I. Foundational Material

Since this is a graduate course in mathematics, it is appropriate to assume some background material as well as some experience in working with it. In particular, we shall presume that the reader has a working knowledge of basic set theory; however, familiarity with a fully developed axiomatic approach to that subject will not be needed. For the sake of completeness, the following files on foundational material have been placed into the course directory:

> `foundations1.*` (where $* =$ `ps` or `pdf`) **:** This is a review of mathematical logic and simple set-theoretic algebra.
>
> `foundations2.*` (where $* =$ `ps` or `pdf`) **:** This discusses some material in set theory that either deviates slightly from the treatment in Munkres or is not covered there. Also included are discussions of the Axiom of Choice and the Generalized Continuum Hypothesis.
>
> `realnumbers.*` (where $* =$ `ps` or `pdf`) **:** This gives a list of axioms for the real numbers and some important properties, including those that are play a major role in this course.
>
> `uniqreals.*` (where $* =$ `ps` or `pdf`) **:** This proves that the axioms for the real numbers completely characterize the latter up to a structure preserving one-to-one correspondence (*i.e*,, an **isomorphism**). One needs to prove a result of this sort in order to talk about **THE** real number system. The uniqueness proof is not particularly difficult but it is rather tedious; given the importance of uniqueness, it is worthwhile to look through this argument sometime and to understand it at least passively.
>
> `categories.*` (where $* =$ `ps` or `pdf`) **:** Category theory will not be used explicitly in this course, but given its fundamental nature a discussion of the basic ideas and some examples is included for reference.

Most of the discussion of foundational material here will concentrate on points that are not worked out in detail in Munkres but are important for the course.

I.1 : Basic set theory

(Munkres, §§ 1, 2, 3)

As indicated on page 3 of Munkres, our approach to questions in logic and set theory is to assume some familiarity with the most elementary ideas and to discuss what else is needed without spending so much time on these subjects that important material in topology must be omitted.

Munkres mentions that an overly casual approach to set theory can lead to logical paradoxes. For example, this happens if we try to consider the "set of all sets." During the early part of the twentieth century mathematicians realized that problems with such things could be avoided by stipulating that sets cannot be "too large," and effective safeguards to eliminate such difficulties were built into the formal axioms for set theory. One simple and reliable way of avoiding such problems with the informal approach to set theory in this course this is to assume that all constructions take place in some extremely large set that is viewed as universal. This is consistent with the formal axiomatic approach, where one handles the problem by considering two types of collections of objects: The **CLASSES** can be fairly arbitrary, but the **SETS** are constrained by a simple logical condition (specifically, they need to belong to *some* other class). Viewing everything as contained in some very large set is the recommended option for this course if difficulties ever arise.

I.2 : Products, relations and functions

(Munkres, §§ 5, 6, 8)

The main differences between this course and Munkres are outlined in the first few pages of `foundations2.*` .

Definitions relations and functions

Since functions play such an important role in topology, it is appropriate to stress the differences between the definition of functions in this course and the definitions that appear in many mathematical textbooks.

Our definition of a function is equivalent to that of Munkres, but the crucial difference between the definitions of relations and functions in Munkres and this course is that the source set (or *domain*) and the target set (or *codomain*) are explicit pieces of data in the definition of a function. Generally the source or domain is unambiguous because usually (but not always!) it is the set on which the function is defined. However, the target or codomain is often ignored. The target must contain the image of the function, but for each set containing the image there is a different function whose target is the given set. For example, this means that we distinguish between the functions $f : \mathbf{R} \to \mathbf{R}$ and $g : \mathbf{R} \to [0, 1]$ defined by the same formula:

$$f(x) = g(x) = \frac{1}{x^2 + 1}$$

One particular context in which it is necessary to make such distinctions is the construction of the fundamental group as in Chapter 9 of Munkres.

Comments on terminology

Mathematicians have two alternate lists of words for discussing functions that are 1–1 and onto, and for set theory and many other branches of mathematics these are used interchangeably. One of these lists is given on page 18 of Munkres: A map if *injective* if it is 1–1, *surjective* if it is onto, and *bijective* if it is both (*i.e.*, a 1–1 correspondence). The terms *monomorphism* and *epimorphism* from category theory are often also used in set theory as synonyms for 1–1 and onto. However, some care is needed when using these category-theoretic terms for functions in discussing morphisms of topological objects, so we shall (try to) avoid using such terminology here.

Disjoint unions

We shall also use an elementary set-theoretic construction that does not appear in Munkres; namely, given two sets A and B we need to have a *disjoint union*, written $A \sqcup B$ or $A \coprod B$, which is a union of two disjoint subsets that. are essentially xerox copies of A and B. One use of this construction will appear in the discussion of cardinal numbers; other applications will be discussed in the main body of the course (in particular, see the section, "Sums and cutting and pasting").

I.3 : Cardinal numbers

(Munkres, §§ 4, 7, 9)

The theory of cardinal numbers for infinite sets illustrates that set theory is not just a more sophisticated approach to well-known mathematical concepts and that it is also a powerful tool for obtaining important new insights into mathematics.

Definition. Two sets A and B have the same cardinality (or cardinal number) if there is a 1–1 onto map (or 1–1 correspondence) $f : A \to B$. Frequently we write $|A| = |B|$ in this case; it follows immediately that this relation is reflexive, symmetric and transitive.

We begin with two basic properties of infinite sets:

PROPOSITION. *Every infinite set contains a subset that is in* $1-1$ *correspondence with the positive integers.*■

GALILEO'S PARADOX. *A set* A *is infinite if and only if there is a* $1-1$ *correspondence between* A *and a proper subset of* A.■

Partial ordering of cardinalities

Definition. If A and B are sets, we write $|A| \leq |B|$ if there is a 1–1 map from A to B.

It follows immediately that this relation is transitive and reflexive, but the proof that it is symmetric is decidedly nontrivial:

SCHRÖDER-BERNSTEIN THEOREM. *If* A *and* B *are sets such that there are* $1-1$ *maps* $A \to B$ *and* $B \to A$*, then* $|A| = |B|$.

Sketch of proof. This is the classic argument from Birkhoff and MacLane's *Survey of Modern Algebra* (see page 340 in the Third Edition).

Let $f : A \to B$ and $g : B \to A$ be the 1–1 mappings. Each $a \in A$ is the image of at most one *parent* element $b \in B$; in turn, the latter (if it exists) has at most one parent element in A, and so on. The idea is to trace back the ancestry of each element as far as possible. For each point in A or B there are exactly three possibilities; namely, the ancestral chain may go back forever, it may end in A, or it may end in B.

Split A and B into three pieces corresponding to these cases, and call the pieces A_1, A_2, A_3 and B_1, B_2, B_3 (the possibilities are ordered as in the previous paragraph).

The map f defines a 1–1 correspondence between A_1 and B_1 (and likewise for g). Furthermore, g defines a 1–1 correspondence from B_2 to A_2, and f defines a 1–1 correspondence from B_3 to A_3. If we combine these 1–1 correspondences $A_1 \longleftrightarrow B_1$, $A_2 \longleftrightarrow B_2$ and $A_3 \longleftrightarrow B_3$, we obtain a 1–1 correspondence between all of A and all of B.■

If we are only considering subsets of some large set, then we can define the "cardinal number" of a subset to the the equivalence class of that set under the relation $A \sim B \leftrightarrow |A| = |B|$. If one assumes the Axiom of Choice it is possible to give an equivalent definition of cardinal number that extends to all sets, but for the time being we shall not worry about this point.

The Schröder-Bernstein Theorem and the preceding observations imply that *if A is an infinite subset of the positive integers $\mathbf{N}^+$, then $|A| = |\mathbf{N}^+|$*. Since we also know that $|A| \leq |\mathbf{N}^+|$ for every infinite set A, it follows that $|\mathbf{N}^+|$ can be viewed as the unique smallest infinite cardinal number and that it is $\leq$ every other infinite cardinal number. Following Cantor's notation this cardinal number is generally denoted by $\aleph_0$ (aleph-null).

In contrast, Cantor's Diagonal Process technique implies that in effect there is no largest infinite cardinal number.

THEOREM. *If A is a set and $\mathbf{P}(A)$ is the set of all subsets of A then $|A| < |\mathbf{P}(A)|$.*

Sketch of proof. The first step is to notice that $\mathbf{P}(A)$ is in 1–1 correspondence with the set of all set-theoretic functions $A \to \{0, 1\}$. Given a subset B let χ_B be the *characteristic function of* B that is 1 on B and 0 on $A - B$. Conversely given a function j of the type described, if we set $B = j^{-1}(\{1\})$ then $j = \chi_B$.

The map $A \to \mathbf{P}(A)$ sending a to $\{a\}$ is a 1–1 mapping, and therefore $|A| \leq |\mathbf{P}(A)|$. Therefore it is only necessary to prove that the cardinalities are unequal. The Diagonal Process argument is worked out in the proof of Theorem 7.7 on pages 49–50 of Munkres in the special case where $A = \mathbf{N}^+$, and the same argument works for an arbitrary nonempty set.■

Notational conventions. If one assumes the Axiom of Choice then one can show that the class of all cardinal numbers is linearly ordered, and in fact it is well-ordered (every nonempty subcollection has a least element). Of course, the minimum element of the class of infinite cardinals is $\aleph_0$, and one proceeds similarly to define $\aleph_1$ to be the least infinite cardinal in the complement of $\{\aleph_0\}$, $\aleph_2$ to be the least infinite cardinal in the complement of $\{\aleph_0, \aleph_1\}$, and similarly one can define $\aleph_n$ recursively for each positive integer n.

We now claim that there is a set S whose cardinality is greater than $\aleph_n$ for all n. Choose A_n such that $|A_n| = \aleph_n$ and consider $S = \cup_n A_n$. By construction we must have $|S| > |A_k|$ for all k. Therefore one can then define $\aleph_\omega$ to be the first infinite cardinal not in $\{\aleph_0, \aleph_1, \aleph_2, \cdots\}$. Similar considerations show that for any **SET** of cardinal numbers one can always find a larger one.

Cardinal arithmetic

One can perform a limited number of arithmetic operations with cardinal numbers, but it is necessary to realize that these do not enjoy all the familiar properties of the corresponding operations on positive integers.

Definition. If A and B are sets, then

(*i*) the **sum** $|A| + |B|$ is equal to $|A \sqcup B|$,

(*ii*) the **product** $|A| \cdot |B|$ is equal to $|A \times B|$,

(*iii*) the **exponentiation** $|A|^{|B|}$ is equal to $|\mathbf{F}(B, A)|$, the set of all set-theoretic functions from B to A.

If A and B are finite, the first two are true by simple counting arguments, and the third is true because the set of functions is in 1–1 correspondence with a product of $|B|$ copies of A. Note that the Diagonal Process argument above shows that $|\mathbf{P}(A)| = 2^{|A|}$.

The following simple result illustrates the difference between finite and infinite cardinals:

PROPOSITION. *If A is finite and B is infinite, then $|A| + |B| = |B|$.*

Sketch of proof. Let $C \subset B$ be a subset in 1–1 correspondence with $\mathbf{N}^+$. Clearly

$$|A \sqcup C| = |A \sqcup \mathbf{N}^+| = |\mathbf{N}^+| = |C|$$

and one can use this to construct a 1–1 correspondence between B and $A \sqcup B$.■

The following standard identities involving $\aleph_0$ were first noted by Galileo and Cantor respectively.

THEOREM. *We have* $\aleph_0 + \aleph_0 = \aleph_0$ *and* $\aleph_0 \cdot \aleph_0 = \aleph_0$.

The first of these is easily seen by splitting $\mathbf{N}^+$ into the even and odd positive integers, while the traditional diagonal argument for proving the latter is worked out in Exercise 2 on page 45 of Munkres.■

We now have the following standard consequences.

COROLLARY. *If* $\mathbf{Z}$ *and* $\mathbf{Q}$ *are the integers and rationals respectively then* $|\mathbf{Z}| = |\mathbf{Q}| = \aleph_0$.■

These lead directly to the following fundamental result.

THEOREM.

$$|\mathbf{R}| = \left|\mathbf{P}\left(\mathbf{N}^+\right)\right| = 2^{\aleph_0}$$

Sketch of proof. Note that Munkres does not prove this result explicitly although a proof is indicated in the exercises and on page 177 the real numbers are shown to have cardinality greater than $\aleph_0$.

Usually this is derived using decimal expansions of real numbers, but we shall give a proof that does not involve decimals (although the idea is similar). The idea is to construct 1–1 maps from $\mathbf{R}$ to $\mathbf{P}\left(\mathbf{N}^+\right)$ and vice versa and then to apply the Schröder-Bernstein Theorem.

Let $D : \mathbf{R} \to \mathbf{P}(\mathbf{Q})$ be the Dedekind cut map sending a real number r to the set of all rational numbers less than r. Since there is always a rational number between any two distinct real numbers, it follows that this map is 1–1. Since $|\mathbf{Q}| = \aleph_0$ there is a 1–1 correspondence from $\mathbf{P}(\mathbf{Q})$ to $\mathbf{P}\left(\mathbf{N}^+\right)$, and the composite of D with this map gives the desired 1–1 map from $\mathbf{R}$ to $\mathbf{P}\left(\mathbf{N}^+\right)$.

Let $\mathbf{P}_\infty\left(\mathbf{N}^+\right)$ denote the set of all *infinite* subsets of $\mathbf{N}^+$, and define a function from $\mathbf{P}_\infty\left(\mathbf{N}^+\right)$ to $\mathbf{R}$ as follows: Given an infinite subset B let χ_B be its characteristic function and consider the infinite series

$$\sum_B = \sum_k \chi_B(k) \cdot 2^{-k} \ .$$

This series always converges because its terms are nonnegative and less than 1 (and the series $\sum_k 2^{-k}$ converges), and different *infinite* subsets yield different values (look at the first value of k that is in one subset but not in the other; if, say, k lies in A but not in B then we have $\sum_A > \sum_B$). Note that $\sum_A \in [0,1]$ for all infinite subsets A since $\sum_k 2^{-k} = 1$.

If A is a *finite* subset, consider the finite sum

$$\Theta_B = 2 + \sum_k \chi_B(k) \cdot 2^{-k} \ .$$

Once again it follows that different finite subsets determine different real (in fact, rational) numbers. Furthermore, since the value associated to a finite set lies in the interval $[2,3]$ it is clear that a finite set and an infinite set cannot go to the same real number. Therefore we have constructed a 1–1 function from $\mathbf{P}\left(\mathbf{N}^+\right)$ to $\mathbf{R}$.

Since we have constructed 1–1 mappings both ways, we can use the Schröder-Bernstein Theorem to complete the proof.$\blacksquare$

Natural question. The Continuum Hypothesis states that $|\mathbf{R}| = \aleph_1$; important results of P. Cohen show that one can construct models of set theory for which this statement is true and other models for which it is false. One an ask which cardinal numbers are possible values for $|\mathbf{R}|$. Results on this and more general questions of the same type can also be settled using the methods introduced by Cohen. In particular, it turns out that $|\mathbf{R}|$ can be equal to $\aleph_n$ for every positive integer n but it cannot be equal to $\aleph_\omega$ (all these are defined above). A proof of the last assertion appears in the exercises on page 66 of the book, *Set Theory and Metric Spaces*, by I. Kaplansky.

Finally, we prove the another fundamental and well known result about the cardinality of $\mathbf{R}$:

PROPOSITION. *For all positive integers n we have $|\mathbf{R}^n| = |\mathbf{R}|$.*

Using the Axiom of Choice one can show that $|A \sqcup A| = |A|$ and $|A^n| = |A|$ for every infinite set A and positive integer n, but we shall outline a direct and relatively standard argument.

Sketch of proof. There is a generalization of a familiar law of exponents for cardinal numbers

$$\gamma^{\alpha+\beta} = \gamma^\alpha \cdot \gamma^\beta$$

that follows from the 1–1 correspondence

$$\mathbf{F}(A \sqcup B, C) \to \mathbf{F}(A, C) \times \mathbf{F}(B, C)$$

which sends a function f to

$$\left(f \circ i_A, f \circ i_B\right)$$

(recall that i_A and i_B are the injections from A and B to $A \sqcup B$).

An inductive argument shows that it suffices to prove the result when $n = 2$. In this case the proof becomes a completely formal exercise involving the exponential law described above:

$$|\mathbf{R}^2| = |\mathbf{R}| \times |\mathbf{R}| = 2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0+\aleph_0} = 2^{\aleph_0}$$

COROLLARY. *We also have $2^{\aleph_0} + 2^{\aleph_0} = 2^{\aleph_0}$ and $\aleph_0 \cdot 2^{\aleph_0} = 2^{\aleph_0}$.*

Proof. These are consequences of the following chain of inequalities:

$$\begin{array}{c} 2^{\aleph_0} \quad \leq \quad 2^{\aleph_0} + 2^{\aleph_0} \quad \leq \aleph_0 \cdot 2^{\aleph_0} \quad \leq \\ 2^{\aleph_0} \cdot 2^{\aleph_0} \quad = \quad 2^{\aleph_0} \ \blacksquare \end{array}$$

Remarks. 1. The following generalizations of the usual laws of exponents also hold for cardinal numbers:

$$\gamma^{\alpha\beta} = \left(\gamma^\alpha\right)^\beta, \qquad \left(\beta \cdot \gamma\right)^\alpha = \beta^\alpha \cdot \gamma^\alpha$$

Since the proofs require some lengthy (but relatively elementary) digressions and we shall not need these results in this course, we shall not verify these relationships.

2. Another natural question about cardinal arithmetic is whether $2^\alpha = 2^\beta$ implies $\alpha = \beta$ as is the case for nonnegative integers. If the Generalized Continuum Hypothesis is true, then the answer is yes. On the other hand, this condition is not strong enough to imply the Generalized Continuum Hypothesis, and one can also construct models of set theory for which $\alpha < \beta$ but $2^\alpha = 2^\beta$. More generally, very strong results on the possible sequences of cardinal numbers that can be written as 2^α for some α are given by results of W. B. Easton that build upon P. Cohen's work on the Continuum Hypothesis; the result essentially states that are few straightforward necessary conditions on such sequences are also sufficient. These results first appeared in in the following paper by Easton: *Powers of regular cardinals*, Ann. Math Logic **1** (1970), 139–178. A more recent paper by T. Jech covers subsequent work on this problem: *Singular cardinals and the PCF theory*, Bull. Symbolic Logic **1** (1995), 408–424.

I.4 : The real number system

(Munkres, § 4)

The first Chapter of Rudin, *Principles of Mathematical Analysis* (Third Edition), contains detailed arguments for many of the results about real numbers from the files in the course directory. Aside from the usual identities involving addition, subtraction, multiplication, division and inequalities, the most crucial properties of the real numbers throughout the course are the following:

(1) **Completeness.** Every nonempty subset of $\mathbf{R}$ that has an upper bound has a **least** upper bound, and every nonempty subset of $\mathbf{R}$ that has a lower bound has a **greatest** upper bound.

(2) **Density of rational numbers.** If we are given two real numbers a and b such that $a < b$, then there is a rational number q such that $a < q < b$.

(3) If $\varepsilon > 0$ then there is a positive integer n such that $\dfrac{1}{n} < \varepsilon$.

II. Metric and topological spaces

The discussion of the main topics in this course begins here.

As noted in a classic text on general topology by W. Franz, " the word 'topology' is derived from the Greek word $\tau \acute{o} \pi o \varsigma$ which means 'place,' 'position,' or 'space' ... it is a subdiscipline of geometry" (see pp.1–3 of the book by Franz; complete bibliographic information is given at the end of these notes).

Although a detailed discussion of the history of point set topology is beyond the scope of these notes, it is useful to mention two important points that motivated the original development of the subject.

(1) Several important theorems about continuous functions of real valued functions of a (single) real variable have analogs in other contexts, and the most effective way to work with such analogs is to develop a unified approach.

(2) When one considers functions of several real variables, more thought must be given to the sets on which functions are defined. For one variable, the emphasis is on functions defined over some interval, which may or may not have end points. On the other hand, for functions of two or more real variables there is an overwhelming variety of shapes to consider (*e.g.*, round, square, triangular, hexagonal, octagonal, with and without some or all boundary points, with or without holes inside, U-shaped, X-shaped, Y-shaped, ...). Clearly there are far too many to enumerate in a relatively simple manner. Therefore, when one wants to discuss concepts like partial differentiation, it is better to begin by considering a reasonable class of **regions** or **domains** to work with. Mathematically, these conditions are given by the definition of an **open set**.

The concepts of point set topology have proven to be useful — in fact indispensable — in a wide range of mathematical contexts where it is meaningful to talk about two objects being close to each other in some algebraic, analytic or geometric sense. In particular, these concepts have played a major and foundational role in the application of geometric ideas to solve analytical questions, the interactions between the two subjects have stimulated each one to a great extent (interactions with algebra have also been mutually beneficial). A graduate course in point set topology should take the important links with algebra and analysis into account, but it also seems important to retain as much of the geometric nature of the subject as possible, particularly for a course that is the first third of a full year sequence, and striking a decent balance from a contemporary perspective is one goal of these notes.

For the sake of completeness, here are some references on the history of topology and related topics:

`www.math.uiuc.edu/~droyster/courses/fall99/math4181/classnotes/notes1.pdf`

`www-gap.dcs.st-and.ac.uk/~history/Hist/Topics/Topology_in_mathematics.html`

`www.wikipedia.org/wiki/Topology`

History of Topology, I. M. James (ed.)

Full information on this and other books cited in these notes appears in the bibliography at the end.

II.1 : Metrics and topologies

(Munkres, §§ 12, 13, 16, 20; Edwards, § I.7)

For both historical and logical reasons one can view the most basic aspects of point set theory as natural generalizations of important properties of certain subsets of the real numbers. One approach to doing this is to base the discussion on an abstract notion of **DISTANCE** that generalizes the usual notion of distance between two numbers in the obvious fashion. Another approach is to take the concept of an **OPEN SUBSET** (or something logically equivalent) as the fundamental abstract structure. It is not difficult to formulate the concept of an open set if one has a notion of distance, so there is a natural progression of abstraction from the real numbers to *metric spaces* (sets with a suitable notion of distance) to *topological spaces* (sets with a suitable notion of open subsets).

Eventually a course in point set topology needs to cover both types of structures, but there is no universal agreement on which should come first and when the other should be introduced. The approach in these notes will be to introduce metric spaces first and topological spaces immediately afterwards. This will allow us to take advantage of the strengths of both approaches throughout the course.

The basic definitions and a few examples

The notion of distance between two points in $\mathbf{R}^n$ is fundamentally important in multivariable calculus and some aspects of linear algebra. It turns out that an extremely short list of properties for distances are enough to prove abstract versions of many important results from advanced calculus and real variables courses.

Definition. A *metric space* is a pair $(X, \mathbf{d})$ consisting of a set X and a function $\mathbf{d} : X \times X \to \mathbf{R}$ (sometimes called the **metric** or **distance function**, with $\mathbf{d}(x, y)$ being called the **distance** from x to y, or between x and y) such that the following properties hold:

(MS1) $\mathbf{d}(x, y) \geq 0$ for all $x, y \in X$.
(MS2) $\mathbf{d}(x, y) = 0$ if and only if $x = y$.
(MS3) $\mathbf{d}(x, y) = \mathbf{d}(y, x) \geq 0$ for all $x, y \in X$.
(MS4) $\mathbf{d}(x, z) \leq \mathbf{d}(x, y) + \mathbf{d}(y, z)$ for all $x, y, z \in X$.

The last property is often called the *triangle inequality* because it generalizes the usual triangle inequality from classical Euclidean geometry.

EXAMPLES. 1. The most important examples are the ordinary coordinate or Euclidean spaces $\mathbf{R}^n$ for which $\mathbf{d}(x, y) = |x - y|$. The four basic properties for an abstract metric are established in undergraduate courses containing linear or vector algebra.

2. If $(X, \mathbf{d})$ is a metric space and A is a subset of X, then one can make A into a metric space using the *subspace metric* given by $\mathbf{d}\big|(A \times A)$; less formally, this means that the distances between points of A are the same as their distances in X itself.

3. If S is an arbitrary set, then one can make S into a metric space with the so-called *discrete metric*, for which $\mathbf{d}(s, t) = 1$ if $s \neq t$ and 0 if $s = t$. It is a routine exercise to verify that this defines a metric on x.

4. One can for an abstract generalization of the first example a follows: If one defines a *norm* on a real vector space V to be a function sending each $v \in V$ to a nonnegative real number $|v|$ such that

(a) $|v| = 0$ if and only if $v = 0$,
(b) $|cv| = |c||v|$ for all $c \in \mathbf{R}$ and $v \in V$,
(c) $|v + w| \leq |v| + |w|$ for all $v, w \in V$,

then the formula $\mathbf{d}(v, w) = |v - w|$ defines a metric on V. It is again an elementary exercise to verify that this satisfies the four conditions required for a metric.

5. We need to give some additional examples in order to show that the preceding construction yields something beyond ordinary Euclidean spaces.

(5A) Actually, this is two examples. Take $V = \mathbf{R}^n$, write a typical vector x in coordinates as $(x_1, \cdots, x_n)$, and consider the functions $|x|_1 = \sum_i |x_i|$ and $|x|_\infty = \max_i\{\ |x_i|\ \}$. It is again elementary to check that each of these define norms. If one wants to distinguish the previous norm from these it is customary to write the latter as $|x|_2$, which reflects the quadratic nature of the latter (as in measure theory, one can interpolate an entire series of norms $|x|_p$ for $1 \leq p \leq \infty$ but we shall not need these examples here).

(5B) This is a much larger example. Let S be a set, and let $\mathbf{C}(S)$ be the space of all bounded real valued functions on S. Then a norm is defined by the formula $|f| = \sup_{x \in S}\{\ |f(x)|\ \}$.

It is also possible to construct a vast array of other norms on vector spaces at this point, but we shall not do so in order to avoid straying too far from the central themes of the course.

The examples above show that one can construct many different metrics on a given set. However, it clearly becomes very cumbersome to write $(X, \mathbf{d})$ every time we are referring to a metric space, so in order to simplify the exposition we shall often simply write X if the metric is clear from the context.

Open sets

The basic definition extends the one for Euclidean spaces.

Definition. Let $(X, \mathbf{d})$ be a metric space. A subset $U \subset X$ is said to be *open* if for each $x \in U$ there is a positive real number ε such that $\mathbf{d}(x, y) < \varepsilon \Longrightarrow y \in U$.

For each $r > 0$ and $x \in X$, the set

$$N_r(y) = \{\ y \in X \mid \mathbf{d}(x, y) < r\ \}$$

is called the open ball (or disk or neighborhood) of radius r centered at x. One can rewrite the definition of open set to say that for all $x \in U$ there exists an $\varepsilon > 0$ such that $N_\varepsilon(x) \subset U$.

The most important properties of open sets in metric spaces are summarized in the following result:

THEOREM. *Let $\mathcal{U}$ be the family of all open subsets of X. The the following hold:*

(i) *Each subset of the form $N_\varepsilon(x)$ is open in X.*
(ii) *The empty set and X itself are both open in X.*
(iii) *If for each $\alpha \in A$ the set U_α is open in X then $\cup_\alpha U_\alpha$ is also open in X.*
(iv) *If U_1 and U_2 are open in X then $U_1 \cap U_2$ is also open in X.*

One can combine (iv) and finite induction to prove that the intersection of any finite collection of open subsets of X is also open in X.

Proof. (i) Let $y \in N_\varepsilon(x)$, and let $s = \mathbf{d}(x,y)$. We claim that $N_{\varepsilon-s}(y) \subset N_\varepsilon(x)$. It may be worthwhile to draw a two-dimensional picture at this point in order to make this assertion plausible; the formal proof of the assertion proceeds as follows. Suppose that $z \in N_{\varepsilon-s}(y)$, so that $\mathbf{d}(y,z) < \varepsilon - s$. We need to prove that $\mathbf{d}(x,z) < \varepsilon$. Applying the triangle inequality we have

$$\mathbf{d}(x,z) \leq \mathbf{d}(x,y) + \mathbf{d}(y,z) = s + \mathbf{d}(y,z) < s + (\varepsilon - s) < \varepsilon$$

as required.

(ii) We shall first consider the case of the empty set. It has no points so the condition on all points in it will automatically be true because it is a statement about nothing. The openness of x follows because $N_\varepsilon(x) = X$ for all x and ε.

(iii) Suppose that $x \in \cup_\alpha U_\alpha$, and choose $\beta \in A$ so that $x \in U_\beta$. Then one can find $\varepsilon > 0$ so that $N_\varepsilon(x) \subset U_\beta$, and since $U_\beta \subset \cup_\alpha U_\alpha$ it also follows that $N_\varepsilon(x) \subset \cup_\alpha U_\alpha$.

(iv) Let $i = 1$ or 2, and let $x \in U_1 \cap U_2$. Then one has $\varepsilon_i > 0$ so that $N_{\varepsilon_i}(x) \subset U_i$ for $i = 1, 2$. If we take ε to be the smaller of ε_1 and ε_2 then $N_\varepsilon(x) \subset U_1 \cap U_2$. Once again, it might be helpful to draw a picture as an aid to understanding this proof.■

It is useful to look at the meaning of open subset for one of the examples described above; namely, the discrete metric on a set. In this case $N_1(x)$ is merely the one point set $\{x\}$. Thus every one point subset of S is open with respect to the discrete metric. But if $W \subset S$, then clearly we have

$$W = \bigcup_{x \in W} \{x\}$$

so by the preceding theorem we see that *every subset of a metric space with a discrete metric is an open subset.* Of course, for examples like Euclidean spaces there are many examples of subsets that are not open. In particular, one point subsets are **NEVER** open in Euclidean spaces (unless one adopts the convention $\mathbf{R}^0 = \{0\}$, in which case this object must be excluded).

Topological spaces

The preceding theorem provides the motivation for the central concept of a course in point set topology:

Definition. A *topological space* is a pair $(X, \mathbf{T})$ consisting of a set X and a collection $\mathbf{T}$ of subsets of X satisfying the following conditions:

(TS1) The empty set and X itself both belong to $\mathbf{T}$.
(TS2) If for each $\alpha \in A$ the set $U_\alpha \subset X$ belongs to $\mathbf{T}$, then $\cup_\alpha U_\alpha$ also belongs to $\mathbf{T}$.
(TS3) If U_1 and U_2 belong to $\mathbf{T}$, then $U_1 \cap U_2$ also belongs to $\mathbf{T}$.

We often say that $\mathbf{T}$ is a topology on X or that $\mathbf{T}$ is the family of open subsets of X, and we say that $U \subset X$ is open if $U \in \mathbf{T}$. As before, if the topology on a set is clear from the context we shall often use the set by itself to denote a topological space.

Note. One can combine **(TS3)** and finite induction to prove that the intersection of any finite collection of subsets in $\mathbf{T}$ is also open in $\mathbf{T}$.

If $(X, \mathbf{d})$ is a metric space and $\mathbf{T}$ denotes the family of open subsets of X, then by the preceding theorem $(X, \mathbf{T})$ is automatically a topological space. We often call this **the metric topology** (associated to $\mathbf{d}$).

In particular, if S is an arbitrary set and we put the discrete metric on S, then we have seen that all subsets of S are open in the corresponding metric topology. More generally, a topological space is said to be *discrete* if the every subset is open (equivalently, every one point subset is open); by previous observations, this is just the metric topology associated to the discrete metric.

On the other hand, there are many examples of topological spaces that do not come from metric spaces.

EXAMPLES. 1. Given a set X, the *indiscrete topology* on X is the family $\mathbf{T}$ consisting only of the empty set and X itself. It is elementary to verify that this defines a topology on X. However, if X contains at least two points then this cannot come from a metric space because if X is a metric space and $p \in X$ then $X - \{p\}$ is open for all $p \in X$ (we shall prove this below).

2. Given a set X, the *finitary topology* on X is the family $\mathbf{T}$ consisting of the empty set and all subsets of the form $X - A$ where A is finite. The verification that $\mathbf{T}$ is a topology can be found in Example 3 on page 77 of Munkres. If X is finite this is equal to the metric topology for the discrete metric. On the other hand, if X is infinite, then if u and v are distinct points of X and U and V, then $U \cap V$ is always infinite (its complement is finite!), and by the Hausdorff separation property below it follows that $\mathbf{T}$ does not come from a metric on X if the latter is infinite.

Here are the results that we need to show that these examples do not come from metrics:

PROPOSITION T1. *If X is a metric space and $p \in X$, then $X - \{p\}$ is open.*

Proof. Let $q \in X - \{p\}$, so that $q \neq p$ and $r = \mathbf{d}(q, p) > 0$. Then clearly $N_r(q) \subset X - \{p\}$, and hence the latter is open.

PROPOSITION T2. (Hausdorff Separation Property) *If X is a metric space and $u, v \in X$ are distinct points, then there exist disjoint open subsets U and V containing u and v respectively.*

Proof. Let $2\varepsilon = \mathbf{d}(x, y) > 0$, and take U and V to be $N_\varepsilon(u)$ and $N_\varepsilon(v)$ respectively. To see that these are disjoint, suppose that they do have some point z in common. Then by the triangle inequality and $z \in N_\varepsilon(u) \cap N_\varepsilon(v)$ we have

$$2\varepsilon = \mathbf{d}(u, v) \leq \mathbf{d}(u, z) + \mathbf{d}(z, v) < \varepsilon + \varepsilon$$

which is a contradiction. Therefore the intersection must be empty. Once again, it might be helpful to draw a picture as an aid to understanding this proof.

The following result is often useful in working with open sets:

LEMMA. *If U is an open subset of a metric space, then one can find numbers $\varepsilon(y) > 0$ for all $y \in U$ such that $U = \cup_y \, N_{\varepsilon(y)}(y)$.*

Proof. If $\varepsilon(y) > 0$ such that $N_{\varepsilon(y)}(y) \subset U$, then we have the chain of inclusions

$$U = \bigcup_y \{y\} \ \subset \ \bigcup_y \ N_{\varepsilon(y)}(y) \ \subset \ U$$

shows that $U = \cup_y \, N_{\varepsilon(y)}(y)$.

Comparing and constructing topologies on a set

We have seen that a given set may have several different metrics and several different topologies. There are a few aspects of the topologies on a space that are worth examining in some more detail at this time.

Since topologies on a set are just families of subspaces, it is meaningful to ask if one is contained in the other. Every topology must contain the indiscrete topology, and since the topology of a discrete metric contains every subset it is clear that every other topology is contained in this one. Frequently one sees statements that one topology $\mathbf{T}$ on a space is stronger or weaker than another topology $\mathbf{S}$ if one contains the other, and the terms coarser or finer are also used in such contexts. Unfortunately, this notation can be hopelessly confusing because, say, there is no consistency about whether a stronger topology has more open sets than a weaker one or vice versa. We shall avoid this by simply saying that one topology is larger or stronger than the other (compare the first few lines on page 78 of Munkres).

The following observation is elementary to verify:

FACT. *The unions and intersections of arbitrary families of topologies on a given set X are also topologies on X.*■

Another abstract feature of topologies is that *given a family $\mathcal{A}$ of subsets of X, there is a unique minimal topology $\mathbf{T}(\mathcal{A})$ on X that contains $\mathcal{A}$.*

In view of the fact stated above, the topology can be described as the intersection of all topologies that contain $\mathcal{A}$; this family is nonempty because the discrete topology is a specific example of a topology containing $\mathcal{A}$. However, for many purposes it is necessary to have a more explicit description of this topology. Define $\mathcal{A}^*$ to be the set of arbitrary unions of sets having the form $A_1 \cap \cdots \cap A_k$ for some finite family of subsets $\{A_1, \cdots, A_k\}$ in $\mathcal{A}$ together with X and the empty set. To show that $\mathcal{A}^*$ is a topology it is necessary to verify that it is closed under arbitrary unions and finite intersections. It will be convenient to let $\mathcal{B}$ denote the set of finite intersections of sets in $\mathcal{A}$; then every element of $\mathcal{A}^*$ except possibly X and the empty set will be a union of subsets in $\mathcal{B}$. Since a union of unions of subsets in $\mathcal{B}$ is again a union of subsets in $\mathcal{B}$, it follows that $\mathcal{A}^*$ is closed under taking arbitrary unions. Suppose now that U and V lie in $\mathcal{A}^*$. Write these sets as $\cup_\beta U_\beta$ and $\cup_\gamma V_\gamma$ respectively; then

$$U \cap V = \bigcup_{\beta,\gamma} U_\beta \cap V_\gamma$$

where each summand $U_\beta \cap V_\gamma$ is a finite intersection of subsets in $\mathcal{A}$, and this implies that $U \cap V \in \mathcal{A}^*$ as required.

Definition. A family $\mathcal{A}$ of subsets of X is called a *subbase* for the topology $\mathbf{T}$ on X if $\mathcal{A}^* = \mathbf{T}$.

Basic open subsets for a topology

There is a special type of subbase known as a *base* that is often useful in constructing topologies.

Definition. A family $\mathcal{B}$ of subsets of X is called a *base* for the topology $\mathbf{T}$ on X if

- **(B0)** $\mathcal{B}^* = \mathbf{T}$,
- **(B1)** each $x \in X$ belongs to at least one $B \in \mathcal{B}$,

(B2) if $x \in B_1 \cap B_2$ where $B_1, B_2 \in \mathcal{B}$, then there is a $B_3 \in \mathcal{B}$ such that $x \in B_3$ and $B_3 \subset B_1 \cap B_2$.

If $\mathcal{B}$ is a base for $\mathbf{T}$ we often refer to the sets in $\mathcal{B}$ as *basic open subsets.*

The following result is elementary to prove:

PROPOSITION. *If $\mathcal{B}$ is a family of subsets of X satisfying* **(B1)** *and* **(B2)** *above, then the smallest topology containing $\mathcal{B}$ is the set of all unions of sets in $\mathcal{B}$ together with the empty set.*■

Lemma 13.2 on page 80 of Munkres is another useful result on bases for topologies.

Subspace topologies

We have already noted that a subset of a metric space can be viewed in a natural way as a metric space in its own right by restricting the metric. There is a parallel way of viewing a subset of a topological space as a topological space in its own right, and it turns out that if X is a metric space, the topologies that one obtains on A in both fashions are identical.

Definition. If $(X, \mathbf{T})$ is a topological space and A is a subset of X, then the *subspace topology* on A is the family $\mathbf{T}|A$ of all intersections $U \cap A$ where $U \in \mathbf{T}$. — It is an elementary set-theoretic exercise to verify that this defines a topology on A.

The following result relates the metric and subspace topologies on a subset of a metric space.

PROPOSITION. *If X is a metric space and $A \subset X$, then the metric topology on A is identical to the subspace topology on A.*

Proof. It will be convenient to distinguish the open disks in A and X by N^A and N^X respectively. By construction we have that $N^A = A \cap N^X$.

Let $\mathbf{T}|A$ denote the subspace topology and let $\mathbf{M}_A$ denote the metric topology. Every set in $\mathbf{T}|A$ has the form $U \cap A$ where $U \in \mathbf{T}$, where $\mathbf{T}$ denotes the metric topology on X. By definition of the metric topology on X, for every point $y \in A \cap U$ there is an $r > 0$ such that $N_r^X(y) \subset U$, and therefore we have

$$N_r^A(y) = A \cap N_r^X(y) \subset A \cap U$$

which shows that every set in $\mathbf{T}|A$ belongs to $\mathbf{M}_A$. Conversely, every open set W in the metric topology is a union of open disks having the form $N_{r(y)}^A(y)$ for $y \in W$ and suitably chosen $r(y) > 0$ (use the lemma stated above), and therefore we have

$$W = \bigcup_y N_{r(y)}^A(y) \ = \ \bigcup_y \left(A \ \cap N_{r(y)}^X(y)\right) \ = \ A \cap \left(\bigcup_y N_{r(y)}^A(y)\right)$$

which shows that W is an intersection of A with an open subset of X.■

Neighborhoods

In a metric space the sets $N_\delta(x)$ are often called *δ-neighborhoods*; for general topological spaces one also uses the term *neighborhood of a point* (say) p to denote an open set containing p (compare the definition on page 96 of Munkres); however, sometimes the term "neighborhood" has a more general meaning of a set N such that N contains some open subset U which in turn contains the point p; one should be aware of this difference when reading other books.

II.2 : Closed sets and limit points

(Munkres, § 17)

The usual discussion of limits for sequences for real numbers extends directly to metric spaces. Given a metric space X and a sequence $\{a_n\}$ in X, we say that $\lim_{n\to\infty} a_n = a$ if for all $\varepsilon > 0$ there is a positive integer M such that $n > M$ implies $\mathbf{d}(a_n, a) < \varepsilon$. As in real variables, a sequence has at most one limit, and the proof in the general case is essentially the same.

Closed subsets of the real line are precisely the subsets that are "closed under taking limits of convergent sequences," so it is clear that one should be able to discuss closed subsets of an arbitrary metric space. At first glance it is less obvious that one can also discuss closed sets for arbitrary topological spaces, but this is indeed the case, and one major objective here to justify this.

Limit points and limits of sequences

We begin with the following definition:

Definition. Let X be a topological space, and let $A \subset X$ be a subset of X. A point $y \in X$ is called a *limit point* of A if for all open sets U containing y the intersection $A \cap (U - \{y\})$ is nonempty. The set of all limit points of A in X is written $\mathbf{L}(A; X)$, and when the ambient space X is clear from the context we shall often write $\mathbf{L}(A)$.

The motivation for the terminology is implicit in the following result:

PROPOSITION. *The following are equivalent for a metric space X, a point $y \in X$ and a subset $A \subset X$:*

(i) $y \in \mathbf{L}(A)$.

(ii) There is a sequence of points $\{a_n\}$ in A such that $a_n \neq y$ for all n but $\lim_{n\to\infty} a_n = y$.

Proof. $\big((i) \Longrightarrow (ii)\big)$ Let n be a positive integer, and consider the open set $N_{1/n}(y)$. By the definition of $\mathbf{L}(A)$ there is a point $a_n \in A$ such that $a_n \neq y$ but $a_n \in N_{1/n}(y)$. Then

$$\mathbf{d}(a_n, y) < \frac{1}{n}$$

for all n and therefore $\lim_{n\to\infty} a_n = y$.

$\big((ii) \Longrightarrow (i)\big)$ Let the sequence $\{a_n\}$ be given as in the statement of (ii), let U be an open subset containing y, let $\varepsilon > 0$ be such that $N_\varepsilon(y) \subset U$, and choose M such that $n > M$ implies $\mathbf{d}(a_n, a) < \varepsilon$. Then we have

$$a_{M+1} \in A \cap (N_\varepsilon(y) - \{y\}) \subset A \cap (U - \{y\})$$

and therefore $y \in \mathbf{L}(A)$.■

The nest result is the key to defining closed subsets in arbitrary topological spaces.

THEOREM. *If A is a subset of a topological space X, then $X - A$ is open if and only if $\mathbf{L}(A) \subset A$.*

Proof. $(\Longrightarrow)$ Suppose that $X - A$ is open and the set $\mathbf{L}(A)$ is not contained in A. Let $y \in \mathbf{L}(A) - A$; clearly $y \in X - A$. By the definition of the set of limit points, it follows that there is a point $x \in X - A$ that is also in A, which is a contradiction. Thus $\mathbf{L}(A) \subset A$ if $X - A$ is open.

($\Longleftarrow$) Suppose that $\mathbf{L}(A) \subset A$ and let $y \in X - A$. By hypothesis y is not a limit point of A and therefore there is some open set U_y containing y such that $U_y - \{y\}$ and A are disjoint (*Note:* You should check that the conclusion is the negation of the condition in the definition of limit point!). Since we also know that $y \notin A$ it follows that U_y and A are also disjoint, so that $U_y \subset X - A$. Therefore we have the string of inclusions

$$X - A \quad = \quad \bigcup_{y \notin A} \{y\} \quad \subset \quad \bigcup_{y \notin A} U_y \quad \subset \quad X - A$$

which shows that $X - A = \cup_y U_y$; since the right hand side is a union of open sets, it follows that the sets on both sides of the equation are open.■

This leads us to a topological definition of closed set that is compatible with the notion of "closure under limits of sequences" for metric spaces.

Definition. A subset F of a topological space X is *closed* if and only if its relative complement $X - A$ is open.

Note. In contrast with the usual usage for the terms "open" and "closed," a subset of a topological space may be open but not closed, closed but not open, neither open nor closed, or both open and closed. Over the real line these are illustrated by the subsets $(0, 1)$, $[0, 1]$, $[0, 1)$ and $\mathbf{R}$ itself (you should verify this for each example).

Closed subsets have the following properties that correspond to the fundamental properties of open subsets.

PROPOSITION. *The family of closed subsets of a topological space X has the following properties:*

(i) *The empty set and X itself are both closed in X.*
(ii) *If for each $\alpha \in A$ the set F_α is closed in X then $\cap_\alpha F_\alpha$ is also closed in X.*
(iii) *If F_1 and F_2 are closed in X then $F_1 \cup F_2$ is also closed in X.*

Note. One can combine (iii) and finite induction to prove that the union of any finite collection of closed subsets in X is also closed in X.

Proof. (i) The empty set and X are complements of each other, so since each is open their complements — which are merely the empty set and X itself — are closed.

(ii) This follows immediately from the complementation formula

$$X - \bigcap_\alpha F_\alpha \quad = \quad \bigcup_\alpha (X - F_\alpha)$$

and the fact that unions of open subsets are open.

(iii) This follows immediately from the complementation formula

$$X - (F_1 \cup F_2) \quad = \quad (X - F_1) \cap (X - F_2)$$

and the fact that the intersection of two open subsets is open.

Clearly one could define mathematical systems equivalent to topological spaces by specifying families of closed subsets satisfying the three properties in the preceding proposition. In fact, there are also many other equivalent ways of describing topological spaces, but we shall not say very much about them here.■

For metric spaces one has the following important fact regarding closed subsets.

PROPOSITION. *If X is a metric space and $x \in X$, then the one point set $\{x\}$ is closed in X.*

This follows immediately from an earlier observation that $X - \{x\}$ is open in X if X is a metric space. As noted previously, the indiscrete topology on a set with at least two elements does not have the corresponding property.■

Closures and interiors of subsets

In many mathematical contexts it is useful and enlightening to have constructions that fill the gaps in a mathematical object. For example, the real numbers are a way of filling the gaps in the rational numbers. Given a subset of, say, the real line, one often wants to expand this set so that it contains all limits of sequences that are defined on that set. This is done by considering the *closure* of the set, and the concept can be formulated in a manner that applies to all topological spaces.

Definition. Given a topological space X and a subset $A \subset X$, the *closure* of A is the set $\overline{A} = A \cup \mathbf{L}(A)$.

The terminology suggests that $\overline{A}$ should be the smallest closed subset of X that contains A. Verifying this will take a little work.

PROPOSITION. *The set $\overline{A}$ is the intersection of all closed subsets containing A, and consequently it is the smallest closed subset containing A.*

Proof. Let F be the intersection described in the statement of the proposition. We need to prove the two inclusions $\overline{A} \subset F$ and $F \subset \overline{A}$.

$\left(\overline{A} \subset F\right)$ Since $A \subset F$ it is only necessary to show that $\mathbf{L}(A) \subset F$. It follows immediately from the definitions that $A \subset F$ implies $\mathbf{L}(A) \subset \mathbf{L}(F)$ (fill in the details here!). But since F is closed we know that $\mathbf{L}(F) \subset F$, and therefore we have that $\mathbf{L}(A) \subset F$ as required.

$\left(F \subset \overline{A}\right)$ This will follow if we can show that $\overline{A}$ is closed in X, or equivalently that $X - \overline{A}$ is open in X. So suppose that $y \in X - \overline{A}$. By definition this means that $y \notin A$ and $y \notin \mathbf{L}(A)$. The latter in turn means that there is an open set $U_y \subset X$ such that $A \cap (U_y - \{y\})$ is empty. But we also know that $y \notin A$,so we can strengthen the latter to say that $A \cap U_y$ is empty. We then have the usual chain of inclusions

$$X - \overline{A} \quad = \quad \bigcup_{y \notin \overline{A}} \{y\} \quad \subset \quad \bigcup_{y \notin \overline{A}} U_y \quad \subset \quad X - A$$

which shows that $X - \overline{A} = \cup_y U_y$ and therefore is open; but this means that $\overline{A}$ is closed.■

There is a complementary concept of the *interior* of a set A, which is the largest open subset U contained in A. Formally, one can define the interior by the formula

$$\text{Int }(A) \quad = \quad X - \overline{X - A}$$

and the proof that this is the union of all open subsets contained in A reduces to an exercise in set theory.

For the sake of completeness, here are the details: One can rewrite the defining equation as $X - \text{Int }(A) = \overline{X - A}$ and since the latter contains $X - A$, by taking complements we have that $\mathrm{Int}(A)$ is an open set that is contained in A. Suppose now that U is any open subset contained in A. Then $X - U$ is a closed set that contains $X - A$ and thus also $\overline{X - A} = X - \mathrm{Int}\ (A)$; taking complements once again we see that U is contained in $\mathrm{Int}(A)$.■

Warning. In some topological contexts the term "interior" has a much different meaning, but in this course the term will always have the meaning given above.

The following result provides an extremely useful relation between the notions of closure and passage to subspaces.

PROPOSITION. *Given a topological space X and subspaces A, Y such that $A \subset Y \subset X$, let $\mathrm{Closure}_Y(A)$ denote the closure of A with respect to the subspace topology on Y. Then*

$$\mathrm{Closure}_Y(A) = \overline{A} \cap Y \ .$$

Proof. Once again we have to prove the inclusions in both directions.

