $a \in \mathbb{N}$, we can write a uniquely in the form $a = 2^r(2s+1)$ for integers $r, s \geq 0$. Suppose $a' = 2^{r'}(2s'+1) \in \mathbb{N}$. Define $a \leq a'$ if $a = a'$, or $r < r'$, or $r = r'$ and $s < s'$. What is the order type of ($\mathbb{N}, \leq$)? Does a nonempty subset of ($\mathbb{N}, \leq$) necessarily contain a smallest element?

E9. Show that it is impossible to define an order $\leq$ on the set $\mathbb{C}$ of complex numbers in such a way that all three of the following are true:

i) for all $z, w \in \mathbb{C}$, exactly one of $x = w$, $x < w$, or $z > w$ holds
ii) for all $u, w, z \in \mathbb{C}$: if $z < w$, then $z + u < w + u$
iii) if x , $y > 0$, then $xy > 0$

(Hint: Begin by showing that if such an order exists, then $-1 > 0$. *But notice that this, in itself, is not a contradiction.*)

E10. Give explicit examples of subsets of $\mathbb{Q}$ which represent the order types:

a) $\omega_0 + 1 + \omega_0$

b) $\omega_0^2 + \omega_0$

c) $\omega_0^2 + 2 \cdot \omega_0 + 3$

d) $\omega_0^2 + \omega_0 \cdot 2 + 3$

e) $\omega_0^2 + \omega_0^2$

f) $\omega_0^3 + \omega_0$

E11. Let η be the order type of $\mathbb{Q}$.

a) Give an example of a set $B \subseteq \mathbb{Q}$ such that neither B nor $\mathbb{Q} - B$ has order type η.

b) Prove that if A is a set of type η and $B \subseteq A$, then either B or $A - B$ contains a subset of type η.

c) Prove or disprove: $\eta + 1 + \eta = \eta$.

Hint: Cantor's characterization of $\mathbb{Q}$ *as an ordered set may be helpful.*

E12. A chain $(X, \leq)$ is called an η_1-set if the following condition holds in X :

(*) Whenever A and B are countable subsets of X such that $a < b$ for every choice of $a \in A$ and $b \in B$, then $\exists c \in X$ such that $a < c < b$ for all $a \in A$ and all $b \in B$

More informally, we could paraphrase condition (*) as: for countable subsets A,B of X, "$A < B$" $\Rightarrow$ $\exists c \in X$ such that "$A < c < B$"

a) Show that $\mathbb{R}$ is not an η_1-set.

b) Prove that every η_1-set is uncountable.
Hint: there is a one line argument; note that $\emptyset$ *is countable.*

c) By b), an η_1-set $(X, \leq)$ satisfies $|X| \geq \aleph_1$, and so $|X| \geq c$ if we assume the continuum hypothesis, CH. Prove that $|X| \geq c$ without assuming CH.
Hint: show how to define a one-to-one function $f : \mathbb{R} \to X$; *begin by defining* f *on* $\mathbb{Q}$.

More generally, a chain $(X, \leq)$ *is called an* η_α*-set if, whenever* A *and* B *are subsets of* X, *both of cardinality* $< \aleph_\alpha$, *then there is a* $x \in X$ *such that "*$A < x < B$*". So, for example, an* η_0*-set is simply an order dense chain.*

5. Well-Ordered Sets and Ordinal Numbers

We now look at a much stronger kind of order $\leq$ on a set.

Definition 5.1 A poset $(X, \leq)$ is called <u>well-ordered</u> if every nonempty subset of X contains a smallest element.

The definition implies that a well-ordered set X is automatically a chain: if $a \neq b \in X$, then set $\{a, b\}$ has a smallest element, so either $a \leq b$ or $b \leq a$.

$\mathbb{N}$ and all its subsets are well-ordered. The set of integers, $\mathbb{Z}$, is not well-ordered since, for example, $\mathbb{Z}$ itself contains no smallest element. $\mathbb{R}$ is not well-ordered since, for example, the nonempty interval $(0, 1)$ contains no smallest element.

Since a well-ordered set X is a chain, it has an order type. These special order types are very nicely behaved and have a special name.

Definition 5.2 An <u>ordinal</u> <u>number</u> (or simply <u>ordinal</u>) is the order type of a well-ordered set.

Since we know how to add and multiply order types, we already know how to add and multiply ordinals and get new ordinals. We also have a Definition 4.2 for $<$ and $\leq$ that applies to ordinals.

Theorem 5.3 If α and β are ordinals, so are $\alpha + \beta$ and $\alpha \cdot \beta$.

Proof Let α and β be represented by disjoint well-ordered sets A and B. Then $\alpha + \beta$ is represented by the ordered sum $A + B$. We must show this set is well-ordered. Since $A + B$ is a chain, we only need to check that a nonempty subset C of $A + B$ must contain a smallest element.

B is well-ordered so, if $C \subseteq B$, then C has a smallest element. Otherwise $C \cap A \neq \emptyset$ and, since A is well-ordered, there is a smallest element $c \in C \cap A$. In that case c is the smallest element of C.

Similarly, we need to show that the lexicographically ordered product $B \times A$ is well-ordered. If C is a nonempty subset of $B \times A$, let b_0 be the smallest first coordinate of a point in C : more precisely, let b_0 be the smallest element in $\{b \in B : \text{for some } a \in A,\ (b, a) \in C\}$. Then let a_0 be the smallest element in $\{a \in A : (b_0, a) \in C\}$. Then (b_0, a_0) is the smallest element in C. (*Intuitively,* (b_0, a_0) *is the point at the "lower left corner" of* C. *The fact that* B *and* A *are well-ordered guarantees that such a point exists.*) $\bullet$

Some examples of ordinals (increasing in size) are

$$0, 1, 2, \dots, \omega_0, \omega_0 + 1, \omega_0 + 2, \dots, \omega_0 + n, \dots, \omega_0 \cdot 2, \omega_0 \cdot 2 + 1, \dots, \omega_0 \cdot 2 + n, \dots,$$

$$\dots \omega_0 \cdot 3, \dots, \omega_0 \cdot n, \dots, \omega_0^2, \omega_0^2 + 1, \dots, \omega_0^2 + \omega_0, \omega_0^2 + \omega_0 + 1, \dots, \omega_0^2 + \omega_0 \cdot 2, \dots$$

All these ordinals can be represented by countable well-ordered sets (in fact, by subsets of $\mathbb{Q}$) so we refer to them as "countable ordinals." We will see later (assuming AC), there are well-ordered sets of arbitrarily large cardinality $-$ so that this list of ordinals barely scratches the surface.

Exercise 5.4 Find a subset of $\mathbb{Q}$ that represents the ordinal $\omega_0^2 + \omega_0 \cdot 3 + 2$.

Here are a few very simple properties of well-ordered sets. Missing details should be checked as exercises.

1) In a well-ordered set X, each element a except the largest (if there is one) has an "immediate successor"— namely, the smallest element of the nonempty set $\{x \in X : x > a\}$. However, an element in a well-ordered set might not have an immediate predecessor: for example in $\{1 - \frac{1}{n} : n \in \mathbb{N}\} \cup \{2 - \frac{1}{n} : n \in \mathbb{N}\} \cup \{2\}$, neither 1 nor 2 has an immediate predecessor. This set represents the ordinal $\omega_0 + \omega_0 + 1$.

2) A subset of a well-ordered set, with the inherited order, is well-ordered.

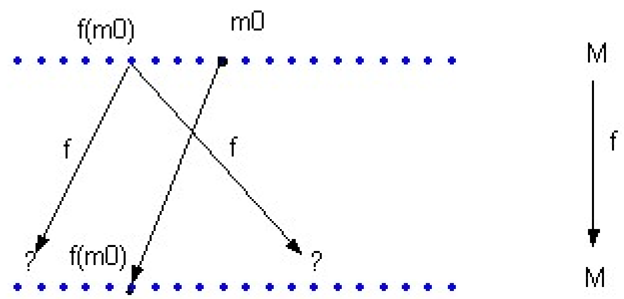
3) Order isomorphisms preserve well-ordering: if a poset is well-ordered, so is any order isomorphic poset. An order isomorphism preserves the smallest element in any nonempty subset.

The following theorems indicate how order isomorphisms between well-ordered sets are much less "flexible" than isomorphisms between chains in general. .

Theorem 5.5 Suppose M is well-ordered. If $f : M \to M$ is a one-to-one, order-preserving map of M into M, then $f(m) \geq m$ for all $m \in M$.

The theorem says that f cannot move an element "to the left." Notice that Theorem 5.5 is false for chains in general: for example, consider $M = \mathbb{R}$ and $f(x) = \frac{1}{2}x$. On the other hand, if you try to construct a counterexample using $M = \mathbb{N}$, you will probably see how the proof of the theorem should go.

Proof Suppose not. Then $A = \{m \in M : f(m) < m\} \neq \emptyset$. Let m_0 be the smallest element in A. We get a contradiction by asking "what is $f(f(m_0))$"?

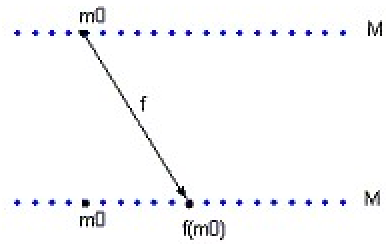


Since $m_0 \in A$, $f(m_0) < m_0$. Since f preserves order and is one-to-one, $f(f(m_0)) < f(m_0)$ which means that $f(m_0) \in A$. But that is impossible because m_0 is the smallest element of A. •

Corollary 5.6 If M is well-ordered, then the only order isomorphism f from M onto M is the identity map $f(m) = m$.

(*Note that the theorem is false for chains in general: if $M = \mathbb{R}$, then $f(x) = x^3$ is an onto order isomorphism.* Corollaries *5.6 and 5.7 indicate that well-ordered sets have a very "rigid" structure.*)

Proof Let $f: M \to M$ be an order isomorphism. By Theorem 5.5, $f(m) \geq m$ for all m. If f is not the identity, then $A = \{m : f(m) > m\} \neq \emptyset$. Let m_0 be the least element of A.



Then m_0 cannot be in ran(f) (*why?*). •

Corollary 5.7 Suppose M and N are well-ordered. If $f : M \to N$ and $g : M \to N$ are order isomorphisms from M onto N, then $f = g$. (*So M and N can be order isomorphic "in only one way."*)

Proof If $f \neq g$, then $f^{-1}f$ and $g^{-1}f$ are two different order isomorphisms from M onto M, and that is impossible by the preceding corollary. •

Exercise 5.8 Find two different order isomorphisms between $\mathbb{R}$ and the set of positive reals $\mathbb{R}^+$.

We have already seen that order types, in general, are not very nicely behaved. Therefore, during this the following discussion about well-ordered sets and ordinal numbers, there is a certain amount of fussiness in the notation – to make sure we do not jump to any false conclusions. Much of this fussiness will drop by the wayside as things become clearer.

In a nonempty well-ordered set M, we will often refer to the smallest element as 0. (*In fact, without loss of generality, we can literally assume* 0 *<u>is</u> the smallest element* M.) If we need to carefully distinguish between the first elements in two well-ordered sets M, N we may write them as 0_M and 0_N. (*This might be necessary if, say,* $N \subseteq M$ *and the smallest elements of* N *and* M *are different.*) But usually this degree of care is not necessary.

Definition 5.9 Suppose $m \in M$, where M is well-ordered. <u>The initial segment of M determined by m</u> $= \{x \in M : x < m\}$. We can write this set using the "interval notation" $[0_M, m)_M$.
If a discussion involves only a single well-ordered set M, we may simply write $[0, m)_M$ or even just $[0, m)$.

Notice that:

i) For every $m \in M$, $M \neq [0, m)$, so M is not an initial segment of itself, and we will see in Theorem 5.10 that much more is true: M cannot even be order isomorphic to an initial segment of itself.

ii) Order isomorphisms preserve initial segments: if $f : M \to N$ is an order isomorphism of M onto N, then $f[[0, m)] = [0, f(m))$

iii) Given any two initial segments in a well-ordered set M, one of them is an initial segment of the other. More precisely, if $m < n \in M$, then "the initial segment in M determined by m" $= \{x \in M : x < m\} = \{x \in [0, n) : x < m\} =$ "the initial segment in $[0, n)$ determined by m."

Theorem 5.10 Suppose M is well-ordered and $N \subseteq M$. M is not order isomorphic to an initial segment of N. In particular (*when* $N = M$), M is not order isomorphic to an initial segment of itself.

Proof Suppose $n \in N \subseteq M$. If $f : M \to N \subseteq M$ is one-to-one and order preserving, then Theorem 5.5 gives us that $f(n) \geq n$ for each $n \in N$. Therefore, for each n, $\text{ran}(f) \neq [0_N, n)$. •

Corollary 5.11 No two initial segments of M are order isomorphic (so each initial segment of M, as well as M itself, represents a different ordinal).

Proof One of the two segments is an initial segment of the other, so by Theorem 5.10 the segments cannot be order isomorphic. •

Definition 5.12 Suppose μ and ν are ordinals represented by the well-ordered sets M and N. We say that $\mu < \nu$ if M is order isomorphic to an initial segment of N – that is $M \simeq [0, n)$ for some $n \in N$. If M is order isomorphic to N we write $\mu = \nu$. We write $\mu \leq \nu$ if $\mu < \nu$ or $\mu = \nu$. (*Check that the definition is independent of the choice of well-ordered sets M and N representing* μ *and* ν.)

> *Note: We already have a different definition (4.2) for* $\mu \leq \nu$ *when we think of* μ *and* ν *as arbitrary order types. It will turn out – for ordinals – that the two definitions are equivalent, that is:*
>
> *for well-ordered sets* M *and* N:
>
> M *is order isomorphic to a proper subset of* N *but not to* N *itself*
> $$\Updownarrow (*)$$
> M *is order isomorphic to an initial segment of* N.
>
> *The equivalence* (*) *is not true for chains in general: for example, each of* $[0, 1]$ *and* $(0, 1)$ *is order isomorphic to a subset of the other, but neither is order isomorphic to an initial segment of the other (why?)*
>
> *Until Corollary 5.19, where we prove the equivalence* (*), *we will be using the new definition 5.12 of* $\mu \leq \nu$ *for ordinals.*

The relation $\leq$ among ordinals is clearly reflexive and transitive. The next theorem implies that $\leq$ is antisymmetric $-$ and therefore any set of ordinals is partially ordered by $\leq$.

Theorem 5.13 If μ and ν are ordinals, then at most one of the relations $\mu < \nu$, $\mu = \nu$, and $\mu > \nu$ can be true.

Proof Let M and N represent μ and ν. If $\mu = \nu$, then $M \simeq N$. In this case, $\mu < \nu$ and $\nu < \mu$ are impossible since a well-ordered set cannot be order isomorphic to an initial segment of itself (Theorem 5.10).

If $\mu < \nu$, then M is isomorphic to an initial segment of N. If $\nu < \mu$ were also true, then N would, in turn, be isomorphic to an initial segment of M. By composing these isomorphisms, we would have M order isomorphic to an initial segment of itself $-$ which is impossible. •

Notation For an ordinal μ, let ord $(\mu) = \{\alpha : \alpha$ is an ordinal and $\alpha < \mu\}$. If $\mu = 0$, then ord $(\mu) = \emptyset$ and if $\mu > 0$, then $0 \in$ ord(μ), so ord$(\mu) \neq \emptyset$. Like any set of ordinals, we know that ord(μ) is partially ordered by $\leq$.

It turns out that much more is true: every set of ordinals is actually well-ordered by $\leq$, but to see that takes a few more theorems. However, ord(μ) is a very special set of ordinals and, for starters, Theorem 5.14 tells us that ord(μ) is well-ordered by $\leq$. Theorem 5.14 also gives us a very nice "standard" way to pick a well-order that represents an ordinal μ.

Theorem 5.14 If μ is an ordinal represented by the well-ordered set M, then ord$(\mu) \simeq M$. Therefore ord(μ) is a well-ordered set of ordinals and ord(μ) represents μ.

Proof For each ordinal $\alpha \in$ ord(μ), we have $\alpha < \mu$, so α can be represented by some initial segment $[0, m_\alpha)$ of M. Define f : ord$(\mu) \to M$ by $f(\alpha) = m_\alpha$. This function f is one-to-one since different ordinals cannot be represented by the same initial segment of M, and f clearly preserves order.

If $m \in M$, then the initial segment $[0, m)$ in M represents some ordinal $\alpha < \mu$. But α is represented by $[0, m_\alpha)$. Since different initial segments of M are not isomorphic, we get $m = m_\alpha = f(\alpha)$ so f is onto. Therefore ord$(\mu) \simeq M$. •

We will often write ord(μ) in "interval" notation:

For an ordinal μ, $[0, \mu) = \{\alpha : \alpha$ is an ordinal and $\alpha < \mu\} =$ ord(μ)

By 5.14, $[0, \mu)$ is well-ordered and represents the ordinal μ; therefore any ordinal μ can be represented by the set of preceding ordinals.

For example,

0 is represented by the set of preceding ordinals, namely $[0,0) = \text{ord}(0) = \emptyset$

1 is represented by $[0,1) = \text{ord}(1) = \{0\}$

2 is represented by $[0,2) = \text{ord}(2) = \{0,1\}$

$\vdots$

ω_0 is represented by $[0,\omega_0) = \text{ord}(\omega_0) = \{0,1,2,...,n,...\}$

$\omega_0 + 1$ is represented by $[0,\omega_0+1) = \{0,1,2,...,n,... \,;\ \omega_0\}$
(*here, "* ; *" indicates that* ω_0 *comes "after" all the natural numbers* n)

$\vdots$

α is represented by the set of previously defined ordinals

etc.

$\vdots$

Some comments about axiomatics

The informal definition of ordinals is good enough for our purposes, However, the preceding list roughly illustrates how one can define ordinals in axiomatic set theory ZFC. For example, in ZFC the ordinal 2 is defined by $2 = \{0,1\}$ (rather than saying that the set $\{0,1\}$ represents the ordinal 2).

Definition

$$\begin{aligned}
0 &= \emptyset \\
1 &= \{0\} = \{\emptyset\} \\
2 &= \{0,1\} = \{\emptyset, \{\emptyset\}\} \\
&\vdots \\
\omega_0 &= \{0,1,2,...,n,...\} \\
\omega_0 + 1 &= \{0,1,2,...,n,...;\omega_0\} \\
&\vdots \\
&\text{etc.}
\end{aligned}$$

and in general, an ordinal $\alpha =$ *the set of previously defined ordinals. Of course, this presentation is still a little vague: in particular, some sort of "induction" in ZFC is needed to justify the "etc." where an ordinal is defined in terms of ordinals already defined.*

Once we have defined ordinals (as sets) in ZFC, we need to say how they are compared, that is, how to define $\leq$. We do this for ordinals α and β by writing $\alpha < \beta$ iff $\alpha \in \beta$. This seems to accomplish what we want. For example:

$1 < 3$ because $1 \in 3$

$0 < 1 < 2 < 3 < ... < \omega_0 < \omega_0 + 1$ because $0 \in 1 \in 2 \in 3 \in ... \in \omega_0 \in \omega_0 + 1 \in ...$

$\omega_0 < \omega_0 + 17$ because $\omega_0 \in \omega_0 + 17$

etc.

If X is any set well-ordered by $\leq$, we can then define its ordinal number ($=$ "the ordinal number associated with X") as follows: from the axioms ZFC one can prove the existence of a function (set) with domain X that is defined "recursively" by :

$$\forall\, y \in X \;\; f(y) = \text{ran}\,(\, f|\{z \in X : z < y\})$$

Then "the ordinal number of X" is defined to be the set ran(f).

For example, for the well-ordered set $X = \{1, 3, 5\}$, what is the function f and what is the ordinal number of X?

$$f(1) = \text{ran}\,(\, f|\{z \in X : z < 1\}) = \text{ran}\,(f|\emptyset\,) = \emptyset$$

$$f(3) = \text{ran}\,(\, f|\{z \in X : z < 3\}) = \text{ran}\,(f|\{1\}) = \{\emptyset\}$$

$$f(5) = \text{ran}\,(\, f|\{z \in X : z < 5\}) = \text{ran}\,(f|\{1,3\}) = \{\emptyset, \{\emptyset\}\}$$

The ordinal number of X is ran $(f) = \{0,\, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = \{0, 1, 2\} = 3$

In axiomatic set theory, cardinals are viewed as certain special ordinals: an ordinal α is called a cardinal if for all ordinals $\beta < \alpha$ there is no bijection between β and α – that is a cardinal is an "initial ordinal" – meaning that it's "the first ordinal with a given size." From that point of view ω_0 is a cardinal because there is a bijection between ω_0 and n for any $n < \omega_0$. Earlier, we gave this cardinal the name $\aleph_0$. But $\omega_0 + 1$ is not a cardinal because there is a bijection between ω_0 and $\omega_0 + 1$.

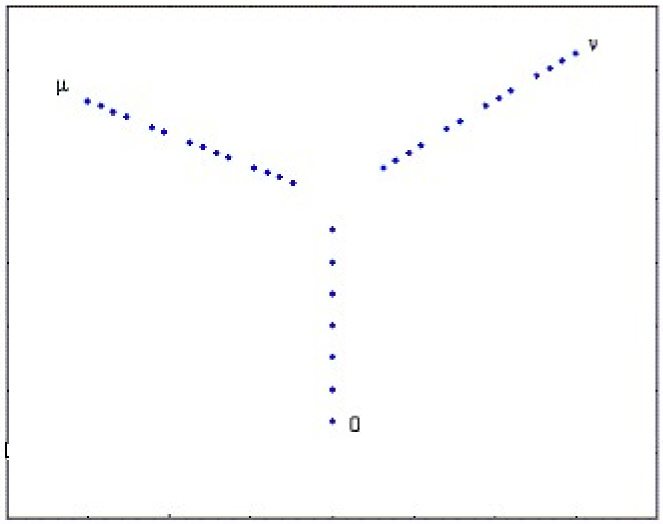
In Theorem I.13.2, we proved that at most one of the relations $<$, $=$, $>$ (as defined for cardinal numbers) can hold between two cardinals. (*Context will make clear whether "* $\leq$ *" refers to the ordering of cardinals or ordinals.*) We also stated in Chapter I that (assuming AC) at least one of the relations $<$, $=$, $>$ must hold between any two cardinals – that is, for any two sets one must be equivalent to a subset of the other. We are almost ready to prove that statement – in fact, this statement about cardinal numbers follows easily (assuming AC) from the corresponding result about ordinal numbers which we now prove.

Theorem 5.15 (Ordinal Trichotomy Theorem) If μ and ν are ordinals, then at least one of the relations $\mu > \nu$, $\mu = \nu$, $\mu < \nu$ must hold (and so, by Theorem 5.13, exactly one of these relations holds).

Proof We already know that certain special sets of ordinals are well-ordered: for example $[0, \mu) = \{\alpha : \alpha$ is an ordinal and $\alpha < \mu\}$ (Theorem 5.14).

The theorem is certainly true if $\mu = 0$ or $\nu = 0$, so we assume both $\mu > 0$ and $\nu > 0$.

Let $D = [0, \mu) \cap [0, \nu) = \{\alpha : \alpha$ is an ordinal for which $\alpha < \mu$ and $\alpha < \nu\}$. Because we do not yet know that μ and ν are comparable, the situation might look something like the following:



D is well-ordered because $D \subseteq [0, \mu)$, and $D \neq \emptyset$ (since $0 \in D$). Therefore D represents some ordinal $\delta > 0$. We claim that $\delta \leq \mu$ and $\delta \leq \nu$.

1) To show that $\delta \leq \mu$, we assume $\delta \neq \mu$ and prove $\delta < \mu$.

$[0, \mu)$ represents μ. Since $D \subseteq [0, \mu)$ and D represents $\delta \neq \mu$, we conclude $D \neq [0, \mu)$, so $[0, \mu) - D \neq \emptyset$. Let γ be the smallest element in $[0, \mu) - D$. $[0, \gamma)$ is an initial segment of $[0, \mu)$.

We claim $D = [0, \gamma)$. If that is true, then D represents γ; but D represents δ and therefore $\delta = \gamma < \mu$.

$[0, \gamma) \subseteq D$: If $\alpha \in [0, \gamma)$, then $\alpha \in [0, \mu)$ and $\alpha < \gamma =$ smallest element in $[0, \mu) - D$, so $\alpha \in D$.

$D \subseteq [0, \gamma)$: If $\alpha \in D$, then α and γ are comparable since both are in the well-ordered set $[0, \mu)$.

We examine the possibilities:

i) $\gamma = \alpha$: impossible, since $\alpha \in D$ and $\gamma \notin D$

ii) $\gamma < \alpha$: impossible, since that would mean $\gamma < \alpha < \mu$ and $\gamma < \alpha < \nu$, forcing $\gamma \in [0, \mu) \cap [0, \nu)$ $= D$ – which is false.

Therefore $\alpha < \gamma$, so $\alpha \in [0, \gamma)$.

2) A similar argument (interchanging "μ" and "ν" throughout) shows that if $\delta \neq \nu$, then $\delta = \gamma < \nu$.

Since $\delta \leq \mu$ and $\delta \leq \nu$, there are only four possibilities:

a) $\delta < \mu$ and $\delta < \nu$, in which case $\delta \in [0, \mu) \cap [0, \nu) = D = [0, \delta)$ – which is impossible.

Therefore one of the remaining three cases must be true:

b) $\delta = \mu$ and $\delta = \nu$, in which case $\mu = \nu$
c) $\delta < \mu$ and $\delta = \nu$, in which case $\mu > \nu$
d) $\delta = \mu$ and $\delta < \nu$, in which case $\mu < \nu$ •

Corollary 5.16 Any set of ordinals is linearly ordered with respect to the ordinal ordering $\leq$. (W*e shall see in Theorem 5.20 that even more is true: every set of ordinals is well-ordered.*)

We simply state the following theorem. A proof of the equivalences can be found, for example, in *Set Theory and Metric Spaces* (Kaplansky) or *Topology* (Dugundji).

Theorem 5.17 The following statements are equivalent. (Moreover, each is consistent with and independent of the axioms ZF for set theory:)

1) (**Axiom of Choice**) If $\{A_\alpha : \alpha \in A\}$ is a family of pairwise disjoint nonempty sets, there is a set $B \subseteq \bigcup A_\alpha$ such that, for all α, $|A_\alpha \cap B| = 1$.

(*This is clearly equivalent to the statement that* $\prod A_\alpha \neq \emptyset$. *If f is in the product, let $B = ran(f)$; on the other hand, if such a set B exists, define f by $f(\alpha) =$ the unique element of $A_\alpha \cap B$. An element $f \in \prod A_\alpha$ is a function that "chooses" one element $f(\alpha)$ from each A_α.*)

2) (**Zermelo's Theorem**) Every set can be well-ordered, i.e., for every set X there is a subset $\leq$ of $X \times X$ such that $(X, \leq)$ is well-ordered.

3) (**Zorn's Lemma**) Suppose $(X, \leq)$ is a nonempty poset. If every chain in X has an upper bound in X, then X contains a maximal element.

We will look at some powerful uses of Zorn's Lemma later. For now, we are mainly interested in Zermelo's Theorem.

It is tradition to call 2) Zermelo's <u>Theorem</u> and 3) Zorn's <u>Lemma</u>. They appear as "proven" results in the early literature, but the "proofs" used some form of the Axiom of Choice (AC). Generally, we have been casual about mentioning when AC is being used. However in the following theorems, for emphasis, $[AC]$ indicates that the Axiom of Choice is used in one of these equivalent forms.

Corollary 5.18 [AC, Cardinal Trichotomy] If m and n are cardinal numbers, <u>at</u> <u>least</u> <u>one</u> (and thus, by Theorem I.13.2 , exactly one) of the relations $m > n$, $m = n$, $m < n$ holds. (*Therefore any set of cardinals is a chain.*)

Proof According to Zermelo's Theorem, we may assume that M and N are well-ordered sets representing the cardinals m and n, so that M and N also represent ordinals μ and ν.
By Theorem 5.15, either $\mu = \nu$, $\mu < n$, or $\mu > \nu$; therefore either M and N are order isomorphic (in which case $m = n$) or one is order isomorphic to an initial segment of the other (so $m < n$ or $n < m$). •

In fact, the Cardinal Trichotomy Corollary, is equivalent to the Axiom of Choice. (See Gillman, *Two Classical Surprises Concerning the Axiom of Choice and the Continuum Hypothesis,* Am. Math. Monthly 109(6), 2002, pp. 544-553 for this and other interesting results that do not depend on techniques of axiomatic set theory.) Over 200 equivalents to the Axiom of Choice are given in *Equivalents of the Axioms of Choice* (Rubin & Rubin, North-Holland Publishing, 1963).

The following corollary tells us that, among ordinals, the two definitions of " $\leq$ " (Definition 4.2 and Definition 5.12) are equivalent.

Corollary 5.19 Suppose M and N are well-ordered sets representing μ and ν. If M is order isomorphic to a subset of N (*so $\mu \leq \nu$ in the sense of Definition 4.2*), then M is order isomorphic to N or to an initial segment of N (*so $\mu \leq \nu$ in the sense of Definition 5.12*).

Proof Without loss of generality, we may assume $M \subseteq N$. If M is not order isomorphic to N, then $\mu \neq \nu$. If M is also not isomorphic to an initial segment of N, then $\mu < \nu$ is also false. Therefore the Trichotomy Theorem 5.15 gives $\mu > \nu$. Then N is order isomorphic to an initial segment of a subset M of itself – which violates Theorem 5.10. •

Theorem 5.20 Every set W of ordinals is well-ordered. In particular, every nonempty set of ordinals contains a smallest element.

Proof Theorem 5.15 implies that W is linearly ordered by $\leq$. We need to show that if A is a nonempty subset of W, then A contains a smallest element. Pick $\alpha \in A$. If α is itself the smallest in A, we are done. If not, then $\{\beta \in A : \beta < \alpha\}$ is nonempty and well-ordered – because it is a subset of $[0, \alpha)$ – so it contains a smallest element β_0, and β_0 is the smallest element in W. •

Corollary 5.21 [AC] Every set C of cardinal numbers is well-ordered. In particular, every nonempty set of cardinal numbers contains a smallest element.

Proof We know that the order $\leq$ (among cardinals) is a linear order. Let D be a nonempty subset of C and, for each cardinal $m \in D$, let M represent m. By Zermelo's Theorem, each set M can be well-ordered, after which it represents some ordinal μ. By Theorem 5.20, the set of all such μ's contains a smallest element μ_0 represented by M_0. Then $m_0 = |M_0|$ is clearly the smallest cardinal in D. •

Example 5.22

1) The set $C = \{m : m$ is a cardinal and $\aleph_0 < m \leq c\}$ has a smallest element. It is called the immediate successor of $\aleph_0$ and is denoted by $\aleph_0^+$ or $\aleph_1$. The statement $c = \aleph_1$ is the Continuum Hypothesis which, we recall, is independent of the axioms ZFC. If CH is assumed as an additional axiom in set theory, then $C = \{\aleph_1\} = \{c\}$.

2) More generally, for any cardinal m we can consider the smallest element m^+ in the set $\{k : k$ is a cardinal and $m < k \leq 2^m\}$. We call m^+ the immediate successor of m. In particular, we write $\aleph_1^+ = \aleph_2$, $\aleph_2^+ = \aleph_3$, and so on.

The Generalized Continuum Hypothesis is the statement

GCH : "for every infinite cardinal m, $m^+ = 2^m$."

GCH and $\sim$ GCH are equally consistent with the axioms ZFC. (*Curiously, ZF + GCH implies AC. This is discussed in the Gillman article cited after Corollary 5.18.*)

3) In Example VI.4.6, we (provisionally) defined the weight of a topological space $(X, \mathcal{T})$ by

$$w(X) = \aleph_0 + \min\{|\mathcal{B}| : \mathcal{B} \text{ is a base for } \mathcal{T}\}.$$

We now see that the definition makes sense $-$ because there must exist a base of smallest cardinality.

Theorem 5.23 If W is a set of ordinals, then there exists an ordinal greater than any ordinal in W. (*Therefore there is no "set of all ordinals."*)

Proof Let $W^* = \{\mu + 1 : \mu \in W\}$, and represent each ordinal $\mu + 1$ in W^* by a well-ordered set $M_{\mu+1}$. We may assume the $M_{\mu+1}$'s are pairwise disjoint. (*If not, replace each $M_{\mu+1}$ with $M_{\mu+1} \times \{\mu + 1\}$, ordered in the obvious way.*) Form the ordered sum: that is, let $S = \bigcup\{M_{\mu+1} : \mu \in W\}$, and order S by

$$x \leq y \text{ if } \begin{cases} x, y \in M_{\mu+1}, \text{ and } x \leq y \text{ in } M_{\mu+1} \\ x \in M_{\mu+1},\ y \in M_{\nu+1} \text{ and } \mu < \nu \end{cases}$$

Clearly, $\leq$ well-orders S, so $(S, \leq)$ represents an ordinal σ. Since each $M_{\mu+1} \subseteq S$, we have $\mu + 1 \leq \sigma$ by Theorem 5.19. Since $\mu < \mu + 1 \leq \sigma$ for each $\mu \in W$, σ is larger than any ordinal in W. •

Corollary 5.24 Every set W of ordinals has a least upper bound, denoted $\sup W$ $-$ that is, there is a smallest ordinal $\geq$ every ordinal in W.

Proof Without loss of generality we may assume that if $\alpha < \mu \in W$, then $\alpha \in W$ (*why? and where is this assumption used in what follows?*). If W contains a largest element, then it is the least upper bound. Otherwise, pick an ordinal σ larger than every ordinal in W. Then $[0, \sigma + 1) - W \neq \emptyset$ (it contains σ) and the smallest element in $[0, \sigma + 1) - W$ is $\sup W$. •

Example 5.25

1) sup $\{0, 1, 2, ..., \} = \omega_0$, and sup$\{0, 1, 2, ..., \omega_0\} = \omega_0$

2) We say that an ordinal "has cardinal m" if it is represented by a well-ordered set with cardinality m. In particular, countable ordinals are those represented by countable well-ordered sets.

3) sup $\{\alpha : \alpha$ is a countable ordinal$\}$ is called ω_1. Since there is no largest countable ordinal (*why?*), we see that ω_1 is the smallest uncountable ordinal. A set representing ω_1 must have cardinal $\aleph_1$ (*why?*). Since $[0, \omega_1)$ represents ω_1, so there are exactly $\aleph_1$ ordinals $< \omega_1$, that is, exactly $\aleph_1$ countable ordinals. Each countable ordinal can be represented by a subset of $\mathbb{Q}$ (*see Example 4.7*), so there are exactly $\aleph_1$ nonisomorphic well-ordered subsets of $\mathbb{Q}$.

Since ω_1 is the smallest uncountable ordinal, each $\alpha < \omega_1$ is a countable ordinal that is represented by $[0, \alpha)$. Therefore α has only countably many predecessors and ω_1 is the first ordinal with uncountably many ($\aleph_1$) predecessors.

The spaces $[0, \omega_1)$ and $[0, \omega_1 + 1) = [0, \omega_1]$, with the order topology, have some interesting properties that we will look at later. These properties hinge on the fact that ω_1 is the smallest uncountable ordinal.

The ordinals ω_0, ... , $\omega_0 + n$, ... , $\omega_0 \cdot 2$, ... , ω_0^2, ... , ω_0^n, ... are all mere countable ordinals. For ordinals α, β it is possible to define "ordinal exponentiation" α^β. (*The definition is sketched in the appendix at the end of this chapter.*) Then it turns out that $\omega_0^{\omega_0}$, $\omega_0^{\omega_0} + 1$, ... , $\omega_0^{(\omega_0^{\omega_0})}$, ... , are still countable ordinals. If you accept that, then you should also believe that $\epsilon_0 = \sup\{\omega_0, \omega_0^{\omega_0}, \omega_0^{\omega_0^{\omega_0}}, \omega_0^{\omega_0^{\omega_0^{\omega_0}}}, ... \}$ (= "ω_0 to the ω_0 power ω_0 times") is still countable. Roughly, each element in the set has only countably many predecessors and the set has only countable many elements, so the least upper bound of the set still has only countably many predecessors—namely, all the predecessors of its predecessors.

But once you get up to ϵ_0, you can then form $\epsilon_0^{\epsilon_0}$, $\epsilon_0^{(\epsilon_0^{\epsilon_0})}$, ..., take the least upper bound again, and still have only a countable ordinal ϵ_1. And so on. So ω_1, the first ordinal with uncountably many predecessors is way beyond all these: "the longer you look at ω_1, the farther away it gets" (*Robert McDowell*).

We now look at some similar results for cardinals.

Lemma 5.26 If $\{k_\alpha : \alpha \in A\}$ and $\{m_\alpha : \alpha \in A\}$ are sets of cardinals and $k_\alpha \geq m_\alpha$ for each $\alpha \in A$, then $\sum k_\alpha \geq \sum m_\alpha$. (*An infinite sum of cardinals is defined in the obvious way: if the K_α's are pairwise disjoint sets with cardinality k_α, then $\sum k_\alpha = |\bigcup K_\alpha|$.*)

Proof Exercise

Theorem 5.27 If a set of cardinals $C = \{k_\alpha : \alpha \in A\}$ contains no largest element, then $\sum k_\alpha > k_{\alpha_0}$ for each $\alpha_0 \in A$.

Proof For any particular $\alpha_0 \in A$, let $\begin{cases} m_{\alpha_0} = k_{\alpha_0} \\ m_\alpha = 0 \text{ for } \alpha \neq \alpha_0 \end{cases}$.
Then the lemma gives $\sum k_\alpha \geq \sum m_\alpha = k_{\alpha_0}$.

If $\sum k_\alpha = k_{\alpha_0}$ for some α_0, k_{α_0} would be the largest element in C which, by hypothesis, does not exist. Therefore $\sum k_\alpha > k_{\alpha_0}$ for every $\alpha_0 \in A$. •

The conclusion <u>*may*</u> *be true even if C has a largest element: for example, suppose* $C = \{1, 2\}$.
Can you give an example involving an infinite set C of cardinals?

Corollary 5.28 If C is a set of cardinals, then there is a cardinal m larger than every member of C (and therefore there is no "set of all cardinals").

Proof If C has a largest element k, then let $m = 2^k$. Otherwise, use the preceding theorem and let $m = \sum\{k : k \in C\}$. •

Corollary 5.29 Every set C of cardinal numbers has a least upper bound, that is, there is a smallest cardinal $\geq$ every cardinal in C.

Proof Without loss of generality, we may assume that if $p \in C$, then all cardinals smaller than p are also in C (*why? and where is this used in what follows?*). If C has a largest element k, then k is the least upper bound. Otherwise, pick a cardinal m greater than all the cardinals in C and let $S = \{p : p \text{ is a cardinal and } p \leq m\}$. Then $S - C \neq \emptyset$ (it contains m) and the smallest cardinal in this set is the least upper bound for C. •

6. Indexing the Infinite Cardinals

By Corollary 5.21, the set of <u>infinite</u> cardinals less than a given cardinal k is well-ordered, so this set is order isomorphic to an initial segment of ordinals. Therefore this set of cardinals can be "faithfully indexed" by that segment of ordinals – that is, indexed in such a way that $k_\alpha < k_\beta$ iff $\alpha < \beta$. When the infinite cardinals are listed in order of increasing size and indexed by ordinals, they are denoted by $\aleph$'s with ordinal subscripts. In this notation, the first few infinite cardinals are

$$\aleph_0, \aleph_1\ (= \aleph_0^+), \aleph_2\ (= \aleph_1^+), \ldots, \aleph_n, \ldots, \aleph_{\omega_0}, \aleph_{\omega_0+1}, \ldots, \aleph_{\omega_0+\omega_0}, \ldots, \aleph_{\omega_0^2}, \ldots, \aleph_{\epsilon_0}, \aleph_{\epsilon_0+1}, \ldots, \aleph_{\omega_1}, \ldots.$$

Thus, $\aleph_{\omega_0} = \sup\{\aleph_n : n < \omega_0\}$ and $\aleph_{\omega_1} = \sup\{\aleph_\alpha : \alpha < \omega_1\}$. $\aleph_{\omega_1}$ is the first cardinal with uncountably many ($\aleph_1$) cardinal predecessors.

In this notation, GCH states that for every ordinal α, $2^{\aleph_\alpha} = \aleph_{\alpha+1}$.

By definition, $c = |\mathbb{R}|$. So where is c is this list of cardinals? The continuum hypothesis states "$c = \aleph_1$." However, it is a fact that for each ordinal α, there exists an ordinal $\beta > \alpha$ such that the assumption "$\aleph_\beta = c$" is consistent with ZFC. (*Perhaps* $\mathbb{R}$ *is more mysterious than you thought.*)

These results imply that not even the simplest exponentiation $2^{\aleph_0}$ involving an infinite cardinal can be "calculated" in ZFC: is $2^{\aleph_0} = \aleph_1?$ $= \aleph_{17}?$ $= \aleph_{\omega_0+1}?$

On the other hand, one <u>cannot</u> consistently assume "$c = \aleph_\alpha$" for an arbitrary choice of α: even in ZFC, certain $\aleph_\alpha$'s are provably excluded. In fact, if $\aleph_\alpha$ is the least upper bound of a strictly increasing sequence of smaller cardinals, we will <u>prove</u> $\aleph_\alpha^{\aleph_0} > \aleph_\alpha$ (and therefore $\aleph_\alpha \neq c$). *It is also true, but we will not prove it, that these are the <u>only</u> excluded $\aleph_\alpha$'s – it is consistent to assume $2^{\aleph_0} = \aleph_\alpha$ for any cardinal $\aleph_\alpha$ for which $\aleph_\alpha^{\aleph_0} = \aleph_\alpha$ (Solovay, 1965).*

At the heart of what we need is a classical theorem about cardinal arithmetic.

Theorem 6.1 (König) Suppose that for each $\alpha \in A$, m_α and n_α are cardinals with $m_\alpha < n_\alpha$. Then $\sum_\alpha m_\alpha < \prod_\alpha n_\alpha$.

Proof Proving " $\leq$ " is straightforward; proving " $<$ " takes a little more work.

Let sets M_α and N_α represent m_α and n_α. We may assume the N_α's are pairwise disjoint and, since $m_\alpha < n_\alpha$, that each M_α is a <u>proper</u> subset of N_α. For each α, pick and fix an element $n_\alpha \in N_\alpha - M_\alpha$ and define $f : \bigcup M_\alpha \to \prod N_\alpha$ by:

$$\text{for } x \in M_{\alpha_0},\ f(x) = z \in \prod N_\alpha, \text{ where } z(\alpha) = \begin{cases} x & \text{if } \alpha = \alpha_0 \\ n_\alpha & \text{if } \alpha \neq \alpha_0 \end{cases}$$

The M_α's are disjoint so f is well defined, and clearly f is one-to-one, so we conclude that $\sum_\alpha m_\alpha \leq \prod_\alpha n_\alpha$.

We now show that $\sum m_\alpha = \prod n_\alpha$ is impossible. We do this by showing that if $h : \bigcup M_\alpha \to \prod N_\alpha$ is one-to-one, then h cannot be onto.

Let $P = \text{ran}(h) = h[\bigcup M_\alpha] = \bigcup h[M_\alpha]$. Let $h[M_\alpha] = P_\alpha$. Since h is one-to-one, $|P_\alpha| = m_\alpha$ so, for each α, there are at most m_α different α^{th} coordinates of points in P_α – that is, $|\{z_\alpha : z \in P_\alpha\}| \leq m_\alpha < n_\alpha$. Then we can pick a point $w_\alpha \in N_\alpha$ with $w_\alpha \neq z_\alpha$ for every $z \in P_\alpha$, i.e., w_α is not the α-th coordinate of any point in P_α.

Define $w \in \prod N_\alpha$ by $w(\alpha) = w_\alpha$. Then $w \notin P_\alpha$ for all α, so $w \notin P = \text{ran}(h)$. •

Example 6.2 Suppose we have a strictly increasing <u>sequence</u> of cardinals

$$0 \neq m_0 < m_1 < \ldots < m_k < \ldots\ .$$

For each k, let $n_k = m_{k+1}$, so $m_k < n_k$. By König's Theorem, $\sum_{k=0}^{\infty} m_k < \prod_{k=0}^{\infty} n_k = \prod_{k=1}^{\infty} m_k \leq \prod_{k=0}^{\infty} m_k$.

