Mathematics 205A

Introduction to Topology — I

Course Notes — Part Two

Revised, Fall 2014

Department of Mathematics

University of California, Riverside

Table of Contents

Introduction

This is the second part of the first entry level graduate courses in topology and geometry, and the goal is to develop algebraic techniques for analyzing closed curves in a topological space. More detailed discussions of the motivation and approach are given below and at the beginning of Unit VII (the first unit in this part of the course). The numbering of these notes is a continuation of the numbering in the notes for the first part of the course (the file `gentopnotes2014.pdf`); for example, Theorem VI.7.8 would refer to Theorem 8 in Section VI.7 of the cited document (however, there is no such section in those notes).

The basic texts for this portion of the course are the following:

[**M**] **J. R. Munkres.** *Topology* (Second Edition), *Prentice-Hall, Saddle River NJ*, 2000. ISBN: 0–13–181629–2. [This is the text for the previous course in the sequence.]

[**H**] **A. Hatcher.** *Algebraic Topology* (Third Paperback Printing), *Cambridge University Press, New York NY*, 2002. ISBN: 0–521–79540–0.

This book can be legally downloaded from the Internet at no cost for personal use, and here is the link to the online version:

`www.math.cornell.edu/~hatcher/AT/ATpage.html`

The material in this part of the course has become fairly standard, and it is directly related to some phenomena involving line integrals and functions of a complex variable. One major theme is the creation of algebraic "pictures" of a topological space which are obtained by studying certain types of topological configurations in the space. Ever since (at least) the beginning of the 17^{th} century, mathematicians and others have recognized the effectiveness and power of algebraic techniques for analyzing geometrical problems by transforming geometric input into algebraic terms, solving the associated algebraic questions, and translating the algebraic results back into the original geometric setting. The central concept in the second part of the course is the *fundamental group* of a space, which is an algebraic object constructed from the 1-dimensional configurations given by closed curves which start and end at a fixed basepoint.

One way to compare the two parts of the course is to describe the conclusions which follow from the respective methods: Using point set topology one can show that $\mathbb{R}$ and $\mathbb{R}^n$ are not homeomorphic if $n \geq 2$, and using point fundamental groups, one can show that $\mathbb{R}^2$ and $\mathbb{R}^n$ are not homeomorphic if $n \geq 3$. — In the second course of the geometry/topology sequence, still other methods are developed to prove that $\mathbb{R}^m$ and $\mathbb{R}^n$ are never homeomorphic if $m \neq n$. Although this may seem obvious intuitively, it is still necessary to give a formal proof because intuition can be misleading.

Comments on the texts

Taking all things into account, the first part of Munkres (on point set topology) is one of the very best accounts of the subject, with an excellent balance of clear exposition, logical completeness and drawings to motivate the underlying geometrical content of the subject (there are some peculiar choices of terms and symbolism, and in a number of instances more motivation would help, but the perfect text is an ideal which is rarely if ever realized). The second part of Munkres, which includes the material on fundamental groups, comes close to meeting this standard. However, there are numerous cases where more motivational comments and drawings would help, and sometimes the logical thoroughness of the exposition interferes with its clarity. To its credit, the second part gives logically complete accounts of several basic applications of topology to basic geometrical results like the Fundamental Theorem of Algebra and the Jordan Curve Theorem (a simple closed curve in the plane separates it into two connected pieces), but the proofs really push the theory in the book to its limits, and consequently the reasoning is often very delicate and difficult to follow. We shall see that homology theory often yields much simpler and more conceptual proofs.

Hatcher's book begins by covering the same topics which appear in the second half of Munkres, and it proceeds to go much further in the subject. The challenges faced in covering the further material are much greater than the corresponding challenges in Munkres. In particular, the gap between abstract formalism and geometrical intuition is much greater, and it is not clear how well any single book can reconcile these complementary factors. More often than not, algebraic topology books stress the former at the expense of the latter, and one important strength of Hatcher's book is that its emphasis tilts very much in the opposite direction. The book makes a sustained effort to include examples that will provide insight and motivation, using pictures as well as words, and it also attempts to explain how working mathematicians view the subject. Because of these objectives, the exposition in Hatcher is significantly more casual than in many books on the subject. Unfortunately, the book's informality is arguably taken too far in numerous places, leading to significant problems in several directions; these include assumptions about prerequisites, clarity, wordiness, thoroughness and some sketchy motivations that are difficult for many readers to grasp.

Prerequisites

This part of the course is based upon material developed in the first part. There is also some background material which is needed here but does not appear in the first part of the course.

Algebra

Some concepts in group theory are needed; most are at the undergraduate level. Several other concepts from group theory are presented in Munkres and will be covered in the course. Material from standard undergraduate linear algebra courses will also be used as needed. Everything we need can be found in the following standard graduate algebra textbook:

> **T. Hungerford.** *Algebra.* (Reprint of the 1974 original edition, Graduate Texts in Mathematics, No. 73.) *Springer-Verlag, New York–Berlin–etc.*, 1980. ISBN: 0–387–90518–9.

At some points in this course we shall invoke the following basic result, which is proved in graduate level algebra courses (for example, Sections II.1 and II.2 of Hungerford).

STRUCTURE THEOREM FOR FINITELY GENERATED ABELIAN GROUPS. *Let G be a finitely generated abelian group (so every element can be written as a monomial in integral powers of some finite subset $S \subset G$). Then G is isomorphic to a direct sum*

$$(H_1 \oplus \cdots \oplus H_b) \oplus (K_1 \oplus \cdots \oplus K_s)$$

where each H_i is infinite cyclic and each K_j is finite of order t_j such that t_{j+1} divides t_j for all j. — For the sake of uniformity set $t_j = 1$ if $j > s$. Then two direct sums as above which are given by $(b; t_1, \cdots)$ and $(b'; t'_1, \cdots)$ are isomorphic if and only if $b = b'$ and $t_j = t'_j$ for all j.

For the purposes of this course, it is enough to understand the statement of the structure theorem; the proof itself is not part of the course or its prerequisites.

Analysis

We shall assume the basic material from an upper division undergraduate course in real variables as well as material from a lower division undergraduate course in multivariable calculus through the theorems of Green and Stokes as well as the 3-dimensional Divergence Theorem. The classic text by W. Rudin (*Principles of Mathematical Analysis*, Third Edition) is an excellent reference for real variables, and the following multivariable calculus text contains more information on the that subject than one can usually find in the usual 1500 page calculus texts (unfortunately, this book is far from perfect, but especially at the graduate level it may be useful for review purposes). Clearly there are also many other sources for this material; the main point is that a reader might have to refresh his or her memory on a few topics at some points in the notes.

> **J. E. Marsden and A. J. Tromba.** *Vector Calculus* (Fifth Edition), *W. H. Freeman & Co., New York NY*, 2003. ISBN: 0–7147–4992–0.

Category theory

The concept of a **category** does not appear explicitly in Munkres, but it is implicit in many places, and at numerous points in these notes it will be useful to formulate things using categories as part of the framework. We have attempted to limit this usage to situations where the formalism seems to simplify the discussion, so only need a few basic ideas are needed, and we shall summarize here the concepts that appear almost immediately in the notes. The course directory file

`categories2014.pdf`

(which is 10 pages long) gives a more organized treatment of the topics discussed here, and **reading the document is an implicit assignment for the course.**

A **category** is an abstract mathematical system which reflects some very basic features of many classes of mathematical objects and the well-behaved morphisms relating them. In set theory the objects and morphisms are sets and functions of sets, and in topology the most basic examples involve topological spaces and continuous mappings. There are many algebraic examples, including groups and morphisms, vector spaces over a fixed field $\mathbb{F}$ and $\mathbb{F}$-linear transformations, or partially ordered sets and monotonically increasing functions. Many other examples appear in the file cited above.

In addition to a family of **objects** as above, the data for a category also include **morphisms** from one object to another with specified objects as their *domains* (sources) and *codomains* (targets); a morphism together with its domain and codomain are generally denoted by notation like $f : A \to B$. There are also binary algebraic operations defined for certain pairs of morphisms, and these behave formally like composition of functions in the following respects:

(*i*) The composition $g \circ f$ of g and f is defined if and only if the target of f is the source of g.

(*ii*) For each object X there is an "identity morphism" $1_X : X \to X$ (sometimes we call this map id_X), and for each morphism $f : A \to B$ we have $1_B \circ f = f = f \circ 1_A$.

(*iii*) There is an associative law $h \circ (g \circ f) = (h \circ g) \circ f$ for threefold compositions.

The most important additional concept is that of an **isomorphism** between two objects A and B. This involves a pair of morphisms $f : A \to B$ and $g : B \to A$ such that $g \circ f = 1_A$ and $f \circ g = 1_B$. As elsewhere in mathematics, if one has such a pair of morphisms we say that f and g are inverse to each other (or inverses of each other).