$\big(\mathrm{Closure}_Y(A) \subset \overline{A} \cap Y\big)$ Note first that the close subsets of Y have the form $Y \cap F$ where F is closed in X (*Proof:* E is closed in $Y \Leftrightarrow Y - E$ is open in $Y \Leftrightarrow Y - E = X \cap U$ for some U open in $X \Leftrightarrow E = Y - Y \cap U$ for some U open in $X \Leftrightarrow E = Y \cap (X - U)$ for some U open in X $\Leftrightarrow E = Y \cap F$ for some F closed in X). — It follows that the right hand side is a closed subset of Y and therefore contains the set $\mathrm{Closure}_Y(A)$.

$\big(\overline{A} \cap Y \subset \mathrm{Closure}_Y(A)\big)$ The right hand side is a closed subset of Y, and therefore by the preceding paragraph it has the form $B \cap Y$ where B is closed in X. By construction $B \supset A$, so B must also contain $\overline{A}$. But this means that

$$\overline{A} \cap Y \subset B \cap Y = \mathrm{Closure}_Y(A)$$

which yields the desired inclusion.■

Convergence in general topological spaces

In general one cannot work with limits of sequences in abstract topological spaces as easily and effectively as one can work with them in metric spaces. The crucial property of metric spaces that allows one to work with sequences is the following:

First Countability Property. *If X is a metric space and $x \in X$ then there is a sequence of decreasing open subsets U_k such that every open subset contains some U_k.*

In fact, we can take U_k to be the open disk of radius $\frac{1}{k}$ centered at x (*i.e.*, $N_{1/k}(x)$).■

There is a somewhat more complicated concept of *net* that serve a similar purpose to sequences for arbitrary topological spaces. Nets for topological spaces are not as important or useful as sequences for metric spaces, but there are some situations where it is convenient to have them. A concise but readable introduction to nets appears on pages 187–188 of Munkres.

II.3 : Continuous functions

(Munkres, §§ 18, 21; Edwards, § I.8)

The standard definitions for continuous and uniformly continuous functions generalize immediately to metric spaces.

Definition. Let $(X, \mathbf{d}_X)$ and $(Y, \mathbf{d}_Y)$ be metric spaces, and let $a \in X$. A set-theoretic function $f : X \to Y$ is said to be *continuous* at a if for each $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon) > 0$ such that $\mathbf{d_X}(x, a) < \delta$ implies $\mathbf{d_Y}(f(x), f(a)) < \varepsilon$. The function f is said to be continuous (on all of X) if it is continuous at every point of X.

As in the case of functions of a real variable, the numbers $\delta(\varepsilon)$ depend upon the point a.

Definition. Let $(X, \mathbf{d}_X)$ and $(Y, \mathbf{d}_Y)$ be metric spaces, and let $a \in X$. A set-theoretic function $f : X \to Y$ is said to be *uniformly continuous* if for each $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon) > 0$ such that for all $u, v \in X$, we have that $\mathbf{d_X}(u, v) < \delta$ implies $\mathbf{d_Y}(f(u), f(v)) < \varepsilon$.

As in real variables, the difference between continuity and uniform continuity is that δ depends upon ε and a for continuity but it depends only upon ε for uniform continuity.

The following characterization of continuity yields a definition that is meaningful for functions on arbitrary topological spaces:

THEOREM. *Let $(X, \mathbf{d}_X)$ and $(Y, \mathbf{d}_Y)$ be metric spaces, and let $f : X \to Y$ be a function. Then f is continuous if and only if for each open set $V \subset Y$, the inverse image $f^{-1}(V)$ is open in X.*

Proof. $(\Longrightarrow)$ Choose $y \in Y$ and $x \in X$ so that $y = f(x)$. There is an $\varepsilon > 0$ so that $N_\varepsilon(y) \subset V$, and by continuity there is a $\delta > 0$ such that f maps $N_\delta(x)$ into $N_\varepsilon(y)$. It follows as in many previous arguments that

$$f^{-1}(V) \quad = \quad \bigcup_x N_\delta(x)$$

(check this out!) and therefore the left hand side is an open subset of X.

$(\Longleftarrow)$ Choose $y \in Y$ and $x \in X$ so that $y = f(x)$, and let $\varepsilon > 0$ be given. By the hypothesis we know that the set

$$W = f^{-1}\left(N_\varepsilon(y)\right)$$

is an open subset of X containing x. If we choose $\delta > 0$ so that $N_\delta(x) \subset W$, then it follows that f maps $N_\delta(x)$ into $N_\varepsilon(y)$.■

In view of the above, if $(X, \mathbf{T}_X)$ and $(Y, \mathbf{T}_Y)$ are topological spaces we may **DEFINE** a set-theoretic map $f : X \to Y$ to be continuous if and only if for each open set $V \subset Y$, the inverse image $f^{-1}(V)$ is open in X.

Several equivalent formulations of continuity are established in Theorem 18.1 on pages 104–105 of Munkres and Lemma 21.3 on page 130 of Munkres. Here is an overlapping list of equivalences:

CHARACTERIZATIONS OF CONTINUITY. *Let X and Y be topological spaces and let $f : X \to Y$ be a set-theoretic map. then the following are equivalent:*

(1) *f is continuous.*

(2) *For every closed subset $F \subset Y$ the inverse image $f^{-1}(F)$ is closed.*

(3) *For all* $A \subset X$ *we have* $f(\overline{A}) \subset \overline{f(A)}$.

(4) *For all* $B \subset Y$ *we have* $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$.

(5) *For all* $A \subset X$ *we have* $\mathrm{Int}\,(f(A)) \subset f\,(\mathrm{Int}(A))$.

(6) *For all* $B \subset Y$ *we have* $f^{-1}\,(\mathrm{Int}(B)) \subset \mathrm{Int}\left(f^{-1}(B)\right)$.

If X *and* Y *are metric spaces then the following is also equivalent to the preceding conditions:*

(7) *For all sequences* $\{\, x_n \,\}$ *in* x *such that* $\lim_{n\to\infty} x_n = a$ *we have* $\lim_{n\to\infty} f(x_n) = f(a)$.

The statements and proofs of the results in Munkres should be read and understood. Verification of the statements not proven in Munkres is left to the reader as an exercise.■

It is not possible to discuss uniform continuity in a topological space unless some extra structure is added; one reference for an abstract treatment of such uniform structures (or uniformities) is Kelley, *General Topology.* Topological spaces with uniform structures are often known as *uniform spaces.* A class of spaces known as *topological groups* have particularly important examples of the uniform structures that exist on uniform spaces. An introduction to the theory of topological groups appears in Appendix A of these notes.

EXAMPLES. 1. A real variables textbook (and even a calculus or precalculus textbook) contains many examples of continuous functions from subsets of the real numbers to the real numbers.

2. If A is a subset of a topological space with the subspace topology, then the inclusion map $i : A \to X$ is continuous because $i^{-1}(U) = U \cap A$ for all open subsets U. In fact, if X is a metric space and A has the subspace metric, then the inclusion map is uniformly continuous; for each $\varepsilon > 0$ we can take $\delta = \varepsilon$.

3. An important special case of the preceding example occurs when $A = X$, and in this case the inclusion is the *identity map* on X.

4. Let X be an arbitrary metric space, and let A be a nonempty subset of X. For each point $x \in X$ the distance from x to A is defined by the formula

$$\mathbf{d}(x, A) = \mathrm{g.}\ell\mathrm{.b.}_{a\in A}\mathbf{d}(x, a)$$

where the greatest lower bound exists and is nonnegative because all distances are nonnegative. We claim that the function $\mathbf{d}(-, A)$ is uniformly continuous. — Here is a **Proof**: By the triangle inequality we have that $\mathbf{d}(x, a) \le \mathbf{d}(x, y) + \mathbf{d}(y, a)$ for all $x, y \in X$ and $a \in A$. Therefore it follows that $\mathbf{d}(x, A) \le \mathbf{d}(x, y) + \mathbf{d}(y, a)$. Subtract $\mathbf{d}(x, y)$ from each side. This yields the inequality $\mathbf{d}(x, A) - \mathbf{d}(x, y) \le \mathbf{d}(y, a)$, which in turn implies that the left hand side is $\le \mathbf{d}(y, A)$. We can now rewrite this in the form $\mathbf{d}(x, A) - \mathbf{d}(y, A) \le \mathbf{d}(x, y)$. If we reverse the roles of x and y in this argument we get the complementary inequality $\mathbf{d}(y, A) - \mathbf{d}(x, A) \le \mathbf{d}(x, y)$. Combining these, we obtain the inequality

$$\left|\mathbf{d}(y, A) - \mathbf{d}(x, A)\right| \le \mathbf{d}(x, y)$$

which shows that the function in question is in fact uniformly continuous because for each $\varepsilon > 0$ we can take $\delta = \varepsilon$.

5. We shall end this list of examples with one that is very simple. Suppose that X and Y are any topological spaces and that $y \in Y$. Then there is a constant map $C_y : X \to Y$ which sends every point of X to y. This map is continuous. To see this, let $V \subset Y$ be open, and consider $f^{-1}(V)$. If $y \in V$ then the inverse image is all of X but if $y \notin V$ then the inverse image is the empty set. In either case the inverse image is open.

In analysis there are theorems stating that sums, products and composites of continuous functions are continuous. Metric and topological spaces usually do not have the algebraic structure needed to construct sums and products. However, one does have the following version of continuity for composite functions.

PROPOSITION. *If X, Y, Z are topological spaces and $f : X \to Y$ and $g : Y \to Z$ are continuous, then so is the composite $g \circ f : X \to Z$.*

Proof. Suppose that W is open in Z; then by continuity it follows that $V = g^{-1}(W)$ is open in Y and $U = f^{-1}(V)$ is open in X. However, we also have

$$u = f^{-1}\left(g^{-1}(W)\right) = (g \circ f)^{-1}(W)$$

and therefore it follows that $g \circ f$ is also continuous.■

COROLLARY. *If $f : X \to Y$ is continuous and $A \subset X$ is equipped with the subspace topology, then the restriction $F|A : A \to Y$ is continuous.*

This is true because the restriction is the composite of f and the inclusion map for $A \subset X$; we have already noted that the latter is continuous.

In addition to the preceding way of constructing continuous functions by restricting the domain, it is also possible to construct new continuous functions by shrinking the codomain if the image of the function is a proper subset.

PROPOSITION. *Let $f : X \to Y$ be a continuous function, let $B \subset Y$ be equipped with the subspace topology, let $j : B \to Y$ denote the inclusion map, and suppose that $f(X) \subset B$. Then there is a unique continuous map $g : X \to B$ such that $j \circ g = f$.*

Proof. On the set-theoretic level one simply defines g by the rule $g(x) = f(x)$. We need to verify that this map is continuous.

Suppose that V is open in B. Then $B = W \cap B$ where W is open in Y. Given an arbitrary subset $A \subset Y$, elementary set-theoretic considerations imply that

$$f^{-1}(A) = f^{-1}(A \cap B) = g^{-1}(A \cap B)$$

with the first equation holding because $f(X) \subset B$ and the second holding because $f(x) = g(x)$ for all x. Therefore if V is open in B and $V = W \cap B$ (where W is open in Y), then

$$g^{-1}(V) = f^{-1}(W \cap B) = f^{-1}(W) \ .$$

Since f is continuous the set on the right hand side of the equation is open in X; therefore the set on the left hand side is also open and the map g is continuous.■

Homeomorphisms and other special mappings

We begin with a natural question:

Continuity of inverses. *Suppose that $f : X \to Y$ is a continuous map of topological spaces that is a $1-1$ correspondence. Is the inverse map f^{-1} also continuous?*

There are many examples to show that the answer to the question is negative. One purely formal approach is to take a $X = Y$ with $f = \mathrm{id}_X$ and the topologies on the domain and codomain equal to the discrete and indiscrete topologies respectively. Then f is continuous (every map into a space with the indiscrete topology is continuous!). What can we say about the continuity of the inverse? By construction the inverse is just the identity map from a space with the indiscrete topology to a space with the discrete topology. If X has more than one element and A is a nonempty proper subset, then A is open in the discrete topology but not in the indiscrete topology, and therefore the inverse map is not continuous.

Here is a more tangible example. Let S^1 be the unit circle in the cartesian plane, and let $f : [0,1) \to S^1$ send t to $(\cos 2\pi t, \sin 2\pi t)$. Then f is clearly continuous and 1–1 onto (it might be helpful to draw a picture of this). However, f^{-1} is not continuous at the point $(1,0) \in \mathbf{R}^2$. Specifically, the set $[0, \frac{1}{2})$ is open in $[0,1)$, but its inverse image in the circle under f^{-1} — which is simply $f\left([0, \frac{1}{2})\right)$ (why?) — is not open in the circle. To see this, note that every open subset of the circle containing $(1,0)$ must have some points whose second coordinates are negative.

We are thus led to the following:

Fundamental Definition. A continuous 1–1 onto map $f : X \to Y$ of topological spaces is a *homeomorphism* if f^{-1} is also continuous.

It follows immediately from the definition that for every topological space X the identity map id_X is continuous (it is understood that X has the same topology whether it is viewed as the domain or the codomain), the inverse of a homeomorphism is a homeomorphism, and the composite of two (composable) homeomorphisms is a homeomorphism.

Here is an alternate characterization of homeomorphisms that may be enlightening:

PROPOSITION. *Let $(X, \mathbf{T}_X)$ and $(Y, \mathbf{T}_Y)$ be topological spaces, and let $f : X \to Y$ be a set-theoretic map that is $1-1$ and onto. Then f is a homeomorphism if and only if for each subset $A \subset X$, we have that A is open in X if and only if $f(A)$ is open in Y.*

Proof. $(\Longrightarrow)$ Let $\subset X$. If $f(A)$ is open in Y, then by continuity of f we have that

$$A = f^{-1}\left(f(A)\right)$$

is open in X. Similarly, if A is open in X, then by continuity of f^{-1} we have that

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

is open in Y.

$(\Longleftarrow)$ The hypotheses on f and the first set-theoretic identity in the previous paragraph imply that f is continuous, and the hypotheses together with the second set-theoretic identity in the previous paragraph imply that f^{-1} is continuous.∎

One can state and prove a similar theorem in which "open" is replaced by "closed" (in fact, the argument is essentially the same with this substitution).

The preceding result and its analog for closed sets lead to some other important classes of mappings on topological spaces.

Definitions. A set-theoretic map $f : X \to Y$ of topological spaces is *open* if for each open set $U \subset X$, the image $f(U)$ is open in Y. Similarly, a set-theoretic map $f : X \to Y$ of topological spaces is *closed* if for each closed set $A \subset X$, the image $f(A)$ is closed in Y.

Here are some instructive examples:

1. The identity map from a set with the indiscrete topology to a set with the discrete topology is both open and closed but not continuous.

2. The previously constructed map from $[0,1)$ to S^1 is continuous but neither open nor closed. Its inverse is open and closed but not continuous.

3. The map from $\mathbf{R}^2$ to $\mathbf{R}$ sending (x,y) to x is continuous and open but not closed. [*Hints:* The proof that the map is open reduces to showing that the image of an open δ-disk is always open. Why is the image of the open δ-disk about (x,y) equal to the open interval $\{t \mid x-\delta < t < x+\delta\}$? To show the map is not closed consider the graph of $1/x$ and its image under the given map.]

4. The map from $\mathbf{R}$ to itself sending x to its absolute value is continuous and closed but not open. [*Hints:* To show the map is not open, consider the image of the whole space. To show it is closed, explain why every closed subset F can be written as a union $F = F_+ \cup F_-$ where $F_\pm$ is a closed set consisting of all points in F that are respectively nonnegative or nonpositive. Why does f take $F_\pm$ to a closed set, and how does this show that f is closed?]

The following result is elementary:

PROPOSITION. *The composite of two open mappings is open, and the composite of two closed mappings is closed. Identity maps are always open and closed.*■

For metric spaces there is an extremely special type of homeomorphism:

Definition. Let $(X, \mathbf{d}_X)$ and $(Y, \mathbf{d}_Y)$ be metric spaces. A set-theoretic function $f : X \to Y$ is said to be an *isometry* if it is onto and $\mathbf{d}_X(u,v) = \mathbf{d}_Y(f(u), f(v))$ for all $u, v \in X$. — Such a map is automatically 1–1 because

$$u \neq v \implies \mathbf{d}_Y(f(u), f(v)) = \mathbf{d}_X(u,v) > 0 \; .$$

By construction such a map is also uniformly continuous and has a uniformly continuous inverse (which is also an isometry). Note that identity maps are always isometries, the composites of isometries are isometries, and inverses to isometries are isometries.

Different metrics determining the same topology

The discussion of diameters and bounded metric spaces on pages 121–122 of Munkres should be read at this point. Theorem 20.1 on page 121 is an important case of an extremely general phenomenon: *If $(X, \mathbf{d})$ is a metric space, then in general there are many different metrics $\mathbf{e}$ such that the identity map from $(X, \mathbf{d})$ to $(X, \mathbf{e})$ is a homeomorphism. Furthermore, in general there are many examples for which the identity map is also uniformly continuous, and in fact one can even find large classes of examples for which the identity map in the opposite direction is also uniformly continuous.* — Important examples of this will arise later in the course.

Metric spaces of functions

We have already noted that the set $\mathbf{BF}(X)$ of bounded functions on a set X has a metric space structure with

$$\mathbf{d}(f,g) = \sup_x \left| f(x) - g(x) \right| \; .$$

In such a space the limit of a sequence of functions corresponds to uniform convergence: We have $\lim_{n\to\infty} f_n = f$ if and only if for all $\varepsilon > 0$ we can find M so that $n > M$ implies $|f_n(x) - f(x)| < \varepsilon$

for all $x \in X$ (this is not quite trivial because the latter inequality only implies $|f_n - f| \le \varepsilon$, but if the condition holds we can also find M' so that $n > M'$ implies $|f_n(x) - f(x)| < \varepsilon/2$ and the latter certainly implies $|f_n - f| \le \varepsilon/2 < \varepsilon$), A large amount of the theory of uniform convergence for functions of real variables carries over to this general setting. In particular, pages 147–151 of Rudin, *Principles of Mathematical Analysis (Third Edition)*, go through with only minor changes. In particular, at certain points in a point set topology course it is necessary to use the following result, which is proved on page 132 of Munkres or pages 149–150 of the book by Rudin mentioned above.

THEOREM. *Let X be a topological space, let $\mathbf{BF}(X)$ be defined as above, and let $\mathbf{BC}(X)$ denote the set of all continuous functions in $\mathbf{BF}(X)$. Suppose that $\lim_{n\to\infty} f_n = f$ in $\mathbf{BF}(X)$ where each of the functions f_n is continuous. Then f is also continuous.*■

The spaces $\mathbf{BC}(X)$ also have a great deal of algebraic structure (for example, addition and multiplication of functions) that one has for continuous functions on, say, the unit interval. This is all discussed in Section 21 of Munkres (pp. 129–133).

Piecing together continuous functions

The previous material provides one powerful method for constructing continuous functions on metric and topological spaces. This is essentially an analytic method for constructing functions. It is often important to have a similarly useful geometric method for obtaining continuous functions.

Geometric piecing problem. *Suppose that we are given topological spaces X and Y, a family of subsets $\{A_\alpha\}$ of X and a continuous function $f_\alpha : A_\alpha \to Y$ for each α. Is it possible to form a continuous function $g : \cup_\alpha A_\alpha \to Y$ such that $g(x) = f_\alpha(x)$ if $x \in A_\alpha$?*

Notation. Given a function $h : A \to B$ and $C \subset B$ we shall denote the composite of h with the inclusion $C \to A$ by $h|C$ and call it the *restriction of h to C*. Note that if h is a continuous map of topological spaces and C has the subspace topology then $h|C$ is automatically continuous.

There is an obvious set-theoretic condition that is necessary if a function g as above actually exists. Namely, for all $y \in A_\alpha \cap A_\beta$ we need the consistency condition $f_\alpha(y) = f_\beta(y)$. In terms of the restriction notation this can be rewritten formally as

$$f_\alpha|\, A_\alpha \cap A_\beta = f_\beta|\, A_\alpha \cap A_\beta .$$

For certain families of subspaces this turns out to be the only condition needed to piece together a continuous function defined on an entire space.

THEOREM. *Let X and Y be topological spaces, let $\mathcal{A} = \{A_\alpha\}$ be a family of subsets of X such that $X = \cup_\alpha A_\alpha$, and for each let $f_\alpha : A_\alpha \to Y$ be a continuous function. Assume that these functions satisfy the consistency condition $f_\alpha|\, A_\alpha \cap A_\beta = f_\beta|\, A_\alpha \cap A_\beta$. If either*

(i) $\mathcal{A}$ is a family of open subsets,

(ii) $\mathcal{A}$ is a **finite** *family of closed subsets,*

then there is a unique continuous function $f : X \to Y$ such that $f|A_\alpha = f_\alpha$ for all α.

To see that a similar result does not hold for arbitrary families of closed subsets, consider the family $\mathcal{A}$ of one point sets $\{x\}$. Given an arbitrary set-theoretic function g from a metric space X to a topological space Y, the restrictions $g|\{x\}$ are all continuous and the consistency condition follows because the sets in the family are pairwise disjoint. Thus any discontinuous function g

satisfies the conditions for the family of closed subsets $\mathcal{A}$, and for most choices of X there are many discontinuous functions to choose from.

Proof of Theorem. *First case.* The consistency condition ensures that there is a well-defined set-theoretic function $f : X \to Y$ with the desired properties, so the real issue is to prove this function is continuous. Let V be an open subset of Y. Then we have

$$f^{-1}(V) = \bigcup_{\alpha} \left(A_\alpha \cap f^{-1}(V)\right)$$

and since $(f|A_\alpha)^{-1}(V) = A_\alpha \cap f^{-1}(V)$ the right hand side of the displayed expression is simply the union of the sets $(f|A_\alpha)^{-1}(V) = f_\alpha^{-1}(V)$. By the continuity of the functions f_α the sets on the right hand side of this expression are open, and therefore the union of these sets, which is just $f^{-1}(V)$, is also open, proving that f is continuous.

Second case. Many of the steps in the argument are the same so we shall concentrate on the differences. First of all, we need to replace the open subset V with a closed subset F. The same argument then shows that

$$f^{-1}(F) = \bigcup_{\alpha} f_\alpha^{-1}(F)$$

where each summand on the right hand side is closed by continuity. Since the union on the right hand side is finite, the union on the right hand side is again a closed subset, and this implies that $f^{-1}(F)$ is closed in X.■

II.4 : Cartesian products

(Munkres, §§ 15, 19)

Product constructions are useful in mathematics both as a means of describing more complicated objects in simpler terms (for example, expressing vectors in terms of magnitude and direction or resolution into x, y and z components) and also as the basis for considering quantities (formally, functions) whose values depend upon several variables.

Topological structures on finite products

Since product structures are less ambiguously defined for topological spaces as opposed to metric spaces, we shall begin with the former.

There are two ways of viewing Cartesian products with finitely many factors. Clearly one wants the product of the sets $A_1, \cdots, A_n$ to be the set of all ordered n-tuples (or lists of length n) having the form $(a_1, \cdots, a_n)$ where $a_i \in A_i$ for all i. Formally these can be described as functions α from $\{1, \cdots, n\}$ to $\cup_i A_i$ such that $\alpha(i) \in A_i$ for all i. Alternatively, one can view finite products as objects constructed inductively from 2-fold products by the recursive formula

$$A_1 \times \cdots \times A_{n+1} = (A_1 \times \cdots \times A_n) \times A_{n+1}$$

for all $n \geq 2$. It is an elementary exercise to see that the two formulations both result in equivalent concepts of ordered n-tuples with the property

$$(a_1, \cdots, a_n) = (b_1, \cdots, b_n) \iff a_i = b_i, \qquad \forall\, i\ .$$

Definition. Let $n \geq 2$ be an integer, and for each integer i between 1 and n let $(X_i, \mathbf{T}_i)$ be a topological space. The *product topology* on $X_1 \times \cdots \times X_n$ is the topology generated by all sets of the form $U_1 \times \cdots \times U_n$, where $U_i \in \mathbf{T}_i$ for all i. Frequently we shall write $\prod_i X_i$ to denote the product of the sets X_i and $\prod_i (X_i, \mathbf{T}_i)$ to denote the product topology; if the topologies on the factors are clear from the context and it is also clear that we want the product topology on $\prod_i X_i$ we shall frequently use the latter to denote the product space.

Before proceeding further we make a simple but useful observation.

PROPOSITION. *Every open subset in the product topology is a union of open subsets of the form $U_1 \times \cdots \times U_n$ where $U_i \in \mathbf{T}_i$ for all i.*

Proof. The topology generated by a family $\mathcal{F}$ of subsets consists of arbitrary unions of finite intersections of sets in $\mathcal{F}$, so it suffices to show that the latter is closed under finite intersections; by associativity and induction it suffices to check this for the intersections of pairs of subsets. But if we are given $\prod_i V_i$ and $\prod_i W_i$ where V_i and W_i are open in X_i for all i, then se have

$$\prod_i V_i \ \bigcap\ \prod_i W_i \quad = \quad \prod_i \left(V_i \cap W_i\right)$$

so that the family of products of open subsets is closed under finite intersections.

COROLLARY. *Let $\mathcal{B}_i$ be a base for the topology on X_i. Then every open subset in the product topology is a union of open subsets of the form $V_1 \times \cdots \times V_n$ where $V_i \in \mathcal{B}_i$ for all i.*

Proof. Since a union of unions is a union, it suffices to show this for open subsets of the form $U_1 \times \cdots \times U_n$, where $U_i \in \mathbf{T}_i$ for all i. Since each $\mathcal{B}_i$ is a base for $\mathbf{T}_i$, we can express each U_i as a union $\cup_{\alpha[i]} V_{\alpha[i]}$, and it follows that $U_1 \times \cdots \times U_n$ is equal to

$$\left(\bigcup_{\alpha[1]} V_{\alpha[1]}\right) \times \cdots \times \left(\bigcup_{\alpha[n]} V_{\alpha[n]}\right) =$$

$$\bigcup_{(\alpha[1], \ldots, \alpha[n])} \prod_i V_{\alpha[i]}$$

and hence $U_1 \times \cdots \times U_n$ is a union of products of the prescribed type.∎

The following result provides some motivation for the definition:

THEOREM. *For each $n \geq 2$ the topology on $\mathbf{R}^n$ with respect to the Euclidean metric is equal to the product topology associated to the family $(X_i, \mathbf{T}_i)$, where each space is the real line with the usual topology.*

Proof. *First step — Open sets in the product topology are open in the metric topology.* By the preceding corollary, every open subset in the product topology is a union of sets of the form $U_1 \times \cdots \times U_n$ where each U_i is an open interval in $\mathbf{R}$, and since arbitrary unions of metrically open sets are metrically open, it suffices to show that each product of open intervals $\prod_i (a_i, b_i)$ is open in the metric topology. Let $x = (x_1, \cdots, x_n)$ be a point in this product, and let $\varepsilon > 0$ be such that

$$\varepsilon \;<\; (x_i - a_i), \;\; (b_i - x_i)$$

for all i. If $y = (y_1, \cdots, y_n)$ satisfies $\mathbf{d}(y, x) < \varepsilon$ with respect to the standard Euclidean metric then we have

$$|y_i - x_i| \;\leq\; \mathbf{d}(y, x) \;<\; \varepsilon$$

for all i. It is an elementary exercise to check that this displayed inequality and the previous one imply $y_i \in (a_i, b_i)$ for all i. Therefore $\prod_i (a_i, b_i)$ is open in the metric topology because $N_\varepsilon(x)$ is contained in this product.

Second step — Open sets in the metric topology are open in the product topology. It suffices to show that for each $x \in \mathbf{R}^n$ and $\varepsilon > 0$ there is a $\delta > 0$ so that $\prod_i (x_i - \delta, x_i + \delta)$ is contained in $N_\varepsilon(x)$. To see this, note that given a metrically open subset U we may write it as

$$\bigcup_{x \in U} N_{\varepsilon(x)}(x)$$

for suitable positive real numbers $\varepsilon(x)$. The latter union then contains the union

$$\bigcup_{x \in U} \; \prod_i \, (x_i - \delta(x, \varepsilon), x_i + \delta(x, \varepsilon))$$

which in turn contains $\cup_x \{x\} = U$. Thus U is a union of sets having the form

$$\prod_i (x_i - \delta, x_i + \delta)$$

and hence is open in the product topology.

The proof of the assertion is best understood using a simple picture in the plane. Consider the open disk in the uv-plane consisting of all points for which $u^2 + v^2 < 1$. How large of an open square centered at the origin can one fit inside this open disk? In particular, one can ask this for a square whose sides are parallel to the coordinate axes. It turns out that the square in question has one vertex of the form $(1/\sqrt{2}, 1/\sqrt{2})$ and the other three vertices given by multiplying either or both coordinates of the latter by -1. Now suppose we are looking inside the open unit disk in coordinate 3-space. What is the largest cube in that case? The coordinates of the vertices turn out to be $\pm 1/\sqrt{3}$. One can then form an educated guess regarding the vertices for a maximal n-dimensional hypercube inside the unit n-dimensional hyperdisk.

Formally, proceed as follows. Given a fixed n and an arbitrary $\varepsilon > 0$, let

$$\delta = \frac{\varepsilon}{\sqrt{n}}$$

and consider the set $\prod_i (x_i - \delta, x_i + \delta)$. If y belongs to this set then $|y_i - x_i| < \delta$ for all i and therefore

$$\mathbf{d}(x,y) \quad = \quad \left(\sum_i |x_i - y_i|^2 \right)^{1/2} \quad < \quad \left(\sum_i \frac{\varepsilon^2}{n} \right)^{1/2} \quad = \quad \varepsilon$$

as required.■

General properties of (finite) product topologies

Given a sequence of sets $X_1, \cdots, X_n$ and an integer j between 1 and n, there is a map

$$p_j : \prod_i X_i \to X_j$$

called *projection onto the* i^{th} *coordinate* defined by the formula

$$p_j(x_1, \cdots, x_n) = x_j \ .$$

The following result characterizes the product topology in terms of these projections:

PROPOSITION. *Let* $(X_i, \mathbf{T}_i)$ *be a topological space for* $1 \le i \le n$, *and let* $\prod_i \mathbf{T}_i$ *denote the product topology on the product* $\prod_i X_i$. *Then* $\prod_i \mathbf{T}_i$ *is the unique smallest topology such that each projection map* p_j *is continuous.*

Proof. *Continuity of projections.* Let W be open in X_j. Then

$$p_j^{-1}(W) = \prod_i W_i$$

where $W_i = X_i$ if $i \neq j$ and $W_j = W$. This product set is open in the product topology and therefore p_j is continuous.

Minimality property. Suppose that $\mathbf{T}$ is a topology on $\prod_i X_i$ such that each p_j is continuous. Let $U = \prod_i U_i$ where U_i is open in X_i; we need to show that U is open with respect to $\mathbf{T}$. By the continuity of the projections we know that each set $p_j^{-1}(U_j)$ is open with respect to $\mathbf{T}$, and therefore the finite intersection

$$\bigcap_i p_i^{-1}(U_i) \;=\; \prod_i U_i$$

is open with respect to $\mathbf{T}$. Since every open set in the product topology is a union of sets of the form $\prod_i U_i$ it follows that every open set in the product topology is also open with respect to $\mathbf{T}$.■

COROLLARY. *Let $(X_i, \mathbf{T}_i)$ be a topological space for $1 \leq i \leq n$, let $(Y, \mathbf{W})$ be a topological space, for each i let $f_i : Y \to X_i$ be a set-theoretic function, and let $f : Y \to \prod_i X_i$ be the unique function such that $f \circ p_i = f_i$ for all i so that $f(y) = (f_1(y), \cdots, f_n(y))$. Then f is continuous (with respect to the product topology on $\prod_i X_i$) if and only if each function f_i is continuous.*

Proof. If f is continuous then the continuity of the projections p_i and the continuity of composites imply that each f_i is continuous because $f_i = p_i \circ f$.

Now suppose that each f_i is continuous. If we can show that the inverse image of each basic open subset $\prod_i U_i$ is under f is open, then since inverse images preserve unions it will follow that the inverse image of every open set under f is open and hence that f is continuous. As before we know that

$$\prod_i U_i = \bigcap_i p_i^{-1}(U_i)$$

and if we take inverse images (and use the fact that inverse images preserve intersections) then we have

$$f^{-1}\left(\prod_i U_i\right) = \bigcap_i f^{-1} \circ p_i^{-1}(U_i) = \bigcap_i f_i^{-1}(U_i)$$

and the latter is open because each f_i is continuous.■

By construction a product of open subsets is open in the product topology (where we are only dealing with finite products). The analogous statement for closed subsets is also true:

PROPOSITION. *Let $(X_i, \mathbf{T}_i)$ be a topological space for $1 \leq i \leq n$, and for each i suppose that F_i is a closed subset of X_i. Then $\prod_i F_i$ is a closed subset of $\prod_i X_i$ with respect to the product topology.*

Proof. This follows from the set-theoretic equation

$$\bigcap_i p_i^{-1}(F_i) \;=\; \prod_i F_i$$

the continuity of the projections p_i and the fact that inverse images of closed subsets with respect to a continuous function are closed.■

COROLLARY. *Let $(X_i, \mathbf{T}_i)$ be a topological space for $1 \leq i \leq n$, and for each i suppose that A_i is a subset of X_i. Then*

$$\prod_i \overline{A_i} \;=\; \overline{\prod_i A_i}\ .$$

Proof. The first set in the display contains the second because the first is a closed set containing the product of the A_i and the second is the smallest such closed subset. To see that the first is contained in the second, let b be a point in the product of the closure, and let U be an open subset of $\prod_i X_i$ that contains b; we need to prove that $U \cap \prod_i A_i \neq \emptyset$. Let $\prod_i V_i$ be a basic open subset that contains b and is contained in U. Since the coordinates of b satisfy $b_j \in \overline{A_j}$ for all j, it follows that $V_j \cap A_j \neq \emptyset$ for all j, and from this we have that $\prod_i V_i \cap \prod_i A_i \neq \emptyset$; since $U \supset \prod_i V_i$ it also follows that $U \cap \prod_i A_i \neq \emptyset$. But this means that b lies in the closure of $\prod_i A_i$.■

Projection maps also have the following important property.

OPENNESS OF PROJECTIONS. *The coordinate projection maps $p_j : \prod_i X_i \to X_j$ are open.*

Proof. The set-theoretic equality

$$g\left(\bigcup W_\alpha\right) \quad = \quad \bigcup_\alpha g(W_\alpha)$$

shows that it suffices to prove $p_j(W)$ is open if W is a basic open subset. But such a set has the form $\prod_i U_i$ where each U_i is open in X_i, and the image of this set under p_j is simply W_j.■

We have already given an example to show that coordinate projections are not necessarily closed; namely projection onto either coordinate is a continuous and open map from $\mathbf{R}^2$ to $\mathbf{R}$, but the image of the closed set of points satisfying the equation $xy = 1$ (geometrically a hyperbola whose asymptotes are the x- and y-axes) is $\mathbf{R} - \{0\}$, which is not a closed subset of the real line.

Products and morphisms

If we are given a sequence of set-theoretic functions $f_i : X_i \to Y_i$, then there one can define the *Cartesian product of morphisms*

$$F = \prod_i f_i \ : \ \prod_i X_i \quad \to \quad \prod_i Y_i$$

by the formula

$$F(x_1, \ \cdots \ , x_n) = (f_1(x_1), \ \cdots \ , f_n(x_n)\,)$$

or alternatively by the conditions

$$\pi_i^Y \circ F = f_i \circ \pi_i^X$$

where π_i^X and π_i^Y denote the i^{th} coordinate projections for $\prod_i X_i$ and $\prod_i Y_i$ respectively. Maps of this sort arise very frequently when one constructs new continuous functions out of old ones. If $n = 2$ one often describes such product maps using notation of the form

$$f_1 \times f_2 : X_1 \times X_2 \longrightarrow Y_1 \times Y_2$$

and similar notation is often used for other small values of n. Here are some properties of the product construction that are extremely elementary but also extremely important in many situations:

PROPOSITION. (i) *In the preceding notation, if each X_i and Y_i are topological spaces the function F is continuous with respect to the product topologies if and only if each f_i is continuous.*

(ii) *If each f_i is an identity map, then so is F.*

(iii) *Suppose we are also given sets Z_i and (set-theoretic) maps $g_i : Y_i \to Z_i$, and we set G equal to $\prod_i g_i$. Then*

$$G \circ F \quad = \quad \prod_i (g_i \circ f_i) \ .$$

The verifications of these statements are left to the reader as exercises.

Another important class of morphisms involving products are the maps that permute coordinates. We shall only discuss the simplest example here. Given two topological spaces X_1 and X_2 the *twist map*

$$\tau(X_1, X_2) : X_1 \times X_2 \longrightarrow X_2 \times X_1$$

is the map sending $(a, b) \in X_1 \times X_2$ to $(b, a) \in X_2 \times X_1$. These maps have the following elementary but important properties:

PROPOSITION. *If X_1 and X_2 are topological spaces, then $\tau(X_1, X_2)$ is continuous with respect to the product topologies. In fact, it is a homeomorphism whose inverse is given by $\tau(X_2, X_1)$.*

Proof. The second statement is purely set-theoretic and is elementary to verify. To check the continuity of the twist map, let p_1 and p_2 be the coordinate projections for the domain and let q_1 and q_2 be the coordinate projections for the codomain. We then have the identities

$$q_1 \circ \tau(X_1, X_2) \ = \ p_2 \ . \qquad q_2 \circ \tau(X_1, X_2) \ = \ p_1$$

and the continuity of $\tau(X_1, X_2)$ follows immediately from these.

Products and metric spaces

If $(X_i, \mathbf{d}_i)$ are metric spaces for $1 \leq i \leq n$, then it is possible to put metrics on $\prod_i X_i$ whose underlying topologies are the product topology. In fact, there are three particularly important product metrics. We shall describe three specific examples that are particularly significant. Let $x, y \in \prod_i X_i$ and express them in terms of coordinates as $(x_1, \ \cdots \ , x_n)$ and $(y_1, \ \cdots \ , y_n)$ respectively. Then the following formulas define metrics on the product:

$$\mathbf{d}^{\langle\infty\rangle}(x, y) = \max_i \{ \ \mathbf{d}_i(x_i, y_i) \ \} \ .$$

$$\mathbf{d}^{\langle 2\rangle}(x, y) = \left(\sum_i \ \mathbf{d}_i(x_i, y_i)^2 \right)^{1/2} \ .$$

$$\mathbf{d}^{\langle 1\rangle}(x, y) = \sum_i \ \mathbf{d}_i(x_i, y_i) \ .$$

The verification that each formula defines a metric is left to the reader as an exercise. We then have the following result:

PROPOSITION. *The topology determined by the metric $\mathbf{d}^{\langle\infty\rangle}$ is the product topology. Furthermore, the identity map from $(\prod_i X_i, \mathbf{d}^{\langle\alpha\rangle})$ to $(\prod_i X_i, \mathbf{d}^{\langle\beta\rangle})$ is uniformly continuous for all choices of $\alpha, \beta \in \{1, 2, \infty\}$.*

Proof. To verify the assertion about $\mathbf{d}^{\langle\infty\rangle}$ note that

$$\mathbf{d}^{\langle\infty\rangle}(x, y) < \varepsilon \quad \Longleftrightarrow \quad y_i \in N_\varepsilon(x_i), \ \forall i \ .$$

Thus the ε-neighborhood of x with respect to the $\mathbf{d}^{\langle\infty\rangle}$ metric is just

$$\prod_i N_\varepsilon(x_i) \ .$$

By previous results, a base for the product topology on $\prod_i X_i$ is given by open sets of the form

$$\prod_i N_{\varepsilon(i)}(x_i)$$

and this implies that the product topology contains the metric topology. On the other hand, for each j the projection map $p_j : \prod_i X_i \to X_j$ is uniformly continuous because

$$\mathbf{d}_j\left(p_j(x), p_j(y)\right) \quad = \quad \mathbf{d}_j(x_j, y_j) \quad \leq \quad \mathbf{d}^{\langle\infty\rangle}(x, y)$$

implies we can take $\delta = \varepsilon$ in the criterion for uniform continuity. This means that the metric topology contains the product topology, and therefore by the previous observations we see that the topologies are equal.

The uniform continuity statements are direct consequences of the following inequalities for nonnegative real numbers α_i for $1 \leq i \leq n$:

$$\max_i\{\ \alpha_i\ \} \quad \leq \quad \left(\sum_i \alpha_i^2\right)^{1/2} \quad \leq \quad \sum_i \alpha_i \quad \leq \quad n \cdot \max_i\{\ \alpha_i\ \}$$

The middle inequality is perhaps the least trivial, and it can be verified by squaring both sides and noting that the corresponding inequality holds for the squares. These inequalities imply that the identity maps

$$\left(\prod_i X_i, \mathbf{d}^{\langle\infty\rangle}\right) \to \left(\prod_i X_i, \mathbf{d}^{\langle 1\rangle}\right) \to \left(\prod_i X_i, \mathbf{d}^{\langle 2\rangle}\right) \to \left(\prod_i X_i, \mathbf{d}^{\langle\infty\rangle}\right)$$

are uniformly continuous (and in fact the δ corresponding to a given ε can be read off explicitly from the inequalities!), and of course the composites of any two or three consecutive maps from this diagram are also uniformly continuous.■

The following basic result on products and metric spaces is also worth mentioning:

PROPOSITION. *If X is a metric space then the distance function $\mathbf{d} : X \times X \to \mathbf{R}$ is continuous (where $\mathbf{R}$ has the usual topology).*

Proof. Let $(x, y) \in X \times X$, and view the product topology as coming from the maximum metric by the preceding discussion. Given $\varepsilon > 0$ suppose that $(u, v) \in X \times X$ satisfies

$$\max\big(\ \mathbf{d}(x, u),\ \mathbf{d}(y, v)\ \big) \ < \ \frac{\varepsilon}{2}\ .$$

Then several applications of the triangle inequality show that $\mathbf{d}(u, v) - \mathbf{d}(x, y) \ < \ \varepsilon$ and therefore $\mathbf{d}$ is in fact *uniformly continuous.*

Products and the Hausdorff Separation Property

We shall say that a topological space X has the *Hausdorff Separation Property* (or more simply, it is Hausdorff) if for each pair of distinct points $u, v \in X$ there are disjoint open subsets $U, V \subset X$ such that $u \in U$ and $v \in V$. As noted before, metric spaces have this property but it does not necessarily hold for an arbitrary topological space.

PROPOSITION. *In a Hausdorff space every one point subset is closed.*

Proof. Given $p \in X$ we shall show that $X - \{p\}$ is open if X is Hausdorff. Suppose that $y \in X - \{p\}$. Then there are disjoint open subsets U_y and V_y such that $p \in U_y$ and $y \in V_y$. Therefore we have

$$X - \{p\} \quad = \quad \bigcup_{y \neq p} \{y\} \quad \subset \quad \bigcup_{y \neq p} V_y \quad \subset \quad X - \{p\}$$

which implies that the last two subsets are equal, and thus $X - \{p\}$ is open because it is a union of open subsets.■

We now come to a result that has appeared on countless examinations:

THEOREM. *Given a set X, let the diagonal δ_X denote the set of all points $(u, v) \in X \times X$ such that $u = v$. Then X is Hausdorff if and only if Δ_X is closed in $X \times X$ with respect to the product topology.*

Proof. This follows because each of the statements listed below is equivalent to the adjacent one(s):

(1) X is Hausdorff.

(2) Given $(u, v) \in X \times X - \Delta_X$ there are open subsets $U, V \subset X$ such that $u \in U$, $v \in V$ and $\big(U \times V\big) \cap \Delta_X = \emptyset$.

(3) $X \times X - \Delta_X$ is open in $X \times X$ with respect to the product topology.

(4) Δ_X is closed in $X \times X$ with respect to the product topology.

Taken together, these prove the result.■

The theorem has an extremely important consequence:

PROPOSITION. *Let Y be a Hausdorff space, and let f and g be continuous functions from a topological space X to Y. Then the set*

$$E = \{\ x \in X \mid f(x) = g(x)\ \}$$

is closed in X.

Proof. If $H : X \to Y \times Y$ is the function defined by

$$H(x) = (f(x), g(x))$$

then we have already noted that H is continuous if f and g are continuous. It follows immediately that $E = H^{-1}(\Delta_Y)$. Since Y is Hausdorff we know that Δ_Y is closed in $Y \times Y$ and therefore its inverse image under H, which is simply E, is closed in X.■

SPECIAL CASE. *If f and g are continuous real valued functions on the unit interval* $[0, 1]$ *and $f(x) = g(x)$ for all rational points of* $[0, 1]$*, then $f = g$.*

Proof. The proposition shows that if f and g are continuous functions from the same space X into a Hausdorff space Y and $f|A = g|A$ then $f|\overline{A} = g|\overline{A}$. In this case A is the set of all rational points in $X = [0, 1]$ and $\overline{A} = X$. More generally this argument shows that if Y is a Hausdorff space, X is any space and $A \subset X$ is a subspace such that $\overline{A} = X$ and $f|A = g|A$, then $f = g$.■

It is easy to construct counterexamples to the conclusion of the proposition if the codomain is not Hausdorff. Suppose that X and Y both have the associated indiscrete topologies where both sets have at least two elements. Then every function from X to Y is continuous, and every nonempty subset $A \subset X$ is dense (*i.e.*, $\overline{A} = X$). Large families of counterexamples can be constructed in this manner; details are left to the reader as an exercise.

Infinite products

In earlier decades infinite products of topological spaces received a great deal of attention. We shall not deal with such objects extensively here, but it seems worthwhile to say a little about them for the sake of completeness and to avoid some natural possibilities for misunderstandings.

Given an indexed family of sets X_α with indexing set A, the set-theoretic cartesian product

$$\prod_{\alpha \in A} X_\alpha$$

may be defined formally as the set of all set-theoretic functions x from A to $\cup_\alpha\ X_\alpha$ such that $x(\alpha) \in X_\alpha$ for all α. This captures the intuitive ideas that the elements of the cartesian product are given by the coordinates x_α and that two elements are equal if and only if all their coordinates are equal.

The Axiom of Choice in set theory is equivalent to the statement that *if each of the sets X_α is nonempty, then so is their cartesian product $\prod_\alpha X_\alpha$.*

As in the case of finite products there are projection maps $p_\beta : \prod_\alpha X_\alpha \to X_\beta$ defined by $p_\beta(x) = x(\beta)$.

Assume now that we have an indexed family of topological spaces $(X_\alpha, \mathbf{T}_\alpha)$. The crucial property of the product topology for $\prod_\alpha X_\alpha$ will be that *it is the unique smallest topology such that every projection map p_β is continuous.*

The preceding condition implies that the product topology should be generated by all sets of the form $p_\beta^{-1}(U_\beta)$ where U_β is open in X_β, and thus the product topology will be arbitrary unions of finite intersections of such sets.

PROPOSITION. *A base for the product topology is given by all open subsets of the form $\prod_\alpha U_\alpha$ where each U_α is open in X_α* **AND** *$U_\alpha = X_\alpha$ for* **all but finitely many** *α.*

Proof. The subsets described in the proposition are finite intersections of sets having the form $p_\beta^{-1}(U_\beta)$; specifically, if Γ is a finite subset of A then the subsets in the proposition have the form

$$\bigcap_{\gamma \in \Gamma} p_\gamma^{-1}(U_\gamma)\ . \blacksquare$$

Another topology on the product is the so-called *box topology* generated by all subsets of the form $\prod_\alpha U_\alpha$ where U_α is an arbitrary open subset of X_α. For finite products these yield the same topology, but this is not true for infinite products. A fairly detailed discussion of the differences appears in Section 19 of Munkres.

Final remark. Theorem 19.6 on page 117 of Munkres gives a fundamentally important property of the product topology (in both the finite and infinite cases).

Finally, here are some facts about finite products that carry over to infinite products. The proofs are essentially the same.

OPENNESS OF PROJECTIONS. *For each β in the indexing set the coordinate projection maps*

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

are open.■

PRODUCTS OF CLOSED SUBSETS. *Let $(X_\alpha, \mathbf{T}_\alpha)$ be a topological space for $\alpha \in A$, and for each α suppose that F_α is a closed subset of X_α. Then $\prod_\alpha F_\alpha$ is a closed subset of $\prod_\alpha X_\alpha$ with respect to the product topology.*■

COROLLARY. *Let $(X_\alpha, \mathbf{T}_\alpha)$ be a topological space for $\alpha \in A$, and for each α suppose that F_α is a closed subset of X_α. Then*

$$\prod_\alpha \overline{A_\alpha} = \overline{\prod_\alpha A_\alpha} \ .\blacksquare$$

MAPPINGS INTO PRODUCTS. *Let A be a set, and for each $\alpha \in A$ let $f_\alpha : X_\alpha \to Y_\alpha$ be a set-theoretic map. Then there is a unique map*

$$F = \prod\nolimits_\alpha f_i \ : \ \prod_\alpha X_\alpha \quad \to \quad \prod\nolimits_\alpha Y_\alpha$$

defined by the conditions

$$\pi_\alpha^Y \circ F = f_\alpha \circ \pi_\alpha^X$$

where π_α^X and π_α^Y denote the i^{th} coordinate projections for $\prod_\alpha X_\alpha$ and $\prod_\alpha Y_\alpha$ respectively. This map is continuous if and only if each f_α is continuous, and it is the identity map if each f_α is an identity map. Finally, if we are also given sets Z_α and (set-theoretic) maps $g_\alpha : Y_\alpha \to Z_\alpha$, and we set G equal to $\prod_\alpha g_\alpha$. Then

$$G \circ F \quad = \quad \prod\nolimits_\alpha (g_\alpha \circ f_\alpha) \ .$$

Finally we mention one more that is an exercise in Munkres (Theorem 19.4, page 116; see also Exercise 3 on page 118). A proof (probably not the best one) for products of two spaces appears in Section III.1 of these notes.