In particular: if $m_k = \aleph_k$, then $\sum_{k=0}^{\infty} \aleph_k < \prod_{k=0}^{\infty} \aleph_k$. Since $\aleph_{\omega_0} \leq \sum_{k=0}^{\infty} \aleph_k$ (*why?*), we have

$$\aleph_{\omega_0} < \prod_{k=0}^{\infty} \aleph_k \leq \aleph_{\omega_0}^{\aleph_0}.$$

Since $c = c^{\aleph_0}$, we conclude that $c \neq \aleph_{\omega_0}$.

Note: A similar argument shows that if a cardinal k is the least upper bound of a sequence of strictly increasing cardinals, then $k^{\aleph_0} > k$, so $k \neq c$.

Exercise 6.3 For a cardinal k, there are how many ordinals with cardinality $\leq k$?

7. Spaces of Ordinals

Let X be a set of ordinals with the order topology. Since X is well-ordered, X is order isomorphic to some initial segment of ordinals $[0, \alpha)$. Therefore X and $[0, \alpha)$ are homeomorphic in their order topologies. Therefore to think about "spaces of ordinals" we only need look at initial segments of ordinals $[0, \alpha)$. We will look briefly at some general facts about these spaces. In Section 8, we will consider the spaces $[0, \omega_1)$ and $[0, \omega_1 + 1) = [0, \omega_1]$ in more detail. These particular spaces have some interesting properties that arise from the fact that ω_1 is the first uncountable ordinal.

According to Definition 3.1, a subbase for the order topology on $X = [0, \alpha)$ consists of all sets

$$\{x \in [0, \alpha) : x < \gamma\} = [0, \gamma), \quad \gamma < \alpha$$

and

$$\{x \in [0, \alpha) : x > \beta\} = (\beta, \alpha), \quad \beta \geq 0$$

The set of finite intersections of such sets is a base, so (*check this!*) a basic open set has one of the following forms:

$[0, \gamma)$, where $\gamma \leq \alpha$ ($\gamma = \alpha$ corresponds to the empty intersection)
(β, γ), where $\beta \geq 0$ and $\gamma \leq \alpha$

If $\alpha = 0$, then $X = [0, \alpha) = \emptyset$; so suppose $\alpha > 0$. What does an efficient neighborhood base at a point $\tau \in [0, \alpha)$ look like? .

If $\tau = 0$: $\{0\} = [0, 1)$ is open in $[0, \alpha)$. Therefore 0 is an isolated point in $[0, \alpha)$ and $\{\{0\}\}$ is an open neighborhood base at 0.

If $0 < \tau < \alpha$, then any basic open set containing τ must contain a set of the form $(\sigma, \tau]$:

if $\tau \in [0, \gamma)$, then $\tau \in (0, \tau] \subseteq [0, \gamma)$
if $\tau \in (\beta, \gamma)$, then $\tau \in (\beta, \tau] \subseteq (\beta, \gamma)$

Each set $(\sigma, \tau] = (\sigma, \tau + 1)$ is open; and each set $(\sigma, \tau]$ is also closed because its complement $[0, \sigma + 1) \cup (\tau, \alpha)$ is open. Therefore $\{(\sigma, \tau] : 0 \leq \sigma < \tau\}$ is a neighborhood base of clopen sets at τ.

Putting together these open neighborhood bases, we get that

$$\mathcal{B} = \{\{0\}\} \cup \{(\sigma, \tau] : 0 \leq \sigma < \tau < \alpha\}$$

is a clopen base for the topology.

We noted in Example 3.4 that any chain with the order topology is Hausdorff. Therefore every ordinal space $[0, \alpha)$ is Hausdorff. Since there is a neighborhood base of closed (in fact, clopen) neighborhoods at each point $\tau \in [0, \alpha)$, we know even more: Theorem VII.2.7 tells us that $[0, \alpha)$ is a T_3-space. But still more is true.

Theorem 7.1 For any ordinal α, $[0, \alpha)$ is T_4.
As remarked earlier, every chain *with the order topology is T_4 – but the proof is much simpler for well-ordered sets.*

Proof We know that $[0, \alpha)$ is T_1, so need to prove that $[0, \alpha)$ is normal. Suppose A and B are disjoint closed sets in $[0, \alpha)$.

If $\tau \in A$, let $U_\tau =$ a basic open set of form $(\sigma, \tau]$ disjoint from B (or, $U_\tau = \{0\}$ if $\tau = 0$)
If $\tau \in B$, let $V_\tau =$ a basic open set of form $(\sigma, \tau]$ disjoint from A (or, $V_\tau = \{0\}$ if $\tau = 0$)

We claim that if $\tau_1 \in A$ and $\tau_2 \in B$, then $U_{\tau_1} \cap V_{\tau_2} = \emptyset$:

The statement is clearly true if τ_1 or $\tau_2 = 0$ so suppose both are > 0. Then $U_{\tau_1} = (\sigma_1, \tau_1]$ and $V_{\tau_2} = (\sigma_2, \tau_2]$. We can assume without loss of generality that $\tau_1 < \tau_2$.
If $(\sigma_1, \tau_1] \cap (\sigma_2, \tau_2] \neq \emptyset$, then we have $\tau_1 \in (\sigma_2, \tau_2]$ which would mean that $V_{\tau_2} \cap A \neq \emptyset$.

If $U = \bigcup_{\tau \in A} U_\tau$ and $V = \bigcup_{\tau \in B} V_\tau$, then U and V are disjoint open sets with $U \supseteq A$ and $V \supseteq B$. •

(*Why doesn't the same proof work for chains with the order topology?*)

Definition 7.2 An ordinal β is called a limit ordinal if $\beta > 0$ and β has no immediate predecessor; β is called a nonlimit ordinal if $\beta = 0$ or β has an immediate predecessor (that is, $\beta = \gamma + 1$ for some ordinal γ).

Example 7.3

1) If β is a limit ordinal in $[0, \alpha)$, then for all $\sigma < \beta$, $(\sigma, \beta] \neq \{\beta\}$. Therefore β is not isolated in the order topology. If β is a nonlimit ordinal, then $\{\beta\} = \{0\}$ or, for some γ, $\{\beta\} = (\gamma, \beta]$. Either way, $\{\beta\}$ is open so β is isolated. Therefore the isolated points in $[0, \alpha)$ are the exactly the points that are not limit ordinals.

2) $[0, \omega_0)$ is discrete: it is homeomorphic to $\mathbb{N}$.

3) $[0, \omega_0 + 1)$ is homeomorphic to $\{1 - \frac{1}{n} : n \in \mathbb{N}\} \cup \{1\}$.

4) For what α's is $[0, \alpha)$ connected?

Theorem 7.4 Suppose $\alpha > 0$. The ordinal space $W = [0, \alpha)$ is compact iff $[0, \alpha) = [0, \beta]$ for some ordinal β – that is, iff W contains a largest element β.

Proof Suppose $[0, \alpha)$ has no largest element. Then $\mathcal{U} = \{[0, \gamma) : \gamma < \alpha\}$ is an open cover of $[0, \alpha)$. The sets in $\mathcal{U}$ are nested so, if there were a finite subcover, there would be a single set $[0, \gamma)$ covering $[0, \alpha)$. That is impossible since $\gamma \in [0, \alpha) - [0, \gamma)$.

Conversely, suppose $[0, \alpha) = [0, \beta]$. If $\beta = 0$, then $[0, \beta] = \{0\}$ is compact, so we assume $\beta > 0$.

Let $\mathcal{U}$ be an open cover of $[0, \beta]$. We can assume $\mathcal{U}$ consists of basic open sets, that is, sets of the form $\{0\}$ or $(\sigma, \tau]$. Necessarily, then, $\{0\} \in \mathcal{U}$.

Let $\beta_1 = \beta > 0$. For some $\sigma_1 < \beta_1$, we have a set $(\sigma_1, \beta_1] \in \mathcal{U}$. If $\sigma_1 = 0$, then $\{\{0\}, (\sigma_1, \beta_1]\}$ is a finite subcover.

If $\sigma_1 > 0$, then for some β_2, $\sigma_1 \in (\sigma_2, \beta_2] \in \mathcal{U}$, so $\sigma_2 < \sigma_1 \leq \beta_2$. If $\sigma_2 = 0$, then $\{\{0\}, (\sigma_2, \beta_2], (\sigma_1, \beta_1]\}$ is a finite subcover.

We proceed inductively. Having chosen $(\sigma_k, \beta_k]$ so that $\sigma_{k-1} \in (\sigma_k, \beta_k]$: if $\sigma_k > 0$ we can choose $(\sigma_{k+1}, \beta_{k+1}] \in \mathcal{U}$ so that $\sigma_k \in (\sigma_{k+1}, \beta_{k+1}]$.

We continue until $\sigma_n = 0$ occurs – and this must happen in a finite number of steps because otherwise, we would have defined an infinite descending sequence of ordinals $\sigma_1 > \sigma_2 > \sigma_3 > \ldots > \sigma_n > \ldots$. This is impossible because a well-ordered set $\{\sigma_1, \sigma_2, \ldots, \sigma_n, \ldots\}$ must have a smallest element.

When we reach $\sigma_n = 0$, we have a finite subcover from $\mathcal{U}$: $\{\{0\}, (\sigma_n, \boldsymbol{\beta_n}], \ldots, (\sigma_1, \beta_1]\}$. •

Example 7.5 $[0, \omega_0)$ is not compact, but $[0, \omega_0 + 1) = [0, \omega_0]$ is compact. In fact, $[0, \omega_0]$ is homeomorphic to $\{1\} \cup \{1 - \frac{1}{n} : n \in \mathbb{N}\}$.

8. The Spaces $[0, \omega_1)$ and $[0, \omega_1]$

Theorem 8.1 For each $n < \omega_0$, suppose $\alpha_n < \omega_1$. Then $\alpha = \sup\{\alpha_n : n < \omega_0\} < \omega_1$ – that is, the sup of a countable set of countable ordinals is countable.

Proof $[0, \alpha_n)$ is countable so $\bigcup_{n<\omega_0}[0, \alpha_n)$ is countable. Since ω_1 has uncountably many predecessors, there is an ordinal $\gamma \in [0, \omega_1) - \bigcup_{n<\omega_0}[0, \alpha_n)$. Then $\gamma > \alpha_n$ for each n, so
$\alpha = \sup\{\alpha_n : n < \omega_0\} \leq \gamma < \omega_1$. •

Corollary 8.2 $[0, \omega_1]$ and $[0, \omega_1)$ are not separable.

Proof If D is a countable subset of $[0, \omega_1]$, then $\alpha = \sup(D - \{\omega_1\}) < \omega_1$. Then
$\operatorname{cl} D \subseteq [0, \alpha] \cup \{\omega_1\} \neq [0, \omega_1]$.

A dense set in $[0, \omega_1)$ is also dense in $[0, \omega_1]$, so $[0, \omega_1)$ is not separable. •

Corollary 8.3 In $[0, \omega_1]$, no sequence from $[0, \omega_1)$ can converge to ω_1.

Proof Suppose (α_n) is a sequence in $[0, \omega_1)$. Let $\alpha = \sup\{\alpha_n : n \in \mathbb{N}\} < \omega_1$. The α_n's are all in the closed set $[0, \alpha]$, so $(\alpha_n) \not\to \omega_1$. •

Example 8.4 **In** Example III.9.8, we saw a rather complicated space L in which sequences are not sufficient to describe the topology. Corollary 8.3 gives an example that may be easier to "see" : $\omega_1 \in \text{cl}\,[0, \omega_1)$, but no sequence (α_n) in $[0, ,\omega_1)$ can converge to ω_1.

This implies that $[0, \omega_1]$ is not first countable. Of course, $\{(\sigma, \tau] : \sigma < \tau\}$ is a countable neighborhood base at each point $\tau < \omega_1$ – so the "problem" point is ω_1. Here, the neighborhood poset $\mathcal{N}_{\omega_1}$ (ordered by reverse inclusion) is very nicely ordered (well-ordered, in fact!) but the chain of all neighborhoods is simply "too long" and we cannot "thin it out" to get a countable neighborhood base at ω_1.

In contrast, the basic neighborhoods of $(0,0)$ in the space L were very badly "entangled". The neighborhood system $\mathcal{N}_{(0,0)}$ had a very complicated order structure – too complicated for us to find a countable subset of $\mathcal{N}_{(0,0)}$ that "goes arbitrarily far out" in the poset. (*See discussion in Example 2.4.6*).

In Theorem IV.8.11 we proved that certain implications hold between various "compactness-like" properties in a topological space X.

$$(*) \begin{cases} X \text{ is compact} \\ \qquad \text{or} \\ X \text{ is sequentially compact} \end{cases} \Rightarrow X \text{ is countably compact} \Rightarrow X \text{ is pseudocompact.}$$

We asserted that, in general, no other implications are valid. The following corollary shows that "sequentially compact" $\not\Rightarrow$ "compact" (and therefore "countably compact" $\not\Rightarrow$ "compact" and "pseudocompact" $\not\Rightarrow$ "compact").

Corollary 8.5 $[0, \omega_1)$ is sequentially compact.

Proof Suppose (α_n) is a sequence in $[0, \omega_1)$. We need to show that (α_n) has a convergent subsequence in $[0, \omega_1)$. Without loss of generality, we may assume that all the α_n's are distinct (*why?*) The sequence (α_n) has either an increasing subsequence $\alpha_{n_1} < \alpha_{n_2} < \ldots < \alpha_{n_k} < \ldots$ or a decreasing subsequence $\alpha_{n_1} > \alpha_{n_2} > \ldots > \alpha_{n_k} > \ldots$

> The argument is completely parallel to the one in Lemma IV.2.10 showing that a sequence in $\mathbb{R}$ has a monotone subsequence. Call α_n a peak point of the sequence if $\alpha_n \geq \alpha_k$ for all $k \geq n$.
>
> If the sequence has only finitely many peak points, then after some α_{n_1} there are no peak points and we can choose an increasing subsequence $\alpha_{n_1} < \alpha_{n_2} < \ldots < \alpha_{n_k} < \ldots$

If (α_n) has infinitely many peak points, then we can choose a subsequence of peak points $\alpha_{n_1}, ...\alpha_{n_k}, ...$ and for these, $\alpha_{n_1} \geq \alpha_{n_2} \geq ... \geq \alpha_{n_k} \geq ...$. Since the α_n's are distinct, we have $\alpha_{n_1} > \alpha_{n_2} > ... > \alpha_{n_k} > ...$.

However, a strictly decreasing sequence of ordinals $\alpha_{n_1} > \alpha_{n_2} > ... > \alpha_{n_k} > ...$ is impossible. Therefore (α_n) has a subsequence of the form $\alpha_{n_1} < \alpha_{n_2} < ... < \alpha_{n_k} < ...$. Setting $\alpha = \sup\{\alpha_{n_k} : k = 1, 2, ...\} < \omega_1$, we see that then $(\alpha_n) \rightarrow \alpha \in [0, \omega_1)$. •

An example of a pseudocompact space that is not countably compact is given in Exercise E31.. In Chapter X, we will discuss a space $\beta\mathbb{N}$ that is compact (therefore countably compact and pseudocompact) but not sequentially compact (see Example X.6.5). That will complete the set of examples showing that "no other implications exist" other than those stated in (*).

Corollary 8.6 In $[0, \omega_1]$, the intersection of a countable collection of neighborhoods of ω_1 is again a neighborhood of ω_1 $-$ that is, the intersection must contain a "tail" $[\alpha, \omega_1]$. Therefore $\{\omega_1\}$ is not a G_δ-set (and therefore not a zero set) in $[0, \omega_1]$.

Proof Let $\{N_k : k = 1, 2, ...\}$ be a collection of neighborhoods of ω_1. For each k, there is a α_k for which $\omega_1 \in (\alpha_k, \omega_1] \subseteq \text{int}\, N_k \subseteq N_k$. Let $\tau = \sup\{\alpha_k : k \in \mathbb{N}\} < \omega_1$. Then $\bigcap_{k=1}^{\infty} N_k \supseteq \bigcap_{k=1}^{\infty} \text{int}\, N_k \supseteq (\tau, \omega_1] = [\alpha, \omega_1]$ where $\alpha = \tau + 1$. In particular, $\bigcap_{k=1}^{\infty} N_k \neq \{\omega_1\}$. •

Since $[0, \omega_1]$ is compact, we know that each $f \in C([0, \omega_1])$ is bounded. In fact, something more is true. (*Why does the corollary state something "more" ?*)

Corollary 8.7 If $f \in C([0, \omega_1])$, then f is constant on a "tail" $[\alpha, \omega_1]$ for some $\alpha < \omega_1$.

Proof Suppose $f(\omega_1) = r$. By Corollary 8.6, $f^{-1}[\{r\}] = \bigcap_{n=1}^{\infty} f^{-1}[(r - \frac{1}{n}, r + \frac{1}{n})]$ contains a tail $[\alpha, \omega_1]$. Therefore $f|[\alpha, \omega_1] = r$. •

Proving Corollary 8.7 was relatively easy because we can see immediately what the constant value must be to make the theorem true: $r = f(\omega_1)$. A more remarkable thing is that the same result holds for $[0, \omega_1)$. But to prove that fact, we have no "initial guess" about what constant value f might have on a tail, so we have to work harder.

Theorem 8.8 If $f \in C([0, \omega_1))$, then f is constant on the "tail" $[\tau, \omega_1)$ for some $\tau < \omega_1$.

Proof Let $T_\alpha = [\alpha, \omega_1) =$ "the α^{th} tail". By Corollary 8.5, $[0, \omega_1)$ is countably compact so the closed set T_α is also countably compact. It is easy to see that a continuous image of a countably compact space is countably compact, so $f[T_\alpha]$ is a countably compact subset of $\mathbb{R}$. Since countable compactness and compactness are equivalent for subsets of $\mathbb{R}$ (Theorem IV.8.17), each $f[T_\alpha]$ is a nonempty compact set: $f[T_\alpha] \subseteq f[T_0] \subseteq \mathbb{R}$.

The T_α's are nested : $f[T_\alpha] \subseteq f[T_\beta] \subseteq f[T_0]$ if $\beta < \alpha$. Therefore $f[T_\alpha]$'s have the finite intersection property, and by compactness $\bigcap_{\alpha < \omega_1} f[T_\alpha] \neq \emptyset$ (*see Theorem IV.8.4*). In fact, we claim the intersection contains a single number : $\bigcap_{\alpha < \omega_1} f[T_\alpha] = \{r\}$.

If $r, s \in \bigcap_{\alpha<\omega_1} f[T_\alpha]$, then f assumes each of the values r, s for arbitrarily large values of α. Therefore we can pick an increasing sequence $\alpha_1 < \beta_1 < \alpha_2 < \beta_2 < \ldots < \alpha_n < \beta_n < \ldots$ such that $f(\alpha_n) = r$ and $f(\beta_n) = s$. Let $\gamma = \sup\{\alpha_n, \beta_n : n \in \mathbb{N}\} < \omega_1$. Then $(\alpha_n) \to \gamma$ and $(\beta_n) \to \gamma$. By continuity $(f(\alpha_n)) = (r) \to \gamma$ and $(f(\beta_n)) = (s) \to \gamma$. and we conclude $\gamma = r = s$.

We claim that $f \equiv r$ on some tail of $[0, \omega_1)$. First notice that if $\beta < \omega_1$, then $\exists\, \gamma_n > \beta$ for which $f[T_{\gamma_n}] \subseteq (r - \frac{1}{n}, r + \frac{1}{n})$.

If not, then $\{\, f[T_\gamma] : \ \beta < \gamma < \omega_1 \,\} \cup \{\, f[[0,\omega_1)] - (r - \frac{1}{n}, r + \frac{1}{n}) : n \in \mathbb{N} \,\}$ would be a family of closed subsets of $f[T_0]$ with the finite intersection property, so this family would have a nonempty intersection. That is impossible since $\bigcap_{\beta<\gamma<\omega_1} f[T_\gamma] = \{r\}$ and $r \notin \bigcap_{n=1}^\infty f[[0,\omega_1)] - (r - \frac{1}{n}, r + \frac{1}{n})$.

Pick γ_1 so that $f[T_{\gamma_1}] \subseteq (r-1, r+1)$. Pick $\gamma_2 > \gamma_1$ so that $f[T_{\gamma_2}] \subseteq (r - \frac{1}{2}, r + \frac{1}{2})$ and continue inductively to pick $\gamma_{n+1} > \gamma_n$ so that $f[T_{\gamma_{n+1}}] \subseteq (r - \frac{1}{n+1}, r + \frac{1}{n+1})$.

Let $\tau = \sup\{\gamma_n : n \in \mathbb{N}\} < \omega_1$. Then $f[T_\tau] \subseteq f[\bigcap T_{\gamma_n}] \subseteq \bigcap f[T_{\gamma_n}] \subseteq \bigcap_{n=1}^\infty (r - \frac{1}{n}, r + \frac{1}{n}) = \{r\}$, so $f|T_\tau = r$. •

(*Since f is bounded on the compact set $[0,\tau]$, we see in a different way that $[0,\omega_1)$ is pseudocompact.*)

Corollary 8.9 Every continuous function $f : [0,\omega_1) \to \mathbb{R}$ can be extended in a unique way to a continuous function $F : [0,\omega_1] \to [0,\omega_1]$.

Proof For some $r \in \mathbb{R}$, $f = r$ on a tail $[\alpha, \omega_1)$. Let $F|[0,\omega_1) = f$ and define $F(\omega_1) = r$. Any continuous extension G of f must agree with F since F and G agree on the dense set $[0,\omega_1)$.•

Note: $[0,\omega_1]$ *is a compact* T_2 *space which contains* $[0,\omega_1)$ *as a dense subspace. We call* $[0,\omega_1]$ *a compactification of* $[0,\omega_1)$. *The property stated in Corollary 8.9 is a very special property for a compactification to have : in fact, it characterizes* $[0,\omega_1]$ *as the so-called Stone-Cech compactification of* $[0,\omega_1)$. *We will discuss compactifications in Chapter 10.*

By way of contrast, notice that $[-1,0]$ *is compactification of* $[-1,0)$ *which, just as above, is obtained by adding a single point to the original space. However, the continuous function* $f : [-1,0) \to \mathbb{R}$ *defined by* $f(x) = \sin(\frac{1}{x})$ *cannot be continuously extended to the point* 0.

We saw in Example VII.5.10 that a subspace of a normal space need not be normal: the Sorgenfrey plane X is not normal however it can be embedded in the T_4-space $[0,1]^m$ for some m. Any space X that is $T_{3\frac{1}{2}}$ but not T_4 works just as well. However, these examples are not very explicit – it is hard to "picture why" the normality of $[0,1]^m$ isn't inherited by the subspace X. The "picturing" may be easier in the following example.

Example 8.10 Let $T^* = [0,\omega_1] \times [0,\omega_0]$, sometimes called the "Tychonoff plank." T^* is compact T_2 and therefore T_4. Discarding the "upper right corner point," we are left with the (open) subspace $T = T^* - \{(\omega_1,\omega_0)\}$ We claim that T is not normal. Let

$A = \{(\omega_1, n) \in T \ : n < \omega_0\} =$ "the right edge of T" and
$B = \{(\alpha,\, \omega_0) \in T : \alpha < \omega_1\} =$ "the top edge of T"

A and B are disjoint sets and closed in T (although not, of course, in T^*).

Suppose U is an open set in T containing the "right edge" A. For each point $(\omega_1, n) \in A \subseteq U$, we can choose a basic open set $(\alpha_n, \omega_1] \times \{n\} \subseteq U$. Let $\alpha = \sup\{\alpha_n : n = 0, 1, 2...\} < \omega_1$. Then $(\alpha, \omega_1] \times \{n\} \subseteq U$ for all n – that is, A is contained in a "vertical strip" $(\alpha, \omega_1] \times [0, \omega_0)$ inside U.

Suppose V is an open set in T containing the "top edge" B. Since $(\alpha+1, \omega_0) \in B$, there is a basic open set $\{\alpha+1\} \times (n, \omega_0] \subseteq V$. But then $(\alpha+1, n+1) \in U \cap V$, so $U \cap V \neq \emptyset$.

Exercise 8.11 Show that every continuous function $f : T \to \mathbb{R}$ can be continuously extended to a function $F : T^* \to \mathbb{R}$.
Hint: $f|[0,\omega_1) \times \{\omega_0\}$ *has constant value* r *on some tail, and for each* $n < \omega_0$, $f|\,[0,\omega_1] \times \{n\}$ *has a constant value* r_n *on a tail. Define* $F((\omega_1,\omega_0)) = r$. *Prove that* $(r_n) \to r$ *and then show that the extension* F *is continuous at* (ω_1, ω_0).

As in the remark following Corollary 8.9, T^* *is a compactification of* T *and the functional extension property in the exercise characterizes* T^* *as the so-called Stone-Cech compactification of* T. *Since* T^* *is compact,* F *must be bounded – so, in retrospect,* f *must have been bounded in the first place. Therefore* T *is pseudocompact. But* T *is not countably compact – because the "right edge"* A *is an infinite set that has no limit point in* T. T *is an example of* *a pseudocompact space that is not countably compact.*

Exercises

E13. Suppose $C \subseteq \mathbb{R}$ and that C is well-ordered (in the usual order on $\mathbb{R}$). Prove that C is countable.

E14. Let ω_0^* denote the order type of the set of nonpositive integers, with its usual ordering.

a) Prove that a chain $(X, \leq)$ is well-ordered iff X contains no subset of order type ω_0^*.

b) Prove that if $(X, \leq)$ is a chain in which every countable subset is well-ordered, then X is well-ordered.

c) Prove that every infinite chain either has a subset of order type ω_0^* or one of order type ω_0.

E15. Prove the following facts about ordinal numbers α, β, γ :

a) if $\beta > 0$, then $\alpha + \beta > \alpha$

b) if $\alpha > \beta$, then there exists a unique γ such that $\alpha = \beta + \gamma$
We might try using b) to define subtraction of ordinals: $\alpha - \gamma = \beta$ if $\alpha = \beta + \gamma$. However this is perhaps not such a good idea. (Consider $\omega_0 = \alpha$, $\beta = 1$). Problems arise because ordinal addition is not commutative.

c) $1 + \alpha = \alpha$ iff $\alpha \geq \omega_0$.

E16. Let $B \subseteq A$, where $(A, \leq)$ be a chain. B is called inductive if

$$\text{for all } t \in A, \quad \{a \in A : a < t\} \subseteq B \Rightarrow t \in B.$$

Prove that if A is the only inductive subset of A, then A is well-ordered.

E17. Let X be a first countable space. Suppose that for each $\alpha < \omega_1$, F_α is a closed subset of X and that $F_{\alpha_1} \subseteq F_{\alpha_2}$ whenever $\alpha_1 \leq \alpha_2 < \omega_1$. Prove that $\bigcup\{F_\alpha : \alpha < \omega_1\}$ is closed in X.

E18. Let A be well-ordered. Order $L = A \times [0,1)$ lexicographically and give the set the order topology.

a) What does a "nice" neighborhood base look like at each point in L? Discuss some other properties of this space.

b) If $A = [0, \omega_1)$, then $L - \{(0,0)\}$ is called is the "long line." Show that L is path connected and locally homeomorphic to $\mathbb{R}$ but it cannot be embedded in $\mathbb{R}$. (See *Topology*, J. Munkres, 2nd edition, p. 159 for an outline of a proof.)

c) Each point in L homeomorphic to $\mathbb{R}$. L is normal but not metrizable.

E19. a) Let A and B be disjoint closed sets in $[0, \omega_1]$. Prove that at least one of A and B is compact and bounded away from ω_1. (*A set C is bounded away from ω_1 if $C \subseteq [0, \alpha]$ for some $\alpha < \omega_1$.*)

b) Characterize the closed sets in $[0, \omega_1)$ that are zero sets.

c) Prove that $[0, \omega_1)$ and $[0, \omega_1]$ are not metrizable.

E20. a) Suppose (x_n) and (y_n) are sequences in $X = [0, \omega_1)$ such that, for all n, $x_n \leq y_n \leq x_{n+1}$. Show that both sequences converge and have the same limit.

b) Show that if $f : X \to X$ is such that $f(x) \geq x$ for every x, then there is an x such that (x, x) is a limit point of the graph of f in $X \times X$.

c) Prove that $X \times [0, \omega_1]$ is not normal.

(*Hint: Let* Δ *be the diagonal of* $X \times X$ *and* $B = X \times \{\omega_1\}$. *Show that if U and V are open with* $\Delta \subseteq U$ *and* $B \subseteq V$, *then* $U \cap V \neq \emptyset$. *To do this: if any point* (x, ω_1) *is in U, we're done. So suppose this is false and define* $f(x)$ *to be the least ordinal* $> x$ *such that* $(x, f(x)) \notin U$. *Use part b).*)

E21. A space X is called σ-compact if X can be written as a countable union of compact sets.

a) Prove $[0, \omega_1)$ is not σ-compact.
b) Using part a) (or otherwise), prove $[0, \omega_1)$ is not Lindelöf.
c) State and prove a theorem of the form: an ordinal space $[0, \alpha)$ is σ-compact iff ...

(*You might begin by thinking about the spaces* $[0, \omega_1)$, $[0, \omega_2)$, *and* $[0, \omega_{\omega_0})$.)

E22. A space X is called <u>functionally</u> <u>countable</u> if every continuous $f : X \to \mathbb{R}$ has a countable range.

a) Show that $X = [0, \omega_1)$ is functionally countable.

b) Let $Y = D \cup \{p\}$ where D is uncountable and $p \notin D$. Give Y the topology for which all the points of D are isolated and for which the basic neighborhoods of p are those cocountable sets containing p. Show Y is functionally countable.

c) Prove that $X \times Y$ is not functionally countable.

(*Hint: Let* $g : X \to D$ *be one-to-one. Consider the set* $H = \{(\alpha, g(\alpha) : \alpha$ *is isolated in* $X\} \subseteq X \times Y$)

E23. A space X is called <u>Y-compact</u> if X is homeomorphic to a closed subspace of the product Y^m for some cardinal m. For example, X is $[0,1]$-compact iff X compact and T_2. An $\mathbb{R}$-compact space is called <u>realcompact</u>.

a) Prove that if X is both realcompact and pseudocompact, then X is compact.

(*Note: the converse is clear*)

b) Suppose that X is Tychonoff and that however X is embedded in $[0,1]^m$, its projection in every direction is compact. (More precisely, suppose that for <u>all</u> possible embeddings $h : X \to [0,1]^m$, we have that $\pi_\alpha[h[X]]$ is compact for all projections π_α. This statement is certainly true, for example, if X is compact.) Prove or disprove that X must be compact.

E24. An infinite cardinal k is called <u>sequential</u> if $k = \sum_{n=1}^{\infty} k_n$ for some sequence of cardinals $k_n < k$.

a) Prove that if k is sequential, then $k^{\aleph_0} > k$.
b) Assume GCH. Prove that if k is infinite and $k^{\aleph_0} > k$, then $k^{\aleph_0} = 2^k$.
c) Assume GCH. Prove that an infinite cardinal k is sequential iff $k^{\aleph_0} > k$.

9. Transfinite Induction and Transfinite Recursion

"...to understand recursion, you have to understand recursion..."

Suppose we have a sequence of propositions P_n that depend on $n = 0, 1, 2, ...$ We would like to show that all the P_n's are true. (*The initial value* $n = 0$ *is not important. We might want to prove the propositions* P_n *for, say,* $13 \leq n < \omega_0$.) For example, we might have in mind the propositions

$$P_n : \qquad 0 + 1 + 2 + ... + n = \tfrac{n(n+1)}{2}$$

As you should already know, the proof can be done by induction. Induction in elementary courses takes one of two forms (stated here using P_0 as the initial proposition):

1) (Principle of Induction) If P_0 is true and if (P_{k-1} is true $\Rightarrow$ P_k is true), then P_n is true for all $n < \omega_0$.

2) (Principle of Complete Induction) If P_0 is true and if (P_j is true for all $j < k \Rightarrow P_k$ is true), then P_n is true for all $n < \omega_0$.

Formally, 2) looks weaker than 1), because it has a stronger hypothesis. But in fact the versions 1) and 2) are equivalent statements about $[0, \omega_0)$. (*Why?*) Sometimes form 2) is more convenient to use. For example, try using both versions of induction to prove that every natural number greater than 1 has a factorization into primes.

The Principle of Induction works because $[0, \omega_0)$ is well-ordered:

> If $\{n \in (0, \omega_0] : P_n \text{ is false}\} \neq \emptyset$, then it would contain a smallest element $k > 0$. This is impossible: since P_{k-1} is true, P_k must be true.

You might expect a principle analogous to 1) could be used in every well-ordered sets $[0, \alpha)$, not just in $[0, \omega_0)$. But an ordinal β might not have an immediate predecessor, so version 1) might not make sense. So we work instead with 2): we can generalize "ordinary induction" if we state it in the form of complete induction.

Theorem 9.1 (Principle of Transfinite Induction) Let α be an ordinal and $T \subseteq [0, \alpha)$. If

1) $0 \in T$ and
2) $\forall \beta \in [0, \alpha) \ \ [\,[0, \beta) \subseteq T \ \Rightarrow \ \beta \in T\,]$,

then $T = [0, \alpha)$.

Proof If $T \neq [0, \alpha)$, then there is a smallest $\beta \in [0, \alpha) - T$. By definition of β, $[0, \beta) \subseteq T$. But then 2) implies $\beta \in T$, contrary to the choice of β.

Using Theorem 9.1 is completely analogous to using complete induction in $[0, \omega_0)$. For each $\gamma < \alpha$, we have a proposition P_γ and we want to show that all the P_γ's are all true. (*For example, we might have a set* $K_\gamma \subseteq X$ *somehow defined for each* $\gamma < \alpha$, *and* P_γ *might be the proposition "*K_γ *is compact."*) Let $T = \{\gamma < \alpha : P_\gamma \text{ is true}\} \subseteq [0, \alpha)$. If we show that P_0 is true, and if, assuming that assume P_γ is true for all $\gamma < \beta$, we can then prove that P_β must be true, then Theorem 9.1 implies that $P_{\gamma.}$ is true for all $\gamma < \alpha$. We will look at several examples in Section 10.

Note: In the statement of Theorem 9.1, part 1) is included only for emphasis. In fact, 1) is automatically true if we know 2) is true: for if we let $\beta = 0$ in 2), then $[0,0) = \emptyset \subseteq T$ is true, so $0 \in T$ – that is, P_0 is true. But in actually using Theorem 9.1 (as described in the preceding paragraph) and trying to prove that 2) is true, the first value $\beta = 0$ requires us to show P_0 is true with "no induction assumptions" since there are no P_γ's with $\gamma < \beta = 0$. Doing that is just verifying that 1) is true.

We can also define objects in a similar way – by transfinite recursion. Elementary definitions by recursion should be familiar – for example, we might say :

let $f(0) = 13$ and,
for each $n > 0$, let $f(n) = 2f(n-1)$ (**)

We than say that "f is defined for all $n = 0, 1, 2, ...$" We draw that conclusion by arguing in the following informal way: if not, then there is a smallest $k \in (0, \omega_0)$ for which f is not defined: this is impossible because then $f(k-1)$ is defined, and therefore (by **) so is $f(k)$.

This argument depends only on the fact that $[0, \omega_0)$ is well-ordered, so it generalizes to the following principle.

Informal Principle 9.2 (Transfinite Recursion) For each $\beta < \alpha$, suppose a rule is given that defines an object P_β in terms of objects P_γ already defined (that is, in terms of P_γ's with $\gamma < \beta$). Then P_β is defined for all $\beta < \alpha$. (*The principle implies that P_0 is defined "absolutely" – that is, without any previous P_γ's to work with – since there are no P_γ's with < 0*}.

This "informal" statement is a reasonably accurate paraphrase of a precise theorem in axiomatic set theory, and the informal proof is virtually identical to the one given above for simple recursion on $[0, \omega_0)$. As stated in 9.2, this principle is strong and clear enough for everyday use, and we will consider it "proven" and useable.

Principle 9.2 is "informal" because it does seem a little vague in spots – "some object", "a rule is given", "if P_γ is defined ... , then P_β is defined ..." Without spending a lot of time on the set theoretic issues, we digress to show how the statement can be made a little more precise.

> In axiomatic set theory, the "objects" P_β will, of course, be sets (since everything is a set). We can think of "defining P_β" to mean choosing P_β from some specified set $\mathcal{E}$. In the context of some problem, $\mathcal{E}$ is the "universal set" in which the objects P_β will all live. For example, we might have $\mathcal{E} = C(X)$ and want to define continuous functions $P_\beta \in C(X)$ for each $\beta < \alpha$.
>
> The sets P_γ already chosen ("defined") in $\mathcal{E}$ for $\gamma < \beta$ can be described efficiently by a function $\psi_\beta \in \mathcal{E}^{[0,\beta)}$: for $\gamma < \beta$, $\psi_\beta(\gamma) = P_\gamma \in \mathcal{E}$.
>
> To define the set P_β in terms of the preceding P_γ's means that we need to define P_β using ψ_β. We need a function ("rule") R_β so that $R_\beta(\psi_\beta)$ gives the new set $P_\beta \in \mathcal{E}$. In other words, we want $R_\beta : \mathcal{E}^{[0,\beta)} \to \mathcal{E}$.

The conclusion that we have "completed" the process and that P_β is defined for all $\beta < \alpha$ means that there is a function $F \in \mathcal{E}^{[0,\alpha)}$ where, for each $\beta < \alpha,\ F(\beta) =$ the P_β selected at the earlier stage by R_β – that is, $F(\boldsymbol{\beta}) = R_\beta(F|\,[0,\beta))$.

This leads us to the following formulation. The "full" formulation of the standard theorem about transfinite recursion in axiomatic set theory needs to be a little stronger still, so we call this version – which we state without proof – the "weak" version. It is more than adequate for our purposes here.

Theorem 9.3 (Transfinite Recursion, Weak Form) Suppose α is an ordinal. Let $\mathcal{E}$ be a set and suppose that, for each $\beta < \alpha$, we have a function $R_\beta : \mathcal{E}^{[0,\beta)} \to \mathcal{E}$. Then there exists a unique function $F \in \mathcal{E}^{[0,\alpha)}$ such that, for each $\beta < \alpha,\ F(\beta) = R_\beta(F|\,[0,\beta))$.

Proof See, for example, *Topology (J. Dugundji)*

The following example illustrates what the Recursion Theorem 9.3 in a concrete example. When all is said and done, it looks just the way an informal, simple definition by recursion (Principle 9.2) should look.

Example 9.4 We want to define numbers E_n for every $n = 0, 1, \ldots$. Informally we might say:

Let $E_0 = 1$ and for $n > 0$, let $E_n = {E_0}^2 + \ldots + E_{n-1}^2$.

The informal Principle 9.2 lets us conclude that E_n is defined for all $n < \omega_0$.

In terms of the more formal Theorem 9.3, we can describe what is "really" happening as follows:

Let $[0,\alpha) = [0,\omega_0)$ and $\mathcal{E} = \mathbb{N}$. For each $n < \omega_0$, define $R_n : \mathbb{N}^{[0,n)} \to \mathbb{N}$ as follows:

For $n = 0$: $\mathbb{N}^{[0,0)} = \mathbb{N}^{\emptyset} = \{\emptyset\}$ and define $R_0(\emptyset) = 1$
For $n > 0$: if $\psi \in \mathbb{N}^{[0,n)}$, define $R_n(\psi) = \psi^2(0) + \ldots + \psi^2(n-1)$

(*Note that the R_n's are defined explicitly for each n, <u>not</u> recursively.*).

The theorem states that there is a unique function $F \in \mathbb{N}^{[0,\omega_0)}$ such that

$$F(0) = R_0(\,F|[0,0)\,) = R_0(\emptyset) = 1$$
$$F(1) = R_1(F|[0,1)\,) = F(0)^2 = 1^2 = 1$$
$$F(2) = R_2(F|[0,2)\,) = F(0)^2 + F(1)^2 = 2$$

... etc. ...

which is just what we wanted: dom $F = [0,\omega_0)$, so $E_n = F(n)$ is defined for all $n < \omega_0$.

10. Using Transfinite Induction and Recursion

This section presents a number of examples using recursion and induction in an essential way. Taken together, they are a miscellaneous collection, but each example has some interest in itself.

Borel Sets in Metric Spaces

The classical theory of Borel sets is developed in metric spaces (X, d). The collection of Borel sets in a metric space is important in analysis and also in set theory. Roughly, Borel sets are the sets that can be generated from open sets by the operations of countable union and countable intersection "applied countably many times." Therefore a Borel set is only a "small" number of operations "beyond" the open sets, and Borel sets are fairly well behaved. We will use transfinite recursion (the informal version) to define the Borel sets and prove a few simple theorems. When objects are defined by recursion, proofs about them often involve induction.

We begin with a simple lemma about ordinals.

Lemma 10.1 Every ordinal α can be written uniquely in the form $\alpha = \beta + n$ where β is either 0 or a limit ordinal, and $n < \omega_0$.

Proof If α is finite, then $\alpha = 0 + n$ where $n < \omega_0$.

Suppose α is infinite. If α is a limit ordinal, we can write $\alpha = \alpha + 0$. If α is not a limit ordinal, then α has an immediate predecessor which, for short, we denote here as "$\alpha - 1$." If $\alpha - 1$ is not a limit ordinal, then it has an immediate predecessor $\alpha - 2$. Continuing in this way, we get to a limit ordinal after a finite number of steps, n – for otherwise $\alpha - 1 > \alpha - 2 > \ldots > \alpha - n > \ldots$ would be an infinite decreasing sequence of ordinals. If $\beta = \alpha - n$ is a limit ordinal, then $\alpha = \beta + n$.

To prove uniqueness, suppose $\alpha = \beta + n = \beta' + n'$ where each of β, β' is 0 or a limit ordinal and n, n' are finite. If β or $\beta' = 0$: say $\beta = 0$. Then $\beta + n$ is finite, so $\beta' = 0$ and $n = n'$. So suppose β and β' are both limit ordinals. We have an order isomorphism f from $[0, \beta + n)$ onto $[0, \beta' + n')$. Since $[0, \beta + n)$ contains $n - 1$ ordinals after its largest limit ordinal β, the same must be true in the range $[0, \beta' + n')$, and therefore $n = n'$. Let $g = f|[0, \beta)$. Then $g : [0, \beta) \to [0, \beta')$ is an order isomorphism, so $\beta = \beta'$. ●

Definition 10.2 Suppose $\alpha = \beta + n$, where β is a limit ordinal or 0, and n is finite. We say that α is <u>even</u> if n is even, and that α is <u>odd</u> if n is odd.

For example, every limit ordinal $\alpha = \alpha + 0$ is even.

Definition 10.3 Suppose (X, d) is a metric space. Let $\mathcal{G}_0 = \mathcal{T}_d$, the collection of open sets. For each $0 < \alpha < \omega_1$, and suppose that $\mathcal{G}_\beta$ has been defined for all $\beta < \alpha$. Then let

$$\begin{cases} \mathcal{G}_\alpha = \{G : G \text{ is a countable intersection of sets from } \bigcup\{\mathcal{G}_\beta : \beta < \alpha\} & \text{(if } \alpha \text{ is odd)} \\ \mathcal{G}_\alpha = \{G : G \text{ is a countable union of sets from } \bigcup\{\mathcal{G}_\beta : \beta < \alpha\} & \text{(if } \alpha \text{ is even)} \end{cases}$$

$\mathcal{B} = \bigcup\{\mathcal{G}_\alpha : \alpha < \omega_1\}$ is the family of <u>Borel sets</u> in (X, d).

The sets in $\mathcal{G}_1$ are the G_δ sets; the sets in $\mathcal{G}_2$ are countable unions of G_δ sets and are traditionally called $G_{\delta\sigma}$ sets; the sets in $\mathcal{G}_3$ are called $G_{\delta\sigma\delta}$-sets, etc.