The ubiquity of categories reflects a basic fact: *If a class of mathematical objects is defined, it is usually possible to define a good concept of mappings or morphisms from one object to another without too much trouble.* — From this perspective, it is natural to speculate about an appropriate notion of morphism relating one category to another. It turns out that there are **two** such notions called **contravariant functors** and **covariant functors**. A *covariant functor* is a system of transformations such that

(*a*) for each object X in the source category there is an associated object $T(X)$ in the target category,

(*b*) for each morphism $f : X \to Y$ in the source category there is an associated morphism $T(f) : T(X) \to T(Y)$ in the target category,

(*c*) the construction on morphisms preserves identity morphisms and compositions; the latter means that $T(g \circ f) = T(g) \circ T(f)$.

Here is an example involving topological spaces: If X is a topological space, let $T(X)$ be the set of continuous curves $\gamma : [0,1] \to X$, and if $f : X \to Y$ is continuous define $T(f)\gamma = f \circ \gamma$. The fundamental group of a pointed space is a more sophisticated example of this sort going from pointed topological spaces to groups.

As noted above, there is also a dual concept of *contravariant functor* from one category to another; the main differences with covariant functors are that a morphism $f : A \to B$ is sent to $T(f) : T(B) \to T(A)$ (*i.e.*, the domain and codomain are switched) and the composition identity is $T(g \circ f) = T(f) \circ T(g)$ (*i.e.*, the order of composition is reversed).

One basic example of a contravariant functor is the dual space construction on a category of vector spaces over some field $\mathbb{F}$. Specifically, a vector space V is sent to the space V^* of $\mathbb{F}$-linear functionals $V \to \mathbb{F}$, and if $T : V \to W$ is a linear transformation then $T^* : W^* \to V^*$ sends a linear functional $f : V \to \mathbb{F}$ to the composite $T \circ f$.

Functors have a simple but far-reaching property which is fairly easy to prove: *If the morphisms f and g are inverse to each other and T is a functor (covariant or contravariant), then $F(f)$ and $F(g)$ are also inverse to each other.*

Since functors are mathematical objects, one can speculate even further about morphisms relating functors and whether such a notion is more than a formal curiosity. It turns out that there is an extremely useful notion called a **natural transformation** of functors (where both the source and target have the same variance). Since this concept is not needed until later in the course sequence, we shall pass on discussing it here.

VII. Topological deformations and approximations

Although Units I – VI develop the formal properties of topological spaces in considerable detail and indicate that such objects are a convenient framework for studying many mathematical topics, the say little to explain the often repeated description of topology as a "rubber sheet geometry." There are at least three interrelated reasons for this description.

First of all, many homeomorphisms of topological spaces can be viewed as elastic deformations of spaces, more general than the maps of one metric space onto another which are rigidly distance-preserving (*isometries*) or mulitply distances by a fixed positive ratio r (*similarities* with ratio of similitude r). This is discussed at greater length, and with some illustrations, in the following documents:

`http://math.ucr.edu/~res/math145A-2014/intro2topA-08.pdf`

`http://math.ucr.edu/~res/math145A-2014/intro2topA-08a.pdf`

One standard and lighthearted comment along this line is that "in topology a doughnut and coffee cup are topologically equivalent." This is illustrated by the following online video:

`http://en.wikipedia.org/wiki/File:Mug_and_Torus_morph.gif`

A second way in which topological spaces formalize the notion of geometric deformation involves curves in an open subset of $\mathbb{R}^n$, where significant examples already arise when $n = 2$ or 3. When one works with a line integral of the form

$$\int_\Gamma \sum_i P_i \, dx_i$$

then standard results in multivariable calculus imply that if the integrands satisfy the compatibility condition

$$\frac{\partial P_i}{\partial x_j} \;=\; \frac{\partial P_j}{\partial x_i}$$

then the value of the line integral does not change if one deforms Γ into another curve Γ' such that the deformation does not move the endpoints. The deformation can be viewed as a 1-parameter family of curves Γ_t, where t is a real variable which may be interpreted as time (we assume that Γ and Γ' are continuous, which is enough to imply that the given line integrals exist provided the functions P_i have continuous partial derivatives). A deformation of this sort is depicted in the following video:

`http://en.wikipedia.org/wiki/File:HomotopySmall.gif`

If Γ_0 and Γ_1 are curves in $\mathbb{R}^n$, then an explicit formula for such a 1-parameter family is given geometrically by the **straight line deformation**

$$\Gamma_t(u) \;=\; (1-t)\Gamma_0(u) \;+\; t\Gamma_1(u)$$

which moves points from $\Gamma_0(u)$ to $\Gamma_1(u)$ along the closed line segment joining these two points.

The preceding is related to a more general problem involving approximations or perturbations of functions. Specifically, if X is a compact subset of $\mathbb{R}^m$ for some m and U is an open subset of $\mathbb{R}^n$ for some n (which need not be equal to m), then for each continuous function $f : X \to U$ we

shall prove that there is some $\delta > 0$ such if $|g - f| < \delta$ (take the norm given by the maximum absolute value of $f - g$), then g is a 1-parameter deformation of f in U. We shall prove these and more in Section 2, which follows the formal description of continuous 1-parameter deformations in Section 1. A deformation of this type is called a HOMOTOPY (pronounced HOME-oh-top-ee, with a light accent on the third syllable), and two maps related by a homotopy are said to be HOMOTOPIC (pronounced home-oh-TOP-ic). In Section 3 we shall develop some basic formal properties of homotopy classes of mappings, and in Section 4 we shall continue the discussion to consider spaces which are homotopically equivalent.

The notion of homotopic equivalence leads to yet another motivation for thinking of topology as a study of flexible geometrical properties. Namely, if we are given a suitably well-behaved subset $A \subset \mathbb{R}^n$, frequently we would like to approximated A by a small neighborhood U which can be deformed nicely back into A. For example, if $A \subset \mathbb{R}^2$ is the unit circle defined by $|z| = 1$, where z is viewed as a complex number, then good choices for U would be the ring shaped regions defined by the inequalities

$$1 \ - \ \varepsilon \ \ < \ \ |z| \ \ < \ \ 1 \ + \ \varepsilon$$

where $0 < \varepsilon < \frac{1}{2}$.

SOME STANDARD TOPOLOGICAL SPACES. It is useful to have consistent notation for certain spaces which arise repeatedly in geometric topology. The **standard unit n-disk** D^n is defined to be the set of all points $x \in \mathbf{R}^n$ such that $|x| \leq 1$, and the **standard n-sphere** S^n is defined to be the set of all points $x \in \mathbf{R}^{n+1}$ such that $|x| = 1$; if $n = 2$, then S^1 is just the unit circle in the plane, and if $n = 1$ then D^1 is just the closed interval $[\,-1,\, 1\,]$.

VII.1 : Homotopic mappings

(Munkres, 51–52; Hatcher, 0–1.1)

Given a continuous mapping $f : X \to Y$, a *continuous deformation* or 1*-parameter perturbation* of f is modeled mathematically by a family of continuous mappings $g_t : X \to Y$, where $0 \leq t \leq 1$ and $g_0 = f$. Intuitively we often think of t as a time parameter, but regardless of the physical motivation the formal definition is as follows:

Definition. If X and Y are topological spaces and f and g are continuous functions from X to Y, then a *homotopy* from f to g is a continuous mapping $H : X \times [0,1] \to Y$ such that

$$f(x) \ \ = \ \ H(x,0) \qquad \text{and} \qquad g(x) \ \ = \ \ H(x,q)$$

for all $x \in X$¿ The maps f and g are said to be *homotopic*, and H is called a *homotopy* from f to g. We shall often write $f \simeq g$ or $f \simeq_H g$ to indicate that f and g are homotopic or that H is a homotopy from f to g.

Examples. 1. If $Y = \mathbb{R}^n$, then the continuous mappings f and g are homotopic by a **straight line homotopy**

$$H(x,t) \ \ = \ \ (1-t)\, f(x) \ + \ t\, g(x) \ .$$

2. If X is an arbitrary topological space, let $i_t : X \to X \times [0,1]$ denote the slice inclusion $i_t(x) = (x,t)$. Then the identity map on $X \times [0,1]$ determines the *cylindrical homotopy* from i_0 to i_1.

3. Define the unit vector retraction mapping on $\mathbb{R}^n - \{\mathbf{0}\}$ to itself by the formula $\rho(\mathbf{x}) = |\mathbf{x}|^{-1} \cdot \mathbf{x}$. Then the image of the straight line map

$$H(\mathbf{x}, t) \quad = \quad (1-t) \cdot \mathbf{x} \ + \ t \cdot \rho(\mathbf{x})$$

lies in $\mathbb{R}^n - \{\mathbf{0}\}$, and therefore H defines a homotopy from the identity map on $\mathbb{R}^n - \{\mathbf{0}\}$ to ρ.