PRODUCTS AND THE HAUSDORFF PROPERTY. *Let A be a nonempty set, and supposed that X_α is a Hausdorff topological space for each $\alpha \in A$. Then $\prod_\alpha X_\alpha$ is also a Hausdorff space.*■

III. Spaces with special properties

We have seen that one can derive a relatively sizable amount of information simply from the axioms for metric and topological spaces. However, it should not be surprising that oned needs to impose further conditions on spaces in order to prove more substantial results, including abstract versions of the Maximum Value Theorem and Intermediate Value Theorem from single variable calculus. It turns out that these two results relies on a separate basic properties of the open subsets in a closed interval. The underlying concepts are known as *compactness* and *connectedness*, and they are treated in this unit.

In calculus it is also important to know that certain infinite series have meaningful sums, and indeed one reasons that mathematicians tightened their standards of logical rigor in the nineteenth century was to analyze the validity of certain strange and unanticipated results that arose from casual manipulations with infinite series; some of the results were justified, but others were not (this was expected because some of the formulas contradicted each other). One abstract version of the basic condition guaranteeing convergence of reasonable infinite series is called *completeness*, and it is also discussed in this unit along with some important geometrical an analytical implications (however, the applications to analysis go far beyond the scope of this course).

III.1 : Compact spaces – I

(Munkres, §§ 26, 27)

One of the most fundamental properties of continuous functions on closed intervals is that they have maximum and minimum values. In contrast, a continuous function on an open or half open interval does not necessarily have this property. In most real variables courses, the existence of maximum and minimum values is established with the help of the Heine-Borel-Lebesgue Theorem (sometimes the third name is dropped when referring to this result, and sometimes the first name is dropped). The conclusion of this result is so important that it has become incorporated into a definition. However, before proceeding to the main result we need a preliminary concept.

Definition. If X is a topological space and $\mathcal{U} = \{U_\alpha\}$ is a family of open subsets of X, we say that $\mathcal{U}$ is an *open covering* of X if $\cup_\alpha U_\alpha = X$. A subfamily $\mathcal{V} \subset \mathcal{U}$ is said to be a *subcovering* if $\cup_\beta V_\beta = X$, where $\mathcal{V} = \{V_\beta\}$.

Definition. A topological space is said to be *compact* if every open covering has a finite subcovering.

The main point of the Heine-Borel-Lebesgue Theorem is that closed intervals in the real line are compact. An abstract version of this result is established as Theorem 27.1 on pages 172–173 of Munkres (see Corollary 27.2 on the second of these pages for the case of interest to us here). In fact, one has the following characterization of compact subsets of the real line.

Characterization of compact subsets. *A subset K of the real line is compact if and only if it is closed and bounded.*

Several portions of the proof are true under much more general conditions, so we shall establish these first.

THEOREM. *If A is a compact subset of a Hausdorff space, then A is closed in X.*

Proof. We use the Hausdorff Separation Property to show that $X - A$ is open.

Let $y \in X - A$, then for each $a \in A$ we have $y \neq a$, and therefore by the Hausdorff Separation Property there are open sets $U_{(a,y)}$ and $V_{(a,y)}$ (in X) containing y and a respectively such that $U_{(a,y)}$ and $V_{(a,y)}$ are disjoint. The family of subsets $\{A \cap V_{(a,y)}\}$ is an open covering of $A$and thus has a finite subcovering

$$A \cap V_{(a_1,y)} \ , \ \cdots \ , \ A \cap V_{(a_k,y)} .$$

By construction we have $A \subset V_{(a_1,y)} \cup \cdots \cup V_{(a_k,y)}$, and thus if we take

$$U_y \ = \ U_{(a_1,y)} \cap \cdots \cap U_{(a_k,y)}$$

then U_y is an open subset containing y and $U_y \cap A = \emptyset$. A (by now) familiar argument shows that $X - A = \cup_y \ U_y$ and hence that $X - A$ is open.■

PROPOSITION. *If A is a closed subset of a compact topological space X, then A is compact.*

Proof. Let $\mathcal{U} = \{U_\alpha\}$ be an open covering of A; choose open sets V_α in X so that $U_\alpha = A \cap V_\alpha$, and let $\mathcal{V}$ be the open covering of X given by the sets V_α together with $X - A$. By compactness of X there is a finite subcovering, which we may as well assume contains $X - A$ as well as open subsets $V_{\alpha(1)}, \ \cdots \ , V_{\alpha(k)}$. It then follows that the corresponding subsets $U_{\alpha(1)}, \ \cdots \ , U_{\alpha(k)}$ for a finite subcovering of A.■

THEOREM. *If $f : X \to Y$ is continuous and X is compact, then its image $f(X)$ is also compact.*

Proof. Let $\{U_\alpha\}$ be an open covering of $f(X)$, and choose open subsets V_α in Y so that $U_\alpha = f(X) \cap V_\alpha$. Then the sets

$$W_\alpha = f^{-1}(U_\alpha) = f^{-1}(V_\alpha)$$

form an open covering of X, so there is a finite subcovering of X having the form $W_1, \cdots, W_k$. But

$$f(W_j) = f\left(f^{-1}(U_j)\right) = U_j$$

and therefore the sets $U_1, \cdots, U_k$ for a finite (open) subcovering of $f(X)$.■

Proof of characterization of compact subsets of the real line. ($\Longrightarrow$) By the first result above a compact subset of the real line is closed. To see that it is bounded, consider the open covering given by the intersections of A with the open intervals $(-n, n)$ where n runs through the positive integers. If A were not bounded, this open covering would not have a finite subcovering, so A must be bounded as claimed.

($\Longleftarrow$) If A is bounded then A is a subset of some closed interval $[-M, M]$. Since A is closed in $\mathbf{R}$, it is also closed in the compact set $[-M, M]$, and therefore A is compact by the second of the results above.■

With a little additional effort one can modify the proof of the Heine-Borel-Lebesgue Theorem to show that every box-shaped subset of $\mathbf{R}^k$ of the form

$$[a_1, b_1] \times \ \cdots \ \times [a_k, b_k]$$

is compact (see Theorem 2.40 on page 39 of Rudin's book); we shall also give an alternate proof later in the course. This in turn yields an extension of the characterization of compact subsets from $\mathbf{R}$ to $\mathbf{R}^k$ for all positive integers k.

COROLLARY. *A subset A of $\mathbf{R}^k$ is compact if and only if it is closed and bounded.*

Sketch of proof. $(\Longrightarrow)$ The subset A is closed for the same reasons as before. If f_j denotes the restriction of the j^{th} coordinate function to A, then $f_j(A)$ is a compact and hence bounded subset of $\mathbf{R}$. If we choose $M > 0$ so that $\cup_j\ f_j(A) \subset [-M, M]$, then $a \in A \Rightarrow a = (a_1, \cdots a_k)$ where $|a_j| \leq M$ for all j. Hence A is bounded.

$(\Longleftarrow)$ If A is closed and bounded then for some $M > 0$ we know that A is a closed subset of the compact set

$$[-M, M] \times \ \cdots \ \times [-M, M]$$

and therefore A is compact.∎

The following consequence is a significant generalization of a fundamental result from calculus.

COROLLARY. *If X is compact and $f : X \to \mathbf{R}$ is continuous, then f attains maximum and minimum values on X.*

Proof. This reduces to showing the following: *If $A \subset \mathbf{R}$ is compact, so that it is closed and bounded, then both the least upper bound and greatest lower bound of A belong to A.* We shall only verify the statement regarding the least upper bound; the other statement follows by reversing the directions of all inequalities.

Let M be the least upper bound of A (which exists because A is bounded). The for every positive integer n we can find a point $a_n \in A$ such that

$$M - \frac{1}{n} < a_n \leq M$$

where the second inequality is true because M is an upper bound for A. It follows immediately that $M = \lim_{n\to\infty} a_n$, and since A is closed it follows that $M \in A$.

The Finite Intersection Property

There is a characterization of compactness in terms of closed subsets. Given a family $\mathcal{A} = \{F_\alpha\}$ of closed subsets of a topological space, we shall say that $\mathcal{A}$ has the *finite intersection property* if

$$F_{\alpha(1)} \cap \ \cdots \ \cap F_{\alpha(k)} \neq \emptyset$$

for all finite subcollections

$$\{\ F_{\alpha(1)},\ \cdots\ F_{\alpha(k)}\ \} \subset \mathcal{A}\ .$$

THEOREM. *A topological space X is compact if and only if for every family of closed subsets $\mathcal{A} = \{F_\alpha\}$ with the finite intersection property we have $\cap_\alpha F_\alpha \neq \emptyset$.*

A proof of this result and some further remarks appear on page 170 of Munkres.∎

Compactness and continuous mappings

We have already noted that continuous map that is 1–1 and onto is not necessarily a homeomorphism. However, if one puts suitable hypotheses on the domain or codomain it is sometimes possible to prove that a 1–1 onto continuous map is a homeomorphism without checking the continuity of the inverse directly. In particular, this holds for compact metric spaces.

PROPOSITION. *Suppose that $f : X \to Y$ is a continuous map from a compact topological space to a Hausdorff space. Then f is a closed mapping.*

Proof. Suppose that A is closed in X. Then A is compact, and therefore $f(A)$ is also compact in Y. But since Y is a Hausdorff space this implies that $f(A)$ is closed in Y.■

COROLLARY. *If $f : X \to Y$ is a continuous and $1-1$ onto map from a compact topological space to a Hausdorff space, then f is a homeomorphism.*■

Products and compactness

The following sort of question arises frequently in mathematics:

PROPERTIES OF PRODUCTS. *If X and Y are systems that have some property* **P** *and there is a reasonable notion of direct product $X \times Y$, does this product also have property* **P***?*

Here are some examples involving topological spaces for which there is a positive answer:

1. Suppose that X and Y are discrete spaces. Then $X \times Y$ is also discrete. (**Proof:** If $(x, y) \in X \times Y$ then $\{x\}$ is open in X and $\{y\}$ is open in Y. Therefore

$$\{(x, y)\} = \{x\} \times \{y\}$$

is open in $X \times Y$, and since x and y are arbitrary this means that every subset of $X \times Y$ is open.)

2. Suppose that X and Y are spaces in which one point subsets are closed. Then the same is true for $X \times Y$; the proof is analogous to the previous one.

3. If X and Y are finite, then the same is true for $X \times Y$.

4. If X and Y are homeomorphic to metric spaces, then the same is true for $X \times Y$. In fact, we have given three ways of constructing a metric on the product.

5. If X and Y are Hausdorff spaces, then $X \times Y$ is also Hausdorff. (**Proof:** One way of doing this is to use the characterization of a Hausdorff space W in terms of the diagonal Δ_W in $W \times W$ being closed. Let **Shuff** be the "middle four shuffle map"

$$X \times X \times Y \times Y \longrightarrow X \times Y \times X \times Y$$

that sends (x_1, x_2, y_1, y_2) to (x_1, y_1, x_2, y_2). This map is continuous because its projections onto the four factors are continuous, and the same is true for the inverse map which sends (x_1, y_1, x_2, y_2) to (x_1, x_2, y_1, y_2). Since **Shuff** is a homeomorphism, it follows that

$$\Delta_{X \times Y} = \mathbf{Shuff}\big(\Delta_X \times \Delta_Y\big)$$

is closed in $(X \times Y) \times (X \times Y)$. This completes the argument.)

In contrast, here is one example where there is a negative answer:

6. If X and Y are homeomorphic to subsets of the real line, the product $X \times Y$ is not necessarily homeomorphic to a subset of the real line. An easy counterexample is given by taking $X = Y = \mathbf{R}$. We shall prove this when we discuss connectedness later in the course.

THEOREM. *The product of finitely many compact spaces is compact.*

Using the canonical homeomorphism

$$(X \times Y) \times Z \cong X \times Y \times Z$$

and finite induction we can reduce the proof to the case of a product of two compact spaces. The proof depends upon the following result which is also useful in other contexts.

TUBE LEMMA. *Let X and Y be topological spaces such that X is compact, let $y \in Y$, and let $\mathcal{W} = \{ W_\alpha \}$ be a family of open subsets of $X \times Y$ such that $X \times \{y\}$ is contained in $\cup_\alpha W_\alpha$. Then there is a finite open covering $\mathcal{U}(y) = \{ U_i \}$ of X and an open subset $V(y)$ of Y containing y such that each product set $U_i \times V(y)$ is contained in some W_α.*

Proof of Tube Lemma. First of all, we claim that $X \times \{y\}$ is homeomorphic to X and therefore is compact. To see this, consider the map $f : X \to X \times \{y\}$ defined by $h(x) = (x, y)$. The projections onto the factors are the identity and the constant map, and therefore h is continuous. Projection onto the X factor yields a continuous inverse to h. Maps of this form are often called *slice inclusions.*

For each $x \in X$ let $W(x)$ be an open subset in $\mathcal{W}$ such that $x \in W(x) \times \{y\}$. Let U_x and V_x be open subsets of X and Y respectively such that

$$(x, y) \in U_x \times V_x \subset W(x) \ .$$

Then $\mathcal{U} = \{ U_x \}$ is an open covering of X and hence there is a finite subcovering

$$\mathcal{U}(y) = \{ U_{x_1}, \ \cdots \ , \ U_{x_n} \} \ .$$

If $V(y) = \cap_i \ V_{x_i}$, it follows that

$$U_{x_i} \times V(y) \quad \subset \quad U_{x_i} \times V_{x_i} \quad \subset \quad W(x_i)$$

which proves the lemma.∎

A picture illustrating this proof is given in the files `tubelemma.*` in the course directory for various formats $*$.

Proof of the Theorem. Let $\mathcal{W} = \{ W_\alpha \}$ be an open covering of $X \times Y$, and for each $y \in Y$ let $\mathcal{W}(y) \subset \mathcal{W}$ be a family that covers $X \times \{y\}$.

Given $y \in Y$, let $\mathcal{U}(y)$ and $V(y)$ be associated to $\mathcal{W}(y)$ as in the Tube Lemma. The sets $V(y)$ form an open covering of Y and therefore there is a finite subcovering $\{ V(y_1), \ \cdots \ , \ V(y_m) \}$. Then the finite family of sets

$$\mathcal{U} = \{ \ U_\beta \times V(y_j) \mid U_\beta \in \mathcal{U}(y_j) \ \}$$

is a finite open covering of $X \times Y$ and for each set in the family there is some $W_{\gamma(\beta,j)}$ in $\mathcal{W}$ such that

$$U_\beta \times V(y_j) \subset W_{\gamma(\beta,j)} \ .$$

The finite collection of sets $W_{\gamma(\beta,j)}$ is the desired finite subcovering of $\mathcal{W}$.■

Compactness and infinite products

The preceding result on compactness of products extends to infinite products provided one assumes the Axiom of Choice (in fact, the statement of the theorem is equivalent to the latter). This was established by A. N. Tychonoff and is known as Tychonoff's Theorem. The result has fundamental applications in several mathematical contexts, and perhaps the most important involve functional analysis (*e.g.*, the Banach-Alaoglu Theorem; see pages 68–69 of Rudin, *Functional Analysis*, for a statement and proof). Proofs of Tychonoff's Theorem and a crucial preliminary result appear on pages 233–235 of Munkres.

Compact metric spaces

If a compact topological space is determined by a metric, then many additional statements can be made. For example, we have the following generalization of the boundedness property:

PROPOSITION. *If X is a compact metric space then there is a constant $K > 0$ such that $\mathbf{d}(x,y) \leq K$ for all $x, y \in X$.*

Proof. By Example 4 in the list of examples of continuous functions in the previous note, for a fixed $z \in X$ the function $f(x) = \mathbf{d}(x,z)$ is (uniformly) continuous. Let M be the maximum value of this function. Given two points $x, y \in X$ the triangle inequality now implies that

$$\mathbf{d}(x,y) \leq \mathbf{d}(x,z) + \mathbf{d}(y,z) \leq M + M = 2M$$

and therefore we can take $K = 2M$.■

Another important and much deeper property of a closed interval in the real line is that every infinite sequence in the interval has a convergent subsequence (the Bolzano-Weierstrass Theorem). This property also holds for compact metric spaces.

THEOREM. *If X is a compact metric space, then every infinite sequence in X has a convergent subsequence.*

Later in this course we shall prove a converse to this result.

Proof. Let $\{a_n\}$ be an infinite sequence in X, and suppose it has no convergent subsequence. If the sequence takes only finitely many values, then at least one of them occurs infinitely many times, and thus one can find a convergent subsequence, so we may as well assume that the sequence takes infinitely many distinct values. Let $A \subset X$ be the set of all these values.

We claim that $\mathbf{L}(A) = \emptyset$; suppose that $b \in \mathbf{L}(A)$. One can then recursively construct a subsequence that converges to b as follows. Suppose that the first r terms of the subsequence $a_{n(k)}$ have been defined so that $\mathbf{d}(b, a_{n(j)}) < \frac{1}{k}$. Let U_{r+1} be the set containing b obtained by taking the open disk $N_{1/(r+1)}(b)$ and removing all elements a_ℓ of the original sequence for $\ell \leq n(k)$ that are not equal to b. Since one point (and hence finite) subsets of a metric space are closed, it follows that U_{r+1} is open. Therefore, by the definition of a limit point there is some $a \in A$ such that $a \neq b$ and $a \in U_{r+1}$. By construction this point has the form a_m for some $m > n(r)$, and we set $m = n(r+1)$. This yields a subsequence whose limit is b.

Since $\mathbf{L}(\mathbf{A})$ is empty it follows that it is contained in A and therefore A is closed. Since X is compact, so is A.

We shall obtain a contradiction by showing that the infinite set A is not compact. If $a \in A$, then since $a \notin \mathbf{L}(A)$ we can find an open subset U_a in X such that $A \cap U_a = \{a\}$. It follows that every one point subset of A is open in the subspace topology and hence that every subset is open in the subspace topology. Since A is infinite, it follows that the open covering of A by one point subsets does not have a finite subcovering, which shows that A is not compact.

The contradiction means that our original assumption — the existence of an infinite sequence in X with no convergent subsequence — is incorrect, and therefore it follows that every infinite sequence in X has a convergent subsequence.■

The next result plays an important role in several analytic and geometric considerations. In particular, we shall use it to show that a continuous map from a compact metric space to another metric space is uniformly continuous.

LEBESGUE'S COVERING LEMMA. *Let X be a compact metric space, and let $\mathcal{U}$ be an open covering of X. Then there is a number $\eta > 0$ such that for every pair of points $x, y \in X$ such that $\mathbf{d}(x, y) < \eta$ there is an open set V in $\mathcal{U}$ such that $x, y \in V$.*

Proof. For each $p \in X$ there is an $\varepsilon(x) > 0$ such that $N_{2\varepsilon(x)}(x)$ is contained in some element of $\mathcal{U}$. Let $W_x = N_{\varepsilon(x)}(x)$.

The family $\mathcal{W} = \{W_x\}$ is an open covering of X, so there is a finite subcovering of the form

$$\{W_{x_1}, \ \cdots \ , \ W_{x_k}\} \ .$$

Let $\varepsilon_j > 0$ be the positive number associated to x_j, and let η be the minimum of the positive numbers $\varepsilon_1, \ \cdots \ , \varepsilon_k$.

Suppose now that $\mathbf{d}(x, y) < \eta$. Choose i so that $x \in W_{x_i}$. Then by the triangle inequality we have

$$\mathbf{d}(y, x_i) \leq \mathbf{d}(y, x) + \mathbf{d}(x, x_i) < \eta + \varepsilon_i \leq \varepsilon_i + \varepsilon_i = 2\varepsilon_i$$

which shows that $y \in N_{2\varepsilon(x_i)}(x_i)$. The latter set is contained in some set V from the family $\mathcal{U}$ and by construction it also contains x.■

A number η satisfying the conditions of the conclusion of the preceding result is called a *Lebesgue number* for the open covering. It is easy to see this result fails for noncompact metric spaces. For example, consider the open covering of the set of the open unit interval $(0, 1)$ given by the open subintervals

$$\left(\frac{1}{2^{k+1}}, \frac{1}{2^{k-1}}\right)$$

where k runs through the positive integers.

The uniform continuity property is an immediate consequence of the Lebesgue Covering Lemma.

THEOREM. *Let X and Y be metric spaces where X is compact, and let $f : X \to Y$ be continuous. Then f is uniformly continuous.*

Proof. Let $\varepsilon > 0$ be arbitrary, and for each $y \in Y$ consider the open set $N_{\varepsilon/2}(y) \subset Y)$. By continuity the sets $f^{-1}\left(N_{\varepsilon/2}(y)\right)$ form an open covering of X, and by the compactness of X this open covering has a Lebesgue number η. Suppose now that $u, v \in X$ satisfy $\mathbf{d}(u, v) < \eta$. Then there is some $y \in Y$ such that $u, v \in f^{-1}\left(N_{\varepsilon/2}(y)\right)$. It follows that $f(u), f(v) \in N_{\varepsilon/2}(y)$, and by the triangle inequality we have

$$\mathbf{d}\left(f(u), f(v)\right) \leq \mathbf{d}\left(f(u), y\right) + \mathbf{d}\left(y, f(u)\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

so that

$$\mathbf{d}(u, v) < \eta \Longrightarrow \mathbf{d}\left(f(u), f(v)\right) < \varepsilon$$

for all $u, v \in X$.■

III.2 : Complete metric spaces

(Munkres, §§43, 45)

Infinite series play an extremely important role in the theory and applications of the real number system; this is particularly apparent in the computational view of real numbers in terms of infinite decimal expansions and the use of power series to work with large families of functions in calculus. However, some care is needed in working with infinite series to ensure the reliability of any calculations done with them; in particular, it is necessary to know whether or not a series actually produces a meaningful real number (in other words, it *converges*). Many of the important criteria for convergence of infinite series in calculus rely on the property of real numbers known as completeness. The definition of this concept requires an important preliminary notion.

Definition. Let X be a metric space. A sequence $\{ a_n \}$ in X is a *Cauchy sequence* if for every $\varepsilon > 0$ there is a positive integer N such that $m, n > M$ implies $\mathbf{d}(a_m, a_n) < \varepsilon$.

PROPOSITION. *Every convergent sequence is a Cauchy sequence.*

Proof. Suppose that $\lim_{n\to\infty} a_n = L$. Given a positive real number ε, choose M such that $n > M$ implies $\mathbf{d}(L, a_n) < \varepsilon/2$. The triangle inequality then implies that

$$\mathbf{d}(a_m, a_n) \leq \mathbf{d}(a_m, L) + \mathbf{d}(L, a_n) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

and therefore $\{ a_n \}$ is a Cauchy sequence.■

It is easy to find examples of Cauchy sequences in metric spaces that do not have limits. For example, take X to be the open interval $(0, 1)$ and consider the sequence $a_n = \frac{1}{n}$. Of course this sequence does have a limit if one expands X to the closed unit interval. This is a special case of the following basic property of the real numbers:

THEOREM. *Every Cauchy sequence in* $\mathbf{R}^k$ *converges for all* $k \geq 1$.■

References for the proof of this fact include Theorem 43.2 on page 265 of Munkres and Theorem 3.11 on pages 53–54 of Rudin. The relevance of this theorem to infinite series is explained in a mathematically rigorous fashion on pages 58–78 of Rudin (see also Exercise 15 on page 81).

Definition. A metric space X is *complete* if every Cauchy sequence in X converges.

One of the main objective of this section is to show that **every** Cauchy sequence in a metric space X converges in some larger metric space Y containing X (isometrically) as a subspace.

Properties of complete metric spaces

Complete metric spaces behave like compact topological spaces in several respects. Of course, there are also some major differences; for example, the real numbers are complete but not compact, and the real numbers are also homeomorphic to the noncomplete subspace $(-1, 1)$, say by the map $f : (-1, 1) \to \mathbf{R}$ defined using the formula

$$f(x) = \frac{x}{1 - |x|}$$

but one can begin the analogies with the following result:

PROPOSITION. *A compact metric space is complete.*

Proof. Let $\{ a_n \}$ be a Cauchy sequence in X. By previous results we know this sequence has a convergent subsequence $\{ a_{n(k)} \}$. Let $\lim_{k\to\infty} a_{n(k)} = L$; we claim that $\lim_{n\to\infty} a_n = L$. Given $\varepsilon > 0$ choose M_1 so that $k > M_1$ implies $\mathbf{d}(L, a_{n(k)}) < \varepsilon/2$, and choose M_2 so that $m, n > M_2$ implies $\mathbf{d}(a_m, a_n) < \varepsilon/2$. Take M to be the larger of M_2 and $n(M_1)$. Then the triangle inequality implies that $\mathbf{d}(L, a_n) < \varepsilon$ if $n > M$ (we have used this sort of argument many times already; at this point the reader should try to fill in the details as an exercise).■

One also has the following useful result about closed subsets and complete metric spaces:

PROPOSITION. *Let X be a metric space and let $A \subset X$. If A is complete in the subspace metric, then A is closed in X. Conversely, if A is closed in X and X is complete, then A is complete in the subspace metric.*

Proof. Suppose first that A is complete with respect to the subspace metric and that $\{ a_n \}$ is a sequence in A with a limit $L \in X$. By a previous result the sequence is a Cauchy sequence, and therefore it has a limit in A. Since the limit of a convergent sequence is unique, this limit in A must be the point L.

Now suppose that A is closed in X where X is complete. If $\{ a_n \}$ is a Cauchy sequence in A then the completeness of X implies that the sequence has a limit $L \in X$. Since A is a closed subspace, this means that $L \in A$, so that the Cauchy sequence in A has a limit in A. Therefore A is complete.■

The following result on completeness and products reflects another similarity with compact topological spaces.

PROPOSITION. *If X and Y are complete metric spaces, then so is $X \times Y$ with respect to each of the three product metrics.*

Proof. Let $\mathbf{d}^{\langle q \rangle}$ be a product metric where $q = 1$, 2 or ∞. By construction each of the projection maps

$$\left(X \times Y, \mathbf{d}^{\langle q \rangle}\right) \to (X, \mathbf{d}_X)\,, \quad (Y, \mathbf{d}_Y)$$

sends points whose distance in $X \times Y$ is ε to points in X and Y with the same property.

It follows that if $\{ (x_n, y_n) \}$ is a Cauchy sequence in $X \times Y$ then $\{ x_n \}$ and $\{ y_n \}$ are Cauchy sequences in X and Y respectively. By the completeness of X and Y each of these sequences has a limit, and we shall call these limits x and y respectively. To complete the proof we need to show that

$$\lim_{n\to\infty} (x_n, y_n) = (x, y)$$

with respect to each of the three product metrics.

The previously established inequalities

$$\mathbf{d}^{\langle \infty \rangle} \leq \mathbf{d}^{\langle 2 \rangle} \leq \mathbf{d}^{\langle 1 \rangle}$$

show that it suffices to prove the limit statement for $\mathbf{d}^{\langle 1 \rangle}$. Given $\varepsilon > 0$, choose M so that $n > M$ implies that both $\mathbf{d}_X(x, x_n)$ and $\mathbf{d}_Y(y, y_n)$ are less than $\varepsilon/2$; then the standard arguments (supply them!) show that $n > M$ implies

$$\mathbf{d}^{\langle 1 \rangle}\Big(\, (x, y)\,,\, (x_n, y_n) \Big) < \varepsilon$$

and this proves the statement(s) about limits.■

Function spaces

The next result provides some important examples of complete metric spaces.

PROPOSITION. *If X is a set then* $\mathbf{BF}(X)$ *is a complete metric space with respect to the norm described previously; furthermore, if X is a topological space then* $\mathbf{BC}(X)$ *is a closed and hence complete subset of* $\mathbf{BF}(X)$*.*

Proof. We shall first prove that $\mathbf{BF}(X)$ is complete. If $\{ f_n \}$ is a Cauchy sequence in $\mathbf{BF}(X)$, then by definition we know that $\{ f_n(x) \}$ is a Cauchy sequence of real numbers for each $x \in X$. Since we know that $\mathbf{R}$ is complete, it follows that for each $x \in X$ there is a real number $f(x)$ to which $\{ f_n(x) \}$ converges. We need to prove two things. First, we need to show that f is bounded. Next we have to show that $\lim_{n\to\infty} f_n = f$ in $\mathbf{BF}(X)$.

To verify that $\{ f_n(x) \}$ is bounded, choose M so that $m, n > M$ implies $|f_m - f_n| < 1$. Then for all n and x we know that

$$|f(x)| \le \max\{|f_i(x)|,\ i < M;\ |f_M(x)| + 1\}$$

and therefore we also have

$$|f(x)| \le \max\{|f_i|,\ i < M;\ |f_M| + 1\}$$

so that the left hand side is bounded by the right hand side for all $x \in X$ and therefore $f \in \mathbf{BF}(X)$.

To show that $\{ f_n \}$ converges to f in $\mathbf{BF}(X)$, let $x \in X$ be arbitrary, and given *varepsilon* > 0 choose M so that $m, n > M$ implies $|f_m - f_n| < \varepsilon/2$. Since $\{ f_n(x) \}$ converges to $f(x)$, there is some $K_x > 0$ such that $m > K_x$ implies $|f_m(x) - f(x)| < \varepsilon/4$. Therefore if $n > M$ and $m > M + K_x$ we have

$$|f(x) - f_n(x)| \le |f(x) - f_m(x)| + |f_m(x) - f_n(x)| < \frac{3\varepsilon}{4}$$

which means that

$$|f - f_n| = \sup\{|f(x) - f_n(x)|\} \le \frac{3\varepsilon}{4} < \varepsilon$$

and thus that $\lim_{n\to\infty} f_n = f$ in $\mathbf{BF}(X)$.

We next need to show that f is continuous if each f_n is continuous. Given $\varepsilon > 0$ and $x \in X$ we shall find an open set U such that $x \in U$ and for all $y \in U$ we have $|f(y) - f(x)| < \varepsilon$.

First of all, choose M such that $n > M$ implies $|f_n - f| < \varepsilon/3$. Next choose an open set U containing x such that $y \in U$ implies $|f_{M+1}(y) - f_{M+1}(x)| < \varepsilon/3$. The triangle inequality for real numbers then implies that

$$|f(y) - f(x)| \le |f(y) - f_{M+1}(y)| + |f_{M+1}(y) - f_{M+1}(x)| + |f(x) - f_{M+1}(x)| < \varepsilon$$

and therefore f is continuous at x; since x was arbitrary, this means that $f \in \mathbf{BC}(X)$.■

The examples in the proposition are special cases of the following important mathematical structure:

Definition. If $(V, |...|)$ is a normed vector space, then V is said to be a *Banach space* if it is complete with respect to the associated metric.

Intersections of nested closed sets

We had previously noted that compact metric spaces are characterized by the fact that families of closed subspaces with the finite intersection property have nonempty intersections. An important special case involves *nested* sequences of closed subsets $\{ A_n \}$ that are nonempty and satisfy $A_{n+1} \subset A_n$ for all n. In this case compactness implies that the intersection $\cap_n A_n$ is nonempty. There is an analog of this property for compact metric spaces. Recall that the *diameter* of a (subset of a) metric space is given by

$$\text{diam } (A) = \sup_{u,v \in A} \mathbf{d}(u,v)$$

if the set of distances has an upper bound and by $+\infty$ if no such upper bound exists.

PROPOSITION. (Nested Intersection Property). *Let X be a complete metric space, and let $\{ A_n \}$ be a nested sequence of nonempty closed subsets of X such that* $\lim_{n\to\infty} \text{diam } (A_n) = 0$. *Then $\cap_n A_n$ consists of one point.*

A proof that the intersection is nonempty is given in Lemma 48.3 on page 297 of Munkres. Suppose that y and z lie in the intersection. Then $y, z \in A_n$ for all n, and therefore $\mathbf{d}(y,z) \leq \text{diam } (A_n)$ for all n. Since the right hand side goes to zero as $n \to \infty$, it follows that the left hand side is ≤ 0; since the left hand side is nonnegative by construction, it must be zero, so that $y = z$ and the intersection contains exactly one point.■

Completions of metric spaces

We have already mentioned that one recurrent theme in theoretical mathematics is the desire to see whether empty spaces in mathematical structures can be filled in some sense, giving the real numbers as an example. In fact, one can view the real numbers as a system obtained from the rational numbers by filling in gaps so that every Cauchy sequence converges. In particular, the finite decimal fraction approximations to a real number form a set of rational numbers converging to the given real number, and as such they are Cauchy sequences that usually do not converge to rational points. Our objective here is to show that every metric space can be expanded to a larger one in which every Cauchy sequence converges. Often this is done by a brute force construction that starts with the set of all Cauchy sequences in the metric space (for example, see Munkres, Exercise 9, page 271). We shall construct this completion by a method that takes just about the same amount of work but also yields some illuminating insights of independent interest.

PROPOSITION. *If $(X, \mathbf{d})$ is a metric space, then there is a $1-1$ isometry from X into a Banach space.*

Proof. Before starting the proof we recall that a distance function satisfies

$$|\mathbf{d}(u,w) - \mathbf{d}(v,w)| \;\leq\; \mathbf{d}(u,v)$$

for all $u, v, w \in X$. This follows from two applications of the triangle inequality.

Let $\mathbf{BC}(x)$ be the Banach space of bounded continuous functions on X, and choose some point $a \in X$. Define a map $\varphi : X \to \mathbf{BC}(X)$ by the formula

$$[\varphi(x)](y) = \mathbf{d}(y, x) - \mathbf{d}(y, a) \ .$$

The right hand side is a continuous function of y because

$$\left|[\varphi(x)](y) - [\varphi(x)](z)\right| = \Big|(\mathbf{d}(y, x) - \mathbf{d}(y, a)) - (\mathbf{d}(z, x) - \mathbf{d}(z, a))\Big| \ \leq$$

$$\left|\mathbf{d}(y, x) - \mathbf{d}(y, x)\right| + \left|\mathbf{d}(z, a) - \mathbf{d}(y, a)\right| \ = \ 2\,\mathbf{d}(y, z)$$

for all $y, z \in X$, and it is a bounded function of y because

$$\Big|\mathbf{d}(y, x) - \mathbf{d}(y, a)\Big| \ \leq \ \mathbf{d}(x, a)$$

for all $y \in X$ if a and x are held fixed.

The estimates of the previous paragraph also show that

$$\left|\varphi(x) - \varphi(y)\right| \ \leq \ \mathbf{d}(x, y)$$

and in fact equality holds because

$$\left|[\varphi(x)](y) - [\varphi(y)](y)\right| \ = \ \Big|(\mathbf{d}(x, y) - \mathbf{d}(a, y)) - (\mathbf{d}(y, y) - \mathbf{d}(y, a))\Big| \ = \ \mathbf{d}(x, y)$$

so that φ is an isometry (hence uniformly continuous).■

Definition. If X is a metric space, then a *completion* of X is a pair (f, Y) consisting of a complete metric space and an isometry $f : X \to Y$ such that $\overline{f(X)} = Y$.

The preceding result implies that completions exist because the closure of $\varphi(X)$ in $\mathbf{BC(X)}$ is a closed subset of a complete metric spaced and therefore is complete. We shall also prove that up to an isometry there is only one way of completing a metric space. Here is a formal statement:

UNIQUENESS THEOREM. *Let X be a metric space, and let (f, Y) and (g, Z) be completions of X. Then there is a unique $1-1$ onto isometry $h : Y \to Z$ such that $h \circ f = g$.*

This result is a consequence of the following more general statement:

THEOREM. *Let X be a metric space, let (f, Y) be a completion of X, let W be a complete metric space, and let $h : X \to W$ be a uniformly continuous function. Then there is a unique uniformly continuous function $H : Y \to W$ such that $H \circ f = h$. Furthermore, if h is an isometry then so is H.*

Proof. The basic idea of the proof is simple. Since $\overline{f(X)} = Y$ it follows that for each $y \in Y$ there is a sequence of points $\{ x_n \}$ in X such that $\lim_{n\to\infty} f(x_n) = y$. The only way we can extend h is if we take $H(y) = \lim_{n\to\infty} h(x_n)$. We need to show this actually works. The first step is to verify that the definition of $H(y)$ makes sense (in particular, the sequence $\{ h(x_n) \}$ actually converges) and does not depend upon the choice of sequence in $f(X)$ converging to Y. Next, we have to show that the function is uniformly continuous. Finally we have to show that H is an isometry if h is.

First step. How do we know that the sequence converges? The hypotheses on h and W suggest a couple of ideas. Since W is complete the sequence $\{ h(x_n) \}$ will converge if it is a Cauchy sequence, and thus one might hope that the uniform continuity of h and the convergence of the sequence $\{ f(x_n) \}$ imply the convergence of $\{ h(x_n) \}$. Thus it suffices to show that *a uniformly continuous map of metric spaces from X to Y takes Cauchy sequences in X to Cauchy sequences in Y.* To see this, let $h : Y \to W$ be a uniformly continuous map of metric spaces, and assume that $\{ y_n \}$ is a Cauchy sequence in Y. Given $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbf{d}(u, v) < \delta$ implies $\mathbf{d}(h(u), h(v)) < \varepsilon$. Since we have a Cauchy sequence it follows that there is an M such that $m, n > M$ implies $\mathbf{d}(y_m, y_n) < \delta$. It follows that $m, n > M$ implies $\mathbf{d}(h(y_m), h(y_n)) < \varepsilon$. — This implies that there is some point $w_0 \in W$ such that $w_0 = \lim_{n\to\infty} h(x_n)$.

Second step. If we have two sequences $\{ u_n \}$ and $\{ v_n \}$ in X such that

$$\lim_{n\to\infty} f(u_n) = y = \lim_{n\to\infty} f(v_n)$$

we need to show that

$$\lim_{n\to\infty} h(u_n) = \lim_{n\to\infty} h(v_n)$$

in order to know that H is well-defined. Denote the limits of these two sequences by α and β respectively; it will suffice to show that $\mathbf{d}(\alpha, \beta) < \varepsilon$ for all $\varepsilon > 0$.

Given $\varepsilon > 0$ one can find a positive integer M such that for all $n > M$ we have $\mathbf{d}(h(u_n), \alpha) < \varepsilon/3$ and $\mathbf{d}(h(v_n), \beta) < \varepsilon/3$. Now choose $\delta > 0$ so that $\mathbf{d}(s, t) < \delta$ implies $\mathbf{d}(h(s), h(t)) < \varepsilon/3$, and choose P so that $p \geq P$ implies $\mathbf{d}(f(u_p), y) < \delta/2$ and $\mathbf{d}(f(v_p), y) < \delta/2$; the latter imply that $\mathbf{d}(f(u_p), f(v_p)) < \delta$, so that $\mathbf{d}(f(u_p), f(v_p)) < \varepsilon/3$. If we choose $q > M + P$ then we have

$$\mathbf{d}(\alpha, \beta) \leq \mathbf{d}(h(u_q), \alpha) + \mathbf{d}(h(u_q), h(v_q)) + \mathbf{d}(h(v_q), \beta)$$

and since $q > M$ the right hand side is less than $\mathbf{d}(h(u_q), h(v_q)) + 2\varepsilon/3$. Finally, since $q > P$ we also know that $\mathbf{d}(h(u_q), h(v_q)) < \varepsilon/3$, and therefore $\mathbf{d}(\alpha, \beta) < \varepsilon$ as required.

Third step. We need to show that the function H is uniformly continuous. Since h is uniformly continuous, for each $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbf{d}(u, v) < \delta$ implies $\mathbf{d}(h(u), h(v)) < \varepsilon/3$. Given $a, b \in Y$, let $\{ u_n \}$ and $\{ v_n \}$ be sequences in X such that $\lim_{n\to\infty} f(u_n) = a$ and $\lim_{n\to\infty} f(v_n) = b$, and suppose that $\mathbf{d}(a, b) < \delta$. Choose M so large that $n > M$ implies

$$\mathbf{d}(u_n, a), \quad \mathbf{d}(v_n, b) \quad < \quad \frac{\delta - \mathbf{d}(a, b)}{2}$$

and also

$$\mathbf{d}\left(h(u_n), H(a)\right), \quad \mathbf{d}\left(h(v_n), H(b)\right) \quad < \quad \frac{\varepsilon}{3}\,.$$

We then have that $\mathbf{d}(u_n, v_n) < \delta$ and that

$$\mathbf{d}\left(H(a), H(b)\right) \leq \mathbf{d}\left(h(u_n), H(a)\right) + \mathbf{d}(h(u_n), h(v_n)) + \mathbf{d}\left(h(v_n), H(b)\right) < 3 \cdot \frac{\varepsilon}{3} = \varepsilon$$

which shows that H is uniformly continuous.

Fourth step. We need to show that H is an isometry if h is an isometry. This will follow from a more general fact:

Suppose that $\{ u_n \}$ and $\{ v_n \}$ are convergent sequences in a metric space E with $\lim_{n\to\infty} u_n = u$ *and* $\lim_{n\to\infty} v_n = v$. *Then* $\lim_{n\to\infty} \mathbf{d}(u_n, v_n) = \mathbf{d}(u, v)$.

Using this we may complete the fourth step as follows: Express $u, v \in Y$ as limits of sequences $\{\, f(u_n)\,\}$ and $\{\, f(u_n)\,\}$ respectively. Then the assertion implies that

$$\mathbf{d}(u, v) = \lim_{n\to\infty} \mathbf{d}(f(u_n), f(v_n)) = \lim_{n\to\infty} \mathbf{d}(u_n, v_n)$$

where the last equation holds because f is an isometry. On the other hand we also have

$$\mathbf{d}(H(u), H(v)) = \lim_{n\to\infty} \mathbf{d}(h(u_n), h(v_n)) = \lim_{n\to\infty} \mathbf{d}(u_n, v_n)$$

because h is also an isometry. Since limits of sequences are unique it follows that $\mathbf{d}(u, v) = \mathbf{d}(H(u), H(v))$ and therefore H is an isometry.

We must now verify the general assertion about limits of distances. Consider the inequality

$$\left|\mathbf{d}(u_n, v_n) - \mathbf{d}(u, v)\right| \leq \left|\mathbf{d}(u_n, v_n) - \mathbf{d}(u, v_n)\right| + \left|\mathbf{d}(u, v_n) - \mathbf{d}(u, v)\right| \leq \mathbf{d}(u_n, u) + \mathbf{d}(v_n, v)\ .$$

If we choose M so that $n > M$ implies $\mathbf{d}(u_n, u),\ \mathbf{d}(v_n, v) < \varepsilon/2$ the inequalities imply that

$$\left|\mathbf{d}(u_n, v_n) - \mathbf{d}(u, v)\right| < \varepsilon$$

which proves the assertion about limits.■

Proof of the Uniqueness Statement. Let X be a metric space, and let (f, Y) and (g, Z) be completions of X. By the preceding theorem there are unique isometries $G : Y \to Z$ and $F : Z \to X$ such that $G \circ f = g$ and $F \circ g = f$. These in turn imply that $F \circ G \circ f = f$ and $G \circ F \circ g = g$ for the maps $FG : Y \to Y$ and $GF : Z \to Z$. Since id_Y and id_Z also satisfy $\mathrm{id}_Y \circ f = f$ and $\mathrm{id}_Z \circ g = g$ it follows that $G \circ F = \mathrm{id}_Z$ and $F \circ G = \mathrm{id}_Y$, showing that the isometries F and G are 1–1 and onto. This completes the proof that any two completions of a given metric space are isometric by an isometry compatible with the isometric inclusions of X.■

III.3 : Implications of completeness

(Munkres, § 48; Edwards, § III.1)

There are two fundamentally important properties of complete metric spaces that arise in numerous analytic and geometric contexts. One of these (Baire's Theorem) can be viewed informally as saying that a complete metric space cannot be decomposed into "thin" pieces. The other (the Banach Contraction Lemma) is a powerful method for finding solutions to various sorts of equations in a wide range of contexts. We shall give both topological and analytic examples in this course.

Nowhere dense and meager subspaces

Definition. A subset A of a topological space X is said to be *nowhere dense in* X if $\mathrm{Int}_X(\overline{A}) = \emptyset$.

The proofs of the following results are left to the reader as exercises.

PROPOSITION. *If A is a subset of X then A is nowhere dense in X if and only if $X - \overline{A}$ is dense in X.*■

PROPOSITION. *Suppose that $A \subset B \subset C \subset X$ and that B is nowhere dense in C. Then A is nowhere dense in C and B is nowhere dense in X.*■

The proof of this is left to the reader as an exercise.

PROPOSITION. *Suppose that A and B are nowhere dense subsets of X. Then $A \cup B$ is nowhere dense in X.*■

It is particularly useful to understand when a one point subset of X is nowhere dense. Since one point subsets are closed, a one point subset $\{x\}$ is nowhere dense in X if and only if it is **not** open; *i.e.*, the point x is not isolated in X.

Definition. A subset $B \subset X$ is said to be *meager* or *of the first (Baire) category* in X if it can be written as a countable union $B = \cup_n A_n$ where each A_n is closed and nowhere dense. A subset C is said to be *nonmeager* or *of the second (Baire) category* in X if it is not meager, and in this case the complement $X - C$ is said to be *residual* or *co-meager*.

Examples.

1. A closed nowhere dense subset is always meager.

2. The rationals are a meager, but definitely not nowhere dense, subset of the real numbers (since the rationals are dense in the reals).

3. Here is a more complicated but still important example. Let $\mathbf{R}^\infty$ denote the inner product space consisting of all sequences $(x_1, x_2, \cdots)$ such that $x_k = 0$ for all but finitely many k, with the inner product given by the convergent (in fact, finite) series

$$\langle x, y \rangle = \sum_j x_j y_j \ .$$

If B_n is the set of all points for which $x_j = 0$ for all $j > n$, then B_n is closed and nowhere dense in B_{n+1} and hence B_n is also closed and nowhere dense in X (why?). Since $\mathbf{R}^\infty = \cup_n B_n$ it follows that $\mathbf{R}^\infty$ is meager in itself.

Note. The concepts of first and second category were so named before category theory was invented; there is no connection between the two meanings of the word "category." In order to avoid confusion many authors have made conscious efforts to avoid terms like first and second category (compare the comments at the bottom of page 295 of Munkres), but this terminology is still very widely used by mathematicians and others. One way of dealing with this is to use the term *Baire category* when referring to concepts involving first and second category as defined above.

Baire spaces

Definition. A topological space X is said to be a *Baire space* if every open subset is nonmeager. Equivalent definitions are given on pages 295–296 of Munkres (in particular, see Lemma 48.1 on page 296).

The following result provides many examples of Baire spaces:

BAIRE CATEGORY THEOREM. *If X is a complete metric space, then X is a Baire space.*

A proof of this result for X itself is given in Theorem 48.2 on page 296 of Munkres (note that this proof uses Lemma 48.3 on the following page).■

The Baire Category Theorem has many extremely important and useful consequences. We shall begin with one that only involves point set theory.

PROPOSITION. *If X is a countable compact metric space, then X has at least one isolated point.*

Proof. Let $X = \{ x_1, x_2, \ ... \}$, and for each i let $F_i = \{ x_i \}$. If X has no isolated points, then each F_i is closed and nowhere dense, and therefore X is meager in itself. By Baire's Category Theorem this is impossible.■

COROLLARY. *If X is a compact metric space such that every point of X is a limit point, then X is uncountable.*■

This is essentially the contrapositive of the theorem.

On pages 41–42 of Rudin's book this corollary is applied to show that the Cantor set is uncountable. One can prove that the Cantor set has the same cardinality as the real numbers by a different argument (see Problems 37–38 on page 46 of Royden, *Real Analysis, Third Edition*, as well as Exercise 6 on page 179 of Munkres).

In Section 49 of Munkres the Baire Category Theorem is used to prove the existence of continuous functions on the unit interval that are not differentiable at any point of the interval. The theorem also plays a crucial role in the foundations of the theory of Banach spaces and functional analysis; in particular, it is the key ingredient of the proofs of the Uniform Boundedness Principle and Open Mapping Theorem (*e.g.*, see Rudin's book on Functional Analysis for more information on this).

The Contraction Lemma

A special case of this result is established as Theorem 1.1 on pages 162–163 of Edwards, and later in that book there are assertions that the proof and conclusion can be generalized. Here is the explicit generalization.

CONTRACTION LEMMA. *Let X be a complete metric space, and let $T : X \to X$ be a map such that $\mathbf{d}(T(x), T(y)) \leq \alpha \cdot \mathbf{d}(x, y)$ for some fixed $\alpha \in (0, 1)$ and all $x, y \in X$ (in particular, T is uniformly continuous). Then there is a unique $z \in X$ such that $T(z) = z$ (in other words, a* **unique fixed point** *for T).*

To see the need for completeness, consider the open interval $(0, 1)$ and let T be multiplication by $\frac{1}{2}$.

Proof. The idea is beautifully simple. One starts with an arbitrary point $x \in X$ and considers the sequence of points $x,\ T(x),\ T^2(x), \cdots$. This sequence is shown to be a Cauchy sequence, and the limit z of this sequence turns out to be the unique fixed point.

More formally, we begin by noting that T has at most one fixed point. If $z, w \in X$ satisfy $T(z) = z$ and $T(w) = w$, then we have

$$0 \quad \leq \quad \mathbf{d}(z, w) \quad = \quad \mathbf{d}\left(T(z), T(w)\right) \quad \leq \quad \alpha\mathbf{d}(z, w)$$

and since $0 < \alpha < 1$ this can only happen if $\mathbf{d}(z, w) = 0$; *i.e.*, if $z = w$.