Theorem 10.4 Suppose (X, d) is a metric space.

1) If $\alpha < \beta < \omega_1$, then $\mathcal{G}_\alpha \subseteq \mathcal{G}_\beta$, so

$$\mathcal{G}_0 \subseteq \mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq ... \subseteq \mathcal{G}_\alpha \subseteq ... \subseteq \mathcal{G}_\beta \subseteq ... \qquad (\alpha < \beta < \omega_1)$$

2) $\mathcal{G}_\alpha$ is closed under countable unions if α is even, and $\mathcal{G}_\alpha$ is closed under countable intersections if α is odd.

3) $\mathcal{B}$ is closed under countable unions, intersections and complements. Also, if B_1 and B_2 are in $\mathcal{B}$, so is $B_1 - B_2$.

Proof 1) Suppose $\alpha < \beta < \omega_1$. $\mathcal{G}_\beta$ is defined as the collection of all countable unions (if β is even) or intersections (if is odd) of sets in the preceding families $\mathcal{G}_\alpha$ $(\alpha < \beta)$. In particular, any one set from $\mathcal{G}_\alpha$ is in $\mathcal{G}_\beta$.

2) Suppose α is even and $G_1, G_2, ..., G_n, ... \in \mathcal{G}_\alpha$. Each G_n is a countable union of sets from $\bigcup\{\mathcal{G}_\beta : \beta < \alpha\}$. Therefore $\bigcup_{n=1}^\infty G_n$ is also a countable union of sets from $\bigcup\{\mathcal{G}_\beta : \beta < \alpha\}$, so $\bigcup_{n=1}^\infty G_n \in \mathcal{G}_\alpha$. The proof is similar if α is odd.

3) Suppose $B_1, B_2, ..., B_n, ... \in \mathcal{B}$. For each n, $B_n \in \mathcal{G}_{\alpha_n}$ for some $\alpha_n < \omega_1$. If $\alpha = \sup\{\alpha_n : n = 1, 2, ...\} < \omega_1$, then $\mathcal{G}_{\alpha_n} \subseteq \mathcal{G}_\alpha \subseteq \mathcal{G}_{\alpha+1}$ for every n. By part 2), one of the collections $\mathcal{G}_\alpha$ or $\mathcal{G}_{\alpha+1}$ is closed under countable intersections and the other under countable unions. Therefore $\bigcup_{n=1}^\infty B_n$ and $\bigcap_{n=1}^\infty B_n$ are both in $\mathcal{G}_{\alpha+1} \subseteq \mathcal{B}$.

To show that $\mathcal{B}$ is closed under complements, we first prove, using transfinite induction, that if $G \in \mathcal{G}_\alpha$, then $X - G \in \mathcal{G}_{\alpha+1}$.

$\alpha = 0$: If $G \in \mathcal{G}_0$, then G is open so $X - G$ is closed. But a closed set in a metric space is a G_δ set, so $X - G \in \mathcal{G}_1$.

Suppose the conclusion holds for all $\beta < \alpha < \omega_1$. We must show it holds also for $\mathcal{G}_\alpha$.

Let $G \in \mathcal{G}_\alpha$. If α is odd, then $G = \bigcap_{n=1}^\infty G_n$, where $G_n \in \mathcal{G}_{\beta_n}$ $(\beta_n < \alpha)$. By the induction hypothesis, $X - G_n \subseteq \mathcal{G}_{\beta_n+1} \subseteq \mathcal{G}_{\alpha+1}$ for all n. Since $\alpha + 1$ is even, we have that $\bigcup_{n=1}^\infty (X - G_n) = X - \bigcap_{n=1}^\infty G_n = X - G \in \mathcal{G}_{\alpha+1}$. (*The case when α is even is similar*).

If $B_2 \in \mathcal{B}$, then $B_2 \in \mathcal{G}_\alpha$ for some $\alpha < \omega_1$, so $X - B_2 \in \mathcal{G}_{\alpha+1} \subseteq \mathcal{B}$. So if $B_1 \in \mathcal{B}$, then $B_1 \cap (X - B_2) = B_1 - B_2 \in \mathcal{B}$. •

Part 3) of the preceding theorem shows why the definition of the Borel sets only uses ordinals $\alpha < \omega_1$. Once we get to $\mathcal{B} = \bigcup\{\mathcal{G}_\alpha : \alpha < \omega_1\}$, the process "closes off" $-$ that is, continuing with additional countable unions and intersections produces no new sets.

Definition 10.3 presents the construction of the Borel sets "from the bottom up." It has the advantage of exhibiting how the sets in $\mathcal{B}$ are constructed step by step. However, it is also possible to define $\mathcal{B}$ "from the top down." This approach is neater, but it gives less insight into which sets are Borel.

Definition 10.5 A family $\mathfrak{S}$ of subsets of X is called a $\underline{\sigma\text{-algebra}}$ if $X \in \mathfrak{S}$ and $\mathfrak{S}$ is closed under complements, countable intersections, and countable unions.

Suppose $\mathcal{A}$ is a collection of subsets of X. Then $\mathcal{P}(X)$ is (the largest) σ-algebra containing $\mathcal{A}$. It is also clear that the intersection of a collection of σ-algebras is a σ-algebra. Therefore the smallest σ-algebra containing $\mathcal{A}$ exists: it is the intersection of all σ-algebras containing $\mathcal{A}$.

The following theorem could be taken as the definition of the family of Borel sets.

Theorem 10.6 The family $\mathcal{B}$ of Borel sets in (X, d) is the smallest σ-algebra containing all the open sets of X.

Proof The rough idea is that our previous construction puts into $\mathcal{B}$ all the sets that need to be there to form a σ-algebra, but no others.

We have already proven that $\mathcal{B}$ is a σ-algebra containing the open sets. We must show $\mathcal{B}$ is the smallest – that is, if $\mathcal{B}'$ is a σ-algebra containing the open sets, then $\mathcal{B} \subseteq \mathcal{B}'$. We show this by transfinite induction.

We are given that $\mathcal{G}_0 \subseteq \mathcal{B}'$.

Suppose that $\mathcal{G}_\beta \subseteq \mathcal{B}'$ for all $\beta < \alpha < \omega_1$. We must show $\mathcal{G}_\alpha \subseteq \mathcal{B}'$.

Assume α is odd. If $G \in \mathcal{G}_\alpha$ then $G = \bigcap_{n=1}^\infty G_n$ where $G_n \in \mathcal{G}_{\beta_n}$ for some $\beta_n < \alpha$. By hypothesis each $\mathcal{G}_{\beta_n} \in \mathcal{B}'$ and $\mathcal{B}'$ is closed under countable intersections so $G \in \mathcal{B}'$. Therefore $\mathcal{G}_\alpha \subseteq \mathcal{B}'$. (*The case when* α *is even is entirely similar*.)

Since $\mathcal{G}_\alpha \subseteq \mathcal{B}'$ for all $\alpha < \omega_1$, we get that $\mathcal{B} = \bigcup_{\alpha<\omega_1} \mathcal{G}_\alpha \subseteq \mathcal{B}'$. •

The next theorem gives us an upper bound for the number of Borel sets in a separable metric space.

Theorem 10.7 If (X, d) is separable metric space, then $|\mathcal{B}| \le c$.

Proof We prove first that for each $\alpha < \omega_1$, $|\mathcal{G}_\alpha| \le c$.

$\alpha = 0$: A separable metric space has a countable base $\mathcal{C}$ for the open sets. Since every open set is the union of a subfamily of $\mathcal{C}$, we have $|\mathcal{T}_d| = |\mathcal{G}_0| \le |\mathcal{P}(\mathcal{C})| \le 2^{\aleph_0} = c$.

Assume that $|\mathcal{G}_\beta| \le c$ for all $\beta < \alpha < \omega_1$. Since α has only countably many predecessors, $|\bigcup_{\beta<\alpha} \mathcal{G}_\beta| \le c$. Since each set in $\mathcal{G}_\alpha$ is a countable intersection or union of a sequence of sets from $\bigcup_{\beta<\alpha} \mathcal{G}_\beta$, we have $|\mathcal{G}_\alpha| \le |\bigcup_{\beta<\alpha} \mathcal{G}_\beta|^{\mathbb{N}} \le c^{\aleph_0} = c$.

Therefore $|\mathcal{B}| = |\bigcup_{\alpha<\omega_1}\mathcal{G}_\alpha| \le \aleph_1 \cdot c = c$. •

(*Can you see a generalization to arbitrary metric spaces?*)

Corollary 10.8 There are non-Borel sets in $\mathbb{R}$.

For those who know a bit of measure theory: every Borel set in $\mathbb{R}$ is Lebesgue measurable. Since a subset of a set of measure 0 is measurable, all 2^c subsets of the Cantor set C are measurable. Therefore there are Lebesgue measurable subsets of C that are not Borel sets.

Example 10.9

1) If (X, d) is discrete, then every subset is open, so every subset is Borel: $\mathcal{B} = \mathcal{T}_d = \mathcal{P}(X)$.

2) If $X = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$, then every subset is a G_δ, but X is not discrete. Therefore

$$\mathcal{G}_0 \underset{\neq}{\subseteq} \mathcal{G}_1 = \mathcal{G}_2 = \ldots = \mathcal{G}\alpha = \ldots = \mathcal{B}.$$

3) The following facts are true but harder to prove:

a) For each $\alpha < \omega_1$, there exists a metric space (X, d) for which

$$\mathcal{G}_0 \underset{\neq}{\subseteq} \mathcal{G}_1 \underset{\neq}{\subseteq} \ldots \underset{\neq}{\subseteq} \mathcal{G}_\alpha = \mathcal{G}_{\alpha+1} = \ldots = \mathcal{G}_\beta = \ldots = \mathcal{B}$$

In other words, the Borel construction continually adds fresh sets until the α^{th} stage but not thereafter.

b) In $\mathbb{R}$, $\mathcal{G}_\alpha \neq \mathcal{G}_\beta$ for all $\alpha < \beta < \omega_1)$ $-$ that is, new Borel sets appear at every stage in the construction.

A New Characterization of Normality

Definition 10.10 A family $\mathcal{F}$ of subsets of a space X is called point-finite if $\forall x \in X$, x is in only finitely many sets from $\mathcal{F}$.

Definition 10.11 An open cover $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ of X is called shrinkable if there exists an open cover $\mathcal{V} = \{V_\alpha : \alpha \in A\}$ of X such that, for each α, cl $V_\alpha \subseteq U_\alpha$. $\mathcal{V}$ is called a shrinkage of $\mathcal{U}$.

Theorem 10.12 X is normal iff every point-finite open cover of X is shrinkable. (*In particular, every finite open cover of a normal space is shrinkable: see Exercise VII.E18.*)

Proof Suppose every point-finite open cover of X is shrinkable and let A and B be disjoint closed sets in X. The open cover $\mathcal{U} = \{X - A, X - B\}$ has a shrinkage $\mathcal{V} = \{V_1, V_2\}$ and the sets $U = X - \text{cl } V_1$ and $V = X - \text{cl } V_2$ are disjoint open sets containing A and B. Therefore X is normal.

Conversely, suppose X is normal and let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ be a point-finite open cover of X. Without loss of generality, we may assume that the index set A is a segment of ordinals $[0, \gamma)$ (*why?*) so that $\mathcal{U} = \{U_\alpha : \alpha < \gamma\}$.

Let $F_0 = X - \bigcup_{\alpha > 0} U_\alpha$. Since $F_0 \subseteq U_0$, we can use normality to choose an open set V_0 such that $F_0 \subseteq V_0 \subseteq \text{cl } V_0 \subseteq U_0$ and $\{V_0\} \cup \{U_\alpha : \alpha > 0\}$ covers X.

Suppose $0 < \alpha < \gamma$ and that for all $\beta < \alpha$ we have defined an open sets V_β such that $\text{cl } V_\beta \subseteq U_\beta$ and such that $\{V_\beta : \beta < \alpha\} \cup \{U_\beta : \beta \geq \alpha\}$ covers X. Letting $F_\alpha = X - (\bigcup_{\beta < \alpha} V_\beta \cup \bigcup_{\beta > \alpha} U_\beta) \subseteq U_\alpha$, we can use normality to choose an open set V_α with $F_\alpha \subseteq V_\alpha \subseteq \text{cl } V_\alpha \subseteq U_\alpha$.

Clearly, $\{V_\beta : \beta < \alpha + 1\} \cup \{U_\beta : \beta \geq \alpha + 1\}$ covers X.

By transfinite recursion, the V_α's are defined for all $\alpha < \gamma$, and we claim that $\mathcal{V} = \{V_\alpha : \alpha < \gamma\}$ is a cover of X.

> Notice that there is something here that needs to be checked: we know that we have a cover $\{V_\beta : \beta < \alpha\} \cup \{U_\beta : \beta \geq \alpha\}$ at each step in the process, but do we still have a cover when we're finished? To see explicitly that there is an issue, consider the following example.
>
> Let X be the set of reals with the "left-ray" topology (a normal space) and consider the open cover $\mathcal{U} = \{(-\infty, n) : n < \omega_0\}$. If we go through the procedure described above, we get $F_0 = X - \bigcup_{n > 0}(-\infty, n) = \emptyset$ so we might have chosen $V_0 = \emptyset$: that would give $F_0 \subseteq V_0 \subseteq \text{cl } V_0 \subseteq U_0$ and $\{V_0\} \cup \{U_n : n > 0\}$ would still be a cover. Continuing, we can see that at every stage we could choose $V_n = \emptyset$ and that $\{V_0, V_1, ... V_n\} \cup \{U_k : k > n\}$ is still a cover. But when we're done, the collection $\mathcal{V} = \{V_n : n < \omega_0\}$ is not a cover! Of course, the cover $\mathcal{U}$ is not point-finite.

Suppose $x \in X$. Then x is in only finitely many sets of $\mathcal{U}$ – say $U_{\alpha_1}, ..., U_{\alpha_n}$. Let α be the largest of these indices so that $x \notin \bigcup_{\beta > \alpha} U_\beta$. If x is in one of the V_β's with $\beta < \alpha$, we're done. Otherwise $x \in X - (\bigcup_{\beta < \alpha} V_\beta \cup \bigcup_{\beta > \alpha} U_\beta) = F_\alpha \subseteq V_\alpha$. Either way x is in a set in $\mathcal{V}$, so $\mathcal{V}$ is a cover. •

Question: what happens if "point-finite" is changed to "point-countable" in the hypothesis? Could the "max" in the argument be replaced by a "sup" ?

Definition 10.13 A cover of X is called locally finite if every point $x \in X$ has a neighborhood that intersects at most finitely many of the sets in the cover.

Corollary 10.14 Suppose $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ is a locally finite open cover of a normal space X. Then there exist continuous functions $f_\alpha : X \to [0, 1]$ such that

i) $f_\alpha \mid X - U_\alpha = 0$ for every $\alpha \in A$
ii) $\sum\{f_\alpha(x) : \alpha \in A\} = 1$ for every $x \in X$.

The collection of functions $\{f_\alpha\}$ is called a partition of unity subordinate to $\mathcal{U}$.

Proof It is easy to see that a locally finite open cover of X is point-finite. By Theorem 10.12, $\mathcal{U}$ has a shrinkage $\mathcal{V} = \{V_\alpha : \alpha \in A\}$. For each α, we can use Urysohn's Lemma (VII.5.2) to pick a continuous function $g_\alpha : X \to [0,1]$ such that $g_\alpha | \text{cl}\, V_\alpha = 1$ and $g_\alpha | X - U_\alpha = 0$.

Each point x is in only finitely many U_α's, so $g_\alpha(x) = 0$ for all but finitely many α's and therefore $g(x) = \sum_{\alpha \in A} g_\alpha(x) \in \mathbb{R}$. Each point x has a neighborhood N_x which intersects only finitely many U_α's so $g|N_x$ is essentially a finite sum of continuous functions and g is continuous (see Exercise III.E21).

Since each x is in some set cl V_{α_0}, so $g_{\alpha_0}(x) = 1$. Therefore $g(x)$ is never 0 so each $f_\alpha(x) = \frac{g_\alpha(x)}{g(x)}$ is continuous. The f_α's clearly satisfy both i) and ii). •

A Characterization of Countable Compact Metric Spaces

In 1920, the journal *Fundamenta Mathematicae* was founded by Zygmund Janiszewski, Stefan Mazurkiewicz and Waclaw Sierpinski. It was a conscious attempt to raise the profile of Polish mathematics thorough a journal devoted primarily to the exciting new field of topology. To reach the international community, it was agreed that published articles would be in one of the most popular scientific languages of the day: French, German or English. *Fundamenta Mathematicae* continues today as a leading mathematical journal with a scope broadened somewhat to cover set theory, mathematical logic and foundations of mathematics, topology and its interactions with algebra, and dynamical systems.

An article by Sierpinski and Mazurkiewicz appeared in the very first volume of this journal characterizing compact, countable metric spaces in a rather vivid way. We will prove only part of this result: our primary purpose here is just to illustrate the use of transfinite recursion and induction and the omitted details are messy.

Suppose (X, d) is a nonempty compact, countable metric space.

Definition 10.15 For $A \subseteq X$, define $A' = \{x \in X : x \text{ is a limit point of } A\}$. A' is called the <u>derived set</u> of A. If A is closed, it is easy to see that $A' \subseteq A$ and that A' is also closed.

We will use the derived set operation $(\,')$ repeatedly in a definition by transfinite recursion.

Let $A_0 = X$ and, for each $\alpha < \omega_1$, define

$$\begin{cases} A_\alpha = A_\beta' & \text{if } \alpha = \beta + 1 \\ A_\alpha = \bigcap\{A_\beta : \beta < \alpha\} & \text{if } \alpha \text{ is a limit ordinal} \end{cases}$$

A_α is closed for all $\alpha < \omega_1$ and $A_0 \supseteq A_1 \supseteq \ldots \supseteq A_\alpha \supseteq \ldots$. This sequence is called <u>the derived sequence</u> of X.

For some $\alpha < \omega_1$, we have $A_{\alpha+1} = A_\alpha$ – because otherwise, for each α, we could choose a point $x_\alpha \in A_\alpha - A_{\alpha+1}$ and $\{x_\alpha : \alpha < \omega_1\}$ would be a subset of X with cardinal $\aleph_1$. Let α_0 be the <u>smallest</u> α for which $A_{\alpha_0} = A_{\alpha_0+1}$. It follows that $A_\beta = A_{\alpha_0}$ for all $\beta > \alpha_0$.

Since X is compact metric, the closed set A_{α_0} is complete, and because $A_{\alpha_0} = A_{\alpha_0+1}$, every point in A_{α_0} is a limit point. Therefore $A_{\alpha_0} = \emptyset$, since a nonempty complete metric space with no isolated points contains at least c points (Theorem IV.3.6).

We know $\alpha_0 = 0$ is impossible (because that would mean $A_{\alpha_0} = A_0 = X = \emptyset$). If α_0 were a limit ordinal, then $A_{\alpha_0} = \bigcap\{A_\beta : \beta < \alpha_0\} \neq \emptyset$ (because $\{A_\beta : \beta < \alpha_0\}$ is a family of nonempty compact sets with the finite intersection property). Therefore α_0 must have an immediate predecessor β_0.

The definition of α_0 implies that $A_{\beta_0} \neq \emptyset$. In fact, A_{β_0} must be finite (if it were an infinite set in the compact space X, A_{β_0} would have a limit point and then $A'_{\beta_0} = A_{\alpha_0} \neq \emptyset$). Let $n = |A_{\beta_0}| < \omega_0$.

In this way, we arrive at a pair (β_0, n), where

> A_{β_0} is the last nonempty derived set of X $(\beta_0 < \omega_1)$, and
> A_{β_0} contains n points $(0 < n < \omega_0)$.

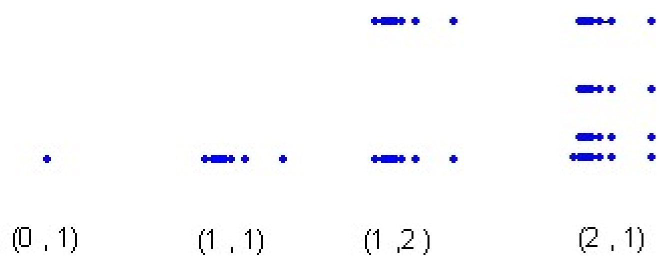
Since homeomorphisms preserve limit points and intersections, it is clear that the construction in any space homeomorphic to X will produce the same pair (β_0, n).

Theorem 10.16 (Sierpinski-Mazurkiewicz) Let $\beta_0 < \omega_1$ and $0 < n < \omega_0$. Two nonempty compact, countable metric spaces X and Y are homeomorphic iff they are associated with the same pair (β_0, n). (*Therefore* (β_0, n) *is a "topological invariant" that characterizes nonempty compact countable metric spaces*.) For any such pair (β_0, n), there exists a nonempty compact countable metric space associated with this pair.

Corollary 10.17 There are exactly $\aleph_1$ nonhomeomorphic compact countable metric spaces.

Proof The number of different compact countable metric spaces is the same as the number of pairs (β_0, n), namely $\aleph_1 \cdot \aleph_0 = \aleph_1$. •

Example 10.18 In the following figure, each column contains a compact countable subspace of $\mathbb{R}^1$ with the invariant pair listed. Each figure is built up using sets order-isomorphic to $\{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}\}$.



Alexandroff's Problem

A compact metric space (X, d) is second countable and therefore satisfies $|X| \leq c$ (Theorem II.5.21). In 1923, Alexandroff and Urysohn conjectured that a stronger result is true:

A first countable compact Hausdorff space X *satisfies* $|X| \leq c$.

This conjecture was not settled until 1969, in a rather famously complicated proof by Arhangel'skii.

Here is a proof of an even stronger result $-$ "compact" is replaced by "Lindelöf" $-$ that comes from a few years after Arhangel'skii's work.

Theorem 10.16 (**Pol, Šapirovski**) If X is first countable, Hausdorff and Lindelöf, then $|X| \leq c$.

Proof For each $p \in X$, choose a countable open neighborhood base $\mathcal{V}_p$ at p and, for each $A \subseteq X$, let $\mathcal{V}_A = \bigcup\{\mathcal{V}_p : p \in A\} =$ the collection of all the basic neighborhoods of all the points in A.

For each countable family of sets $\mathcal{V} \subseteq \mathcal{V}_A$ for which $X - \bigcup\mathcal{V} \neq \emptyset$, pick a point $q_{\mathcal{V}} \in X - \bigcup\mathcal{V}$. Define $P(A) =$ cl $(A \cup \{$all such $q_{\mathcal{V}}$'s chosen for $A\})$.

Notice that if $|A| = c$, then $|P(A)| = c$.

Since $|A| = c$, we have $|\mathcal{V}_A| = c$. There are at most $c^{\aleph_0} = c$ countable families $\mathcal{V} \subseteq \mathcal{V}_A$, so $|A \cup \{$all such $q_{\mathcal{V}}$'s chosen for $A\}| = c$. Since X is first countable, sequences suffice to describe the topology, and since X is Hausdorff, sequential limits are unique. Therefore $|P(A)|$ is no larger than the number of sequences in $A \cup \{$all such $q_{\mathcal{V}}$'s chosen for $A\}$ – namely $c^{\aleph_0} = c$.

Now fix, once and for all, a set $A \subseteq X$ with $|A| = c$. (*If no such A exists, we are done!*) By recursion, we now "build up" some new sets A_α from A. The idea is that the new A_α's always have cardinality $\leq c$ and that the new sets eventually include all the points of X.

Let $A_0 = A$.
For each ordinal $\alpha < \omega_1$, define

$$A_\alpha = \begin{cases} P(A_\beta) & \text{if } \alpha = \beta + 1 \\ \bigcup\{A_\beta\colon \beta < \alpha\} & \text{if } \alpha \text{ is a limit ordinal} \end{cases}.$$

For each $\alpha < \omega_1$, $|A_\alpha| = c$:

$|A_0| = c$.
Suppose $|A_\beta| = c$ for all $\beta < \alpha$.

If $\alpha = \beta + 1$, then $A_\alpha = P(A_\beta)$, so $|A_\alpha| = c$.
If α is a limit ordinal, then $|A_\alpha| = |\bigcup\{A_\beta : \beta < \alpha\}| = c \cdot \aleph_0 = c$.

Let $B = \bigcup\{A_\alpha : \alpha < \omega_1\}$. Then $c \leq |B| = |\bigcup\{A_\alpha : \alpha < \omega_1\}| \leq c \cdot \aleph_1 = c$.

We claim that $B = X$ and, if so, we are done.

B is closed in X: if $x \in \operatorname{cl} B$, then (using first countability) there is a sequence (x_n) in B with $(x_n) \to x$. If $x_n \in A_{\alpha_n}$ and we let $\alpha = \sup\{\alpha_n\} < \omega_1$, then every x_n is in A_α. Therefore $x \in \operatorname{cl} A_\alpha = A_{\alpha+1} \subseteq B$.

Since B is a closed subspace of the Lindelöf space X, B is Lindelöf.

If $B \neq X$, then we can pick a point $q \in X - B$. For each $p \in B$, choose an open neighborhood $V_p \in \mathcal{V}_p$ such that $q \notin V_p$. The V_p's form an open cover of B, so a countable collection of these sets, say $\mathcal{V} = \{V_p : p \in C$, for some countable $C \subseteq B\}$ covers B. Thus $B \subseteq \bigcup\mathcal{V}$ and $q \notin \bigcup\mathcal{V}$. But this is impossible:

Since C is countable, we have that $C \subseteq A_\alpha$ for some $\alpha < \omega_1$ and $\mathcal{V}$ is a countable subfamily of $\mathcal{V}_{A_\alpha}$ for which $X - \bigcup\mathcal{V} \neq \emptyset$. By definition of $P(A_\alpha) = A_{\alpha+1}$), a point $q_{\mathcal{V}} \in X - \bigcup\mathcal{V}$ was put into the set $A_{\alpha+1} \subseteq B$. This contradicts the fact that $\mathcal{V}$ covers B. •

The Mazurkiewicz 2-set

The circle S^1 is a subset of the plane which intersects every straight line in at most two points. We use recursion to construct something more bizarre.

Theorem 10.17 (Mazurkiewicz) There exists a set $A \subseteq \mathbb{R}^2$ such that $|A \cap L| = 2$ for every straight line L.

Proof Let δ be the first ordinal of cardinal c. Since the well-ordered segment $[0, \delta)$ represents δ, there are c ordinals $\alpha < \delta$.

In fact, there are c limit ordinals $< \delta$. There are certainly infinitely many (say m) limit ordinals $< \delta$ (*why?*) and, by Lemma 10.1, every ordinal $\alpha < \delta$ can be written uniquely in the form $\alpha = \beta + n$, where β is a limit ordinal (or 0) and n is finite. Therefore there are $m \cdot \aleph_0 = m$ ordinals $< \delta$. But δ has c predecessors. Therefore $m = c$.

Since $\mathbb{R}^2$ has exactly c points and exactly c straight lines, we can index both $\mathbb{R}^2$ and the set of straight lines using the ordinals less than δ: $\mathbb{R}^2 = \{p_\xi : \xi < \delta\}$ and $\{L_\xi : \xi < \delta\}$.

We will define points a_α for each $\alpha < \delta$, and the set $A = \{a_\alpha : \alpha < \delta\}$ will be the set we want.

Let a_0 be the first point of $\mathbb{R}^2$ (as indexed above) not on L_0.

Suppose that we have defined points a_ξ for all $\xi < \alpha < \delta$. We need to define a_α. Let

$A_\alpha = \{a_\xi : \xi < \alpha\}$ *Note* $|A_\alpha| < c$
$T_\alpha = \{L : L$ is a straight line containing ≥ 2 points of $A_\alpha\}$ *Note* $|T_\alpha| < c$
$\beta_\alpha =$ least ordinal so that $L_{\beta_\alpha} \notin T_\alpha$
(that is, L_{β_α} is the first line listed that is not in T_α)
$S_\alpha = \{p : p$ is a point of intersection of L_{β_α} with a line in $T_\alpha\}$ *Note* $|S_\alpha| < c$

Since $|A_\alpha \cup S_\alpha| < c$, there are points on L_{β_α} not in $A_\alpha \cup S_\alpha$: let a_α be the first p_ξ listed that is on L_{β_α} but not in $A_\alpha \cup S_\alpha$.

By recursion, we have now defined a_α for all $\alpha < \delta$. Let $A = \{a_\alpha : \alpha < \delta\}$.

We claim that for each straight line L, $|A \cap L| \leq 2$.

If $|A \cap L| > 2$, then we could pick 3 points $a_\beta, a_\gamma, a_\alpha \in L \cap A$, where, say, $\beta < \gamma < \alpha$. Then $a_\beta, a_\gamma \in A_\alpha$, so that $L \in T_\alpha$. Since $a_\alpha \in L_{\beta_\alpha}$ (by definition) and $a_\alpha \in L$, we have $a_\alpha \in L \cap L_{\beta_\alpha}$. Therefore $a_\alpha \in S_\alpha$ – which contradicts the definition of a_α.

To complete the proof, we will show that for each straight line L, $|A \cap L| \geq 2$.

We begin with a series of observations:

a) If $\alpha_1 \leq \alpha_2 < \delta$, then $\beta_{\alpha_1} \leq \beta_{\alpha_2}$. [Clearly $T_{\alpha_1} \subseteq T_{\alpha_2}$. But $L_{\beta_{\alpha_1}}$ is the first line not in T_{α_1} and $L_{\beta_{\alpha_2}}$ is the first line not in T_{α_2}, so $\beta_{\alpha_1} \leq \beta_{\alpha_2}$.]

b) If $\alpha_1 < \alpha_2 < \alpha_3 < \delta$, then $\beta_{\alpha_1} < \beta_{\alpha_3}$. [Otherwise, by a), $\beta_{\alpha_1} = \beta_{\alpha_2} = \beta_{\alpha_3}$. Then a_{α_1}, $a_{\alpha_2}, a_{\alpha_3}$ are on $L_{\beta_{\alpha_1}} = L_{\beta_{\alpha_2}} = L_{\beta_{\alpha_3}}$. Then $L_{\beta_{\alpha_3}}$contains 3 points of A which is impossible.] In particular, if α_1, α_3 are distinct limit ordinals, there is a third ordinal α_2 between them so β_{α_1} and β_{α_3} must be distinct.

c) For any $\gamma < \delta$, there is $\alpha < \delta$ such that $\beta_\alpha > \gamma$. [Since there are c limit ordinals $< \delta$, there are c distinct values for β_α. They cannot all be $\leq \gamma$, since γ has fewer than c predecessors.]

Finally, if $L =$ some L_γ, pick an α so that $\beta_\alpha > \gamma$. Since L_{β_α} is the first line $\notin T_\alpha$, we get that $L = L_\gamma \in T_\alpha$. Therefore L contains ≥ 2 points of $A_\alpha \subseteq A$. •

More generally, it can be shown more that if for each line L we are given a cardinal number m_L with $2 \leq m_L \leq c$, then there exists a set $A \subseteq \mathbb{R}^2$ such that for each L, $|A \cap L| = m_L$.

11. Zorn's Lemma

Zorn's Lemma (ZL) states that if every chain in a nonempty poset $(X, \leq)$ has an upper bound in X, then X contains a maximal element. We remarked in Theorem 5.17 that the Axiom of Choice (AC) and Zermelo's Theorem are equivalent to Zorn's Lemma.

As a first example using Zorn's Lemma, we prove part of Theorem 5.17, in two different ways.

Theorem 11.1 ZL $\Rightarrow$ AC

Proof 1 Let $\{A_\alpha : \alpha \in A\}$ be a collection of pairwise disjoint nonempty sets. Consider the poset set $\mathcal{P} = \{S \subseteq \bigcup_{\alpha \in A} A_\alpha :$ for all α, $|S \cap A_\alpha| \leq 1\}$, ordered by inclusion. $\mathcal{P}$ is a nonempty poset because $\emptyset \in \mathcal{P}$.

Suppose $\{S_\beta : \beta \in I\}$ is a chain in $\mathcal{P}$ and let $B = \bigcup\{S_\beta : \beta \in I\}$. We claim that $B \in \mathcal{P}$.

Suppose $|B \cap A_{\alpha_0}| \geq 2$ for some α_0. Consider two points $x \neq y \in B \cap A_{\alpha_0}$ Then $x \in S_{\beta_1}$and $y \in S_{\beta_2}$ for some $\beta_1, \beta_2 \in I$. Since the S_β's are a chain, either $S_{\beta_1} \subseteq S_{\beta_2}$ or $S_{\beta_2} \subseteq S_{\beta_1}$. Therefore one of these sets, say S_{β_1}, contains both x, y. Therefore $|S_{\beta_1} \cap \mathrm{A}_{\alpha_0}| > 1$. But this is impossible since $S_{\beta_1} \in \mathcal{P}$. Therefore B contains at most one point from each A_α so $B \in \mathcal{P}$.

Since B is an upper bound for the chain in $\mathcal{P}$, Zorn's Lemma says that there is a maximal element $M \in \mathcal{P}$.

Since $M \in \mathcal{P}$, $|M \cap A_\alpha| \leq 1$ for every α. If $M \cap A_{\alpha_0} = \emptyset$ for some α_0, then we could choose an $a \in A_{\alpha_0}$ form the set $M' = M \cup \{a\} \underset{\neq}{\supseteq} M$. Then $|M \cap A_\alpha| \leq 1$ would still be true for every α, so we would have $M' \in \mathcal{P}$, which is impossible because M is maximal. Therefore $|M \cap A_\alpha| = 1$ for every α. •

Notice that the function $f : A \to \bigcup A_\alpha$ given by $f = \{(\alpha, y) \in A \times \bigcup_{\alpha \in A} A_\alpha : y \in M \cap A_\alpha\}$ is in the product $\prod\{A_\alpha : \alpha \in A\}$, so we see that $\prod(A_\alpha : \alpha \in A\} \neq \emptyset$.

Proof 2 Let $\{A_\alpha : \alpha \in A\}$ be a collection of nonempty sets. Let

$$\mathcal{P} = \{g : B \to \bigcup A_\alpha : B \subseteq A \text{ and } f(\beta) \in A_\beta \text{ for each } \beta \in B\}$$

If $g_1, g_2 \in \mathcal{P}$, then the function g_1, g_2 are sets of ordered pairs. So we can order $\mathcal{P}$ by inclusion: $g_1 \leq g_2$ iff $g_1 \subseteq g_2$. (*This relation is just "functional extension": $g_1 \leq g_2$ iff dom$(g_1) \subseteq$ dom(g_2) and $g_2|$dom$(g_1) = g_1$.*) $(\mathcal{P}, \leq)$ is a nonempty poset because $\emptyset \in \mathcal{P}$.

Suppose $\{g_i : i \in I\}$ is a chain in $\mathcal{P}$. Define $g = \bigcup\{g_i : i \in I\}$. Since the g_i's form a chain, their union is a function $g : B \to \bigcup A_\alpha$, where $B = \bigcup_{i \in I} \text{dom}(g_i)$. Moreover, if $\beta \in B$, then $\beta \in \text{dom}(g_i)$ for some i, so $g(\beta) = g_i(\beta) \in A_\beta$. Therefore $g \in \mathcal{P}$, and g is an upper bound on the chain. By Zorn's Lemma, $\mathcal{P}$ has a maximal element, f.

By maximality, the domain of f is $A -$ if not, we could extend the definition of f by adding to its domain a point α from $A - \text{dom}(f)$ and defining $f(\alpha)$ to be a point in the nonempty set A_α. Therefore $f \in \prod\{A_\alpha : \alpha \in A\}$, so $\prod\{A_\alpha : \alpha \in A\} \neq \emptyset$. •

In principle, it should be possible to rework any proof using transfinite induction into a proof that uses Zorn's Lemma and vice-versa. However, sometimes one is much more natural to use than the other.

We now present several miscellaneous examples that further illustrate how Zorn's Lemma is used.

The Countable Chain Condition and ϵ-discrete sets in (X, d)

Definition 11.2 Suppose (X, d) is a metric space and $\epsilon > 0$. A set $A \subseteq X$ is called ϵ-discrete if $d(x, y) \geq \epsilon$ for every pair $x \neq y \in A$.

Theorem 11.3 For every $\epsilon > 0$, (X, d) has a maximal ϵ-discrete set.

For example, $\mathbb{Z}$ is a maximal 1-discrete set in $\mathbb{R}$.

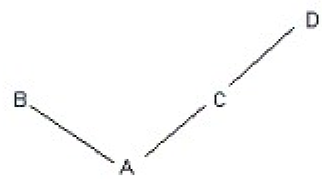
Proof Let $\epsilon > 0$. The theorem is clearly true if $X = \emptyset$, so we assume $X \neq \emptyset$.
Let $\mathcal{P} = \{A \subseteq X : A \text{ is } \epsilon\text{-discrete}\}$ and partially order $\mathcal{P}$ by inclusion $\subseteq$.
If $a \in X$, then $\{a\} \in \mathcal{P}$, so $\mathcal{P} \neq \emptyset$.

Suppose $\{C_\alpha : \alpha \in A\}$ is a chain in $(\mathcal{P}, \leq)$. We claim $\bigcup_{\alpha \in A} C_\alpha$ is ϵ-discrete.

If $x, y \in \bigcup_{\alpha \in A} C_\alpha$, then C_{α_1} and $y \in C_{\alpha_2}$ for some α_1, α_2. Since the C_α's form a chain, so $C_{\alpha_1} \subseteq C_{\alpha_2}$ or $C_{\alpha_2} \subseteq C_{\alpha_1}$. Without loss of generality, $C_{\alpha_1} \subseteq C_{\alpha_2}$. Then $x, y \in C_{\alpha_2}$, and this set is ϵ-discrete. Therefore $d(x, y) \geq \epsilon$.

Therefore $\bigcup_{\alpha \in A} C_\alpha \in \mathcal{P}$. Clearly, $\bigcup_{\alpha \in A} C_\alpha$ is an upper bound for the chain $\{C_\alpha : \alpha \in A\}$. By Zorn's Lemma $(\mathcal{P}, \leq)$ has a maximal element. •

Notice that when we use Zorn's Lemma, an upper bound that we produce for a chain in $(\mathcal{P}, \leq)$ – for example, $\bigcup_{\alpha \in A} C_\alpha$ in the preceding paragraph – might not itself be a maximal element in $\mathcal{P}$. For example, suppose $\mathcal{P}$ is a poset that contains exactly four sets A, B, C, D ordered by inclusion. A particular $\mathcal{P}$ is indicated in the following diagram:



One chain in $\mathcal{P}$ is $\{A, C\}$. Then $\bigcup\{A, C\} = C$ is an upper bound for the chain in $\mathcal{P}$. However, $C \subseteq D$ so C is not a maximal element in $\mathcal{P}$.

It is always true that an upper bound for a maximal chain in $\mathcal{P}$ (such as $\{A, C, D\}$, above) is a maximal element in $\mathcal{P}$ (why?). The statement that "in any poset, every chain is contained in a maximal chain" is called the Hausdorff Maximal Principle – and it is yet another equivalent to the Axiom of Choice.

Definition 11.4 A space X satisfies the countable chain condition (CCC) if every family of disjoint open sets in X is countable.

We already know that every separable space satisfies CCC. The following theorem shows that CCC is equivalent to separability among metric spaces.

Theorem 11.5 Suppose (X, d) is a metric space satisfying CCC. Then (X, d) is separable.

Proof For each $n \in \mathbb{N}$, we can use Theorem 11.3 to get a maximal $\frac{1}{n}$-discrete subset D_n. For $x \neq y \in D_n,\ B_{\frac{1}{2n}}(x) \cap B_{\frac{1}{2n}}(y) = \emptyset$, so, by CCC, the family $\{B_{\frac{1}{2n}}(x) : x \in D_n\}$ must be countable. Therefore D_n is countable, and we claim that the countable set $D = \bigcup_{n=1}^{\infty} D_n$ is dense.

Suppose $z \in X - D$ and, for $\epsilon > 0$, choose n so that $\frac{1}{n} < \epsilon$. Since D_n is a maximal $\frac{1}{n}$-discrete set, $D_n \cup \{z\}$ is not $\frac{1}{n}$-discrete, so there is a point $x \in D_n \subseteq D$ with $d(x, z) < \frac{1}{n} < \epsilon$. ●

Sets of cardinals are well-ordered

We proved this result earlier (Corollary 5.21) using ordinals. The following proof, due to Metelli and Salce, avoids any mention of ordinals $-$ but it makes heavy use of Zorn's Lemma and the Axiom of Choice. The statement that any set of cardinals is well-ordered is clearly equivalent to the following theorem.

Theorem 11.6 If $\{X_\alpha : \alpha \in A\}$ is a nonempty collection of sets, then $\exists\, \overline{\alpha} \in A$ such that $|X_{\overline{\alpha}}| \leq |X_\alpha|$ for every $\alpha \in A$ $-$ in other words, for each $\alpha \in A$ there exists a one-to-one map $\phi_{\overline{\alpha}\alpha} : X_{\overline{\alpha}} \to X_\alpha$.

Proof Assume all the X_α's are nonempty (otherwise the theorem is obviously true).

Let $\mathcal{P} = \{B \subseteq \prod X_\alpha :$ for all $\alpha \in A$, $\pi_\alpha|B$ is one-to-one$\}$, ordered by inclusion. $(\mathcal{P}, \subseteq)$ is a nonempty poset since $\emptyset \in \mathcal{P}$.

If $\{B_i : i \in I\}$ is any chain in $\mathcal{P}$, we claim that $B = \bigcup B_i \in \mathcal{P}$. Otherwise there would be points $x \neq y \in B$ and an α for which $\pi_\alpha(x) = \pi_\alpha(y)$. Since the B_i's form a chain, we would have both $x, y \in B_i$ for some i, and this would imply that $\pi_\alpha|B_i$ is not one-to-one.

B is an upper bound for the chain $\{B_i : i \in I\}$ in $\mathcal{P}$, so by Zorn's Lemma $\mathcal{P}$ contains a maximal element, M.

We claim that for some $\overline{\alpha} \in A$, $\pi_{\overline{\alpha}}|M : M \to X_{\overline{\alpha}}$ is onto.

> If not, then, since each X_α is nonempty, $X_\alpha - \pi_\alpha[M] \neq \emptyset$ for every α. By [AC], we can choose a point $y = (y_\alpha) \in \prod_{\alpha \in A}(X_\alpha - \pi_\alpha[M]) \neq \emptyset$. Since $y_\alpha \notin \pi_\alpha[M]$, $\pi_\alpha|(M \cup \{y\})$ would be one-to-one for all α so that $M \cup \{y\} \in \mathcal{P}$. Since $y \notin M$, $M \cup \{y\}$ is strictly larger than M and that is impossible because M is maximal.

Therefore $\pi_{\overline{\alpha}}|M : M \to X_{\overline{\alpha}}$ is a bijection and the map $\phi_{\overline{\alpha}\alpha} = \pi_\alpha \circ (\pi_{\overline{\alpha}}|M)^{-1} : X_{\overline{\alpha}} \to X_\alpha$ is one-to-one for every $\alpha \in A$. $\bullet$

Maximal ideals in a commutative ring with unit 1 Suppose K is a commutative ring with a unit element. (*If "commutative ring with unit" is unfamiliar, then just let $K = C(X)$ throughout the whole discussion. In that case, the "unit" is the constant function* $\mathbf{1}$.).

Definition 11.7 Suppose $I \subseteq K$, where K is a commutative ring with unit. A subset I of K is called an ideal in K if

1) $I \neq K$
2) $a, b \in I \;\Rightarrow\; a + b \in I$
3) $a \in I$ and $k \in K \;\Rightarrow\; ka \in I$

In other words, an ideal in K is a proper subset of K which is closed under addition and "superclosed" under multiplication.

An ideal I in K is called a maximal ideal if: whenever J is an ideal and $I \subseteq J$, then $I = J$. A maximal ideal is a maximal element in the poset of all ideals of K, ordered by inclusion.

For example, two ideals in $K = C(\mathbb{R})$ are:

$I = \{g : g = fi$ where $f, i \in C(\mathbb{R})$ and i is the identity function $i(x) = x\}$
For example, $g(x) = xe^x \in I$ (using $f(x) = e^x$).