4. As a special case of the first example, it follows that every mapping $f : X \to \mathbb{R}^n$ is homotopic to the constant map whose value is $\mathbf{0}$ everywhere. More generally, a map is said to be *nullhomotopic* if it is homotopic to a constant map, and a homotopy from a continuous map $f : X \to Y$ to a constant mapping is called a *nullhomotopy*.

We shall describe many other examples in these notes.

PROPOSITION 1. *Given two topological spaces X and Y, the relation $f \simeq g$ is an equivalence relation on the set of all continuous mappings from X to Y.*

Proof. We shall establish the three defining properties separately.

The relation is reflexive. If $f : X \to Y$ then a homotopy $H : X \times [0,1] \to Y$ is given by $H(x,t) = f(x)$; this type of homotopy is often called a **constant homotopy**.

The relation is reflexive. If H is a homotopy $f \simeq g$, then the "reverse homotopy" from g to f is given by $H^*(x,t) = H(x, 1-t)$. This mapping is continuous because it is a composite of the H with the mapping $\mathrm{id}_X \times L$, where $L : [0,1] \to [0,1]$ sends t to $1-t$.

The relation is reflexive. If P is a homotopy $f \simeq g$ and Q is a homotopy $g \simeq h$, we can get from f to h by first using P to get from f to g and then using Q to get from g to h. Formally, we define the homotopy $K : X \times [0,1] \to Y$ by $K(x,t) = P(x, 2t)$ if $t \leq \frac{1}{2}$ and $K(x,t) = Q(x, 2t-1)$ if $t \geq \frac{1}{2}$. These formulas determine a well-defined continuous mapping because the two definitions agree on the points satisfying both conditions (*i.e.*, both formulas send $(x, \frac{1}{2})$ to $g(x)$ for all x).∎

It follows that the relation $\simeq$ partitions the set of continuous mappings $X \to Y$ into pairwise disjoint equivalence classes, and these are called *homotopy classes* of mappings from X to Y. The set of all such equivalence classes is usually denoted by $[X, Y]$. We shall close this section with one simple example:

PROPOSITION 2. *If Y is a topological space and θ is the standard 1–1 correspondence between points of Y and continuous mappings from the one point space $X = \{0\}$ to Y, then the homotopy classes of continuous mappings from X to Y are the same as the arc components of Y.*

Note that every mapping from a one point set into Y is continuous because the topology on the one point set must be the indiscrete topology.

Proof. The definitions imply that if $a, b \in Y$ and C_a, C_b are the associated constant mappings, then C_a is homotopic to C_b if and only if a and b can be joined by a continuous curve in Y; *i.e.*, $C_a \simeq C_b$ if and only if a and b lie in the same arc component of Y.∎

Relative homotopies

Given two topological spaces X and Y together with a subspace $A \subset X$ which is usually closed, it is often useful or even essential to have a refined notion of homotopy for continuous mappings $X \to Y$ which have the same restriction to A.

Definition. In the setting above, suppose that f and g are continuous mappings $X \to Y$ whose restrictions satisfy $f|A = g|A$; denote this common restriction by f_0. Then f and g are said to be

relatively homotopic with respect to A or *homotopic* rel A if there is a homotopy $H : X \times [0,1] \to Y$ such that $H(a,t) = f_0(a)$ for all $a \in A$; in other words, a homotopy is a relative homotopy with respect to A if and only if its restriction to A is the constant homotopy.

Most of the preceding discussion extends routinely to relative homotopies. For example, if f and g are continuous mappings such that $f|A = g|A$ and $Y = \mathbb{R}^n$, then the straight line homotopy from f to g is in fact a relative homotopy with respect to A. Similarly, the statement and proof Proposition 1 generalize in a straightforward manner to the concept of relative homotopy.

One reason for interest in relative homotopies is that one can use them to piece together homotopies defined on subsets of spaces.

PROPOSITION 3. *Suppose that f and g are continuous mappings from one space X to another space Y, and suppose that $X = X_1 \cup X_2$ where each X_i is closed in X. If $f|X_1$ and $g|X_1$ are homotopic* rel $X_1 \cap X_2$ *and $f|X_2$ and $g|X_2$ are homotopic* rel $X_1 \cap X_2$*, then f and g are homotopic* rel $X_1 \cap X_2$*.*

Sketch of proof. If H_1 and H_2 are the relative homotopies on X_1 and X_2, then they can be pieced together to construct a relative homotopy on

$$X \times [0,1] \quad = \quad X_1 \times [0,1] \ \cup \ X_2 \times [0,1]$$

because each $X_i \times [0,1]$ is closed in $X \times [0,1]$ and the two functions agree on the overlapping piece, which is $(X_1 \cap X_2) \times [0,1]$.∎

VII.2 : Some examples

(*No references to Munkres or Hatcher*)

One purpose of this section is to justify some of the statements that. in at least some reasonable cases, mappings which are sufficiently close approximations to a given mapping f_0 must be homotopic to f_0. Another goal is to prove a result which shows that, in some reasonable cases, homotopy classes of mappings take the uncountably large sets of continuous functions from one space X to another space Y and simplify them to countable, discrete sets of homotopy classes $[X, Y]$.

Homotopy and close approximations

In this discussion we shall assume that X is a compact subset of some Euclidean space $\mathbb{R}^m$ and Y is an open subset of some (possibly different) Euclidean space $\mathbb{R}^n$.

PROPOSITION 1. *Let X and Y be as above, and let $f : X \to Y$ be continuous. Then there is some $\varepsilon > 0$ such that if $g : X \to Y$ is a continuous mapping satisfying $\mathbf{d}(f, g) < \varepsilon$ (with respect to the uniform metric), then f and g are homotopic.*

Proof. We know that the image $f[X]$ is a compact subset of Y. For each $z \in f[X]$ there is some $\varepsilon_z > 0$ such that the open disk W_z of radius ε_z centered at z is contained in Y. Let ε be a Lebesgue number for the open covering of $f[X]$ by the sets W_z. It follows that if $z \in f[X]$ and $\mathbf{d}(z, y) < \varepsilon$,

then the entire closed line segment joining z to y is contained in Y. Hence if $g : X \to Y$ is a continuous mapping satisfying $\mathbf{d}(f,\, g) < \varepsilon$ (with respect to the uniform metric), then the image of the straight line homotopy $H(x,t) \;=\; t\, g(x) \;+\; (1-t)\, f(x)$ is contained in Y. But this means that f and g are homotopic as mappings from X to Y.∎

WARNING. Frequently in this course we show that two specific continuous mappings $f, g : X \to Y$ where $Y \subset \mathbb{R}^n$ are homotopic by a straight line homotopy. *It is* always essential *to verify that the images of such straight line homotopies are contained in* Y. See `straightline.pdf` for further discussion.

Countability criterion for homotopy classes

In principle, the preceding result shows that the homotopy relation is the equivalence relation generated by the binary relation $f \sim g$ if and only if for each $x \in X$ the line segment joining $f(x)$ to $g(x)$ lies entirely inside the open set Y. The next result shows that one has only countably many homotopy equivalence classes of mappings for X and Y as above.

PROPOSITION 2. *Let X and Y be as above. Then the set $[X, Y]$ of homotopy classes of continuous mappings from X to Y is a countable set.*

Proof. We shall use the Stone-Weierstrass Approximation Theorem (Rudin, *Principles of Mathematical Analysis*, Theorem 7.32, pp. 162–164) and the preceding result. More precisely, we shall prove that each continuous mapping f is homotopic to a mapping g whose coordinate functions are all given by polynomials in m variables. Since the set of all such maps is countable, it follows that the collection of all homotopy classes must also be countable.

Let $\mathcal{C}(X)$ denote the space of all continuous real valued functions on X; then the Stone-Weierstrass Theorem implies that the subalgebra $\mathcal{A}$ of all (restrictions of) polynomial functions on X is a dense subset. Given a continuous function $f : X \to Y$, denote its coordinate functions by f_j for $1 \leq j \leq n$.

By the previous result there is some $\varepsilon > 0$ such that $\mathbf{d}(f, g) < \varepsilon$ implies that f and g are homotopic, and in fact by the construction it follows that Y contains all points z such that $\mathbf{d}(\, f(x),\, z) < \varepsilon$ for some $x \in X$. By the observations of the preceding paragraph there are polynomial functions g_j such that

$$\mathbf{d}(f_j, g_j) \;\; < \;\; \frac{\varepsilon}{\sqrt{n}}$$

for each j, and it follows that the continuous function g with coordinate functions g_j maps X into U. Therefore we know that f is homotopic to a mapping g whose coordinate functions are given by polynomials.