We now follow the idea described in the first paragraph of the proof. By induction on n we have

$$\mathbf{d}\left(T^n(x), T^{n+1}(x)\right) \quad \leq \quad \alpha^n \mathbf{d}\left(x, T(x)\right)$$

and therefore by the triangle inequality for $m > n$ we also have

$$\mathbf{d}\left(T^n(x), T^m(x)\right) \quad \leq \quad \sum_{i=n+1}^{m} \alpha^i \mathbf{d}\left(x, T(x)\right) \quad =$$

$$\frac{\alpha^{n+1}(1 - \alpha^{m-n})}{1 - \alpha}\,\mathbf{d}\left(x, T(x)\right) \quad \leq \quad \frac{\alpha^{n+1}}{1 - \alpha}\mathbf{d}\ \left(x, T(x)\right)$$

which implies that the sequence $\{\, T^n(x)\, \}$ is a Cauchy sequence. By the completeness of X there is a point z such that $z = \lim_{n\to\infty} T^n(x)$.

By Theorem 23.1 on page 130 of Munkres we have

$$T(z) \quad = \quad \lim_{n\to\infty} T\left(T^n(x)\right) \quad = \quad \lim_{n\to\infty} T^{n+1}(x)$$

and by a change of variable (specifically, take $k = n+1$) the right hand side is equal to $\lim_{k\to\infty} T^k(x)$, which by construction is z. Therefore we have $T(z) = z$.∎

Applications of the Contraction Lemma to solving equations in one and two real variables are discussed in Section III.1 on pages 160–172 of Edwards as well as an addendum to this section (`cubicroots.*` in the course directory), and further applications from Edwards will be discussed later. For the time being we shall merely use the Contraction Lemma to prove the basic existence and uniqueness theorem for solutions of first order differential equations in one variable.

PICARD SUCCESSIVE APPROXIMATION METHOD FOR THE SOLUTIONS OF DIFFERENTIAL EQUATIONS. *Let $F(x, y)$ be a real valued function of two variables on an open set U such that F has continuous partial derivatives on U. Then for each $(a, b) \in U$ there is a positive real number $\delta > 0$ such that there is a unique solution of the differential equation*

$$\frac{dy}{dx} = \mathbf{F}(x, y)$$

on the interval $(a - \delta, a + \delta)$ *satisfying the initial condition* $y(a) = b$.

Proof. To motivate the proof, note first that a function f is a solution of the differential equation with the given initial value condition if and only if

$$f(x) = b \ + \int_a^x \mathbf{F}\,(t, f(t))\ dt$$

where as usual the integral is zero if $x = a$, while if $x < a$ the integral from a to x is defined to be the negative of the integral from x to a.

The idea of the proof is to use the right hand side to define a map of bounded continuous functions and then to apply the Contraction Lemma. However, one needs to be a bit careful in order to specify exactly which sorts of functions form the space upon which the mapping is defined and in order to ensure that the map has the contraction property.

Choose $h, k > 0$ so that

$$S = [a - h, a + h] \times [b - k, b + k] \quad \subset \quad U$$

so that $\mathbf{F}$ and its (first) partial derivatives are bounded on S. Let L be an upper bound for $\mathbf{F}$. By the Mean Value Theorem we have that

$$\left|\mathbf{F}(x, y_1) - \mathbf{F}(x, y_2)\right| \leq \max_{(u,v)\in S} \left(\left| \frac{\partial \mathbf{F}}{\partial y}(u, v) \right| \cdot |y_1 - y_2| \right)$$

for all x, y_1, y_2; let A be the maximum value of the absolute value of the second partial derivative on S.

Choose $\delta > 0$ so that $\delta \leq h$, $L\,\delta < k$ and $A\,\delta < 1$. Define M to be the metric space of all bounded continuous functions g on $(a - \delta, a + \delta)$ for which $|g - b| \leq L\,\delta$, where as usual we identify a real number with the constant function whose value is that number.

For every metric space Z, every $z \in Z$ and every positive real number B, the set of points w with $\mathbf{d}(z, w) \leq B$ is closed (why?), and therefore M is a complete metric space. We need to show that the map

$$[T(g)](x) \quad = \quad b \ + \int_a^x \mathbf{F}\,(t, g(t))\ dt$$

is defined for all $g \in X$, it maps

$$X \subset \mathbf{BC}\,(\ (a - \delta, a + \delta)\ )$$

into itself, and it satisfies the hypothesis of the Contraction Lemma on M.

First of all, it follows immediately that $T(g)$ is continuous whenever g is continuous (fill in the details here). Next, by the boundedness of $\mathbf{F}$ on the closed solid rectangle S we have

$$|T(g) - b| \quad = \quad \left| \int_a^x \mathbf{F}\,(t, g(t))\ dt \right| \quad \leq \quad L \cdot \left| \int_a^x \ dt \right| \quad \leq \quad L\,\delta$$

so that $g \in M$ implies $T(g) \in M$.

Finally, let $g_1, g_2 \in X$ and consider $|T(g_1) - T(g_2)|$. By definition the latter is equal to the least upper bound of the numbers

$$\left| \int_a^x \left(\mathbf{F}\left(t, g_1(t)\right) - \left(t, g_2(t)\right)\right) \, dt \right| \quad \leq$$

$$\int_a^x \left| \mathbf{F}\left(t, g_1(t)\right) - \left(t, g_2(t)\right)\right| dt \quad \leq \quad A\,\delta \cdot |g_1 - g_2| \ .$$

Since $A\,\delta < 1$, all the hypotheses of the Contraction Lemma apply so that there is a unique fixed point, and as noted above this unique fixed point must be the (necessarily unique) solution of the original differential equation with the prescribed boundary condition.■

Note. One can prove an existence theorem with the weaker hypothesis that $\mathbf{F}$ is continuous (compare Exercise 25 on pages 170–171 of Rudin's book), but uniqueness does not follow. For example, if $\mathbf{F}(x, y) = y^{1/2}$, then the zero function and $x^2/4$ are both solutions to the differential equation with initial condition $y(0) = 0$.

III.4 : Connected spaces

(Munkres, §§ 23, 24, 25)

One of the most basic results in single variable calculus is the *Intermediate Value Theorem*, which states that if a real valued function f is continuous on an interval $J \subset \mathbf{R}$ and $a, b \in J$ are points such that $f(a) \neq f(b)$, then for each real number y between $f(a)$ and $f(b)$ there is a real number x between a and b such that $f(x) = y$.

There are many other situations where one has such a conclusion, and it is useful to have a systematic understanding of when one has such intermediate value results. This requires one to find the abstract concept which underlies the Intermediate Value Theorem, and this notion is called *connectedness.*

Definition. A *separation* of topological space X is a pair of disjoint closed proper subspaces A and B whose union is X. A space is said to be *connected* if it does not have any separations. A space is said to be *disconnected* if it is not connected.

Of course we want intervals in the real line to be connected, but before addressing this point we give two equivalent formulations of the concept of separation.

PROPOSITION. *If X is a topological space, the the following are equivalent:*

(i) One can write $X = A \cup B$ where A and B are nonempty disjoint closed subsets.

(ii) One can write $X = A \cup B$ where A and B are nonempty disjoint open subsets.

(iii) There is a nonempty proper subset $A \subset X$ that is both open and closed (sometimes one says that such a subset is **clopen***).*

Proof. $(i) \Longrightarrow (ii)$ By construction we have $B = X - A$ and $A = X - B$, so that the subsets A and B **are also open in** X.

$(ii) \Longrightarrow (iii)$ The set A is nonempty and it is a proper subset because $B = X - A$ is nonempty. By hypothesis A is closed, and since $A = X - B$ we know that A is also open.

$(iii) \Longrightarrow (i)$ If $B = X - A$, then B is closed since A is open and B is open since A is closed. By hypothesis, A is nonempty, and since it is a proper subset we also know that B is nonempty. The conditions $A \cap B = \emptyset$ and $A \cup B = X$ follow immediately from the definition of B.■

Connectedness and the real line

There are some immediate examples of spaces that are connected and spaces that are not connected. Every space with at most one point is connected because there are no nonempty proper subsets, and hence there are no subsets for which condition (iii) is meaningful. Every set with an indiscrete topology is connected because there are no nonempty proper subsets that are either open or closed. On the other hand, if a set S has at least two elements and is given the discrete topology, then every subset of S is open (hence by complementation every subset is also closed!), and therefore **every** nonempty proper subset of S will be open and closed.

In particular, the preceding discussion implies that *a topological space with the discrete topology is connected if and only if it contains at most one element.*

Without further discussion we proceed to the single most important family of examples of connected sets.

THEOREM. *Let A be a subset of $\mathbf{R}$ with at least two elements. Then A is connected if and only if for each $a, b \in A$ such that $a < b$ the entire interval $[a, b]$ is contained in A.*

Proof. ($\Longrightarrow$) Suppose that the conclusion is false, so that there is some c satisfying $a < c < b$ and $c \notin A$. Let

$$B = A \cap (c, \infty) = A \cap [c, \infty)$$

where the second equality holds because $c \notin A$. We know that $b \in B$ so that the latter is nonempty, and the two descriptions of B in the displayed formula show it is both open and closed in $\mathbf{R}$. Finally, since $a \in A - B$ we know that B is a proper subset, and therefore A is not connected. We have thus shown the contrapositive of what we wanted to prove (and this proves the latter).

($\Longleftarrow$) Suppose that C is a nonempty open and closed subset of A. Without loss of generality we may as well assume that $a \in C$; if this is false, then a lies in the nonempty open closed subset $A - C$ and we can go through the same argument reversing the roles of C and $A - C$ at each point.

By hypothesis we know that $[a, b] \subset A$. Since C is open in A there is a $\delta > 0$ such that

$$[a, a + \delta) \quad \subset \quad (a - \delta, a + \delta) \ \cap C \ \cap [a, b] \quad \subset \quad C$$

and thus the set

$$K = \{ \ y \in (a, b] \ \big| \ [a, y] \quad \subset \quad C \ \}$$

is nonempty with an upper bound (specifically, b), Let y^* be the least upper bound of K; we claim that $y^* = b$. In order to do this we need to show that $y^* < b$ is impossible.

By the definition of least upper bound, for each $n > 0$ there is some $y_n \in K$ such that

$$y^* - \frac{1}{n} \quad < \quad y_n \quad \leq \quad y^*$$

and for this sequence we have $y^* = \lim_{n \to \infty} y_n$. Since each of the points of the sequence lies in C and the latter is closed, it follows that $y^* \in K$. Suppose now that $y^* < b$. On the other hand, since C is open there is some $\eta > 0$ such that $y^* + \eta < b$ and

$$(y^* - \eta, y^* + \eta) \quad \subset \quad C \ .$$

The latter in turn implies

$$\left[a, y^* + \frac{\eta}{2}\right] \quad \subset \quad C$$

which means that y^* is NOT an upper bound for K. This contradiction forces the conclusion $y^* = b$.

The preceding argument shows that *if C is a nonempty open and closed subset of A containing a point a and b is another point in A such that $a < b$, then $b \in C$.* — To prove that $C = A$ and hence that A is connected, it suffices to verify that the corresponding statement holds for all $b \in A$ such that $b < a$. The only way this could fail would be if $b \in A - C$. But in this case our argument would imply that $a \in A - C$, which is false. Therefore we have shown that a nonempty open and closed subset of A must be all of A if the latter has the intermediate point property described in the statement of the theorem.■

COROLLARY. *The connected subsets of* $\mathbf{R}$ *are all given by the following list:*

(1) *Closed intervals* $[a,b]$ *where* $a,b \in \mathbf{R}$.

(2) *Open intervals* (a,b) *where* $a \in \mathbf{R} \cup \{-\infty\}$ *and* $b \in \mathbf{R} \cup \{+\infty\}$.

(3) *Half open intervals* $[a,b)$ *or* $(a,b]$ *where* $a \in \mathbf{R}$ *and* $b \in \mathbf{R} \cup \{+\infty\}$ *in the first case and* $a \in \mathbf{R} \cup \{-\infty\}$ *and* $b \in \mathbf{R}$ *in the second.*■

It follows that the cardinality of the set of all connected subspaces of $\mathbf{R}$ is equal to the cardinality of $\mathbf{R}$ itself. The verifications of this statement and the corollary are left to the reader as exercises.

The proof of the Intermediate Value Theorem is essentially a combination of the characterization of connected subsets of the real line and the following abstract result, which establishes the conclusion of the Intermediate Value Theorem for arbitrary continuous real valued functions on connected topological spaces:

PROPOSITION. *If* $f : X \to Y$ *is continuous and* X *is connected, then* $f(X)$ *is also connected.*

Proof. Let $C \subset f(X)$ be a nonempty subset that is both open and closed. Then $C = U \cap f(X)$ where U is open in Y and $C = E \cap f(X)$ where E is closed in Y. We then have that

$$f^{-1}(C) = f^{-1}\big(U \cap f(X)\big) = f^{-1}(U)$$

is open in X and

$$f^{-1}(C) = f^{-1}\big(E \cap f(X)\big) = f^{-1}(E)$$

is closed in X. Since C is a nonempty subset of $f(X)$ it follows that $f^{-1}(C)$ is nonempty and therefore by the connectedness of X we have $f^{-1}(C) = X$. In particular, for all $x \in X$ this means that $f(x) \in C$, which in turn means that $C \supset f(X)$; by assumption the reverse inequality holds so that $C = f(X)$. Therefore we have shown that the only nonempty subset of $f(X)$ that is open and closed is $f(X)$ itself, which means that $f(X)$ is connected.■

Finding connected (sub)sets

The Intermediate Value Theorem for connected spaces is a very powerful statement on the existence of solutions to equations of the form $y = f(x)$, and therefore it is important to recognize when it applies, particularly to subspaces of the plane or other Euclidean spaces. One expects that the theorem will apply to open and closed rectangles that are products of two open or closed intervals respectively. These facts will follow quickly from our abstract discussion below.

PROPOSITION. *Suppose that* X *is a topological space, that* C *is an open and closed subset of* X*, and that* A *is a connected subset of* X*. Then either* $A \subset C$ *or* $A \subset X - C$*.*

Proof. The intersection $A \times C$ is an open and closed subset of A, so by connectedness of the latter it must either be all of A or empty. In the first case we have either $A \subset C$, and in the second we have $A \subset X - C$.■

PROPOSITION. *Suppose that* X *is a topological space and that* A *and* B *are connected subsets of* X *with a nonempty intersection. Then* $A \cup B$ *is connected.*

Proof. Suppose that C is a nonempty open and closed subset of $A \cup B$. Without loss of generality we may as well assume that some point $x_0 \in A \subset B$ belongs to C; if instead we have $x_0 \in A \cup B - C$ then we can switch the roles of C and $A \cup B - C$ in the proof of the first case.

Since A is a connected subset of $A \cup B$ we must have $A \subset C$ or $A \subset A \cup B - C$. Since $x_0 \in C$ the latter is impossible, and hence $A \subset C$. If we interchange the roles of A and B in this argument we also conclude that $B \subset C$, so that

$$C \subset A \cup B \subset C$$

which implies that the two sets are equal and hence that $A \cup B$ must be connected.■

Remark. If A and B are connected subsets of a topological space X it does **NOT** follow that $A \cap B$ is connected. Here is a counterexample when $X = \mathbf{R}^2$:

Let A and B be the semicircles in the unit circle (with equation $x^2 + y^2 = 1$) whose second coordinates are positive and negative respectively. Each subset is connected because the semicircles are the continuous images of the interval $[-1, 1]$ under the continuous mappings

$$\gamma_{\pm}(t) = \left(t, \pm\sqrt{1-t^2}\right)$$

but the intersection is the pair of points with coordinates $(\pm 1, 0)$ and this set is not connected (it has the discrete topology).

Definition. Given a topological space X and $a, b \in X$, define a binary relation $\sim_{[\mathrm{CONN}]}$ by $a \sim_{[\mathrm{CONN}]} b$ (for $a, b \in X$) if and only if there is a connected subset of X that contains both a and b.

ELEMENTARY FACT. *The preceding binary relation is an equivalence relation, and its equivalence classes are called the* **(connected) components** *of* X.■

PROPOSITION. *The connected components of X are maximal connected subsets of X.*

Proof. Let A be a connected component of X, and let C be a nonempty open and closed subset of A. Take a to be a point in $C \cap A$.

If $y \in A$, then by definition there is a connected subset A_y of X that contains both a and y. The definition of the equivalence relation implies that $A_y \subset A$. By an earlier result we know that either $A_y \subset C$ or $A_y \subset A - C$. The latter is impossible because $a \in A_y \cap C$, and therefore $A_y \subset C$ for all $y \in A$. Since $y \in A_y$ for all y, this means that $A \subset C$ so that $C = A$ and therefore A must be connected.

To verify maximality, suppose that B is a connected set such that $B \supset A$. Then for each $b \in B$ the set B itself is a connected subset of X that contains a and b, and therefore all points of B are in the same component as a, which is merely A. Therefore $B \subset A$ and A is maximal.■

In a discrete space the connected components are just the one point subsets and as such they are open and closed. One can ask whether connected components in arbitrary spaces have similar properties. It turns out that these subsets are always closed but not necessarily open. The first of these will be an immediate consequence of the following result.

PROPOSITION. *If X is a topological space and $A \subset X$ is nonempty and connected, then its closure $\overline{A}$ is also connected.*

Since a component is a maximal connected subset, the preceding result shows that a component must be equal to its own closure and therefore must be closed.

Proof of proposition. Let C be a nonempty open and closed subset of $\overline{A}$ It follows that there is some point $y \in \overline{A}$ and since C is open it also follows that either $A \cap C \neq \emptyset$. Since A is connected it follows that $A \subset C$, and since C is closed in $\overline{A}$ it also follows that $\overline{A} \subset C$.■

COROLLARY. *In the notation of the previous proposition, if $B \subset X$ satisfies $A \subset B \subset \overline{A}$, then B is connected.*

Proof. The proposition implies that Closure(A, B) is connected, and the latter is just the set $\overline{A} \cap B = B$.■

Example. There are many examples to show that connected components are not necessarily open subsets. In particular, the rational numbers with the subspace topology inherited from the real numbers have this property. — Let $a \in \mathbf{Q}$ and consider the open sets $N^{\mathbf{Q}}_{\sqrt{2}/2n}(a)$. We claim these sets are both open and closed in $\mathbf{Q}$; openness is immediate, and they are closed because a rational number b lies in such a set if and only if $|b - a| \leq \sqrt{2}/2n$ because

$$a \pm \frac{\sqrt{2}}{2n}$$

is never a rational number. Therefore it follows that if A is the connected component of a in $\mathbf{Q}$ we must have that $A \subset N^{\mathbf{Q}}_{\sqrt{2}/2n}(a)$ for all $n > 0$. But this forces A to be equal to $\{a\}$, which is definitely not open in $\mathbf{Q}$ (each ε-neighborhood contains infinitely may other rational numbers).

Products and connectedness

The next result provides an important tool for recognizing connected subset in Euclidean spaces.

THEOREM. *If X and Y are connected spaces then so is their product $X \times Y$.*

Proof. The result is trivial if either X or Y is empty, so assume that both are nonempty. Let $(x_0, y_0) \in X \times Y$, and let C be the connected component of (x_0, y_0). We shall show that $C = X \times Y$.

For each $(u, v) \in X \times Y$ we have slice homeomorphisms from X and Y to $X \times \{v\}$ and $\{u\} \times Y$ respectively, and therefore the latter subspaces are all connected. This means that for each $(x, y) \in X \times Y$ the points (x, y) and (x_0, y) lie in the same connected component, and similarly the points (x_0, y) and (x_0, y_0) also lie in the same connected component. By the transitivity of the relation of belonging to the same component it follows that (x, y) and (x_0, y_0) also lie in the same connected component, and hence that $C = X \times Y$.■

COMPLEMENT. *The same conclusion holds for arbitrary finite products.*

In fact, an arbitrary product of connected spaces is connected. An outline of the proof appears in Exercise 10 on page 152 of Munkres.

Proof. This follows from the theorem by induction and the canonical homeomorphism

$$\big(X_1 \times \cdots \times X_n\big) \times X_{n+1} \cong \prod_{i=1}^{n+1} X_i \ .■$$

COROLLARY. *For each positive integer n the space $\mathbf{R}^n$ is connected, and for each sequence of closed intervals $[a_i, b_i]$ (where $1 \leq i \leq n$) the product $\prod_i [a_i, b_i]$ is connected.*■

The results proven thus far have the following noteworthy consequence: *The cardinality of the set of connected subsets of $\mathbf{R}^2$ is the same as the cardinality of the set of all subsets of $\mathbf{R}^2$ (or* *equivalently the cardinality of the set of all subsets of* $\mathbf{R}$*).* — To see this, begin by noting that the open rectangular region $(0,1)^2$ is connected by the theorem as is the closed rectangle $[0,1]^2$, and the closed rectangle is the closure of the open rectangle (it is easy to find infinite sequences in the open rectangle that converge to an arbitrary point of the closed rectangle; alternatively, one can use the general rule

$$\overline{A \times B} = \overline{A} \times \overline{B}$$

to show this). Given a subset S of $(0,1) \cong \mathbf{R}$. consider the set

$$C_S = [0,1) \times [0,1] \cup \{1\} \times S \ .$$

By the previous results on closures of connected sets it follows that each set C_S is connected and by construction $C_S \neq C_T$ if $S \neq T$. Therefore there are at least as many connected subsets of $\mathbf{R}^2$ as we have claimed. On the other hand, there are at most as many of these as there are subsets of $\mathbf{R}^2$, and therefore by the Schröder-Bernstein Theorem it follows that the cardinalities are the same.

The preceding yields an example for an assertion made earlier: $\mathbf{R}^2$ *is not homeomorphic to a subset of* $\mathbf{R}$*.* If A is an arbitrary subset of $\mathbf{R}$, the characterization of connected subsets of the real line shows that the cardinality of the set of connected subsets of A is at most the cardinality of the real numbers themselves, but we know that the cardinality of the set of connected subsets of $\mathbf{R}^2$ is greater than this.

Distinguishing homeomorphism types

Connectedness provides an effective means for showing that certain pairs of spaces are not homeomorphic to each other.

PROPOSITION. *No two sets in the following list are homeomorphic:*

(*i*) *The closed unit interval* $[0,1]$*.*
(*ii*) *The open unit interval* $(0,1)$*.*
(*iii*) *The half-open unit interval* $(0,1]$*.*
(*iv*) *The circle* $S^1 \subset \mathbf{R}^2$ *defined by the equation* $x^2 + y^2 = 1$*.*

Proof. The first and last sets are compact while the second and third are not, so it suffices to show that S^1 is not homeomorphic to $[0,1]$ and $(0,1)$ is not homeomorphic to $(0,1]$ (Would the result remain true if we added the half-open interval $[0,1)$ to the list? Why or why not?).

Our reasoning relies on the following observation: *If* $f : X \to Y$ *is a homeomorphism and* A *is a finite subset, then* $f|X - A$ *maps* $X - A$ *homeomorphically to* $Y - f(A)$*; in particular,* $X - A$ *is connected if and only if* $Y - f(A)$ *is connected.* If one removes two points from S^1 the resulting space is disconnected (supply the details!), but if one removes the endpoints of $[0,1]$ the resulting space is still connected. Therefore the observation shows that S^1 and $[0,1]$ cannot be homeomorphic.

Similarly, if one removes the endpoint from $[0,1)$ then the resulting space is connected, but if one removes any point from $(0,1)$ the resulting space is disconnected.■

Further discussion along these lines yields complete topological characterizations of 1-dimensional objects like the unit circle or a closed interval. Textbook discussions of this appear in books by Hocking and Young (*Topology*, Section 2–5 on pages 52–55 with background material in the preceding section) and Christensen and Voxman (*Aspects of Topology* (First Edition), Section 9.A on pages 227–232 with accompanying exercises on page 251, and closely related material in Section 5.A on pages 127–128).

III.5 : Variants of connectedness

(Munkres, §§ 23, 24, 25)

If U is an open subset of some Euclidean space, then U is connected if and only if each pair of points in U can be joined by a broken line curve that lies entirely in U. This fact, which we shall prove at the end of the present section, reflects two important refinements of the concept of connectedness.

Locally connected spaces

Definition. A topological space X is said to be *locally connected* if for each $x \in X$ and each open set $U \subset X$ such that $x \in U$ there is a **connected** open set V such that $x \in V \subset U$.

Example. If U is open in some Euclidean space then U is locally connected. Suppose that $W \subset U$ is an open set and that $x \in W$. Choose $\delta > 0$ so that $N_\delta(x) \subset W$. To see that $N_\delta(x)$ is connected, given y in the latter consider the image J_y of the curve

$$\gamma(t) = x + t(y - x)$$

for $0 \leq t \leq 1$. The set J_y is a connected set containing x and y, and therefore y lies in the same connected component of $N_\delta(x)$ as x, and since y was arbitrary this implies that the set in question is connected.

If we take a disconnected open subet of some Euclidean space, we have an example of a locally connected space that is not connected. Finding examples of spaces that are connected but not locally connected takes more work. Before doing this we state a basic charactererization of local connectedness that is established in Theorem 25.3 on page 161 of Munkres.

PROPOSITION. *A topological space X is locally connected if and only if for each open subset U, the components of U are open.*■

Example. *A connected space that is not locally connected.* Let A be the graph of the function $f(x) = \sin(1/x)$ for $x > 0$, and let $B = \overline{A}$. It follows that B is the union of A with the set $\{0\} \times [-1, 1]$ and that B is connected because A is connected. We claim that B is not locally connected. Let W be the set of all points in B for which the second coordinate lines in $(-1, 1)$; then W is open and therefore it suffices to find a component of W that is not open. By construction W does not contain any points whose first coordinates have the form $2/m\pi$ where m is an odd positive integer. Therefore if C is the connected component in W containing the connected set $\{0\} \times (-1, 1)$, it follows that C cannot have any points whose first coordinates are $\geq 2/m\pi$ for all m and thus that the first components of points in C must be zero. In other words, we must have $C = \{0\} \times (-1, 1)$. This set is not open. In particular, for $\delta > 0$ the set $N_\delta([\mathbf{origin}])$ contains points of W that are not in C — specifically all points of the form $(1/k\pi, 0)$ for all k sufficiently large.

The proof of the following result is left to the reader as an exercise:

PROPOSITION. *A finite product of locally connected spaces is locally connected.*

There is an analog for infinite products with a curious extra condition. Namely, an arbitrary product of locally connected spaces is locally connected if and only if all but at most finitely many of the factors are also connected. To see the need for this condition, note that if $x \in \prod_\alpha X_\alpha$ and U is an open set containing x, then $p_\alpha(U) = X_\alpha$ for all but finitely many α (filling in the details and the proof of the original assertion are left to the reader as exercises).

Path or arcwise connectedness

The connectedness of the real interval leads to an important and useful criterion for recognizing connected spaces.

Definition. A topological space X is said to be *path connected* or *arcwise connected* if for each pair of points $x, y \in X$ there is a continuous function (or curve or path) $\gamma : [a, b] \to X$ such that $\gamma(a) = x$ and $\gamma(b) = y$.

PROPOSITION. *An arcwise connected space is connected.*

Proof. Let $x \in X$, and let C be the connected component of x in X. Given $y \in Y$, let γ be as in the definition, and let J_y be the image of γ. Then J_y is a connected set containing x and y, and consequently we must have $y \in C$. Since $y \in X$ was arbitrary, this means that $X = C$.

Example. The previous example B constructed from the graph of $f(x) = \sin(1/x)$ is a connected space that is not arcwise connected. — To see this, suppose that γ is a continuous curve in B defined on a closed interval $[a, b]$ that joins a point of the form $(0, y_0)$ to a point (x_1, y_1) with $x_1 > 0$. Since $\{0\} \times [-1, 1]$ is compact, it follows that the closed set

$$\gamma^{-1}\big(\{0\} \times [-1, 1]\big) \subset [a, b)$$

is a closed subset of $[a, b]$ and thus has a maximum point $c < b$. At least one of the open sets

$$V_- \;=\; B \cap (\mathbf{R} \times (-1, 1]\,)\,, \qquad V_+ \;=\; B \cap (\mathbf{R} \times [-1, 1)\,)$$

contains $\gamma(c)$ depending upon whether the last coordinate is ± 1 or neither so choose V_ε to be such an open subset. Then there is a $\delta > 0$ such that $\delta < b - c$ and $0 \leq \delta$ implies that $\gamma(c+t) \in V_\varepsilon$. It follows that for all but finitely many positive integers n there are points $t_n \in (0, \delta)$ such that the first coordinate of $\gamma(t_n)$ is equal to $2/n\pi$; specifically, pick any value $t_0 \in (0, \delta)$, so that the first coordinate η_0 of $\gamma(t_0)$ will be positive, and then notice that for all sufficiently large n we have

$$\frac{2}{n\pi} \;<\; t_0 \;.$$

However, by construction V_- does not have any points whose first coordinates have the form $2/n\pi$ where n is an integer of the form $4k+3$ (the values of x for which $\sin(1/x) = -1$), and V_+ does not have any points whose first coordinates have the form $2/n\pi$ where n is an integer of the form $4k+1$ (the values of x for which $\sin(1/x) = +1$). Therefore it is not possible to construct a continuous curve in B joining a point with zero first coordinate to a point with positive first coordinate, and therefore the connected set B is not arcwise connected.

Analogies with connectedness

There is a concept of *path component* or *arc(wise) component* that is analogous to the concept of connected component.

Definition. Given a topological space X and $x, y \in X$, define a binary relation $\sim_{[\mathrm{ARC}]}$ by $x \sim_{[\mathrm{ARC}]} y$ (for $a, b \in X$) if and only if there is a continuous function $\gamma : [a, b] \to X$ such that $\gamma(a) = x$ and $\gamma(b) = y$. We often say that γ is a curve joining x and y if this condition holds.

It is obvious that this relation is reflexive (use the constant curve) and symmetric (consider the curve δ defined on $[-b, -a]$ with $\delta(t) = \gamma(-t)$. To see that the relation is transitive it is convenient to introduce a concept that arises frequently in mathematics.

Definition. Let X be a topological space, and suppose that $\gamma : [a, b] \to X$ and $\delta : [c, d] \to X$ are continuous curves such that $\gamma(b) = \delta(c)$. The *sum* or *concatenation* (stringing together) $\gamma + \delta$ is the continuous curve

$$\sigma : [a, b + d - c] \to X$$

defined by $\sigma(t) = \gamma(t)$ if $t \in [a, b]$ and $\sigma(t) = \delta(t + b - c)$ if $t \in [b, b + d - c]$. An illustration of this appears in the course directory as files of the form `concat.`$*$.

By construction the sum of curves is an associative operation whenever it is defined.

Transitivity of $\sim_{[\mathrm{ARC}]}$ follows immediately from this definition, for if γ joins x and y and δ joins y and z, then $\gamma + \delta$ joins x and z.

Definition. The equivalence classes of $\sim_{[\mathrm{ARC}]}$ are called *path components* or *arc components* of X. It follows that every arc component is arcwise connected, but the preceding example shows that arc components need not be closed or open in the ambient space X.

The statements of the next two results are parallel to those for connected spaces, but the proofs are entirely different.

PROPOSITION. *If X is an arcwise connected space and $f : X \to Y$ is continuous, then $f(X)$ is arcwise connected.*

Proof. Given $a, b \in f(X)$ write $a = f(c)$ and $b = f(d)$ for $c, d \in X$. Since X is arcwise connected there is a continuous curve γ in X joining c to d, and the composite $f \circ \gamma$ is a continuous curve in $f(X)$ joining a to b.■

PROPOSITION. *A (finite) product of arcwise connected spaces is arcwise connected.*

In fact, the finiteness condition is completely unnecessary in the statement and proof of this result.

Proof. Let $X_1, \cdots, X_n$ be the spaces in question, let $u, v \in \prod_i X_i$ and for each j between 1 and n let u_j and v_j be the j^{th} coordinates of u and v respectively. Then for each j one can join u_j to v_j by a continuous curve γ_j. Suppose that γ_j is defined on the closed interval $[a_j, b_j]$ and let $L_j : [0, 1] \to [a_j, b_j]$ be the unique linear function that sends 0 to a_j and 1 to b_j. Then there is a unique continuous function $\alpha : [0, 1] \to \prod_i X_i$ whose projection onto the j^{th} coordinate is $\gamma_j \circ L_j$ for each j, and by construction α joins a to b.■

There is also a corresponding notion of local path or arcwise connectedness (see Munkres, page 161), and as noted in Theorem 25.5 on that page (the proof continues to page 162), if a space is locally arcwise connected then its components and path components are identical.

Open subsets of Euclidean spaces

The discussion of local connectedness for open sets in Euclidean spaces actually proves that such sets are locally arcwise connected and hence their components, which are open, are the same as their arc components. In particular, an open subset of a Euclidean space is connected if and only if it is arcwise connected, and the discussion at the beginning of this section asserts that one can choose the curves in question to be of a special type. In order to prove this we need to give a formal definition of a broken line curve.

Definition. A *closed line segment curve* in $\mathbf{R}^n$ is a continuous curve γ defined on $[0,1]$ by an equation of the form

$$\gamma(t) = a + t \cdot (b - a)$$

for some $a, b \in \mathbf{R}^n$. A *broken line curve* is a finite iterated concatenation of closed line segment curves.

PROPOSITION. *Let U be open in $\mathbf{R}^n$. Then U is connected if and only if every pair of points in U can be joined by a broken line curve that lies entirely in U.*

Proof. If the conclusion is true then U is arcwise connected. To prove the $(\Longrightarrow)$ implication, define a binary relation $\sim$ by $u \sim v$ if and only if there is a broken line curve in U joining u to v. This is an equivalence relation, and in fact the equivalence classes are open subsets (if $x \in U$ and $N_\delta(x) \subset U$, then every point in $N_\delta(x)$ can be joined to x by a closed line segment curve). It follows that the union of any family of equivalence classes is also open, and in particular, if W is an equivalence class this means that $U - W$, which is the union of all the equivalence classes except W, is also open. The latter implies that W is closed, and since U is connected it follows that there can be only one equivalence class for the equivalence relation described above; this proves that each pair of points in U can be joined by a broken line curve in U.■

Given an open connected subset of $\mathbf{R}^n$ one can ask many different questions about the continuous curves joining two arbitrary points in U, including the following: *Given two points in U, can they be joined by a curve whose coordinate functions are infinitely differentiable? If so, can one find such a function such that the tangent vector at every point is nonzero?*

The answer to both questions is yes, but the proofs are more complicated than the ones given above. Such results can be established using techniques from an introductory graduate course on smooth manifolds.

IV. Smooth Functions

In many mathematical contexts one considers functions with properties that are stronger than continuity. For example, if X and Y are open in Euclidean spaces it is often necessary or desirable to consider functions with good differentiability properties. The latter often have important metric or topological consequences, and in this course we shall be interested in certain results of this type.

IV.1 : Linear approximations

(Edwards, §§ II.1, II.2)

For functions of one real variable, a function is continuous at a point if it has a derivative at that point, but there are standard examples of functions of two variables that have partial derivatives defined near a point but are not continuous there. On the other hand, a basic result in multivariable calculus shows that functions that have continuous partial derivatives near a point are necessarily continuous at the point in question.

We shall begin by establishing a version of this result that holds for functions of an arbitrary (finite) number of real variables. It will be convenient to adopt some notation first. The unit vector in $\mathbf{R}^n$ whose i^{th} coordinate is 1 and whose other coordinates are 0 will be denoted by $\mathbf{e}_i$. If U is an open set in $\mathbf{R}^n$ and $f : U \to \mathbf{R}^n$ is a (not necessarily continuous) function such that all first partial derivatives exist at some point $p \in U$, then the *gradient* $\nabla f(p)$ will denote the vector whose i^{th} coordinate is the partial derivative of f with respect to the i^{th} variable at p. We shall use $\langle u, v\rangle$ to denote the usual dot product of two vectors $u, v \in \mathbf{R}^n$.

PROPOSITION. *Let U be an open subset in $\mathbf{R}^n$, let $x \in \mathbf{R}^n$, and let $f : U \to \mathbf{R}^n$ be a (not necessarily continuous) function such that f has continuous partial derivatives on some open subset of U containing x. Then for all sufficiently small $h \neq 0$ in $\mathbf{R}^n$ one can define a function $\theta(h)$ such that*

$$f(x+h) - f(x) = \langle \nabla f(x), h\rangle + |h|\,\theta(h)$$

where $\lim_{h\to 0} \theta(h) = 0$.

Proof. Write $h = \sum_i t_i h_i$ for suitable real numbers t_i, take $\delta > 0$ so that f has continuous partial derivatives on $N_\delta(x)$, and assume that $0 < |h| < \delta$. Define h_i for $0 \leq i \leq n$ recursively by $h_0 = 0$ and $h_{i+1} = h_i + t_{i+1}\mathbf{e}_{i+1}$. Then $h_n = h$ and we have

$$f(x+h) - f(x) = \sum_i f(x+h_i) - f(x_{i-1})$$

and if we apply the ordinary Mean Value Theorem to each summand we see that the right hand side is equal to

$$\sum_i \frac{\partial}{\partial x_i} f\left(x + h_{i-1} + K_i(x) t_i\right) \cdot t_i$$

for some numbers $K_i(x) \in (0,1)$. The expression above may be further rewritten in the form

$$\langle \nabla f(x), h\rangle + \left(\sum_i \frac{\partial}{\partial x_i} f\left(x + h_{i-1} + K_i(x)t_i\right) - \frac{\partial}{\partial x_i} f(x)\right) \cdot t_i$$

and therefore an upper estimate for $|f(x+h) - f(x) - \langle \nabla f(x), h\rangle \mid$ is given by

$$\sum_i \left|\frac{\partial}{\partial x_i} f\left(x + h_{i-1} + K_i(x)t_i\right) - \frac{\partial}{\partial x_i} f(x)\right| |t_i|.$$

Since the partial derivatives of f are all continuous at x, for every $\varepsilon > 0$ there is a $\delta_1 > 0$ such that $\delta_1 < \delta$ and $|h| < \delta_1$ implies that each of the differences of partial derivatives has absolute value less than ε/n. Since $h \neq 0$, if we define $\theta(h)$ as in the statement of the conclusion, the preceding considerations imply show that

$$|\theta(h)| < \sum_i \frac{\varepsilon \cdot |t_i|}{n \cdot |h|}$$

and since

$$|t_i| \leq \left(\sum_i t_i^2\right)^{1/2}$$

it follows that $|\theta(h)| < \varepsilon$ when $0 < |h| < \delta_1$.■

COROLLARY. *If f satisfies the conditions of the proposition, then f is continuous at x.*■

The conclusion of the theorem indicates the right generalization of differentiability to functions of more than one variable:

Definition. Let U be an open subset in $\mathbf{R}^n$ and let $f : U \to \mathbf{R}^n$ be a function (with no further assumptions at this point). The function f is said to be *differentiable* at the point $x \in U$ if there is a linear transformation $L : \mathbf{R}^n \to \mathbf{R}^m$ such that for all sufficiently small vectors h we have

$$f(x+h) - f(x) = L(h) + |h|\,\theta(h)$$

where $\lim_{h \to 0} \theta(h) = 0$.

Immediate Consequence. *If $m = 1$ in the definition above and we restrict h so that $h = t\mathbf{e}_i$ for $|t|$ sufficiently small then if f is differentiable at x it follows that all first partial derivatives of f are defined at x and*

$$\frac{\partial f(x)}{\partial x_i} = L(\mathbf{e}_i)\ .$$

In particular, the preceding shows that there is at most one choice of L for which the differentiability criterion is true, at least if $m = 1$.■

Of course the proposition above implies that a function is differentiable if it has continuous partial derivatives.

The differentiability of a function turns out to be determined completely by the differentiability of its coordinate functions:

PROPOSITION. *let U be open in $\mathbf{R}^n$, let $f : U \to \mathbf{R}^m$ be a function, and express f in coordinates as $f(x) = \sum_j y_j(x)\mathbf{e}_j$. Then f is differentiable at x if and only if each y_j is differentiable at x, and in this case the linear transformation L is given by*

$$L(u) = \sum_i \langle \nabla y_i(x), u\rangle \mathbf{e}_i \ .$$

It follows that there is at most one choice of L for an arbitrary value of m. If we write $u = \sum_j u_j\mathbf{e}_j$ then this yields the fundamental identity

$$L(u) = \sum_i \sum_j \frac{\partial y_i(x)}{\partial x_j} u_j \mathbf{e_i}$$

for the derivative linear transformation. In words, the (i,j) entry of the matrix representing L is the j^{th} partial derivative of the i^{th} coordinate function.

Proof of proposition. Suppose that f is differentiable at x. For i between 1 and m, let $\mathbf{P}_i$ be projection onto the i^{th} coordinate. If we apply $\mathbf{P_i}$ to the formula for $f(x+h) - f(x)$ we obtain the following relationship:

$$y_i(x+h) - y_i(x) = \mathbf{P_i}L(h) + |h|\mathbf{P_i}\theta(h)$$

The composite $\mathbf{P}_iL$ is linear because both factors are linear, and the relation

$$\left|\mathbf{P_i}\theta(h)\right| \leq \left|\theta(h)\right|$$

shows that $\lim_{h\to 0} \mathbf{P_i}\theta(h) = 0$, so that $y_i = \mathbf{P}_if$ is differentiable at X and

$$\nabla y_i(x) = \mathbf{P}_iL$$

as required.

Now suppose that each y_i is differentiable at x, and write

$$f(x+h) - f(x) = \sum_i \left(y_i(x+h) - y_i(x)\right)\mathbf{e_i} =$$

$$\sum_i \langle \nabla y_i(x), h\rangle \mathbf{e_i} + \sum_i |h|\theta_i(h)\mathbf{e_i}$$

where $h = \sum_j t_j\mathbf{e}_j$ and $\lim_{h\to 0} \theta_i(h) = 0$ for all i. Choose $\delta > 0$ so that $N_\delta(x) \subset U$ and $0 < |h| < \delta$ implies $|\theta_i(h)| < \varepsilon/n$ for all i. Then $0 < |h| < \delta$ implies $|\theta(h)| < \varepsilon$, whosing that f is differentiable at x and the linear transformation L has the form described in the proposition.■

Smoothness classes of functions

If U and V are open sets in Euclidean spaces and $f : U \to V$ is a function, then we say that f is *(smooth of class)* $\mathbf{C}^1$ if Df exists everywhere and is continuous. For $r \geq 2$ we inductively define f to be *(smooth of class)* $\mathbf{C}^r$ if Df is (smooth of class) $\mathbf{C}^{r-1}$, and we say that f is *(smooth of class)* $\mathbf{C}^\infty$ if it is smooth of class $\mathbf{C}^r$ for all positive integers r. For the sake of notational uniformity we often say that every continuous function is of class $\mathbf{C}^0$.

It is elementary to check that a function is of class $\mathbf{C}^r$ if and only if all its coordinate functions are and that

$$\mathbf{C}^\infty \Longrightarrow \mathbf{C}^r \Longrightarrow \mathbf{C}^{r-1} \Longrightarrow \mathbf{C}^0$$

for all r. Every polynomial function is obviously of class $\mathbf{C}^\infty$, and for each r there are many examples of functions that are $\mathbf{C}^r$ but not $\mathbf{C}^{r+1}$. If $r = 0$ the absolute value function $f_0(x) = |x|$ is an obvious example, and inductively one can construct an example f_r which is $\mathbf{C}^r$ but not $\mathbf{C}^{r+1}$ by taking an antiderivative of f_{r-1}.

One important point in ordinary and multivariable courses is that standard algebraic operations on differentiable (or smooth) functions yield differentiable (or smooth) functions. In particular, this applies to addition, subtraction, multiplication, and division (provided the denominator is nonzero in this case). We shall use these facts without much further comment. The smoothness properties of composites of differentiable and smooth functions will be discussed reasonably soon.

Matrix operations

Plenty of examples of smooth functions can be found in multivariable calculus books, so we concentrate here on some basic examples that will be needed shortly.

PROPOSITION. *Addition addition of $m \times n$ matrices is a $\mathbf{C}^\infty$ map from*

$$\mathbf{R}^{2mn} \cong (\mathbf{M}(m,n))^2$$

to $\mathbf{R}^{mn} \cong \mathbf{M}(m,n)$, scalar multiplication of $m \times n$ matrices is a $\mathbf{C}^\infty$ map from

$$\mathbf{R}^{mn+1} \cong \mathbf{R} \times \mathbf{M}(m,n)$$

to $\mathbf{R}^{mn} \cong \mathbf{M}(m,n)$, and matrix multiplication from $\mathbf{M}(m,n) \times \mathbf{M}(n,p)$ to $\mathbf{M}(m,p)$ is also a $\mathbf{C}^\infty$ map.

This simply reflects the fact that that the entries of a matrix sum or product are given by addition and multiplication operations on the entries of the original matrices (or matrix and scalar).■

The next result is slightly less trivial but still not difficult.

PROPOSITION. *The set of invertible $n \times n$ matrices $GL(n,\mathbf{R})$ is an open subset of $\mathbf{R}^{n^2} \cong \mathbf{M}(n,n)$, and the map from $GL(n,\mathbf{R})$ to itself sending a matrix to its inverse is a $\mathbf{C}^\infty$ map.*

Proof. We shall prove this using coordinates; it is possible to prove the result without using coordinates, but the proof using coordinates is shorter.

Recall that the determinant of a square matrix is a polynomial function in the entries of the matrix and that a matrix is invertible if and only if its determinant is nonzero. The former implies that the determinant function is continuous (and in fact $\mathbf{C}^\infty$), while the second observation and the continuity of the determinant imply that the set of invertible matrices, which is equal to the set $\det^{-1}(\mathbf{R} - \{0\})$, is open. But Cramer's Rule implies that the entries of the inverse to a matrix are rational expression in the entries of the original matrix, and thus the entries of an inverse matrix are $\mathbf{C}^\infty$ functions of the entries of the original matrix. Therefore the matrix inverse is a $\mathbf{C}^\infty$ function from $GL(n,\mathbf{R})$ to itself.■

Matrix norms

Before discussing the metric and topological properties of $\mathbf{C}^r$ functions it is necessary to know a little about the corresponding properties of their linear approximations. The most basic property is a strong form of uniform continuity.

PROPOSITION. *If $L : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, then there is a constant $b > 0$ such that $|L(u)| \leq b|u|$ for all $U \in \mathbf{R}^n$.*

Proof. The easiest way to see this is to note that L is continuous, and therefore the restriction of the function $h(u) = |L(u)|$ to the (compact!) unit sphere in $\mathbf{R}^n$ assumes a maximum value, say c, so that $|u| = c$. Every vector u may be written as a product cv where c is nonnegative and $|v| = 1$. It then follows that

$$|L(u)| = |cL(v)| = c \cdot |L(v)| = |u| \cdot |L(v)| \leq b \cdot |u|$$

as required.∎

Notation. The maximum value b is called the *norm* of L and written $\| L \|$.

PROPOSITION. *The norm of a matrix (or linear transformation) makes the space of $m \times n$ matrices into a normed vector space.*

Proof. By definition the norm is nonnegative, and if it is zero then $L(v) = 0$ for all $v \in \mathbf{R}^n$ satisfying $|v| = 1$; it follows that $L(v) = 0$ for all v (why?). If a is a scalar and L is a linear transformation, then the maximum value $\| L \|$ of $|aL(v)|$ for $|v| = 1$ is simply $|a| \cdot \| L \|$. Finally, if L_1 and L_2 are linear transformations and v is a unit vector such that $|[L_1 + L_2](v)| = \| L_1 + L_2 \|$, then we have

$$\| L_1 + L_2 \| = |[L_1 + L_2](v)| \ \leq \ |L_1(v)| + |L_2(v)| \ \leq \| L_1 \| + \| L_2 \| \ .$$

Thus the norm as defined above satisfies the conditions for a normed vector space.∎

The matrix norm has the following additional useful property:

PROPOSITION. *If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then $\| AB \| \leq \| A \| \cdot \| B \|$.*

Proof. Let v be a unit vector in $\mathbf{R}^p$ at which the function $f(x) = ABx$ takes a maximum value. Then we have

$$|ABx| \quad \leq \quad \| A \| \cdot |Bx| \quad \leq \quad \| A \| \cdot \| B \|$$

as required.∎

Comparisons of norms

Although there are many different norms that can be defined on $\mathbf{R}^n$, the following result shows that they all yield the same open sets.

THEOREM. *Let $|...|$ denote the standard Euclidean norm on $\mathbf{R}^n$, and let $\| ... \|$ denotes some other norm. Then there are positive constants A and B such that*

$$\| x \| \leq \ A\,|x| \ , \quad |x| \ \leq \ B \,\| x \|$$

for all $x \in \mathbf{R}^n$. In particular, the identity maps of normed vector spaces from Euclidean space $(\mathbf{R}^n, |...|)$ to $(\mathbf{R}^n, \| ... \|)$ and vice versa are uniformly continuous.

Proof. Given a typical vector x, write it as $\sum_i x_i \mathbf{e}_i$.

Choose M such that $\| \mathbf{e}_i \| \leq M$ for all i. Then we have

$$\| x \| \;\leq\; \sum_i |x_i| \cdot \; \| \mathbf{e}_i \| \;\leq\; nM \sum_i |x_i|$$

and by the Cauchy-Schwarz-Buniakovsky Inequality the summation is less than or equal to $n^{1/2}|x|$. Therefore $\| x \| \leq n^{3/2} M |x|$.