$M = \{f : f(0) = 0\}$
In fact, M is a maximal ideal in $C(\mathbb{R})$ (T*his take a small amount of work to verify; see Exercise E33.*)

Maximal ideals, M, are important in ring theory – for example, if M is a maximal ideal in K, then the quotient ring K/M is actually a field.

Using Zorn's Lemma, we can prove

Theorem 11.8 Let K be a commutative ring with unit. Every ideal I in K is contained in a maximal ideal M. (*M might not be unique.)*

Proof Let $\mathcal{P} = \{J : J$ is an ideal and $I \subseteq J\}$. $(\mathcal{P}, \subseteq)$ is a nonempty poset since $I \in \mathcal{P}$. We want to show that $\mathcal{P}$ contains a maximal element.

Suppose $\{J_\alpha : \alpha \in A\}$ is a chain in $\mathcal{P}$. Let $J = \bigcup\{J_\alpha : \alpha \in A\}$. Since $I \subseteq J$, we only need to check that J is an ideal to show that $J \in \mathcal{P}$.

If $a, b \in J$, then $a \in J_\gamma$ and $b \in J_\beta$ for some $\gamma, \beta \in A$. Since the J_α's form a chain, either $J_\gamma \subseteq J_\beta$ or $J_\beta \subseteq J_\gamma$: say $J_\gamma \subseteq J_\beta$. Then a, b are both in the ideal J_β, so $a + b \in J_\beta \subseteq J$. Moreover, if $k \in K$, $ka \in J_\beta \subseteq J$. Therefore J is closed under addition and superclosed under multiplication.

Finally, $J \neq K$: If $J = K$, then $1 \in J$ so $1 \in J_\alpha$ for some α. Then, for all $k \in K$, $k = k \cdot 1 \in J_\alpha$. So $J_\alpha = K$, which is impossible since J_α is an ideal.

By Zorn's Lemma, we conclude that $\mathcal{P}$ contains a maximal element M. •

Basis for a Vector Space It is assumed here that you know the definition of a vector space V over a field K. (*K is the set of "scalars" which can multiply the vectors in $\mathcal{V}$. If "field" is unfamiliar, then you may just assume $K = \mathbb{R}$, $\mathbb{Q}$, or $\mathbb{C}$ in the result.*) Beginning linear algebra courses usually only deal with finite-dimensional vector spaces: the number of elements in a basis for V is called the dimension of V. But some vector spaces are infinite dimensional. Even then, a basis exists, as we now show.

Definition 11.9 Suppose V is a vector space over a field K. A collection of vectors $B \subseteq V$ is called a vector space basis for V if each nonzero $v \in V$ can be written in a unique way as a finite linear combination of elements of B using nonzero coefficients from K. More formally, B is a basis if for each $v \in V$, $v \neq 0$, there exist unique nonzero $\alpha_1, ..., \alpha_n \in K$ and unique $b_1, ..., b_n \in B$ such that $v = \sum_{i=1}^{n} \alpha_i b_i$. V is called finite dimensional if a finite basis B exists.

Definition 11.10 A set of vectors $C \subseteq V$ is called linearly independent if, whenever $\alpha_1, ..., \alpha_n \in K$ and $c_1, ..., c_n \in C$ and $\sum_{i=1}^{n} \alpha_i c_i = 0$, then $\alpha_1 = ... = \alpha_n = 0$ (in other words, the only linear combination of elements of C adding to 0 is the trivial combination).

Theorem 11.11 Every vector space $V \neq \{0\}$ over a field K has a basis. (*The trivial vector space* $\{0\}$ *cannot have a basis under our definition: the only nonempty subset* $\{0\}$ *is not linearly independent*.)

Proof We will use Zorn's Lemma to show that there is a maximal linearly independent subset of $\mathcal{V}$ and that it must be a basis.

Let $\mathcal{P} = \{C \subseteq V : C$ is linearly independent$\}$, ordered by $\subseteq$. For any $0 \neq v \in V$, we have $\{v\} \in \mathcal{P}$, so $(\mathcal{P}, \subseteq)$ is a nonempty poset.

Let $\{C_\alpha : \alpha \in I\}$ be a chain in $\mathcal{P}$. We claim $C = \bigcup C_\alpha$ is linearly independent, so that $C \in \mathcal{P}$.

> If $\alpha_1, ..., \alpha_n \in K$ and $c_1, ..., c_n \in C$ and $\sum_{i=1}^{n} \alpha_i c_i = 0$, then each c_i is in some C_{α_i}. The C_α's form a chain with respect to $\subseteq$, so one of the C_{α_i}'s – call it C_{α^*} – contains all the others. Then $0 = \sum_{i=1}^{n} \alpha_i c_i$ is a linear combination of elements from C_{α^*}, and C_{α^*} is linearly independent. So all the α_i's must be 0.

Therefore, $C \in \mathcal{P}$ and C is an upper bound for the chain $\{C_\alpha : \alpha \in I\}$. By Zorn's Lemma, $\mathcal{P}$ has a maximal element B.

We claim that B is a basis for V.

> First we show that if $v \in V$, then v can be written as a finite linear combination of elements of $\mathcal{B}$.
>
> > If $v \in B$, then $v = 1 \cdot v$. If $v \notin B$, then $B \subsetneqq B \cup \{v\}$ so, by maximality, $B \cup \{v\}$ is not linearly independent. That means there is a nontrivial linear combination of elements of $B \cup \{v\}$ (*necessarily involving* v) with sum 0 :
> >
> > $$\exists\, b_1, ..., b_n \text{ and } \exists\, \alpha_1, ..., \alpha_n, \beta \in K \text{ with } \beta \neq 0 \text{ and } \sum_{i=1}^{n} \alpha_i b_i + \beta v = 0.$$
> >
> > Since $\beta \neq 0$, we can solve the equation and write $v = \sum_{i=1}^{n} \frac{-\alpha_i}{\beta} b_i$.
>
> We complete the proof by showing that such a representation for v is unique.
>
> > Suppose that we have $\sum_{i=1}^{n} \alpha_i b_i = v = \sum_{i=1}^{m} \alpha_i' b_i'$ with $\alpha_i, \alpha_i' \in K$ and $b_i, b_i' \in B$. By allowing additional b's with 0 coefficients as necessary, we may assume that the same elements of B are used on both sides of the equation, so that $\sum_{i=1}^{k} \beta_i b_i'' = v = \sum_{i=1}^{k} \gamma_i b_i''$. Then $\sum_{i=1}^{k} (\beta_i - \gamma_i) b_i'' = 0$ and B is linearly independent, so $\beta_i = \gamma_i$ for $i = 1, ..., k$.

Example 11.12 In Example II.5.14 we showed that the only continuous functions $f : \mathbb{R} \to \mathbb{R}$ satisfying the functional equation

$$f(x + y) = f(x) + f(y) \text{ for all } x, y \in \mathbb{R} \qquad (*)$$

are the linear functions $f(x) = cx$ for some $c \in \mathbb{R}$. Now we can see that there are other functions, necessarily discontinuous, that satisfy (*).

Consider $\mathbb{R}$ as a vector space over the field $\mathbb{Q}$, and let $B \subseteq \mathbb{R}$ be a basis. Choose any $b_1 \in B$. Then for each $x \in \mathbb{R}$, there is a unique expression $x = q_x b_1 + \sum_{i=2}^{n} q_i b_i$ for some $b_2, ..., b_n \in B$ and $q_x, q_2, ..., q_n \in \mathbb{Q}$. (We can insist that b_1 be part of this sum by allowing $q_x = 0$ when necessary.). Define $f : \mathbb{R} \to \mathbb{Q} \subseteq \mathbb{R}$ by $f(x) = q_x$. Clearly $f(x + y) = f(x) + f(y)$.

Although the definition of f looks complicated, perhaps it could happen, for a cleverly chosen $c \in \mathbb{R}$, that $f(x) = cx$ for all x? No: we show that $f(x) = cx$ is not possible (and therefore f is not continuous).

1) If $f(x) = cx$ for some constant c, we would have:

$f(\sqrt{2}) = c \cdot \sqrt{2} \in \mathbb{Q}$, and
$f(1) = c \cdot 1 = c \in \mathbb{Q}$. But $c \neq 0$, or else $f(x) = 0$ for every x, whereas $f(b) = 1 \cdot b = 1$. So we can divide by c getting
$c\sqrt{2}/c = \sqrt{2} \in \mathbb{Q}$, which is false.

2) Here is a different argument to the same conclusion: For every $b \in \mathcal{B}$ where $b \neq b_1$, we have $b = 0 \cdot b_1 + 1 \cdot b$, so $f(b) = 0$. Therefore the equation $f(x) = 0$ has infinitely many solutions so $f(x)$ is not linear.

As we remarked earlier in Example II.5.14, it can be shown that discontinuous solutions f for (*) must be "not Lebesgue measurable" (a nasty condition that implies that f must be "extremely discontinuous.")

A "silly" example from measure theory (optional) It is certainly possible for an uncountable union of sets of measure 0 to have measure 0. For example, let $I_p = \{x \in \mathbb{Q} \cap [0, 1] : x < p\}$ for each irrational $p \in [0, 1]$. There are uncountably many I_p's and each one has measure 0. In this case, $\bigcup I_p \subseteq \mathbb{Q}$, so $\bigcup I_p$ also has measure 0.

Might it be true that every union of sets of measure 0 (say, in $[0, 1]$) must have measure 0? (It is easy to answer this question: how?) What follows is an "unnecessarily complicated" answer using Zorn's Lemma.

If every such union had measure zero, we could apply Zorn's Lemma to the poset consisting of all measure-0 subsets of $[0, 1]$ and get a maximal subset M with measure 0 in $[0, 1]$. Since $[0, 1]$ does not have measure 0, there is a point $p \in [0, 1] - M$. Then $M \cup \{p\}$ also has measure 0, contradicting the maximality of M.

Exercises

E24. Let (X, d) be a metric space. In Definition 10.3, we defined collections $\mathcal{G}_\alpha$ $(\alpha < \omega_1)$ and the collection of Borel sets $\mathcal{B} = \bigcup\{\mathcal{G}_\alpha : \alpha < \omega_1\}$.

a) Let $\mathcal{F}_0$ be the family of closed sets in X. For $\alpha < \omega_1$, define families

$$\mathcal{F}_\alpha = \begin{cases} \{\bigcap_{n=1}^\infty F_n : F_n \in \mathcal{F}_{\beta_n}, \beta_n < \alpha\} & \text{if } \alpha \text{ is even} \\ \{\bigcup_{n=1}^\infty F_n : F_n \in \mathcal{F}_{\beta_n}, \beta_n < \alpha\} & \text{if } \alpha \text{ is odd} \end{cases}$$

$\mathcal{F}_1$ is the collection of countable unions of closed sets (called F_σ-sets) and $\mathcal{F}_2$ is the family of countable intersections of F_σ sets (called $F_{\sigma\delta}$-sets).

Prove that $\mathcal{F}_\alpha \subseteq \mathcal{G}_\beta$ and $\mathcal{G}_\alpha \subseteq \mathcal{F}_\beta$ for all $\alpha < \beta < \omega_1$.

It follows that $\mathcal{B} = \bigcup\{\mathcal{F}_\alpha : \alpha < \omega_1\}$. We can build the Borel sets "from the bottom up" beginning with either the open sets or the closed sets. Would this be true if we defined Borel sets the same way in an arbitrary topological space?)

b) Suppose X and Y are separable metric spaces. A function $f: X \to Y$ is called <u>Borel-measurable</u> (or <u>B-measurable</u>, for short) if $f^{-1}[B]$ is a Borel set in X whenever B is a Borel set in Y. Prove that f is B-measurable iff $f^{-1}[O]$ is Borel in X whenever O is open in Y.

c) Prove that there are $\leq c$ B-measurable maps f from X to Y.

E25. a) Is a locally finite cover of a space X necessarily point finite? Is a point finite cover necessarily locally finite (*see Definitions 10.10, 10.13*)? Give an example of a space X and an open cover $\mathcal{U}$ that satisfies one of these properties but not the other.

b) Suppose F_i $(i = 1, ..., n)$ are closed sets in the normal space X with $\bigcap_{i=1}^n F_i = \emptyset$. Prove that there exist open sets V_i $(i = 1, ..., n)$ such that $F_i \subseteq V_i$ and $\bigcap_{i=1}^n \overline{V}_i = \emptyset$.
Hint: Use the characterization of normality in Theorem 10.12.

E26. Prove that the continuum hypothesis ($c = \aleph_1$) is true iff $\mathbb{R}^2$ can be written as $A \cup B$ where A has countable intersection with every horizontal line and B has countable intersection with every vertical line.
Hints: $\Rightarrow$: *See Exercise I.E44. If CH is true, $\mathbb{R}$ can be indexed by the ordinals $< \omega_1$.*
$\Leftarrow$: *If CH is false, then $c > \aleph_1$. Suppose A meets every horizontal line only countably often and that $A \cup B = \mathbb{R}^2$. Show that B meets some vertical line uncountably often by letting Q be the union of any $\aleph_1$ horizontal lines and examining $\pi_X[Q \cap A]$.*)

E27. A cover $\mathcal{U}$ of the space X is called irreducible if it has no proper subcover.

a) Give an example of an open cover of a noncompact space which has no irreducible subcover.

b) Prove that X is compact iff every open cover has an irreducible subcover.

Hint: Let $\mathcal{U}$ be any open cover and let $\mathcal{A}$ be a subcover with the smallest possible cardinality m. Let γ be the least ordinal with cardinality m. Index $\mathcal{A}$ using the ordinals less than γ, so that $\mathcal{A} = \{U_\alpha : \alpha < \gamma\}$. Then consider $\{V_\beta : \beta < \gamma\}$, where $V_\beta = \bigcup\{U_\alpha : \alpha < \beta\}$.

E28. Two set theory students, Ray and Deborah, are arguing.

> Ray: "There must be a maximal countable set of real numbers. Look: partially order the countable infinite subsets of $\mathbb{R}$ by inclusion. Now every chain of such sets has an upper bound (remember, a countable union of countable sets is countable), so Zorn's Lemma gives us a maximal element."
>
> Deborah: "I don't know anything about Zorn's Lemma but it seems to me that you can always add another real to any countable set of real numbers and still have a countable set. So how can there be a largest one?"
>
> Ray: "I didn't say largest! I said maximal!"

Resolve their dispute.

E29. Suppose $(P, \leq)$ is a poset. Prove that $\leq$ can be "enlarged" to a relation $\leq^*$ such that $(P, \leq^*)$ is a chain. ("$\leq^*$ enlarges $\leq$" means that "$\leq \subseteq \leq^* \subseteq P \times P$")
Hint: Suppose $\leq$ is a linear ordering on P. Can $\leq$ be enlarged?

E30. Let m be a cardinal. A space X has caliber m if, whenever $\mathcal{U}$ is a family of open sets with $|\mathcal{U}| = m$, there is a family $\mathcal{V} \subseteq \mathcal{U}$ such that $|\mathcal{V}| = m$ and $\bigcap\{V : V \in \mathcal{V}\} \neq \emptyset$.

a) Prove that every separable space has caliber $\aleph_1$.

b) Prove that any product of separable spaces has caliber $\aleph_1$.
Hint: Recall that a product of c separable spaces is separable (Theorem VI.3.5).

c) Prove that if X has caliber $\aleph_1$, then X satisfies the countable chain condition (*see Definition 11.4*).

d) Let X be a set of cardinal $\aleph_1$ with the cocountable topology. Is X separable? Does X satisfy the countable chain condition? Does X have caliber $\aleph_1$? (*For notational convenience, you can assume, without loss of generality, that $X = [0, \omega_1)$.*)

E31. a) Prove that there exists an infinite maximal family $\mathcal{E}$ of infinite subsets of $\mathbb{N}$ with the property that the intersection of any two sets from $\mathcal{E}$ is finite.

b) Let $D = \{\omega_E : E \in \mathcal{E}\}$ be a set of distinct points such that $D \cap \mathbb{N} = \emptyset$. Let $Z = \mathbb{N} \cup D$, with the following topology:

i) points of $\mathbb{N}$ are isolated
ii) a basic neighborhood of ω_E is any set containing ω_E and all but at most finitely many points of E.

Prove Z is Tychonoff.

c) Prove that Z is not countably compact. *Hint: Consider the set D*

d) Prove that Z is pseudocompact. *Hint: This proof uses the maximality of $\mathcal{E}$.*

E32. According to Theorem V.5.10, the closed interval $[0,1]$ cannot be written nontrivially as a countable union of pairwise disjoint nonempty closed sets. Of course, $[0,1]$ can certainly be written as the union of c such sets: for example, $[0,1] = \bigcup_{x \in [0,1]} \{x\}$.

Prove that $[0,1]$ can be written as the union of uncountably many pairwise disjoint closed sets each of which is countably infinite.

Hint: Use Zorn's Lemma to choose a maximal family $\mathcal{F}$ of subsets of $[0,1]$ each homeomorphic to $\{\frac{1}{n} : n \in \mathbb{N}\} \cup \{0\}$. Let $A = [0,1] - \bigcup \mathcal{F}$. A is relatively discrete and therefore countable. For each $x \in A$, choose a different $C_x \in \mathcal{F}$ and replace C_x by $C_x \cup \{x\}$.)

E33. Suppose X is Tychonoff. For $p \in X$, let $M_p = \{f \in C(X) : f(p) = 0\}$. Clearly, M_p is an ideal in $C(X)$.

a) Prove that M_p is a maximal ideal in $C(X)$.

b) Prove that is X is compact iff every maximal ideal in $C(X)$ is of the form M_p for some $p \in X$.

E34. Prove that there exists a subset A of $\mathbb{R}$ that has only countably many distinct "translates" $A_r = \{a + r : a \in A\}$ – that is only countably many of the sets $A_r = \{a + r : a \in A\}$, $r \in \mathbb{R}$ are distinct).

Hint: Consider $\mathbb{R}$ as a vector space over the field $\mathbb{Q}$, and pick a basis B. Pick a point $b \in B$ and consider all reals whose expression as a finite linear combination of elements of B does not involve b.

Appendix
Exponentiation of Ordinals: A Sketch

The appendix gives a brief sketch about exponentiation of ordinals. Some of the details are omitted. The main point is to explain why ordinals like $\omega_0^{\omega_0}$ are still countable ordinals. (See Example 5.25, part 3)

If μ and α are ordinals, let

$$\mu^\alpha = \{f \in [0,\mu)^{[0,\alpha)} : f(\xi) = 0 \text{ for all but finitely many } \xi < \alpha\}.$$

We can think of a point f in μ^α as a "transfinite sequence" in $[0,\mu)$, where f has well-ordered domain $[0,\alpha)$ rather than $\mathbb{N}$. Using "sequence-like" notation, we can write:

$$\mu^\alpha = \{(a_\xi) : 0 \le \xi < \alpha, 0 \le a_\xi < \mu, \text{ and } a_\xi = 0 \text{ for all but finitely many } \xi\}$$

We put an ordering on μ^α by:

Given $(a_\xi) \neq (\mathrm{b}_\xi)$ in μ^α, let ν be the largest index for which $a_\nu \neq b_\nu$. We write $(a_\xi) < (b_\xi)$ if $a_\nu < b_\nu$.

Example For $\alpha = \mu = 2$, μ^α consists of 4 pairs ordered as follows: $(0,0) < (1,0) < (0,1) < (1,1)$.

We also use μ^α to denote the order type associated with $(\mu^\alpha, \le)$. It turns out that $(\mu^\alpha, \le)$ is well-ordered, so this order type is actually an ordinal number.

Example $\omega_0^{\omega_0}$ is represented by the set of all sequences in $\{0,1,2,...\}$ which are eventually 0. (*The sequences in the set $\omega_0^{\omega_0}$ turn out to be those which are "eventually 0" because each element of ω_0 has has only finitely many predecessors. In general, the condition that members of μ^α be 0 for all but finitely many $\xi < \alpha$ is much stronger than merely saying "eventually 0."*)

The order relation on $\omega_0^{\omega_0}$ between two sequences is determined by comparing the largest term at which the sequences differ. For example,

$$(5,0,0,1,2,0,0,0,0,...) < (107,12,1,3,5,0,0,0,0,...) < (103,7,0,0,6,0,0,0,0...,)$$

The initial segment of $\omega_0^{\omega_0}$ representing ω_0 consists of:

$$(0,0,0,...,...) < (1,0,0,...,...) < (2,0,0,...,...) < \ ... \ < (n,0,0,...,...) < \ ...$$

$\omega_0^{\omega_0}$ is clearly a countable ordinal. A little thought shows that $\omega_0^{\omega_0} = \omega_0 + \omega_0^2 + \omega_0^3 + ...$

Once μ^α is defined, it is easy to see that the ordinals $\omega_0^{\omega_0}$, $\omega_0^{\omega_0^{\omega_0}}$, $\omega_0^{\omega_0^{\omega_0^{\omega_0}}}$, ... are all countable ordinals. (*Here, α^{β^γ} is understood, as usual, to mean $\alpha^{(\beta^\gamma)}$*).

We can then define $\epsilon_0 = \sup\ \{\omega_0^{\omega_0}, \omega_0^{\omega_0^{\omega_0}}, \omega_0^{\omega_0^{\omega_0^{\omega_0}}}, \ldots\}$. Roughly, ϵ_0 is "ω_0 raised to the ω_0 power ω_0 times". As noted earlier, the sup of a countable set of countable ordinals is still countable: ϵ_0 is a countable ordinal, that is, $\epsilon_0 < \omega_1$!

Exercise Prove that $\mu^1 = \mu$. Prove that for ordinals μ and α, μ^α is also an ordinal, i.e., $(\mu^\alpha, \leq)$ is a well-ordered set (not trivial!). For ordinals $\mu, \alpha, \beta > 0$, and $\mu^{\alpha+\beta} = \mu^\alpha \cdot \mu^\beta$.

For further information, see Sierpinski, Cardinal and Ordinal Numbers, pp. 309 ff.

Chapter VIII Review

Explain why each statement is true, or provide a counterexample.

1. Let $X = [0, \omega_1)$. For each $\alpha \in X$, let $\alpha \sim \alpha + 1$. In the quotient space $X/\sim$, exactly one point is isolated.

2. $\aleph_{\omega_0}^{\aleph_0}$ can be written as a sum of countably many smaller cardinals.

3. Let $n \in \mathbb{N}$. A continuous function $f : [0, \omega_1) \to \mathbb{R}^n$ is constant on a tail of $[0, \omega_1)$.

4. Every order-dense chain with more than 1 point contains a subset order isomorphic to $\mathbb{Q}$.

5. If $x \in O$, where O is an open set in $X = [0, \omega_1] \times [0, \omega_0] - \{(\omega_0 + 1, \omega_0)\}$, then there must be a continuous function $f \colon X \to \mathbb{R}$ such that $x \in \text{coz}\,(f) \subseteq O$.

6. Let C be a non-Borel subset of the unit circle S^1 and let $f : S^1 \to \mathbb{R}$ be the characteristic function of C. Then f is not Borel measurable but f is Borel measurable in each variable separately (that is, for each $a \in [-1, 1]$, the functions ϕ and ψ defined by $\phi(x) = f(x, a)$ and $\psi(a, y) = f(a, y)$ are Borel measurable.

7. Let C be a dense subset of $\mathbb{R}$. Then C is not well-ordered (in the usual order on $\mathbb{R}$).

8. Let $D = \{f \in \mathbb{Q}^{\mathbb{N}} : f(n) = 0$ for all but finitely many $n\}$, with the lexicographic ("dictionary") ordering $\leq$. Then $(D, \leq)$ is order isomorphic to $\mathbb{Q}$.

9. Suppose $\leq_1$ and $\leq_2$ are linear orders on a set X. If $\leq_1 \subseteq \leq_2$, then $\leq_1 = \leq_2$.

10. Consider the ordinals 1, ω_0 and $\omega_0 \cdot 2$. Considering all possible sums of these ordinals (in the six different possible orders) produces exactly 3 distinct values.

11. If α and β are nonzero ordinals and $a + \beta = \omega_0$, then $\alpha\beta = \omega_0$.

12. The order topology on $(1, 3) \cup (3, 5)$ is the same as the subspace topology from $\mathbb{R}$.

13. If $(X, \mathcal{T})$ is a finite topological space, then there exists a partial ordering $\leq$ on X for which $\mathcal{T}$ is the order topology.

14. If O is open and F is closed in $T = [0, \omega_1] \times [0, \omega_0] - \{(\omega_1, \omega_0)\}$ and $F \subseteq O$, then there must exist an open set V such that $F \subseteq V \subseteq \text{cl}\, V \subseteq O$.

15. There are exactly $\aleph_\alpha$ limit ordinals $< \omega_\alpha$.

16. If α denotes the order type of the irrationals, then $\alpha^2 = \alpha$.

17. If $\alpha > \omega_1$, then the ordinal space $X = [0, \alpha]$ is not metrizable because X is not normal.

18. Suppose $\mathcal{P}$ is a nonempty collection containing all subsets of X with a certain property (*). Suppose $\{P_\alpha : \in A\}$ is a chain of subsets of $\mathcal{P}$ and that $\bigcup P_\alpha \in \mathcal{P}$. By Zorn's Lemma, $\bigcup P_\alpha$ is a maximal subset of X (with respect to $\subseteq$) having property (*).

19. For any infinite cardinal k, there are exactly k^+ ordinals with cardinality k.

20. Suppose α is an infinite order type. If $1 + \alpha = \alpha + 1$, then there is an order type ξ (possibly 0) such that $\alpha = \omega_0 + \xi + \omega_0^*$.

21. If $\leq$ a linear order on X, then the "reversed relation" $\leq^*$ defined by $x \leq^* y$ iff $y \leq x$ is also a linear order on X.

22. For any infinite cardinal k, $2^k = k^k$ (without GCH).

23. If $\leq$ is a linear ordering on a finite set S and $|S| = n$, then $|\leq| = \frac{n(n+1)}{2}$.

24. If A and B are disjoint closed sets in $[0, \omega_1)$, then at least one of them is countable.

25. A locally finite open cover (by nonempty sets) of a compact space must be finite.

26. $[0, \omega_1]$ is homeomorphic to a subspace of the Cantor set.

27. Let $\mathbb{N}^{\aleph_0}$ have the lexicographic order $\leq$. $(\mathbb{N}^{\aleph_0}, \leq)$ is not well-ordered, because the set $A = \{x \in \mathbb{N}^{\aleph_0} : \exists k \ \forall n > k \ \ x(n) = 2\}$ contains no smallest element.

28. Every subset of $\mathbb{Q}$ is a Borel set in $\mathbb{R}$.

29. With the order topology, the set $\{0\} \cup \{x \in \mathbb{R} : |x| > 1\}$ is homeomorphic to $\mathbb{R}$.

30. If $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ is a point-finite open cover of the Sorgenfrey plane $S \times S$, then there must be an open cover $\mathcal{V} = \{V_\alpha : \alpha \in A\}$ of $S \times S$ such that, for each $\alpha \in A$, $\mathrm{cl}\,(V_\alpha) \subseteq U_\alpha$.

31. There is a countably infinite compact connected metric space.

32. If $\leq$ is a linear ordering on X, then there can be linear ordering $\leq'$ for which

$$\leq \ \underset{\neq}{\subseteq} \ \leq' \subseteq X \times X.$$

Chapter IX
Theory of Convergence

1. Introduction

In Chapters II and III, we discussed convergence of sequences. One result was, true in any space X, was: if (a_n) is a sequence in $A \subseteq X$ and $(a_n) \to x$, then $x \in \text{cl}\, A$.

Sequential convergence is especially useful for work in first countable spaces – in particular, in metric spaces. In a first countable space, $x \in \text{cl}\, A$ iff there is a sequence (a_n) in A with $(a_n) \to x$, and therefore convergent sequences determine the set cl A. It follows we can check whether or not a set A is closed (whether $A = \text{cl}\, A$) by using sequences. So in a first countable space "sequences determine the topology." (*See Theorem III.9.6.*)

However, in general sequences are not sufficient to describe closures. For example in $[0, \omega_1]$ we know that $\omega_1 \in \text{cl}\, [0, \omega_1)$ but no sequence (α_n) in $[0, \omega_1,)$ converges to ω_1. The basic neighborhoods $(\alpha, \omega_1]$ of ω_1 are very nicely ordered, but there are just "too many" of them: no mere sequence (α_n) in $[0, \omega_1)$ is "long enough" to be eventually inside every neighborhood of ω_1.

A second example is the space L (*see Example III.9.8*) in which we have $(0,0) \in \text{cl}\, (L - \{(0,0)\})$ but no sequence in $(L - \{(0,0)\})$ converges to $(0,0)$. Here the problem is different. L is a "small" space ($|L| = \aleph_0$) but the neighborhood system $\mathcal{N}_{(0,0)}$ (ordered by reverse inclusion) is a very complicated poset – so complicated that a mere sequence (a_n) in $L - \{(0,0)\}$ cannot eventually be in every neighborhood of $(0,0)$.

In this chapter we develop a theory of convergence that is sufficient to describe the topology in any space X. We define a kind of "generalized sequence" called a net. A sequence is a function $f \in X^{\mathbb{N}}$, and we write $f(n) = x_n$. A net is a function $f \in X^{\Lambda}$, where $(\Lambda, \le)$ is a more general kind of ordered set. Informally, we write a sequence f as (x_n); similarly, when f is a net, then we informally write the net as (x_λ).

Thinking of $X = [0, \omega_1]$, we might hope that it would be a sufficient generalization to replace $\mathbb{N}$ with an initial segment of ordinals $\Lambda = [0, \alpha)$: in other words, to replace sequences with "transfinite sequences" with domain some well-ordered set "longer" than $\mathbb{N}$. In fact, this is a sufficient generalization to deal with a case like $[0, \omega_1]$: if (x_α) is the "transfinite sequence" in $[0, \omega_1)$ given by $x_\alpha = \alpha$ $(\alpha < \omega_1)$, then $(x_\alpha) \to \omega_1$ (in the sense that (x_α) is eventually in every neighborhood of ω_1). But as we will see below, such a generalization does not go far enough (*see Example 2.9*). We need a generalization that uses some kind of ordered set Λ more complicated than just initial segments $[0, \alpha)$ of ordinals.

The theory of nets turns out to have a "dual" formulation in the theory of filters. In this chapter we will discuss both formulations. It turns out that nets and filters are fully equivalent formulations of convergence, but sometimes one is more natural to use than the other.

2. Nets

Definition 2.1 A nonempty ordered set $(\Lambda, \leq)$ is called a directed set if

1) $\leq$ is transitive and reflexive
2) for all $\lambda_1, \lambda_2 \in \Lambda$, there exists $\lambda_3 \in \Lambda$ such that $\lambda_3 \geq \lambda_1$ and $\lambda_3 \geq \lambda_2$.

Example 2.2

1) Any chain (for example, $\mathbb{N}$) is a directed set.

2) In $\mathbb{R}$, define $x \leq^* y$ if $|x| \geq |y|$. Then $(\mathbb{R}, \leq^*)$ is a directed set and $x \leq^* y$ means that "x is at least as far from the origin as y." Notice that $-1 \leq^* 1$ and $1 \leq^* -1$, so that $\leq^*$ is not antisymmetric. A directed set might not be a poset.

3) Let $\mathcal{N}_x$ be the neighborhood system at x in X, ordered by "reverse inclusion" – that is, $N_1 \leq N_2$ iff $N_1 \supseteq N_2$. $\mathcal{N}_x \neq \emptyset$ and condition 1) in the definition clearly holds. Condition 2) is satisfied because for $N_1, N_2 \in \mathcal{N}_x$, we have $N_1 \cap N_2 = N_3 \in \mathcal{N}_x$ and $N_3 \geq N_1$ and $N_3 \geq N_2$. Therefore $(\mathcal{N}_x, \leq)$ is a directed set.

This example hints at why replacing $\mathbb{N}$ (in the definition of a sequence) with a directed set Λ (in the definition of net) will give a tool strong enough to describe the topology in any space X: the directed set Λ can chosen to be as complicated as the most complicated system of neighborhoods at a point x.

4) Let $\Lambda = \{F \subseteq [0,1] : 0 \in F, 1 \in F$ and F is finite$\}$. In analysis, F is called a partition of the interval $[0,1]$. Order Λ by inclusion: a "larger" partition is a "finer" one – one with more subdivision points. Then Λ is a directed set.

Definition 2.3 i) A net in a set X is a function $f : (\Lambda, \leq) \to X$, where $(\Lambda, \leq)$ is directed. We write $f(\lambda) = x_\lambda$ and, informally, denote the net by (x_λ).

ii) A net (x_λ) in a space X converges to $x \in X$ if (x_λ) is eventually in every neighborhood of x – that is, $\forall\, N \in \mathcal{N}_x\ \exists\, \lambda_0 \in \Lambda$ such that $x_\lambda \in N$ wherever $\lambda \geq \lambda_0$.

iii) A point x in a space X is a cluster point of a net (x_λ) if the net is frequently in every neighborhood of x, that is: for all $N \in \mathcal{N}_x$ and all $\lambda_0 \in \Lambda$ there is a $\lambda \geq \lambda_0$ for which $x_\lambda \in N$.

Clearly, every sequence is a net, and when $\Lambda = \mathbb{N}$, the definitions ii), and iii) are the same as the old definitions for convergence and cluster point of a sequence.

Example 2.4 $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ and define an order in which $\lambda_3 \geq \lambda_1$ and $\lambda_3 \geq \lambda_2$ but λ_1 and λ_2 are not comparable. $(\Lambda, \leq)$ is a directed set. We can define a net $f : \Lambda \to \mathbb{R}$ by $f(\lambda_3) = 0$ and assigning any real values to $f(\lambda_1)$ and $f(\lambda_2)$. Since $x_\lambda \in (-\epsilon, \epsilon)$ for all $\lambda \geq \lambda_3$, we have $(x_\lambda) \to 0$. (*Notice that a net can have a finite domain and can have a "last term."*)

Example 2.5 The following examples indicate how several different kinds of limit that come up in analysis can all be reformulated in terms of net convergence. In other words, many different "limit" definitions in analysis are "unified" by the concept of net convergence.

1) Let $\Lambda = \mathbb{R} - \{a\}$ and define $x \leq^* y$ (in Λ) iff $|x - a| \geq \backslash y - a|$ (in $\mathbb{R}$). Suppose $f : \Lambda \to \mathbb{R}$ is a net. Then the net $(x_\lambda) \to L \in \mathbb{R}$

iff for all $\epsilon > 0$ there exists $\lambda_0 \in \Lambda$ such that $x_\lambda \in (L - \epsilon, L + \epsilon)$ if $\lambda \geq^* \lambda_0$
iff for all $\epsilon > 0$ there exists $\lambda_0 \in \Lambda$ such that $x_\lambda \in (L - \epsilon, L + \epsilon)$ if $0 < |\lambda - a| \leq |\lambda_0 - a|$
iff for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(\lambda) - L| < \epsilon$ if $0 < |\lambda - a| < \delta$
iff $\lim_{x \to a} f(x) = L$ (in the usual sense of analysis).

2) Let $\Lambda = \mathbb{R}$ with the usual order $\leq$. If $f : \mathbb{R} \to \mathbb{R}$, we can think of f as a net, and write $f(r) = x_r$. Then the net $(x_r) \to L \in \mathbb{R}$

iff for all $\epsilon > 0$ there exists $r_0 \in \Lambda = \mathbb{R}$ such that $x_r \in (L - \epsilon, L + \epsilon)$ if $r \geq r_0$
iff for all $\epsilon > 0$ there exists $r_0 \in \mathbb{R}$ such that $|f(r) - L| < \epsilon$ if $r \geq r_0$
iff $\lim_{x \to \infty} f(x) = L$ (in the usual sense of analysis).

Parts 1) and 2) show how two different limits in analysis can each be expressed as convergence of a net. In fact each of the limits $\lim_{x \to a} f(x) = L$, $\lim_{x \to a^+} f(x) = L$, $\lim_{x \to a^-} f(x) = L$, $\lim_{x \to \infty} f(x) = L$, and $\lim_{x \to -\infty} f(x) = L$ can be expressed as the convergence of some net. In each case, the trick is to choose the proper directed set.

3) Integrals can also be defined in terms of the convergence of nets.

Let g be a bounded real valued function defined on $[0,1]$ and let $\Lambda = \{F \subseteq [0,1] : F$ is finite and $0 \in F,\ 1 \in F\}$. Order Λ by inclusion: $F_1 \leq F_2$ iff $F_1 \subseteq F_2$ iff F_2 is a "finer" partition than F_1.

For $F \in \Lambda$, enumerate F as $0 = x_0 < x_1 < \ldots < x_{n-1} < x_n = 1$. Let $\Delta x_i = x_i - x_{i-1}$ and

$$f(F) = \sum_{i=1}^{n} (\inf_{[x_{i-1}, x_i]} g) \cdot \Delta x_i = x_F \in \mathbb{R}.$$

$f : \Lambda \to \mathbb{R}$ is a net in $\mathbb{R}$ and we can write $f(F) = x_F$. Then

$(x_F) \to L$ iff for all $\epsilon > 0$ there is a partition F_0 such that
for all partitions F finer than F_0, $|x_F - L| < \epsilon$

One can show that such an L always exists: L is called the lower integral of g over $[0,1]$, denoted $L = \underline{\int}_0^1 g$.

If we replace "inf" by "sup" in defining f, then the limit U of the net is called the upper integral of g over $[0,1]$, denoted $U = \overline{\int}_0^1 g$. If $L = U$, we say that g is (Riemann) integrable on $[0,1]$ and write $\int_0^1 g = L\ (= U)$.

A sequence can have more than one limit: for example, if $|X| > 1$ and X has the trivial topology, <u>every</u> sequence (x_n) in X converges to <u>every</u> point in X. Since a sequence <u>is</u> a net (with $\Lambda = \mathbb{N}$), a net can have more than one limit. We also proved that a sequence in a Hausdorff space can have at most one limit (Theorem III.9.3). This theorem still holds for nets, but then even more is true: uniqueness of net limits actually <u>characterizes</u> Hausdorff spaces.

Theorem 2.6 A space X is Hausdorff <u>iff</u> every net in X has at most one limit.

Proof Let $x \neq y \in X$ with $(x_\lambda) \to x$ and $(x_\lambda) \to y$. Suppose U and V are open sets with $x \in U$ and $y \in V$. Then $(x_\lambda) \in U$ for all $\lambda \geq$ some λ_1 and $(x_\lambda) \in V$ for all $\lambda \geq$ some λ_2. If $\lambda_3 \geq \lambda_1$ and $\lambda_3 \geq \lambda_3$, then $x_{\lambda_3} \in U \cap V$, so $U \cap V \neq \emptyset$. Therefore X is not Hausdorff.

Conversely, suppose X is not Hausdorff. Then there are points $x \neq y \in X$ such that $U \cap V \neq \emptyset$ whenever $U \in \mathcal{N}_x$ and $V \in \mathcal{N}_y$. Let $\Lambda = \mathcal{N}_x \times \mathcal{N}_y$ and order Λ by defining $(U, V) \leq (U', V')$ iff $U' \subseteq U$ and $V' \subseteq V$ (reverse inclusion in both coordinates). For each $(U, V) \in \Lambda$, let $f((U, V)) =$ any point $x_{(U,V)} \in U \cap V$. We claim that $(x_{(U,V)}) \to x$ and $(x_{(U,V)}) \to y$.

> If $V \in \mathcal{N}_y$, let $\lambda_0 = (X, V) \in \Lambda$. Then if $\lambda = (U', V') \geq \lambda_0$, we have $x_\lambda = x_{(U',V')} \in U' \cap V' \subseteq X \cap V = V$. Therefore $(x_\lambda) \to y$.
> The proof that $(x_\lambda) \to x$ is similar. $\bullet$

Example 2.7 The implication ($\Leftarrow$) in Theorem 2.6 is <u>false</u> for sequences. Let $X = [0, \omega_1] \cup \{\omega_1^*\}$ where $\omega_1^* \notin [0, \omega_1]$. Let points in $[0, \omega_1]$ have their usual neighborhoods and let a basic neighborhood of ω_1^* be a set of the form $(\alpha, \omega_1] - \{\omega_1\} \cup \{\omega_1^*\}$, where $\alpha < \omega_1$. (*In effect, the basic neighborhoods of ω_1 and ω_1^* are identical except for replacing ω_1 by ω_1^* or vice versa; it is as if ω_1^* is a "shadow" of ω_1 which can't be separated from ω_1).*

This space is <u>not</u> Hausdorff, but we claim that a sequence (α_n) in X has at most one limit.

> If (α_n) contains infinitely many terms $\alpha_{n_k} < \omega_1$, then let $\alpha = \sup\{\alpha_{n_k} : k = 1, 2, ...\} < \omega_1$, so that the subsequence (α_{n_k}) is in $[0, \alpha]$. If $(\alpha_n) \to \beta$ and $(\alpha_n) \to \gamma$, then $(\alpha_{n_k}) \to \beta$ and $(\alpha_{n_k}) \to \gamma$. Therefore β, γ are both in the closed set $[0, \alpha]$. But $[0, \alpha]$ is T_2 so a limit for (α_{n_k}) must be unique: $\beta = \gamma$.
>
> If only finitely many α_n's are less than ω_1, we can assume without loss of generality that for all n, $\alpha_n \in \{\omega_1, \omega_1^*\}$. It is easy to see that if $\alpha_n = \omega_1$ for only finitely many n, then $(\alpha_n) \to \omega_1^*$ only. Similarly, if $\alpha_n = \omega_1^*$ for only finitely many n, then $(\alpha_n) \to \omega_1$ only. If $\alpha_n = \omega_1$ and $\alpha_n = \omega_1^*$ each for infinitely many n, then (α_n) has no limit.

The next theorem tells us that nets are sufficient to describe the topology in any space X.

Theorem 2.8 Suppose $A \subseteq (X, \mathcal{T})$. Then $x \in \text{cl}\, A$ iff there is a net (a_λ) in A for which $(a_\lambda) \to x$.

Proof Suppose $(a_\lambda) \to x$, where (a_λ) is a net in A. For each $N \in \mathcal{N}_x$, (a_λ) is eventually in N so $N \cap A \neq \emptyset$. Therefore $x \in \text{cl}\, A$.

We can prove the converse because we can choose a directed set "as complicated as" the neighborhood system of x: Suppose $x \in \mathrm{cl}\, A$ and let $\Lambda = \mathcal{N}_x$, ordered by reverse inclusion. For $\lambda = N \in \Lambda$, let $f(\lambda) = a_\lambda =$ a point chosen in $N \cap A$. If $\lambda_0 = N_0 \in \Lambda$ and $\lambda = N \geq \lambda_0$ then $a_\lambda \in N \cap A \subseteq N_0 \cap A \subseteq N_0$. Therefore $(a_\lambda) \to x$. •

Example 2.9 In general, "sequences aren't sufficient" to describe the topology of a space X. However we might ask whether something simpler than nets will do. For example, suppose we consider "transfinite sequences" – that is, very special nets $f : [0, \alpha) \to X$, where α is some ordinal. It turns out that these are not sufficient.

Suppose $X = [0, \omega_1) \times [0, \omega_0) \subseteq T^* = [0, \omega_1] \times [0, \omega_0]$. Let $p =$ the "upper right corner point" (ω_1, ω_0). Then $p \in \mathrm{cl}\, X$, but we claim that no transfinite sequence in X can converge to $p \in T^*$.

Let $f : [0, \alpha) \to X$ and write $f(\lambda) = (x_\lambda, y_\lambda)$. We claim that $((x_\lambda, y_\lambda)) \not\to p$ (*no matter how large an* α *we use!*). So suppose $((x_\lambda, y_\lambda)) \to p$:

ran(f) must be uncountable.

If ran(f) were countable, then the set $C = \{x_\lambda : \lambda < \alpha\}$ would be countable and $\beta = \sup \{x_\lambda : \lambda < \alpha\} < \omega_1$. Then ran$(f) \subseteq [0, \beta] \times [0, \omega_0)$ and so (x_λ, y_λ) could not converge to p.