To complete the argument, we need to show that g is homotopic to a mapping h whose coordinate functions are given by polynomials with rational coefficients. Let $\delta > 0$ be the number as in the preceding proposition, so that $\mathbf{d}(g, h) < \delta$ implies that G and h are homotopic and if $y \in \mathbb{R}^n$ satisfies $\mathbf{d}(y,\, f(x)\,) < \delta$ for some x then $y \in Y$.

Let d be the maximum degree of the coordinate functions g_j for g; then each g_j is uniquely expressible as a linear combination of monomials $\sum_\alpha \; b_{\alpha,j}\, x^\alpha$, where x^α runs through all monomials that are products of the fundamental indeterminates $x_1, \; \cdots \; , x_m$ such that $\deg(x^\alpha) \leq d$. Let A be

the number of such monomials with degree $\leq d$, let M_α be the maximum of the monomial function x^α on X, and let M be the largest of these maxima M_α (where again the degree is $\leq d$).

If we now choose rational numbers $c_{\alpha,j}$ such that

$$|c_{\alpha,j} - b_{\alpha,j}| \;\; < \;\; \frac{\delta}{A \cdot M \cdot \sqrt{n}}$$

for all α and j, and we take $h = \sum_\alpha \; c_\alpha \, x^\alpha$, then a standard estimation argument as in 205A or real analysis shows that the rational polynomial functions h_j satisfy

$$\mathbf{d}(h_j, g_j) \;\; < \;\; \frac{\delta}{\sqrt{n}}$$

which in turn implies that $\mathbf{d}(g, h) < \delta$, so that h maps X into Y and g and h are homotopic as continuous mappings from X to Y.∎

A simple variant of the preceding result is often useful. Given a topological space Y and a space U containing Y as a subspace, we shall say that Y is a *retract* of U if there exists a continuous mapping $r : U \to Y$ such that $r|Y$ is the identity. If we let j denote the inclusion of Y in U, the restriction condition can be rewritten as $r \circ j = \mathrm{id}_Y$; in other words, the mapping r is a left inverse to j. As in linear algebra, one-sided inverses to continuous maps are not unique; in topology it is customary to use the term *retraction* to denote a left inverse maps for a retract.

COROLLARY 3. *Suppose that X is a compact subset of some Euclidean space and Y is a retract of an open subset of some Euclidean space. Then the set of homotopy classes $[X, Y]$ is countable.*

Proof. Let $j : Y \to U$ be the inclusion of Y into the open subset U in some Euclidean space. Since $[X, U]$ is countable, it suffices to show that if f and g are continuous mappings from X to Y such that $j \circ f$ is homotopic to $j \circ g$, then f is homotopic to g. This is less trivial than it may seem; later on we shall see that if $i : Y \to Z$ is an arbitrary inclusion map then it is possible to have $i \circ f \simeq i \circ g$ even when f and g are not homotopic.

Suppose that $j \circ f \simeq j \circ g$, and let H be a homotopy from the first map to the second. Let $r : U \to Y$ be a retraction. Then the composite $r \circ H$ is a homotopy from $r \circ j \circ f$ to $r \circ j \circ g$. Since $r \circ j$ is the identity, the latter mappings are merely f and g respectively, and therefore $r \circ H$ defines a homotopy from f to g. By the comments in the preceding paragraph, this completes the proof.∎

Here is an important special case:

PROPOSITION 4. *If X is a compact subset of some Euclidean space and S^n is the unit sphere in $\mathbb{R}^n$ defined by $|x| = 1$, then the set of homotopy classes $[X, S^n]$ is countable.*

Proof. We only need to check that S^n is a retract of an open subset of $\mathbb{R}^{n+1}$. But if $U = \mathbb{R}^{n+1} - \{\mathbf{0}\}$, then the map $r : U \to S^n$ sending x to $|x|^{-1} \, x$ is a continuous map whose restriction to S^n is the identity.∎

Section 54 of Munkres contains a proof that $[S^1, S^1]$ is countably infinite (we shall go through the proof in the next unit of this course), so in general the cardinality estimate in the proposition is the best possible.

VII.3 : Homotopy classes of mappings

(Munkres, 51–52, 58; Hatcher, 0)

According to Exercise 51.1 on page 330 of Munkres, if X, Y, Z are topological spaces and we are given homotopic mappings $h \simeq h' : X \to Y$ and $k, k' : Y \to Z$, then $k \circ h$ and $k' \circ h'$. Therefore composition of mappings passes to a well-defined *homotopy composition* operation (sometimes also called a *pairing*) on homotopy classes

$$\text{“}\circ\text{”}\colon [X,Y] \ \times \ [Y,Z] \ \longrightarrow \ [X,Z]$$

which sends $([h],[k])$ to $[k \circ h]$. In this section we shall describe some features of this and a closely related construction for topological spaces with a small piece of additional structure.

Standard properties of functions imply that the homotopy composition operations satisfy the following identities:

(1) (*Identity conditions*) If $u \in [X,Y]$, then $u \circ [\mathrm{id}_X] = u = [\mathrm{id}_Y] \circ u$.

(2) (*Associativity*) If $t \in [X,Y]$, $u \in [Y,Z]$ and $v \in [Z,W]$, then $v \circ (u \circ t) = (v \circ u) \circ t$.

(3) (*Null conditions*) If $C_p : A \to A$ denotes the constant map whose image is $\{p\} \subset A$, then for all $p \in Y$ and $u \in [X,Y]$ we have $[C_{f(p)}] \circ u = u \circ [C_p]$, and this is the homotopy class of the constant map from X to Y whose value everywhere is $f(p)$.

In the language of category theory, the first two properties imply that one can define a *homotopy category* whose objects are topological spaces and whose morphisms are given by **homotopy classes of mappings**. Furthermore, the construction sending a space to itself and a continuous mapping to its homotopy class yields a *covariant functor* from the category of topological spaces and continuous mappings to the category of topological spaces and homotopy classes of continuous mappings.

Homotopy equivalences

For every example of a category, it is useful to understand its isomorphisms. It is not difficult to describe the isomorphisms in the homotopy category:

Definition. A continuous mapping $f : X \to Y$ is a *homotopy equivalence* if there is a (homotopy inverse) mapping $g : Y \to X$ such that $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$. Homotopy inverses for the same mapping are generally not unique, but the next result shows that they must be homotopic.

PROPOSITION 1. *If $f : X \to Y$ is a homotopy equivalence and $g, g' : Y \to X$ are homotopy inverses to f, then $g \simeq g'$.*

Proof. Consider the composite $g' \circ f \circ g$. We then have

$$[g] \ = \ [g \circ \mathrm{id}_Y] \ = \ [g \circ f \circ g'] \ = \ [\mathrm{id}_Y \circ g'] \ = \ [g']$$

which is what we wanted to prove.∎

We also have the following:

PROPOSITION 2. (Model category equivalence property) *Let $f : X \to Y$ and $g : Y \to Z$ be continuous mappings. If two of the mappings $\{f, g, g \circ f\}$ are homotopy equivalences, then so is the third.*

Proof. There are three cases, depending upon which two maps are known to be homotopy equivalences. Suppose first that f and g are homotopy equivalences, and let F and G be homotopy inverses for these respective mappings. Then one can check directly that $F \circ G$ is a homotopy inverse for $g \circ f$. Similarly, if f and $g \circ f$ are homotopy equivalences, and let F and H be homotopy inverses for these respective mappings. Then one can check directly that $f \circ H$ is a homotopy inverse for g. Finally, if g and $g \circ f$ are homotopy equivalences, and let G and H be homotopy inverses for these respective mappings. Then one can check directly that $H \circ g$ is a homotopy inverse for f.∎

Every homeomorphism is automatically a homotopy equivalence, and in order to justify the latter concept we shall give examples of homotopy equivalences which are not homeomorphisms.

PROPOSITION 3. *(i) If $K \subset \mathbb{R}^n$ is a nonempty convex subset and $p \in K$, then the constant map $C : \{p\} \to K$ is a homotopy equivalence.*

(ii) If $U \subset \mathbb{R}^{n+1}$ is the complement of $\{\mathbf{0}\}$, then the inclusion of S^n in U is a homotopy equivalence.

Proof. *(i)* Let $\rho : K \to \{p\}$ be the constant map. Then it follows that $\rho \circ C$ is the identity on $\{p\}$, and $C \circ \rho$ is homotopic to the identity by a straight line homotopy.

(ii) Let $\rho : U \to S^n$ be the map sending the nonzero vector x to $|x|^{-1} \cdot x$. Then $\rho | S^n$ is the identity, and if $j : S^n \to U$ is the inclusion, then

$$H(x,t) \;\; = \;\; \left(1 - t + \frac{t}{|x|}\right) \cdot x$$

defines a homotopy from ρ to the identity on U. Geometrically, the homotopy starts at $\rho(x)$ and moves to x along the open ray joining x to the origin.∎

Composition functors

General topological spaces A and B, it one often studies the set of homotopy classes $[A, B]$ by means of the following next two results. Here is the first one.