The preceding paragraph implies that the function $f(x) = \| x \|$ is a continuous function on $\mathbf{R}^n$ with respect to the usual Euclidean metric. Let $c > 0$ be the minimum value of f on the unit sphere defined by $|x| = 1$. It then follows that $\| x \| \geq c \cdot |x|$ for all x and hence that

$$|x| \;\leq\; \frac{1}{c} \, \| x \|$$

for all $x \in \mathbf{R}^n$.■

COROLLARY. *The same conclusion holds if the Euclidean norm is replaced by a second arbitrary norm.*

Proof. Let α and β denote arbitrary norms on $\mathbf{R}^n$ and let $|...|$ denote the Euclidean norm. Then there are positive constants A_α, A_β, B_α B_β such that the following hold for all $x \in \mathbf{R}^n$:

$$\alpha(x) \;\leq A_\alpha |x| \;, \quad |x| \;\leq\; B_\alpha \, \alpha(x)$$

$$\beta(x) \;\leq A_\beta |x| \;, \quad |x| \;\leq\; B_\beta \, \beta(x)$$

These immediately imply $\alpha(x) \leq A_\alpha B_\beta \beta(x)$ and $\beta(x) \leq A_\beta B_\alpha \alpha(x)$.■

The following observation will be useful later.

PROPOSITION. *If A is an $n \times n$ matrix such that $\| A \| < 1$, then $I - A$ is invertible.*

Proof. Suppose that the conclusion is false, so that $I - A$ is not invertible. Then there is a nonzero vector $v \in \mathbf{R}^n$ such that $(I - A)v = 0$. The latter implies that $Ax = x$ for some x such that $|x| = 1$, which in turn implies that $\| A \| \geq 1$.■

Note. If $\| A \| < 1$, then the the previously stated inequalities for the matrix norm show that

$$\| A^k \| \;\leq\; \| A \|^k$$

and the latter implies that the inverse to $I - A$ may be computed using the geometric series:

$$(I - A)^{-1} = \sum_k A^k$$

IV.2 : Properties of smooth functions

(Edwards, § II.3)

In this section we shall establish generalizations of two important principles from elementary calculus. One is a version of the Chain Rule, and the other is a general form of a basic consequence of the Mean Value Theorem. As an application of these results we shall show that the restriction of a smooth map to a compact set satisfies a metric inequality called a *Lipschitz condition* that generalizes the strong form of uniform continuity which holds for linear maps of (finite-dimensional) Euclidean spaces.

The Chain Rule

Undergraduate multivariable calculus courses generally state one or more extensions of the ordinary Chain Rule

$$[g \circ f]'(x) = g'(f(x)) \cdot f'(x)$$

to functions of several real variables. Linear transformations provide a conceptually simple way of summarizing the various generalizations of this basic fact from single variable calculus:

CHAIN RULE. *Let U and V be open in $\mathbf{R}^n$ and $\mathbf{R}^m$ respectively, let $f : U \to V$ be a map that is differentiable at x, and let $g : V \to \mathbf{R}^p$ be differentiable at $f(x)$. Then $g \circ f$ is differentiable at x and*

$$D[g \circ f](x) = D(g)(f(x)) \circ Df(x) \ .$$

Proof. By the definition of differentiability for g at $y = f(x)$, for $|k|$ sufficiently small we have

$$g(y + k) - g(y) = [Dg(y)](k) + |k| \cdot \alpha(k)$$

where $\lim_{k \to 0} \alpha(k) = 0$. If we take k so that $y + k = f(x + h)$ for $|h|$ sufficiently small, then we have $k = f(x + h) - f(x)$, and by the differentiability of f at x we have

$$k = f(x + h) - f(x) = [Df(x)](h) + |h| \cdot \beta(h)$$

where $\lim_{h \to 0} \beta(h) = 0$. If we make this substitution into the first equation in the proof we obtain the relation

$$[g \circ f](x + h) - [g \circ f](x) = [Dg(f(x))]([Df(x)](h) + |h| \cdot \beta(h)) + |k|\alpha(f(x + h) - f(x))$$

and for $h \neq 0$ the right hand side may be rewritten as follows:

$$[Dg(f(x))]([Df(x)](h)) + [Dg(f(x))](|h| \cdot \beta(h)) + |h| \cdot \frac{|k|}{|h|}\, \alpha\big(f(x + h) - f(x)\big)$$

Let $\varepsilon > 0$ be given. We need to show there is a $\delta > 0$ such that $|h| < \delta$ implies the following inequalities:

$$\| \, Dg(f(x)) \, \| \cdot |\beta(h)| \ \ < \ \ \frac{\varepsilon}{2}$$

$$\frac{|k|}{|h|} \quad < \quad \| \, Df(x) \, \| + 1$$

$$\bigl|f(x+h) - f(x)\bigr| \quad < \quad \frac{\varepsilon}{2(\| \, Df(x) \, \| + 1)}$$

The first of these can be realized by the limit condition on β, and the third can be realized by the limit condition on α and the continuity of f at x. To deal with the second condition, note that

$$|k| = \bigl|f(x+h) - f(x)\bigr| = \bigl|[Df(x)](h) + |h|\beta(h)\bigr| \leq$$

$$\| \, Df(x) \, \| \cdot |h| \quad + \quad |h|\bigl|\beta(h)\bigr|$$

so that

$$\frac{|k|}{|h|} \quad \leq \quad \| \, Df(x) \, \| \quad + \quad \bigl|\beta(h)\bigr|$$

and thus we may realize the second inequality if we choose δ so that $0 < |h| < \delta$ implies

$$\bigl|\beta(h)\bigr| \quad < \quad \frac{\varepsilon}{2(\| \, Df(x) \, \| + 1)} \ .$$

If we combine all these we see that

$$[g \circ f](x+h) \ - \ [g \circ f](x) \quad = \quad [Dg\,(f(x))]\,([Df(x)](h)) \ + \ |h| \ \gamma(h)$$

where $\lim_{h \to 0} \gamma(h) = 0$.■

COROLLARY. *In the notation of the preceding result, if f is $\mathbf{C}^r$ on U and g is $\mathbf{C}^r$ on V, then $g \circ f$ is $\mathbf{C}^r$.*

Proof. Suppose that $r = 1$. Then the Chain Rule formula, the continuity of the derivatives of f and g and the continuity of f show that $D(g \circ f)$ is continuous.

Suppose now that $r \geq 2$ is an integer and we have shown the Corollary inductively for $\mathbf{C}^s$ functions for $1 \leq s \leq r - 1$. Then the functions $Dg \circ f$ and Df are $\mathbf{C}^{r-1}$ by the induction hypothesis and the fact that f is $\mathbf{C}^r$, and the matrix product of these functions is also $\mathbf{C}^r$ because matrix multiplication is $\mathbf{C}^\infty$.■

Since the result is true for all finite r, it follows immediately that the conclusion is also true if $r = \infty$.

Example. Suppose that U is open in $\mathbf{R}^n$ and $f : U \to \mathbf{R}^m$ is $\mathbf{C}^r$ for some $r \geq 1$; let $a, x \in U$ be such that $N_{2|x-a|}(a) \subset U$. Then the function

$$g(t) = f\,(a + (x - a))$$

is a $\mathbf{C}^r$ function on some interval $(-\delta, 1 + \delta)$ and

$$g'(t) = \bigl[Df\,(a + t(x-a))\,\bigr](x - a) \ .$$

The Mean Value Theorem for real-valued differentiable functions of one real variable does not generalize directly to other situations, but some of its important consequences involving derivatives and definite integrals can be extended. The following example is important in many contexts:

PROPOSITION. *Let U be open in $\mathbf{R}^n$, let $f : U \to \mathbf{R}^m$ be a $\mathbf{C}^1$ function, and suppose that $a \in U$ and $\delta > 0$ are such that $|x - a|, |y - a| \leq \delta$ implies $x, y \in U$. Then for all such x we have the following inequality:*

$$|f(x) - f(y)| \quad \leq \quad \max_{|z-a|\leq\delta} \parallel Df(z) \parallel \cdot |x - y|$$

Proof. Let $h = x - y$, and set $g(t) = f(y + th)$; since open disks are convex we know that $y + th$ also satisfies $|y + th - a| \leq \delta$. Then we have

$$f(x) - f(y) = \int_0^1 g'(t)\, dt$$

and therefore

$$\big|f(x) - f(y)\big| \;\leq\; \int_0^1 \big|g'(t)\big|\, dt\ .$$

As indicated before, by the Chain Rule we know that

$$g'(t) = \big[Df\,(y + t(x - y))\,\big](x - y)$$

and therefore we have the estimate

$$\int_0^1 \big|g'(t)\big|\, dt \quad \leq \quad \max_{0\leq t\leq 1} \parallel Df\,(y + t(x - y)) \parallel \cdot \big|x - y\big|$$

which immediately yields the inequality in the proposition. ∎

Lipschitz conditions

The restriction of a smooth function (say of class $\mathbf{C}^r$) to a compact set satisfies a strong form of uniform continuity that generalizes the matrix inequality $|Ax| \leq \parallel A \parallel \cdot |x|$.

THEOREM. *Let U be open in $\mathbf{R}^n$, let $f : U \to \mathbf{R}^m$ be a $\mathbf{C}^1$ function, and let $K \subset U$ be compact. Then there is a constant $B > 0$ such that*

$$|f(u) - f(v)| \;\leq\; B\,|u - v|$$

for all $u, v \in K$ such that $u \neq v$.

The displayed inequality is called a *Lipschitz condition* for f. This strong form of uniform continuity associates to each $\varepsilon > 0$ a corresponding δ equal to ε/B. An example of a function not satisfying any Lipschitz condition is given by $h(x) = \sqrt{x}$ on the closed unit interval $[0, 1]$ (use the Mean Value Theorem and $\lim_{t\to 0^+} h'(t) = +\infty$). Incidentally, the inverse of this map is a homeomorphism that does satisfy a Lipschitz condition (*e.g.*, we can take $B = 2$).

The inequality

$$\Big|u - v\Big| \quad \geq \quad \Big||u| - |v|\Big|$$

for $u, v \in \mathbf{R}$ shows that $f(x) = |x|$ is a function that satisfies a Lipschitz condition but is not $\mathbf{C}^1$.

A *Lipschitz constant* for f on a set K (not necessarily compact) is a number $B > 0$ such that $|f(u) - f(v)| \leq B\,|u - v|$ for all $u, v \in K$ such that $u \neq v$. Note that Lipschitz constants are definitely nonunique; if B is a Lipschitz constant for f on a set K and $C > B$, then C is also a Lipschitz constant for f on a set K.

Proof of theorem. For each $x \in K$ there is a $\delta(x) > 0$ such that $N_{2\delta(x)}(x) \subset U$. By compactness there are finitely many points $x_1, \cdots, x_q$ such that the sets $N_{\delta(x_i)}(x_i)$ cover K. Let B_i be the maximum of $\| \, Df \, \|$ for $|y - x_i| \leq \delta(x_i)$. If $B_i = 0$ for all i then $Df = 0$ on an open set containing K and therefore f is constant on K, so that the conclusion of the theorem is trivial. Therefore we shall assume some $B_i > 0$ for the rest of the proof.

By the Mean Value Estimate we know that $y, z \in N_{\delta(x_i)}(x_i)$ implies that $|f(y) - f(z)| \leq B_i|y - z|$.

Let $\eta > 0$ be a Lebesgue number for the open covering of K determined by the sets $N_{\delta(x_i)}(x_i)$, and let $M \subset K \times K$ be the set of all points $(u, v) \in K \times K$ such that $|u - v| \geq \eta/2$. The function $\Delta(u, v) = |u - v|$ is continuous on $K \times K$, and consequently it follows that M is a closed and thus compact subset of $K \times K$. Consider the continuous real-valued function on M defined by

$$h(u, v) = \frac{|f(u) - f(v)|}{|u - v|} \; .$$

Since the denominator is positive on M, this is a continuous function and therefore attains a maximum value A.

Let B be the maximum of the numbers $A, B_1, \cdots, B_k$, and suppose that $(u, v) \in K \times K$. If $(u, v) \in M$, then by the preceding paragraph we have $|f(u) - f(v)| \leq A \cdot |u - v|$. On the other hand, if $(u, v) \notin M$, then $|u - v| < \eta/2$ and thus there some i such that $u, v \in N_{\delta(x_i)}(x_i)$. By the Mean Value Estimate we know that $|f(u) - f(v)| \leq B_i \cdot |u - v|$ in this case. Therefore $|f(u) - f(v)| \leq B \cdot |u - v|$ for all u and v.■

IV.3 : Inverse Function Theorem

(Edwards, §§ III.2, III.3)

We have already mentioned the general question of recognizing when a 1–1 onto continuous function from one space to another has a continuous inverse. There are also many situations where it is useful to know simply whether a **local inverse** exists. For real valued functions on an interval, the Intermediate Value Property from elementary calculus implies that local inverses exist for functions that are strictly increasing or strictly decreasing (we have not actually proved this yet, however). Since the latter happens if the function has a derivative that is everywhere positive or negative close to a given point, one can use the derivative to recognize very quickly whether local inverses exist in many cases, and in these cases one can even compute the derivative of the inverse function using the standard formula:

$$g = f^{-1} \implies g'(y) = \frac{1}{f'(g(y))}$$

Of course this formula requires that the derivative of f is not zero at the points under consideration.

There is a far-reaching generalization of the single variable inverse function theorem for functions of several real variables. It is covered in many but not all courses on multivariable calculus or undergraduate courses on the theory of functions of a real variable, but even when it is covered the treatment is sometimes incomplete (for example, only worked out for functions of two or at most three variables).

INVERSE FUNCTION THEOREM. *Let U be open in $\mathbf{R}^n$, let $a \in U$, and let $f : U \to \mathbf{R}^n$ be a $\mathbf{C}^r$ map (where $1 \leq r \leq \infty$) such that $Df(a)$ is invertible. Then there is an open set W containing a such that the following hold:*

(i) The restriction of f to W is $1-1$ and its image is an open subset V.

(ii) There is a $\mathbf{C}^r$ inverse map $g : V \to U_0$ such that $g(f(x)) = x$ on U_0.

Proof. The proof given here uses the Contraction Lemma.

It is convenient to reduce the proof to the special case where $a = f(a) = 0$ and $Df(a) = I$. Suppose we know the result in that case. Let $A = Df(a)$, and define f_1 by the formula

$$f_1(x) = A^{-1}\left(f(x+a) - f(a)\right) \ .$$

Then $f_1(0) = 0$ and by the Chain Rule we have $Df_1(0) = I$. Then assuming the conclusion of the theorem is known for f_1, we take W_1, V_1, g_1 as in that conclusion. If we take $W = a + W_1$, $V = AV_1 + f(a)$, and

$$g(z) = g_1\left(A^{-1}\left(y - f(a)\right)\right) + a$$

then it follows immediately that the function $f(y) = Af_1(y-a) + f(a)$ satisfies the conclusions of the theorem.

Since f has a continuous derivative, there is a $\delta > 0$ such that $|x| \leq \delta$ implies $\| Df(x) - I \| < \frac{1}{2}$. For each y on the closed disk D of radius $\delta/2$ about the origin, define a map T on D by the formula $T(x) = x + y - f(x)$, and observe that $T(x) = x$ if and only if $y = f(x)$.

We want to apply the Contraction Lemma to T. The first step is to show that T maps D to itself. Let φ be the function $\varphi(x) = x - f(x)$; then $\varphi(0) = 0$ and $D\varphi = I - Df$, and consequently by the Mean Value Estimate we have that $|\varphi(x)| \leq |x|/2$ if $|x| \leq \delta$. Since $T(x) = y + \varphi(x)$ and $|y| < \delta/2$ it follows that $|T(x)| \leq \delta$ and hence $T(D) \subset D$.

We now need to estimate $|T(x_1) - T(x_0)|$ in terms of $|x_1 - x_0|$. Since T and φ differ by a constant it follows from the Mean Value Estimate that

$$\left|T(x_1) - T(x_0)\right| = \left|\varphi(x_1) - \varphi(x_0)\right| \leq \frac{1}{2}|x_1 - x_0|$$

and therefore the Contraction Lemma implies the existence of a unique point $x \in D$ such that $T(x) = x$, which is equivalent to $f(x) = y$.

Let $g : N_{\delta/2}(0) \to D$ be the inverse map sending a point y to the unique x such that $f(x) = y$. We claim that g is continuous. The first step is to show that $|x_0|,\ |x_1| \leq \delta/2$ implies $|f(x_1)-f(x_0)| \geq \frac{1}{2}|x_1 - x_0|$. To see this, use the identity $f(x) = x - \varphi(x)$ and use the equation and inequalities

$$|f(x_1) - f(x_0)| \geq |x_1 - x_0| - \left|\varphi(x_1) - \varphi(x_0)\right| \geq |x_1 - x_0| - \frac{1}{2}|x_1 - x_0| = \frac{1}{2}|f(x_1) - f(x_0)| \ .$$

If we set $y_i = f(x_i)$ so that $x_i = g(y_i)$ then we have

$$|x_1 - x_0| = |g(y_1) - g(y_0)| \ \leq \ 2 \cdot |y_1 - y_0|$$

and thus g is uniformly continuous.

Let U_0 be the image of g; we claim that U_0 is open. Suppose that $x \in U_0$, so that $f(x) = y$ where $|y| < \delta/2$. Then one can find some $\eta > 0$ so that $|z - x| < \eta$ implies $|f(z)| < \delta/2$ (why?) and the identity $g(f(z)) = z$ then implies that $z \in \mathrm{Image}(g)$. Thus we may take $V = N_{\delta/2}(0)$ and $U_0 = g(V)$.

Finally, we need to show that g is a $\mathbf{C}^r$ function if f is a $\mathbf{C}^r$ function. Given $y \in V$ and k such that $y + k \in V$, write $y = f(x)$ and $y + k = f(x + h)$. Since $\| \, Df(x) - I \, \| < 1$ it follows that $Df(x)$ is invertible. Let L be its inverse. Then we have

$$g(y + k) - g(y) - L(k) = h - L(k) = -L\left(f(x + h) - f(x) - Df(x)h\right)$$

and the right hand side is equal to

$$L\left(|h| \cdot \theta(h)\right)$$

where $\lim_{|h|\to 0} \theta(h) = 0$. Since $h = g(y + k) - g(y)$ we know that $|h| \leq 2|k|$ and therefore we also have

$$\lim_{|k|\to 0} \frac{1}{|k|} \cdot L\left(|h| \cdot \theta(h)\right) = 0$$

(where $h = g(y + k) - g(y)$ as above), which shows that g is differentiable at y and satisfies a familiar looking formula:

$$Dg(y) = \left(Df\left(g(y)\right)\right)^{-1}$$

Since the entries of an inverse matrix are rational expressions in the inverse of the original matrix, the continuity of g and the $\mathbf{C}^1$ property of f imply that g is also $\mathbf{C}^1$.

If f is a $\mathbf{C}^r$ function, one can now prove that g is a $\mathbf{C}^s$ function for all $s \leq r$ inductively as follows: Suppose we know that f is $\mathbf{C}^r$ and g is $\mathbf{C}^s$ for $1 \leq s < r$. By the formula for the derivative of g we know that Dg is formed by the composite of g, Df and matrix inversion. We know that g is $\mathbf{C}^s$, that Df is too because f is $\mathbf{C}^{s+1}$ (recall that $s + 1 \leq r$), and that inversion is $\mathbf{C}^\infty$ because its entries are given by rational functions, and therefore it follows that $D[g \circ f]$ is also $\mathbf{C}^s$, which means that $g \circ f$ is $\mathbf{C}^{s+1}$.■

COROLLARY. *Let U and V be open in $\mathbf{R}^n$, and let $f : U \to V$ be $1-1$ onto and $\mathbf{C}^r$ where $1 \leq r \leq \infty$. Then f^{-1} is also $\mathbf{C}^r$.*■

A similar result holds when $r = 0$, but the proof requires entirely different methods which come from algebraic topology.

The Inverse Function Theorem also has the following purely topological consequence for $\mathbf{C}^1$ mappings:

COROLLARY. *Let $f : U \to \mathbf{R}^n$ be a $\mathbf{C}^1$ function ($r \geq 1$), where U is open in $\mathbf{R}^n$, and assume that $Df(x)$ is invertible for all $x \in U$. Then f is open.*■

There is also an extension of this result to the $\mathbf{C}^0$ case provided f is locally 1–1 (this is Brouwer's *Invariance of Domain* Theorem); the proof again requires methods from algebraic topology.

Proof. Let W be open in U and let $x \in W$. Then the Inverse Function Theorem implies that there is an open subset $W_0(x) \subset W$ containing x such that f maps $W_0(x)$ onto an open subset $V(x)$ in $\mathbf{R}^n$. Therefore it follows that

$$f(W) \;=\; \bigcup_x f\big(W_0(x)\big) \;=\; \bigcup_x V_x$$

which is open in $\mathbf{R}^n$.

Examples. Consider the complex exponential mapping f from $\mathbf{R}^2$ to itself sending (x, y) to $(e^x \cos y, e^x \sin y)$. The derivative of this map is invertible at every point but the map is not 1–1 because every nonzero point in $\mathbf{R}^2$ is the image of infinitely many points; specifically, for every (x, y) and integer k we have $f(x, y + 2k\pi) = f(x, y)$.

Another example of this type is the complex square mapping $f : \mathbf{R}^2 - \{0\} \to \mathbf{R}^2$ sending (x, y) to $(x^2 - y^2, 2xy)$, which has the property that $f(-x, -y) = f(x, y)$ for all (x, y); note that if we write $z = u + iv$, then $f(z) = z^2$.

Final Remark. Given a $\mathbf{C}^1$ function $f : U \to \mathbf{R}^n$ with U open in $\mathbf{R}^n$, if we write the coordinate functions of f as $y_1, \cdots y_n$ then $\det Df(p)$ is just the classical *Jacobian function*

$$\frac{\partial(y_1, \cdots, y_n)}{\partial(x_1, \cdots, x_n)}(p)$$

and with this terminology the condition on $Df(p)$ in the Inverse Function Theorem may be rephrased to state that the Jacobian at p is nonzero.

The Implicit Function Theorem

There is a close relation between the Inverse Function Theorem and the standard Implicit Function Theorem from ordinary and multivariable calculus. In its simplest form the Implicit Function Theorem states that **locally** one can solve an equation $F(x, y) = 0$ uniquely for y in terms of x; more precisely, if $F(a, b) = 0$ and the second partial derivative of F is nonzero at (a, b), then on some open interval $(a-\delta, a+\delta)$ there is a unique function $f(x)$ such that $y = f(x) \iff F(x,y) = 0$ (hence $f(a) = b$) and

$$\frac{df}{dx} \quad = \quad - \frac{\left(\frac{\partial F}{\partial x}\right)}{\left(\frac{\partial F}{\partial y}\right)} \quad .$$

Here is a general version of this result:

IMPLICIT FUNCTION THEOREM. *Let U and V be open in $\mathbf{R}^n$ and $\mathbf{R}^m$ respecitvely, and let $f : U \times V \to \mathbf{R}^m$ be a smooth function such that for some $(x,y) \in U \times V$ we have $f(x,y) = 0$ and the partial derivative of f with respect to the last m coordinates is invertible. Then there is an open neighborhood U_0 of x and a smooth function $g : U_0 \to V$ such that $g(x) = y$ and for all $u \in U_0$ we have $f(u,v) = 0$ if and only if $v = g(u)$.*

For the sake of completeness we note that the partial derivative of f with respect to the last m coordinates is the derivative of the function $f^*(v) = f(x,v)$, and that smooth means smooth of class $\mathbf{C}^r$ for some r such that $1 \leq r \leq \infty$.

Proof. Define $h : U \times V \to \mathbf{R}^m \times \mathbf{R}^n$ by $h(u,v) = (f(u,v), u)$. Then the hypotheses imply that $Dh(x,y)$ is invertible, and therefore by the Inverse Function Theorem there is a local inverse

$$k : \mathrm{Int}\ (\varepsilon D^m) \times U_0 \longrightarrow U \times V\ .$$

Since the second coordinate of $h(u,v)$ is u, it follows that the first coordinate of the inverse $k(z,w)$ is w so that we may write $k(z,w) = (w,\ Q(z,w)\,)$ for some smooth function Q.

On one hand we have $g\,(\,k(z,w)\,) = (z,w)$, but on the other hand we also have

$$g\,(k(z,w)\,) \quad = \quad g\,(w, Q(z,w)) \quad = \quad \big(f\,(w, Q(z,w))\,, w\big).$$

In particular, this means that

$$z \quad = \quad f\big(u, Q(z,u)\big)$$

for all z and u. If we take $g(u) = Q(0,u)$ it follows that $y = g(x)$ and $f(u,v) = 0$ if and only if $v = g(u)$.∎

One can use the Chain Rule to calculuate $Dg(u)$ as follows: If $\varphi(u) = f(u, g(u))$ and $\mathbf{p} \in \mathbf{R}^n$, then the Chain Rule yields the formula

$$[D\varphi(u)](\mathbf{p}) \quad = \quad [D_1 f\,(u, g(u)\,)]\,(\mathbf{p}) \;+\; [D_2 f\,(u, g(u)\,)]\,(\,[Dg(u)](\mathbf{p})\,)$$

where D_1 and D_2 refer to partial derivatives with respect to the first and last sets of variables. Since $\varphi = 0$ it also follows that $[D\varphi(u)](\mathbf{p}) = 0$. Furthermore, if u and v are sufficiently close to x and y then the second partial derivative is invertible. Therefore one obtains the formula

$$Dg \quad = \quad -\,(D_2 f)^{-1} \circ D_1 f$$

which generalizes the formula in elementary multivariable calculus.∎

V. Constructions on Spaces

This material in this unit concerns two basic themes that run throughout geometry — all the way from children's toys to the frontiers of research. One of these is the construction of new objects from old ones by gluing certain subsets together. For example, one can form a circle (actually, a space homeomorphic to a circle) from a closed interval by gluing the two endpoints together. A second theme is the formation of new objects from old ones by first creating several disjoint copies of the original objects and then gluing pieces together in an appropriate fashion. For example, one can form a cube from six pairwise disjoint squares with sides of equal length by gluing the latter together in a suitable way along the edges.

Two widely known examples of such constructions are the Möbius strip and the Klein bottle, and we shall indicate how they can be formed using the ideas presented here. The latter can also be used to show that one can construct a Klein bottle from two Möbius strips by gluing the latter together along their (homeomorphic) edges.

V.I : Quotient spaces

(Munkres, § 22)

We shall adopt a somewhat different approach from the one appearing in Munkres. In mathematics it is often useful to create a quotient object from a mathematical structure and a well-behaved equivalence relation on such an object. For example, if n is a positive integer greater than 1, then one can construct the ring $\mathbf{Z}_n$ of integers modulo n using the equivalence classes of the relation

$$a \equiv b(n) \iff a - b = kn, \text{ some } k \in \mathbf{Z}$$

and the projection from $\mathbf{Z}$ to the set of equivalence classes $\mathbf{Z}_n$ is compatible with the addition and multiplication on both systems. One often says that $\mathbf{Z}_n$ is a *quotient ring* of $\mathbf{Z}$, and further constructions of this sort are indispensable in most of abstract algebra.

More generally, quotient constructions arise naturally for many types of mathematical systems, so the following question about topological spaces arises naturally at least from a formal perspective:

Question. *Let X be a topological space and let $\mathcal{R}$ be an equivalence relation on X. Is there a reasonable definition of a topology on the set of equivalence classes $X/\mathcal{R}$?*

One obvious requirement is that the projection map

$$\pi_{\mathcal{R}} : X \to X/\mathcal{R}$$

taking a point x to its $\mathcal{R}$-equivalence class $[x]_{\mathcal{R}}$ (frequently abbreviated to $[x]$) is continuous. One trivial way of achieving this is to take the indiscrete topology on $X/\mathcal{R}$, but something this easy should seem too good to be true (and it is). For example, if X is Hausdorff we would like the topology on $X/\mathcal{R}$ to be Hausdorff, at least as often as possible (there are fundamental examples to show that one cannot always have a Hausdorff topology).

It turns out that the right topology to take for $X/\mathcal{R}$ is the unique maximal topology for which $\pi_{\mathcal{R}}$ is continuous, and it is useful to formulate things somewhat more generally.

Definition. Let $(X, \mathbf{T})$ be a topological space, let Y be a set, and let $f : X \to Y$ be a map of sets. The *quotient topology* $f_*\mathbf{T}$ on Y is defined by the condition

$$V \in f_*\mathbf{T} \text{ if and only if } f^{-1}(V) \in \mathbf{T}.$$

Before proceeding, we need to check that *the construction above yields a topology on $X/\mathcal{R}$.* — The inverse image of the empty set is the empty set and the inverse image of $X/\mathcal{R}$ is X, so $f_*\mathbf{T}$ contains the empty set and $X/\mathcal{R}$. If U_α lies in $X/\mathcal{R}$ for all α, then $f^{-1}(\cup_\alpha U_\alpha) = \cup_\alpha f^{-1}(U_\alpha)$ where each term on the right hand side lies in $\mathbf{T}$ by the definition of $f_*\mathbf{T}$; since the union of open sets in X is again an open subset, it follows that $f^{-1}(\cup_\alpha U_\alpha)$ is open in X which in turn implies that $\cup_\alpha U_\alpha$ belongs to $f_*\mathbf{T}$. Likewise, if U_1 and U_2 are in $f_*\mathbf{T}$, then $f^{-1}(U_1 \cap U_2) = f^{-1}(U_1) \cap f^{-1}(U_2)$, and each term of the right hand side lies in $\mathbf{T}$; since the latter is a topology for x it follows that the right hand side also lies in $\mathbf{T}$, and therefore it follows that $U_1 \cap U_2$ lies in $f_*\mathbf{T}$.

The following are immediate consequences of the definition and the preceding paragraph.

PROPOSITION. *(i) f defines a continuous map from $(X, \mathbf{T})$ to $(Y, f_*\mathbf{T})$,*

(ii) $f_*\mathbf{T}$ contains **every** *topology $\mathbf{U}$ for which $f : (X, \mathbf{T}) \to (Y, \mathbf{U})$ is continuous.*

(iii) A subset $B \subset Y$ is closed with respect to $f_*\mathbf{T}$ if and only if its inverse image $f^{-1}(B)$ is closed with respect to $\mathbf{T}$.

Proof. The first statement follows because a set $V \subset Y$ is open if and only if $f^{-1}(V)$ is open in X. The second is verified by noting that if $\mathbf{U}$ is given as above and $W \in \mathbf{U}$, then $f^{-1}(W) \subset \mathbf{T}$ by continuity, and hence $\mathbf{U} \subset \mathbf{T}$. The third statement holds because B is closed with respect to the quotient topology if and only if $Y - B$ is open, which is true if and only if $f^{-1}(Y - B) = X - f^{-1}(B)$ is open in X, which in turn is true if and only if $f^{-1}(B)$ is closed in X.■

Quotients and morphisms

It is helpful to deal first with some aspects of equivalence class projections that are entirely set-theoretic and relate the discussion here to the approach in Munkres.

If X is a set and $\mathcal{R}$ is an equivalence relation on X, then the equivalence class projection $\pi_{\mathcal{R}}$ is onto. In fact, every onto map can be viewed as an equivalence class projection as follows: If $f : X \to Y$ is an arbitrary onto map of sets, and then one can define an equivalence relation $\mathcal{R}_f$ on X by $u \sim v$ if and only if $f(u) = f(v)$. There is a canonical 1–1 correspondence $h : Y \to X/\mathcal{R}_f$ which sends the equivalence class $[x]$ to $f(x)$. It is an elementary exercise to show that h is well-defined, 1–1 and onto (verify this!).

The following set-theoretic observation describes an fundamental property of quotient constructions with respect to functions.

PROPOSITION. *Let $f : X \to Y$ be a function, let $\mathcal{R}$ be an equivalence relation on X, and let $p : X \to X/\mathcal{R}$ be the map sending an element to its equivalence class. Suppose that whenever $u \sim_{\mathcal{R}} v$ in X we have $f(u) = f(v)$. Then there is a unique function $g : X/\mathcal{R} \to Y$ such that $g \circ p = f$.*

The point is that a well-defined function is obtained from the formula $g([x]) = f(x)$. Analogous results hold for a wide range of mathematical structures; a version for topological spaces is given immediately after the proof below.

Proof. Suppose that $[x] = [y]$; then by definition of the equivalence relation we have $f(x) = f(y)$, and therefore the formula defining g yields

$$g([x]) = f(x) = f(y) = g([y])$$

which shows that g is indeed well-defined.$\blacksquare$

COROLLARY. *Let f, X, Y, $\mathcal{R}$ be as in the proposition, suppose they satisfy the conditions given there, suppose there are topologies on X and Y such that f is continuous, and put the quotient topology on $X/\mathcal{R}$. Then the unique map g is continuous.*

Suppose that W is open in Y. Then $g^{-1}(W)$ is open in $X/\mathcal{R}$ if and only if $\pi_{\mathcal{R}}^{-1}(g^{-1}(W)$ is open in X. But $g\circ\pi_{\mathcal{R}} = f$, and hence $\pi_{\mathcal{R}}^{-1}(g^{-1}(W) = f^{-1}(W)$. Since f is continuous, the latter is in fact open in X. Therefore $g^{-1}(W)$ is indeed open, so that g is continuous as required.$\blacksquare$

The following relates our approach to that of Munkres.

DEFINITION. Let $f : X \to Y$ be continuous and onto, let $\mathcal{R}_f$ be the equivalence relation described above, and let $h : X/\mathcal{R} \to Y$ be the standard 1–1 onto map described above. By construction, h is continuous if $X/\mathcal{R}$ is given the quotient topology. We say that f *is a quotient map* if h is a homeomorphism.

By the definitions, f is a quotient map if and only if for every $B \subset Y$ we have that B is open in Y if and only if $f^{-1}(B)$ is open in X. — The statement remains true if one replaces "open" by "closed" everywhere.

Remark. In many books the quotient topology is only defined for continuous maps that are onto, so we shall comment on what happens if f is not onto. In this case, if $y \notin f(X)$ then $\{y\}$ is an open and closed subset because $f^{-1}(\{y\}) = \emptyset$ and $\emptyset$ is open and closed in X; more generally, every subset of $Y - f(X)$ is open and closed for the same reason. It also follows that $f(X)$ is open and closed because its inverse image is the open and closed subset X. The quotient topology on the entire space Y is given by the quotient topology on $f(X)$ with respect to the onto map $g : X \to f(X)$ determined by f and the discrete topology on $Y - f(X)$.

Recognizing quotient maps

The following result provides extremely useful criteria for concluding that a continuous onto map is a quotient map.

PROPOSITION. *Suppose that $f : X \to Y$ is continuous and onto, and also assume that f is* **either** *an open mapping or a closed mapping. Then f is a quotient map.*

Proof. We shall only do the case where f is open; the other case follows by replacing "open" with "closed" everywhere in the argument.

We need to show that V is open in Y if and only if $f^{-1}(V)$ is open in X. The ($\Longrightarrow$) implication is true by continuity, and the other implication follows from the hypothesis that f is open because if $f^{-1}(V)$ is open in X then

$$V \;=\; f\big(f^{-1}(V)\big)$$

must be open in Y.$\blacksquare$

Exercise 3 on page 145 of Munkres gives an example of a quotient map that is neither open nor closed. We have already given examples of continuous open onto maps that are not closed and continuous closed onto maps that are not open.

COROLLARY. *If X is compact, Y is Hausdorff and $f : X \to Y$ is continuous, then the quotient space $X/\mathcal{R}_f$ is homeomorphic to the subspace $f(X)$ (and hence the quotient is Hausdorff).*

Proof. The preceding arguments yield a continuous 1–1 onto map h from the space $X/\mathcal{R}_f$, which is compact, to the space $f(X)$, which is Hausdorff. Earlier results imply that h is a closed mapping and therefore a homeomorphism.■

Important examples

Throughout this discussion D^n will denote the set of all points x in $\mathbf{R}^n$ satisfying $|x| \leq 1$ (the unit n-disk) and S^{n-1} will denote the subset of all point for which $|x| = 1$ (the unit $(n-1)$-sphere).

We start with the first example at the top of this unit; namely, the circle S^1 is homeomorphic to the quotient of $[0, 1]$ modulo the equivalence relation $\mathcal{R}$ whose equivalence classes are the one point sets $\{t\}$ for $t \in (0, 1)$ and the two point set $\{0, 1\}$. The construction of a homeomorphism is fairly typical; one constructs a continuous onto map from $[0.1]$ to S^1 for which the inverse images of points are the equivalence classes of $\mathcal{R}$. Specifically, let f be the map $[0, 1] \to S^1$ defined by $f(t) = \exp(2\pi i t)$.

Non-Hausdorff quotients. We have already mentioned that one can find equivalence relations on Hausdorff spaces for which the quotient spaces are not Hausdorff. The example we shall consider is one with only finitely many equivalence classes: Take the equivalence relation $\mathcal{A}$ on the real line $\mathbf{R}$ whose equivalence classes are all positive real, all negative reals and zero (one verbal description of this relation is that two real numbers are $\mathcal{A}$-related if and only if one is a positive real multiple of another. Then there are three equivalence classes that we shall call $+$, $-$ and 0, and the closed subsets are precisely the following:

$$\emptyset,\ \mathbf{R}/\mathcal{A},\ \{0\},\ \{+, 0\},\ \{-, 0\}$$

Since the one point subsets $\{\pm\}$ are not closed in this topology, it is not Hausdorff. Since the quotient topology is the largest topology such that the projection map is continuous, it follows that in this case there is **NO** Hausdorff topology on $\mathbf{R}/\mathcal{A}$ for which the projection

$$\pi_{\mathcal{A}} : \mathbf{R} \to \mathbf{R}/\mathcal{A}$$

is continuous.

Exercise 6 on page 145 of Munkres gives an example of a quotient map on a Hausdorff space where one point subsets in the quotient space are always closed but the quotient is not Hausdorff.

We now come to the constructions of the Möbius strip and Klein bottle. Our description of the former will be designed to reflect the usual construction by gluing together the two short ends of a rectangle whose length is much larger than its width. Let $K > 0$ be a real number, and consider the equivalence relation $\mathcal{M}_K$ on $[-K, K] \times [-1, 1]$ whose equivalence classes are the one point sets $\{\, (s, t)\,\}$ for $|s| < K$ and the two point sets $\{\ (-K, -t).\ (K, t)\ \}$ for each $t \in [-1, 1]$. Mathematically this corresponds to gluing $\{K\} \times [-1, 1]$ to $\{-K\} \times [-1, 1]$ with a twist. A Möbius strip may be viewed as the associated quotient space; note that any two models constructed above are homeomorphic (one can shrink or stretch the first coordinate; details are left to the reader). — Similarly, one can construct a Klein bottle from the cylinder $[-K, K] \times S^1$ by means of the equivalence relation whose equivalence classes are the one point sets $\{\, (s, z)\,\}$ for $|s| < K$ and the two point sets $\{\ (-K, \overline{z}).\ (K, z)\ \}$ for each $z \in S^1$. Mathematically this corresponds to gluing

$\{K\} \times S^1$ to $\{-K\} \times S^1$ by the reflection map that interchanges the upper and lower arcs with endpoints ± 1. As in the previous case, the quotient spaces obtained for different values of K are all homeomorphic to each other.

It is physically clear that one can construct the Möbius strip in $\mathbf{R}^3$, and although one cannot find a space homeomorphic to the Klein bottle in $\mathbf{R}^3$ (one needs some algebraic topology to prove this!), some thought strongly suggests that the Klein bottle should be homeomorphic to a subset of $\mathbf{R}^4$ (this has been exploited by numerous science fiction authors). For the sake of completeness (and to prove that the spaces constructed are Hausdorff) we shall prove these realization statements.

The first step is a simple geometric observation:

LEMMA. *For all positive integers p and q the products $S^p \times S^q$ and $S^p \times D^{q+1}$ are homeomorphic to subsets of $\mathbf{R}^{p+q+1}$.*

Proof. We know that $S^p \times \mathbf{R}$ is homeomorphic to the nonzero vectors in $\mathbf{R}^{p+1}$ by the map sending (x, t) to tx because

$$P(v) = \big(|v|^{-1}v, |v|\big)$$

is the inverse. Taking products with $\mathbf{R}^q$ shows that $S^p \times \mathbf{R}^{q+1}$ is homeomorphic to a subset of $\mathbf{R}^{p+q+1}$, and the lemma follows because the former clearly contains $S^p \times S^q$ and $S^p \times D^{q+1}$ ∎

The next step is to observe that one can write the spaces in question as quotients of $[a, b] \times X$ for $a < b \in \mathbf{R}$ and $X = [0, 1]$ or S^1 depending upon whether we are constructing the Möbius strip or Klein bottle; it is only necessary to consider the increasing linear homeomorphism from $[-K, K]$ to $[a, b]$ and substitute a and b for $-K$ and K in the description of the equivalence relations.

The final step is to construct continuous maps from the Möbius strip and Klein bottle to $S^1 \times D^2$ and $S^1 \times S^2$ respectively such that the inverse images of points in the codomains are merely the equivalence classes of the defining relations for the quotient spaces. Since maps from compact spaces into Hausdorff spaces are closed, this means that the images are homeomorphic to the quotient spaces of the domains. For the Möbius strip the map $f : [0, 1] \times [-1, 1] \to S^1 \times D^2$ is given by

$$f(u, v) = \big(\exp(2\pi i u), v \exp(\pi i u)\big)$$

and for the Klein bottle the map $g : [0, 1] \times S^1 \to S^1 \times S^2$ is given by

$$g(u, v) = \big(\exp(2\pi i u), A_u(J(v))\big)$$

where $J : S^1 \to S^2$ is the standard inclusion of $S^1 = S^2 \cap \big(\mathbf{R}^2 \times \{0\}\big)$ and A_u is the orthogonal rotation matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \pi u & -\sin \pi u \\ 0 & \sin \pi u & \cos \pi u \end{pmatrix}$$

for which $A_1(J(v)) = J(\overline{v})$. Verifications that these maps have all the desired properties are left to the reader as an exercise.

Final remark. A very extensive treatment of quotient topologies is given in Chapter VI of the text, *Topology*, by J. Dugundji.

V.2 : Sums and cutting and pasting

Most texts and courses on set theory and point set topology do not say much about disjoint union constructions, one reason being that everything is fairly elementary when one finally has the right definitions (two references in print are Sections I.3 and III.4–III.7 of Jänich, *Topology*, and Section 8.7 of Royden, *Real Analysis*). However, these objects arise immediately in a wide range of geometrical and topological constructions of the sort described at the beginning of this unit, including some fundamental examples from later courses in this sequence. A brief but comprehensive treatment seems worthwhile to make everything more precise and to eliminate the need to address the underlying issues in contexts that also involve more sophisticated concepts.

Disjoint union topologies

We have already defined the disjoin union (or set-theoretic sum) of two sets A and B to be the set

$$A \coprod B = (A \times \{1\}) \bigcup (B \times \{2\}) \quad \subset \quad (A \cup B) \times \{1, 2\}$$

with injection maps $i_A : A \to A \coprod B$ and $i_B : A \to A \coprod B$ given by $i_A(a) = (a, 1)$ and $i_B(b) = (b, 2)$. The images of these injections are disjoint copies of A and B, and the union of the images is $A \coprod B$.

Definition. If X and Y are topological spaces, the *disjoint union topology* or *(set-theoretic) sum topology* consists of all subsets having the form $U \coprod V$, where U is open in X and V is open in Y.

We claim that this construction defines a topology on $X \coprod Y$, and the latter is a union of disjoint homeomorphic copies of X and Y such that each of the copies is an open and closed subset. Formally, all this is expressed as follows:

ELEMENTARY PROPERTIES. *The family of subsets described above is a topology for $X \coprod Y$ such that the injection maps i_X and i_Y are homeomorphisms onto their respective images. These images are pairwise disjoint, and they are also open and closed subspaces of $X \coprod Y$. Each injection map is continuous, open and closed.*

Sketch of proof. This is all pretty elementary, but we include it because the properties are so fundamental and the details are not readily available in the standard texts.

Since X and Y are open in themselves and $\emptyset$ is open in both, it follows that $X \coprod Y$ and $\emptyset = \emptyset \coprod \emptyset$ are open in $X \coprod Y$. Given a family of subsets $\{ U_\alpha \coprod V_\alpha \}$ in the so-called disjoint union topology, then the identity

$$\bigcup_\alpha \left(U_\alpha \coprod V_\alpha \right) \quad = \quad \left(\bigcup_\alpha U_\alpha \right) \coprod \left(\bigcup_\alpha V_\alpha \right)$$

shows that the so-called disjoint union topology is indeed closed under unions, and similarly the if $U_1 \coprod V_1$ and $U_2 \coprod V_2$ belong to the so-called disjoint union topology, then the identity

$$\bigcap_{i=1,2} \left(U_i \coprod V_i \right) \quad = \quad \left(\bigcap_{i=1,2} U_i \right) \coprod \left(\bigcap_{i=1,2} V_i \right)$$

shows that the so-called disjoint union topology is also closed under finite intersections. In particular, we are justified in calling this family a topology.

By construction U is open in X if and only if $i_X(U)$ is open in $i_X(X)$, and V is open in Y if and only if $i_Y(V)$ is open in $i_Y(Y)$; these prove the assertions that i_X and i_Y are homeomorphisms onto their images. Since $i_X(X) = X \coprod \emptyset$, it follows that the image of i_X is open, and of course similar considerations apply to the image of i_Y. Also, the identity

$$i_X(X) = \left(X \coprod Y\right) \; - \; i_Y(Y)$$

shows that the image of i_X is closed, and similar considerations apply to the image of i_Y.

The continuity of i_X follows because every open set in $X \coprod Y$ has the form $U \coprod V$ where U and V are open in X and Y respectively and

$$i_X^{-1}\left(U \coprod V\right) \;\; = \;\; U$$

with similar conditions valid for i_Y. The openness of i_X follows immediately from the identity $i_X(U) = U \coprod \emptyset$ and again similar considerations apply to i_Y. Finally, to prove that i_X is closed, let $F \subset X$ be closed. Then $X - F$ is open in X and the identity

$$i_X(F) = F \coprod \emptyset = \left(X \coprod Y\right) - \left((X-F) \coprod Y\right)$$

shows that $i_X(F)$ is closed in $X \coprod Y$; once more, similar considerations apply to i_Y.■

IMMEDIATE CONSEQUENCE. *The closed subsets of $X \coprod Y$ with the disjoint union topology are the sets of the form $E \coprod F$ where E and F are closed in X and Y respectively.*■

If the topologies on X and Y are clear from the context, we shall generally assume that the $X \coprod Y$ is furnished with the disjoint union topology unless there is an explicit statement to the contrary.

Since the disjoint union topology is not covered in many texts, we shall go into more detail than usual in describing their elementary properties.

FURTHER ELEMENTARY PROPERTIES. (i) *If X and Y are discrete, then so is $X \coprod Y$.*

(ii) *If X and Y are Hausdorff, then so is $X \coprod Y$.*

(iii) *If X and Y are homeomorphic to metric spaces, then so is $X \coprod Y$.*

(iv) *If $f : X \to W$ and $g : Y \to W$ are continuous maps into some space W, then there is a unique continuous map $h : X \coprod Y \to W$ such that $h \circ i_X = f$ and $h \circ i_Y = g$.*

(v) *The spaces $X \coprod Y$ and $Y \coprod X$ are homeomorphic for all X and Y. Furthermore, if Z is a third topological space then there is an "associativity" homeomorphism*

$$\left(X \coprod Y\right) \coprod Z \cong X \coprod \left(Y \coprod Z\right)$$

(in other words, the disjoint sum construction is commutative and associative up to homeomorphism).

Sketches of proofs. (i) A space is discrete if every subset is open. Suppose that $E \subset X \coprod Y$. Then E may be written as $A \coprod B$ where $A \subset X$ and $B \subset Y$. Since X and Y are discrete it follows that A and B are open in X and Y respectively, and therefore $E = A \coprod B$ is open in $X \coprod Y$. Since E was arbitrary, this means that the disjoint union is discrete.

(ii) If one of the points p, q lies in the image of X and the other lies in the image of Y, then the images of X and Y are disjoint open subsets containing p and q respectively. On the other hand, if both lie in either X or Y, let V and W be disjoint open subsets containing the preimages of p and q in X or Y. Then the images of V and W in $X \coprod Y$ are disjoint open subsets that contain p and q respectively.

(iii) As noted in Theorem 20.1 on page 121 of Munkres, if the topologies on X and Y come from metrics, one can choose the metrics so that the distances between two points are ≤ 1. Let $\mathbf{d}_X$ and $\mathbf{d}_Y$ be metrics of this type.

Define a metric $\mathbf{d}^*$ on $X \coprod Y$ by $\mathbf{d}_X$ or $\mathbf{d}_Y$ for ordered pairs of points (p, q) such that both lie in the image of i_X or i_Y respectively, and set $\mathbf{d}^*(p, q) = 2$ if one of p, q lies in the image of i_X and the other lines in the image of i_Y. It follows immediately that $\mathbf{d}^*$ is nonnegative, is zero if and only if $p = q$ and is symmetric in p and q. All that remains to check is the Triangle Inequality:

$$\mathbf{d}^*(p, r) \leq \mathbf{d}^*(p, q) + \mathbf{d}^*(q, r)$$

The verification breaks down into cases depending upon which points lie in the image of one injection and which lie in the image of another. If all three of p, q, r lie in the image of one of the injection maps, then the Triangle Inequality for these three points is an immediate consequence of the corresponding properties for $\mathbf{d}_X$ and $\mathbf{d}_Y$. Suppose now that p and r lie in the image of one injection and q lies in the image of the other. Then we have $\mathbf{d}^*(p, r) \leq 1$ and

$$\mathbf{d}^*(p, q) + \mathbf{d}^*(q, r) = 2 + 2 = 4$$

so the Triangle Inequality holds in these cases too. Finally, if p and r lie in the images of different injections, then either p and q lie in the images of different injections or else q and r lie in the images of different injections. This means that $\mathbf{d}^*(p, r) = 2$ and $\mathbf{d}^*(p, q) + \mathbf{d}^*(q, r) \geq 2$, and consequently the Triangle Inequality holds for **all** ordered pairs (p, r).