For every countable set $E \subseteq [0, \alpha)$, sup $E < \alpha$:

Since E well-ordered, E represents an ordinal $\beta \leq \alpha$, so there is an order isomorphism $g : [0, \beta) \to E$ for some $\beta \leq \alpha$.

Certainly sup $E \leq \alpha$.

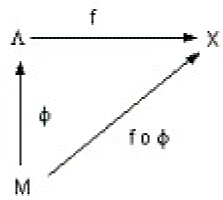
If $\sup(E) = \alpha$, then $f \circ g : [0, \beta) \to X$ would be a transfinite sequence with countable range converging to p (since f does). But we have shown that a net converging to p must have uncountable range. Therefore sup $E < \alpha$.

For each $m < \omega_0$, let $E_m = \{\lambda < \alpha : y_\lambda = m\}$ and $\lambda_m = \sup E_m \leq \alpha$ (" $\leq$ " is the most we can say since E_m might be uncountable). For each $\beta < \alpha$, β is in some E_m – so, for that m, $\lambda_m \geq \beta$. Therefore $\sup \{\lambda_m : m < \omega_0\} = \alpha$. By the preceding paragraph, we conclude $\lambda_{m_0} = \alpha$ for some m_0.

Since $(x_\lambda, y_\lambda) \to p$, we conclude that $(x_\lambda, y_\lambda)_{\lambda \in E_{m_0}} \to p$. But $(x_\lambda, y_\lambda)_{\lambda \in E_{m_0}} \to (\omega_1, m)$, so this is impossible. •

The definition of a subnet is analogous to the definition of a subsequence (*see Definition III.10.1*).

Definition 2.10 Let Λ and M be directed sets Suppose $f : \Lambda \to X$, $\phi\colon M \to \Lambda$. Suppose that for each $\lambda_0 \in \Lambda$, there exists a $\mu_0 \in M$ such that $\phi(\mu) \geq \lambda_0$ whenever $\mu \geq \mu_0$,



then $f \circ \phi$ is called a subnet of f. We write $\phi(\mu) = \lambda_\mu$ and $f(\phi(\mu)) = x_{\lambda_\mu}$. Informally, the subnet is (x_{λ_μ}). The definition of subnet guarantees that $\phi(\mu) = \lambda_\mu \geq \lambda_0$ whenever $\mu \geq \mu_0$.

> *Note: an alternate definition for subnet is used in some books. It requires that*
>
> *1) ϕ is increasing: if $\mu_1 \leq \mu_2$, then $\phi(\mu_1) \leq \phi(\mu_2)$ and*
>
> *2) ϕ is cofinal in Λ: for each $\lambda_0 \in \Lambda$, there is a $\mu_0 \in M$ for which $\phi(\mu_0) \geq \lambda_0$.*
>
> *A subnet in the sense of this definition is also a subnet in the sense of Definition 2.10, but the definitions are not equivalent. Definition 2.10 is "more generous" – it allows more subnets because Definition 2.10 does not require ϕ to be increasing. For most purposes, the slight disagreement in the definitions doesn't matter. However, the full generality of Definition 2.10 is required to develop the full duality between nets and filters that we discuss later.*

We will state the following theorem, for now, without proof. The proof will be easier after we talk about filter convergence. For now, we simply want to use the theorem to highlight an observation in Example 2.12.

Theorem 2.11 In $(X, \mathcal{T})$, if x is a cluster point of the net (x_λ), then there is a subnet $(x_{\lambda_\mu}) \to x$.

Proof See Corollary 4.8 later in this chapter.

Example 2.12 (*See Example III.9.8*) In $L = \{(m, n) : m, n$ are nonnegative integers$\}$, all points except $(0, 0)$ are isolated. Basic neighborhoods of $(0, 0)$ are sets containing $(0, 0)$ and "most of the points from most of the columns" (where "most" means "all but finitely many").

Let (x_n) be an enumeration of $L - \{(0, 0)\}$. Although $(0, 0)$ is a cluster point of (x_n), we proved in Example III.9.8 that no sequence in $L - \{(0, 0)\}$ can converge to $(0, 0)$. In particular, no subsequence of (x_n) can converge to $(0, 0)$. However, Theorem 2.11 implies that there is a subnet of (x_n) that converges to $(0, 0)$.

This might seem surprising, but it simply highlights something is actually clear from the definition of subnet: even if a net is a sequence ($\Lambda = \mathbb{N}$), the directed set M in the definition of subnet need not be $\mathbb{N}$. Therefore a subnet of a sequence might not be a sequence.

3. Filters

$X^{\mathbb{N}}$ is the set of all sequences in X, but the "set of all nets in X" makes no sense in ZFC. We can put an order $\leq$ on any set Λ to create a directed set $(\Lambda, \leq)$: for example (using Zermelo's Theorem) we could let $\leq$ be a well-ordering of Λ. Therefore there at least as many directed sets $(\Lambda, \leq)$ as there are sets Λ. The "set" of all nets in X = "$\bigcup\{X^{(\Lambda, \leq)} : (\Lambda, \leq)$ is a directed set$\}$" is "too big" to be a set in ZFC. This is not only an aesthetic annoyance; it is also a serious set-theoretic disadvantage for certain purposes.

Therefore we look at an equivalent way to describe convergence, one which is strong enough to describe the topology of any space X but which doesn't have this set-theoretic drawback. The theory of filters does this, and it turns out to be a theory "dual" to the theory of nets $-$ that is, there is a natural "back-and-forth" between nets and filters that converts each theorem about nets to a theorem about filters and vice-versa.

Definition 3.1 A filter $\mathcal{F}$ in a set X is a nonempty collection of subsets of X such that

i) $\emptyset \notin \mathcal{F}$
ii) $\mathcal{F}$ is closed under finite intersections
iii) If $F \in \mathcal{F}$ and $G \supseteq F$, then $G \in \mathcal{F}$.

A nonempty family $\mathcal{B} \subseteq \mathcal{F}$ is called a filter base for $\mathcal{F}$ if $\mathcal{F} = \{F \subseteq X : F \supseteq B$ for some $B \in \mathcal{B}\}$.

Assuming, as we have stated, that filters will provide an equivalent theory of convergence in a space X, we see that there is no longer a set-theoretic issue. The set of all filters in X makes perfectly good sense: each filter $\mathcal{F} \in \mathcal{P}(\mathcal{P}(X))$, so the set of all filters in X is just a subset of $\mathcal{P}(\mathcal{P}(X))$.

In a topological space X, there is a completely familiar example of a filter : $\mathcal{N}_x$, the neighborhood system at x. A base for this filter is what we have been calling a neighborhood base, $\mathcal{B}_x$. In fact, $\mathcal{N}_x$ and $\mathcal{B}_x$ are what inspired the general definitions of filter and filter base in the first place.

If we start with a filter $\mathcal{F}$, then a base $\mathcal{B}$ for a filter $\mathcal{F}$ must have certain properties: $\mathcal{B}$ is nonempty (or else $\mathcal{F} = \emptyset$), and each $B \in \mathcal{B}$ is nonempty (or else $\emptyset \in \mathcal{F}$). Moreover, if B_1, $B_2 \in \mathcal{B} \subseteq \mathcal{F}$, then $B_1 \cap B_2 \in \mathcal{F}$ so (by definition of a base) $B_1 \cap B_2 \supseteq B_3$ for some $B_3 \in \mathcal{B}$.

If, on the other hand, we start with any nonempty collection $\mathcal{B}$ of nonempty sets in X such that

whenever $B_1, B_2 \in \mathcal{B}$, there is a $B_3 \in \mathcal{B}$ such that $B_1 \cap B_2 \supseteq B_3$ (*)

and define $\mathcal{F} = \{F : F \supseteq B$ for some $B \in \mathcal{B}\}$, then $\mathcal{F}$ is a filter and $\mathcal{B}$ is a base for $\mathcal{F}$ (*check!*). $\mathcal{F}$ is called is called the filter generated by $\mathcal{B}$; it is the smallest filter containing $\mathcal{B}$.

If we begin with a nonempty collection $\mathcal{S}$ of nonempty sets in X with the finite intersection property, then

$\mathcal{B} = \{B : B$ is a finite intersection of sets from $\mathcal{S}\}$ has property (*)

so $\mathcal{B}$ is a base for a filter $\mathcal{F}$ called the filter generated by $\mathcal{S}$; it is smallest filter containing $\mathcal{S}$.

There can be many neighborhood bases $\mathcal{B}_x$ for a neighborhood system $\mathcal{N}_x$ in a space X. Similarly, a filter $\mathcal{F}$ can have many different filter bases $\mathcal{B}$. In particular, notice that $\mathcal{F}$ itself is a filter base – but usually we want to choose the simplest base possible.

Definition 3.2 A filter base $\mathcal{B}$ converges to $x \in X$ if for every $N \in \mathcal{N}_x$, there is a $B \in \mathcal{B}$ such that $B \subseteq N$. In this case we write $\mathcal{B} \rightarrow x$. (Since a filter $\mathcal{F}$ is also a filter base, we have also just defined the meaning of $\mathcal{F} \rightarrow x$.)

If $\mathcal{F}$ is the filter generated by $\mathcal{B}$, then clearly $\mathcal{B} \rightarrow x$ iff $\mathcal{F} \supseteq \mathcal{N}_x$. (**)

The following theorem is very simple. It is stated explicitly just to make sure that there are no confusions.

Theorem 3.3 Let $\mathcal{F}$ be a filter in the space X. For any $x \in X$, the following are equivalent:

i) $\mathcal{F} \rightarrow x$
ii) $\mathcal{F} \supseteq \mathcal{N}_x$
iii) $\mathcal{F}$ has a base $\mathcal{B}$ where $\mathcal{B} \rightarrow x$.
iv) Every base $\mathcal{B}'$ for $\mathcal{F}$ satisfies $\mathcal{B}' \rightarrow x$.

Proof From Definition 3.2 and the observation (**) shows that i) $\Leftrightarrow$ ii) $\Leftrightarrow$ iii).

iii) $\Rightarrow$ iv) Suppose $\mathcal{B}$ is a base for $\mathcal{F}$ and that $\mathcal{B} \rightarrow x$. Then for each $N \in \mathcal{N}_x$, there is a $B \in \mathcal{B}$ for which $N \supseteq B$. If $\mathcal{B}'$ is another base for $\mathcal{F}$ then, since $B \in \mathcal{F}$, B contains some set $B' \in \mathcal{B}'$. Therefore $N \supseteq B \supseteq B'$, so $\mathcal{B}' \rightarrow x$. •

iv) $\Rightarrow$ i) because $\mathcal{F}$ is a base for $\mathcal{F}$. •

Example 3.4

1) Let $x \in (X, \mathcal{T})$. If $\mathcal{B}_x$ is a neighborhood base at x, then $\mathcal{B}_x \to x$. In particular, $\mathcal{N}_x \to x$.

2) Suppose $\mathcal{F}$ is a filter in X. If $A \subseteq X$ and $A \cap F \neq \emptyset$ for all $F \in \mathcal{F}$, then the collection $\mathcal{F} \cup \{A\} = \mathcal{S}$ is a nonempty collection with the finite intersection property. Therefore $\mathcal{S}$ generates a filter $\mathcal{F}' \supseteq \mathcal{F}$. So: if $A \notin \mathcal{F}$ and $A \cap F \neq \emptyset$ for all $F \in \mathcal{F}$, we can enlarge $\mathcal{F}$ to a filter $\mathcal{F}'$ that also contains the set A.

3) If $x_0 \in X$, then $\mathcal{F} = \{A \subseteq X : x_0 \in A\}$ is a filter. Since $\mathcal{F} \supseteq \mathcal{N}_{x_0}$, we have $\mathcal{F} \to x_0$. The simplest base for $\mathcal{F}$ is $\mathcal{B} = \{\{x_0\}\}$ and $\mathcal{B} \to x_0$.

Suppose $B \notin \mathcal{F}$. Then $x_0 \notin B$ so $\mathcal{F}$ cannot be enlarged to a filter $\mathcal{F}'$ that also contains the set B – such an $\mathcal{F}'$ would contain both $\{x_0\}$ and B, and therefore also contain $\emptyset = \{x_0\} \cap B$, which is impossible. (*This also follows from Theorem 3.5, below.*)

Therefore, $\mathcal{F}$ is a maximal filter in X. A maximal filter is called an ultrafilter.

For each $x_0 \in X$, $\mathcal{F} = \{A \subseteq X : x_0 \in A\}$ is called a trivial ultrafilter. In general, it takes some more work (and AC) to decide whether a set X also contains "nontrivial" ultrafilters – that is, ultrafilters not of this form. (*See Theorem 5.4 and the examples that follow*).

The next theorem gives us a simple characterization of ultrafilters. It is completely set-theoretic, not topological.

Theorem 3.5 A filter $\mathcal{F}$ in X is an ultrafilter iff for every $A \subseteq X$, either $A \in \mathcal{F}$ or $X - A \in \mathcal{F}$.

Proof ($\Rightarrow$) If $A \subseteq X$ and $A \notin \mathcal{F}$, then $A \supseteq F$ is false for every $F \in \mathcal{F}$, so $F \cap (X - A) \neq \emptyset$ for all $F \in \mathcal{F}$. Therefore the collection $\mathcal{S} = \mathcal{F} \cup \{X - A\}$ has the finite intersection property, so $\mathcal{S}$ generates a filter $\mathcal{F}' \supseteq \mathcal{F}$. So if $\mathcal{F}$ is an ultrafilter, we must have $\mathcal{F}' = \mathcal{F}$ and therefore $X - A \in \mathcal{F}$.

($\Leftarrow$) Suppose that for every $A \subseteq X$, either A or $X - A$ is in $\mathcal{F}$. If $\mathcal{F}'$ is a filter and $\mathcal{F}' \supseteq \mathcal{F}$, then $\mathcal{F}' = \mathcal{F}$: for if $A \in \mathcal{F}' - \mathcal{F}$, then $X - A \in \mathcal{F}$ which would mean $A \cap (X - A) = \emptyset \in \mathcal{F}'$. •

Note: The proof ($\Rightarrow$) shows that if a filter $\mathcal{F}$ contains neither A nor $X - A$, then $\mathcal{F}$ can be enlarged to a new filter $\mathcal{F}'$ containing either A or $X - A$ – whichever of the two you wish.

Definition 3.6 A point $x \in X$ is a cluster point for a filter base $\mathcal{B}$ if $N \cap B \neq \emptyset$ for every $N \in \mathcal{N}_x$ and every $B \in \mathcal{B}$. (Since a filter $\mathcal{F}$ is also a filter base, we have also just defined a cluster point for a filter $\mathcal{F}$.)

It is immediate from the definition that x is a cluster point of a filter base $\mathcal{B}$ iff x is in the closure of each set in $\mathcal{B}$, that is, if and only if $x \in \bigcap\{\text{cl}\, B : B \in \mathcal{B}\}$.

Clearly, if $\mathcal{B} \to x$, then x is a cluster point of $\mathcal{B}$ (*explain!*).

Theorem 3.7 Suppose $\mathcal{F}$ is a filter in a space X. For $x \in X$, the following are equivalent:

i) x is a cluster point of $\mathcal{F}$
ii) $x \in \bigcap\{\text{cl}\, F : F \in \mathcal{F}\}$
iii) there is a filter base $\mathcal{B}$ for $\mathcal{F}$ such that x is a cluster point of $\mathcal{B}$
iv) for every filter base $\mathcal{B}$ for $\mathcal{F}$, x is a cluster point of $\mathcal{B}$.

Proof It is clear from the definition and the following remarks that i) $\Leftrightarrow$ ii) $\Leftrightarrow$ iii).

iii) $\Rightarrow$ iv) Suppose x is a cluster point of $\mathcal{B}$ and $\mathcal{B}'$ is another base for $\mathcal{F}$. If $B' \in \mathcal{B}' \subseteq \mathcal{F}$ then $B' \supseteq B$ for some set $B \in \mathcal{B}$ – because $\mathcal{B}$ is a base for $\mathcal{F}$. But every neighborhood of x intersects B, so every neighborhood of x also intersects B'. Therefore x is a cluster point of $\mathcal{B}'$.

iv) $\Rightarrow$ i) This follows since $\mathcal{F}$ is a base for $\mathcal{F}$. •

Example 3.8

1) Let $\mathcal{F} = \{F \subseteq \mathbb{R} : F \supseteq [0,1]\}$. Each real number r has a neighborhood N for which $N \not\supseteq [0,1]$ and therefore $N \notin \mathcal{F}$. So $\mathcal{F} \not\supseteq \mathcal{N}_r$ so $\mathcal{F} \not\to r$.

If $r \in [0,1]$, then $r \in A \subseteq \text{cl}\, A$ for each $A \in \mathcal{F}$, so r is a cluster point of $\mathcal{F}$.

If $r \notin [0,1]$, then $N = \mathbb{R} - [0,1]$ is a neighborhood of r disjoint from $[0,1] \in \mathcal{F}$. Therefore r is not a cluster point of $\mathcal{F}$. So the set of cluster points of $\mathcal{F}$ is precisely the interval $[0,1]$.

2) Let $\mathbb{N}$ have the usual topology. $\mathcal{F} = \{A \subseteq \mathbb{N} : \mathbb{N} - A \text{ is finite}\}$ is a filter (*check!*). For each $n \in \mathbb{N}$, we have $F = \{n+1, n+2, ...\} \in \mathcal{F}$ and $n \notin F = \text{cl}\, F$. Therefore n is not a cluster point of $\mathcal{F}$. And since $\mathcal{F}$ has no cluster points, certainly $\mathcal{F}$ does not converge.

3) Let $\mathcal{F}$ be a filter in a space X. If $\mathcal{F} \to x$ and $\mathcal{F} \to y$, then $\mathcal{F} \supseteq \mathcal{N}_x$ and $\mathcal{F} \supseteq \mathcal{N}_y$. Since $\emptyset \notin \mathcal{F}$, every neighborhood of x must intersect every neighborhood of y. Therefore, if X is Hausdorff, we must have $x = y$, that is, <u>a filter in a Hausdorff space can have at most one limit</u>. (*In fact, X is Hausdorff* <u>*iff*</u> *every filter in $\mathcal{F}$ in X has at most one limit. We could prove this directly,here and now (try it!). But this fact follows later "for free" from Theorem 2.6 via the duality between nets and filters: see Corollary 4.5 below.*)

4. The Relationship Between Nets and Filters

Nets and filters are dual to each other in a natural way – it is possible to move back-and-forth between them. Although this back-and-forth process is not perfectly symmetric, it is still very useful because the process preserves limits and cluster points.

Definition 4.1 Let $(\Lambda, \leq)$ be a directed set. For a net (x_λ) in X, we define its associated filter $\mathcal{F}$ (or, the filter generated by (x_λ)) :

Let $T_\lambda = \{x_\mu : \mu \in \Lambda, \mu \geq \lambda\}$ = the λ^{th} tail of the net, and let $\mathcal{B} = \{T_\lambda : \lambda \in \Lambda\}$. $\mathcal{B}$ is nonempty since $\Lambda \neq \emptyset$, and each $T_\lambda \neq \emptyset$ since $x_\lambda \in T_\lambda$. Moreover, if $\lambda_3 \geq \lambda_1$ and $\lambda_3 \geq \lambda_2$, then $T_{\lambda_1} \cap T_{\lambda_2} \supseteq T_{\lambda_3}$. Therefore the collection of tails $\mathcal{B} = \{T_\lambda : \lambda \in \Lambda\}$ is a filter base and the filter it generates is the associated filter of (x_λ).

Definition 4.2 For a filter $\mathcal{F}$ in X, we define its associated net (or, the net generated by $\mathcal{F}$) :

Let $\Lambda = \{(x, F) : x \in F \in \mathcal{F}\}$ and define $(x, F) \leq (x', F')$ if $F \supseteq F'$. Then $(\Lambda, \leq)$ is directed. (*Why?*) The net $f : \Lambda \to X$ defined by $f((x, F)) = x$ is the associated net of $\mathcal{F}$.

Notice that $(\Lambda, \leq)$ is not a poset: for example if $x \neq x' \in F \in \mathcal{F}$, then $(x, F) \geq (x', F)$ and $(x', F) \geq (x, F)$ but $(x, F) \neq (x', F)$.

If we begin with a filter $\mathcal{F}$, form its associated net, and then form its associated filter $\mathcal{F}'$, we are back to where we started:

$$\mathcal{F} \quad \rightsquigarrow \quad \text{associated net } (x_\lambda) \quad \rightsquigarrow \quad \text{associated filter } \mathcal{F}' = \mathcal{F}$$

To see this: a base for $\mathcal{F}'$ is the collection of tails $\{T_{(x,F)} : (x, F) \in \Lambda\}$. But
$T_{(x,F)} = \{f(x', F') : (x', F') \geq (x, F) = \{x' : (x', F') \geq (x, F)\}$
$= \{x' : x' \in F' \text{ and } F' \subseteq F, \text{ where } F, F' \in \mathcal{F}\} = \bigcup\{F' : F' \subseteq F \in \mathcal{F}\} = \mathcal{F}$. So the collection of tails is $\mathcal{F}$ itself!

On the other hand, if we begin with a net $(x_\lambda : \lambda \in \Lambda\}$, form its associated filter $\mathcal{F}$, and then form the net associated with $\mathcal{F}$, we do not return to the original net (x_λ) :

$$(x_\lambda) \quad \rightsquigarrow \quad \text{associated filter } \mathcal{F} \quad \rightsquigarrow \quad \text{net associated to } \mathcal{F} \neq (x_\lambda)$$

First of all, the net associated with $\mathcal{F}$ has for its directed set
$\Lambda' = \{(x, F) : x \in F \in \mathcal{F}\} \neq \Lambda$. But the problem runs even deeper than that:

Consider the net (x_n) in $\mathbb{R}$ with directed set $\mathbb{N}$, where $x_n = 0$ for all n. The collection of tails is $\mathcal{B} = \{\{0\}\}$, which generates the associated $\mathcal{F} = \{A \subseteq \mathbb{R} : 0 \in A\}$ (a trivial ultrafilter). In turn, $\mathcal{F}$ generates a net whose directed set $\Lambda = \{(x, A) : x \in A \in \mathcal{F}\}$ and in Λ, $(0, \{0\})$ is a maximal element. The directed set for the new net is not even order isomorphic to $\mathbb{N}$!

Nevertheless, the following theorem shows that this back-and-forth process between associated nets and associated filters is good enough to be very useful for topological purposes.

Theorem 4.3 In any space X,

1) x is a cluster point of a net (x_λ) iff x is a cluster point of the associated filter $\mathcal{F}$.
2) x is a cluster point of a filter $\mathcal{F}$ iff x is a cluster point of the associated net (x_λ)

Proof 1) x is a cluster point of (x_λ) iff (x_λ) is frequently in every neighborhood of x
iff $T_\lambda \cap N \neq \emptyset$ for every tail T_λ and every $N \in \mathcal{N}_x$
iff x is a cluster point of the filter base $\mathcal{B} = \{T_\lambda : \lambda \in \Lambda\}$
iff x is a cluster point of $\mathcal{F}$.

2) We use 1). Given a filter $\mathcal{F}$, consider its associated net (x_λ). By 1), x is a cluster point of (x_λ) iff x is a cluster point of its associated filter $\mathcal{F}'$. But $\mathcal{F}' = \mathcal{F}$. •

Notice: if we had somehow proved part 2) first, and then tried to use 2) to prove part 1), we would run into trouble – we would end up looking at a net different from the one we started with. The "asymmetry" in the back-and-forth process between nets and filters shows up here. .

Theorem 4.4 In any space X,

1) a net $(x_\lambda) \to x$ iff its associated filter $\mathcal{F} \to x$.
2) a filter $\mathcal{F} \to x$ iff its associated net $(x_\lambda) \to x$.

Proof The proof is left as an exercise. As in Theorem 4.3, part 2) follows "for free" from 1) using the duality between nets and filters. •

Corollary 4.5 A space X is Hausdorff iff every filter $\mathcal{F}$ has at most one limit in X.

Proof The result follows immediately by duality: use Theorem 4.4 and Theorem 2.6. •

The following theorem shows how subnets and larger filters are related: subnets generate larger filters and vice-versa.

Theorem 4.6 Suppose $\mathcal{F}$ is a filter in X generated by some net $f : \Lambda \to X$. (*This is not a restriction on $\mathcal{F}$ because every filter is generated by a net: for example, by its associated net.*)

1) Each subnet of f generates a filter $\mathcal{G} \supseteq \mathcal{F}$
2) Each filter $\mathcal{G} \supseteq \mathcal{F}$ is generated by a subnet of f.

Proof 1) A base for $\mathcal{F}$ is $\{B_\lambda : \lambda \in \Lambda\}$, where B_λ is the λ^{th} tail of f. Suppose $\phi : M \to \Lambda$ and that $f \circ \phi : M \to X$ is a subnet of f. The filter base $\{C_\mu : \mu \in M\}$, where C_μ is the μ^{th} tail of $f \circ \phi$, generates a filter $\mathcal{G}$.

Let $F \in \mathcal{F}$. Then $F \supseteq B_{\lambda_0}$ for some $\lambda_0 \in \Lambda$. Pick $\mu_0 \in M$ so that $\mu \geq \mu_0 \Rightarrow \phi(\mu) \geq \lambda_0$. Then the subnet tail $C_{\mu_0} = \{f \circ \phi(\mu) : \mu \geq \mu_0\} \subseteq B_{\lambda_0} \subseteq F$, so F is in the filter $\mathcal{G}$ generated by $f \circ \phi$. Therefore $\mathcal{F} \subseteq \mathcal{G}$.

2) Conversely, let $\mathcal{G}$ be a filter containing $\mathcal{F}$. We claim $\mathcal{G}$ is generated by a subnet of f. Let B_λ denote the λ^{th} tail of f. We claim that $\mathcal{B} = \{G \cap B_\lambda : G \in \mathcal{G}, \lambda \in \Lambda\}$ is a base for $\mathcal{G}$:

$\mathcal{B}$ is clearly base for some filter $\mathcal{F}'$ and $\mathcal{G} \subseteq \mathcal{F}'$ (since $G \supseteq G \cap B_\lambda$). On the other hand, each $B_\lambda \in \mathcal{F} \subseteq \mathcal{G}$, so each $G \cap B_\lambda \in \mathcal{G}$ so $\mathcal{F}' \subseteq \mathcal{G}$

Let $M = \{(\xi, G \cap B_\lambda) : G \in \mathcal{G},\ \lambda, \xi \in \Lambda,\ \xi \geq \lambda$ and $f(\xi) = x_\xi \in G \cap B_\lambda\}$. (For each $G \cap B_\lambda$ there is at least one such ξ because $G \cap B_\lambda \neq \emptyset$). Order M by defining $(\xi, G \cap B_\lambda) \leq (\xi', G' \cap B_{\lambda'})$ if $\lambda' \geq \lambda$ and $G' \cap B_{\lambda'} \subseteq G \cap B_\lambda$. It is easy to check that $\leq$ is transitive and reflexive. In fact, $(M, \leq)$ is a directed set:

> If $\lambda'' \geq \lambda, \lambda'$, we have $B_{\lambda''} \subseteq B_{\lambda'} \cap B_\lambda$. Since $G'' = G \cap G' \in \mathcal{G}$, we have $G'' \cap B_{\lambda''} \neq \emptyset$, so there exists a $\xi'' \geq \lambda''$ such that $f(\xi'') = x_{\xi''} \in G'' \cap B_{\lambda''}$. Thus $(\xi'', G'' \cap B_{\lambda''}) \in M$, and we have $(\xi'', G'' \cap B_{\lambda''}) \geq (\xi, G \cap B_\lambda)$ and $(\xi'', G'' \cap B_{\lambda''}) \geq (\xi', G' \cap B_{\lambda'})$.

Define $\phi : M \to \Lambda$ by $\phi((\xi, G \cap B_\lambda)) = \xi$. Then $f \circ \phi$ is a subnet of f:

> Suppose $\lambda_0 \in \Lambda$. Pick any $G_0 \in \mathcal{G}$ and any $\xi_0 \geq \lambda_0$ such that $f(\xi_0) = x_{\xi_0} \in G_0 \cap B_{\lambda_0}$. Let $\mu_0 = (\xi_0, G_0 \cap B_{\lambda_0}) \in M$. If $\mu = (\xi, G \cap B_\lambda) \geq \mu_0$ in M, then $\phi(\mu) = \xi \geq \xi_0 \geq \lambda_0$.

The filter generated by the subnet $f \circ \phi$ has the set of tails C_{μ_0} a base. If $\mu_0 = (\xi_0, G_0 \cap B_{\lambda_0}) \in M$, we claim that the tail $C_{\mu_0} = G_0 \cap B_{\lambda_0}$. <u>If this is true, we are done</u>, since the sets of this form, as noted above, are a base for the filter $\mathcal{G}$.

> $C_{\mu_0} = \{f \circ \phi(\mu) : \mu \geq \mu_0\}$. If $\mu \in M$ and $\mu = (\xi, G \cap B_\lambda) \geq \mu_0 = (\xi_0, G_0 \cap B_{\lambda_0})$, then $G \cap B_\lambda \subseteq G_0 \cap B_{\lambda_0}$, so $f \circ \phi(\mu) = f(\xi) \in G \cap B_\lambda \subseteq G_0 \cap B_{\lambda_0}$.
> Therefore $C_{\mu_0} \subseteq G_0 \cap B_{\lambda_0}$.
>
> Conversely, if $y \in G_0 \cap B_{\lambda_0}$, then $y = f(\xi) = x_\xi$ for some $\xi \in \Lambda$, $\xi \geq \lambda_0$.
> Then $\mu = (\xi, G_0 \cap B_\xi) \in M$ and $\mu \geq \mu_0$ so $y = f(\xi) = f \circ \phi(\mu) \in C_{\mu_0}$.
> Therefore $C_{\mu_0} \supseteq G_0 \cap B_{\lambda_0}$. •

The remark following Definition 2.10 is relevant here. To prove part 2) of Theorem 4.6: we need the more generous definition of subnets to be sure that we have "enough" subnets to generate all possible filters containing $\mathcal{F}$. In particular, the subnet defined in proving part 2) is <u>not</u> a "subnet" using the more restrictive definition for subnet.

Theorem 4.7 A point $x \in X$ is a cluster point of the filter $\mathcal{F}$ iff there exists a filter $\mathcal{G} \supseteq \mathcal{F}$ such that $\mathcal{G} \to x$.

Proof ($\Rightarrow$) If such a filter $\mathcal{G}$ exists, then $\mathcal{N}_x \subseteq \mathcal{G}$. Therefore each neighborhood of x intersects each set in $\mathcal{G}$, and therefore, in particular, intersects each set in $\mathcal{F}$. Therefore x is a cluster point of $\mathcal{F}$.

($\Leftarrow$) If x is a cluster point of $\mathcal{F}$, then the set $\mathcal{B} = \{N \cap F : N \in \mathcal{N}_x$ and $F \in \mathcal{F}\}$ is a filter base that generates a filter $\mathcal{G}$. For $F \in \mathcal{F}$, we have $F = X \cap F \in \mathcal{B}$, so $\mathcal{F} \subseteq \mathcal{G}$; and for each $N \in \mathcal{N}_x$, we have $N = N \cap X \in \mathcal{B}$, so $\mathcal{N}_x \subseteq \mathcal{G}$ and therefore $\mathcal{G} \to x$. •

Corollary 4.8 x is a cluster point of the net (x_λ) in X iff there exists a subnet $(x_{\lambda_\mu}) \to x$.
(*This result was stated earlier, without proof, as Theorem 2.11.*)

Proof x is a cluster point of (x_λ) iff x is a cluster point of the associated filter $\mathcal{F}$
iff there exists a filter $\mathcal{G} \supseteq \mathcal{F}$ with $\mathcal{G} \to x$
iff (x_λ) has a subnet converging to x. •

Example 4.9 Think about each of the following parallel statements about nets and filters. Which ones follow "by duality" from the others?

1) If $x_\lambda = a$ for all $\lambda \geq \lambda_0$, then $(x_\lambda) \to a$	$1'$) If $\mathcal{F}$ consists of all sets containing a, then $\mathcal{F} \to a$
2) $(x_\lambda) \to a$ iff every subnet converges to a	$2'$) If $\mathcal{F} \to a$ iff $\mathcal{G} \to a$ for every filter $\mathcal{G} \supseteq \mathcal{F}$
3) If a subnet of (x_λ) has cluster point a, then a is a cluster point of (x_λ)	$3'$) If a is a cluster point of $\mathcal{G}$ and $\mathcal{G} \supseteq \mathcal{F}$, then a is a cluster point of $\mathcal{F}$.

5. Ultrafilters and Universal Nets

Nets and filters are objects that can be defined in any set X. The same is true for ultrafilters and universal nets. No topology is needed unless we want to talk about convergence, cluster points and other ideas that involve "nearness." So many results in this section are purely set-theoretic.

Definition 5.1 An ultrafilter $\mathcal{U}$ in X is called fixed (or trivial) if $\bigcap \mathcal{U} \neq \emptyset$. $\mathcal{U}$ is called free (or nontrivial) if $\bigcap \mathcal{U} = \emptyset$.

For example (*see Example 3.4*)) the ultrafilter $\mathcal{F} = \{A \subseteq X : p \in A\}$ is fixed (trivial).

Theorem 5.2 For an ultrafilter $\mathcal{U}$ in X, then the following are equivalent.

i) for some $p \in X$, $\mathcal{U} = \{A \subseteq X : p \in A\}$

ii) $\bigcap \mathcal{U} = \{p\}$ for some $p \in X$

iii) $\mathcal{U}$ is fixed – that is, $\bigcap \mathcal{U} \neq \emptyset$.

Proof It is clear that i) $\Rightarrow$ ii) $\Rightarrow$ iii)

iii) $\Rightarrow$ i) Suppose i) is false. Then $\{x\} \notin \mathcal{U}$ for every $x \in X$ (*why?*) so by Theorem 3.5, $X - \{x\} \in \mathcal{U}$ for every x. Therefore $\bigcap \mathcal{U} \subseteq \bigcap \{X - \{x\} : x \in X\} = \emptyset$. •

Example 5.3 If $\mathcal{F}$ is a filter and $\bigcap \mathcal{F} \neq \emptyset$, $\mathcal{F}$ might not be an ultrafilter.

Let $\mathcal{F}$ be the filter in $\mathbb{N}$ generated by $\{B_n : n \in \mathbb{N}\}$, where $B_n = \{1\} \cup \{k : k \geq n\}$. Then $\bigcap \mathcal{F} = \bigcap_{n=1}^{\infty} B_n = \{1\}$. Since neither $\mathbb{E} = \{2, 4, 6, ...\} \subseteq \mathbb{N}$ nor $\mathbb{N} - \mathbb{E}$ contains one of the sets B_n, neither $\mathbb{E}$ nor $\mathbb{N} - \mathbb{E}$ is in $\mathcal{F}$. Therefore $\mathcal{F}$ is not an ultrafilter.

$\mathcal{F} \cup \{\mathbb{E}\}$ is a filter base that generates a filter $\mathcal{F}'$ strictly larger than $\mathcal{F}$. Similarly, $\mathcal{F} \cup \{\mathbb{N} - \mathbb{E}\}$ generates a filter $\mathcal{F}''$ strictly larger than $\mathcal{F}$. Then $\mathcal{F}' \neq \mathcal{F}''$ (because $\mathbb{E}$ and $\mathbb{N} - \mathbb{E}$ cannot be in the same filter). So $\mathcal{F}$ can be "enlarged" in at least two different ways.

Theorem 5.4 If $\mathcal{F}$ is a filter in X, then $\mathcal{F} \subseteq \mathcal{U}$ for some ultrafilter $\mathcal{U}$.

Proof Let $\mathcal{P} = \{\mathcal{G} : \mathcal{G}$ is a filter and $\mathcal{G} \supseteq \mathcal{F}\}$, ordered by inclusion. $\mathcal{P}$ is a nonempty poset since $\mathcal{F} \in \mathcal{P}$. Let $\{\mathcal{G}_\alpha : \alpha \in I\}$ be a chain in $\mathcal{P}$. We claim that $\bigcup \mathcal{G}_\alpha \in \mathcal{P}$.

Clearly, $\bigcup \mathcal{G}_\alpha \supseteq \mathcal{F}$; and $\bigcup \mathcal{G}_\alpha$ is a filter:

$\bigcup \mathcal{G}_\alpha \neq \emptyset$ since each $\mathcal{G}_\alpha \neq \emptyset$, and each set in $\bigcup \mathcal{G}_\alpha$ is nonempty.
If $A, B \in \bigcup \mathcal{G}_\alpha$, then both A, B are in a single filter $\mathcal{G}_{\alpha_0}$. Therefore $A \cap B \in \mathcal{G}_{\alpha_0}$, so $A \cap B \in \bigcup \mathcal{G}_\alpha$. In addition, if $C \supseteq A$, then $C \in \mathcal{G}_{\alpha_0}$, so $C \in \bigcup \mathcal{G}_\alpha$.

Therefore the chain $\{\mathcal{G}_\alpha : \alpha \in I\}$ has an upper bound $\bigcup \mathcal{G}_\alpha$ in $\mathcal{P}$. By Zorn's Lemma, $\mathcal{P}$ contains a maximal element $\mathcal{U}$. •

Example 5.5 For each $n \in \mathbb{N}$, there is a fixed (trivial) ultrafilter $\mathcal{U}_n = \{A \subseteq \mathbb{N} : n \in A\}$. It is the ultrafilter for which $\bigcap \mathcal{U}_n = \{n\}$. By Theorem 5.2, the $\mathcal{U}_n$'s are <u>the only</u> fixed ultrafilters in $\mathbb{N}$. There are also nontrivial (free) ultrafilters in $\mathbb{N}$:

Let $B_n = \{k \in \mathbb{N} : k \geq n\}$. The collection $\mathcal{B} = \{B_n : n \in \mathbb{N}\}$ is a filter base and $\mathcal{B}$ generates a filter $\mathcal{F}$. By Theorem 5.4 there is an ultrafilter $\mathcal{U} \supseteq \mathcal{F}$, and $\bigcap \mathcal{U} \subseteq \bigcap_{n=1}^{\infty} B_n = \emptyset$.

$\mathcal{U}$ is <u>not the only</u> ultrafilter containing $\mathcal{F}$. If $\mathbb{E}$ and $\mathbb{O}$ are the sets of even and odd natural numbers, then $\mathcal{B} \cup \{\mathbb{E}\}$ and $\mathcal{B} \cup \{\mathbb{O}\}$ are filter bases that generate filters $\mathcal{F}'$ and $\mathcal{F}''$. Then there are ultrafilters $\mathcal{U}' \supseteq \mathcal{F}'$ and $\mathcal{U}'' \supseteq \mathcal{F}''$. Since $\mathbb{E} \in \mathcal{U}'$, $\mathbb{O} \in \mathcal{U}''$ and $\mathbb{E} \cap \mathbb{O} = \emptyset$, we have $\mathcal{U}' \neq \mathcal{U}''$. Since $\mathcal{B} \subseteq \mathcal{U}'$ and $\mathcal{B} \subseteq \mathcal{U}''$, we know that $\bigcap \mathcal{U}' = \emptyset$ and $\bigcap \mathcal{U}'' = \emptyset$, so $\mathcal{U}'$ and $\mathcal{U}'$ are free ultrafilters.

It is a nice exercise to prove that <u>if there is only one ultrafilter $\mathcal{U}$ containing</u> $\mathcal{F}$, then $\mathcal{F} = \mathcal{U}$: that is, if $\mathcal{F}$ is <u>not</u> an ultrafilter, then $\mathcal{F}$ can always be enlarged to an ultrafilter in more than one way.

If $\mathcal{F}$ is a filter in $\mathbb{N}$, then $\mathcal{F} \in \mathcal{P}(\mathcal{P}(\mathbb{N}))$, so there are <u>at most</u> $|\mathcal{P}(\mathcal{P}(\mathbb{N}))| = 2^{2^{\aleph_0}} = 2^c$ filters in $\mathbb{N}$.

<u>Comment without proof</u>: There are <u>exactly</u> 2^c different filters in $\mathbb{N}$. In fact, there are 2^c different ultrafilters in $\mathbb{N}$, and since $\mathbb{N}$ contains only countably many fixed ultrafilters $\mathcal{U}_n$, there are in fact 2^c free ultrafilters !

Theorem 5.6 An ultrafilter $\mathcal{U}$ in a space X converges to each of its cluster points.

Proof If x is a cluster point of $\mathcal{U}$, then there is a filter $\mathcal{G} \supseteq \mathcal{U}$ such that $\mathcal{G} \to x$. But $\mathcal{G} = \mathcal{U}$ since $\mathcal{U}$ is an ultrafilter. •

Corollary 5.7 An ultrafilter $\mathcal{U}$ in a T_2 space X has at most one cluster point.

We now define the analogue of ultrafilters for nets.

Definition 5.8 A net (x_λ) in X is called universal net (or ultranet) if for every $A \subseteq X$, (x_λ) is either eventually in A or eventually in $X - A$.

For example, a net (x_λ) which is eventually constant, say $x_\lambda = p$ for all $\lambda \geq \lambda_0$, is a universal net. It is referred to as a trivial universal net because its associated filter is the trivial ultrafilter $\{A \subseteq X : p \in A\}$.

Theorem 5.9 In any set X,

1) a net (x_λ) is universal iff its associated filter is an ultrafilter, and
2) a filter $\mathcal{F}$ is an ultrafilter iff its associated net is universal.

Proof 1) (x_λ) is universal iff for every $A \subseteq X$, (x_λ) is eventually in A or $X - A$
iff for every $A \subseteq X$, A or $X - A$ contains a tail of (x_λ)
iff for every $A \subseteq X$, A or $X - A$ is in the associated filter
iff the associated filter is an ultrafilter.

2) We use duality. For a given a filter $\mathcal{F}$, consider its associated net (x_λ). By a), (x_λ) is universal iff its associated filter $\mathcal{F}'$ is an ultrafilter. But $\mathcal{F}' = \mathcal{F}$. •

Corollary 5.10 In any space X,

1) a subnet of a universal net is universal
2) a universal net converges to each of its cluster points
3) every net has a universal subnet.

Proof 1) If (x_λ) is universal, then (x_λ) generates an ultrafilter $\mathcal{U}$. By Theorem 4.6, each subnet (x_{λ_μ}) generates a filter $\mathcal{G} \supseteq \mathcal{U}$. So $\mathcal{G} = \mathcal{U}$. Because the filter associated to (x_{λ_μ}) is an ultrafilter, (x_{λ_μ}) is universal.

2) If x is a cluster point of the universal net (x_λ), then x is a cluster point of the associated ultrafilter $\mathcal{U}$. By Theorem 5.6 $\mathcal{U} \to x$ and therefore $(x_\lambda) \to x$.

3) Let $\mathcal{F}$ be the associated filter for the net (x_λ) and let $\mathcal{U}$ be an ultrafilter containing $\mathcal{F}$. By Theorem 4.6(2), $\mathcal{U}$ is generated by a subnet (x_{λ_μ}) of (x_λ). Since the filter associated with (x_{λ_μ}) is an ultrafilter, (x_{λ_μ}) is a universal net. •

Corollary 5.11 A universal net in a T_2 space has at most one cluster point.

Both nets and filters are sufficient to describe the topology in any space, so we should be able to use them to describe continuous functions.

Theorem 5.12 Suppose X and Y are topological spaces, $f : X \to Y$ and $a \in X$. The following are equivalent:

1) f is continuous at a
2) whenever a net $(x_\lambda) \to a$ in X, then $(f(x_\lambda)) \to f(a)$ in Y
3) whenever a universal net $(x_\lambda) \to a$ in X, then $(f(x_\lambda)) \to f(a)$ in Y
4) whenever a is a cluster point of a net (x_λ) in X, then $f(a)$ is a cluster point of $(f(x_\lambda))$ in Y
5) whenever a filter base $\mathcal{B} \to a$ in X, then the filter base $f[\mathcal{B}] = \{f[B] : B \in \mathcal{B}\} \to f(a)$ in Y
6) whenever an ultrafilter $\mathcal{U} \to a$ in X, then the filter base $f[\mathcal{U}] \to f(a)$ in Y.
7) whenever a is a cluster point of a filter base $\mathcal{B}$ in X, then $f(a)$ is a cluster point of the filter base $f[\mathcal{B}]$ in Y.