PROPOSITION 4. *Let X, Y, Z be topological spaces, and let $f : Y \to Z$ be continuous. Then there is a well-defined mapping of homotopy classes $f_* : [X, Y] \to [X, Z]$ such that if $v \in [X, Y]$ is represented by the function h, then $f_*(v)$ is represented by the function $f \circ h$. Furthermore, this construction has the following properties:*

(i) If F is homotopic to f, then $F_* = f_*$.

(ii) If f is the identity mapping on Y, then f_* is the identity mapping on $[X, Y]$.

(iii) If $g : Y \to Z$ is another continuous mapping, then $(g \circ f)_* \;=\; g_* \circ f_*$.

(iv) If f is a constant mapping, then so is f_*.

Proof. Throughout the discussion below, v will denote an element of $[X, Y]$ and the notation $v = [h]$ will indicated that h is a representative for the equivalence class v.

The main point needed to justify the definition of f_* is to show that the construction $f_*(v)$ does not depend upon the choice of function representing v. In other words, if h and h' are homotopic, we need to know that $f \circ h$ is homotopic to $f \circ h'$; but this follows from the Exercise 51.1 in Munkres (cited previously at the top of this section).

Property (i) also follows directly from the exercise in Munkres, and Property (ii) merely reflects the identity chain

$$v \;=\; [h] \;=\; [\mathrm{id}_Y \circ h] \;=\; (\mathrm{id}_Y)_*[h] \;=\; (\mathrm{id}_Y)_*(v) \; .$$

Finally, Property (iii) follows from another simple chain of identities:

$$(g \circ f)_*(v) \;=\; [g \circ f \circ h] \;=\; g_*([f \circ h]) \;=\; g_*\big(f_*(v) \big) \;=\; g_* \circ f_*(v) \; .$$

Finally, Property (iv) follows because $f \circ h$ is a constant map if f is a constant map.∎

There is a similar **dual result** involving composition on the other side; *i.e.*, given a map from some space W into X, there is an associated map of homotopy classes from $[X,Y]$ to $[W,Y]$. We shall merely state the results and leave the details to the reader as an exercise.

PROPOSITION 5. *Let W, X, Y be topological spaces, and let $g : W \to X$ be continuous. Then there is a well-defined mapping of homotopy classes $g^* : [X,Y] \to [W,Y]$ such that if $v \in [X,Y]$ is represented by the function h, then $g^*(v)$ is represented by the function $h \circ g$. Furthermore, this construction has the following properties:*

(i) If G is homotopic to g, then $G^* = g^*$.

(ii) If g is the identity mapping on X, then g^* is the identity mapping on $[X,Y]$.

(iii) If $f : V \to W$ is another continuous mapping, then $(f \circ g)^* \;=\; g^* \circ f^*$.

(iv) If g is is a constant mapping, then so is g^*.∎

Pointed spaces and basepoint-preserving mappings

One goal in this part of the course is to start with a topological object and to define an associated algebraic structure (such as a group) whose algebraic structure reflects the geometrical properties of the original object (one might call this an "algebraic picture of the space"). The description of certain key objects will require a refinement of the notion of topological space.

Definition. A *pointed topological space* (or space with basepoint) is a pair (X,p) consisting of a space X and a point $p \in X$ which is called the **basepoint**. A continuous mapping $f : (X,p) \to (Y,q)$ is said to be *basepoint preserving* if $f(p) = q$, and a homotopy H from one basepoint preserving mapping f to another such mapping g is said to be basepoint preserving if $H(p,t) = q$ for all $t \in [0,1]$. Identity mappings automatically preserve basepoints, and it is an elementary exercise to verify that the composite of two basepoint preserving mappings is also basepoint preserving, and this implies we have a category of pointed spaces whose morphisms are basepoint preserving continuous mappings.

As in the unpointed case, basepoint preserving homotopy of basepoint preserving mappings is an equivalence relation, so it is possible to discuss the set of basepoint preserving homotopy classes $[(X,p),(Y,q)]$. **One can then generalize everything done in this section to the category of pointed spaces and basepoint preserving continuous mappings.**∎

VII.4 : Homotopy types

(Munkres, 58; Hatcher, 0)

Two spaces X and Y are said to have the same *homotopy type* if there is a homotopy equivalence from X to Y (equivalently, there is a homotopy equivalence from Y to X. If $\mathcal{A}$ is a family of topological spaces, then the concept of homotopy type defines an equivalence relation on $\mathcal{A}$, and of course two spaces lie in the same equivalence class if and only if they have the same homotopy type. Similar considerations hold for pointed spaces and the analogous notion of *pointed homotopy type*.

The first result summarizes a fundamentally important fact about spaces with the same homotopy type.

PROPOSITION 1. *Suppose that we are given spaces X, Y, Z, and let X' and Y' be spaces such that X and X' have the same homotopy type. Then there are $1-1$ correspondences from $[X,Y]$ to $[X',Y]$ and from $[Z,X]$ to $[Z,X']$. A similar statement holds in the category of pointed spaces.*

Proof. We shall only do the unpointed case and leave the pointed case to the reader. Let $f : X \to X'$ be a homotopy equivalence, and let $g : X' \to X$ be a homotopy inverse. Then the final two propositions in the preceding section imply that the induced maps $f_* : [Z,X] \to [Z,X']$, $g_* : [Z,X'] \to [Z,X]$, $f^* : [X',Y] \to [X,Y]$ and $g^* : [X,Y] \to [X',Y]$ satisfy

$$g_* \circ f_* \;=\; \text{identity}\,, \qquad f_* \circ g_* \;=\; \text{identity}\,, \qquad g^* \circ f^* \;=\; \text{identity}\,, \qquad f^* \circ g^* \;=\; \text{identity}$$

which combine to imply that both f_* and f^* are isomorphisms.∎

COROLLARY 2. *If $K \subset \mathbb{R}^n$ and $p \in K$, then for every space X the set $[X,K]$ consists of a single point, and for every arcwise connected space X the set $[K,X]$ consists of a single point. Similarly, for every pointed space (X,x) the set $[(X,x),(K,p)]$ consists of a single point, and for every arcwise connected pointed space (X,x) the set $[(K,p),(X,x)]$ consists of a single point.*

Proof. If $K = \{p\}$ then the conclusion follows immediately, and by the Proposition one has a similar conclusion for every convex subset K because the inclusion $\{p\} \subset K$ is a (basepoint preserving) homotopy equivalence (in the construction of the straight line homotopy from the previous section, the homotopy is fixed on $\{p\}$).∎

Note. Up to this point we have not shown that there are arcwise connected spaces X and Y such that $[X,Y]$ contains more than one point. We shall do so in the next unit.

Deformation retracts

We have seen that the inclusions $\{p\} \subset K$ and $S^n \subset \mathbb{R}^{n+1} - \{\mathbf{0}\}$ are examples of homotopy equivalences which are not homeomorphisms. In fact, they both satisfy the condition in the next definition.

Definition. Let A be a subset of the topological space X, and let $i : A \to X$ denote the inclusion mapping. Then A is said to be a *strong deformation retract* of X if there is a continuous mapping $r : X \to A$ such that $r|A$ is the identity and $i \circ r$ is homotopic to the identity relative to A (*i.e.*, the homotopy is fixed on A). If the final condition is weakened to $i \circ \simeq \mathrm{id}_X$ the we shall simply say that A is a deformation retract of X.

Here are some additional examples:

1. The unit disk D^n is a strong deformation retract of $\mathbb{R}^n$. In this case the mapping r sends x to itself if $|x| \leq 1$ and to $|x|^{-1} \cdot x$ if $|x| \geq 1$. Note that these definitions agree if $|x| = 1$. A relative homotopy from $i \circ r$ is given by the straight line homotopy.

2. Strictly speaking, this is a method for constructing new examples of deformation retracts out of old ones. Suppose that A is a deformation retract of X and B is a (strong) deformation retract of Y. We claim that $A \times B$ is a (strong) deformation retract of $X \times Y$. The reverse mapping $X \times Y \to A \times B$ is just the product of the mappings $X \to A$ and $Y \to B$, and if $H : X \times [0,1] \to X$ and $K : Y \times [0,1] \to Y$ are the homotopies to the identities on X and Y respectively, then the associated homotopy to the identity on $X \times Y$ is given by $L(u,v;t) = (\, H(u,t),\, K(v,t)\,)$.

Still other (more complicated) examples are given on page 362 of Munkres.