(iv) Define $h(x, 1) = f(x)$ and $h(y, 2) = g(y)$ for all $x \in X$ and $y \in Y$. By construction $h \circ i_X = f$ and $h \circ i_Y = g$, so it remains to show that h is continuous and there is no other continuous map satisfying the functional equations. The latter is true for set theoretic reasons; the equations specify the behavior of h on the union of the images of the injections, but this image is the entire disjoint union. To see that h is continuous, let U be an open subset of X, and consider the inverse image $U^* = h^{-1}(U)$ in $X \coprod Y$. This subset has the form $U^* = V \coprod W$ for some subsets $V \subset X$ and $W \subset Y$. But by construction we have

$$V = i_X^{-1}(U^*) = i_X^{-1} \circ h^{-1}(U) = f^{-1}(U)$$

and the set on the right is open because f is continuous. Similarly,

$$W = i_Y^{-1}(U^*) = i_Y^{-1} \circ h^{-1}(U) = g^{-1}(U)$$

so that the set on the right is also open. Therefore $U^* = V \coprod W$ where V and W are open in X and Y respectively, and therefore U^* is open in $X \coprod Y$, which is exactly what we needed to prove the continuity of h.

(v) We shall merely indicate the main steps in proving these assertions and leave the details to the reader as an exercise. The homeomorphism τ from $X \coprod Y$ to $Y \coprod X$ is given by sending $(x, 1)$

to $(x, 2)$ and $(y, 2)$ to $(y, 1)$; one needs to check this map is 1–1, onto, continuous and open (in fact, if τ_{XY} is the map described above, then its inverse is τ_{YX}). The "associativity homeomorphism" sends $((x, 1), 1)$ to $(x, 1)$, $((y, 2), 1)$ to $((y, 1), 2)$, and $(z, 2)$ to $((z, 2), 2)$. Once again, one needs to check this map is 1–1, onto, continuous and open.■

COMPLEMENT. *There is an analog of Property (iv) for untopologized sets.*

Perhaps the fastest way to see this is to make the sets into topological spaces with the discrete topologies and then to apply (*i*) and (*iv*).■

Property (*iv*) is dual to a fundamental property of product spaces. Specifically, oredered pairs of maps from a fixed object A to objects B and C correspond to maps from A into $B \times C$, while oredered pairs of maps going **TO** a fixed object A and coming **FROM** objects B and C correspond to maps from $B \coprod C$ into A. For this reason one often refers to $B \coprod C$ as the *coproduct* of B and C (either as sets or as topological spaces); this is also the reason for denoting disjoint unions by the symbol $\coprod$, which is merely the product symbol $\prod$ turned upside down.

Copy, cut and paste constructions

Frequently the construction of spaces out of pieces proceeds by a series of steps where one takes two spaces, say A and B, makes disjoint copies of them, finds closed subspaces C and D that are homeomorphic by some homeomorphism h, and finally glues A and B together using this homeomorphism. For example, one can think of a rectangle as being formed from two right triangles by gluing the latter along the hypotenuse. Of course, there are also many more complicated examples of this sort.

Formally speaking, we can try to model this process by forming the disjoint union $A \coprod B$ and then factoring out by the equivalence relation

$$x \sim y \Longleftrightarrow x = y \quad \text{or}$$

$$x = i_A(a),\ y = i_B(h(a)) \quad \text{for some } a \in A \quad \text{or}$$

$$y = i_A(a),\ x = i_B(h(a)) \quad \text{for some } a \in A.$$

It is an elementary but tedious exercise in bookkeeping to to verify that this defines an equivalence relation (the details are left to the reader!). The resulting quotient space will be denoted by

$$A \bigcup_{h:C \equiv D} B.$$

As a test of how well this approach works, consider the following question:

Scissors and Paste Problem. *Suppose we are given a topological space X and closed subspaces A and B such that $X = A \cup B$. If we take $C = D = A \cap B$ and let h be the identity homeomorphism, does this construction yield the original space X?*

One would expect that the answer is yes, and here is the proof:

Retrieving the original space. Let Y be the quotient space of $A \coprod B$ with respect to the equivalence relation, and let $p : A \coprod B \to Y$ be the quotient map. By the preceding observations, there is a unique continuous map $f : A \coprod B \to X$ such that $f \circ i_A$ and $f \circ i_B$ are the inclusions

$A \subset X$ and $B \subset X$ respectively. By construction, if $u \sim v$ with respect to the equivalence relation described above, then $f(u) = f(v)$, and therefore there is a unique continuous map $h : Y \to X$ such that $f = h \circ p$. We claim that h is a homeomorphism. First of all, h is onto because the identities $h \circ p \circ i_A = \text{inclusion}_A$ and $h \circ p \circ i_B = \text{inclusion}_B$ imply that the image contains $A \cup B$, which is all of X. Next, h is 1–1. Suppose that $h(u) = h(v)$ but $u \neq v$, and write $u = p(u')$, $v = p(v')$. The preceding identities imply that h is 1-1 on both A and B, and therefore one of u', v' must lie in A and the other in B. By construction, it follows that the inclusion maps send u' and v' to the same point in X. But this means that u' and v' correspond to the same point in $A \cap B$ so that $u = p(u') = p(v') = v$. Therefore the map h is 1–1. To prove that h is a homeomorphism, it suffices to show that h takes closed subsets to closed subsets. Let F be a closed subset of Y. Then the inverse image $p^{-1}(F)$ is closed in $A \coprod B$. However, if we write write $h(F) \cap A = P$ and $h(f) \cap B = Q$, then it follows that $p^{-1}(F) = i_A(P) \cup i_B(Q)$. Thus $i_A(P) = p^{-1}(F) \cap i_A(A)$ and $i_B(Q) = p^{-1}(F) \cap i_B(B)$, and consequently the subsets $i_A(P)$ and $i_B(Q)$ are closed in $A \coprod B$. But this means that P and Q are closed in A and B respectively, so that $P \cup Q$ is closed in X. Therefore it suffices to verify that $h(F) = P \cup Q$. But if $x \in F$, then the surjectivity of p implies that $x = p(y)$ for some $y \in p^{-1}(F) = i_A(P) \cup i_B(Q)$; if $y \in i_A(P)$ then we have

$$h(x) = h(p(y)) = f(y) = f \circ i_A(y) = y$$

for some $y \in P$, while if $y \in i_B(Q)$ the same sorts of considerations show that $h(x) = y$ for some $y \in Q$. Hence $h(F)$ is contained in $P \cup Q$. On the other hand, if $y \in P$ or $y \in Q$ then the preceding equations for P and their analogs for Q show that $y = h(p(y))$ and $p(y) \in F$ for $y \in P \cup Q$, so that $P \cup Q$ is contained in $h(F)$ as required.■

One can formulate an analog of the scissors and paste problem if A and B are open rather than closed subset of X, and once again the answer is that one does retrieve the original space. The argument is similar to the closed case and is left to the reader as an exercise.

Examples. Many examples for the scissors and paste theorem can be created involving subsets of Euclidean 3-space. For example, as noted before one can view the surface of a cube as being constructed by a sequence of such operations in which one adds a solid square homeomorphic to $[0, 1]^2$ to the space constructed at the previous step. Our focus here will involve examples of objects in 4-dimensional space that can be constructed by a single scissors and paste construction involving objects in 3-dimensional space.

1. The *hypersphere* $S^3 \subset \mathbf{R}^4$ is the set of all points (x, y, z, w) whose coordinates satisfy the equation

$$x^2 + y^2 + z^2 + w^2 \quad = \quad 1$$

and it can be constructed from two 3-dimensional disks by gluing them together along the boundary spheres. An explicit homeomorphism

$$D^3 \bigcup_{\mathrm{id}(S^2)} D^3 \longrightarrow S^3$$

can be constructed using the maps

$$f_{\pm}(x, y, z) = \left(x, y, z, \sqrt{1 - x^2 - y^2 - z^2}\right)$$

on the two copies of D^3. The resulting map is well defined because the restrictions of $f_{\pm}$ to S^2 are equal.

2. We shall also show that the Klein bottle can be constructed by gluing together two Möbius strips along the simple closed curves on their edges. Let $g_\pm : [-1,1] \to S^1$ be the continuous 1–1 map sending t to $\left(\pm\sqrt{1-t^2}, t\right)$. It then follows that the images $F_\pm$ of the maps $\mathrm{id}_{[0,1]} \times [-1,1]$ satisfy $F + \cup F_- = [0,1] \times S^1$ and $F_+ \cap [0,1] \times \{-1,1\}$. If $\varphi : [0,1] \times S^1 \to \mathbf{K}$ is the quotient projection to the Klein bottle, then it is relatively elementary to verify that each of the sets $\varphi(F_\pm)$ is homeomorphic to the Möbius strip (look at the equivalence relation given by identifying two points if they have the same images under $\varphi \circ g_\pm$) and the intersection turns out to be the set $\varphi(F_+) \cap \varphi(F_-)$, which is homeomorphic to the edge curve for either of these M̈bius strips.

Disjoint unions of families of sets

As in the case of products, one can form disjoint unions of arbitrary finite collections of sets or spaces recursively using the construction for a pair of sets. However, there are also cases where one wants to form disjoint unions of infinite collections, so we shall sketch how this can be done, leaving the proofs to the reader as exercises.

Definition. If A is a set and $\{ X_\alpha \mid \alpha \in A \}$ is a family of sets indexed by A, the *disjoint union* (or set-theoretic sum)

$$\coprod_{\alpha \in A} X_\alpha$$

is the subset of all

$$(x, \alpha) \in \left(\bigcup_{\alpha \in S} X_\alpha \right) \times A$$

such that $x \in X_\alpha$.

This is a direct generalization of the preceding construction, which may be viewed as the special case where $A = \{1, 2\}$. For each $\beta \in A$ one has an injection map

$$i_\beta : X_\beta \to \coprod_{\alpha \in A} X_\alpha$$

sending x to (x, β); as before, the images of i_β and i_γ are disjoint if $\beta \neq \gamma$ and the union of the images of the maps i_α is all of $\coprod_\alpha X_\alpha$.

Notation. In the setting above, suppose that each X_α is a topological space with topology $\mathbf{T}_\alpha$. Let $\sum_\alpha \mathbf{T}_\alpha$ be the set of all disjoint unions $\coprod_\alpha U_\alpha$ where U_α is open in X_α for each α.

As in the previous discussion, this defines a topology on $\coprod_\alpha X_\alpha$, and the basic properties can be listed as follows:

[1] *The family of subsets $\sum_\alpha \mathbf{T}_\alpha$ defines a topology for $\coprod_\alpha X_\alpha$ such that the injection maps i_α are homeomorphisms onto their respective images. the latter are open and closed subspaces of $\coprod_\alpha X_\alpha$, and each injection is continuous, open and closed.*

[2] *The closed subsets of $\coprod X_\alpha$ with the disjoint union topology are the sets of the form $\coprod F_\alpha$ where F_α is closed in X_α for each α.*

[3] *If each X_α is discrete then so is $\coprod_\alpha X_\alpha$.*

[4] *If each X_α is Hausdorff then so is $\coprod_\alpha X_\alpha$.*

[5] *If each X_α is homeomorphic to a metric space, then so is $\coprod_\alpha X_\alpha$.*

[6] *If for each α we are given a continuous function $f : X_\alpha \to W$ into some fixed space W, then there is a unique continuous map $h : \coprod_\alpha X_\alpha \to W$ such that $h \circ i_\alpha = f_\alpha$ for all α.*

The verifications of these properties are direct extensions of the earlier arguments, and the details are left to the reader.■

In linear algebra one frequently encounters vector spaces that are isomorphic to direct sums of other spaces but not explicitly presented in this way, and it is important to have simple criteria for recognizing situations of this type. Similarly, in working with topological spaces one frequently encounters spaces that are homeomorphic to disjoint unions but not presented in this way, and in this context it is also convenient to have a simple criterion for recognizing such objects.

RECOGNITION PRINCIPLE. *Suppose that a space Y is a union of pairwise disjoint subspaces X_α, each of which is open and closed in Y. Then Y is homeomorphic to $\coprod_\alpha X_\alpha$.*

Proof. For each $\alpha \in A$ let $j_\alpha : X_\alpha \to Y$ be the inclusion map. By [6] above there is a unique continuous function

$$J : \coprod_\alpha X_\alpha \to Y$$

such that $J \circ i_\alpha = j_\alpha$ for all α. We claim that J is a homeomorphism; in other words, we need to show that J is 1–1 onto and open. Suppose that we have $(x_\alpha, \alpha) \in i_\alpha(X_\alpha)$ and $(z_\beta, \beta) \in i_\beta(X_\beta)$ such that $J(x_\alpha, \alpha) = J(z_\beta, \beta)$. By the definition of J this implies $i_\alpha(x_\alpha) = i_\beta(z_\beta)$. Since the images of i_α and i_β are pairwise disjoint, this means that $\alpha = \beta$. Since i_α is an inclusion map, it is 1–1, and therefore we have $x_\alpha = z_\beta$. The proof that J is onto drops out of the identities

$$J\left(\coprod_\alpha X_\alpha\right) = J\left(\bigcup_\alpha i_\alpha(X_\alpha)\right) = \bigcup_\alpha J\left(i_\alpha(X_\alpha)\right) = \bigcup_\alpha j_\alpha(X_\alpha) = Y \ .$$

Finally, to prove that J is open let W be open in the disjoint union, so that we have

$$W = \coprod_\alpha U_\alpha$$

where each U_α is open in the corresponding X_α. It then follows that $J(W) = \cup_\alpha U_\alpha$. But for each α we know that U_α is open in X_α and the latter is open in Y, so it follows that each U_α is open in Y and hence that $J(W)$ is open.■

VI. Spaces with additional properties

This unit is essentially a continuation of Unit III, and it deals with two main issues:

(1) Topological properties of spaces that are not compact but still have some important properties in common with compact spaces (open subsets of $\mathbf{R}^n$ are particularly important examples in this connection).

(2) Recognition of topological spaces that come from metric spaces. We shall concentrate on questions involving spaces constructed from reasonable pieces and merely state the general results with references to Munkres for the proofs.

In a ten week beginning graduate course it is not possible to cover everything about topological spaces that is useful in a broad range of mathematical contexts and/or for further courses in geometry and topology. Two particularly worthwhile topics of this sort are paracompact spaces (Munkres, § 41) and the compact-open topologies on spaces of continuous functions (Munkres, § 46). Another interesting topic, known as dimension theory (Munkres, § 50), deals with the following natural question: *How can one use topology to define the dimension of a topological space (as an integer* ≥ -1 *or* ∞*) such that the topological dimension of* $\mathbf{R}^n$ *is precisely* n*?* It seems reasonable to expect that $\mathbf{R}^m$ and $\mathbf{R}^n$ are not homeomorphic if $m \neq n$, and such a definition would yield this as a simple corollary. Proper mappings are another important topic that definitely would be worth discussing; they are discussed at a number of points in Munkres (where they are called *perfect* maps), and the files `proper.*` in the course directory contain further information.

An extremely comprehensive listing of properties of topological spaces, along with theorems and examples to describe the logical interrelationships between the concepts, is contained in the book, *Counterexamples in Topology*, by L. A. Steen and J. A. Seebach; the reference charts at the end are particularly helpful for obtaining a good overview of this area. Another book containing a very substantial amount of information on different properties of topological spaces is the text, *Topology*, by J. Dugundji.

In analysis one considers a large variety of topological vector spaces (each one point subset is closed, and both addition and scalar multiplication are continuous), and questions about the metrizability of these spaces (and the properties of such metrics) arise naturally. This topic is discussed specifically in Chapter 1 of Rudin, *Functional Analysis.*

VI.1 : Second countable spaces

(Munkres, § 30)

We have already noted that continuous functions from the unit interval to a Hausdorff space are completely determined by their restrictions to the rational points of the interval. In fact, this property holds for **all** subsets of Euclidean spaces. The proof of this depends upon the existence of a countable dense subset and the fact that the topology comes from a metric. There are two useful equivalent characterizations of such metric spaces, and one of the conditions implies the other two for arbitrary topological spaces.

Definitions. If X is a topological space then X is said to be

(i) *separable* if it has a countable dense subset,

(ii) *second countable* (or to satisfy the second countability axiom) if there is a countable base for the topology ($i.e.$, a countable family $\mathcal{B}$ of open sets such that every open set is a union of sets in $\mathcal{B}$,

(iii) *Lindelöf* (or to have the Lindelöf property) if every open covering has a countable subcovering.

The logical relations between these concepts are given as follows:

THEOREM. *A topological space that is second countable is also separable and Lindelöf, and a metric space that is either separable or Lindelöf is also second countable.*

In particular, all three concepts are equivalent for metric spaces. Examples exist to show that separable or Lindelöf spaces need not be second countable, and for arbitrary topological spaces there is no relation between separability and the Lindelöf property (there are examples where each is true and the other is false).

The book, *Counterexamples in Topology*, by Steen and Seebach, is a standard reference for examples of topological spaces which have one property but not another.

Implications of second countability

We shall begin by showing that second countability implies the other properties.

Separability. Let $\mathcal{B} = \{U_1, U_2, \cdots\}$ be a countable base for the topology, and form the countable subset $A \subset X$ by picking a point $a_i \in U_i$ for each i. To show that A is dense in X we need to show that every open subset of X contains a point in A. Since every open set $W \subset X$ is a union of sets in $\mathcal{B}$ there is at least one U_j that is contained in W. We then have $a_j \in U_j \subset W$. Therefore $\overline{A} = X$.

Lindelöf property. Once again let $\mathcal{B} = \{U_1, U_2, \cdots\}$ be a countable base for the topology. Given an open covering $\mathcal{U} = \{\mathcal{W}_\alpha\}$ of X, let

$$\mathcal{B}_0 = \{V_1, V_2, \cdots\}$$

be the (countable) family of all basic open sets that are contained in some element of $\mathcal{W}$. It follows that $\mathcal{B}_0$ is an open covering of X (because X is a union of the W_α and each W_α is a union of sets in $\mathcal{B}_0$). If for each j we pick $\alpha(j)$ such that $V_j \subset W_{\alpha(j)}$ it follows that

$$\mathcal{W}_O = \{W_{\alpha(1)}, W_{\alpha(2)}, \cdots\}$$

is a countable subcovering of $\mathcal{W}$.■

For the time being we shall simply note that $\mathbf{R}^n$ is an example of a separable metric space; a countable dense subset is given by the subset $\mathbf{Q}^n$ of points whose coordinates are all rational numbers. The results below will show that $\mathbf{R}^n$ and its subspaces also have the other two properties.

Second countability behaves well with respect to some standard operations on topological spaces:

PROPOSITION. *A subspace of a second countable space is second countable, and the product of two (hence finitely many) second countable spaces is second countable.*

This is stated and proved as Theorem 30.2 on page 191 of Munkres.

The reverse implications for metric spaces

We shall show that each of the other two properties implies second countability for metric spaces. The first of these implications will prove that $\mathbf{R}^n$ is a separable metric space.

Separable metric spaces are second countable. Let $A = \{a_1, a_2, \cdots\}$ be a countable dense subset and consider the countable family of open sets $W_{m,n} = N_{1/m}(a_n)$. Given an open set U in X and a point $p \in X$, let $\varepsilon > 0$ be chosen so that $N_\varepsilon(x) \subset U$. Choose m and n such that $1/2m < \varepsilon$ and $\mathbf{d}(x, a_n) < 1/2m$. If we set $W(x) = W_{m,n}$ it then follows that $x \in W(x) \subset U$, and consequently we also have $U = \cup_x W(x)$, which shows that the countable family $\mathcal{W} = \{ W_{m,n} \}$ is a base for the topology.■

COROLLARY. *The space $\mathbf{R}^n$ is a second countable space.*■

The preceding implications have the following useful consequence.

PROPOSITION. *If X is a separable metric space and $A \subset X$, then A is also separable.*■

This follows because separable metric implies second countable and the latter implies separable. One can construct examples of separable topological spaces that have nonseparable subspaces.

COROLLARY. *Every subset S of $\mathbf{R}^n$ has a countable dense subset.*■

Note that the subset S might not contain any points at all from some arbitrary countable dense subset $D \subset \mathbf{R}^n$.

Proof that Lindelöf metric spaces are second countable. For each positive integer n let $\mathcal{U}_n$ be the family of all sets $N_{1/n}(x)$ where x runs through all the points of X. Then $\mathcal{U}_n$ is an open covering of X and consequently has a countable subcovering $\mathcal{W}_n$. We claim that $\mathcal{W} = \cup_n \mathcal{W}_n$ is a base for the metric topology.

Let V be an open subset of X and let $x \in V$. Then there is some positive integer M such that $N_{1/n}(x) \subset V$ for all $n > M$. Choose an open set $N'(x)$ from $\mathcal{W}_{2n}$ such that $x \in N'$. It then follows that $N'(x) \subset N_{1/n}(x) \subset V$, and therefore we have $\cup_x N'(x) = V$, which shows that $\mathcal{W}$ is a countable base for the metric topology on X.

A compact space is automatically Lindelöf, and therefore we have the following consequence for compact metric spaces:

PROPOSITION. *A compact metric space is second countable and has a countable dense subset.*■

VI.2 : Compact spaces – II

(Munkres, §§ 26, 27, 28)

This unit contains some additional results on compact spaces that are useful in many contexts.

Sequential compactness for metric spaces

We have shown that infinite sequences in compact metric spaces always have convergent subsequences and noted that the converse is also true. Here is a proof of that converse:

THEOREM. *If X is a metric space such that every infinite sequence has a convergent subsequence, then X is compact.*

Proof. The idea is to show first that X is separable, then to use the results on second countability to show that X is Lindelöf, and finally to extract a finite subcovering from a countable subcovering.

Proof that a metric space is separable if every infinite sequence has a convergent subsequence. Let $\varepsilon > 0$ be given. We claim that there is a finite collection of points $Y(\varepsilon)$ such that the finite family

$$\{ N_\varepsilon(y) \mid y \in Y(\varepsilon) \}$$

is an open covering of X. — Suppose that no such finite set exists. Then one can recursively construct a sequence $\{ x_n \}$ in X such that

$$N_\varepsilon(x_n) \cap \Big(\bigcup_{i<n} N_\varepsilon(x_i)\Big) = \emptyset .$$

By construction we have that $\mathbf{d}(x_p, x_q) \geq \varepsilon$ if $p \neq q$, and therefore $\{ x_n \}$ has no Cauchy (hence no convergent) subsequence.

If $A = \cup_n Y(1/n)$ then A is countable and for every $\delta > 0$ and $x \in X$ there is a point of A whose distance to x is less than δ, and therefore A is dense in X.

Proof that open coverings have finite subcoverings. Since X is separable and metric, it is second countable, and therefore it also has the Lindelöf property. Therefore given an open covering $\mathcal{W}$ of X we can fine a countable subcovering $\mathcal{U} = \{U_1, U_2, \cdots\}$. We need to extract a finite subcovering from $\mathcal{U}$.

For each positive integer n let

$$E_n = X - \Big(\bigcup_{i\leq n} U_n\Big) .$$

Then each E_n is closed and $E_n \supset E_{n+1}$ for all n. If some E_n is empty then the first n sets in $\mathcal{U}$ form a finite subcovering; in fact, if some E_n is finite, then there still is a finite subcovering (take the first n sets in $\mathcal{U}$ and add one more set from $\mathcal{U}$ for each point in the intersection). Therefore the proof reduces to finding a contradiction if one assumes that each E_n is infinite.

If each E_n is infinite, then one can find a sequence of *distinct* points y_n such that $y_j \in E_j$ for each j. The assumption on X implies that the infinite sequence $\{ y_j \}$ has a convergent subsequence, say $\{ y_{k(j)} \}$. Let y^* be the limit of this subsequence. By construction, $y_m \in E_n$ if $m \geq n$, and because each E_i is closed it follows that $y^* \in E_{k(j)}$ for all j. Since the sequence of closed sets $\{ E_n \}$ is decreasing, it follows that

$$\bigcap_n E_n \quad = \quad \bigcap_j E_{k(j)}$$

and that y^* belongs to this intersection. But by construction the set $\cap_n E_n$ is empty because $\mathcal{U}$ is a countable open covering for X, and therefore we have a contradiction; it follows that $\mathcal{U}$ must have a finite subcovering, and therefore X must be compact.■

Wallace's Theorem

The following result is related to the proof that a product of two compact spaces is compact, and it turns out to be extremely useful in many contexts.

WALLACE'S THEOREM. *Let X and Y be topological spaces, let A and B be compact subsets of X and Y, and let W be an open subset of $X \times Y$ that contains A and B. Then there are open subsets U and V of X and Y respectively such that $A \subset U$, $B \subset V$ and*

$$A \times B \ \subset \ U \times V \ \subset \ W \ .$$

Proof. Given $p = (x, y) \in A \times B$ one can find open sets U_p and V_p in X and Y respectively such that $x \in U_p$, $y \in V_p$ and $U_p \times V_p \subset W$.

For a fixed $b \in B$ the open sets $U_p \times V_p$ define an open covering of the compact subset $A \times \{b\}$ and hence there is a finite subcovering associated to the family of sets

$$\mathcal{A}_b = \{U_{p(1)} \times V_{p(1)}, \ \cdots \ , U_{p(M(b))} \times V_{p(M(b))}\} \ .$$

If we take $V_b^{\#}$ to be the intersection of the sets $V_{p(j)}$ and $U_b^{\#}$ to be the union of the sets $U_{p(j)}$, it follows that

$$A \times \{b\} \ \subset \ U_b^{\#} \times V_b^{\#} \ \subset \ W$$

for each $b \in B$

The family of open subsets $\{ V_b^{\#} \}$ in Y defines an open covering of the compact subspace b, and therefore there is a finite subcovering associated to some family of sets

$$\mathcal{F} = \{U_{b(1)}^{\#} \times V_{b(1)}^{\#}, \ \cdots \ , U_{b(N)}^{\#} \times V_{b(N)}^{\#}\} \ .$$

If we now take U to be the intersection of the sets $U_{b(i)}^{\#}$ and V to be the union of the sets $V_{b(i)}^{\#}$, it follows that

$$A \times B \ \subset \ U \times V \ \subset \ W$$

as required.■

VI.3 : Separation axioms

(Munkres, §§ 31, 32, 33, 35)

In some sense the definitions for topological spaces and metric spaces present an interesting contrast. While it is clear that one does not need the full force of the properties of a metric space to prove many basic results in point set theory, it is also apparent that one needs something more than the austere structure of a topological space to go beyond a certain point. Up to this point we have introduced several conditions like the Hausdorff Separation Property which suffice for proving a number of basic results. This property is just one of a list of increasingly stronger properties that lie between a topological space with no further conditions at all and a topological space that comes from a metric space. There are many important examples of topological spaces in topology, geometry and analysis that are not homeomorphic to metric spaces (*e.g.*, many infinite-dimensional objects in algebraic topology and the so-called weak topologies on Banach spaces), and such objects are one motivation for introducing concepts somewhere between metric spaces and arbitrary topological spaces. A second motivation, which can be viewed as interesting for its own sake as well as its usefulness in certain situations, is the following:

METRIZABILITY PROBLEM. *What sorts of topological conditions are necessary or sufficient for a topological space to be homeomorphic to a metric space?*

We shall consider this problem in some detail as the final topic of the course.

The $\mathbf{T_i}$ *conditions*

The traditional way of organizing the separation properties that may or may not hold in a topological space involves a list of statements $\mathbf{T_i}$ where the subscript is some rational number and the strength of the condition increases with the index (so if $\mathbf{i} > \mathbf{j}$ then $\mathbf{T_i} \Longrightarrow \mathbf{T_j}$). We shall only deal with the statements for $\mathbf{i} = 0,\ 1,\ 2,\ 3,\ 3\frac{1}{2}$ and 4. Definitions for certain other values of i (and a great deal more) may be found online in the web sites, the paper and the reference to Munkres listed below:

`www.wikipedia.org/wiki/Separation_axiom` — This list is pretty comprehensive.

`at.yorku.ca/i/d/e/b75.htm` — This points to electronic copies of a paper, *"Definition bank" in general topology*, by G. V. Nagalagi, and one has many choices of format for downloading this paper. The listing of properties $\mathbf{T_i}$ for $\mathbf{i} < 1$ is particularly extensive.

The paper, *Espaces* $\mathbf{T}_{1\frac{1}{2}}$, by Carlos A. Infantozzi [Proceedings of the International Symposium on Topology and its Applications (Budva, 1972), pp. 116–122; Savez Društava Mat. Fiz. i Astronom., Belgrade, 1973], deals with the case $\mathbf{i} = 1\frac{1}{2}$.

Exercise 6(b) on page 213 of Munkres describes a natural candidate for $\mathbf{T_6}$.

To answer obvious questions about the choice of terminology, the symbolism $\mathbf{T_i}$ comes from the German word *Trennung*, which means separation (the terms were introduced in the classic book of Alexandroff and Hopf on topology, which was published in the nineteen thirties and written in German). In any case, here are the most important of the separation properties for this course:

Definitions. **(T0)** A topological space X is said to be a $\mathbf{T_0}$ space if for each pair of points $x,\ y \in X$ there is an open set containing one but not the other.

(T1) A topological space X is said to be a $\mathbf{T_1}$ space if for each $x \in X$ the one point set $\{x\}$ is closed in X.

(T2) A topological space X is said to be a $\mathbf{T_2}$ space if it has the Hausdorff Separation Property.

(T3) A topological space X is said to be a $\mathbf{T_3}$ space if it is a $\mathbf{T_1}$ space and is also *regular*: Given a point $x \in X$ and an open set U containing x, there is an open subset V such that

$$x \in V \subset \overline{V} \subset U$$

or equivalently that if $x \in X$ and F is a closed subset not containing X, then there are disjoint open subsets V and W such that $x \in V$ and $F \subset W$.

(T3$\frac{1}{2}$) A topological space X is said to be a $\mathbf{T_{3\frac{1}{2}}}$ space if it is a $\mathbf{T_1}$ space and is also *completely regular*: Given a point $x \in X$ and a closed subset $F \subset X$ not containing x, there is a continuous function $f : X \to [0,1]$ such that $f(x) = 0$ and $f = 1$ on F.

(T4) A topological space X is said to be a $\mathbf{T_4}$ space if it is a $\mathbf{T_1}$ space and is also *normal*: Given a closed subset $E \subset X$ and an open set U containing x, there is an open subset V such that

$$E \subset V \subset \overline{V} \subset U$$

or equivalently that if E and F are disjoint closed subsets of X, then there are disjoint open subsets V and W such that $E \subset V$ and $F \subset W$.

Most of the implications

$$\text{if } \mathbf{i} > \mathbf{j} \text{ , then } \mathbf{T_i} \Longrightarrow \mathbf{T_j}$$

are clear or have already been established, the main exceptions involving the case $\mathbf{i} = 3\frac{1}{2}$. To see that $\mathbf{T_{3\frac{1}{2}}}$ implies $\mathbf{T_3}$, given x and F let f be the continuous function and take V and W to be $f^{-1}\big([0, \frac{1}{2})\big)$ and $f^{-1}\big((\frac{1}{2}, 1]\big)$ respectively. The proof that $\mathbf{T_4}$ implies $\mathbf{T_{3\frac{1}{2}}}$ is considerably more difficult and in fact relies on the following deep result:

URYSOHN'S LEMMA. *If X is a $\mathbf{T_1}$ space, then X is $\mathbf{T_4}$ if and only if for each pair of disjoint closed subspaces $E,\ F \subset X$ there is a continuous function $f : X \to [0,1]$ such that $f = 0$ on E and $f = 1$ on F.*

The proof of the ($\Longleftarrow$) implication is similar to the proof that $\mathbf{T_{3\frac{1}{2}}}$ implies $\mathbf{T_3}$; the reader should fill in the details.

The remaining implication is an immediate consequence of this result. We shall not need the result in this generality (a simple proof for metric spaces is given below); Section 33 of Munkres (pages 207–212) give a detailed proof.■

There are two things one would like for the preceding list of separation properties. First of all, every metric space should be a $\mathbf{T_4}$ space. Second, to avoid redundancies one would like to know that if $\mathbf{i} > \mathbf{j}$ then $\mathbf{T_i}$ and $\mathbf{T_j}$ are not logically equivalent; *i.e.*, there is an example of a topological space that is a $\mathbf{T_j}$ space but not a $\mathbf{T_i}$ space. Some examples appear in Munkres, and there are many other examples of this sort in the book by Steen and Seebach. The proof that metric spaces are $\mathbf{T_4}$ turns out to be fairly straightforward.

PROPOSITION. *Every metric space is a $\mathbf{T_4}$ space.*

PROOF. We have already shown that metric spaces are $\mathbf{T_1}$ (and even $\mathbf{T_2}$). To prove the normality condition of Urysohn's Lemma involving continuous functions, consider the function

$$f(x) = \frac{\mathbf{d}(x, E)}{\mathbf{d}(x, E) + \mathbf{d}(x, F)} \ .$$

We know that the distance functions in the formula are continuous, so the formula will define a continuous real valued function if the denominator is nonzero. But since E and F are disjoint it follows that for each $x \in X$ either $x \notin E$ or $x \notin F$ is true (and maybe both are true). This means that at least one of the numbers $\mathbf{d}(x, E)$, $\mathbf{d}(x, F)$ is positive and hence their sum is always positive. Since the numerator is nonnegative and less than or equal to the denominator, it follows that $f(x) \in [0, 1]$ for all $x \in X$. If $x \in E$ then $\mathbf{d}(x, E) = 0$ and therefore $f(x) = 0$, while if $x \in F$ then $\mathbf{d}(x, F) = 0$ and therefore $f(x) = 1$.∎

Compactness and separation axioms

The following result has appeared on many examinations. When combined with Urysohn's Lemma it provides a powerful means for constructing continuous real valued functions on compact Hausdorff spaces with far-reaching consequences, particularly in functional analysis. Stronger forms of this result exist for spaces that are $\mathbf{T_2}$ and *paracompact* (a condition implied by compactness; see Section 41 on pages 252–260 of Munkres, and particularly see Theorem 41.1 on pages 253–254 for an analog of the theorem below). The result for paracompactness also has far-reaching consequences in topology and differential geometry.

THEOREM. *If a topological space is compact and* $\mathbf{T_2}$*, it is also* $\mathbf{T_4}$.

Proof. We shall give a quick proof that uses Wallace's Theorem. The more traditional proof is given by combining Lemma 26.4 on page 166 of Munkres (with the reasoning given on the previous page), which shows that compact and $\mathbf{T_2}$ implies $\mathbf{T_3}$, with Theorem 32.3 on page 202 of Munkres, which shows that compact and $\mathbf{T_3}$ implies $\mathbf{T_4}$.

Recall that a space X is Hausdorff if and only if the diagonal Δ_X is closed in $X \times X$. If A and B are disjoint subsets of a set S, it is immediate that $A \cap B = \emptyset$ is true if and only if $(A \times B) \cap \Delta_X = \emptyset$. Combining these observations, we see that if E and F are disjoint closed subspaces of a Hausdorff space X then $E \times F$ is contained in the open set $X \times X - \Delta_X$.

If X is compact, then so are E and F, and therefore by Wallace's Theorem it follows that there are open subsets $U, V \subset X$ such that

$$E \times F \subset U \times V \subset X \times X - \Delta_X \ .$$

This means that U and V are disjoint open subsets containing E and F respectively.∎

References for examples

As noted before, if $i < j$ it is not always easy to find examples of topological spaces that are $\mathbf{T_i}$ but not $\mathbf{T_j}$, and the examples in Munkres are spread out over Sections 31 through 33. Therefore we shall give an index to those examples here.

When analyzing the logical relations among separation properties it is often useful to consider the following related questions:

(1) *If A is a subspace of a $\mathbf{T_i}$ space, is A also a $\mathbf{T_i}$ space?*

(1) *If X and Y are $\mathbf{T_i}$ spaces, is their product $X \times Y$ also a $\mathbf{T_i}$ space?*

In particular, if $i < j$ and either **(1)** or **(2)** is true for $\mathbf{T_i}$ spaces but not for $\mathbf{T_j}$ spaces, then it follows that $\mathbf{T_i}$ is strictly weaker than $\mathbf{T_j}$. Specifically, the following answers to **(1)** and **(2)** show that $\mathbf{T_i}$ is strictly weaker than $\mathbf{T_4}$ if $i < 4$:

FACTS. *The answers to both* **(1)** *and* **(2)** *are positive for* $\mathbf{T_i}$ *spaces if* $i < 4$ *and negative for* $\mathbf{T_4}$ *spaces,*

Given that the proofs for $i < 4$ are fairly direct, the failures of these results to hold for $\mathbf{T_4}$ are a bit surprising at first. However, they become less surprising in light of Urysohn's Lemma, which has many far-reaching implications (compare the first paragraph of Section 32 in Munkres).

It is relatively straightforward to show that the answers to both **(1)** and **(2)** are positive for $\mathbf{T_0}$ and $\mathbf{T_1}$, and the proofs are left to the reader as exercises. It is also straightforward to produce examples to show that $\mathbf{T_0}$ does not imply $\mathbf{T_1}$ and $\mathbf{T_1}$ does not imply $\mathbf{T_2}$. In the first case, one can use the Sierpiński space whose underlying set is $\{0, 1\}$ and whose open subsets are the empty set, the set itself and $\{0\}$. In the second case one can use the finite complement topology on an infinite set. We shall now list the references to Munkres for the remaining cases.

Subspaces and products of $\mathbf{T_i}$ *spaces are* $\mathbf{T_i}$ *if* $i = 1, 2, 3, 3\frac{1}{2}$. — The first two cases are treated in Theorem 31,2 on pages 196–197 of Munkres, and the other case is treated in Theorem 33.2 on pages 211-212 of Munkres.■

Subspaces and products of $\mathbf{T_4}$ *spaces are not necessarily* $\mathbf{T_4}$. — The reference here is a combination of Theorem 32.1 and Example 2 on pages 202–204 of Munkres.■

$\mathbf{T_2}$ *does not imply* $\mathbf{T_3}$. — The reference is Example 1 on pages 197–198 of Munkres.■

$\mathbf{T_3}$ *does not imply* $\mathbf{T_{3\frac{1}{2}}}$. — The reference is Exercise 11 on pages 214 of Munkres.■

$\mathbf{T_{3\frac{1}{2}}}$ *does not imply* $\mathbf{T_4}$. — One can extract this from Example 2 on pages 203–204 either by taking a subspace of a $\mathbf{T_4}$ space that is not $\mathbf{T_4}$ or by taking a product of $\mathbf{T_4}$ spaces that is not $\mathbf{T_4}$. Since $\mathbf{T_4}$ implies $\mathbf{T_{3\frac{1}{2}}}$ and this property is preserved under taking subspaces and products, it follows that the spaces in the given example are $\mathbf{T_{3\frac{1}{2}}}$ but not $\mathbf{T_4}$.■

As we have already noted, the book by Steen and Seebach is a comprehensive summary of many further results and examples for the logical interrelationships of various special properties of topological spaces.

Non-Hausdorff topologies

As noted in Munkres and numerous other references, the addition of the Hausdorff Separation Property to the axioms for a topological space yields a class of objects that are more general than metric spaces but are relatively closed to one's geometric intuition in many ways. One additional motivation is that most of the spaces that are important to mathematicians satisfy the Hausdorff Separation Property. In some respects the examples of non-Hausdorff spaces that one sees in an introductory topology course may be viewed as instructive, showing that non-Hausdorff spaces often have very strange properties and pointing out that in some cases this property does not automatically hold, even if one makes simple constructions starting with Hausdorff spaces.

However, there are mathematical situations in which non-Hausdorff spaces arise, and in some branches of mathematics these examples turn out to be extremely important. One particularly noteworthy class of examples is given by the Zariski topologies from algebraic geometry. Here is the basic idea in the most fundamental cases: Let $\mathbf{k}$ be an algebraically closed field (every nonconstant polynomial factors completely into a product of linear polynomials, as in the complex numbers). A set $A \subset \mathbf{k}^n$ is said to be *Zariski closed* if A is the set of solutions for some finite system or polynomial equations in n variables (where the coefficients lie in $\mathbf{k}$). It is an easy exercise in algebra to show that the family of all such subsets satisfies the conditions for closed subsets of a topological space (see nearly any textbook on algebraic geometry or commutative algebra), and the resulting topological structure is called the Zariski topology. Other objects in algebraic geometry admit similar notions of Zariski topologies, but the latter quickly reach beyond the scope of this course. If $n = 1$ then the Zariski topology on $\mathbf{k}$ is the finite complement topology; since algebraically closed fields are always infinite (this follows immediately from the theory of finite fields), the Zariski topology on $\mathbf{k}$ is not Hausdorff although it is $\mathbf{T_1}$. Similarly, most other Zariski topologies are not Hausdorff (and not necessarily even $\mathbf{T_1}$). Two of the exercises for earlier sections (one on irreducible spaces, one on noetherian spaces) contain elementary results that arise naturally when one works with Zariski topologies; in fact, $\mathbf{k}^n$ with the Zariski topology is both irreducible and noetherian. The introduction of these topological structures in the nineteen forties was an elementary but far-reaching step in formulating the present day mathematical foundations for algebraic geometry. Dieudonné's book on the history of algebraic geometry provides some further information on these points.

During the past 25 to 30 years, non-Hausdorff topological spaces have also been used in certain areas of theoretical computer science. Although many of the basic ideas in such studies come from topology and branches of the "foundations of mathematics," the basic structures of the work and its goals differ substantially from those of the traditional core of mathematics, and the relevant spaces are much less geometrically intuitive than the Zariski topologies.

Two introductory references for this material are Section 3.4 of the Book, *Practical Foundations of Mathematics*, by P. Taylor, and a survey article by M. W. Mislove, *Topology, Domain Theory and Theoretical Computer Science* (Topology Atlas Preprint #181), which is available online at `http://at.yorku.ca/p/a/a/z/15.htm`$\sim$. The book *Domains and Lambda Calculi*, by R. Amadio and P.-L. Curien, presents this material specifically in connection with its applications to topics

VI.4 : Local compactness and compactifications

(Munkres, §§ 29, 37, 38)

As in the case of connectedness, it is often useful to have a variant of compactness that reflects a basic property of open subsets in Euclidean spaces: Specifically, if U is such a set and $x \in U$, then one can find an $\varepsilon > 0$ such that the closure of $N_\varepsilon(x)$ is compact; in fact if we choose $\delta > 0$ such that $N_\delta(x) \subset U$, then the compact closure property will hold for all ε such that $\varepsilon < \delta$.

Definitions. A topological space X is said to be *locally compact in the weak sense* if for each $x \in X$ there is a compact subset N such that $x \in \mathrm{Int}(N) \subset N$.

A topological space X is said to be *locally compact in the strong sense* if for each $x \in X$ and each open subset U containing x there is a compact subset N such that $x \in \mathrm{Int}(N) \subset N \subset U$.

Clearly the second condition implies the first, but in many important cases these concepts are equivalent:

PROPOSITION. *If X is a Hausdorff space, then X is locally compact in the weak sense if and only if X is locally compact in the strong sense. Furthermore, in this case if $x \in X$ and U is an open set containing x, then there is an open set W such that $x \in W \subset \overline{W} \subset U$ and $\overline{W}$ is compact.*

Proof. It is only necessary to show that the weak sense implies the strong sense and that the additional condition holds in this case. Suppose that X is locally compact in the weak sense and that $x \in U$ where U is open in X. Let N be the compact set described above, and let $V = \mathrm{Int}(N) \cap U$. Then $x \in V$ where V is open in N. Since the latter is regular (it is compact Hausdorff), there is an open subset $W \subset V$ such that

$$x \in W \subset \mathrm{Closure}(W, N) \subset V \ .$$

The set W is in fact open in X (because $W = N \cap W'$ where W' is open in X and $W = W \cap V = V \cap W'$ since $W \subset V \subset N$), Furthermore, since N is closed in X it follows that $\mathrm{Closure}(W, N) = \overline{W} \cap N$ must be equal to $\overline{W}$.

Not all subspaces of a locally compact Hausdorff space are locally compact. For example, the set of all rational numbers in the real line is not locally compact. (**Proof:** Suppose that $a \in \mathbf{Q}$ and that $B \subset \mathbf{Q}$ is an open subset in the subspace topology such that $\overline{B} \cap \mathbf{Q}$ is compact. Without loss of generality we may assume that B is an open interval centered at a. The compactness assumption on the closure implies that $\overline{B} \cap \mathbf{Q}$ is in fact a compact, hence closed and bounded, subset of the real line. This is impossible since there are many irrational numbers that are limit points of B.). However, a large number of interesting subspaces are locally compact.

PROPOSITION. *If X is a locally compact Hausdorff space in the strong sense and Y is* **either** *an open or a closed subset of X, then Y is locally compact (and Hausdorff).*

Proof. The proof for open subsets follows because if Y is open in X and U is open in Y, then U is open in X; one can then use the strong form of local compactness to prove the existence of an open subset of U with the required properties.

Suppose now that Y is closed, let $y \in Y$, and let U be an open subset of Y containing y. Write $U = Y \cap U_1$ where U_1 is open in X. Then there is an open set W_1 in X such that

$$x \ \in \ W_1 \ \subset \ \overline{W_1} \ \subset \ U_1$$

and $\overline{W_1}$ is compact. Let $W = W_1 \cap Y$; then

$$x \in W \subset \overline{W_1} \cap Y \subset U$$

where $\overline{W_1} \cap Y$ is compact because it is closed in X and contained in the compact subspace $\overline{W_1}$. Since

$$\text{Closure}(W, Y) = \overline{W} \cap Y \subset \overline{W_1} \cap Y$$

it follows that Closure(W, Y) is compact and that

$$x \in W \subset \text{Closure}(W, Y) \subset \overline{W_1} \cap Y \subset U$$

and therefore Y is locally compact in the strong sense.

COROLLARY. *If X is locally compact Hausdorff $B = U \cap F$ where $U \subset X$ is open and $F \subset X$ is closed, then B is locally compact.*

Proof. By the proposition we know that F is locally compact Hausdorff. Since $B = U \cap F$ is open in F, the proposition then implies the same conclusion for b.■

COROLLARY. *If U is open in a compact Hausdorff space, then U is locally compact.*

Proof. A compact Hausdorff space is clearly locally compact in the weak sense and hence locally compact in the strong sense. Therefore U must also be locally compact in the strong sense.

COROLLARY. *If X is a locally compact Hausdorff space, then X is $\mathbf{T_3}$.*

The defining conditions for locally compact Hausdorff in the strong sense imply that such a space is regular. In fact, one can go further using Urysohn's Lemma to prove that a locally compact Hausdorff space is in fact completely regular (hence it is $\mathbf{T_{3\frac{1}{2}}}$).

Note. A locally compact Hausdorff space is not necessarily $\mathbf{T_4}$. For example, Example 2 on pages 203–204 of Munkres actully describes an *open* subset of a compact Hausdorff space that is not $\mathbf{T_4}$, and by the proposition above this subset is locally compact Hausdorff.

Compactifications of noncompact spaces

Frequently in mathematics it is helpful to add points at infinity to a mathematical system to deal with exceptional cases. For example, when dealing with limits in single variable calculus this can be done using an extended real number system that consists of the real line together with two additional points called $\pm\infty$. In some other cases, it is preferable to add only a single point at infinity; for example, this is necessary if one wants to have something equal to

$$\lim_{x \to 0} \frac{1}{x}$$

and in the theory of functions of a complex variable it is also natural to have only one point at infinity.

In other situations it is desirable to add many different points at infinity. Projective geometry is perhaps the most basic example. In this subject one wants to add a point at infinity to each line in such a way that two lines are parallel if and only if their associated extended lines contain the same point at infinity. This turns out to be useful for many reasons; in particular, it allows one to state certain results in a uniform manner without detailed and often lengthy lists of special cases (however, you do not actually get something for nothing — one must first invest effort into the construction of points at infinity in order to obtain simplified arguments and conclusions).

In all these cases and many others, one basic property of the enriched spaces with added points at infinity is that these enriched spaces are compact and contain the original spaces as dense subspaces. In order to simplify the discussion but include the most interesting examples, we shall only consider (original and enriched) spaces that are Hausdorff.

Definition. If X is a topological space, then a *compactification* of X is a pair (Y, f) where Y is compact and $f : X \to Y$ is a continuous map that is a homeomorphism onto a dense subspace. Two compactifications (Y, f) and (Z, g) are *equivalent* if there is a homeomorphism $h : Y \to Z$ such that $h \circ f = g$.

The compactification (Y, f) is said to *dominate* the compactification (Z, g) if there is a continuous map $h : Y \to Z$ such that $h \circ f = g$, and in this case we write $(Y, f) \geq (Z, g)$.

There is also a corresponding notion of *abstract closure* in which there is no compactness assumption on Y, and one can define equivalence and domination in a parallel manner. Given an abstract closure (Y, f) of X, its *residual set* is the subset $Y - f(X)$.

Here are some basic properties of Hausdorff compactifications and abstract closures.