Proof 1) $\Rightarrow$ 2) Suppose V is a neighborhood of $f(a)$. Since f is continuous at a, there is a neighborhood U of a such that $f[U] \subseteq V$. If $(x_\lambda) \to a$, then (x_λ) is eventually in U so $(f(x_\lambda))$ is eventually in V. Therefore $(f(x_\lambda)) \to f(a)$.

2) $\Rightarrow$ 3) This is immediate.

3) $\Rightarrow$ 4) If a is a cluster point of (x_λ), then there is a subnet $(x_{\lambda_\mu}) \to a$. Let $(x_{\lambda_{\mu_\nu}})$ be a universal subnet of (x_{λ_μ}). Then $(x_{\lambda_{\mu_\nu}}) \to a$ and by iii), $(f(x_{\lambda_{\mu_\nu}})) \to f(a)$. Then $(f(x_\lambda))$ has a subnet converging to $f(a)$, so $f(a)$ is a cluster point of $(f(x_\lambda))$.

4) $\Rightarrow$ 1) Suppose f is not continuous at a. Then there is a neighborhood V of $f(a)$ such that $f[N] \not\subseteq V$ for every $N \in \mathcal{N}_a$. Let $\Lambda = \mathcal{N}_a$, ordered by reverse inclusion, and define a net $g : \Lambda \to X$ by $g(N) = x_N =$ a point in N for which $f(x_N) \notin V$. Since $(x_N) \to a$, (x_N) has a cluster point at a. But the net $(f(x_N))$ does not have a cluster point at $f(a)$ because $(f(x_N))$ is <u>never</u> in V. Therefore 4) fails.

Knowing that 1)-4) are equivalent, we could show that each of 2)-4) is equivalent to its filter counterpart – e.g., that 2) $\Leftrightarrow$ 5), etc. This involves a little more than simply saying "by duality" because, in each case, the function f also comes into the argument. Instead, for practice, we will show directly that 1) $\Rightarrow$ 5) $\Rightarrow$ 6) $\Rightarrow$ 7) $\Rightarrow$ 1).

It is easy to check that if $\mathcal{B}$ is a filter base in X, then $f[\mathcal{B}] = \{f[B] : B \in \mathcal{B}\}$ is a filter base in Y.

1) $\Rightarrow$ 5) Suppose $\mathcal{B} \to a$ in X and let $f[B] \in f[\mathcal{B}]$. If N is any neighborhood of $f(a)$ in Y, then by continuity $f^{-1}[N]$ is a neighborhood of a in X. Therefore $f^{-1}[N] \supseteq B$ for some $B \in \mathcal{B}$, so $N \supseteq f[f^{-1}[N]] \supseteq f[B]$. Therefore $f(\mathcal{B}) \to f(a)$.

5) $\Rightarrow$ 6) This is immediate.

6) $\Rightarrow$ 7) Suppose a is a cluster point of the filter base $\mathcal{B}$. We can choose an ultrafilter $\mathcal{U} \supseteq \mathcal{B}$ with $\mathcal{U} \to a$. By 6) , $f[\mathcal{U}] \to f(a)$. Therefore the filter $\mathcal{U}'$ generated by $f[\mathcal{U}]$ converges to $f(a)$, so $\mathcal{U}' \supseteq \mathcal{N}_{f(a)}$. Since $f[\mathcal{B}] \subseteq \mathcal{U}'$, each set in $f[\mathcal{B}]$ intersects every set in $\mathcal{N}_{f(a)}$. Therefore $f(a)$ is a cluster point of $f[\mathcal{B}]$.

7) $\Rightarrow$ 1) Suppose f is not continuous at a. Then there is a neighborhood V of $f(a)$ such that $f[N] \not\subseteq V$ for all $N \in \mathcal{N}_a$, that is, $N - f^{-1}[V] \neq \emptyset$ for all $N \in \mathcal{N}_a$. The collection $\mathcal{B} = \{N - f^{-1}[V] : N \in \mathcal{N}_a\}$ is a filter base (*why?*) that has a cluster point at a. However $f[\mathcal{B}]$ does not cluster at $f(a)$ since no set in $f[\mathcal{B}]$ intersects V. •

Corollary 5.13 Let (x_λ) be a net in the product $X = \prod\{X_\alpha : \alpha \in A\}$. Then $(x_\lambda) \to x$ in X iff $(\pi_\alpha(x_\lambda)) \to \pi_\alpha(x)$ in X_α for each $\alpha \in A$.

Proof If $(x_\lambda) \to x \in X$, then by continuity $(\pi_\alpha(x_\lambda)) \to \pi_\alpha(x) \in X_\alpha$ for every $\alpha \in A$.

Conversely, suppose $(\pi_\alpha(x_\lambda)) \to \pi_\alpha(x)$ for every α and let $U = \,< U_{\alpha_1}, U_{\alpha_2}, ..., U_{\alpha_n} >$ be a basic open set containing x in X. For each $i = 1, ..., n$ we have $(\pi_{\alpha_i}(x_\lambda)) \to \pi_{\alpha_i}(x) \in U_{\alpha_i}$. Therefore we can choose a $\lambda_i \in \Lambda$ so that $\pi_{\alpha_i}(x_\lambda) \in U_{\alpha_i}$ when $\lambda \geq \lambda_i$. Pick $\lambda^* \geq \lambda_1, \ldots, \lambda_n$. If $\lambda \geq \lambda^*$, we have $\pi_{\alpha_i}(x_\lambda) \in U_{\alpha_i}$ for each $i = 1, ..., n$. Therefore $x_\lambda \in U$ for $\lambda \geq \lambda^*$ so $(x_\lambda) \to x$. •

6. Compactness Revisited and the Tychonoff Product Theorem

With nets and filters available, we can give a nice characterization of compact spaces in terms of convergence.

Theorem 6.1 For any space X, the following are equivalent:

1) X is compact
2) if $\mathcal{F}$ is a family of closed sets with the finite intersection property, then $\bigcap \mathcal{F} \neq \emptyset$
3) every filter $\mathcal{F}$ has a cluster point
4) every filter can be enlarged to a filter that converges
5) every net has a cluster point
6) every net has a convergent subnet
7) every universal net converges
8) every ultrafilter converges.

Proof We proved earlier that 1) and 2) are equivalent (*see Theorem IV.8.4*).

2) $\Rightarrow$ 3) If $\mathcal{F}$ is a filter in X, then $\mathcal{F}$ has the finite intersection property, so $\{\text{cl}\, F : F \in \mathcal{F}\}$ is a family of <u>closed</u> sets also with the finite intersection property. By ii), $\exists x \in \bigcap\{\text{cl}\, F : F \in \mathcal{F}\}$ so x is a cluster point of $\mathcal{F}$.

3) ⇒ 4) If $\mathcal{F}$ is a filter, then $\mathcal{F}$ has a cluster point x so, by Theorem 4.7, there is a filter $\mathcal{G} \supseteq \mathcal{F}$ such that $\mathcal{G} \to x$.

4) ⇒ 5) If (x_λ) is a net, consider the associated filter $\mathcal{F}$. By 4) there is a filter $\mathcal{G} \supseteq \mathcal{F}$ where $\mathcal{G} \to x \in X$. $\mathcal{G}$ is generated by a subnet (x_{λ_μ}) and by duality, $(x_{\lambda_\mu}) \to x$.

5) ⇒ 6) If (x_λ) has a cluster point x, then by Corollary 4.8, there is a subnet $(x_{\lambda_\mu}) \to x$.

6) ⇒ 7) If (x_λ) is a universal net, then 6) gives that (x_λ) has a subnet that converges to a point x. Then x is a cluster point of (x_λ). Since (x_λ) is universal, $(x_\lambda) \to x$ by Corollary 5.10.

7) ⇒ 8) This is immediate from the duality between universal nets and ultrafilters (Theorems 5.9 and 4.4)

8) ⇒ 1) Suppose X is not compact and let $\mathcal{U} = \{U_\alpha : \alpha \in A\}$ be an open cover with no finite subcover. Then for any $\alpha_1, \alpha_2, ..., \alpha_n \in A$, $X \neq U_{\alpha_1} \cup U_{\alpha_2} \cup ... \cup U_{\alpha_n}$; so, by complements, $\emptyset \neq (X - U_{\alpha_1}) \cap ... \cap (X - U_{\alpha_n})$. Therefore $\mathcal{S} = \{X - U_\alpha : \alpha \in A\}$ is a collection of closed sets with the finite intersection property. The set of all finite intersections of sets from $\mathcal{S}$ is a filter base which generates a filter $\mathcal{F}$, and we can find an ultrafilter $\mathcal{V} \supseteq \mathcal{F}$.

Every point x is in U_α for some α, so $x \notin X - U_\alpha = \text{cl}\,(X - U_\alpha)$. Therefore x is not a cluster point of $\mathcal{V}$. In particular, this implies $\mathcal{V} \not\to x$. Since x was arbitrary, this means $\mathcal{V}$ does not converge. •

Theorem 6.1 gives us a fresh look at the relationship between some of the "compactness-like" properties that we defined in Chapter IV:

X is compact iff every net has a convergent subnet (Theorem 6.1)	X is sequentially compact iff every sequence has a convergent subsequence (Definition IV.8.7)

↘ ↙

X is countably compact iff
every sequence has a cluster point (Theorem IV.8.10) iff
every sequence has a convergent subnet (Corollary 4.8)

It is now easy to prove that every product of compact spaces is compact.

Theorem 6.2 (Tychonoff Product Theorem) Suppose $X = \prod\{X_\alpha : \alpha \in A\} \neq \emptyset$. X is compact iff each X_α is compact.

Proof For each α, $X_\alpha = \pi_\alpha[X]$ so if X is compact, then each X_α is compact.

Conversely, suppose each X_α is compact and let (x_λ) be a universal net in $X = \prod\{X_\alpha : \alpha \in A\}$.
For each α, $(\pi_\alpha(x_\lambda))$ is a universal net in X_α (*Check: if* $A \subseteq X_\alpha$, *then* (x_λ) *is eventually in* $\pi_\alpha^{-1}[A]$ *or* $\pi_\alpha^{-1}[X_\alpha - A]$, *so* $(\pi_\alpha(x_\lambda))$ *eventually in* A *or* $X - A$.) But X_α is compact, so by Theorem 6.1 $(\pi_\alpha(x_\lambda)) \to$ some point $z_\alpha \in X_\alpha$. Let $z = (z_\alpha) \in X$. By Corollary 5.13, $(x_\lambda) \to z$. Since every universal net in X converges, X is compact by Theorem 6.1. •

Remark: A quite different approach to the Tychonoff Product Theorem is to show first that a space X *is compact iff every open cover by* *sub*basic *open sets has a finite subcover. This is called the* *Alexander Subbase Theorem* *and the proof is nontrivial: it involves an argument using Zorn's Lemma or one of its equivalents.*

After that, it is fairly straightforward to show that any cover of $X = \prod\{X_\alpha : \alpha \in A\}$ *by sets of the form* $\pi_\alpha^{-1}[U_\alpha]$ *has a finite subcover. See Exercise E10.*

At this point we restate a result which we stated earlier but without a complete proof (Corollary VII.3.16).

Corollary 6.3 A space X is Tychonoff iff it is homeomorphic to a subspace of a compact Hausdorff space. (*In other words, the Tychonoff spaces are exactly the subspaces of compact Hausdorff spaces.*)

Proof A compact Hausdorff space is T_4, and therefore Tychonoff. Since the Tychonoff property is hereditary, every subspace of a compact T_2 space is Tychonoff.

Conversely, every Tychonoff space X is homeomorphic to a subspace of some cube $[0,1]^m$. This cube is Hausdorff and it is compact by the Tychonoff Product Theorem. •

Remark Suppose X is embedded in some cube $[0,1]^m$. To simplify notation, assume $X \subseteq [0,1]^m$. Then $X \subseteq \text{cl}\, X = K \subseteq [0,1]^m$. K is a compact T_2 space containing X as a dense subspace and K is called a compactification of X. Since every Tychonoff space can be embedded in a cube, we have therefore shown that every Tychonoff space X has a compactification.

Conversely, if K is a compactification of X, then K is Tychonoff and its subspace X is also Tychonoff. Therefore X has a compactification iff X is Tychonoff.

Our proof of the Tychonoff Product Theorem used the Axiom of Choice (AC) in the form of Zorn's Lemma (to get the necessary universal nets or ultrafilters). The following theorem shows that, in fact, the Tychonoff Product Theorem and AC are equivalent. This is perhaps somewhat surprising since AC is a purely set theoretical statement while Tychonoff's Theorem is topological. On the other hand, if "all mathematics can be embedded in set theory" then every mathematical statement is purely set theoretical.

Theorem 6.5 (Kelley, 1950) The Tychonoff Product Theorem implies the Axiom of Choice (so the two are equivalent).

Proof Suppose $\{X_\alpha : \alpha \in A\}$ is a collection of nonempty sets. The Axiom of Choice is equivalent to the statement that $\prod_{\alpha \in A} X_\alpha \neq \emptyset$ (*see Theorem 6.2.2*).

Let $Y_\alpha = X_\alpha \cup \{p\}$ where $p \notin \bigcup_{\alpha \in A} X_\alpha$ and give Y_α the very simple topology $\mathcal{T}_\alpha = \{Y_\alpha, \{p\}, \emptyset\}$. Then Y_α is compact, so $Y = \prod_{\alpha \in A} Y_\alpha$ is compact by the Tychonoff Product Theorem.

$\{p\}$ is open in Y_α, so X_α is closed in Y_α. Therefore $\mathcal{F} = \{\pi_\alpha^{-1}[X_\alpha] : \alpha \in A\}$ is a family of closed sets in Y. We claim that $\mathcal{F}$ has the finite intersection property.

> Suppose $\alpha_1, \alpha_2, ..., \alpha_n \in A$. Since the X_α's are nonempty, there exist points $x_{\alpha_1} \in X_{\alpha_1}, ...,$ $x_{\alpha_n} \in X_{\alpha_n}$.
>
> Define $f : A \to \bigcup_{\alpha \in A} Y_\alpha$ by:
>
> $$\text{for } \alpha \in A,\ f(\alpha) = \begin{cases} x_{\alpha_i} & \text{if } \alpha = \alpha_i \\ p & \text{if } \alpha \neq \alpha_1, \alpha_2, ..., \alpha_n \end{cases}$$
>
> To be more formal $-$ since this is the crucial set-theoretic issue in the argument $-$ we can formally and precisely define f in ZF by:
>
> $$f = \{(\alpha, y) \in A \times \bigcup_{\alpha \in A} Y_\alpha : (\alpha = \alpha_1 \wedge y = x_{\alpha_1}) \vee ... \vee (\alpha = \alpha_n \wedge y = x_{\alpha_n}) \\ \vee\ ((\alpha \neq \alpha_1) \wedge ... \wedge (\alpha \neq \alpha_n)) \wedge y = p)\}$$
>
> Then $f \in \pi_{\alpha_1}^{-1}[X_{\alpha_1}] \cap ... \cap \pi_{\alpha_n}^{-1}[X_{\alpha_n}]$.

Since $\mathcal{F}$ has the finite intersection property and Y is compact, $\bigcap \mathcal{F} \neq \emptyset$. If $g \in \bigcap \mathcal{F}$, then $g \in \prod_{\alpha \in A} X_\alpha$ and therefore $\prod_{\alpha \in A} X_\alpha \neq \emptyset$. ∙

If $\prod X_\alpha \neq \emptyset$ and any X_α's are noncompact, then $\prod X_\alpha$ is noncompact. And we note that if infinitely many X_α's are noncompact, then $\prod X_\alpha$ is "dramatically" noncompact as the following theorem indicates.

Theorem 6.6 Let $X = \prod \{X_\alpha : \alpha \in A\} \neq \emptyset$. If infinitely many of the X_α's are not compact, then every compact closed subset of X is nowhere dense. (*Thus, all closed compact subsets of X are "very skinny" and "far from" being all of X.*)

Proof Suppose B is a compact closed set in X and that B is not nowhere dense. Then there is a point x and indices $\alpha_1, \alpha_2, ..., \alpha_n$ such that $x \in\ < U_{\alpha_1}, U_{\alpha_2}, ... U_{\alpha_n} > \ \subseteq \text{int}(\text{cl}\, B) = \text{int}\, B \subseteq B$. Then $\pi_\alpha[B] = X_\alpha$ for $\alpha \neq \alpha_1, \alpha_2, ..., \alpha_n$, so X_α is compact if $\alpha \neq \alpha_1, \alpha_2, ..., \alpha_n$. ∙

7. Applications of the Tychonoff Product Theorem

We have already used the Tychonoff Theorem in several ways (*see, for example, Corollary 6.3 and the remarks following.*) It's a result that is useful in nearly all parts of analysis and topology, although its full generality is not always necessary. In this section we sketch how it can be used in more "unexpected" settings. The following examples also provide additional insight into the significance of compactness.

The Compactness Theorem for Propositional Calculus

Propositional calculus is a part of mathematical logic that deals with expressions such as $p \wedge q$, $p \vee q$, $p \Rightarrow q$, $\sim p$, $(p \Rightarrow q) \vee r$, etc. Letters such as $p, q, r, \ldots$ are used to represent "propositions" that can have "truth values" T (true) or F (false). These letters are the "alphabet" for propositional calculus. For example, we could think of p as representing the (false) proposition "$2+2=5$" or the (true) proposition "$\forall x \in \mathbb{R}\,(x \geq 0 \Rightarrow \exists y \in \mathbb{R}\ (y^2 = x))$". However, p could not represent an expression like "$x = 5$", because this expression has no truth value: it contains a "free variable" x.

In propositional calculus, propositions $p, q, \ldots$ are thought of as "atoms" – that is, the internal structure of the propositions $p, q, \ldots$ (such as variables and quantifiers) is ignored. Propositional calculus deals with "basic" or "atomic" propositions such as $p, q, \ldots$, with compounds built up from them such as $(p \vee q)$ and $\sim(\sim p \vee q)$, and with the relations between their truth values. We want to allow the possibility of infinitely many propositions, so we will use $A_1, \ldots, A_n, \ldots$ as our alphabet instead of the letters $p, q, \ldots$ that one usually sees in beginning treatments of propositional calculus.

Here is a slightly more formal description of propositional calculus.

Propositional calculus has an alphabet $\mathcal{A} = \{A_1, A_2, \ldots, A_n, \ldots\}$. We will assume $\mathcal{A}$ is countable, although that restriction is not really necessary for anything we do. Propositional calculus also has connective symbols: (,) , $\vee$, and $\sim$.

Inductively we define a collection $\mathcal{W}$ of well-formed formulas (called wffs, for short) that are the "legal expressions" in propositional calculus:

1) for each n, A_n is a wff
2) if ϕ and ψ are wffs, so are $(\phi \vee \psi)$ and $\sim \phi$.

For example, $((A_{13} \vee A_2) \vee \sim A_4)$ and $\sim (A_1 \vee A_4)$ are wffs but the string of symbols ($\vee$ A$_3$ $\vee$ $\vee$) is not a wff. If we like, we can add additional connectives $\wedge$ and $\Rightarrow$ to our propositional calculus, defining them as follows:

Given wffs ϕ and ψ,

$$\phi \wedge \psi = \sim(\sim \phi \vee \sim \psi) \quad \text{and}$$
$$\phi \Rightarrow \psi = \psi \vee \sim \phi.$$

Since $\wedge$ and $\Rightarrow$ can be defined in terms of $\vee$ and $\sim$, it is simpler and involves no loss to develop the theory using just the smaller set of connectives.

A truth assignment is a function s that assigns a truth value to each A_n. More formally, $s : \mathcal{A} \to \{T, F\}$, so $s \in \{T, F\}^{\aleph_0} = C$. We give C the product topology. By the Tychonoff Theorem, C is compact (in fact, more directly, by Theorem VI.2.19 which shows that C is homeomorphic to the Cantor set).

A truth assignment s can be used to assign a unique truth value to every wff in $\mathcal{W}$, that is, we can extend s to a function $\overline{s} : \mathcal{W} \to \{T, F\}$ as follows:

For any wff σ we define:

$$\text{if } \sigma = A_n, \text{ then } \overline{s}\,(\sigma) = s(\sigma)$$

$$\text{if } \sigma = \phi \vee \psi, \text{ then } \overline{s}\,(\sigma) = \begin{cases} T & \text{if } \overline{s}\,(\phi) = T \text{ or } \overline{s}\,(\psi) = T \\ F & \text{otherwise} \end{cases}$$

$$\text{if } \sigma = \ \sim \phi, \text{ then } \overline{s}\,(\sigma) = \begin{cases} T & \text{if } \overline{s}\,(\phi) = F \\ F & \text{if } \overline{s}\,(\phi) = T \end{cases}$$

We say that a truth assignment s satisfies a wff σ if $\overline{s}\,(\sigma) = T$. A set of wffs Σ is called satisfiable if there exists a truth assignment $s \in \{T, F\}^{\aleph_0}$ such that s satisfies σ for every $\sigma \in \Sigma$.

For example,

$\Sigma = \{A_1, A_1 \vee A_2, A_2\}$ is satisfiable (*We can use any s for which* $s(A_1) = s(A_2) = T$)

$\Sigma = \{\ A_1,\ \sim (A_1 \vee\ \sim A_2)\}$ is not satisfiable (*If $s(A_1) = T$, then* $\overline{s}\,(\ \sim (A_1 \vee\ \sim A_2)) = F$)

Theorem 7.1 (Compactness Theorem for Propositional Calculus) Let Σ be a set of wffs. If every finite subfamily of Σ is satisfiable, then Σ is satisfiable.

Proof Let $C = \{T, F\}^{\aleph_0}$ and suppose that every finite subset of Σ is satisfiable. Then for each $\sigma \in \Sigma$, $A_\sigma = \{s \in C : s \text{ satisfies } \sigma\} \neq \emptyset$. We claim that A_σ is closed in C:

Suppose $s \notin A_\sigma$. We need to produce an open set U containing s for which $U \cap A_\sigma = \emptyset$.

For a sufficiently large n, the list $A_1, \ldots, A_n$ will contain all the letters that occur in σ. Let $U = \{t \in C\colon\ t_i = s_i, i = 1, \ldots, n\} = \{s_1\} \times \ldots \times \{s_n\} \times \{T, F\}^{\aleph_0}$. In other words, U is the set of truth assignments that agree with s for all the letters $A_1, \ldots, A_n$ that may occur in σ. Since s fails to satisfy σ, each $t \in U$ also fails to satisfy σ, so $\sigma \in U \subseteq C - A_\sigma$.

The A_σ's have the finite intersection property – in fact, this is precisely equivalent to saying that every finite subset of Σ is satisfiable. Since C is compact, $\bigcap\{A_\sigma : \sigma \in \Sigma\} \neq \emptyset$, i.e., Σ is satisfiable. ∙

If we assume the Alexander Subbase Theorem (*see the remarks following Theorem 6.2, as well as Exercise E10*), then we can also prove that the Compactness Theorem 7.1 is equivalent to the statement that $C = \{T, F\}^{\aleph_0}$ is compact.

> Suppose the Compactness Theorem 7.1 is true. To show that C is compact it is sufficient, by the Alexander Subbase Theorem, to show that every cover of C by subbasic open sets has a finite subcover. Each subbasic set has the form $\pi_n^{-1}[\{T\}]$ or $\pi_n^{-1}[\{F\}]$.
>
> Taking complements, we see that it is sufficient to show that:
>
> > if $\mathcal{F}$ is any family of closed sets with the finite intersection property and each set in $\mathcal{F}$ has form $C - \pi_n^{-1}[\{T\}] = \pi_n^{-1}[\{F\}]$ or $C - \pi_n^{-1}[\{F\}] = \pi_n^{-1}[\{T\}]$, then $\bigcap \mathcal{F} \neq \emptyset$.
>
> Let $\mathcal{F} = \{F_\alpha : \alpha \in A\}$ be any such family. For each $\alpha \in A$, define a wff
>
> $$\sigma_\alpha = \begin{cases} A_n & \text{if } F_\alpha = \pi_n^{-1}[\{T\}] \\ \sim A_n & \text{if } F_\alpha = \pi_n^{-1}[\{F\}] \end{cases}$$
>
> Clearly, σ_α is satisfied precisely by the truth assignments in F_α.
>
> Let $\Sigma = \{\sigma_\alpha : \alpha \in A\}$. Since the F_α's have the finite intersection property, any finite subset of Σ is satisfiable. By the Compactness Theorem, Σ is satisfiable, so $\bigcap \mathcal{F} \neq \emptyset$. •

Note: If the propositional calculus is allowed to have an uncountable alphabet of cardinality m, then the compactness theorem is equivalent to the statement that $\{T, F\}^m$ is compact; the proof requires only minor notational changes.

A "map-coloring" theorem

Imagine that M is a (geographical) map containing infinitely many countries $C_1, C_2, ..., C_n, ...$.
A valid coloring c of M with 4 colors (red, white, blue, and green, say) is a function

$$c : \{C_1, C_2, ..., C_n, ...\} \to \{R, W, B, G\}$$

such that no two adjacent countries are assigned the same color.

Intuitively, if M doesn't have a valid covering, it must be because some "finite piece" of the map M has a configuration of countries for which a valid coloring can't be done. That is the content of the following theorem.

Theorem 7.2 Suppose M is a (geographical) map with infinitely countries $C_1, ..., C_n, ...$. If every finite submap of M has a valid coloring, then M has a valid coloring. (*Any reasonable definition of "adjacent" and "submap" will work in the proof.*)

Proof Consider set of all colorings $C = \{R, W, B, G\}^{\aleph_0}$ with the product topology. For each finite submap F, let $V_F = \{c \in C : c \text{ is a valid coloring of } F\} \neq \emptyset$. We claim that V_F is closed in C.

Suppose $c \notin V_F$. We need to produce an open set U with $c \in U \subseteq C - V_F$. If n is large enough, the list $C_1, ...C_n$ will include all the countries in the submap F. Then the open set

$$U = \{c(C_1)\} \times ... \times \{c(C_n)\} \times \{R, W, B, G\}^{\aleph_0}$$

works – any coloring in U is invalid because it colors the countries in F the same way c does.

If F_1 and F_2 are finite submaps of M, then so is $F_1 \cup F_2$. Since $F_1 \cup F_2$ has a valid coloring by hypothesis, $V_{F_1 \cup F_2} \subseteq V_{F_1} \cap V_{F_2} \neq \emptyset$. Therefore $\mathcal{F} = \{V_F : F \text{ a finite submap of } M\}$ has the finite intersection property. Since C is compact, $\bigcap \mathcal{F} \neq \emptyset$, and any $c \in \bigcap \mathcal{F}$ is a valid covering of the whole map M. •

(*It is clear that a nearly identical proof would work for any finite number of colors and for maps with uncountably many countries.*)

Exercises

E1. Let X be a topological space and $a \in X$. For each $N \in \mathcal{N}_a$, pick a point $x_N \in N$. If we order $\mathcal{N}_a$ by reverse inclusion, then (x_N) is a net in X. Prove that $(x_N) \to a$.

E2. a) Let $(C, \leq)$ be an uncountable <u>chain</u> in which each element has only countably many predecessors. Suppose $f : C \to \mathbb{R}$ and that $f(\lambda) > 0$ for each $\lambda \in C$. Show that the net f does not converge to 0 in $\mathbb{R}$.

b) Give an example to show that part a) <u>may</u> be false if $(C, \leq)$ is an uncountable <u>directed set</u> in which each element has only countably many predecessors.

b) Is it possible to have a net $f : [0, \omega_2) \to \mathbb{R}$ and that $f(\lambda) > 0$ for each λ? Is it possible that f converges to 0 in $\mathbb{R}$? (*Recall that ω_α denotes the first ordinal with $\aleph_\alpha$ predecessors*.)

E3. Suppose X is a compact Hausdorff space and that $(\Lambda, \leq)$ is a directed set. For each $\lambda \in \Lambda$, let A_λ be a nonempty closed subset in X such that $A_{\lambda_2} \subseteq A_{\lambda_1}$ iff $\lambda_1 \leq \lambda_2$. Prove that $\bigcap\{A_\lambda : \lambda \in \Lambda\} \neq \emptyset$.

E4. a) Let $f : \Lambda \to X$ be a net in a space X and write $f(\lambda) = x_\lambda$. For each $\alpha \in \Lambda$, let $T_\alpha = \{x_\lambda : \lambda \geq \alpha\} =$ "the α^{th} tail of the net." Show that a point $x \in X$ is a cluster point of (x_λ) iff $x \in \bigcap\{\text{cl}\, T_\alpha : \alpha \in \Lambda\}$.

b) Suppose x is a cluster point of the net (x_λ) a product $\prod\{X_\alpha : \alpha \in A\}$. Show that for each α, $\pi_\alpha(x) \in X_\alpha$ is a cluster point of the net $(\pi_\alpha(x_\lambda))$.

c) Give an example to show that the converse to part b) is false.

d) Let (X, d) be a metric space and $f : [0, \omega_1) \to X$ a function given by $f(\alpha) = x_\alpha$. Show that the net (x_α) converges iff (x_α) is eventually constant.

E5. a) Suppose X is infinite set with the cofinite topology. Let $\mathcal{F}$ be the filter generated by the filter base consisting of all cofinite sets. To what points does $\mathcal{F}$ converge?

b) Translate the work in part a) into statements about nets.

E6. Show that if a filter $\mathcal{F}$ is contained in a unique ultrafilter $\mathcal{U}$, then $\mathcal{F} = \mathcal{U}$.
(*Thus, if $\mathcal{F}$ is not an ultrafilter, it can enlarged to an ultrafilter in more than one way.*)

E7. a) State and prove a theorem of the form: Suppose x is a point in a space X. Then $\mathcal{N}_x$ is an ultrafilter $\Leftrightarrow$ ....

b) Prove or disprove: Suppose $X \neq \emptyset$ and that $\mathcal{F}$ is a <u>maximal</u> family of subsets of X with the finite intersection property. Then $\mathcal{F}$ is an ultrafilter.

E8. a) Let $\mathcal{F}$ be a filter in a set X. Prove that $\mathcal{F}$ is the intersection of all ultrafilters containing $\mathcal{F}$. (*Note that this implies the result in E6.*)

b) Let $\mathcal{U}$ be an ultrafilter in X and suppose that $A_1 \cup ... \cup A_n \in \mathcal{U}$. Prove that at least one A_i must be in $\mathcal{U}$. (*This is the filter analogue for a fact in ring theory: in a commutative ring with a unit, every maximal ideal is a prime ideal.*)

c) Give an example to show that part b) is not true for infinite unions.

d) "By duality," there is a result similar to b) about universal nets. State the result and prove it directly.

E9. Let $\mathcal{U}$ be a free ultrafilter in $\mathbb{N}$ and let $\Sigma = \mathbb{N} \cup \{\sigma\}$, where $\sigma \notin \mathbb{N}$. Define a topology $\mathcal{T}$ on Σ by $\mathcal{T} = \{O : O \subseteq \mathbb{N} \text{ or } O = U \cup \{\sigma\} \text{ where } U \in \mathcal{U}\}$

a) Prove that Σ is T_4 and that $\mathbb{N}$ is dense in Σ.

b) Prove that a free ultrafilter $\mathcal{U}$ on $\mathbb{N}$ cannot have a countable base.
Hint: Since $\mathcal{U}$ is free, each set in $\mathcal{U}$ must be infinite. Why?

c) Prove that no sequence in $\mathbb{N}$ can converge to σ (and therefore there can be no countable neighborhood base at σ in Σ)
(*Hint: Work you did in b) might help.)*

Thus, Σ is another example of a countable space where only one point is not isolated and which is not first countable, See the space L in Example III.9.8.

d) How would the space Σ be different if $\mathcal{U}$ were a fixed ultrafilter?

Note: if $\mathcal{U}'$ is a free ultrafilter on $\mathbb{N}$ and $\mathcal{U}' \neq \mathcal{U}$, then the corresponding spaces Σ' and Σ may not be homeomorphic: the neighborhood systems of σ may look quite different. In this sense, free ultrafilters in $\mathbb{N}$ do not all "look alike."

E10. Suppose $(X, \mathcal{T})$ has a some property P. $\mathcal{T}$ is called a maximal-P topology (or minimal-P topology) if any larger (smaller) topology on X fails to have property P. Prove that if $(X, \mathcal{T})$ is a compact Hausdorff space, then $\mathcal{T}$ is maximal-compact and minimal-Hausdorff. (*Compare Exercise IV E23.*)

In one sense, this "justifies" the choice of the product topology over the box topology: for a product of compact Hausdorff spaces, a larger topology would not be compact and a smaller one would not be Hausdorff. The product topology is "just right" to ensure that the property "compact Hausdorff" is productive.

E11. Show that the map coloring Theorem 7.2 is equivalent to the statement that $\{R, W, B, G\}^{\aleph_0}$ is compact.

E12. A family $\mathcal{B}$ of subsets of X is called inadequate if it does not cover X, and $\mathcal{B}$ is called finitely inadequate if no finite subfamily covers X.

a) Use Zorn's Lemma to prove that any finitely inadequate family $\mathcal{B}$ is contained in a maximal finitely inadequate family.

b) Prove (by contradiction) that a maximal finitely inadequate family $\mathcal{B}$ has the following property: if $C_1, ..., C_n$ are subsets of X and $C_1 \cap ... \cap C_n \in \mathcal{B}$, then at least one set $C_i \in \mathcal{B}$.

c) The following are equivalent (Alexander's Subbase Theorem)

i) X has a subbase $\mathcal{S}$ such that each cover of X by members of $\mathcal{S}$ has a finite subcover

ii) X has a subbase $\mathcal{S}$ such that each finitely inadequate collection from $\mathcal{S}$ is inadequate

iii) every finitely inadequate family of open sets in X is inadequate

iv) X is compact

d) Use c) to prove the Tychonoff Product Theorem.

E13. This exercise gives still another proof of the Tychonoff Product Theorem. Suppose X_α is compact for all $\alpha \in A$. We want to prove that $\prod X_\alpha$ is compact. We proceed assuming X is not compact.

a) Show that there is a maximal open cover $\mathcal{U}$ of X having no finite subcover.

b) Show that if O is open in X and $O \notin \mathcal{U}$, then there are sets $U_1, ..., U_n \in \mathcal{U}$ such that $\{U_1, ..., U_n, O\}$ covers X.

c) Show that for each α, $\{V_\alpha \subseteq X_\alpha : V_\alpha$ is open and $< V_\alpha > \in \mathcal{U}\}$ cannot cover X_α. Conclude that for each α we can choose $x_\alpha \in X_\alpha$ so that $x_\alpha \notin$ any open set V_α for which $< V_\alpha > \in \mathcal{U}$.

d) Let $x = (x_\alpha) \in X$ and suppose $x \in U \in \mathcal{U}$. Pick open sets $V_{\alpha_i} \subseteq X_{\alpha_i}$ so that $x \in < V_{\alpha_1}, ..., V_{\alpha_k} > \subseteq U$. Explain why each $< V_{\alpha_i} > \notin \mathcal{U}$.

e) Show that for each $i = 1, ..., k$, there is a finite family $\mathcal{U}_i \subseteq \mathcal{U}$ such that $\mathcal{U}_i \cup < V_{\alpha_i} >$ covers X.

f) Show that $\bigcup_{i=1}^k \mathcal{U}_i \cup \{ < V_{_1}, ..., V_{\alpha_k} > \}$ covers X, and then arrive at the contradiction that $\bigcup_{i=1}^k \mathcal{U} \cup \{U\}$ covers X.

Chapter IX Review

Explain why each statement is true, or provide a counterexample.

1. Order $C(X)$ by $f \leq g$ iff $f(x) \leq g(x)$ (in $\mathbb{R}$) for every $x \in X$. Then $(C(X), \leq)$ is a directed set.

2. x is a limit point of A in the space X iff there exists a filter $\mathcal{F}$ such that $A - \{x\} \in \mathcal{F}$ and $\mathcal{F} \to x$.

3. A set U is open in X iff U belongs to every filter $\mathcal{F}$ which converges to a point of U.

4. In a space X, let $\Phi_x = \{\mathcal{F} : \mathcal{F}$ is a filter converging to $x\}$. Then $\bigcap\{\mathcal{F} : \mathcal{F} \in \Phi_x\} = \mathcal{N}_x$.

5. Suppose $\mathcal{A}$ is a family of subsets of a space X such that if $A, B \in \mathcal{A}$ then $A \cap B \supseteq C$ for some $C \in \mathcal{A}$. Suppose (x_λ) is a net which is frequently in each set of $\mathcal{A}$. Then (x_λ) has a subnet which is eventually in each set in $\mathcal{A}$.

6. If a net $(x_\lambda) \to x$ in X and $|X| = \aleph_0$, then (x_λ) has a subsequence (i.e., a subnet whose directed set is $\mathbb{N}$) which converges to x.

7. If X is a nonempty finite set and $|X| = n$, then there are exactly $2^n - 1$ different filters and exactly n different ultrafilters on X.

8. A universal <u>sequence</u> must be eventually constant.

9. Suppose X is infinite. The collection $\mathcal{S} = \{A \subseteq X : |X - A| < |X|\}$ is an ultrafilter.

10. In $\mathbb{R}$, a filter $\mathcal{F} \to x$ iff $\forall\, \epsilon > 0\ \exists F \in \mathcal{F}$ such that $x \in F$ and $\operatorname{diam}(F) < \epsilon$.

11. If X is compact, then every net in X has a convergent subsequence. (*Note: a "subsequence of a net" is a subnet whose directed set is* $\mathbb{N}$.)

12. If $x \neq y$ in a space X and if $\mathcal{N}_x \cup \mathcal{N}_y$ generates a filter, then X is not T_1.

13. If $\mathcal{F}$ is a filter in $[0, \omega_1]$, then there must be a filter $\mathcal{G} \supseteq \mathcal{F}$ such that $\mathcal{G}$ has a cluster point.

14. Call a set in $[0, 1]$ an endset if it has the form $[0, \epsilon)$ or $(1 - \epsilon, 1]$ for some $\epsilon > 0$. To show $[0, 1]$ is compact, it is sufficient to show that any cover by endsets has a finite subcover.

15. Every separable metric space (X, d) has an equivalent totally bounded metric.

16. Suppose Λ is the collection of finite subsets of $[0, 1]$, directed by $\subseteq$, and $f(F) = 1$ for all $F \in \Lambda$. The net f converges to 1 in $[0, 1]$.

Chapter X
Compactifications

1. Basic Definitions and Examples

Definition 1.1 Suppose $h : X \to Y$ is a homeomorphism of X into Y, where Y is a compact T_2 space. If $h[X]$ is dense in Y, then the pair (Y, h) is called a compactification of X.

By definition, only Hausdorff spaces X can (possibly) have a compactification.

If we are just working with a single compactification Y of X, then we can usually just assume that $X \subseteq Y$ and that h is the identity map $-$ so that the compactification is just a compact Hausdorff space that contains X as a dense subspace. In fact, if $X \supseteq Y$, we will always assume that h is the identity map unless something else is stated. We made similar assumptions in discussing of the completion of a metric space (X, d) in Chapter IV.

However, we will sometimes want to compare different compactifications of X (in a sense to be discussed later) and then we may need to know how X is embedded in Y. We will see that different dense embeddings h of X into the same space Y can produce "nonequivalent" compactifications. Therefore, strictly speaking, a "compactification of X" is the pair (Y, h).

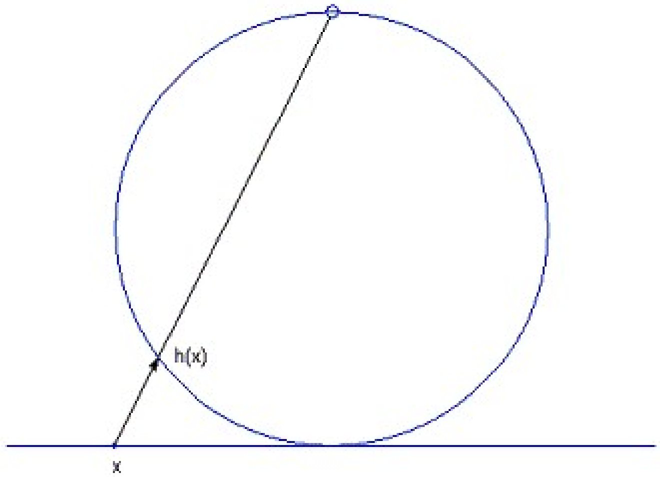
If, in Definition 1.1, X is already a compact Hausdorff space, then $h[X]$ is closed and dense in Y and therefore $h[X] = Y$. Therefore, topologically, the only possible compactification of X is X itself.

The next theorem restates exactly which spaces have compactifications.

Theorem 1.2 A space X has a compactification iff it is a Tychonoff space.

Proof See the remarks following Corollary IX.6.3. $\bullet$

Example 1.3 The circle S^1 can be viewed as a compactification of the real line, $\mathbb{R}$. Let h be the "inverse projection" pictured below: here $h[\mathbb{R}] = S^1 - \{\text{North Pole}\}$. We can think of $h[\mathbb{R}]$ as a "bent" topological copy of $\mathbb{R}$, and the compactification is created by "tying together" the two ends of $\mathbb{R}$ by adding one new "point at infinity" (the North Pole).



Since $|S^1 - h[\mathbb{R}]| = 1$, (S^1, h) is called a one-point compactification of $\mathbb{R}$. (*We will see in Example 4.2 that we can call S^1 the one-point compactification of $\mathbb{R}$.*)

Example 1.4

1) $[-1,1]$ is a compact Hausdorff space containing $(-1,1)$ are a dense subspace, so $[-1,1]$ is a "two-point" compactification of $(-1,1)$ (with embedding $h = i$).

2) If $h : \mathbb{R} \to (-1,1)$ is a homeomorphism, then $i \circ h : \mathbb{R} \to [-1,1]$ gives a "two-point" compactification of $\mathbb{R}$. *It is true (but not so easy to prove) that there is no n-point compactification of $\mathbb{R}$ for $2 < n < \omega_0$.*

Example 1.5 Suppose $Y = X \cup \{p\}$ is a one-point compactification of X. If O is an open set containing p in Y, then $K = Y - O \subseteq X$ and K is compact. Therefore the open sets containing p are the complements of compact subsets of X. (*Look at open neighborhoods of the North Pole p in the one-point compactification S^1 of $\mathbb{R}$; a base for the open neighborhoods of p consists of complements of the closed (compact!) arcs that do not contain the North Pole.*)

Suppose $x \in U$, where U is open in X. Because Y is Hausdorff, we can find disjoint open sets V and W in Y with $x \in V$ and $p \in W$. Since $p \notin V$, we have that $V \subseteq X$ and therefore $V \cap X = V$ is also open in X. Since $x \in U \cap V$, we can use the regularity of X to choose an open set G in X for which $x \in G \subseteq \mathrm{cl}_X G \subseteq U \cap V \subseteq U$. But $G \subseteq V \subseteq Y - W$ (a closed set in Y), so $\mathrm{cl}_Y G \subseteq Y - W \subseteq X$. Therefore $\mathrm{cl}_X G = X \cap \mathrm{cl}_Y G = \mathrm{cl}_Y G$, so $\mathrm{cl}_X G$ is also closed in Y. So $\mathrm{cl}_X G$ is a compact neighborhood of x inside U. This shows that each point $x \in X$ has a neighborhood base in X consisting of compact neighborhoods.

The property in the last sentence is important enough to deserve a name: such spaces are called locally compact.

2. Local Compactness

Definition 2.1 A Hausdorff space X is called locally compact if each point $x \in X$ has a neighborhood base consisting of compact neighborhoods .

Example 2.2

1) A discrete space is locally compact.

2) $\mathbb{R}^n$ is locally compact: at each point x, the collection of closed balls centered at x is a base of compact neighborhoods. On the other hand, neither $\mathbb{Q}$ nor $\mathbb{P}$ is locally compact. (*Why?*)

3) If X is a compact Hausdorff space, then X is regular so there is a base of closed neighborhoods at each point – and each of these neighborhoods is compact. Therefore X is locally compact.