By the preceding results of this section, if A is a strong deformation retract of X then for every space Y the map i_* defines an isomorphism from $[Y,A]$ to $[Y,X]$ and the map i^* defines an isomorphism from $[X,Y]$ to $[A,Y]$, and similar statements hold for pointed spaces, assuming that the basepoint for X lies in A.

Since an inclusion which is a deformation retract is a homotopy equivalence, it is natural to ask whether some sort of converse holds. Section 58 of Munkres describes an open subset U in $\mathbb{R}^2$ and subspaces $A, B \subset U$ such that A and B are both deformation retracts of U — and hence A and B are homotopy equivalent — but neither A nor B is homeomorphic to a deformation retract of the other. A detailed proof is given in the following document:

`http://math.ucr.edu/~res/math205B-2012/graph-embed.pdf`

Irregularly behaved examples exist even for contractible spaces. Exercise 8 on page 366 of Munkres and Exercises 6 – 7 in Hatcher give examples of contractible spaces X and points $p \in X$ such that $\{p\}$ is not a deformation retract of X (and in Exercise 7 there is no point p such that $\{p\}$ is a deformation retract of X). On a more positive note, Munkres cites a result of M. Fuchs, which states that if two spaces X and Y are homotopy equivalent, then there is some third space W containing subspaces $X' \cong X$ and $Y' \cong Y$ such that both X' and Y' are deformation retracts of W (see pages 365 and 517). Here is the reference to the paper by Fuchs:

> **M. Fuchs.** *A note on mapping cylinders.* Michigan Mathematical Journal **18** (1971), 289–290.

A more complicated example

Example 2 on page 362 of Munkres contains an illustrated explanation of why a Figure 8 Space $S^1 \vee S^1$ — namely, a union of two circles which have a single point in common — is a deformation retract of the doubly punctured plane $\mathbb{R}^2 - \{\mathbf{p}, \mathbf{q}\}$; if $\mathbf{p}$ and $\mathbf{q}$ are the points $(\pm \frac{1}{2}, 0)$, we can choose the Figure 8 explicitly to be the pair of circles defined by the equations $(x \pm \frac{1}{2})^2 + y^2 = \frac{1}{4}$. In order to indicate how one translates the explanation in Munkres into a written argument, we shall provide more details for the steps suggested by Figure 58.2 on the cited page. More precisely, these steps involve finding a chain of subsets

$$S^1 \vee S^1 \quad \subset \quad A \quad \subset \quad B \quad \subset \quad C \quad \subset \quad \mathbb{R}^2 - \{\mathbf{p}, \mathbf{q}\}$$

and showing that X is a (strong) deformation retract of Y for each pair of successive subspaces $X \subset Y$ in this chain.

We shall use the explicit model described above. The first step in Figure 58.2 suggests that $D^2 - \{\mathbf{p}, \mathbf{q}\}$ should be a deformation retract of $\mathbb{R}^2 - \{\mathbf{p}, \mathbf{q}\}$. This is fairly simple to check. Let

$$r : \mathbb{R}^2 - \{p, q\} \;\longrightarrow\; D^2 - \{\mathbf{p}, \mathbf{q}\}$$

be the map which sends $\mathbf{x}$ to itself if $|\mathbf{x}| \leq 1$ and to $|\mathbf{x}|^{-1} \cdot \mathbf{x}$ if $|\mathbf{x}| \geq 1$. If i_0 is the inclusion map of the doubly punctured disk into the doubly punctured plane, then $r \circ i_0$ is the identity, and the map $i_0 \circ r$ is homotopic to the identity by the straight line homotopy $H_0(\mathbf{x}, t) = t\mathbf{x} + (1-t)r(\mathbf{x})$ because the image of the latter lies in the doubly punctured plane.

The second step in Figure 58.2 is to show that if E is the union of two closed disks

$$\{\, \mathbf{x} \in \mathbb{R}^2 \mid |\mathbf{x} - \mathbf{p}| \leq \tfrac{1}{2} \,\} \quad \text{or} \quad \{\, x \in \mathbb{R}^2 \mid |\mathbf{x} - \mathbf{q}| \leq \tfrac{1}{2} \,\}$$

(note that the intersection of the disks is the origin) then $E - \{\mathbf{p}, \mathbf{q}\}$ is a strong deformation retract of $D^2 - \{\mathbf{p}, \mathbf{q}\}$. In this case the definition of the retraction is more complicated and we must divide into cases, depending upon whether the first coordinate of $\mathbf{x} = (u, v)$ is nonnegative or nonpositive. Specifically, if $0 \leq u \leq 1$ then let $r(u, v) = (u, v)$ if $(u - \frac{1}{2})^2 + v^2 \leq \frac{1}{4}$ and if the reverse inequality holds let

$$r(u,v) \;\; = \;\; \left(u,\; \text{sign}(v) \cdot \sqrt{\tfrac{1}{4} \;-\; \left(u - \tfrac{1}{2}\right)^2} \right) \;\; .$$

(Although the function sign(v) is discontinuous at 0, a direct check shows that the function defined by the displayed formula turns out to be continuous.) If i_1 denotes the associated inclusion, then $r \circ i_1$ is the identity and once again there is a straight line homotopy from the identity to $i_1 \circ r$ which is the constant homotopy on $E - \{\mathbf{p}, \mathbf{q}\}$.

Finally, in the last step we need to show that the Figure 8 space given by the union of the circles with equations $|\mathbf{x} - \mathbf{p}| = \frac{1}{2}$ and $|\mathbf{x} - \mathbf{q}| = \frac{1}{2}$ is a strong deformation retract of $E - \{\mathbf{p}, \mathbf{q}\}$. Once again the definition of the retraction splits into cases depending upon the sign of the first coordinate of $\mathbf{x}$. Specifically, if $\mathbf{x} = (u, v)$ satisfies $u \geq 0$, then

$$r(\mathbf{x}) \;\; = \;\; \mathbf{p} \; + \; \frac{1}{2\,|\mathbf{x} - \mathbf{p}|} \cdot (\mathbf{x} - \mathbf{p})$$

while if $u \leq 0$ then

$$r(\mathbf{x}) \;\; = \;\; \mathbf{q} \; + \; \frac{1}{2\,|\mathbf{x} - \mathbf{q}|} \cdot (\mathbf{x} - \mathbf{q}) \;\; .$$

One can then check that this mapping is well-defined, its restriction to the Figure 8 is the identity, and there is a straight line homotopy from the composite of inclusion following retraction to the identity on $E - \{\mathbf{p}, \mathbf{q}\}$.∎

A related example. Example 3 on page 362 of Munkres asserts that specific example of a Figure Theta — specifically, the union of the standard unit circle with the closed segment joining $(0, -1)$ to $(0, 1)$ — is also a strong deformation retract of $\mathbb{R}^2 - \{\mathbf{p}, \mathbf{q}\}$. In the file `secVII.04-add.pdf` the approach in Munkres and these notes is modified to prove a result of this type, in which Munkres' subspace is replaced by a topologically equivalent Figure Theta; namely, the subspace $\left([-1, 1] \times \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right) \cup \left(\{-1, 0, 1\} \times \left[-\frac{1}{2}, \frac{1}{2}\right]\right)$.

VIII. The fundamental group

If X is an arcwise connected topological space and $p \in X$, the fundamental group $\pi_1(X,p)$ is a group which provides an rough algebraic picture of the closed curves in X which start and end at p. One indication of the need for such an object arises in connection with line integrals in multivariable calculus, where one encounters the following issue:

PATH DEPENDENCE QUESTION. *Let $U \subset \mathbb{R}^n$ be an open subset, where $n = 2$ or 3, and let $P_i : U \to \mathbb{R}$ be functions with continuous partial derivatives for $1 \leq i \leq n$ which satisfy the conditions*

$$\frac{\partial P_i}{\partial x_j} \;=\; \frac{\partial P_j}{\partial x_i}$$

for all i and j such that $i \neq j$ (i.e., the integrand is closed*). Given a closed, continuous rectifiable curve $\gamma : [0,1] \to X$ which starts and ends at p, to what extent does the line integral*

$$\int_\gamma \sum_i P_i \, dx_i$$

depend upon the choice of γ?

If there is a smooth function f with continuous second partial derivatives whose gradient is equal to $(P_1, \, \cdots \, , P_n)$, and the coordinates of γ have continuous derivatives (which implies rectifiability), then the Fundamental Theorem of Calculus implies that the line integral is zero for all choices of γ, but if $U = \mathbb{R}^2 - \{\mathbf{0}\}$ then the line integral

$$\int_\gamma \frac{x\,dy \;-\; y\,dx}{x^2 \;+\; y^2}$$

is zero if γ is the clockwise circle of radius 1 centered at $(2,0)$ — which can be parametrized as $\gamma(t) = (2 - \cos 2\pi\, t, \, -\sin 2\pi\, t)$ — and it is 2π if γ is the standard counterclockwise circle of radius 1 centered at the origin. One can evaluate the first line integral using Green's Theorem and the fact that the circle in question bounds a disk in U. The second integral can be computed directly, and in view of Green's Theorem the result reflects the fact that the unit circle centered at the origin does not bound a disk in U. The definitive result on the path dependence question, which is stated in many multivariable calculus texts without proof or precise definitions, is that if the integrand is closed then the line integral only depends upon the basepoint preserving homotopy class of γ in U.