PROPOSITION. *(i) If (Y, f) and (Z, g) are Hausdorff compactifications or abstract closures of the same Hausdorff space X, then there is at most one $h : Y \to Z$ such that $h \circ f = g$.*

(ii) If (Y, f) and (Z, g) are Hausdorff compactifications such that $(Y, f) \geq (Z, g)$ and h is the continuous map defining the domination, then h is onto.

(iii) There is a set of equivalence classes of Hausdorff compactifications or abstract closures of a Hausdorff spaces such that every Hausdorff compactification of X is equivalent to a compactification in that set.

(iv) The relation of domination makes the equivalence classes of Hausdorff compactifications or abstract closures into a partially ordered set.

PROOF. *Proof of (i).* Let $h : Y \to Z$ be a homeomorphism such that $h \circ f = g$. If h' is another such map then the restrictions of h and h' to $f(X)$ are equal. Since $f(X)$ is dense and the set of points where two functions into a Hausdorff space are equal is a closed subset of the domain, it follows that this subset is all of Y and thus $h = h'$ everywhere.

Proof of (ii). By construction the image of h contains the dense subspace $g(X)$, and since $h(Y)$ is compact and Z is Hausdorff it also follows that $h(Y)$ is closed. Therefore we must have $h(Y) = Z$.

Note that the analogous result for abstract closures is false; by definition the identity map on a Hausdorff space X is an abstract closure in the sense of the definition, and in general there are many abstract closures (Y, f) such that $f(X) \neq Y$.

Proof of (iii). We need to introduce some set-theoretic notation. Given a set S and a nonempty family of subsets $\mathcal{M} \subset \mathbf{P}(S)$, a family $\mathcal{A} \subset \mathbf{P}(S)$ is called a *filter* provided

[*a*] if $B \in \mathcal{A}$, $B \subset C$ and $C \in \mathcal{M}$ then $C \in \mathcal{A}$,

[*b*] if $B \in \mathcal{A}$ and $C \subset \mathcal{A}$ then $B \cap C \in \mathcal{A}$.

Formally, this concept is dual to the concept of an ideal in a Boolean algebra where union and intersection are interpreted as addition and multiplication, but for our purposes the important point is that in a topological space X the set $\mathcal{N}_x$ of all open subsets containing a given point $x \in X$ is a filter.

Given a family of subsets $\mathcal{B} \subset \mathcal{M}$ that is closed under finite intersections, the smallest filter $\mathcal{B}^*$ containing $\mathcal{B}$ (or the filter generated by $\mathcal{B}$) consists of $\mathcal{B}$ together with all subsets containing some element of $\mathcal{B}$.

If X is a Hausdorff topological space then $\mathcal{F}(X)$ will denote the set of all filters of open subsets in X. For Hausdorff spaces it is immediate that $p \neq q$ implies $\mathcal{N}_p \neq \mathcal{N}_q$. Given a 1–1 continuous map g from a Hausdorff space X to another Hausdorff space Y, there is an associated map f^* from Y to $\mathcal{F}(X)$ that sends y to the filter generated by $f^{*-1}(\mathcal{N}_y)$.

We claim that f^* is 1–1 provided f is 1–1, Y is Hausdorff and $f(X)$ is dense in Y. Suppose that $u, v \in Y$. Then there are disjoint open subset $U, V \in Y$ such that $u \in U$ and $v \in V$ It suffices to show that $f^{-1}(U) \in f^*(u) - f^*(v)$ and $f^{-1}(V) \in f^*(v) - f^*(u)$. In fact, if we can prove the first, then we can obtain a proof of the second by reversing the roles of u and v and U and V throughout the argument.

By construction we have $f^{-1}(U) \in f^*(u)$ so verification of the claim reduces to showing that $f^{-1}(U) \notin f^*(v)$. Assume that we do have $f^{-1}(U) \in f^*(v)$. Then there is some open set W in Y containing v such that $f^{-1}(W) \subset f^{-1}(U)$, and it follows that we also have $f^{-1}(W \cap V) \subset f^{-1}(U)$. Since $f^{-1}(U) \cap f^{-1}(V) = f^{-1}(U \cap V) = \emptyset$ this means that $f^{-1}(W \cap V)$ must be empty. To see that this is impossible, note that $f(X) \cap W \cap V \neq \emptyset$ because $v \in W \cap V$ and $f(X)$ is dense in Y.

It follows that Y can be identified with a subset of $\mathcal{F}(X)$. By construction the latter is a subset of $\mathbf{P}(X)$ and therefore the number of points in Y is at most the cardinality of $\mathbf{P}(\,\mathbf{P}(X)\,)$. Since each topology on Y is a family of subsets of Y it follows that there is a specific bound on the cardinality of equivalence classes of Hausdorff spaces that contain a homeomorphic copy of X as a dense subspace.

Proof of (iv). Reflexivity is trivial (take the identity map on the compact space) and transitivity is also trivial (take the composite of the mappings on the compact spaces). Suppose now that $(Y, f) \geq (Z, g)$ and vice versa. Then there are continuous maps $h : Y \to Z$ and $k : Z \to Y$ such that $g = h \circ f$ and $f = k \circ g$. Therefore we have $f = k \circ h \circ f$ and $g = h \circ k \circ g$, so that $k \circ h$ and id_Y agree on the dense subset $f(X)$ and $h \circ k$ and id_Z agree on the dense subset $g(X)$. These imply that h and k are inverses to each other and hence that the two compactifications are equivalent.■

The Alexandroff one point compactification

In general there are many compactifications of a Hausdorff space. For example, there are many closed bounded subsets of the plane that contain open dense subsets homeomorphic to $\mathbf{R}^2 \cong (0,1)^2$ (these include solid rectangles with an arbitrary finite number of open holes removed; see any of the files `swisscheese.*` for more about this). Furthermore, a Hausdorff compactification does not necessarily inherit certain "good" properties of the original space; in particular, Example 3 on page 238 of Munkres, shows that a compactification of the real line is not necessarily locally connected or path connected.

In view of the examples in the previous paragraph and many others, it seems advisable to begin with simple questions about the structure of the partially ordered set of equivalence classes of Hausdorff compactifications. Two obvious questions are whether this set has maximal or minimal elements. It turns out that every $\mathbf{T}_{\mathbf{3}\frac{1}{2}}$ space has a maximal compactification (the Stone-Čech compactification) that is unique up to equivalence. This object is constructed directly in Section 38 of Munkres.

Note. Since a compact $\mathbf{T_2}$ space is $\mathbf{T_4}$ and every subspace of a $\mathbf{T}_{\mathbf{3}\frac{1}{2}}$ space is again $\mathbf{T}_{\mathbf{3}\frac{1}{2}}$, it follows that a topological space has a Hausdorff compactification if and only if it is $\mathbf{T}_{\mathbf{3}\frac{1}{2}}$.

Our main purpose here is to consider the minimal type of compactification where the added set $Y - f(X)$ consists of a single point. If $X = \mathbf{R}^n$ there is a standard and important visualization of this compactification; Y turns out to be homeomorphic to the n-dimensional unit sphere in $\mathbf{R}^{n+1}$.

If a space X has a Hausdorff compactification (Y, f) such that $Y - f(X)$ is a single point (or more generally a closed subset!) then by a corollary stated above the space X must be locally compact. This will explain our assumption in the basic construction.

Definition. Let X be a locally compact Hausdorff topological space that is not compact, and let ∞ be some point not in X (for example, in the standard axiomatic model for set theory one has $X \notin X$ so we can take $\infty = \{X\}$). The *one point* or *Alexandroff* compactification of X is the set $X^\bullet = X \cup \{\infty\}$ with open sets given as follows:

(1) *"Bounded open sets"* that are open subsets of X itself.
(2) *"Open neighborhoods of* ∞*"* that are unions $\{\infty\} \cup X - K$ where K is a compact subset of X.

It is necessary to verify that this family of sets defines a topology on X. The empty set is open in $X^\bullet$ because it is open in X, and $X^\bullet$ is open because it is equal to $\{\infty\} \cup X - \emptyset$ and $\emptyset$ is compact. Suppose that we have a family of open sets in $X^\bullet$, and split it into the bounded open sets U_α and the open neighborhoods of infinity $\{\infty\} \cup X - K_\beta$ where K_β is compact. Some elementary set-theoretic manipulation shows that the union of this family is either the bounded open set $\cap_\alpha U_\alpha$ if there are no open neighborhoods of infinity in the family or else it is

$$\{\infty\} \bigcup X - \big(K - U\big)$$

where $U = \cup_\alpha U_\alpha$ and $K = \cap_\beta K_\beta$ (note that an arbitrary intersection of compact subsets in a Hausdorff space is compact). The details of the set-theoretic algebra are described on the top of page 184 of Munkres. The latter also gives the arguments to verify that the family of open subsets defined above is closed under (finite) intersection.

CLAIM. *The inclusion of X in $X^\bullet$ is a compactification.*

Indication of proof. We need to show that the inclusion of X in $X^\bullet$ is 1–1 continuous and open (this will show it is a homeomorphism onto its image), that the image of X is dense in $X^\bullet$ and that $X^\bullet$ is compact Hausdorff. A proof that $X^\bullet$ is compact Hausdorff appears near the bottom of page 184 of Munkres.

To see that X is dense in $X^\bullet$, it suffices to verify that ∞ is a limit point of X. Let

$$\{\infty\} \bigcup (X - K)$$

(with K compact) be an open set containing the point at infinity. Since X is noncompact we must have $X - K \neq \emptyset$, and this proves the limit point assertion.

By construction the inclusion map from X to $X^\bullet$ is 1–1 and open; we need to show it is also continuous. This follows because the inverse image of a finite open set is just the set itself, and the latter is open in X by construction, while the inverse image of a open neighborhood of infinity has the form $X - K$ where K is compact; since compact subsets are closed it follows that the inverse image $X - K$ is open in X.■

The uniqueness of this one point compactification is given by the following result:

PROPOSITION. *If X is locally compact Hausdorff and (Y, f) is a Hausdorff compactification such that $Y - f(X)$ is a single point, then there is a unique homeomorphism $h : X^\bullet \to Y$ such that $h|X = f$.*

The proof of this is given as Step 1 on the bottom of page 183 in Munkres.∎

One point compactifications of Euclidean spaces

Since the spaces $\mathbf{R}^n$ are some of the most fundamental examples of locally compact spaces that are not compact, it is natural to ask if some additional insight into the nature of the one point compactification can be obtained in these cases. The following result shows that the one point compactification of $\mathbf{R}^n$ is homeomorphic to the n-dimensional sphere

$$S^n = \{ \ x \in \mathbf{R}^{n+1} \mid |x|^2 = 1 \ \} \ .$$

PROPOSITION. *Let $\mathbf{e}_{n+1} \in \mathbf{R}^{n+1}$ be the unit vector whose last coordinate is 1 (and whose other coordinates are zero). Then there is a canonical homeomorphism from the subspace $S^n - \{ \mathbf{e}_{n+1} \}$ to $\mathbf{R}^n$.*

Sketch of proof. The homeomorphism is defined by *stereographic projection*, whose physical realization is the polar projection map of the earth centered at the south pole. Mathematically this is given as follows: View $\mathbf{R}^n$ as the linear subspace spanned by the first n unit vectors, and given a point $v \in S^n$ such that $v \neq \mathbf{e}_{n+1}$ let w be the unique point of $\mathbf{R}^n$ such that $w - \mathbf{e}_{n+1}$ lies on the straight line joining $\mathbf{e}_{n+1}$ to v. The explicit formula for this map is

$$f(v) = 2\,\mathbf{e}_{n+1} \ + \ \frac{2}{1 - \langle v, \mathbf{e}_{n+1} \rangle} \cdot \big(v - \mathbf{e}_{n+1} \big)$$

and illustrations of this appear in the files `sterepic2.*` in the course directory.

The stereographic projection map f is continuous by the formula given above, and elementary considerations from Euclidean geometry show that this map defines a 1–1 correspondence between $S^n - \{ \mathbf{e}_{n+1} \}$ and $\mathbf{R}^n$. In order to give a rigorous proof that f is a homeomorphism, it suffices to verify that the map

$$g : \mathbf{R}^n \to S^n - \{ \mathbf{e}_{n+1} \}$$

defined by the formula

$$g(w) = \mathbf{e}_{n+1} = \frac{4}{|w|^2 + 4} \cdot \big(w - 2\,\mathbf{e}_{n+1} \big)$$

is an inverse to f; *i.e.*, we have $g(f(v)) = v$ and $f(g(w)) = w$ for all $v \in S^n - \{ \mathbf{e}_{n+1} \}$ and $w \in \mathbf{R}^n$. In principle the verification of these formulas is entirely elementary, but the details are tedious and therefore omitted.∎

The following important geometrical property of stereographic projection was essentially first established by Hipparchus of Rhodes (*c.* 190 B.C.E. – *c.* 120 B.C.E.):

CONFORMAL MAPPING PROPERTY. *Let $\alpha, \beta : [0,1] \to \mathbf{R}^n$ be differentiable curves with $\alpha(0) = \beta(0)$ and $\alpha'(0), \beta'(0) \neq 0$. Then the image curves $g \circ \alpha$ and $g \circ \beta$ in S^n satisfy the conditions $g \circ \alpha(0) = g \circ \beta(0)$ and $[g \circ \alpha]'(0), [g \circ \beta]'(0) \neq 0$, and*

$$\mathbf{angle}\big(\alpha'(0), \beta'(0) \big) = \mathbf{angle}\big([g \circ \alpha]'(0), [g \circ \beta]'(0) \big) \ .$$

This result will be established in the appendix on stereographic projection and inverse geometry.

VI.5 : Metrization theorems

(Munkres, §§ 39, 40, 41, 42)

As noted before, it is natural to ask for necessary and sufficient conditions that a topology on a space comes from a metric. Most point set topology texts, including Munkres, cover this material in considerable detail.

Our approach here is somewhat different; namely, we want to show that compact Hausdorff spaces built out of compact subsets of finite-dimensional Euclidean spaces are also homeomorphic to subsets of such spaces. Examples of such objects arise in many geometric and topological contexts. Our results deal with subsets of finite-dimensional Euclidean spaces and the general metrization results involve finding homeomorphic copies of a space in various infinite-dimensional spaces, so the results given here are not actually contained in the more general ones that are stated and proved in Munkres.

Since a metric is by definition a real valued continuous function on $X \times X$, it is not surprising that the proofs of metrization theorems rely heavily on constructing continuous real valued functions on a space. Therefore it is necessary to begin with results about continuous functions on compact metric spaces.

Constructions for continuous functions

The basis for all these constructions is *Urysohn's Lemma*, which is true for arbitrary $\mathbf{T_4}$ spaces and was verified earlier in the notes for metric spaces: *Given two nonempty disjoint closed subsets E and F in X, there is a continuous function $f : X \to [0,1]$ such that $f = 0$ on E and $f = 1$ on F.*

Notation. A Hausdorff space will be called a **UL**-space if Urysohn's Lemma is true. We have already noted that this is equivalent to the $\mathbf{T_4}$ condition and that every metric space is a **UL**-space.

The following basic result is in fact logically equivalent to Urysohn's Lemma:

TIETZE EXTENSION THEOREM. *Let X be a* **UL***-space, let A be a closed subset of X and let $f : A \to [0,1]$ be a continuous function. Then f extends to a continuous function from all of X to $[0,1]$.*

It is easy to prove a converse result that X is a **UL**-space if for every closed subset $A \subset X$ and every continuous function $A \to [0,1]$ there is an extension to X, for if E and F are closed subsets of X, then the function that is 0 on E and 1 on F is continuous and defined on a closed subset; the extension to X shows that the **UL**-space condition is fulfilled.

Proof. Since the closed intervals $[0,1]$ and $[-1,1]$ are homeomorphic, we may as well replace the former by the latter in the proof. The main idea of the proof is to construct a sequence of functions $\{\, \varphi_n \,\}$ such that

$$\lim_{n\to\infty} \varphi_n|A = f$$

in $\mathbf{BC}(A)$, and the basis for the recursive step in the construction is the following:

LEMMA. *Let $r > 0$ and let $h : A \to [-r,r]$ be continuous. Then there is a continuous function $g : X \to \left[-\frac{1}{3}r, \frac{1}{3}r\right]$ such that $\| \, g|A - h \, \| \leq \frac{2}{3}r$.*

Proof of Lemma. Let $B = h^{-1}\big([-r, -\frac{1}{3}r]\big)$ and $C = h^{-1}\big([\frac{1}{3}r, r]\big)$, so that B and C are disjoint closed subsets of A. Take $g : X \to \left[-\frac{1}{3}r, \frac{1}{3}r\right]$ to be a continuous function such that $g = -\frac{1}{3}r$ on B and $g = \frac{1}{3}r$ on C. The verification that $\| g|A - h \| \leq \frac{2}{3}r$ separates into three cases depending upon whether a point $a \in A$ lies in B, C or

$$h^{-1}\big([-\tfrac{1}{3}r, \tfrac{1}{3}r]\big)$$

(at least one of these must hold). In each case one can show directly that $|g(a) - h(a)| \leq \frac{2}{3}r$.■

Proof of the Tietze Extension Theorem continued. Start off with $f : A \to [-1, 1]$, and apply the lemma to get a function $g_1 : X \to [-\frac{1}{3}, \frac{1}{3}]$ with the properties stated in the lemma. Consider now the function $f_1 = f - (g_1|A)$, which is a continuous function that takes values in $[-\frac{2}{3}, \frac{2}{3}]$. Let g_2 be the continuous function on X associated to f_1 as in the lemma, define f_2 to be $f_1 - (g_2|A)$, and note that

$$\| f_2 \| \leq \left(\frac{2}{3}\right)^2.$$

We can now continue recursively to define sequences of continuous real valued functions g_n on X and $f_n = f_{n-1} - (g_n|A)$ such that

$$\| g_n \| \ \leq \ \left(\frac{1}{3}\right) \cdot \left(\frac{2}{3}\right)^{n-1}, \quad \| f_n \| \ \leq \ \left(\frac{2}{3}\right)^{n-1}.$$

Note that we have

$$f_n = f - \sum_{i=1}^{n} g_n|A$$

for all positive integers n.

We want to define a continuous function $G(x)$ by an infinite series

$$\sum_n g_n$$

and this will be possible if

$$\sum_n \| g_n \|$$

converges. But the latter sum is dominated by the convergent series

$$\frac{1}{3} \cdot \sum_n \left(\frac{2}{3}\right)^{n-1}$$

so there is no problem with constructing the continuous function G. To see that $G|A = f$, note that $G|A = \sum_n g_n|A$ and for each n we have that

$$\| f - \sum_{i=1}^{n} (g_i|A) \| \ = \ \| f_n \| \ \leq \ \left(\frac{2}{3}\right)^n$$

so that we also have $f = \sum_n \ g_n|A$. Finally, G maps X into $[-1, 1]$ because

$$\| G \| \ \leq \ \frac{1}{3} \cdot \sum_n \left(\frac{2}{3}\right)^{n-1} \ = 1 .■$$

COROLLARY. *Let X be a* **UL***-space, let A be a closed subset of X and let $f : A \to (-1, 1)$ be a continuous function. Then f extends to a continuous function from all of X to $(-1, 1)$.*

Proof. By the theorem we have a continuous function $G : X \to [-1, 1]$ such that $G = f$ on A. We need to modify this function to something that still extends f but only takes values in $(-1, 1)$.

Let $D \subset X$ be $G^{-1}(\{-1, 1\})$. By construction D and A are disjoint closed subsets, so there is a continuous function $k : X \to [0, 1]$ that is 0 on D and 1 on A. If we set F equal to the product of G and k, then it follows that F takes values in $(-1, 1)$ and $F|A = f$.∎

COROLLARY. *Let X be a* **UL***-space, let A be a closed subset of X and let $f : A \to \mathbf{R}^n$ be a continuous function. Then f extends to a continuous function from all of X to $\mathbf{R}^n$.*

Proof. If $n = 1$ this follows from the previous corollary because $(-1, 1)$ is homeomorphic to $\mathbf{R}$ (specifically, take the map $h(x) = x/(1 - |x|)$). If $n \geq 2$ let $f_1, \cdots, f_n$ be the coordinate functions of f, and let F_i be a continuous extension of f_i for each i. If F is defined by the formula

$$F(x) = \big(F_1(x), \ \cdot, F_n(x)\big)$$

then $F|A = f$.∎

Piecewise metrizable spaces

The following result provides a useful criterion for recognizing that certain compact Hausdorff spaces that are homeomorphic to subsets of Euclidean spaces.

PROPOSITION. *Let X be a compact Hausdorff space such that X is a union of closed subsets $A \cup B$, where A and B are homeomorphic to subsets of some finite-dimensional Euclidean space(s). Then X is also homeomorphic to a subset of some finite-dimensional Euclidean space.*

Proof. We may as well assume that both A and B are homeomorphic to subsets of the same Euclidean space $\mathbf{R}^n$ (take the larger of the dimensions of the spaces containing A and B respectively). Let $f : A \to \mathbf{R}^n$ and $g : B \to \mathbf{R}^n$ be 1–1 continuous mappings (hence homeomorphisms onto their images).

Let $F_0 : B \to \mathbf{R}^n$ and $G : A \to \mathbf{R}^n$ be continuous functions that extend $f|A \cap B$ and $g|A \cap B$ respectively. We can then define $F, G : X \to \mathbf{R}^n$ by piecing together f and F_0 on A and B in the first case and by piecing together G_0 and g on A and B in the second.

We shall also need two more continuous functions to construct a continuous embedding (a 1–1 continuous map that is a homeomorphism onto its image) on $X = A \cup B$. Let $\alpha : X \to \mathbf{R}$ be defined by $\mathbf{d}_B(x, A \cap B)$ on B and by 0 on A; this function is well defined and continuous because the two definitions agree on $A \cap B$. Note also that $\alpha(x) = 0$ if and only if $x \in A$. Similarly, let $\beta : X \to \mathbf{R}$ be defined by $\mathbf{d}_A(x, A \cap B)$ on A and by 0 on B; as before, this function is well defined, and furthermore $\beta(x) = 0$ if and only if $x \in B$.

Define a continuous function

$$h : X \to \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R} \times \mathbf{R} \cong \mathbf{R}^{2n+2}$$

by the formula

$$h(x) = \big(F(x), G(x), \alpha(x), \beta(x)\big) \ .$$

By construction h is continuous, and since X is compact the map h will be an embedding if and only if h is 1–1. Suppose that $h(y) = h(z)$. If $y \in A$ then we know that $\alpha(y) = 0$ and therefore we must also have $\alpha(z) = 0$ so that y and z both belong to A. Likewise, if $h(y) = h(z)$ and $y \in B$ then we must also have $z \in B$. We then also have $F(y) = F(z)$ and $G(y) = G(z)$. If $y \in A$, then the fact that z also lies in A combines with the first equation to show that $y = z$, while if $y \in B$, then the fact that z also lies in B combines with the first equation to show that $y = z$. In either case we have that $h(y) = h(z)$ implies $y = z$.■

COROLLARY. *If X is a compact Hausdorff space that is a finite union of the closed metrizable subspaces A_i that are homeomorphic to subsets of some finite-dimensional Euclidean space, then X is also homeomorphic to subsets of some finite-dimensional Euclidean space and hence metrizable.*■

In particular, this gives an alternate proof of Theorem 36.2 on pages 226–227 of Munkres (the details are left to the reader, but here is a hint — show that the space in the theorem is a finite union of subspaces homeomorphic to closed disks in $\mathbf{R}^n$).■

The next result is also useful for showing that certain compact Hausdorff spaces are homeomorphic to subsets of Euclidean spaces. Some preliminaries are needed.

Definition. If X and Y are topological spaces, and $f : X \to Y$ is continuous, then the *mapping cylinder* $\mathcal{M}_f$ of f is the quotient of $Y \amalg (X \times [0,1])$ modulo the equivalence relation generated by the condition

$$(x,1) \in X \times \{1\} \quad \sim \quad f(x) \in Y \ .$$

The equivalence classes of this relation are the one point sets $\{\,(x,t)\,\}$ for $t < 1$ and the hybrid sets

$$\{y\} \amalg f^{-1}\big(\{y\}\big) \times \{1\} \ .$$

This space is a quotient of a compact space (the disjoint union of two compact spaces) and therefore is compact.

PROPOSITION. *If X and Y are homeomorphic to subspaces of $\mathbf{R}^n$ for some n, then the mapping cylinder $\mathcal{M}_f$ is also homeomorphic to a subspace of some Euclidean space.*

Outline of proof. The details are left to the reader as an exercise, but the underlying idea is as follows. Start out with embeddings α and β of X and Y in $\mathbf{R}^n$ and construct a map

$$H : \mathcal{M}_f \to \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}^n \cong \mathbf{R}^{2n+1}$$

that is equal to $(0, 0, \beta(y)\,)$ on Y and given by

$$\big(\,(1-t)\alpha(x), 1-t, t\cdot\beta(f(x))\,\big)$$

on $X \times [0,1]$. This yields a well defined map on $\mathcal{M}_f$ because it is consistent with the equivalence relation, and the proof that h is a homeomorphism onto its image reduces to showing that h is 1–1; the latter is an elementary exercise.■

Example. In algebraic topology one often encounters the following construction called *adjoining or attaching a k-cell*: Given a space A and a continuous map $f : S^{k-1} \to A$, we define

$$B = A \cup_f e^k$$

to be the disjoint union of A and the disk D^k modulo the equivalence relation generated by $x \in S^{k-1}$ with $f(x) \in A$ for all x. Let $E \subset B$ be the set of all points that come from $A \cup \{\, x \in D^k \mid |x| \geq \frac{1}{2}\,\}$

and let F be the image of $\frac{1}{2}D^k$ in B. Then E is homeomorphic to $\mathcal{M}_f$, and then one can apply both results above to show that B is homeomorphic to a subset of some Euclidean space if A is homeomorphic to a compact subset of some Euclidean space.

A *finite cell complex* (also called a finite CW complex) is a compact Hausdorff space that is obtained from a one point space by a finite number of attachings of k_i-cells for varying values of i. We allow the case where $i = 0$ with the conventions that $S^{-1} = \emptyset$ and adjoining a 0-cell is simply the disjoint union of the original space with a one point space. The preceding results imply that each finite cell complex is homeomorphic to a subset of some Euclidean space.

An alternate approach

The metrizability proofs given above are direct and complete, and they show that the examples are in fact homeomorphic to subsets of ordinary finite dimensional Euclidean spaces. Another approach to proving the metrizability of quotients of compact metrizable spaces is by means of the *Hausdorff metric* on the family of closed subsets. The precise definitions and formal properties of the Hausdorff metric are given in Exercises 7 and 8 on pages 280–281 of Munkres. As noted in part (a) of Exercise 7, this metric makes the set of all closed subsets of a metric space X into another metric space that Munkres denotes by $\mathcal{H}$.

PROPOSITION. *Let X be a compact metric space, and let Y be a $\mathbf{T_1}$ quotient space of X. Then Y is metrizable. in the subspace topology.*

The two classes of examples above are homeomorphic to quotient spaces of compact metric spaces (presented as disjoint unions of other compact metric spaces), and in each case every equivalence class is a closed subset of the disjoint union. It follows that the quotient spaces are $\mathbf{T_1}$ in this case and therefore the proposition implies metrizability of the space constructed from the pieces (although it does not yield the embeddability in some Euclidean space if each piece is so embeddable). The need for the $\mathbf{T_1}$ condition is illustrated by a variant of an earlier example: Take $X = [-1, 1]$ and consider the equivalence relation $x \sim y$ if and only if x and y are positive real multiples of each other; as in Section 1 of Unit V, the quotient space is a non-Hausdorff space consisting of three points (in fact, this quotient space is not $\mathbf{T_1}$ because the equivalence classes of -1 and $+1$ are not closed subsets).

Proof of Proposition. If Y is $\mathbf{T_1}$ then every equivalence class in X is a closed subset, and therefore we have a map $F : X \to \mathcal{H}$ sending x to $f^{-1}(\{f(x)\})$. If we can show that F is continuous, then the metrizability of Y may be estabilished as follows: Let $\pi : X \to Y$ be the quotient projection. Since $\pi(u) = \pi(v)$ implies $F(u) = F(v)$ there is a unique continuous map $g : Y \to \mathcal{H}$ such that $F = g \circ \pi$. By construction g is a 1–1 continuous map from the compact space Y (recall it is the image of a compact space) to the metric space $\mathcal{H}$, and therefore g is a homeomorphism onto its image.

To verify that F is continuous, let $\varepsilon > 0$, let $\mathbf{D}$ denote the Hausdorff metric on $\mathcal{H}$. What does it mean to say that $\mathbf{D}(F(u), F(v)) < \varepsilon$? If one defines $U(A, \varepsilon) \subset X$ as in Munkres to be the set of all points whosed distance from a subset A is less than ε, then the contidition on the Hausdorff metric is that $F(u) \subset U(F(v), \varepsilon)$ and $F(v) \subset U(F(u), \varepsilon)$. Suppose now that $\mathbf{d}(u, v) < \varepsilon$ where as usual $\mathbf{d}$ denotes the original metric on X. Then the distance from u to $F(v)$ is less than ε and likewise the distance from v to $F(u)$ is less than ε. Therefore we have $F(u) \subset U(F(v), \varepsilon)$ and $F(v) \subset U(F(u), \varepsilon)$, and as noted above this implies $\mathbf{D}(F(u), F(v)) < \varepsilon$ so that F is uniformly continuous (in fact, Lipschitz). We can now use the argument of the first paragraph of the proof to show that Y is metrizable.∎

General metrization theorems

We begin with an early and powerful result. The proof is given in Theorem 34.1 on pages 215–218 of Munkres, with important preliminary material appearing in Theorem 20.5 on pages 125–126 and Theorem 32.1 on pages 200–201.

URYSOHN METRIZATION THEOREM. *Let X be a second countable space. Then X is homeomorphic to a metric space if and only if X is* $\mathbf{T_3}$.■

The necessary and sufficient conditions for the metrizability of arbitrary topological spaces require some additional concepts.

Definition. A family of subsets $\mathcal{A} = \{ A_\alpha \}$ in a topological space X is said to be *locally finite* if for each $x \in X$

General metrization theorems

We begin with an early and powerful result. The proof is given in Theorem 34.1 on pages 215–218 of Munkres, with important preliminary material appearing in Theorem 20.5 on pages 125–126 and Theorem 32.1 on pages 200–201.

URYSOHN METRIZATION THEOREM. *Let X be a second countable space. Then X is homeomorphic to a metric space if and only if X is* $\mathbf{T_3}$.■

The necessary and sufficient conditions for the metrizability of arbitrary topological spaces require some additional concepts.

Definition. A family of subsets $\mathcal{A} = \{ A_\alpha \}$ in a topological space X is said to be *locally finite* if for each $x \in X$ there is an open neighborhood U such that $U \cap A_\alpha \neq \emptyset$ for only finitely many A_α.

Examples. Aside from finite families, perhaps the most basic examples of locally finite families are given by the following families of subsets of $\mathbf{R}^n$.

(1) For each positive integer n let A_n be the closed annulus consisting of all points x such that $n - 1 \leq |x| \leq n$. Then for each $y \in \mathbf{R}^n$ the set $N_{1/2}(y)$ only contains points from at most two closed sets in the family (verify this!).

(2) For each positive integer n let V_n be the open annulus consisting of all points x such that $n - 2 < |x| < n + 1$. The details for this example are left to the reader as an exercise.

Locally finite families are useful in many contexts. For example, we have the following result:

PROPOSITION. *If X is a topological space and $\mathcal{A} = \{ A_\alpha \}$ is a locally finite family of closed subsets (not necessarily finite), then $\cup_\alpha A_\alpha$ is also closed.*

Proofs of this and other basic results on locally finite families of subsets appear on pages 112 and 244–245 of Munkres.■

Here is the ultimate result on metrization.

NAGATA-SMIRNOV METRIZATION THEOREM. *A topological space is metrizable if and only if it is* **T3** *and there is a base that is a countable union of locally finite families (also known as a σ-locally finite base).*

This is Theorem 40.3 on page 250 of Munkres, and the a proof including preliminary observations is contained in Section 40 on pages 248–252. Note that the Urysohn Metrization Theorem is an immediate consequence of this result because a countable base is a countable union of families such that each has exactly one element.■

A somewhat different metrization theorem due to Smirnov is in some sense the ultimate result on finding a metric on spaces built out of metrizable pieces: *A Hausdorff space is metrizable if it is paracompact and locally metrizable.* The converse to this result is also true and is due to A. H. Stone (metrizable $\Longrightarrow$ paracompact). Further information on these results appears in the files `smirnov.*` in the course directory.

Appendix A : Topological groups

(Munkres, Supplementary exercises following $ 22)

As the name suggests, a topological group is a mathematical structure that is both a topological space and a group, with some sort of compatibility between the topological and algebraic structures. Such objects lie at the point where two different areas of pure mathematics meet.

One motivation for looking at such combination structures is pure intellectual curiosity. Scientists are always interested in learning what happens if you combine **A** with **B**. Sometimes the results are not particularly useful (or do not seem so at the time), but very often these combinations can lead to important new insights into our knowledge of the original structures and to powerful new methods for analyzing questions that had previously been relatively difficult to study.

It turns out that topological groups form a rich family of interesting, relatively accessible and fundamentally important topological spaces. One reason for this is that the group structure turns out to impose some severe restrictions on the topology of the underlying space. In the other direction, the structure theory of an important special case of topological groups — namely, the compact connected Lie (pronounced "lee") groups — foreshadowed the classification of finite simple groups that was completed during the second half of the twentieth century (*Note:* A finite group is said to be **SIMPLE** if it is nonabelian and the only normal subgroups are the trivial subgroup and the group itself; the alternating groups on $n \geq 5$ letters are the most basic examples, there is a very large class of such groups that are related to Lie groups, and there is a list of 26 other groups that are called sporadic).

Another reason for the importance of topological groups is that such structures arise in a wide range of mathematical contexts. In particular, many of the most important objects studied in mathematical analysis come from topological groups (usually with some additional structure). Topological groups also arise play crucial roles in many areas of geometry, topology and algebra, and they are quite useful in application of mathematics to the sciences as well, and physics is an especially prominent example.).

Having discussed the role of topological groups, the next step is to describe them formally.

Definition. A *topological group* is a quadruple

$$(G, \mathbf{T}, m, \mathbf{inv})$$

consisting of a nonempty topological space $(G, \mathbf{T})$ and a group $(G, m, \mathbf{inv})$ such that the multiplication map $m : G \times G \to G$ and the inverse map $\mathbf{inv} : G \to G$ are continuous.

EXAMPLES. 1. The real numbers $\mathbf{R}$ form a topological group with respect to the addition operation, and the nonzero real numbers form a topological group with respect to multiplication. Similar statements hold for the complex numbers $\mathbf{C}$. In fact, one can the real and complex numbers with the addition and multiplication operations as examples of a suitably defined *topological field.*

2. The unit circle S^1 is a topological group because it is a subgroup of the multiplicative group of nonzero complex numbers.

3. If $(G, \cdots)$ and $(H, \cdots)$ are topological groups then one can make their product $G \times H$ into a topological space and a group, and it turns out that the product topology and the product group operations define a topological group structure on $G \times H$. If we specialize this to the case

$G = H = S^1$ we obtain the group structure of the 2-*dimensional torus* T^2. By induction the k-fold product of S^1 with itself, which is equivalent to $T^{k-1} \times S^1$ is also a topological group which is called the k-torus and denoted by T^k.

4. Let $\mathbf{F} = \mathbf{R}$ or $\mathbf{C}$, and let $\mathbf{GL}(n, \mathbf{F})$ be the group of invertible $n \times n$ matrices over $\mathbf{F}$. Suppose first that $\mathbf{F} = \mathbf{R}$. Then $\mathbf{GL}(n, \mathbf{R})$ is an open subset of the set of all $n \times n$ matrices (with the topology inherited from $\mathbf{R}^{n^2}$, and the coordinates of the multiplication and inverse maps are respectively polynomial and rational functions of the entries of the matrices (multiplication) or matrix (inversion). Therefore the conditions for a topological group are satisfied. Suppose now that $\mathbf{F} = \mathbf{C}$. Then $\mathbf{GL}(n, \mathbf{C})$ consists of all complex matrices whose determinants are nonzero. The natural topology for the space of $n \times n$ matrices over $\mathbf{C}$ is given by identifying the matrices with points of $\mathbf{R}^{2n^2}$, and from this viewpoint the real and imaginary parts of the complex determinant are polynomial functions of the real and imaginary parts of the matrix entries. Therefore we see that $\mathbf{GL}(n, \mathbf{C})$ corresponds to an open subset of $\mathbf{R}^{2n^2}$, and as in the real case it follows that multiplication and inversion are polynomial and rational functions so that $\mathbf{GL}(n, \mathbf{C})$ is also a topological group.

5. The subgroups of *orthogonal* and *unitary* matrices are important compact subgroups of $\mathbf{GL}(n, \mathbf{R})$ and $\mathbf{GL}(n, \mathbf{C})$ respectively. The verification that they are subgroups is carried out in linear algebra courses, Why are they compact? Recall that orthogonal and unitary matrices are characterized by the fact that their rows (equivalently, columns) form an orthonormal set. Since the rows (or columns) are all unit vectors, it follows that the groups O_n and U_n of orthogonal and unitary matrices are subsets of the compact set

$$\prod{}^{n} S^{\alpha n - 1} \quad \subset \quad \mathbf{R}^{\alpha n^2}$$

where $\alpha = 1$ for $\mathbf{R}$ and $\alpha = 2$ for $\mathbf{C}$. Since the orthonormality conditions reduce to equations involving certain polynomials in the real and imaginary parts of the matrix entries, it follows that the orthogonal and unitary groups are closed subsets of the products of spheres displayed above, and therefore these groups are compact.

6. In Exercises I.1.1 and VI.2.4 we considered a metric $\mathbf{d}_p$ on the integers for each prime p, and we noted that the completion $\widehat{\mathbf{Z}_p}$ was a compact metric space. In fact, one can extend the usual addition and multiplication maps on $\mathbf{Z}$ to continuous maps

$$\widehat{\mu} : \widehat{\mathbf{Z}_p} \times \widehat{\mathbf{Z}_p} \to \widehat{\mathbf{Z}_p}$$

$$\widehat{\alpha} : \widehat{\mathbf{Z}_p} \times \widehat{\mathbf{Z}_p} \to \widehat{\mathbf{Z}_p}$$

that make the space in question into a *compact topological commutative ring with unit.* The existence of these extensions follows directly from the results on extending uniformly continuous functions on metric spaces to the completions of the latter and the uniform continuity of ordinary addition and multiplication on $\mathbf{Z}$ with respect to the metrics $\mathbf{d}_p$ (verify this!). The systems obtained in this manner are known as the ***p*-adic integers**.

Especially in topology, whenever some type of mathematical structure is defined, one should also define the mappings or morphisms from one such object to another. For topological groups, the notion of *continuous homomorphism* is an obvious choice. Formally, these are continuous functions $\varphi : G \to H$ such that $\varphi(ab) = \varphi(a) \cdot \varphi(b)$ or alternatively satisfy the morphism identity

$$m_H \circ (\varphi \times \varphi) = \varphi \circ m_G \; : G \times G \to H$$

where m_G and m_H are the multiplication maps for G and H respectively.

One particularly important continuous homomorphism is the *exponential map* from the additive topological group of real or complex numbers to the multiplicative topological group of nonzero real or complex numbers. Another example on $\mathbf{GL}(n, \mathbf{F})$ is the map sending an invertible matrix to its transposed inverse. Over the complex numbers one also has the conjugation map on both the additive and nonzero multiplicative groups of complex numbers and the conjugate of the transposed inverse for invertible complex matrices.

All the nonexponential examples in the preceding paragraph are in fact *topological automorphisms* of the groups in question; *i.e.*, they are continuous homomorphisms that have continuous inverses. In fact, for all these examples the map is equal to its own inverse.

Properties of topological groups

We had previously stated that the group structure on a topological group (and especially its continuity!) implies strong restrictions on the topology of the underlying space. We shall discuss the most basic properties here.

HOMOGENEITY OF TOPOLOGICAL GROUPS

A topological space X is said to be *homogeneous* if for each pair of points $u\, v \in X$ there is a homeomorphism $h : X \to X$ such that $h(u) = v$. It is necessary to be somewhat careful when using this term, because the expression "homogeneous space" has a special meaning that is described below. There are many examples of spaces that are homogeneous, and there are many spaces that do not satisfy this condition.

EXAMPLES. 1. Every normed vector space V is homogeneous. Given two vectors $u,\, v \in V$ the translation map $T(x) = x + (v - u)$ is an isometry and sends u to v.

2. The closed unit interval $[0, 1]$ is not homogeneous; more precisely, it is impossible to construct a homeomorphism taking an end point to a point in the open interval $(0, 1)$. If such a homeomorphism h existed then for all open sets U containing 1 there would be homeomorphisms between $U - \{1\}$ and $h(U) - \{h(1)\}$. On the other hand, sets of the latter type are disconnected if $h(1) \in (0, 1)$ while $U - \{1\}$ is the connected set $[0, 1)$ if $U = [0, 1]$.

3. Since the open interval $(0, 1)$ is homeomorphic to $\mathbf{R}$, it follows from the first example that $(0, 1)$ is homogeneous. In fact, *if U is an open connected subset of $\mathbf{R}^n$, then U is homogeneous.* The proof is relatively elementary, but since the argument is a bit lengthy it will be given separately in Appendix C.

The homogeneity of topological groups is essentially a generalization of the first example given above:

PROPOSITION. *If G is a topological group, let $L_a : G \to G$ be the continuous map $L_a(g) = a \cdot g$ ("left multiplication"), and let $R_a : G \to G$ be the continuous map $R_a(g) = g \cdot a$ ("left multiplication"). Then L_a and R_a are homeomorphisms.*

Sketch of proof. Let $b = a^{-1}$. Then the associativity and continuity of multiplication imply that L_b and R_b are inverses to L_a and R_a respectively.■

COROLLARY. *A topological group is homogeneous.*

Proof. Given distinct points $a\,, b \in G$ let $c = b\, a^{-1}$ and note that $L_c(a) = b$.■

SEPARATION PROPERTIES OF TOPOLOGICAL GROUPS

All of the examples of topological groups that we have explicitly described are Hausdorff. It turns out that weaker separation properties imply the Hausdorff Separation Property, and the Hausdorff Separation Property itself implies even stronger separation properties. We shall begin by proving the first of these statements.

PROPOSITION. *If a topological group is a $\mathbf{T_0}$ space, then it is a Hausdorff (or $\mathbf{T_2}$) space.*

Proof. The proof splits into two parts, one showing the implication $\mathbf{T_0} \implies \mathbf{T_1}$ and the other showing the implication $\mathbf{T_1} \implies \mathbf{T_2}$.

($\mathbf{T_0} \implies \mathbf{T_1}$). It suffices to prove that the one point set $\{1\}$ consisting only of the identity element is closed in G; for every other element g, one can apply the homeomorphism L_g to see that $\{g\} = L_g(\{1\})$ is also closed in G. Therefore we need only show that $G - \{1\}$ is open. If x belongs to the latter, then the $\mathbf{T_0}$ condition implies that there is n open subset W that contains either 1 or x but not both. If $x \in W_x$ let $U_x = W_x$. Suppose now that $1 \in W_x$ but $x \notin W_x$; if we can find a homeomorphism $h : G \to G$ such that h switches 1 and x, then $h(W_x)$ is an open set containing x but not 1, and if we set $U_x = h(W_x)$ this will express $G - \{1\}$ as a union of open sets and thus prove that $G - \{1\}$ is open. The desired homeomorphism is merely the composite $L_x \circ \mathbf{inv}$.

($\mathbf{T_1} \implies \mathbf{T_2}$). It will suffice to show that the diagonal is closed in $G \times G$. Let $D : G \times G \to G$ be the composite $m \circ (\mathbf{id}_G \times \mathbf{inv})$. As noted in Exercise 1 on page 145 of Munkres, this map is continuous. But $\Delta_G = D^{-1}(\{1\})$ and since $\{1\}$ is closed in G by assumption, it follows that Δ_G is closed in $G \times G$ so that G is Hausdorff, which is the same as saying that G is $\mathbf{T_2}$.■

If a topological group is Hausdorff, then one can prove that it satisfies even stronger separation properties. In particular, Exercise 7 on page 146 of Munkres shows that a Hausdorff (equivalently, a $\mathbf{T_0}$) topological group is $\mathbf{T_3}$, and Exercise 10 on pages 213–214 of Munkres shows that such a topological group is also $\mathbf{T_{3\frac{1}{2}}}$.

CLOSED SUBGROUPS OF TOPOLOGICAL GROUPS

In the theory of topological groups it is important to know whether or not a subgroup is closed, and if a subgroup is not closed it is important to have the following additional information about its closure:

PROPOSITION. *If G is topological group and H is a subgroup, then its closure $\overline{H}$ is also a subgroup. Furthermore, if H is normal then so is $\overline{H}$.*

Before proceeding further, here are some examples when $G = \mathbf{R}$ (with addition as the group operation). The rationals are clearly not a closed subgroup (since they are dense in the real line), but the integers are, and one way of seeing this is to note that the complement of integers is equal to the union of the open intervals $(n, n+1)$ where n ranges over all the elements of $\mathbf{Z}$. Over the complex numbers, the additive groups of real an purely imaginary numbers are closed subgroups (why?) and in the groups $\mathbf{GL}(n, \mathbf{F})$ the orthogonal and unitary groups are closed subgroups because the latter are compact.

Proof of proposition. If H is a subgroup then we know that $m(H \times H) = H$ and $\mathbf{inv}(H) = H$. Recall that a function f is continuous if and only if it satisfies the condition

$$f(\overline{A}) \quad \subset \quad \overline{f(A)} \ .$$

If we apply this to multiplication and the inverse map we find that

$$m\left(\overline{H} \times \overline{H}\right) \;=\; m\left(\overline{H \times H}\right) \;\subset\; \overline{m(H \times H)} \;=\; \overline{H}$$

and similarly

$$\mathbf{inv}\left(\overline{H}\right) \;\subset\; \overline{H}$$

so that $\overline{H}$ is a subgroup of G.

Suppose now that H is normal in G; in other words, for all $a \in G$ the inner automorphism $I_a(x) = axa^{-1}$ maps H to itself. This mapping is continuous because it is the composite $L_a \circ R_{a^{-1}}$; by associativity it can also be written as $R_{a^{-1}} \circ L_a$. As in the previous paragraph the continuity of this map and the hypothesis $I_a(H) \subset H$ imply that $I_a\left(\overline{H}\right) \subset \overline{H}$.∎

In group theory, if one has a group G and a subgroup H, then it is possible to form the set G/H of cosets, and it H is a normal subgroup then one can construct a quotient group structure on G/H. For topological groups one can carry out the same constructions and topologize G/H using the quotient topology; this quotient is $\mathbf{T_1}$ if and only if H is a closed subgroup. Some additional information on this can be found in Exercises 5 and 6 on page 146 of Munkres (see also Exercise 7(d) on that page). We shall also note that the standard isomorphism theorems known from ordinary group theory have topologized analogs for topological groups.

Example. Although closed subgroups play an extremely important role in the theory of topological groups, the non-closed subgroups are too important to be ignored, so we shall give one more example involving T^2: Given an irrational real number $a > 0$, consider the continuous homomorphism

$$\varphi : \mathbf{R} \longrightarrow T^2$$

defined by

$$\varphi(t) \;=\; (\,\exp(2\pi i t),\ \exp(2\pi i a t)\,)$$

and let H be the image of φ. To see that H is not closed, consider its intersection with $\{1\} \times S^1$, and call this intersection L. By definition it follows that L consists of all elements of the form $(1, \exp(2\pi i a n)$ for some $n \in \mathbf{Z}$. If H is a closed subgroup of T^2 then so is L. By the previous description L is countable; we claim it is also infinite; if not, then there would be two integers $p \neq q$ such that $\exp(2\pi i a p) = \exp(2\pi i a q)$, and the latter would imply that $a(p-q)$ is an integer, contradicting the irrationality of a. Since T^2 is a compact metric space and L is a closed subset, it follows that L is complete with respect to the subspace metric. Therefore a corollary of Baire's Theorem implies that L has an isolated point. But L is a topological group and therefore every point is isolated by homogeneity. This and compactness force L to be finite, contradicting our previous conclusion that L was infinite. The contradiction arose from the assumption that the subgroup H was closed, and therefore we see that H cannot be a closed subgroup.∎

CONNECTED SUBGROUPS OF TOPOLOGICAL GROUPS

It is also important to know whether a subgroup of a topological group is open, but the reasons for this are fundamentally different, and the following result indicates why this is the case:

PROPOSITION. *An open subgroup of a topological group is also closed.*

In contrast, a closed subgroup of a topological group is not necessarily open, and the simplest example is probably the inclusion of $\mathbf{R} \times \{0\}$ in $\mathbf{R}^2$.

Proof. Let H be an open subgroup of G. Then there exist elements $g_\alpha \in G$ such that the cosets Hg_α are pairwise disjoint and their union is G; we may as well suppose that $g_\beta = 1$ if g_β is the unique element such that $H = Hg_\beta$. Since the right translation maps R_{g_α} are homeomorphisms it follows that each coset is open. Therefore

$$H \quad = \quad G \; - \; \bigcup_{g_\alpha \neq 1} Hg_\alpha$$

expresses H as the complement of a union of open subsets, and it follows that H is closed.∎

Here are a few consequences of this relatively elementary but far-reaching observation.