4) Each ordinal space $[0, \alpha)$ is locally compact. The space $[0, \alpha]$ is a (one-point) compactification of $[0, \alpha)$ iff α is a limit ordinal.

5) Example 1.5 shows that if a space X has a one-point compactification, it must be locally compact (and, of course, noncompact and Hausdorff). Therefore neither $\mathbb{Q}$ nor $\mathbb{P}$ has a one-point compactification. The following theorem characterizes the spaces with one-point compactifications.

Theorem 2.3 A space X has a one-point compactification iff X is a noncompact, locally compact Hausdorff space. (*The one-point compactification of X for which the embedding h is the identity is denoted X^*.*)

Proof Because of Example 1.5, we only need to show that a noncompact, locally compact Hausdorff space X has a one-point compactification. Choose a point $p \notin X$ and let $X^* = X \cup \{p\}$. Put a topology on X^* by letting each point $x \in X$ have its original neighborhood base of compact neighborhoods, and by defining basic neighborhoods of p be the complements of compact subsets of X :

$$\mathcal{B}_p = \{N \subseteq X^* : p \in N \text{ and } X^* - N \text{ is compact}\}.$$

(*Verify that the conditions of the Neighborhood Base Theorem III.5.2 are satisfied.*)

If $\mathcal{U}$ is an open cover of X^* and $p \in U \in \mathcal{U}$, then there exists an $N \in \mathcal{B}_p$ with $p \in N \subseteq U$. Since $X^* - N$ is compact, we can choose $U_1, ..., U_n \in \mathcal{U}$ covering $X^* - N$. Then $\{U, U_1, ..., U_n\}$ is a finite subcover of X^* from $\mathcal{U}$. Therefore X^* is compact.

X^* is Hausdorff. If $a \neq b \in X$, then a and b can be separated by disjoint open sets in X and these sets are still open in X^*. Furthermore, if K is a compact neighborhood of a in X, then K and $(X^* - K)$ are disjoint neighborhoods of a and p in X^*.

Finally, notice that $\{p\}$ is not open in X^* – or else $\{p\} \in B_p$ and then $X^* - \{p\} = X$ would be compact. Therefore every open set containing p intersects X, so X is dense in X^*.

Therefore X^* is a one-point compactification of X. •

What happens if the construction for X^* in the preceding proof is carried out starting with a space X which is already compact? What happens if X is not locally compact? What happens if X is not Hausdorff?

Corollary 2.4 A locally compact Hausdorff space X is Tychonoff.

Proof X is either compact or X has a one-point compactification X^*. Either way, X is a subspace of a compact T_2 space which (by Theorem VII.5.9) is Tychonoff. Therefore X is Tychonoff. •

The following theorem about locally compact spaces is often useful.

Theorem 2.5 Suppose $A \subseteq X$, where X is Hausdorff.

a) If X is locally compact and $A = F \cap G$ where F is closed and A is open in X, then A is locally compact. In particular, an open (or, a closed) subset of a locally compact space X is locally compact.

b) If A is a locally compact and X is Hausdorff, then A is open in $\text{cl}_X A$.

c) If A is a locally compact subspace of a Hausdorff space X, then $A = F \cap G$ where F is closed and A is open in X.

Proof a) It is easy to check that if F is closed and G is open in a locally compact space X, then F and G are locally compact. It then follows easily that $F \cap G$ is also locally compact. (*Note: Part a) does not require that X be Hausdorff.*)

b) Let $a \in A$ and let K be a compact neighborhood of a in A. Then $a \in \text{int}_A K = U$. Since A is Hausdorff, K is closed and therefore $U \subseteq \text{cl}_A U \subseteq K$, so $\text{cl}_A U$ is compact.

Because U is open in A, there is an open set V in X with $A \cap V = U$ and we have:

$$\text{cl}_X(A \cap V) \cap A = (\text{cl}_X U) \cap A = \text{cl}_A U \subseteq A$$

so $(\text{cl}_X(A \cap V)) \cap A$ is compact and therefore closed in X (since X is Hausdorff).

Since $A \cap V \subseteq (\text{cl}_X(A \cap V)) \cap A$, we have $\text{cl}_X(A \cap V) \subseteq (\text{cl}_X U) \cap A = \text{cl}_A U \subseteq A$.

Moreover, since V is open, then $V \cap \text{cl}_X A \subseteq \text{cl}_X(V \cap A)$ (*this is true in <u>any</u> space X: why?*).

So $W = V \cap \text{cl}_X A \subseteq \text{cl}_X(A \cap V) \subseteq (\text{cl}_X U) \cap A = \text{cl}_A U \subseteq A$.

Then $a \in W \subseteq A$ and W is open in $\text{cl}_X A$ so $a \in \text{int}\,\text{cl}_X A$. Therefore A is open in $\text{cl}_X A$.

c) Since A is locally compact, part b) gives that A is open in $\text{cl}_X A$, so $A = \text{cl}_X A \cap G$ for some open set G in X. Let $F = \text{cl}_X A$. ∙

Corollary 2.6 A dense locally compact subspace of a Hausdorff space X is open in X.

Proof This follows immediately from part b) of the theorem ∙

Corollary 2.7 If X is a locally compact, noncompact Hausdorff space, then X is open in <u>any</u> compactification Y that contains X.

Proof This follows immediately from Corollary 2.6. ∙

Corollary 2.8 A locally compact metric space (X, d) is completely metrizable.

Proof Let $(\widetilde{X}, \widetilde{d}\,)$ be the completion of (X, d). X is locally compact and dense in $\widetilde{X}$ so X is open in $\widetilde{X}$. Therefore X is a G_δ-set in $\widetilde{X}$ so it follows from Theorem IV.7.5 that X is completely metrizable. ∙

Theorem 2.9 Suppose $X = \prod_{\alpha \in A} X_\alpha \neq \emptyset$ is Hausdorff. Then X is locally compact iff

i) each X_α is locally compact
ii) X_α is compact for all but at most finitely many $\alpha \in A$.

Proof Assume X is locally compact. Suppose U_α is open in X_α and $x_\alpha \in U_\alpha$. Pick a point $z \in \pi_\alpha^{-1}[U_\alpha]$ with $z_\alpha = x_\alpha.$ Then z has a compact neighborhood K in X for which $z \in K \subseteq \pi_\alpha^{-1}[U_\alpha]$. Since π_α is an open continuous map, $\pi_\alpha[K]$ is a compact neighborhood of x_α with $x_\alpha \in \pi_\alpha[K] \subseteq U_\alpha$. Therefore X_α is locally compact, so i) is true.

To prove ii), pick a point $x \in X$ and let K be a compact neighborhood of x. Then $x \in U = \, < U_{\alpha_1}, ..., U_{\alpha_n} > \, \subseteq K$ for some basic open set U. If $\alpha \neq \alpha_1, ..., \alpha_n$, we have $\pi_\alpha[K] \supseteq \pi_\alpha[U] = X_\alpha$. Therefore X_α is compact if $\alpha \neq \alpha_1, ..., \alpha_n$.

Conversely, assume i) and ii) hold. If $x \in U \subseteq X$, where U is open, then we can choose a basic open set $V = \, < V_{\alpha_1}, ..., V_{\alpha_n} >$ so that $x \in V \subseteq U$. Without loss of generality, we can assume that X_α is compact for $\alpha \neq \alpha_1, ..., \alpha_n$ (*why?*). For each i we can choose a compact neighborhood K_{α_i} of x_{α_i} so that $x_{\alpha_i} \in K_{\alpha_i} \subseteq V_{\alpha_i} \subseteq X_{\alpha_i}$. Then $< K_{\alpha_1}, ..., K_{\alpha_n} >$ $= K_{\alpha_1} \times ... \times K_{\alpha_n} \times \prod X_{\alpha \neq \alpha_1, ...\alpha_n} X_\alpha$ is a compact neighborhood of x and $x \in \, < K_{\alpha_1}, ..., K_{\alpha_n} > \, \subseteq \, < V_{\alpha_1}, ..., V_{\alpha_n} > \, \subseteq U$. So X is locally compact. ∙

3. The Size of Compactifications

Suppose X is a Tychonoff space, that $X \subseteq Y$, and that Y is a compactification of X. How large can $|Y - X|$ be? In all the specific example so far, we have had $|Y - X| = 1$ or $|Y - X| = 2$.

Example 3.1 This example illustrates a compactification of a discrete space created by adding c points.

Let $I_0 = \{(x, 0) : x \in [0, 1]\}$ and $I_1 = \{(x, 1) : x \in [0, 1]\}$, two disjoint "copies" of $[0, 1]$. Let Define a topology on $Y = I_0 \cup I_1$ by using the following neighborhood bases:

i) points in I_1 are isolated: for $p \in I_1$, a neighborhood base at p is $\mathcal{B}_p = \{\{p\}\}$

ii) if $p = (x, 0) \in I_0$: a basic neighborhood of p is any set of form $V \cup \{(z, 1) : (z, 0) \in V, z \neq x\}$, where V is an open neighborhood of p in $[0, 1]$

(*Check that the conditions in the Neighborhood Base Theorem III.5.2 are satisfied.*)

Y is called the "double" of the space $[0, 1] = I_0$.

Clearly, Y is Hausdorff, and we claim that Y is compact. Let $\mathcal{U}$ be a covering of Y by basic open sets. It is sufficient to check that $\mathcal{U}$ has a finite subcover.

Let $\mathcal{W} = \{W \in \mathcal{U} : W \cap I_0 \neq \emptyset\}$. $\mathcal{W}$ covers I_0 and each $W \in \mathcal{W}$ has form $V \cup \{(z, 1) : (z, 0) \in V, z \neq x\}$, where V is open in I_0. The "V-parts" of the sets $W \in \mathcal{W}$ cover the compact space I_0, so finitely many $W_1, ..., W_n \in \mathcal{W}$ cover I_0. These sets also cover I_1, <u>except</u> for possibly finitely many points $p_1, ..., p_k \in I_1$. For each such point p_i choose a set $U_i \in \mathcal{U}$ containing p_i . Then $\{W_1, ..., W_n, U_1, ..., U_k\}$ is a finite subcover from $\mathcal{U}$..

Every neighborhood of a point in I_0 intersects I_1, so $\mathrm{cl}\, I_1 = Y$. Therefore Y is a compactification of the discrete space I_1 and $|Y - I_1| = c$.

Since I_1 is locally compact, I_1 also has another quite different compactification I_1^* for which $|I_1^* - I_1| = 1$. In fact, it is true (depending on X) that can be many different compactifications Y, each with a different size for $|Y - X|$.

But, for a given space X and a compactification Y, there <u>is</u> an upper bound for how large $|Y - X|$ can be. We can find it using the following two lemmas.

Recall that the weight $w(Y)$ of a space $(Y, \mathcal{T})$ is defined by $w(Y) = \aleph_0 + \min\{|\mathcal{B}| : \mathcal{B}$ is a base for $\mathcal{T}\}$. (*Example VI.4.6*)

Lemma 3.2 If Y is a T_0 space, then $|Y| \leq 2^{w(Y)}$.

Proof Let $\mathcal{B}$ be <u>any</u> base for Y, and for each point $y \in Y$, let $\mathcal{B}_y = \{U \in \mathcal{B} : y \in U\}$. Since Y is T_0, we have $\mathcal{B}_{y'} \neq \mathcal{B}_y$ if $y' \neq y$. Therefore the map $y \to \mathcal{B}_y \subseteq \mathcal{B}$ is one-to-one, so $|Y| \leq |\mathcal{P}(\mathcal{B})| = 2^{|\mathcal{B}|}$. In particular, if we pick $\mathcal{B}$ to be a base with the least possible cardinality, minimal cardinality, then $|Y| \leq 2^{|\mathcal{B}|} \leq 2^{w(Y)}$. •

Lemma 3.3 Suppose Y is an infinite T_3 space and that X is a dense subspace of Y. Then $w(Y) \leq 2^{|X|} \leq 2^{|Y|}$.

Proof A T_3 space with a finite base must be finite, so <u>every</u> base for Y must be infinite. Let $\mathcal{B} = \{U_\alpha : \alpha \in A\}$ be a base for Y. Each U_α is open so we have i) $U_\alpha \subseteq \operatorname{int} \operatorname{cl} U_\alpha \subseteq \operatorname{cl} U_\alpha$, <u>and</u> ii) because X is dense in Y, $\operatorname{cl} U_\alpha = \operatorname{cl}(U_\alpha \cap X)$ (*see Lemma IV.6.4*).

For each α, define $V_\alpha = \operatorname{int} \operatorname{cl}(U_\alpha \cap X)$, so that $U_\alpha \subseteq \operatorname{int} \operatorname{cl} U_\alpha = \operatorname{int} \operatorname{cl}(U_\alpha \cap X) = V_\alpha$.

Then $\mathcal{B}' = \{V_\alpha : \alpha \in A\}$ is also a base for Y: to see this, suppose $y \in O \subseteq Y$ where O is open. By regularity, there is a U_α such that $y \in U_\alpha \subseteq \operatorname{int} \operatorname{cl} U_\alpha = V_\alpha \subseteq \operatorname{cl} U_\alpha \subseteq O$.

Since each $U_\alpha \cap X \subseteq X$, there are no more distinct V_α's than there are subsets of X, that is $|\mathcal{B}'| \leq |\mathcal{P}(X)|$. Since $\mathcal{B}'$ must be infinite, we have $w(Y) \leq |\mathcal{B}'| \leq |\mathcal{P}(X)| = 2^{|X|} \leq 2^{|Y|}$. •

Theorem 3.4 If Y is a compactification of X and D is dense in X, then $|Y| \leq 2^{2^{|D|}}$.

Proof Y is Tychonoff. If Y is finite, then $D = X = Y$ so $|Y| \leq 2^{2^{|Y|}} = 2^{2^{|D|}}$. Therefore we can assume Y is infinite. Since D is dense in Y, $w(Y) \leq 2^{|D|}$ (by Lemma 3.3), and therefore so $|Y| \leq 2^{w(Y)} \leq 2^{2^{|D|}}$ (by Lemma 3.2) •

Example 3.5 An upper bound on the size of a compactification of $\mathbb{N}$ is $2^{2^{\aleph_0}} = 2^c$. More generally, a compactification of any separable Tychonoff space – such as $\mathbb{N}, \mathbb{Q}, \mathbb{P}$ or $\mathbb{R}$ – can have no more than 2^c points.

We will see in Section 6 that Theorem 3.4 is "best possible" upper bound. For example, there actually exists a compactification of $\mathbb{N}$, called $\beta\mathbb{N}$, with cardinality $2^{2^{\aleph_0}} = 2^c$! (It is difficult to imagine how the "tiny" discrete set $\mathbb{N}$ can be dense in a such a large compactification $\beta\mathbb{N}$.

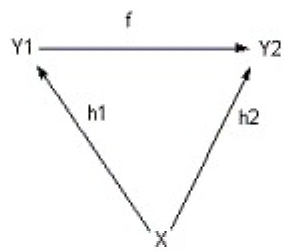
> *Assume such a compactification $\beta\mathbb{N}$ exists. Since $\mathbb{N}$ is dense, each point σ in $\beta\mathbb{N} - \mathbb{N}$ is the limit of a net in $\mathbb{N}$, and this net has a universal subnet which converges to σ.*
>
> *Since $\beta\mathbb{N}$ is Hausdorff, a universal net in $\mathbb{N}$ has at most one limit in $\beta\mathbb{N} - \mathbb{N}$, so there are at least as many universal nets in $\mathbb{N}$ as there are points in $\beta\mathbb{N} - \mathbb{N}$, namely 2^c. None of these universal nets can be trivial (that is, eventually constant). Therefore each of these universal nets is associated with a free (= nontrivial) ultrafilter in $\mathbb{N}$. So there must be 2^c free ultrafilters in $\mathbb{N}$.*

4. Comparing Compactifications

We want to compare compactifications of a Tychonoff space X. We begin by defining an equivalence relation $\simeq$ between compactifications of X. Then we define a relation $\geq$. It will turn out that $\geq$ can also be used to compare equivalence classes of compactifications of X. When applied in a set equivalence classes of compactifications of X, $\geq$ will turn out to be a partial ordering.

The definition of $\simeq$ requires that we use the formal definition of a compactification as a pair.

Definition 4.1 Two compactifications (Y_1, h_1) and (Y_2, h_2) of X are called equivalent, written $(Y_1, h_1) \simeq (Y_2, h_2)$, if there is a homeomorphism f of Y_1 onto Y_2 such that $f \circ h_1 = h_2$.



In the special case where $X \subseteq Y_1$, $X \subseteq Y_2$, and $h_1 = h_2 =$ the identity map on X, then the condition $f = f \circ h_1 = h_2$ simply states that $f(x) = x$ for $x \in X$ – that is, points in X are fixed under the homeomorphism f.

It is obvious that $(Y_1, h_1) \simeq (Y_1, h_1)$ and that $\simeq$ is a transitive relation among compactifications of X. Also, if $(Y_1, h_1) \simeq (Y_2, h_2)$, then $f^{-1} : Y_2 \to Y_1$ is a homeomorphism and $f^{-1} \circ h_2 = f^{-1}(f \circ h_1) = h_1$ so that $(Y_2, h_2) \simeq (Y_1, h_1)$. Therefore $\simeq$ is a symmetric relation, so $\simeq$ is an equivalence relation on any set of compactifications of X.

Example 4.2 Suppose X is a locally compact, noncompact Hausdorff space. We claim that all one-point compactifications of X are equivalent. Because $\simeq$ is transitive, it is sufficient to show that each one-point compactification (Y_1, h_1) is equivalent to the one-point compactification (Y^*, i) constructed in Theorem 2.3.

Let $Y^* = X \cup \{p\}$ and $Y_1 - h_1[X] = \{p_1\}$. Define $f : Y^* \to Y_1$ by

$$f(y) = \begin{cases} h_1(y) & \text{if } y \in X \\ p_1 & \text{if } y = p \end{cases}$$

f is clearly a bijection and $f \circ i = h_1$. We claim f is continuous.

> If $y \in X$: Let V be an open set in Y_1 with $f(y) = h_1(y) \in V$. Then $V' = V - \{p_1\}$ is also open in $h_1[X]$. Since $h_1 : X \to h_1[X]$ is a homeomorphism, $U = h_1^{-1}[V_1]$ is open in X and X is open in Y^*. Then $y \in U$, U is open in Y^* and $f[U] = h_1[U] = V' \subseteq V$. Therefore f is continuous at y.
>
> If $y = p$: Let V be an open set in Y_1 with $f(p) = p_1 \in V$. Then $Y_1 - V = K_1$ is a compact in $h_1[X]$, so $h_1^{-1}[K_1] = K$ is a compact (therefore closed) set in Y^*. Then $U = Y^* - K$ is a neighborhood of p and $f[U] \subseteq V$. Therefore f is continuous at p.

Since f is a continuous bijection from a compact space to a T_2 space, f is closed and therefore f is a homeomorphism.

Therefore (up to equivalence) we can talk about <u>the</u> one-point compactification of a noncompact, locally compact Hausdorff space X. Topologically, it makes no difference whether we think of the one-point compactification of $\mathbb{R}$ geometrically as S^1, with the North Pole p as the "point at infinity," or whether we think of it more abstractly as the result of the construction in Theorem 2.3.

Question: Are all two point compactifications of $(-1, 1)$ equivalent to $[-1, 1]$?

Example 4.3 Suppose (Y_1, h_1) is a compactification of X. Then (Y_1, h_1) is equivalent to a compactification (Y, i) where $X \subseteq Y$ and i is the identity map. We simply define $Y = (Y_1 - h_1[X]) \cup X$, topologized in the obvious way – in effect, we are simply giving each point $h_1(x)$ in Y_1 a new "name" x. We can then define $f : Y_1 \to Y$ by

$$f(z) = \begin{cases} z & \text{if } z \in Y_1 - h_1[X] \\ i \circ h_1^{-1}(z) = h_1^{-1}(z) & \text{if } z \in h_1[X] \end{cases}$$

Clearly, $f \circ h_1 = i$, so $(Y_1, h_1) \simeq (Y, i)$.

Example 4.3 shows means that whenever we work with only one compactification of X, or are discussing properties that are shared by all equivalent compactifications of X, we might as well (for simplicity) replace (Y_1, h_1) with an equivalent compactification Y where Y contains X as a dense subspace.

Example 4.4 Homeomorphic compactifications are not necessarily equivalent. In this example we see two dense embeddings h_1, h_2 of $\mathbb{N}$ into the same compact Hausdorff space Y that produce nonequivalent compactifications.

Let $Y = \{(\frac{1}{n}, i) : i = 1, 2 \text{ and } n \in \mathbb{N}\} \cup \{(0,1),\ (0,2)\} \subseteq \mathbb{R}^2$.

Let $h_1 : \mathbb{N} \to Y$ by $\begin{cases} h_1(2n) & = (\frac{1}{n}, 1) \\ h_1(2n-1) & = (\frac{1}{n}, 2) \end{cases}$. (Y, h_1) is a 2-point compactification of $\mathbb{N}$.

Let $h_2 : \mathbb{N} \to Y$ by $\begin{cases} h_2(n) = (\frac{1}{j}, 1) & \text{if } n \text{ is the } j^{\text{th}} \text{ element of } \{1, 2, 4, 5, 7, 8, 10, 11, ...\} \\ h_2(n) = (\frac{1}{j}, 2) & \text{if } n \text{ is the } j^{\text{th}} \text{ element of } \{3, 6, 9, 12, ...\} \end{cases}$

For example, $h_2(7) = (\frac{1}{5}, 1)$ and $h_2(9) = (\frac{1}{3}, 2)$. (Y, h_2) is also a two-point compactification of $\mathbb{N}$.

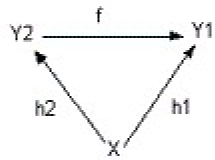
Topologically, each compactification is the same space Y, but (Y, h_1) and (Y, h_2) are <u>not equivalent</u> compactifications of $\mathbb{N}$:

> Suppose $f : Y \to Y$ is any (onto) homeomorphism.
>
> $(h_1(2n)) \to (0, 1)$, so $f(h_1(2n)) \to f((0, 1))$, and $f((0, 1)) =$ <u>either</u> $(0, 1)$ <u>or</u> $(0, 2)$ (*why?*).
>
> But the sequence $(h_2(2n)) = ((\frac{1}{2}, 1), (\frac{1}{3}, 1), (\frac{1}{2}, 2), (\frac{1}{6}, 1), (\frac{1}{7}, 1), (\frac{1}{4}, 2), ...)$. does <u>not</u> converge to either $(0, 1)$ or $(0, 2)$. Therefore $f \circ h_1 \neq h_2$, so $(Y_1, h_1) \not\simeq (Y_2, h_2)$

By adjusting the definitions of h_1 and h_2, we can create infinitely many nonequivalent 2-point compactifications of $\mathbb{N}$ all using different embeddings of $\mathbb{N}$ into the same space Y.

We now define a relation $\geq$ between compactifications of a space X.

Definition 4.5 Suppose (Y_2, h_2) and (Y_1, h_1) are compactifications of X. We say that $(Y_2, h_2) \geq (Y_1, h_1)$ if there exists a continuous function $f : Y_2 \to Y_1$ such that $f \circ h_2 = h_1$.



Notice that:

i) Such a mapping f is necessarily onto Y_1: $f[Y_2] \supseteq f[h_2[X]] = h_1[X]$, so $f[Y_2]$ is dense in Y_1. But $f[Y_2]$ is compact, so $f[Y_2] = Y_1$.

ii) If $X \subseteq Y_2$, $X \subseteq Y_1$ and $h_1 = h_2 =$ the identity map i, then the condition $f \circ h_2 = h_1$ simply states that $f|X = i$.

iii) $f[Y_2 - h_2[X]] \subseteq Y_1 - h_1[X]$: that is, the "points added" to create Y_2 are mapped onto the "points added" to create Y_1. To see this, let $z \in Y_2 - h_2[X]$. We want to show $f(z) \in Y_1 - h_1[X]$. So suppose that $f(z) = h_1(x) \in h_1[X]$.

Since $h_2[X]$ is dense in Y_2, there is a net in $h_2[X]$ converging to z :

$$(h_2(x_\lambda)) \to z \qquad (*)$$

f is continuous, so $\quad h_1(x_\lambda) = f(h_2(x_\lambda)) \to f(z) = h_1(x)$.
But $h_1 : X \to h_1[X]$ is a homeomorphism so

$$(x_\lambda) = (h_1^{-1}h_1(x_\lambda)) \to h_1^{-1}h_1(x) = x \in X,$$

and therefore $\quad (h_2(x_\lambda)) \to h_2(x) \in h_2[X] \qquad (**)$

A net in Y_2 has at most one limit, so $(*)$ and $(**)$ give that $z = h_2(x)$. This is impossible since $z \notin h_2[X]$.

iv) From iii) we conclude that if $(Y_2, h_2) \geq (Y_1, h_1)$, then $|Y_2 - h_2[X]| \geq |Y_1 - h_1[X]|$

Suppose $(Y_2, h_2) \geq (Y_1, h_1)$. The next theorem tells us that the relation " $\geq$ " is unaffected if we replace these compactifications of X with equivalent compactifications – so we can actually compare <u>equivalence classes</u> of compactifications of X by comparing <u>representatives</u> of the equivalence classes. The proof is very easy and is omitted.

Theorem 4.6 Suppose (Y_2, h_2) and (Y_1, h_1) are compactifications of X and that $(Y_2, h_2) \geq (Y_1, h_1)$. If $(Y_2, h_2) \simeq (Y_2', h_2')$ and $(Y_1, h_1) \simeq (Y_1', h_1')$, then $(Y_2', h_2') \geq (Y_1', h_1')$.

The ordering " $\geq$ " is well behaved on the equivalence classes of compactifications of X.

Theorem 4.7 Let $\mathcal{C}$ be a set of equivalence classes of compactifications of X. Then $(\mathcal{C}, \geq)$ is a poset.

Proof It is clear from the definition that $\geq$ is both reflexive and transitive. We need to show that $\geq$ is also antisymmetric. Suppose $[(Y_1, i)]$ and $[(Y_2, i)]$ are equivalence classes of compactifications of X (*By Theorem 4.6, we are free to choose from each equivalence class representative compactifications with $X \subseteq Y_i$ and embeddings $h = i =$ the identity map*).

If both $(Y_1, i) \geq (Y_2, i)$ and $(Y_2, i) \geq (Y_1, i)$ hold, then we have the following maps:

$$\begin{tikzcd}
Y1 \arrow[rr, shift left, "f"] & & Y2 \arrow[ll, shift left, "g"] \\
& X \arrow[ul, "i"] \arrow[ur, "i"'] &
\end{tikzcd}$$

with $f \circ i = i = g \circ i$. For $x \in X$, $g(f(x)) = g(f(i(x)) = g(i(x)) = i(x) = x$, so the maps $g \circ f$ and the identity $i : Y_1 \to Y_1$ agree on the <u>dense</u> subspace X. Since Y_1 is Hausdorff, it follows that $g \circ f = i$ everywhere in Y_1. (*See Theorem 5.12 in Chapter II, and its generalization in Exercise E9 of Chapter III.*) Similarly $f \circ g$ and $i : Y_2 \to Y_2$ agree on the dense subspace X so $f \circ g = i$ on Y_2.

Since $f \circ g = i$ and $g \circ f = i$, f and g are inverse functions and f is a homeomorphism. Therefore $(Y_1, i) \simeq (Y_2, i)$. So we have shown that if $[(Y_1, i)] \geq [(Y_2, i)]$ and $[(Y_1, i)] \leq [(Y_2, i]$, then $[(Y_1, i)] = [(Y_2, i]$ •

An equivalence <u>class</u> of compactifications of a space X is "too big" to be a set in ZFC set theory. (It is customary to refer informally to such collections "too big" to be sets in ZFC as "classes.")

<u>However</u>, suppose (Y, i) represents one of these equivalence classes. If X has weight m, then X contains a dense set D with $|D| \leq m$. It follows from Lemma 3.3 that $w(Y) \leq 2^m$ so, by Theorem VII.3.17, Y can be embedded in the cube $[0,1]^{2^m}$. Therefore <u>every</u> compactification of X can be <u>represented</u> by a subspace of the <u>one</u> fixed cube $[0,1]^{2^m}$.

Therefore we can form a <u>set</u> consisting of one representative from each equivalence class of compactifications of X : this set is just a certain set of subspaces of $[0,1]^{2^m}$. This set is partially ordered by $\geq$.

In fact, we can even given a bound on the number of different equivalence classes of compactifications of X: since every compactification of X can be represented as a subspace of $[0,1]^{2^m}$, the number of equivalence classes of compactifications of X is no more than $|\mathcal{P}([0,1]^{2^m})| = 2^{(c^{2^m})} = 2^{2^{2^m}}$. In other words, there are no more than $2^{2^{2^m}}$ different compactifications of X.

Example 4.8 Let (Y_1, h_1) be a 1-point compactification of X. For every compactification (Y, h) of X, $(Y, h) \geq (Y_1, h_1)$. (*So, among equivalence classes of compactifications of X, the equivalence class* $[(Y_1, h_1)]$ *is <u>smallest</u>*.)

By Theorem 4.6, we may assume $X \subseteq Y_1$, $X \subseteq Y$ and that h, h_1 are the identity maps; in fact, we may as well assume $Y_1 = X^*$ (the one-point compactification constructed in Theorem 2.3).

Since X has a one-point compactification, X is locally compact (*see Example 1.5*). By Corollary 2.7, X is open in both Y and X^*.

Let $X^* - X = \{p\}$ and define

$$f : Y \to X^* \text{ by } f(y) = \begin{cases} y & \text{if } y \in X \\ p & \text{if } z \in Y - X \end{cases}$$

To show that $(Y, i) \geq (X^*, i)$, we only need to check that f is continuous each point $z \in Y$.

> If $y \in X$ and V is an open set containing $f(y) = y$ in X^*, then $y \in U = V - \{p\}$ which is open in X and therefore also open in Y. Clearly, $f[U] = U \subseteq V$.
>
> If $z \in Y - X$ and V is an open neighborhood of $f(z) = p$ in X^*, then $X^* - V = K$ is a compact subset of X. Therefore K is closed in Y so $U = Y - K$ is an open set in Y containing z and $f[U] \subseteq V$.

5. The Stone-Cech Compactification

Example 4.8 shows that the one-point compactification of a space X, when it exists, is the smallest compactification of X. Perhaps it is surprising that every Tychonoff space X has a largest compactification and, by Theorem 4.7, this compactification is unique up to equivalence. In other words, a poset which consists of one representative of each equivalence class of compactifications of X has a largest (not merely maximal!) element. This largest compactification of X is called the Stone-Cech (pronounced "check") compactification and is denoted by βX.

ce

Theorem 5.1 1) Every Tychonoff Space X has a largest compactification, and this compactification is unique up to equivalence. (*We may represent the largest compactification by* $(\beta X, i)$ *where* $\beta X \supseteq X$ *and* i *is the identity map. We do this in the remaining parts of theorem.*)

2) Suppose X is Tychonoff and that Y is a compact Hausdorff space. Every continuous $f : X \to Y$ has a unique continuous extension $f^\beta : \beta X \to Y$. (*The extension* f^β *is called the Stone extension of* f. *The property of* βX *in 2) is called the Stone Extension Property.*)

3) Up to equivalence, βX is the only compactification of X with the Stone Extension Property. (*In other words, the Stone Extension Property characterizes* βX *among all compactifications of* X.)

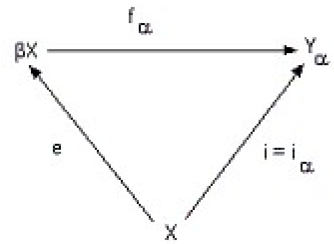
Example 5.2 (assuming Theorem 5.1) $[0,1]$ is a compactification of $(0,1]$. However the continuous function $f : (0,1] \to Y = [-1,1]$ given by $f(x) = \sin(\frac{1}{x})$ cannot be continuously extended to a map $f^\beta : [0,1] \to Y$. Therefore $[0,1] \neq \beta(0,1]$. Is it possible that $S^1 = \beta\mathbb{R}$?

Proof of Theorem 5.1

1) Since X is Tychonoff, X has at least one compactification. Let $\{(Y_\alpha, i_\alpha) : \alpha \in A\}$ be a set of compactifications of X, where $X \subseteq Y_\alpha$, $i_\alpha : X \to Y_\alpha$ is the identity, and one Y_α is chosen from each equivalence class of compactifications of X. (*As noted in the remarks following Theorem 4.7, this is a legitimate set since every compactification of X can be represented as subset of one fixed cube* $[0,1]^k$.)

Define $e : X \to Y = \prod\{Y_\alpha : \alpha \in A\}$ by $e(x)(\alpha) = i(x) = x$. This "diagonal" map e sends each x to the point in the product all of whose coordinates are x, and e is the evaluation map generated by the collection of maps $i_\alpha : X \to Y_\alpha$. X is a subspace of Y_α, and the subspace topology is precisely the weak topology induced on X by each i_α (*see Example VI.2.5*). It follows from Theorem VI.4.4 that e is an embedding of X into the compact space Y. If we define $\beta X = \mathrm{cl}_Y\, e[X]$, then $(\beta X, e)$ is a compactification of X.

Every compactification of X is equivalent to one of the (Y_α, i_α). Therefore, to show βX is the largest compactification we need only show that $(\beta X, e) \geq (Y_\alpha, i_\alpha)$ for each $\alpha \in A$. This, however, is clear: in the diagram below, simply let $f_\alpha = \pi_\alpha | \beta X$.



Then $f_\alpha \circ e = i$ because $f_\alpha(e(x)) = \pi_\alpha(e(x)) = e(x)(\alpha) = i(x) = x$.

Therefore $(\beta X, e) \geq (Y_\alpha, i)$.

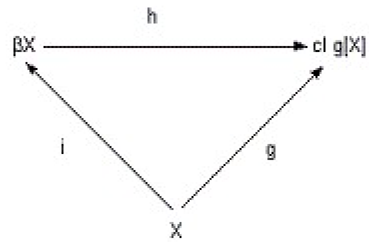
Note: now that the construction is complete, we can replace $(\beta X, e)$ *with an equivalent largest compactification actually containing* X : $(\beta X, i)$.

Since $\geq$ is antisymmetric among the compactifications (Y_α, i_α), the largest compactification of X is unique (up to equivalence).

2) Suppose $f : X \to Y$ where Y is a compact Hausdorff space. First, we need to produce a continuous extension $f^\beta : \beta X \to Y$.

Define $g : X \to \beta X \times Y$ by $g(x) = (x, f(x))$. Clearly, g is $1-1$ and continuous, and X has the weak topology generated by the maps $i : X \to X$ and $f : X \to Y$, so g is an embedding. Since $\beta X \times Y$ is compact, $(\mathrm{cl}_{\beta X \times Y}\, g[X],\ g)$ is a compactification of X.

But $(\beta X, i) \geq (\mathrm{cl}\, g[X], g)$, so we have a continuous map $h : \beta X \to \mathrm{cl}\, g[X]$ for which $h \circ i = g$ $-$ that is $h(x) = g(x)$ for $x \in X$ (*see the following diagram*)



For $z \in \beta X$, define $f^\beta(z) = \pi_Y \circ h(z)$. Then f^β is continuous and for $x \in X$ we have $f^\beta(x) = \pi_Y(h(x)) = \pi_Y(h(i(x))) = \pi_Y(g(x)) = \pi_Y(x, f(x)) = f(x)$, so $f^\beta|X = f$.

If $k : \beta X \to Y$ is continuous and $k|X = f$, then k and f^β agree on the dense set X, so $k = f^\beta$ Therefore the Stone extension f^β is unique. (*See Theorem II.5.12, and its generalization in exercise E9 of Chapter III.*.)

3) Suppose (Y, i) is a compactification of X with the Stone Extension Property. Then the identity map $i : X \to \beta X$ has an extension $i^* : Y \to \beta X$ such that $i^* \circ i = i$, so $(Y, i) \geq (\beta X, i)$. Since βX is the largest compactification of X, $(Y, i) \simeq (\beta X, i)$. ●

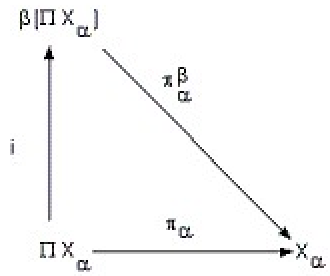
The Tychonoff Product Theorem is equivalent to the Axiom of Choice AC (*see Theorem IX.6.5*). Our construction of βX used the Tychonoff Product Theorem – but only applied to a collection of compact Hausdorff spaces. In fact, as we show below, the existence of a largest compactification βX is equivalent to "the Tychonoff Product Theorem for compact T_2 spaces."

The "Tychonoff Product Theorem for compact T_2 spaces" also cannot be proven in ZF, but it is strictly weaker than AC. (*In fact, the "Tychonoff Product Theorem for compact T_2 spaces" is equivalent to a statement called the "Boolean Prime Ideal Theorem."*)

The main point is that the very existence of βX involves set-theoretic issues and any method for constructing βX must, in some form, use something beyond ZF set theory – something quite close to the Axiom of Choice.

Theorem 5.3 If every Tychonoff space X has of a largest compactification βX, then any product of compact Hausdorff spaces is compact.

Proof Suppose $\{X_\alpha : \alpha \in A\}$ is a collection of compact T_2 spaces. Since $\prod_{\alpha\in A} X_\alpha$ is Tychonoff, it has a compactification $\beta\,[\prod_{\alpha\in A} X_\alpha]$ and for each α the projection map π_α can be extended to $\pi_\alpha^\beta : \beta\,[\prod_{\alpha\in A} X_\alpha] \to X_\alpha$.



For each $x \in \beta\,[\prod_{\alpha\in A} X_\alpha]$, define a point $f(x) \in X$ with coordinates $f(x)(\alpha) = \pi_\alpha^\beta(x)$.

$$f : \beta[\textstyle\prod_{\alpha\in A} X_\alpha] \to \prod_{\alpha\in A} X_\alpha$$

f is continuous because each coordinate function π_α^β is continuous. If $x \in \prod_{\alpha\in A} X_\alpha$ $\subseteq \beta[\prod_{\alpha\in A} X_\alpha]$, then $f(x)(\alpha) = \pi_\alpha^\beta(x) = \pi_\alpha(x) = x(\alpha) = x_\alpha$ for each α, so $f(x) = x$. Therefore $\prod_{\alpha\in A} X_\alpha$ is a continuous image of the compact space $\beta\,[\prod_{\alpha\in A} X_\alpha]$, so $\prod_{\alpha\in A} X_\alpha$ is compact. •

We want to consider some other ways to recognize βX. Since βX can be characterized by the Stone Extension Property, the following technical theorem about extending continuous functions will be useful.

Theorem 5.4 (Taimonov) Suppose C is a dense subspace of a Tychonoff space X and let Y be a compact Hausdorff space. A continuous function $f : C \to Y$ has a continuous extension $\widetilde{f} : X \to Y$ iff

whenever A and B are disjoint closed sets in Y, $\mathrm{cl}_X f^{-1}[A] \cap \mathrm{cl}_X f^{-1}[B] = \emptyset$.

Proof $\Rightarrow$: If $\widetilde{f}$ exists and A and B are disjoint closed sets in Y, then $\widetilde{f}^{-1}[A] \cap \widetilde{f}^{-1}[B] = \emptyset$. But these sets are closed in X, so $\widetilde{f}^{-1}[A] = \mathrm{cl}_X f^{-1}[A]$ and $\widetilde{f}^{-1}]B] = \mathrm{cl}_X f^{-1}[B]$. Therefore $\mathrm{cl}_X f^{-1}[A] \cap \mathrm{cl}_X f^{-1}[B] = \emptyset$.

$\Leftarrow$: We must define a function $\widetilde{f}\ : X \to Y$ such that $\widetilde{f}\,|C = f$ and then show that $\widetilde{f}$ is continuous. For $x \in X$, let $\mathcal{N}_x$ be its neighborhood filter in X. Define a collection of closed sets $\mathcal{F}_x$ in Y by

$$\mathcal{F}_x = \{\text{cl}\, f[C \cap U] : U \in \mathcal{N}_x\}$$

Then $\text{cl}\, f[C \cap U_1] \cap \text{cl}\, f[C \cap U_2] \supseteq \text{cl}\, f[C \cap U_1 \cap U_2] \neq \emptyset$ (since C is dense in X). Therefore $\mathcal{F}_x$ is a family of closed sets in Y with the finite intersection property so $\bigcap \mathcal{F}_x \neq \emptyset$ (because Y is compact) .

We claim that $\bigcap \mathcal{F}_x$ contains only one point: $\bigcap \mathcal{F}_x = \{y\}$ for some $y \in Y$.

Suppose $y, z \in \bigcap \mathcal{F}_x$. If $y \neq z$, then (since Y is T_3) we can pick open sets U, V so that $y \in U$ and $z \in V$ and $\text{cl}\, U \cap \text{cl}\, V = \emptyset$. Then $\text{cl}_X f^{-1}[\text{cl}\, U] \cap \text{cl}_X f^{-1}[\text{cl}\, V] = \emptyset$ so, of course, $\text{cl}_X f^{-1}[U] \cap \text{cl}_X f^{-1}[V] = \emptyset$. Taking complements, we get

$$(X - \text{cl}_X f^{-1}[U]) \cup (X - \text{cl}_X f^{-1}[V]) = X$$

so x is in one of these open sets: say $x \in W = X - \text{cl}_X f^{-1}[U]$. Since $W \in \mathcal{N}_x$, $\text{cl}\, f[C \cap W] \in \mathcal{F}_x$. We claim $\text{cl}\, f[C \cap W] \subseteq Y - U$, from which will follow the contradiction that $y \notin \bigcap \mathcal{F}_x$.

To check this inclusion, simply note that $C \cap W = C - \text{cl}_X f^{-1}[U]$. Therefore, if $u \in C \cap W$, we have $u \notin \text{cl}_X f^{-1}[U]$, so $u \notin f^{-1}[U]$, so $f(u) \notin U$. Thus, $f[C \cap W] \subseteq Y - U$ (a closed set) so $\text{cl}\, f[C \cap W] \subseteq Y - U$.

Define $\widetilde{f}\ (x) = y$. We claim that $\widetilde{f}$ works.

$\widetilde{f}\,|C = f$: Suppose $x \in C$. $\mathcal{B} = \{C \cap U : U \in \mathcal{N}_x\}$ is the neighborhood filter of x <u>in C</u> so $\mathcal{B} \to x$ in C. Since f is continuous, the filter base $f[\mathcal{B}] = \{f[C \cap U] : U \in \mathcal{N}_x\}$ $\to f\,(x)$ in Y. In particular, $f(x)$ is a cluster point of $f[\mathcal{B}]$, so $f(x) \in \bigcap \text{cl}(f[C \cap U])$ $= \bigcap \mathcal{F}_x = \{\widetilde{f}\ (x)\}$. So $f(x) = \widetilde{f}\ (x)$.