The goal of this unit is to develop algebraic formalism for working with basepoint preserving homotopy classes of closed curves in a space. In the first section we give the formal definitions and show that the set of such homotopy classes is a group with respect to a binary operation on closed curves called *concatenation* (which means stringing together); the resulting structure is the **fundamental group** (sometimes called the *Poincaré group*) of the pointed space (X,p).

It follows immediately that the fundamental group is trivial if X is a convex subset of $\mathbb{R}^n$ for some n, and in Section 2 we prove that this group is nontrivial if $X = S^1$ or $\mathbb{R}^2 - \{\mathbf{0}\}$. The method of proof is nearly as important as the result itself, and a setting for some far-reaching generalizations is presented in Section 3. In Section 4 we compute the fundamental groups of a few other basic examples, and in Section 5 we discuss simply connected spaces, which are arcwise connected spaces with trivial fundamental groups. Finally, in Section 6 we give a link to a mathematically complete treatment of the results on path dependence and homotopy which were stated earlier.

The file `knots.pdf` describes on branch of geometric topology in which fundamental groups play a central role (namely, the theory of knotted curves in $\mathbb{R}^3$).

VIII.0 : Default hypotheses

Unless stated otherwise explicitly, we shall assume that all spaces which arise in this unit are Hausdorff. This has an important implication for the constructions we shall make: *In order to apply the results of this unit to constructed examples, we need to know that the latter satisfy the default hypotheses.* Sometimes problems arise because many of our constructions will involve quotient topologies and the Hausdorff Separation Property is not preserved under taking quotients, and in such cases we must it will be verify that we obtain Hausdorff spacee unless we explicitly state something to the contrary (*e.g.*, see some of the exercises for Section 3).

Further study of the concepts introduced in Section VIII.3 (in 205B) also requires that spaces be locally arcwise connected, but that is not needed here; the files `polishcircle.pdf` and `polishcircleA.pdf` describe a subset of the plane which is not locally arcwise connected and indicate some things which go wrong if we do not assume local arcwise connectedness.

VIII.1 : Definitions and basic properties

(Munkres, 52; Hatcher, 1.1)

Given an arcwise connected space X and points $u, v \in X$, there are numerous reasons why one might wish to consider the set of all continuous curves starting at u and ending at v (formally, all continuous $\gamma : [0,1] \to X$ such that $\gamma(0) = u$ and $\gamma(0) = v$). The appropriate notion of *endpoint preserving homotopy* for such curves is easy to formulate at this point; namely, there is a homotopy $H : [0,1] \times [0,1] \to X$ such that $H(t,0)$ is the initial curve, $H(t,1)$ is the final curve, $H(0,s) = u$ for all s, and $H(1,s) = v$ for all s. The arguments in the preceding unit imply that the relation of endpoint preserving homotopy is an equivalence relation on all continuous curves joining u to v.

If $u = v$ so that we are considering closed curves, then a continuous curve $\gamma : [0,1] \to X$ joining u to itself is equivalent to a continuous basepoint preserving map from $(S^1, 1) \to (X, u)$.

> PROOF: Let $\varphi : [0,1] \to S^1$ be the map sending t to $\exp(2\pi\, i\, t) = \cos 2\pi\, t \, + \, i \sin 2\pi\, t$. Then a continuous mapping $\alpha : S^1 \to X$ yields a closed curve $gamma = \alpha \circ p$. Conversely, if we are given a closed curve γ, let ρ be the quotient space projection from $[0,1]$ to the quotient space obtained by identifying the two end points. Then general considerations involving quotient topologies imply that $\varphi = \psi \circ \rho$, where ψ is a continuous mapping from the compact space K to the Hausdorff space S^1 which is 1–1 and onto. It follows that ψ is a homeomorphism. Since γ is a closed curve, we also have a factorization $\gamma = \gamma^* \circ \rho$, and from this we obtain the map $\alpha = \gamma^* \psi^{-1}$.

One can push things further and observe that endpoint preserving homotopies of closed curves are equivalent to basepoint preserving homotopies of mappings $(S^1, 1) \to (X, u)$; for this, we need to observe that the map $\varphi \times \mathrm{id} : [0,1] \times [0,1] \to S^1 \times [0,1]$ passes to a homeomorphism from the quotient space

$$[0,1] \times [0,1] / \{(0,t) \ \equiv \ (1,t) \quad \text{for each} \quad t\}$$

to $S^1 \times [0,1]$ (once again the map from the quotient space into the codomain is continuous, 1–1 and onto with a compact domain and a Hausdorff codomain).

We shall often pass back and forth between the two characterizations of closed curves and homotopies of closed curves.

Concatenation of curves

If X is a topological space and we are given two continuous curves $\alpha : [0,1] \to X$ and $\beta : [0,1] \to X$ such that $\alpha(1) = \beta(0)$, then we can lay them end to end, first moving along α and then along β, and the resulting curve is called the *concatenation* of α and β. We shall denote this curve by $\alpha + \beta$ (other writers use a variety of symbols for this construction; see the paragraph below concerning remarks on the notation). Formally, we define $\alpha + \beta(t)$ to be $\alpha(2t)$ if $t \leq \frac{1}{2}$ and to be $\beta(2t-1)$ if $t \geq \frac{1}{2}$. The two formulas agree on the overlapping value $t = \frac{1}{2}$ because $\alpha(1) = \beta(0)$, and therefore $\alpha + \beta$ is a well-defined continuous curve. For rectifiable curves in $\mathbb{R}^n$, it is a routine exercise to verify that this concatenation operation has the following two additivity properties:

(1) It is additive with respect to length: $L(\alpha + \beta) = L(\alpha) + L(\beta)$

(2) It is additive with respect to line integrals which have a fixed integrand $\omega = \sum_i \ P_i \, dx_i$:

$$\int_{\alpha+\beta} \omega \ = \ \int_{\alpha} \omega \ + \ \int_{\beta} \omega$$

WARNING: Concatenation does **NOT** satisfy a commutativity law $\alpha + \beta = \beta + \alpha$ (for example, the concatenation in one order does not imply that the curves can be concatenated in the opposite order) or an associativity law $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$, but we shall see that the construction is associative up to homotopy.

SECOND WARNING. Some books and papers define $\alpha + \beta$ so that the first part of the curve is β and the last part is α. Each convention has advantages and disadvantages, but in any case it is good to recognize which convention is used in a particular reference in order to avoid misinterpreting some statements.

Remarks on notation. We have chosen to use a plus sign (+) for concatenation of curves because of the clear analogies between this concept and concatenation of string variables in some computer languages; since the later operation is not commutative (if `A$ = ‘‘a’’` and if `B$ = ‘‘b’’` then `A$ + B$` $\neq$ `B$ + A$`), there is a strong precedent for using a plus sign to denote such a noncommutative operation.

For our purposes the following homotopy invariance property is fundamentally important.

PROPOSITION 1. *Let X be a space with $u, v, w \in X$ (some or all of these points may be equal). If α_0 and α_1 are endpoint preserving homotopic curves from u to v and β_0 and β_1 are endpoint preserving homotopic curves from v to w, then $\alpha_0 + \beta_0$ and $\alpha_1 + \beta_1$ are endpoint preserving homotopic curves from u to w.*

In particular, this means that we have a well-defined concatenation operation on endpoint preserving homotopy classes of curves in X.

Proof. Let H be an endpoint preserving homotopy from α_0 to α_1, and let K be an endpoint preserving homotopy from β_0 to β_1. Define $L : [0,1] \times [0,1] \to X$ such that $L(t,s) = H(t,2s)$ if $s \leq \frac{1}{2}$ and $L(t,s) = K(t,2s-1)$ if $s \geq \frac{1}{2}$. These definitions overlap at the points where $s = \frac{1}{2}$, and they agree on these points because $H(t,1) = v = K(t,0)$ for all t. Therefore L defines an endpoint preserving homotopy from $\alpha_0 + \beta_0$ to $\alpha_1 + \beta_1$. ∎

We have already mentioned that concatenation of curves is not necessarily an associative operation, but the next result shows that it is associative up to homotopy. This sort of thing occurs repeatedly in the use of algebraic methods to study topological spaces. In particular, two additional results of this type will also be established here.