PROPOSITION. (i) *The smallest subgroup which contains a fixed open subset of a topological group is both open and closed.*

(ii) *The connected component of the identity in a topological group is a normal closed subgroup.*

(iii) *If a topological group is connected, then it is generated by every open neighborhood of the identity.*

Proof. (i) Let W be an open neighborhood of the identity; by taking the union of W with its inverse we may assume that $\mathbf{inv}(W) = W$. Consider the sequence of sets W_n defined inductively so that $W_1 = W$ and $W_{n+1} = W_n \cdot W$, where the raised dot denotes group multiplication. By construction W_n is mapped to itself by the inverse mapping; we claim that it is also open. To see this, note that if U is an open subset of G and A is an arbitrary set then the identity $A \cdot U = \cup_a \, L_a(U)$ shows that $A \cdot U$ is open. Consider the set $W_\infty = \cup_n \, W_n$. If H is an arbitrary subgroup of G containing W, then H certainly contains W_∞. We claim that W_∞ itself is a subgroup; at this point we only need to verify that W_∞ is closed under multiplication, but this is an elementary exercise (work out the details!). Thus W_∞ is the smallest subgroup containing W, and it is an open subset. By our previous result it is also a closed subset.

Proof of (ii). As usual, write $a \sim b$ if a and b lie in the same connected component of G. Since left multiplication is continuous we know that $a \sim b$ implies $ga \sim gb$ for all $g \in G$. Therefore if $a \sim 1$ and $b \sim 1$ we obtain the relation $ab \sim a$, so that $ab \sim 1$ by transitivity. Therefore the component of the identity is closed under multiplication. Similarly, if $a \sim 1$ then left multiplication by a^{-1} yields $1 \sim a^{-1}$ and hence that the component of the identity is also closed under taking inverses. Thus this component G° is a subgroup, and it is closed because connected components are closed.

Finally to see that G° is normal, let $x \in G$ and let I_x denote conjugation by x. Since I_x is a homeomorphism and takes 1 to itself, it follows that I_x must also take the connected component of 1, which is G°, to itself, and therefore the latter must be a normal subgroup.

Proof of (iii). If U is a neighborhood of the identity we have seen that the smallest subgroup H containing U is open and closed. Since $1 \in H$, the set is also nonempty, and therefore by connectedness we must have $H = G$.∎

Note that the additive groups of p-adic integers and the rationals are examples for which the component of the identity is not an open subgroup.

At this point we shall only state and prove one more result which illustrates the strong properties of topological groups.

PROPOSITION. *If G is a connected topological group and D is a discrete normal subgroup of G, then D is central.*

The normality is crucial in this result. The group $\mathbf{GL}(n, \mathbf{C})$ is connected if $n \geq 2$ (see the exercises for hints on proving this) but it contains many discrete non-central subgroups. One simple example is the subgroup with two elements given by the identity and the diagonal matrix with a -1 in the upper left corner and $+1$ in every other diagonal entry.

Proof. Let $J : G \times D \to D$ be the continuous map sending (g, x) to gxg^{-1}; the image lies in D because the latter is a normal subgroup. What is the image of $G \times \{x\}$? We know it is connected and that it contains x. Since $\{x\}$ is the connected component of x in D because the latter is discrete, it follows that J maps $G \times \{x\}$ onto x. The latter means that $gxg^{-1} = x$ for all $g \in G$, which can be rewritten in the form $gx = xg$ for all $g \in G$. Therefore x lies in the center of G. Since x was arbitrary it follows that all of D lies in the center.∎

Analysis on topological groups

This subsection discusses how some important concepts from real variables (uniform continuity and a good theory of integration) can be developed for topological groups. A reader who would prefer to continue with the topological and geometrical discussion may skip this section without missing any topological or geometrical material.

As noted in Section I.3 of the notes, topological groups have a uniform structure that allows one to formulate a useful notion of uniform continuity. In principle, the idea is to consider a family of symmetric neighborhoods of the identity (*i.e.*, neighborhoods mapped to themselves under inversion) and to formulate a uniform concept of closeness using these neighborhoods. For example, if we are given two functions f and g from a set X to the topological group G and a symmetric open neighborhood V of the identity, one can say that f and g are uniformly within V of each other if $f(x) \cdot g(x)^{-1} \in V$ for all $x \in X$

In fact, every topological group can be viewed as a uniform space in *two* ways; the left uniformity turns all left multiplications into uniformly continuous maps while the right uniformity turns all right multiplications into uniformly continuous maps. If G is not abelian, then these two structures need not coincide. These uniform structures allow to talk about notions such as completeness, uniform continuity and uniform convergence on topological groups.

INTEGRATION ON LOCALLY COMPACT TOPOLOGICAL GROUPS

In mathematical analysis, the locally compact groups are of particular importance because they admit a natural notion of measure and integral that was introduced by Alfréd Haar in the nineteen thirties. The idea is to assign a "translation invariant volume" to subsets of a locally compact Hausdorff topological group (*e.g.* if one takes a reasonable subset S and considers its left translate $L_g(S)$ for some $g \in G$, then the volumes of the two sets are equal) and subsequently to define an integral for functions on those groups.

If G is a locally compact Hausdorff topological group, one considers the algebra $\mathcal{M}$ of subsets generated by all compact subsets of G and including all countable unions, countable intersections and complements. If g is an element of G and S is a set in $\mathcal{M}$, then the set $L_a(S)$ is also in $\mathcal{M}$. It turns out that there is, up to a positive multiplicative constant, only one left-translation-invariant measure on $\mathcal{M}$ which is finite on all compact sets, and this is the Haar measure on G (*Note:* There is also an essentially unique right-translation-invariant measure on $\mathcal{M}$, but the two measures need not coincide; the difference between these two measures is completely understood). Using the general Lebesgue integration approach, one can then define an integral for all measurable functions $f : G \to \mathbf{R}$ (or $\mathbf{C}$).

The Internet site

`http://www.wikipedia.org/wiki/Haar_measure`

contains further information about integration on locally compact groups and its uses in mathematics (and physics!).

Remarks on Lie groups

We have already mentioned the particular importance of Lie groups in mathematics. It is beyond the scope of these notes to describe the objects fully (this is material for the third course in the geometry/topology sequence), but we have developed enough ideas to discuss a few important points.

We begin with a discussion of the exponential map for $n \times n$ matrices over the real numbers. Our results on norms for finite-dimensional real vector spaces imply that $\mathbf{R}^k$ is complete with respect to every norm. Take $k = n^2$ and view $\mathbf{R}^{n^2}$ as the space of all $n \times n$ matrices over $\mathbf{R}$. The exponential mapping

$$\exp : \mathbf{M}(n; \mathbf{R}) \longrightarrow \mathbf{GL}(n, \mathbf{R})$$

is the map defined by the familiar infinite series

$$\exp(A) = \sum_{n \geq 0} \frac{1}{n!} A^n$$

and this series converges absolutely because

$$\sum_{n \geq 0} \frac{1}{n!} \parallel A^n \parallel \ \leq \ \sum_{n \geq 0} \frac{1}{n!} \parallel A \parallel^n$$

and the right hand side converges to $\exp(\parallel A \parallel)$.

THEOREM. *The exponential map defines a homeomorphism from an open neighborhood of $0 \in \mathbf{M}(n; \mathbf{R})$ to an open neighborhood of $I \in \mathbf{GL}(n, \mathbf{R})$.*

Proof. The key to this is the Inverse Function Theorem. By construction we know that

$$\exp(A) \ - \ \exp(0) \ = \ \exp(A) \ - \ I \ = \ A \ + \ A \cdot \theta(A)$$

where $\lim_{\parallel A \parallel \to 0} \theta(A) = 0$. Therefore it follows that $D\exp(0) = I$. If exp is a $\mathbf{C}^1$ mapping then we can apply the Inverse Function Theorem to complete the proof.

One direct way to verify that exp is $\mathbf{C}^1$ is to use power series. If A is an $n \times n$ matrix and B is another matrix of the same size that is close to zero, then we may write

$$\exp(A + B) - \exp(A) = \sum_{n > 0} \frac{1}{n!} \ [(A + B)^n - A^n]$$

where each of the terms in brackets is a sum of degree n monomials in the noncommuting matrices A and B, and each monomial in the summand has at least one factor equal to B. The n^{th} bracketed term can be written as a sum $f_n(A, B) + g_n(A, B)$ where each term in $f_n(A, B)$ has exactly one B

factor and each term in $g_n(A, B)$ has at least two B factors. The first of these may be written in the form

$$f_n(A, B) \quad = \quad \sum_{r=1}^{n} A^r \, B \, A^{n-r-1}$$

and one can check that the series

$$\Phi(A, B) \quad = \quad \sum_{n>1} \frac{1}{n!} \, f_n(A, B)$$

converges absolutely because $\| \; f_n(A, B) \; \| \leq \; n \, \|A\|^{n-1} \cdot \|B\|$. Furthermore, it follows that Φ is continuous in A (by uniform convergence of the series) and is linear in B. Therefore we will have that $D \exp(A)(B) = \Phi(A, B)$ and exp is $\mathbf{C}^1$ if and only if for $B \neq 0$ we can show that

$$\lim_{B \to 0} \frac{1}{\|B\|} \sum_{n>0} \frac{1}{n!} \, g_n(A, B) \quad = \quad 0$$

An upper estimate for the norm of the right hand side is given by

$$\frac{1}{\|B\|} \sum_{n>0} \frac{1}{n!} \quad \| \; g_n(A, B) \; \|$$

where each term $\|g_n(A, B)\|$ is bounded from above by

$$\sum_{r=2}^{n} \binom{n}{r} \|B\|^r \, \|A\|^{n-r} \; .$$

The standard considerations then show that the series

$$\Theta(A, B) = \frac{1}{\|B\|} \sum_{n>0} \frac{1}{n!} \, g_n(A, B)$$

converges absolutely for all A and B and that

$$\lim_{B \to 0} \Theta(A, B) = 0$$

if A is held fixed. Thus we have shown directly that the matrix exponential is a $\mathbf{C}^1$ function.■

This result yields an important structural property.

PROPOSITION. *There is a neighborhood U of the identity in $\mathbf{GL}(n, \mathbf{R})$ such that the only subgroup contained in U is the trivial subgroup.*

The conclusion is often abbreviated to say that $\mathbf{GL}(n, \mathbf{R})$ is a *topological group with no small subgroups* (NSS).

Proof. Choose $\delta > 0$ so that exp maps the disk of radius $2\,\delta$ about the origin homeomorphically to an neighborhood W of the identity in $\mathbf{GL}(n, \mathbf{R})$. Suppose that U is the image of the neighborhood of radius δ in $\mathbf{GL}(n, \mathbf{R})$, and let H be a subgroup that lies entirely in U. Suppose further that H is nontrivial, and choose a matrix A so that $\|A\| < \delta$ and $1 \neq \exp(A) \in U$.

Consider the $\mathbf{C}^\infty$ map $\varphi_A : \mathbf{R} \longrightarrow \mathbf{GL}(n, \mathbf{R})$ sending t to $\exp(tA)$. Since $\exp(P+Q) = \exp(P)\exp(Q)$ if P and Q commute, it follows that φ_A is a continuous homomorphism. The sequence $\{ \|nA\| = n\|A\| \}$ is unbounded and therefore there is a first n such that $\|nA\| = n\|A\| > \delta$. Since $\|A\| < \delta$ it follows that $n\|A\| = \|nA\| < 2\,\delta$. But this means that

$$\exp(nA) \quad = \quad \exp(A)^n \quad \in \quad W - U \ ;$$

on the other hand, since $\exp(A)$ was supposed to belong to a subgroup $H \subset U$, we also have that $\exp(A)^n = \exp(nA) \in U$. The contradiction arises from the assumption that there was a nontrivial subgroup H contained in U, and therefore no such subgroup exists.■

COROLLARY. *If H is a subgroup of $\mathbf{GL}(n, \mathbf{R})$, then H has no small subgroups.*■

Results of A. Gleason and (jointly and independently) D. Montgomery and L. Zippin provide *a completely topological characterization of Lie groups as locally compact Hausdorff groups with no small subgroups.* These results are proved in the book Montgomery and Zippin as well as Kaplansky's *Lie Algebras and Locally Compact Groups.*

Note that a locally compact Hausdorff topological group that is locally connected and second countable does not necessarily satisfy the "no small subgroups" condition. A countably infinite product of copies of S^1 is a relatively simple counterexample.

Appendix B : Stereographic projection and inverse geometry

The conformal property of stereographic projections can be established fairly efficiently using the concepts and methods of inverse geometry. This topic is relatively elementary, and it has important connections to complex variables and hyperbolic (= Bolyai-Lobachevsky noneuclidean) geometry.

Definition. Let $r > 0$ be a real number, let $y \in \mathbf{R}^n$, and let $S(r;y)$ be the set of all points $x \in \mathbf{R}^n$ such that $|x - y| = r$. The *inversion map with respect to the sphere* $S(r;y)$ is the map T on $\mathbf{R}^n - \{0\}$ defined by the formula

$$T(x) = y \ + \ \frac{r^2}{|x-y|} \cdot \big(x - y\big) \ .$$

Altternatively, $T(x)$ is defined so that $T(x) - y$ is the unique positive scalar multiple of $x - y$ such that

$$\big|T(x) - y\big) \cdot \big|x - y\big) = r^2 \ .$$

Another way of saying this is that inversion interchanges the exterior points to $S(r;y)$ and the interior points with the center deleted. If $r = 1$ and $y = 0$ then inversion simply takes a nonzero vector x and sends it to the nonzero vector pointing in the same direction with length equal to the reciprocal of $|x|$ (this should explain the term "inversion").

If $n = 2$ then inversion corresponds to the *conjugate* of a complex analytic function. Specifically, if a is the center of the circle and r is the radius, then inversion is given in complex numbers by the formula

$$T(z) \quad = \quad r^2 \cdot \big(\overline{z} - a\big)^{-1} \quad = \quad r^2 \cdot \overline{\big(z - \overline{a}\big)^{-1}}$$

where the last equation holds by the basic properties of complex conjugation. If $r = 1$ and $a = 0$ then inversion is just the conjugate of the analytic map sending z to z^{-1}.

The geometric properties of the analytic inverse map on the complex plane are frequently discussed in complex variables textbooks. In particular, this map has nonzero derivative wherever the function is defined, and accordingly the map is conformal. Furthermore, the map $z \rightarrow z^{-1}$ is an involution (its composite with itself is the identity), and it sends circles not containing 0 to circles of the same type. In addition, it sends lines not containing the origin into circles containing the origin and vice versa (*Note:* This means that 0 lies on the circle itself and NOT that 0 is the center of the circle!). Since complex conjugation sends lines and circles to lines and circles, preserves the angles at which curves intersect and also sends 0 to itself, it follows that the inversion map with respect to the unit circle centered at 0 also has all these properties).

It turns out that all inversion maps have similar properties. In particular, they send the every point of sphere $S(r;y)$ to itself and interchange the exterior points of that sphere with all of the interior points except y (where the inversion map is not defined), they preserve the angles at which curves intersect, they are involutions, they send hyperspheres not containing the central point y to circles of the same type, and they send hyperplanes not containing $\{y\}$ into hyperspheres containing $\{y\}$ and vice versa. We shall limit our proofs to the properties that we need to study stereographic projections; the reader is encouraged to work out the proofs of the other assertions.

We begin by recalling(?) some simple observations involving isometries and similarity transformations from $\mathbf{R}^n$ to itself. Proofs or hints for proofs in the cases of isometries can be found in many standard linear algebra texts.

FACT 1. *If $b \in \mathbf{R}^n$ and F is translation by b (formally, $F(x) = x + b$), then F is an isometry from $\mathbf{R}^n$ to itself and F sends the straight line curve from x to y defined by*

$$\alpha(t) = ty + (1-t)x$$

to the straight line curve from $F(x)$ to $F(y)$ defined by

$$\alpha(t) = tF(y) + (1-t)F(x)\ .$$

Definition. If $r > 0$ then a *similarity transformation with ratio of similitude* r on a metric space X is a 1–1 correspondence f from X to itself such that $\mathbf{d}\big(f(x), f(y)\big) = r \cdot \mathbf{d}(x, y)$ for all $x,\ y \in X$.

Note that every isometry (including every identity map) is a similarity transformation with ration of similitude 1 and conversely, the inverse of a similarity transformation with ratio of similitude r is a similarity transformation with ratio of similitude r^{-1}, and the composite of two similarity transformations with ratios of similitude r and s is a similarity transformation with ratio of similitude rs. Of course, if $r > 0$ then the invertible linear transformation rI on $\mathbf{R}^n$ is a similarity transformation with ratio of similitude r.■

FACT 2. *In $\mathbf{R}^n$ every similarity transformation with ratio of similitude r satisfying $F(0) = 0$ has the form $F(x) = rA(x)$ where A is given by an $n \times n$ orthogonal matrix.*■

FACT 3. *In $\mathbf{R}^n$ every similarity transformation F is conformal; specifically, if α and β are differentiable curves in $\mathbf{R}^n$ that are defined on a neighborhood of $0 \in \mathbf{R}$ such that $\alpha(0) = \beta(0)$ such that both $\alpha'(0)$ and $\beta'(0)$ are nonzero, then the angle between $\alpha'(0)$ and $\beta'(0)$ is equal to the angle between $[F\circ\alpha]'(0)$ and $[F\circ\beta]'(0)$.*■

The verification of the third property uses Facts 1 and 2 together with the additional observation that if A is given by an orthogonal transformation then A preserves inner products and hence the cosines of angles between vectors.

The key to relating inversions and stereographic projections is the following result:

PROPOSITION. *Let $e \in \mathbf{R}^n$ be a unit vector, and let T be inversion with respect to the sphere $S(1; 0)$. Then T interchanges the hyperplane defined by the equation $\langle x, e\rangle = -1$ with the nonzero points of the sphere $S(\frac{1}{2};\ -\frac{1}{2}e)$.*

Proof. By definition we have

$$T(x) = \frac{1}{\langle x, x\rangle} \cdot x$$

and therefore the proof amounts to finding all x such that

$$\left|\, T(x) + \tfrac{1}{2}e \,\right| = \tfrac{1}{2}.$$

This equation is equivalent to

$$\tfrac{1}{4} = \left|\, T(x) + \tfrac{1}{2}e \,\right|^2 = \langle T(x) + \tfrac{1}{2}e,\ T(x) + \tfrac{1}{2}e\rangle$$

and the last expression may be rewritten in the form

$$\langle T(x), T(x)\rangle + \langle T(x), e\rangle + \frac{1}{4} =$$

$$\frac{\langle x, x\rangle}{\langle x, x\rangle^2} + \frac{\langle x, e\rangle}{\langle x, x\rangle} + \frac{1}{4}$$

which simplifies to

$$\frac{1}{\langle x, x\rangle} + \frac{\langle x, e\rangle}{\langle x, x\rangle} + \frac{1}{4} \ .$$

Our objective was to determine when this expression is equal to $\frac{1}{4}$, and it follows immediately that the latter is true if and only if $1 + \langle x, e\rangle = 0$; *i.e.*, it holds if and only if $\langle x, e\rangle = -1$.■

An llustration of the preceding result appears in the files `stereopic1.*` in the course directory.

COROLLARY. *Let e be as above, and let W be the $(n-1)$-dimensional subspace of vectors that are perpendicular to e. Then the stereographic projection map from $S(1;0) - \{e\}$ to W is given by the restriction of the composite*

$$G \circ t \circ H$$

to $S(1;0) - \{e\}$, where H is the similarity transformation $H(u) = \frac{1}{2}(u - e)$ and G is the translation isometry $G(v) = v + e$.

Proof. It will be convenient to talk about the *closed ray* starting at a vector a and passing through a vector b; this is the image of the parametrized curve

$$\gamma(t) = (1-t)a + tb = a + t(b - a)$$

where $t \geq 0$.

First note that the composite $T \circ H$ sends the ray starting at the point e and passing through a point x with $\langle x, e\rangle < 1$ to the ray starting at e and passing through the point $H(x)$, and the latter satisfies $\langle H(x), e\rangle < 0$. This is true for the mapping H by Fact 1 and the equation $H(e) = 0$, and it is true for $T \circ H$ because T is inversion with respect to a sphere centered at 0. Furthermore, by construction H sends the sphere $S(1;0)$ to $S\left(\frac{1}{2};\ -\frac{1}{2}e\right)$, and it also sends the hyperplane P defined by $\langle x, e\rangle = -1$ to itself.

By construction, stereographic construction sends the point $y \in S(1;0) - \{e\}$ to the point $z \in W$ such that $z - e$ is the unique point at which the ray starting at e and passing through y meets the hyperplane P, and inversion sends the point $\eta \in S\left(\frac{1}{2};\ -\frac{1}{2}e\right) - \{0\}$ to the unique point α at which the ray starting at 0 and passing through η meets the hyperplane P.

Combining these, we see that $T \circ H$ maps y to the unique point where the ray passing through 0 and y meets the hyperplane P. This point may be written uniquely in the form $w - e$ where $w \in W$, and in fact we have $w = G \circ T \circ H(\mathrm{y})$. On the other hand, by the preceding paragraph we also know that w is given by the stereographic projection.■

The conformal property for inversions

The following is an immediate consequence of the Chain Rule and Fact 3 stated above:

PROPOSITION. *Let U be open in $\mathbf{R}^n$, let $x \in U$ and let $f : U \to \mathbf{R}^n$ be a $\mathbf{C}^1$ mapping. Then f is conformal at x if $Df(x)$ is a nonzero scalar multiple of an orthogonal transformation.*

The key observation behind the proposition is that if $\gamma : (-\delta, \delta) \to U$ is a differentiable curve with $\gamma(0) = x$ then

$$[f \circ \gamma]'(0) = Df(x)[\gamma'(0)]$$

by the Chain Rule.■

We are now ready to prove the result that we wanted to establish.

THEOREM. *Every inversion map is conformal.*

Proof. It is convenient to reduce everything to the case where the sphere is $S(1; 0)$. Given an arbitrary sphere $S(r; y)$ there is a similarity transformation sending $S(r; y)$ to $S(1; 0)$ that is defined by the formula $F(x) = r^{-1}(x - y)$, and if T' and T are the associated inversions then we have

$$T' = F^{-1} \circ T \circ F \ .$$

By construction $DF(x) = r^{-1}I$ for all x and therefore it follows that $DT' = DT$ (as usual, "D" denotes the derivative of a function). Therefore, by the proposition it will suffice to show that DT is always a scalar multiple of an orthogonal map.

Let $x \neq 0$ and write an arbitrary vector $v \in \mathbf{R}^n$ as a sum $v = cx + u$ where $c \in \mathbf{R}$ and $\langle x, u \rangle = 0$. We have already noted that $T(x) = \rho(x)^{-2} \cdot x$ where $\rho(x) = |x|$, and we wish to use this in order to compute the value of $DT(x)$ at some vector $h \in \mathbf{R}^n$. The appropriate generalization of the Leibniz Rule for products and elementary multivariable calculus show that

$$DT(x)[h] = \rho(x)^{-2}h + \langle (\nabla[\rho(x)]^{-2}, h \rangle \cdot x$$

where

$$\nabla[\rho(x)]^{-2} = -[\rho(x)]^{-4} \cdot \nabla[\rho(x)]^2 = -2[\rho(x)]^{-4}x \ .$$

If $\langle x, h \rangle = 0$ this shows that $DF(x)[h] = [\rho(x)]^{-2}h$, while if $h = x$ it follows that $DF(x)[x] = -[\rho(x)]^{-2}x$. In particular, this shows that there is an orthonormal basis for $\mathbf{R}^n$ consisting of eigenvectors for $DT(x)$ with associated eigenvalues $\pm\rho(x)^{-2}$, and therefore it follows that $DT(x)$ is a positive scalar multiple of an orthogonal map, which in turn implies that T is conformal at x. Since x was arbitrary, this proves the theorem.■

Appendix C : Homogeneity of open sets in Euclidean spaces

The purpose of this section is to prove a result that was stated in Appendix A:

HOMOGENEITY THEOREM. *If U is an open connected subset of $\mathbf{R}^n$ and a and b are distinct points of U, then there is a homeomorphism $h : U \to U$ such that $h(a) = b$.*

There are two major steps in the proof; the first is to show that connected open sets are locally homogeneous, and the second is to use local homogeneity and connectedness to prove global homogeneity.

A local construction

The following result provides one way to verify local homogeneity:

PROPOSITION. *Let D^n be the solid unit disk in $\mathbf{R}^n$, and let $v \in D^n$ be an interior point with $|v| < 1$. Then there is a homeomorphism $f : D^n \to D^n$ such that f is the identity on S^{n-1} and $f(0) = v$.*

Sketch of Proof. The geometric motivation is simple. Every point on D^n lies on a closed segment joining the origin to a point on S^{n-1}. One maps such a segment linearly to the segment joining v to the same point on S^{n-1}. As is often the case with such geometrical ideas, it takes a fair amount of algebraic manipulation to show that this actually works. An illustration of the basic idea is given in the files `radialproj.*` in the course directory.

The following result will be useful in the course of the proof:

LEMMA. *D^n is homeomorphic to the quotient space of $S^{n-1} \times [0,1]$ whose equivalence classes are the one point subsets for all points with positive first coordinates together with the set $S^{n-1} \times \{0\}$.*

Proof of lemma. Consider the continuous map $g : S^{n-1} \times [0,1] \to D^n$ sending (x,t) to tx. The inverse images of points are precisely the classes described above, and since g is a closed and surjective mapping it follows that the quotient space with the given equivalence classes is homeomorphic to the image, which is D^n.■

Proof of proposition continued. Define $F : S^{n-1} \times [0,1] \to D^n$ by the formula

$$F(u,t) = tu + (1-t)v \ .$$

If $t = 0$ then $F(u,t) = F(u,0) = v$ and therefore the lemma implies that F passes to a continuous mapping $f : D^n \to D^n$. This map corresponds to the geometric idea proposed in the first paragraph of the proof. We need to prove that f is 1–1 and onto.

By construction we have $f(0) = v$; it will be useful to start by proving that no other point maps to v. But suppose that we have $f(tu) = v$ where $t \in (0,1]$ and $|u| = 1$. By construction this means that

$$v = tu + (1-t)\,v$$

which is equivalent to the equation $t(u - v) = 0$. Since $|v| < 1 = |u|$ we clearly have $u \neq v$ so this forces the conclusion that $t = 0$, which contradicts our original assumption that $t > 0$.

Suppose now that we have $w \in D^n - \{v\}$. Consider first the case where $w \in S^{n-1}$. Then by construction $w = f(w)$, and there is no other point w^* on S^{n-1} such that $w^* = f(w)$. On the other hand, if $|\xi| < 1$ we claim that $f(\xi) \notin S^{n-1}$ because

$$|f(\xi)| \;=\; \big|\, |\xi| u + (1 - |\xi|) v \,\big| \;\leq\; |\xi| + (1 - |\xi|) \cdot |v| \;<\; |\xi| + (1 - |\xi|) \;=\; 1 \;.$$

Therefore we only need to show that *if* $|w| < 1$ *then there is a unique point* ξ *such that* $|\xi| < 1$ *and* $f(\xi) = w$. This in turn reduces to showing that there is a unique point $u \in S^{n-1}$ such that w lies on the open segment joining v and u; the latter is equivalent to showing that *there is a unique real number* t *such that* $t > 1$ *and*

$$\big| v + t(w - v) \big| = 1 \;.$$

Rather than prove this by working out an explicit but complicated formula for t in terms of v and w, we shall approach the assertion by analyzing the behavior of the quadratic function

$$f(t) \quad = \quad \big| v + t(w - v) \big|^2$$

which represents the square of the distance between the point $v + t(w - v)$ and the origin. The coefficient of t^2 for this function is the positive number $|w - v|^2$ and therefore $\lim_{t \to \pm\infty} f(t) = +\infty$. Furthermore, $f(0) = |v|^2 < 1$ and $f(1) = |w|^2 < 1$, and therefore there are exactly two values of t such that $f(t) = 1$, one of which is greater than 1 and one of less is less than 0.

If s is chosen so that $s > 1$ and $f(s) = 1$ in the notation of the preceding paragraphs and we let $u = v + s(v - w)$, then $f(x) = w$ if and only if

$$x \quad = \quad w \;+\; \left(\frac{1 - s}{s} \right) \; v \;.\blacksquare$$

The local homogeneity of an open subset of $\mathbf{R}^n$ is an direct consequence of the proposition. Note first that the proposition above remains true for every closed disk in $\mathbf{R}^n$ of the form

$$\mathbf{D}(y; r) \quad = \quad \{ \; y \in \mathbf{R}^n \; | \; \; |y - a| = r \; \} \;.$$

Local homogeneity

The following local result is an immediate consequence of the preceding proposition:

LOCAL HOMOGENEITY PROPERTY. *If* U *is open in* $\mathbf{R}^n$ *and* $a \in U$, *then there is an open neighborhood* v *of* x *such that* $V \subset U$ *and for all* $x \in V$ *there is a homeomorphism* $h : U \to U$ *sending* a *to* x.$\blacksquare$

Proof. Note first that the proposition above remains valid for every closed disk in $\mathbf{R}^n$ of the form

$$\mathbf{D}(y; r) \quad = \quad \{ \; y \in \mathbf{R}^n \; | \; \; |y - a| \leq r \; \} \;.$$

To see this, observe that this disk is the image of D^n under the self-homeomorphism h of $\mathbf{R}^n$ sending x to $rx + a$. Thus if $b \in \mathbf{D}(y; r)$ with $|b - a| < r$ we can construct a homeomorphism sending a to b and fixing the boundary sphere by the formula

$$g(y) = h\left(f\left(h^{-1}(y)\right)\right)$$

where h is the homeomorphism in the proposition sending 0 to $h^{-1}(b)$. Given an open set $U \subset \mathbf{R}^n$ and a point $a \in U$, one can find some $r > 0$ so that $\mathbf{D}(y; r) \subset U$. For each b in the interior of this disk we can construct a homeomorphism of the closed disk sending a to b such that the restriction of the homeomorphism to the boundary is the identity. Given two points $b\lambda$, b' in the interior of this disk one can use such homeomorphisms and their inverses to construct a homeomorphism of the disk that is again fixed on the boundary sphere and sends b to b'. This homeomorphism can be extended to the entire open set U by defining it to be the identity on the complement of the disk (why does this work?). This proves the desired property.■

A global homogeneity theorem

Suppose now that $U \subset \mathbf{R}^n$ is open and connected. The following result allows us to extend the local homogeneity property to a global result:

PROPOSITION. *Let X be a connected topological space, and for each $x \in X$ suppose that there is an open neighborhood V of X such that for each $v \in V$ there is a homeomorphism $h : X \to X$ such that $h(x) = v$. Then X is homogeneous; in other words, for each pair of distinct points x, $y \in X$ there is a homeomorphism $h : X \to X$ such that $h(x) = y$.*

Proof. Consider the binary relation on X given by $a \sim b$ if and only if there is a homeomorphism $h : X \to X$ such that $h(a) = b$. It follows immediately that this is an equivalence relation. Furthermore, if C is an equivalence class of $\sim$ and $x \in C$, then there is an open neighborhood V of x such that $V \subset C$. In particular, it follows that C is open. Since this is true for all equivalence classes it follows that the latter decompose X into pairwise disjoint open subsets. Therefore the union of all equivalence classes except C is also open, and hence C is closed. But X is connected, and therefore we must have $C = X$, so that all points of X are equivalent under $\sim$. By the definition of the latter, it follows that X is homogeneous.■

Appendix D : Normal forms for orthogonal transformations

The Spectral Theorem in linear algebra implies that a normal linear transformation on a complex inner product space (one that commutes with its adjoint) has an orthonormal basis of eigenvectors. In particular, since the adjoint of a unitary transformation is its inverse, the result implies that every unitary transformation has an orthonormal basis of eigenvectors.

It is clear that one cannot have a direct generalization of the preceding result to orthogonal transformations on real inner product spaces. In particular, plane rotations given by matrices of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

do not have such a basis over the reals except in the relatively trivial cases when θ is an integral multiple of π and the matrices reduce to $(-I)^k$ for some integer k. However, if one takes these into account it is possible to prove the following strong result on the existence of a "good" orthonormal basis for a given orthogonal transformation.

NORMAL FORM. *Let V be a finite-dimensional real inner product space, and let $T : V \to V$ be an orthogonal transformation of V. Then there is an orthonormal direct sum decomposition of V into T-invariant subspaces W_i such that the dimension of each W_i is either* 1 *or* 2*.*

In particular, this result implies that there is an ordered orthonormal basis for V such that the matrix of T with respect to this ordered orthonormal basis is a block sum of 2×2 and 1×1 orthogonal matrices.

It is beyond the scope of these notes to go into detail about the results from a standard linear algebra course that we use in the proof of the result on normal forms. Virtually all of the background information can be found in nearly any linear algebra text that includes a proof of the Spectral Theorem (*e.g.*, the book by Fraleigh and Beauregard cited in the bibliography).

Small dimensional orthogonal transformations

Since orthogonal transformations preserve the lengths of vectors it is clear that a 1-dimensional orthogonal transformation is just multiplication by ± 1. It is also not difficult to describe 2-dimensional orthogonal transformations completely using the fact that their columns must be orthonormal. In particular, it follows that the matrices representing 2-dimensional orthogonal transformations of $\mathbf{R}^2$ with respect to the standard inner product have the form

$$\begin{pmatrix} \cos\theta & \mp\sin\theta \\ \sin\theta & \pm\cos\theta \end{pmatrix}$$

for some real number θ. We have already noted that one class of cases corresponds to rotation through θ, and it is an exercise in linear algebra to check that each of the remaining matrices

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

has an orthonormal basis of eigenvectors, with one vector in the basis having eigenvalue -1 and the other having eigenvalue $+1$. The details of verifying this are left to the reader as an exercise (*Hint:* First verify that the characteristic polynomial is $t^2 - 1$).

One can use this to prove a geometrically sharper version of the result on normal form.

STRONG NORMAL FORM. *Let V be a finite-dimensional real inner product space, and let $T : V \to V$ be an orthogonal transformation of V. Then there is an orthonormal direct sum decomposition of V into T-invariant subspaces W_i such that the dimension of each W_i is either* 1 *or* 2 *and T operates by a plane rotation on each* 2*-dimensional summand.*

This is an immediate consequence of the preceding discussion, for if T operates on an invariant 2-dimensional subspace with by a map that is not a rotation, then we can split the subspace into two 1-dimensional eigenspaces.

Complexification

If $V = \mathbf{R}^n$ and T is represented by the orthogonal matrix A, it is clear how one can extend T to a unitary transformation on $\mathbf{C}^n$. We shall need a version of this principle that works for an arbitrary real inner product space V. The idea is simple but a little inelegant; it can be done better if one uses tensor products, but we want to prove the result without introducing them.

One defines the complexification of V formally to be very much as one defines the complex numbers. The underlying set of complex vectors $V_{\mathbf{C}}$ is given by $V \times V$, addition is defined in a coordinatewise fashion, multiplication by a complex scalar $a + bi$ is given by the formula

$$[a + bi](v, w) = (av - bw, bv + aw)$$

and the complexified inner product is defined by

$$\langle\, (v, w), (v', w')\, \rangle_{\text{complex}} =$$

$$(\langle v, v' \rangle + \langle w, w' \rangle) + i \cdot (\langle w, v' \rangle - \langle v, w' \rangle) \ .$$

Both intuitively and formally the pair (v, w) can be viewed as $v + i\, w$ if one identifies a vector v in the original space with $(v, 0)$ in the complexification. Verification that the structure defined above is actually a complex inner product space is essentially an exercise in bookkeeping and will not be carried out here.

If we are given a linear transformation $T : V \to W$ of real inner product spaces, then

$$T \times T : V \times V \to W \times W$$

defines a complex linear transformation $T_{\mathbf{C}}$ on the complexification. Moreover, this construction is compatible with taking adjoints and composites:

$$(T^*)_{\mathbf{C}} = (T_{\mathbf{C}})^*$$

$$(S \circ T)_{\mathbf{C}} = S_{\mathbf{C}} \circ T_{\mathbf{C}}$$

In particular, *if T is orthogonal then $T_{\mathbf{C}}$ is unitary.*

The main ideas behind the proof are

(1) to extract as much information as possible using the Spectral Theorem for the unitary transformation $T_{\mathbf{C}}$,

(2) prove the result by induction on the dimension of V.

In order to do the latter we need the following observation that mirrors one step in the proof of the Spectral Theorem.

PROPOSITION. *Let T be as above, and assume that $W \subset V$ is T-invariant. Then the orthogonal complement $W^{\perp}$ is also T-invariant.*■

In the Spectral Theorem the induction proof begins by noting that there is an invariant 1-dimensional subspace corresponding to an eigenvector for T. One then obtains a transformation on the orthogonal complement of this subspace and can apply the inductive hypothesis to the associated transformation on this invariant subspace. We would like to do something similar here, and in order to begin the induction we need the following result.

KEY LEMMA. *If V is a finite dimensional real inner product space and $T : V \to V$ is an orthogonal transformation, then there is a subspace $W \subset V$ of dimension 1 or 2 that is T-invariant.*

If we have this subspace, then we can proceed as before, using the induction hypothesis to split $W^{\perp}$ into an orthogonal direct sum of T-invariant subspaces of dimension 1 or 2, and this will complete the derivation of the normal form.■

Proof of the key lemma. Suppose that λ is an eigenvalue of $T_{\mathbf{C}}$ and x is a nonzero eigenvector; since $T_{\mathbf{C}}$ is unitary we know that $|\lambda| = 1$. Express both λ and x in terms of their real and imaginary components, using the fact that $|\lambda| = 1$:

$$\lambda \quad = \quad \cos\theta + i\,\sin\theta, \qquad x \quad = \quad (v, w) \quad = \quad v + i\,w$$

Then the eigenvalue equation $T_{\mathbf{C}}(x) = \lambda\,x$ may be rewritten in the form

$$T(v) + i\,T(w) \quad = \quad T_{\mathbf{C}}(x) \quad = \quad \lambda\,x \quad = \quad (\cos\theta + i\,\sin\theta)\cdot(v + i\,w) \quad =$$

$$\Big(\cos\theta\,v - \sin\theta\,w\Big) + i\,\Big(\sin\theta\,v + \cos\theta\,w\Big)$$

which yields the following pair of equations:

$$T(v) \quad = \quad \cos\theta\,v - \sin\theta\,w$$

$$T(w) \quad = \quad \sin\theta\,v + \cos\theta\,w$$

It follows that the subspace W spanned by v and w is a T-invariant subspace, and since it is spanned by two vectors its dimension is at most 2. On the other hand, since $x \neq 0$ we also know that at least one of v and w is nonzero and therefore the dimension of W is at least 1.■

References

The following are the texts for the course:

J. R. Munkres. Topology. (Second Edition), *Prentice-Hall, Saddle River NJ*, 2000. ISBN: 0–13–181629–2.

C. H. Edwards, Jr. Advanced Calculus of Several Variables. (Corrected Reprint of 1973 Edition). *Dover, Mineola NY*, 1994. ISBN: 0–496–68336–2.

Further references

Alexandroff, Pavel S.; Hopf, Heinz. Topologie, Band 1. (German. Grundlehren der mathematischen Wissenschaften *etc.* Nr. 45.) *Julius Springer, Berlin*, 1935.

Amadio, Roberto M.; Curien, Pierre-Louis. Domains and Lambda Calculi. (Cambridge Tracts in Theoretical Computer Science, 46.) *Cambridge University Press, Cambridge*, 1998. ISBN: 0-521-62277-8.

Birkhoff, Garrett; MacLane, Saunders. A Survey of Modern Algebra. (Reprint of the Third 1968 Edition.) *Chelsea, New York NY*, 1988. ISBN: 0-023-74310-7.

Birkhoff, Garrett; MacLane, Saunders. A Survey of Modern Algebra. (Fourth Edition.) *Prentice-Hall, Englewood Cliffs NJ*, 1977. ISBN: 0-023-10070-2.

Bourbaki, Nicolas. General Topology, Chapters 1–4. (Translated from the French. Reprint of the 1989 English translation. Elements of Mathematics.) *Springer-Verlag, Berlin-Heidelberg-New York, etc.*, 1998. ISBN: 3-540-64241-2.

Bourbaki, Nicolas. General Topology, Chapters 5–10. (Translated from the French. Reprint of the 1989 English translation. Elements of Mathematics.) *Springer-Verlag, Berlin-Heidelberg-New York, etc.*, 1998. ISBN: 3-540-64563-2 .

Christenson, Charles O.; Voxman, William L. Aspects of Topology. (Pure and applied Mathematics, Vol. 39.) *Marcel Dekker, New York-Basel*, 1977.

Christenson, Charles O.; Voxman, William L. Aspects of Topology. (Second Edition.) *BCS Associates, Moscow ID*, 1998. ISBN: 0-914351-07-9; 0-914351-08-7.

Dieudonné, Jean A. History of algebraic geometry. An outline of the history and development of algebraic geometry. (Translated from French by Judith D. Sally. Wadsworth Mathematics Series.) *Wadsworth International, Belmont, CA*, 1985. ISBN: 0-534-03723-2.

Dugundji, James. Topology. (Reprint of the 1966 Edition. Allyn and Bacon Series in Advanced Mathematics. *Allyn and Bacon, Boston MA-London (U.K.)-Sydney (Austr.)*, 1978. ISBN: 0-205-00271-4.

Fraleigh, John B.; Beauregard, Raymond A. Linear Algebra (Third Edition. Historical notes by Victor J. Katz) *Addison-Wesley, Reading MA etc.*, 1995. ISBN: 0-2-1-52675-1.

Franz, Wolfgang. General Topology. (Translated from the German by Leo F. Boron.) *Frederick Ungar, New York NY*, 1965.

Goffman, Casper; Pedrick, George. First Course in Functional Analysis. (Second Edition.) *Chelsea, New York NY*, 1983. ISBN: 0-828-40319-8.

Hewitt, Edwin; Stromberg, Karl. Real and Abstract Analysis. A Modern Treatment of the Theory of Functions of a Real Variable. (Graduate Texts in Mathematics Vol. 25.) *Springer-Verlag, Berlin-Heidelberg-New York-etc.*, 1997. ISBN: 0-387-90138-8.

Hocking, John G.; Young, Gail S. Topology. (Second edition.) *Dover, New York NY*, 1988. ISBN: 0-486-65676-4.

James, I. M. (*ed.***).** History of topology. *North-Holland, Amsterdam (NL)*, 1999. ISBN: 0-444-82375-1.

Jänich, Klaus. Topology. (With a chapter by Theodor Bröcker. Translated from the German by Silvio Levy. Undergraduate Texts in Mathematics.) *Springer-Verlag, Berlin-Heidelberg-New York*, 1984. ISBN: 0-387-90892-7.

Kaplansky, Irving. Set Theory and Metric Spaces. (Second edition.) *Chelsea, New York*, 1977. ISBN 0-8284-0298-1.

Kaplansky, Irving. Lie Algebras and Locally Compact Groups. (Chicago Lectures in Mathematics Series.) *University of Chicago Press, Chicago IL*, 1995. ISBN: 0-226-42453-7.

Kasriel, Robert H. Undergraduate Topology. (Reprint of the 1971 Edition.) *Robert E. Krieger, Huntington NY*, 1977. ISBN: 0-721-65298-0

Kelley, John L. General Topology. (Reprint of the 1955 Edition. Graduate Texts in Mathematics, No. 27.) *Springer-Verlag, Berlin-Heidelberg-New York*, 1975. ISBN: 0-387-90125-6.

Krantz, Steven G. Real Analysis and Foundations. (Studies in Advanced Mathematics.) *CRC Press, Boca Raton FL*, 1991. ISBN: 0-8493-7156-2.

McCarty George S. Topology : An Introduction with Application to Topological Groups. (Second Edition.) *Dover, New York NY*, 1988. ISBN: 0-486-65633-0.

Montgomery, Deane; Zippin, Leo. Topological Transformation Groups. (Reprint of the 1955 Edition.) *Robert E. Krieger, Huntington NY*, 1974. ISBN: 0-88275-169-7

Pontryagin, Lev S. Topological Groups. (Translated from the Second Russian Edition with a preface by Arlen Brown, with additional material translated by P. S. V. Naidu. Classics of Soviet Mathematics) *Taylor and Francis, London (U. K.)*, 1987. ISBN: 2-881-24133-6

Rickart, Charles E. General Theory of Banach Algebras. (Reprinted with Corrections.) *Robert E. Krieger, Huntington NY*, 1974. ISBN: 0-88275-091-7.

Royden, Halsey L. Real Analysis. (Third Edition.) *Macmillan, New York NY*, 1988. ISBN: 0-02-404151-3.

Rudin, Walter. Principles of Mathematical Analysis. (Third Edition. International Series in Pure and Applied Mathematics.) *McGraw-Hill, New York-Auckland-Düsseldorf*, 1976. ISBN: 0-07-054235-X.

Rudin, Walter. Functional Analysis. (Second Edition. International Series in Pure and Applied Mathematics.) *McGraw-Hill, New York NY, etc.*, 1991. ISBN: 0-07-054236-8.

Steen, Lynn Arthur; Seebach, J. Arthur, Jr. Counterexamples in Topology. (Reprint of the Second (1978) Edition.) *Dover, Mineola NY*, 1995. ISBN: 0-486-68735-X.

Taylor,Paul. Practical Foundations of Mathematics. (Cambridge Studies in Advanced Mathematics 59.) *Cambridge University Press, Cambridge*, 1998. ISBN: 0-521-63107-6.

Wolf, Robert S. Proof, Logic, and Conjecture: The Mathematician's Toolbox. *W. H. Freeman, New York NY*, 1998. ISBN: 0-7167-3050-2.

Summary of Files in the Course Directory

The course directory is $\sim$`res/math205A` on the `math.ucr.edu` network.

`braintest.pdf`
`braintest.ps`

This is definitely not a serious piece of course material, but it does illustrate the importance of staying focused on the main points when learning or doing mathematics.

`categories.pdf`
`categories.ps`

A brief survey of category theory, not needed for the course but included as background.

`concat.pdf`
`concat.ps`

A picture illustrating the stringing together, or concatenation, of two curves, where the ending point of the first is the starting point of the second.

`contents.dvi`
`contents.pdf`
`contents.ps`

The title page and table of contents for this document.

`coursehw.dvi`
`coursehw.pdf`
`coursehw.ps`

The electronic files containing the homework assignments for this course; some of these come from the two texts and others are also included. Solutions will be provided in the course directory. Preliminary versions with names of the form `prelimhw.`$*$ will initially be in the course directory.

`coursetext.dvi`
`coursetext.pdf`
`coursetext.ps`

The electronic files for viewing or printing this document. Preliminary versions with names of the form `prelimtext.`$*$ will initially be in the course directory.

`cubicroots.dvi`
`cubicroots.pdf`
`cubicroots.ps`

This note discusses the application of the Contraction Lemma to finding the roots of cubic polynomials with real coefficients.

`foundations1.pdf`
`foundations1.ps`

A brief discussion of basic logic and some very elementary set theory. This is meant as review and/or background.

foundations2.pdf
foundations2.ps

This discusses topics on relations and functions from the perspective of this course to the extent that it differs from the main text; it also includes comments on axioms for the positive integers and discussions of the roles of the Axiom of Choice and the Generalized Continuum Hypothesis in set theory. The latter discussions are not needed for the course but may be helpful for obtaining a broader perspective of the mathematical foundations underlying the course.

general205A.dvi
general205A.pdf
general205A.ps

This is the handout from the first day of class.

math205Afall03.pdf
math205Afall03.ps

The course outline.

nicecurves.dvi
nicecurves.pdf
nicecurves.ps

This an appendix to the course notes, and it proves an assertion from Section III.5 about joining two points in a connected open subset of Euclidean space by a curve that is infinitely differentiable and has nonzero tangent vectors at every point.

ordinals.dvi
ordinals.pdf
ordinals.ps

Background material on the ordinal numbers. The latter may be viewed as equivalence classes of well-ordered sets, and the class of all ordinals determines a natural indexing for the class of all cardinal numbers. This summary, which was written by N. Strickland of the University of Sheffield, is available online from

http://www.shef.ac.uk/ pm1nps/courses/topology/ordinals.∗

and is included in the course directory only for the convenience of students enrolled in this course. The site also includes several other interesting and informative documents related to this course.

polya.pdf
polya.ps

A one page summary of the advice for solving problems in the classic book, *How to Solve It*, by G. Pólya (with an extra piece of advice added at the end).

proper.dvi
proper.pdf
proper.ps

A discussion of the basic facts about proper maps, with emphasis on their properties that are relevant to algebraic geometry. This is supplementary material not covered in the course, and the level of exposition is slightly higher than in the course notes.

radproj.pdf
radproj.ps

A picture illustrating the radial projection map that is used to prove a result in Appendix C.

realnumbers.pdf
realnumbers.ps

A summary of the basic properties of the real numbers.

smirnov.dvi
smirnov.pdf
smirnov.ps

A proof of Smirnov's Theorem on constructing global metrics for topological spaces whose topologies locally come from metric spaces. This is supplementary material not covered in the course, and the level of exposition is slightly higher than in the course notes.

solutions∗.dvi
solutions∗.pdf
solutions∗.ps

Solutions to the assigned exercises in the course, where $*$ ranges from 1 to 5. The first file contains solutions through Section II.1, the second contains solutions through the remainder of Unit II, the third contains solutions for Unit III, the fourth contains solutions for Units IV and V, and the fifth contains solutions for Unit VI. No solutions are provided for the exercises listed for the appendices.

stereopic1.pdf
stereopic1.ps

A picture showing the relationship between stereographic projections and inversions with respect to spheres.

stereopic2.pdf
stereopic2.ps

More pictures involving stereographic projections.

swisscheese.pdf
swisscheese.ps

A picture illustrating how a square with two holes can be viewed as a compactification of the Euclidean plane.