$\widetilde{f}$ is continuous: Let $x \in X$ and let V be open in Y with $y = \widetilde{f}\ (x) \in V$. Since $\bigcap \mathcal{F}_x = \{y\} \subseteq V$, there exist $U_1, ..., U_n \in N_x$ such that

$$\text{cl}\, f[C \cap U_1] \cap ... \cap \text{cl}\, f[C \cap U_n] \subseteq V$$

(*If V is an open set in a compact space and $\mathcal{F}$ is a family of closed sets with $\bigcap \mathcal{F} \subseteq V$, then some finite subfamily of $\mathcal{F}$ satisfies $F_1 \cap ... \cap F_n \subseteq V$. Why?*)

Let $W = U_1 \cap ... \cap U_n \in \mathcal{N}_x$. If $z \in W$, then

$$\widetilde{f}\ (z) \in \text{cl}\, f\,[C \cap W] \subseteq \text{cl}\, f[C \cap U_1] \cap ... \cap \text{cl}\, f\,[C \cap U_n] \subseteq V$$

so $\widetilde{f}\ [W] \subseteq V$. Therefore $\widetilde{f}$ is continuous at x. $\bullet$

Corollary 5.5 Suppose Y_1 and Y_2 are compactification of X where the embeddings are the identity map. Then $(Y_1, i) \simeq (Y_2, i)$ iff : for every pair of disjoint closed sets in X,

$$\text{cl}_{Y_1} A \cap \text{cl}_{Y_1} B = \emptyset \Leftrightarrow \text{cl}_{Y_2} A \cap \text{cl}_{Y_2} B = \emptyset) \qquad (*)$$

Proof If $(Y_1, i) \simeq (Y_2, i)$, it is clear that (*) holds. If (*) holds, then Taimonov's Theorem guarantees that the identity maps $i_1 : X \to Y_1$ and $i_2 : X \to Y_2$ can be extended to maps $f_1 : Y_2 \to Y_1$ and $f_2 : Y_1 \to Y_2$. It is clear that $f_1 \circ f_2 | X$ and $f_2 \circ f_1 | X$ are the identity maps on the dense subspace X. Therefore $f_1 \circ f_2$ and $f_2 \circ f_1$ are each the identity everywhere so f_1, f_2 are homeomorphisms so $(Y_1, i) \simeq (Y_2, i)$. •

For convenience, we repeat here a definition included in the statement of Theorem VII.5.2

Definition 5.6 Suppose A and B are subspaces of X. A and B are completely separated if there exists $f \in C(X)$ such that $f|A = 0$ and $f|B = 1$. (It is easy to see that $0, 1$ can be replaced in the definition by any two real numbers a, b.)

Urysohn's Lemma states that disjoint closed sets in a normal space are completely separated.

Using Taimonov's theorem, we can characterize βX in several different ways. In particular, condition 4) in the following theorem states that βX is actually characterized by the "extendability" of continuous functions from X into $[0, 1]$ – a statement which looks weaker than the full Stone Extension Property.

Theorem 5.7 Suppose (Y, i) is a compactification of X, where $Y \supseteq X$ and i is the identity. The following are equivalent:

1) Y is βX (that is, Y is the largest compactification of X)
2) every continuous $f : X \to K$, where K is a compact Hausdorff space, can be extended to a continuous map $\widetilde{f} : Y \to K$
3) every continuous $f : X \to [a, b]$ can be extended to a continuous $\widetilde{f} : Y \to [a, b]$
4) every continuous $f : X \to [0, 1]$ can be extended to a continuous $\widetilde{f} : Y \to [0, 1]$
5) completely separated sets in X have disjoint closures in Y
6) disjoint zero sets in X have disjoint closures in Y
7) if Z_1 and Z_2 are zero sets in X, then $\text{cl}_Y(Z_1 \cap Z_2) = \text{cl}_Y Z_1 \cap \text{cl}_Y Z_2$,

Proof Theorem 5.1 gives that 1) and 2) are equivalent, and the implications 2) $\Rightarrow$ 3) $\Rightarrow$ 4) are trivial.

4) $\Rightarrow$ 5) If A and B are completely separated in X, then there is a continuous $f : X \to [0, 1]$ with $f|A = 0$ and $f|B = 1$. By 4), f extends to a continuous map $\widetilde{f} : Y \to [0, 1]$. Then $\text{cl}_Y A \subseteq \text{cl}_Y \widetilde{f}^{-1}(0) = \widetilde{f}^{-1}(0)$ and $\text{cl}_Y B \subseteq \text{cl}_Y \widetilde{f}^{-1}(1) = \widetilde{f}^{-1}(1)$, so $\text{cl}_Y A \cap \text{cl}_Y B = \emptyset$.

5) $\Rightarrow$ 6) Disjoint zero sets $Z(f)$ and $Z(g)$ in X are completely separated (for example, by the function $h = \frac{f^2}{f^2 + g^2}$) and therefore, by 5), have disjoint closures in Y.

6 $\Rightarrow$ 7) A zero set neighborhood of X is a zero set Z with $x \in \operatorname{int} Z$. It is easy to show that in a Tychonoff space X, the zero set neighborhoods of x form a neighborhood base at x (*check this!*).

Suppose Z_1 and Z_2 are zero sets in X. Certainly, $\mathrm{cl}_Y(Z_1 \cap Z_2) \subseteq \mathrm{cl}_Y Z_1 \cap \mathrm{cl}_Y Z_2$, so suppose $x \in \mathrm{cl}_Y Z_1 \cap \mathrm{cl}_Y Z_2$. If V is a zero set neighborhood of x, then $x \in \mathrm{cl}_Y(Z_1 \cap V)$ and $x \in \mathrm{cl}_Y(Z_2 \cap V)$ (*why?*). $Z_1 \cap V$ and $Z_2 \cap V$ are zero sets in X and $x \in \mathrm{cl}_Y(Z_1 \cap V) \cap \mathrm{cl}_Y(Z_2 \cap V)$ so, by 6). $(Z_1 \cap V) \cap (Z_2 \cap V) = Z_1 \cap Z_2 \cap V \neq \emptyset$.

Since every zero set neighborhood V of x intersects $Z_1 \cap Z_2$, and the zero set neighborhoods of x are a neighborhood base, we have $x \in \mathrm{cl}_Y(Z_1 \cap Z_2)$.

7) $\Rightarrow$ 2) Suppose that $f : X \to K$ is continuous. K is T_4 so if A and B are disjoint closed sets in K, there is a continuous $g : K \to [0,1]$ such that $A \subseteq \{x : g(x) = 0\} = Z_1$ and $B \subseteq \{x : g(x) = 1\} = Z_2$.

Then $f^{-1}[A] \subseteq f^{-1}[Z_1]$ and $f^{-1}[B] \subseteq f^{-1}[Z_2]$ and $f^{-1}[Z_1]$ and $f^{-1}[Z_2]$ are disjoint zero sets in X. By 7), $\mathrm{cl}_Y f^{-1}[A] \cap \mathrm{cl}_Y f^{-1}[B] \subseteq \mathrm{cl}_Y f^{-1}[Z_1] \cap \mathrm{cl}_Y f^{-1}[Z_2] = \mathrm{cl}_Y f^{-1}[Z_1 \cap Z_2]$ $= \emptyset$. By Taimonov's Theorem 5.4, f has a continuous extension $\widetilde{f} : Y \to K$. ●

Example 5.8

1) By Theorem VIII.8.8, every continuous function $f : [0, \omega_1) \to [0,1]$ is "constant on a tail" so f can be continuously extended to $\widetilde{f} : [0, \omega_1] \to [0,1]$. By Theorem 5.7, $[0, \omega_1] = \beta[0, \omega_1)$.

In this case the largest compactification of $[0, \omega_1)$ is the same as the smallest compactification – the one-point compactification. Therefore, up to equivalence, $[0, \omega_1]$ is the <u>only</u> compactification of $[0, \omega_1)$.

A similar example of this phenomenon is $T^* = [0, \omega_1] \times [0, \omega_0] = \beta T$, where $T = T^* - \{(\omega_1, \omega_0)\}$ (see the "Tychonoff plank" in Example VIII.8.10 and Exercise VIII.8.11).

2) The one-point compactification $\mathbb{N}^*$ of $\mathbb{N}$ is <u>not</u> $\beta\mathbb{N}$ because the function $f : \mathbb{N} \to \{0,1\}$ given by $f(n) = \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 & \text{if } n \text{ is odd} \end{cases}$ cannot be continuously extended to $\widetilde{f} : \mathbb{N}^* \to \{0,1\}$. (*Why? It might help think of* $\mathbb{N}$ *(topologically) as* $\{\frac{1}{n} : n \in \mathbb{N}\} \subseteq \mathbb{R}$.)

Theorem 5.9 βX is metrizable iff X is a compact metrizable space (i.e., iff X is metrizable and $X = \beta X$).

Proof $\Leftarrow$: Trivial

$\Rightarrow$: βX is metrizable $\Rightarrow \begin{cases} X \text{ is metrizable} \Rightarrow X \text{ is } T_4 \\ \beta X \text{ is first countable} \end{cases}$.

If X is not compact, there is a sequence (x_n) in X with $(x_n) \to p \in \beta X - X$. Without loss of generality, we may assume the x_n's are distinct (*why?*).

Let $O = \{x_1, x_3, ..., x_{2n+1}, ...\}$ and $E = \{x_2, x_4, ..., x_{2n}, ...\}$. O and E are disjoint closed sets in X so Urysohn's Lemma gives us a continuous $f : X \to [0,1]$ for which $f|O = 0$ and $f|E = 1$. Let $f^\beta : \beta X \to [0,1]$ be the Stone Extension of f. Then $f^\beta(p) = \lim_{n\to\infty} f^\beta(x_{2n+1})$
$= \lim_{n\to\infty} f(x_{2n+1}) = 0 \neq 1 = \lim_{n\to\infty} f(x_{2n}) = \lim_{n\to\infty} f^\beta(x_{2n}) = f^\beta(p)$, which is impossible. •

6. The space $\beta\mathbb{N}$

The Stone-Cech compactification of $\mathbb{N}$ is a strange and curious space.

Example 6.1 $\beta\mathbb{N}$ is a compact Hausdorff space in which $\mathbb{N}$ is a countable dense set. Since $\beta\mathbb{N}$ is separable, Theorem 3.4 gives us the upper bound $|\beta\mathbb{N}| \leq 2^{2^{\aleph_0}} = 2^c$.

On the other hand, suppose $f : \mathbb{N} \to [0,1] \cap \mathbb{Q}$ is a bijection and consider the Stone extension $f^\beta : \beta\mathbb{N} \to [0,1]$. Since $f^\beta[\beta\mathbb{N}]$ is compact, it is a closed set in $[0,1]$ and it contains the dense set $\mathbb{Q}$. Therefore $f^\beta[\beta\mathbb{N}] = [0,1]$ so we have $c \leq |\beta\mathbb{N}| \leq 2^c$.

A similar argument makes things even clearer. By Pondiczerny's Theorem VI.3.5, there is a countable dense set $D \subseteq [0,1]^{[0,1]}$. Pick a bijection $f : \mathbb{N} \to D$ and consider the extension $f^\beta : \beta\mathbb{N} \to [0,1]^{[0,1]}$. Just as before, f^β must be onto. Therefore $[\beta\mathbb{N}] \geq |[0,1]^{[0,1]}| = c^c = 2^c$.

Combining this with our earlier upper bound, we conclude that $|\beta\mathbb{N}| = 2^c$. $\beta\mathbb{N}$ is quite large but it contains the dense discrete set $\mathbb{N}$ that is merely countable.

Every set $A \subseteq \mathbb{N}$ is a zero set in $\mathbb{N}$ so we can write

$$\beta\mathbb{N} = \mathrm{cl}_{\beta\mathbb{N}}\,\mathbb{N} = \mathrm{cl}_{\beta\mathbb{N}} A \cup \mathrm{cl}_{\beta\mathbb{N}}(\mathbb{N} - A),$$

and by Theorem 5.7(6) these sets are disjoint. Therefore for each $A \subseteq \mathbb{N}$, $\mathrm{cl}_{\beta\mathbb{N}} A$ is a clopen set in $\beta\mathbb{N}$. In particular, each singleton $A = \{n\}$ is open in $\beta\mathbb{N}$ (that is, n is isolated in $\beta\mathbb{N}$), so $\mathbb{N}$ is open in $\beta\mathbb{N}$. Therefore $\beta\mathbb{N} - \mathbb{N}$ is compact.

At each $x \in \beta\mathbb{N}$, there is a neighborhood base $\mathcal{B}_x$ consisting of clopen neighborhoods:

i) if $x \in \mathbb{N}$, we can use $\mathcal{B}_x = \{\{x\}\}$
ii) if $x \in \beta\mathbb{N} - \mathbb{N}$, we can use $\mathcal{B}_x = \{\mathrm{cl}_{\beta\mathbb{N}} A : A \subseteq \mathbb{N} \text{ and } x \in \mathrm{cl}_{\beta\mathbb{N}} A\}$

If U is an open set in $\beta\mathbb{N}$ containing x, we can use regularity to choose an open set W such that $x \in W \subseteq \mathrm{cl}_{\beta\mathbb{N}} W \subseteq U$. If $A = W \cap \mathbb{N}$, then $x \in \mathrm{cl}_{\beta\mathbb{N}} A = \mathrm{cl}_{\beta\mathbb{N}}(W \cap \mathbb{N}) = \mathrm{cl}_{\beta\mathbb{N}} W \subseteq U$. (*Why?*)

Definition 6.2 Suppose $A \subseteq X$. A is said to be C^*-embedded in X if every $f \in C^*(A)$ has a continuous extension $\widetilde{f} \in C^*(X)$.

To illustrate the terminology:

i) Tietze's Theorem states that every closed subspace of a normal space is C^*-embedded.

ii) For a Tychonoff space X, βX is <u>the</u> compactification (up to equivalence) in which X is C^*-embedded.

The following theorem is very useful in working with βX.

Theorem 6.3 Suppose $A \subseteq X \subseteq \beta X$, and that A is C^*-embedded in X. Then $\mathrm{cl}_{\beta X} A = \beta A$.

Proof If $f : A \to [0,1]$ is continuous, then f extends continuously to $\overline{f} : X \to [0,1]$, and, in turn, $\overline{f}$ extends continuously to $\widetilde{f} : \beta X \to [0,1]$. Then $\widetilde{f}\,|\mathrm{cl}_{\beta X} A$ is a continuous extension of f to $\mathrm{cl}_{\beta X} A$. Since $\mathrm{cl}_{\beta X} A$ has the extension property in Theorem 5.7 (4), $\mathrm{cl}_{\beta X} A = \beta A$. •

Example 6.4 Since $\mathbb{N}$ is discrete, every $A \subseteq \mathbb{N}$ is C^*-embedded in $\mathbb{N}$ and so, by Theorem 6.3, $\mathrm{cl}_{\beta\mathbb{N}} A = \beta A$.

Of course if A is finite, $\mathrm{cl}_{\beta\mathbb{N}} A = A = \beta A$. But if A is infinite, then A is homeomorphic to $\mathbb{N}$, so $\mathrm{cl}_{\beta\mathbb{N}} A = \beta A$ is homeomorphic to $\beta\mathbb{N}$.

In particular, if $\mathbb{E}$ and $\mathbb{O}$ are the sets of even and odd natural numbers, we have $\mathbb{N} = \mathbb{E} \cup \mathbb{O}$, so $\beta\mathbb{N} = \mathrm{cl}_{\beta\mathbb{N}}\mathbb{E} \cup \mathrm{cl}_{\beta\mathbb{N}}\mathbb{O}$ – so $\beta\mathbb{N}$ is the union of two disjoint, clopen copies of itself. It is easy to modify this argument to show that, for any natural number k, $\beta\mathbb{N}$ <u>can be written as the union of k disjoint clopen copies of itself</u>.

If we write $\mathbb{N} = \bigcup_{k=1}^{\infty} A_k$, where each A_k's are pairwise disjoint infinite subsets of $\mathbb{N}$, then we have $\beta\mathbb{N} = \mathrm{cl}_{\beta\mathbb{N}} \bigcup_{k=1}^{\infty} A_k \supseteq \bigcup_{k=1}^{\infty} \mathrm{cl}_{\beta\mathbb{N}} A_k$, and these sets $\mathrm{cl}_{\beta\mathbb{N}} A_k$ are pairwise disjoint copies of $\beta\mathbb{N}$. Moreover, $\bigcup_{k=1}^{\infty} \mathrm{cl}_{\beta\mathbb{N}} A_k$ is dense in $\beta\mathbb{N}$ since the union contains $\mathbb{N}$. (*If we choose the A_k's properly chosen, can we have $\beta\mathbb{N} = \bigcup_{k=1}^{\infty} \mathrm{cl}_{\beta\mathbb{N}} A_k$? Why or why not?*)

Example 6.5 No sequence (n_k) in $\mathbb{N}$ can converge to a point of $\beta\mathbb{N} - \mathbb{N}$. In particular, the sequence (n) has no convergent subsequence in $\beta\mathbb{N}$ so $\beta\mathbb{N}$ is not sequentially compact.

Define $f : \mathbb{N} \to \{0,1\}$ by $f(x) = \begin{cases} 1 & \text{if } x = n_{2k} \text{ for some } k \\ 0 & \text{otherwise} \end{cases}$. Consider the Stone extension $f^{\beta} : \beta\mathbb{N} \to \{0,1\}$. If $(n_k) \to p \in \beta\mathbb{N} - \mathbb{N}$, then $(f^{\beta}(n_k)) = (f(n_k)) \to f^{\beta}(p) \in \{0,1\}$, so $(f(n_k))$ must be eventually constant – which is false.

Therefore $\beta\mathbb{N}$ is an example showing that "compact $\not\Rightarrow$ sequentially compact." (*See the remarks before and after corollary VIII.8.5.*)

Theorem 6.6 Every infinite closed set F in $\beta\mathbb{N}$ contains a copy of $\beta\mathbb{N}$ and therefore satisfies $|F| = 2^c$.

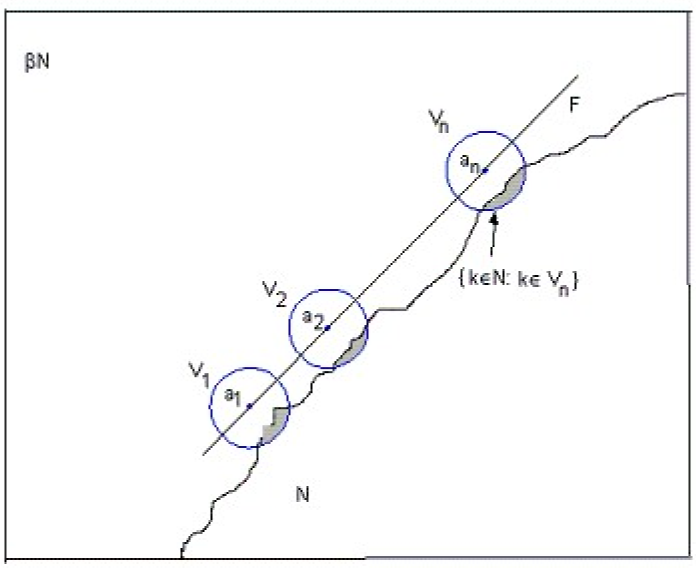
Proof Pick an infinite discrete set $A = \{a_n : n = 1, 2, ...\} \subseteq F$. (*See Exercise III E9*). Using regularity, pick pairwise disjoint open sets V_n in $\beta\mathbb{N}$ with $a_n \in V_n$.

Suppose $g : A \to [0, 1]$ (g is continuous since A is discrete). Define $G : \mathbb{N} \to [0, 1]$ by

$$G(k) = \begin{cases} g(a_n) & \text{for } k \in \mathbb{N} \cap V_n \\ 0 & \text{for } k \in \mathbb{N} - \bigcup_{n=1}^{\infty} V_n \end{cases}$$

Extend G to a continuous map $G^\beta : \beta\mathbb{N} \to [0, 1]$.

The following diagram gives a very "distorted" image of how the sets in the argument are related.



We have $G^\beta | \mathbb{N} \cap V_n = g(a_n)$. Since $\mathbb{N} \cap V_n$ is dense in V_n (*why?*), we have $G^\beta | V_n = g(a_n)$ so $G^\beta | A = g$.

Thus, $g : A \to [0, 1]$ has an extension $G^\beta : \beta\mathbb{N} \to [0, 1]$, so A is C^*-embedded in $\beta\mathbb{N}$. By Theorem 6.3, $\text{cl}_{\beta\mathbb{N}} A = \beta A$ and since A is a countably infinite discrete space, βA is homeomorphic to $\beta\mathbb{N}$.

Since F is closed, $\text{cl}_{\beta\mathbb{N}} A = \beta A \subseteq F$, so $|F| = 2^c$. ∙

Theorem 6.6 illustrates a curious property of $\beta\mathbb{N}$: there is a "gap" in the sizes of closed subsets. That is, every closed set in $\beta\mathbb{N}$ is either finite or has cardinality 2^c – no sizes in-between! This "gap in the possible sizes of closed subsets" can sometimes occur, however, even in spaces as nice as metric spaces – although not if the Generalized Continuum Hypothesis is assumed. (See A.H. Stone, *Cardinals of Closed Sets*, Mathematika 6 (1959), pp. 99-107.)

Example 6.7 $\beta\mathbb{N}$ is separable, but its subspace $\beta\mathbb{N}-\mathbb{N}$ is not; $\beta\mathbb{N}-\mathbb{N}$ does not even satisfy the weaker countable chain condition CCC (see Definition VIII.11.4). Specifically, we will show that $\beta\mathbb{N}-\mathbb{N}$ contains c pairwise disjoint clopen (in $\beta\mathbb{N}-\mathbb{N}$) subsets, each of which is homeomorphic to $\beta\mathbb{N}-\mathbb{N}$.

Let $\{N_t : t \in [0,1]\}$ be a collection of c infinite subsets of $\mathbb{N}$ with the property that any two have finite intersection. (*See Exercise I.E41.*) Let $U_t = (\beta\mathbb{N}-\mathbb{N}) \cap \mathrm{cl}_{\beta\mathbb{N}} N_t = \mathrm{cl}_{\beta\mathbb{N}} N_t - N_t$. Each $U_t \neq \emptyset$ (*why?*) and U_t is a clopen set in $\beta\mathbb{N}-\mathbb{N}$ homeomorphic to $\beta\mathbb{N}-\mathbb{N}$.

Moreover, the U_t's are disjoint:

> Suppose $t \neq t'$. If $z \in U_t \cap U_{t'}$, then $z \in \mathrm{cl}_{\beta\mathbb{N}} N_t \cap \mathrm{cl}_{\beta\mathbb{N}} N_{t'}$. In a T_1 space, deleting finitely many points from an infinite set A does not change the set $\mathrm{cl}\, A - A$ (*why?*), so $z \in \mathrm{cl}_{\beta\mathbb{N}}(N_t - (N_t \cap N_{t'}))$ and $z \in \mathrm{cl}_{\beta\mathbb{N}}(N_{t'} - (N_t \cap N_{t'}))$. But $N_t - (N_t \cap N_{t'})$ and $N_{t'} - (N_t \cap N_{t'})$ are disjoint zero sets in $\mathbb{N}$ and must have disjoint closures.

An additional tangential observation:

If we choose points $x_t \in U_t$ and let $X = \mathbb{N} \cup \{x_t : t \in [0,1]\}$, then X is not normal – since a separable normal space cannot have a closed discrete subset $\{x_t : t \in [0,1]\}$ of cardinality c. (See the "counting continuous functions" argument in Example VII.5.6.)

The following example shows us that countable compactness and pseudocompactness are not even finitely productive.

Example 6.8 There is a countably compact space X for which $X \times X$ is not pseudocompact (so $X \times X$ is also not countably compact).

Let $\mathbb{E} = \{2, 4, 6, ...\}$ and $\mathbb{O} = \{1, 3, 5, ...\}$ and write $\beta\mathbb{N} = \mathrm{cl}_{\beta\mathbb{N}}\mathbb{E} \cup \mathrm{cl}_{\beta\mathbb{N}}\mathbb{O} = \beta\mathbb{E} \cup \beta\mathbb{O}$. $\beta\mathbb{E}$ and $\beta\mathbb{O}$ are disjoint clopen copies of $\beta\mathbb{N}$. Choose any homeomorphism $f : \beta\mathbb{E} \longrightarrow \beta\mathbb{O}$ (n*ecessarily, $f[\mathbb{E}] = \mathbb{O}$: why?*) and define $g : \beta\mathbb{N} \to \beta\mathbb{N}$ by $g(x) = \begin{cases} f(x) & \text{if } x \in \beta\mathbb{E} \\ f^{-1}(x) & \text{if } x \in \beta\mathbb{O} \end{cases}$.

The map g is a homeomorphism since g and g^{-1} are continuous on the two disjoint clopen sets $\beta\mathbb{E}$ and $\beta\mathbb{O}$ whose union is $\beta\mathbb{N}$. Clearly, $g|\mathbb{N} : \mathbb{N} \to \mathbb{N}$, g has no fixed points, and $g \circ g$ is the identity map.

Let $\mathcal{C} = \{A \subseteq \beta\mathbb{N} : A \text{ is countably infinite}\}$. $|\,\mathcal{C}\,| = (2^c)^{\aleph_0} = 2^c$. Let λ be the first ordinal with cardinality 2^c and index $\mathcal{C}$ as $\{A_\alpha : \alpha < \lambda\}$. For each α, $|\mathrm{cl}_{\beta\mathbb{N}} A_\alpha|$ is an infinite closed set so, by Theorem 6.6, $|\mathrm{cl}_{\beta\mathbb{N}} A_\alpha| = 2^c$. Therefore $\mathrm{cl}_{\beta\mathbb{N}} A_\alpha - A_\alpha \neq \emptyset$.

Pick p_0 to be a limit point of A_0 not in A_0. Proceeding inductively, assume that for all $\alpha < \beta < \lambda$ we have chosen a limit point p_α of A_α that is not in A_α and that, for the points p_α, p_γ $(\alpha < \gamma < \beta)$ already defined :

$$\begin{cases} p_\alpha \neq p_\gamma \\ p_\alpha \neq g(p_\gamma) \\ p_\gamma \neq g(p_\alpha) \end{cases} \qquad (*)$$

For the "next step", we want to define p_β. Since $|[0,\beta)| < 2^c$, we have so far defined fewer than 2^c points p_α. Therefore

$$|\{p_\alpha : \alpha < \beta\} \cup \{g(p_\alpha) : \alpha < \beta\} \cup \{g^{-1}(p_\alpha) : \alpha < \beta\}| < 2^c.$$

But $|\mathrm{cl}_{\beta\mathbb{N}} A_\beta - A_\beta| = 2^c$, so we can chose a limit point p_β of A_β with $p_\beta \notin A_\beta$ so that the conditions $(*)$ continue to hold for $\alpha < \gamma < \beta + 1$.

Therefore, by transfinite recursion, we have defined distinct points p_α $(\alpha < \lambda)$ in such a way that for $\alpha \neq \beta < \lambda$, $g(p_\alpha) \neq p_\beta$ and $g(p_\beta) \neq p_\alpha$.

Let $X = \mathbb{N} \cup \{p_\alpha : \alpha < \lambda\}$. By construction, X is countably compact because every infinite set in X (for that matter, even every infinite set in $\beta\mathbb{N}$) has a limit point in X. But we claim that $X \times X$ is not pseudocompact.

To see this, consider $Z = \{(n, g(n) : n \in \mathbb{N}\} \subseteq X \times X$. We claim Z is clopen in $X \times X$.

Since $(n, g(n))$ is isolated in $X \times X$, Z is a discrete open subset of $X \times X$.

On the other hand, the graph of $g = \{(x, g(x)) : x \in \beta\mathbb{N}\}$ is closed in $\beta\mathbb{N} \times \beta\mathbb{N}$ so that

$\{(x, g(x)) : x \in \beta\mathbb{N}\} \cap (X \times X)$ is closed in $X \times X$.

and we claim that $\{(x, g(x) : x \in \beta\mathbb{N}\} \cap (X \times X) = Z$.

Indeed, it is clear that

$$Z \subseteq \{(x, g(x) : x \in \beta\mathbb{N}\} \cap (X \times X)$$

and the complicated construction of the p_α's was done precisely to guarantee the reverse inclusion:

If $(x, g(x)) \in X \times X$, then $x \in \mathbb{N}$ – for otherwise we would have $x = p_\alpha$ for some α, and then $g(x) = g(p_\alpha) \notin X$ by construction.

Therefore Z is closed in $X \times X$.

Therefore function $h : X \times X \to \mathbb{N}$ defined by

$$h(u) = \begin{cases} n & \text{if } u = (n, g(n)) \in Z \\ 0 & \text{if } u \in (X \times X) - Z \end{cases}$$

continuous. But h is unbounded, so $X \times X$ is not pseudocompact.

7. Alternate Constructions of βX

We constructed βX by defining an order $\geq$ between certain compactifications of X and showing that there must exist a largest compactification (unique up to equivalence) in this ordering. Theorem 5.7, however, shows that there are many different characterizations of βX and some of these characterizations suggest other ways to construct βX. .

For example, Theorem 5.7 shows that the zero sets in a Tychonoff space X play a special role in βX. Without going into the details, one can construct βX as follows:

Let $\mathcal{Z}$ be the collection of zero sets in X. A filter $\mathcal{F}$ in $\mathcal{Z}$ (also called a z-filter) means a nonempty collection of nonempty zero sets such that

i) if $F_1, F_2 \in \mathcal{F}$, then $F_1 \cap F_2 \in \mathcal{F}$, and
ii) if $F \in \mathcal{F}$ and $G \supseteq F$ where G is a zero set, then $G \in \mathcal{F}$.

A z-ultrafilter in X is a maximal z-filter.

Define a set $\beta X = \{\mathcal{U} : \mathcal{U}$ is a z-ultrafilter in $X\}$. For each $p \in X$, the collection $\mathcal{U}_p = \{Z : Z$ is a zero set containing $p\}$ is a (trivial) z-ultrafilter, so $\mathcal{U}_p \in \beta X$. The map $h(p) = \mathcal{U}_p$ is a $1-1$ map of X into the set βX.

It turns out that X compact iff every z-ultrafilter is of the form $\mathcal{U}_p$ for some $p \in X$. Therefore the set $\beta X - X = \emptyset$ iff X is compact. Each z-ultrafilter $\mathcal{U}$ in X that is not of the trivial form $\mathcal{U}_p$ is a point in $\beta X - X$.

The details of putting a topology on βX to create the largest compactification of X are a bit tricky and we will not go into them here.

The situation is simpler in the case $X = \mathbb{N}$. Since every subset of $\mathbb{N}$ is a zero set, a "z-ultrafilter" in $\mathbb{N}$ is just an ordinary ultrafilter in $\mathbb{N}$.

Then, to be a bit more specific,

let $\beta\mathbb{N} = \{\mathcal{U} : \mathcal{U}$ is an ultrafilter in $\mathbb{N}\}$ and for $A \subseteq \mathbb{N}$, define
$\text{cl}\, A = \{\mathcal{U} : A \in \mathcal{U}\}$

Give $\beta\mathbb{N}$ the topology for which $\{\text{cl}\, A : A \subseteq \mathbb{N}\}$ is a base for the open sets.

This topology makes $\beta\mathbb{N}$ into a compact T_2 and we can embed $\mathbb{N}$ into $\beta\mathbb{N}$ using the mapping $h(n) = \mathcal{U}_n$ ($=$ the trivial ultrafilter "fixed" at n). This "copy" of $\mathbb{N}$ is dense in $\beta\mathbb{N}$, so $\beta\mathbb{N}$ is a compactification of $\mathbb{N}$. It can be shown that "this $\beta\mathbb{N}$" is the largest compactification of $\mathbb{N}$ (and therefore equivalent to the $\beta\mathbb{N}$ constructed earlier).

The free ultrafilters in $\mathbb{N}$ are the points in $\beta\mathbb{N} - \mathbb{N}$. Since $|\beta\mathbb{N}| = 2^c$ and there are only countably many trivial ultrafilters $\mathcal{U}_n$, we conclude that there are 2^c free ultrafilters in $\mathbb{N}$

It turns out that the z-ultrafilters in a Tychonoff space X are associated in a natural $1-1$ way with the maximal ideals of the ring $C(X)$, so it is also possible to construct βX by putting an appropriate topology on the set

$$\beta X = \{M : M \text{ is a maximal ideal in } C(X)\}$$

It turns out that if $p \in X$, then $M_p = \{f \in C(X) \; : f(p) = 0\}$ is a (trivial) maximal ideal and the mapping $h(p) = M_p$ gives a natural way to embed X in βX. X is not compact iff there are maximal ideals in $C(X)$ that are not of the form M_p (that is, nontrivial maximal ideals) and these are the points of $\beta X - X$.

More information about these constructions can be found in the beautifully written classic *Rings of Continuous Functions* (Gillman & Jerison).

In this section, we give one alternate construction of βX in detail. It is essentially the construction used by Tychonoff, who was the first to construct βX for arbitrary Tychonoff spaces. In his paper *Über die topologische Erweiterung von Räumen* (Math. Annalen 102(1930), 544-561) Tychonoff also established the notation "βX." The construction involves a specially chosen embedding of X into a cube.

Suppose X is a Tychonoff space. For each $f \in C^*(X)$, choose a closed interval $I_f \subseteq \mathbb{R}$ such that $\operatorname{ran}(f) \subseteq I_f$. If $\mathcal{F} \subseteq C^*(X)$ is a family that $\mathcal{F}$ separates points from closed sets, then according to Theorem VI.4.10 the evaluation map $e_{\mathcal{F}} : X \to \prod_{f \in \mathcal{F}} I_f$ given by $e_{\mathcal{F}}(x) = f(x)$ is an embedding. In this way, every such family $\mathcal{F} \subseteq C^*(X)$ generates a compactification $(\operatorname{cl} e_{\mathcal{F}}[X], e_{\mathcal{F}})$ of X. In fact, the following theorem states that every compactification of X can be obtained by choosing the correct family $\mathcal{F} \subseteq C^*(X)$.

Theorem 7.1 Every compactification of X can be achieved using the construction in the preceding paragraph. More precisely, if Y is a compactification containing X (with embedding i), then there exists a family $\mathcal{F} \subseteq C^*(X)$ such that $\mathcal{F}$ separates points and closed sets and $(\operatorname{cl} e_{\mathcal{F}}[X], e_{\mathcal{F}}) \simeq (Y, i)$.

Proof Let $\mathcal{F} = \{f \in C^*(X) : f$ can be continuously extended to $\widetilde{f} \; : Y \to I_f\}$. (*Note that* $\widetilde{f}$ *is unique if it exists – since any two extensions would agree on the dense set* X.)

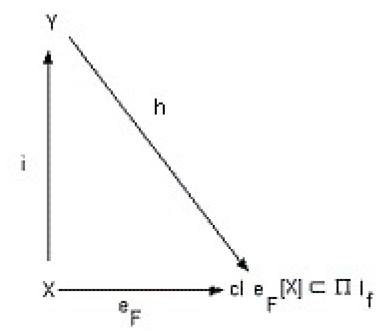
The family $\mathcal{F}$ separates points from closed sets:

If F is a closed set in X and $x \notin F$, then there is a closed set $K \subseteq Y$ with $x \notin K \cap X = F$. By complete regularity there is a continuous function $g : Y \to [0,1]$ such that $g(x) = 0$ and $g|K = 1$. Since Y is compact, g must be bounded and therefore $f = g|X \in C^*(X)$. Moreover, $f \in \mathcal{F}$ (because g is the required extension). Clearly, $f(x) = g(x) = 0 \notin \operatorname{cl} f[F] \subseteq \operatorname{cl} g[K] = \{1\}$.

Therefore $(\operatorname{cl} e_{\mathcal{F}}[X], e_{\mathcal{F}})$ is a compactification of X.

Define $h : Y \to \prod_{f \in \mathcal{F}} I_f$ by $h(p)(f) = \widetilde{f}\ (p)$. Then h is continuous and, for $x \in X$, $h(x)(f) = \widetilde{f}\ (x) = f(x) = e_{\mathcal{F}}(x)(f)$. Therefore $h[X] = e_{\mathcal{F}}[X]$ and $h \circ i = e_{\mathcal{F}}$.

Clearly, $e_{\mathcal{F}}[X] = h[X] \subseteq h[Y]$ and $h[Y]$ is compact Hausdorff, so $\text{cl}\, e_{\mathcal{F}}[X] \subseteq h[Y]$. On the other hand, by continuity, $h[Y] = h[\text{cl}\, X] \subseteq \text{cl}\, h[X] = \text{cl}\, e_{\mathcal{F}}[X]$. Therefore $h[Y] = \text{cl}\, e_{\mathcal{F}}[X]$.



Since $h : Y \to \text{cl}\, e_{\mathcal{F}}[X]$ is continuous and onto, $(Y, i) \geq ((\text{cl}\, e_{\mathcal{F}}[X], e_{\mathcal{F}}))$.

We claim h is also $1-1$:

> If $p \neq q \in Y$, then there is a continuous map $g : Y \to [0,1]$ such that $g(p) = 0$ and $g(q) = 1$. Then $f = g|X \in \mathcal{F}$ and $\widetilde{f}\ (p) = g(p) \neq g(q) = \widetilde{f}\ (q)$. Therefore $h(p)(f) \neq h(q)(f)$, so $h(p) \neq h(q)$.

Since Y is compact and $\text{cl}\, e_{\mathcal{F}}[X]$ is Hausdorff, h is a homeomorphism and, as mentioned above, $h \circ i = e_{\mathcal{F}}$. Therefore $(Y, i) \simeq ((\text{cl}\, e_{\mathcal{F}}[X], e_{\mathcal{F}})$.

Theorem 7.2 Suppose $\mathcal{F} \subseteq \mathcal{F}' \subseteq C^*(X)$ and that both $\mathcal{F}$ and $\mathcal{F}'$ separate points from closed sets. Then $(\mathrm{cl}\, e_{\mathcal{F}'}[X], e_{\mathcal{F}'}) \geq (\mathrm{cl}\, e_{\mathcal{F}}[X], e_{\mathcal{F}})$.

$$\begin{tikzcd}
\mathrm{cl}\, e_{F'}[X] \subset \prod_{f \in F'} I_f \arrow[dr, "h"] & \\
X \arrow[u] \arrow[r] & \mathrm{cl}\, e_F[X] \subset \prod_{f \in F} I_f
\end{tikzcd}$$

For $p \in \mathrm{cl}\, e_{\mathcal{F}'}[X]$, define $h(p) \in \mathrm{cl}\, e_{\mathcal{F}}[X]$ by $h(p)(f) = p(f) = p_f$. (*Informally,* $h(p)$ *is just the result of deleting from* p *all the coordinates corresponding to functions in* $\mathcal{F}' - \mathcal{F}$.) Clearly $e_{\mathcal{F}} = h \circ e_{\mathcal{F}'}$ so $(\mathrm{cl}\, e_{\mathcal{F}'}[X], e_{\mathcal{F}'}) \geq (\mathrm{cl}\, e_{\mathcal{F}}[X], e_{\mathcal{F}})$. •

Corollary 7.3 A Tychonoff space X has a largest compactification.

Proof Combining Theorems 7.1 and 7.2, we see that the largest compactification corresponds to taking $\mathcal{F} = C^*(X)$ in the preceding construction. •

Of course we can do the construction (from the paragraph preceding Theorem 7.1) simply using $\mathcal{F} = C^*(X)$ *in the first place (that is what Tychonoff did) and define the resulting compactification to be* βX. *We would then need to prove that it has one of the features that make it interesting – for example, the Stone Extension Property. Instead, using Theorems 7.1 and 7.2, what we did was first to argue that* $\mathcal{F} = C^*(C)$ *produces the largest compactification of* X*; then Theorem 5.7 told us that the compactification we constructed is the same as our earlier* βX.

Exercises

E1. Show that the Sorgenfrey line (Example III.5.3) is not locally compact.

E2. Suppose X is a locally compact T_2 space that is separable and not compact. Show that the one-point compactification X^* is metrizable.

E3. Suppose C and K are disjoint compact subsets in a locally compact Hausdorff space X. Prove that there exist disjoint open sets $U \supseteq C$ and $V \supseteq K$ such that $\operatorname{cl} U$ and $\operatorname{cl} V$ are compact.

E4. a) Let K be a compact subspace of a Tychonoff space X. Prove that for each $g \in C(K)$ there is an $f \in C(X)$ that $g = f|K$ – that is, every continuous real valued function on K can be extended to X. (*A subspace of* X *with this property is said to be* C*-embedded in* X. *Compare Definition 6.2; for a compact since* K *is compact, "*C*-embedded" and "*C^**-embedded" mean the same thing.*)

b) Suppose A is a dense C-embedded subspace of a Tychonoff space X. If $f \in C(X)$ and $f(x) = 0$ for some $x \in X$, prove that $f(a) = 0$ for some $a \in A$. *Hint: if* $f|\,A$ *is never* 0*, then* $\frac{1}{f} \in C(A)$

c) Every bounded function $f : \mathbb{N} \to \mathbb{R}$ has a continuous extension $f^\beta : \beta\mathbb{N} \to \mathbb{N}$. In particular, the function $f(n) = \frac{1}{n}$ can be extended. If $p \in \beta\mathbb{N} - \mathbb{N}$, what is $f^\beta(p)$? Why does this not contradict part b) ?

E5. Prove that $|\beta\mathbb{R}| = |\beta\mathbb{Q}| = 2^c$.

E6. Prove that a Tychonoff space X is connected iff βX is connected. Is it true that X is connected iff every compactification of X is connected?

E7. a) Show that $\beta\mathbb{R} - \mathbb{R}$ has two components A and B.

b) $[0, \infty)$ has a limit point in $\beta\mathbb{R} - \mathbb{R}$, say in the set B. Is $\beta[0, 1) = B$?

E8. Let $\mathcal{U}$ be a free ultrafilter in $\mathbb{N}$.

a) Choose a point $\sigma \in \beta\mathbb{N} - \mathbb{N}$ and let $\mathcal{U} = \{A \subseteq \mathbb{N} : \sigma \in \operatorname{cl}_{\beta\mathbb{N}} A\}$. Show that $\mathcal{U}$ is a free ultrafilter on $\mathbb{N}$.

b) Using the ultrafilter $\mathcal{U}$ from a), construct the space Σ as in Exercise IX.E8. Prove that

Σ is homeomorphic to $\mathbb{N} \cup \{\sigma\}$ with the subspace topology from $\beta\mathbb{N}$.

c) Define an equivalence relation on $\beta\mathbb{N} - \mathbb{N}$ by $x \sim y$ if $\mathbb{N} \cup \{x\}$ is homeomorphic to $\mathbb{N} \cup \{y\}$. For $x \in \beta\mathbb{N} - \mathbb{N}$, let $[x]$ be the equivalence class of x. Prove that each equivalence class satisfies $|[x]| \leq c$ (so there must be 2^c different equivalence classes.)

Note: Part c) says that, in some sense, there are 2^c topologically different points $\sigma \in \beta\mathbb{N} - \mathbb{N}$. By part a), each of these points σ is associated with a free ultrafilter $\mathcal{U}$ in $\mathbb{N}$ that determines the topology on $\mathbb{N} \cup \{\sigma\}$. Therefore there are 2^c "essentially different" free ultrafilters $\mathcal{U}$ in $\mathbb{N}$.

Chapter X Review

Explain why each statement is true, or provide a counterexample.

1. Every Tychonoff space has a one-point compactification.

2. If X is Tychonoff and βX is first countable, then $|\beta X| \leq c$.

3. $\mathbb{R}$ has a compactification of cardinal 2^{2^c}.

4. $\mathbb{R}$ has a compactification $Y \supseteq \mathbb{R}$ where $Y - \mathbb{R}$ is infinite and Y is metrizable.

5. Suppose that X is a compact Hausdorff space and that each $x \in X$ has a metrizable neighborhood (i.e., X is *locally metrizable*). Then X is metrizable.

6. Let $\mathbb{N}^*$ be the 1-point compactification of $\mathbb{N}$. Every subset of $\mathbb{N}^*$ is Borel.

7. $\beta\mathbb{N} - \mathbb{N}$ is dense in $\beta\mathbb{N}$.

8. If $X = [0, \omega_0 + \omega_0)$, then $\beta X = [0, \omega_0 + \omega_0]$.

9. Every point in $\beta\mathbb{N}$ is the limit of a sequence from $\mathbb{N}$.

10. The one-point compactification of $\mathbb{R}$ is completely metrizable.

11. If X and Y are locally compact Hausdorff spaces with homeomorphic one-point compactifications, then X must be homeomorphic to Y.

12. Let $n \in \mathbb{N}$. All n-point compactifications of the Tychonoff space X are equivalent.

13. Every subset of $\mathbb{R}$ is C^*-embedded in $\mathbb{R}$.

14. If X is compact Hausdorff and $a \in X$, then $\beta(X - \{a\}) = X$.

15. Every compact Hausdorff space is separable.

16. A metric space (X, d) has a metrizable compactification iff X is separable.

17. $\mathbb{Q} = U \cap F$ for some open U and closed F in $\mathbb{R}$.