PROPOSITION 2. *Let X be a topological space, let $p, q, r, s \in X$, and suppose that*

α is a curve in X joining p to q,

β is a curve in X joining q to r, and

γ is a curve in X joining r to s.

Then $(\alpha + \beta) + \gamma$ is endpoint preserving homotopic to $\alpha + (\beta + \gamma)$.

Proof. If $0 \leq a < b \leq 1$, let $\lambda[a, b]$ be the unique monotonically increasing linear map from $[0, 1]$ to itself which maps 0 to a and 1 to b.

$$\lambda[a, b](t) \;\; = \;\; (1 - t)\, a \;\; + \;\; t\, b$$

Then the curve $(\alpha + \beta) + \gamma$ is given by $\alpha \circ \lambda[0, \frac{1}{4}]^{-1}$ on $[0, \frac{1}{4}]$, by $\beta \circ \lambda[\frac{1}{4}, \frac{1}{2}]^{-1}$ on $[\frac{1}{4}, \frac{1}{2}]$, and by $\gamma \circ \lambda[\frac{1}{2}, 1]^{-1}$ on $[\frac{1}{2}, 1]$. Similarly, the curve $\alpha + (\beta + \gamma)$ is given by $\alpha \circ \lambda[0, \frac{1}{2}]^{-1}$ on $[0, \frac{1}{2}]$, by $\beta \circ \lambda[\frac{1}{2}, \frac{3}{4}]^{-1}$ on $[\frac{1}{2}, \frac{3}{4}]$, and by $\gamma \circ \lambda[\frac{3}{4}, 1]^{-1}$ on $[\frac{3}{4}, 1]$.

If we compare the preceding descriptions, we see that

$$(\alpha \;+\; \beta) \;+\; \gamma \;\; = \;\; \big(\, \alpha \;+\; (\beta \;+\; \gamma)\, \big) \circ h$$

where $h : [0.1] \to [0, 1]$ is defined as follows:

On $[0, \frac{1}{4}]$ it is the increasing linear map sending that interval to $[0, \frac{1}{2}]$.

On $[\frac{1}{4}, \frac{1}{2}]$ it is the increasing linear map sending that interval to $[\frac{1}{2}, \frac{3}{4}]]$.

On $[\frac{1}{2}, 1]$ it is the increasing linear map sending that interval to $[\frac{3}{4}, 1]$.

Therefore the conclusion of the proposition will follow if h is homotopic to the identity on $[0, 1]$ leaving the endpoints fixed. Since the straight line homotopy from h to the identity has this property, the proposition follows immediately.∎

The second result on the homotopy algebra of paths is that left or right concatenation with constant maps does not change the endpoint preserving homotopy class of a curve.

PROPOSITION 3. *Let X be a topological space, let $p, q, \in X$, and suppose that α is a curve in X joining p to q. If C_p and C_q denote the constant curves at p and q respectively, then both $C_p + \alpha$ and $\alpha + C_q$ are endpoint preserving homotopic to α.*

Proof. We shall adopt the approach and notation of the preceding result.

Then the curve $C_p + \alpha$ is given by $C_p \circ \lambda[0, \frac{1}{2}]^{-1}$ on $[0, \frac{1}{2}]$ and by $\alpha \circ \lambda[\frac{1}{2}, 1]^{-1}$ on $[\frac{1}{2}, 1]$. Similarly, the curve $\alpha + C_q$ is given by $\alpha \circ \lambda[0, \frac{1}{2}]^{-1}$ on $[0, \frac{1}{2}]$ and by $\gamma \circ \lambda[\frac{1}{2}, 1]^{-1}$ on $[\frac{1}{2}, 1]$.

If we compare the preceding descriptions, we see that

$$C_p \;+\; \alpha \;\; = \;\; \alpha \circ h_L \qquad \text{and} \qquad \alpha \;+\; C_q \;\; = \;\; \alpha \circ h_R$$

where h_L is 0 on $[0, \frac{1}{2}]$ and is the increasing linear map sending $[0, \frac{1}{2}]$ to $[0, 1]$, and h_R is the increasing linear map sending $[0, \frac{1}{2}]$ to $[0, 1]$ and is 1 on $[\frac{1}{2}, 1]$. Both h_L and h_R are homotopic to the identity by straight line homotopies, and therefore the same considerations used in the previous proposition imply the present one.∎

The third result in this sequences concerns homotopy inverses. It turns out that a homotopy inverse to α is given by simply reversing the direction of α.

PROPOSITION 4. *Let X be a topological space, let $p, q, \in X$, and suppose that α is a curve in X joining p to q. Let $-\alpha$ be the curve joining q to p such that $-\alpha(t) = \alpha(1-t)$. If C_p and C_q denote the constant curves at p and q respectively, then $\alpha + (-\alpha)$ is endpoint preserving homotopic to C_p and $(-\alpha) + \alpha$ is endpoint preserving homotopic to C_q.*

Proof. We shall adopt the approach and notation of the preceding two results. Since $-\alpha$ is a curve joining q to p and $-(-\alpha) = \alpha$, it will suffice to prove that $\alpha + (-\alpha)$ is endpoint preserving homotopic to C_p (the other conclusion follows from replacing α by $-\alpha$ in the argument.

Note that the function $|2x-1|$ linearly decreases from 1 to 0 over the interval $[0, \frac{1}{2}]$ and linearly increases from 0 to 1 over $[\frac{1}{2}, 1]$, so that $1 - |2x-1|$ linearly increases from 0 to 1 over the interval $[0, \frac{1}{2}]$ and linearly increases from 1 to 0 over $[\frac{1}{2}, 1]$. This means that $\alpha + (-\alpha)(t) = \alpha(1 - |2t-1|)$, and therefore the mapping $H(t,s) \;=\; \alpha\big((1-s) \cdot (1 - |2t-1|) \big)$ defines a homotopy from $\alpha + (-\alpha)$ to C_p.∎

We can use the language of category theory to summarize the preceding four results as follows:

COROLLARY 5. *If X is a topological space then there is a category $\mathcal{G}(X)$ such that the objects are the points of X and the morphisms from a point p to a point q are the endpoint preserving homotopy classes of curves joining p to q.*∎

One further conclusion is important enough to deserve more emphasis.

THEOREM 6. *If X is a topological space and $x \in X$, then the set of endpoint preserving homotopy classes of closed curves from x to itself is a group with respect to the binary operation defined by concatenation.*

This group is called the **fundamental group** of X and denoted by $\pi_1(X, x)$; the symbol π is an acknowledgment that this group was first considered explicitly by H. Poincaré, and sometimes the group is also called the Poincaré group of (X, x). In fact, there is a sequence of *homotopy groups* $\pi_n(X, x)$ defined for each positive integer n (see Hatcher), but we shall not need the groups defined for $n \geq 2$ in this course. Given a pointed space one also defines $\pi_0(X, x)$ to be the set of arc components of X, with a basepoint given by the arc component of x.

Proof of Theorem 6. The preceding corollary implies that $\pi_1(X, x)$ is a monoid (associative binary operation from concatenation, with the class of the constant curve as the identity, pronounced MON-oid), and this monoid is a group because we have shown that if u is represented by α, then $-\alpha$ represents a class which is a multiplicative inverse to u.∎

Finally, we have the following result which includes the two preceding ones.

COROLLARY 7. *In the category described above, every morphism is an isomorphism.*

A category with this property is called a **groupoid**. One can think of a group as a groupoid which contains exactly one object. The category in this and the previous corollary is called the *fundamental groupoid* of the space X.

Proof of Corollary 7. If γ is a closed curve representing a morphism u in the category, then the class of $-\gamma$ defines an inverse to u.∎

If K consists of a single point and Y is an arbitrary topological space, then there is a unique continuous mapping from Y to K, and this simple observation implies that if $K = \{p\}$ then $\pi_1(\{p\}, p)$ must be the trivial group. Our short term goal is to obtain a weak criterion for recognizing some spaces X for which $\pi_1(X, x)$ is trivial, and in the course of doing this we shall derive some formal properties which are important in their own right.

The first of these properties can be summarized to state that that the fundamental group extends to a group valued covariant functor on pointed topological spaces.

THEOREM 8. *Let $f : (X, p) \to (Y, q)$ be a continuous basepoint preserving continuous mapping of pointed topological spaces. Then there is a homomorphism $f_* : \pi_1(X, p) \to \pi_1(Y, q)$ such that if $z \in \pi_1(X, p)$ is represented by γ, then $f_*(z)$ is represented by $f \circ \gamma$. This construction satisfies the following (functorial) properties:*

(i) If f is the identity map on X, then f_* is the identity map on $\pi_1(X, p)$.

(ii) If f is as above and $g : (Y, q) \to (Z, r)$ is another basepoint preserving continuous mapping then $(g \circ f)_* = g_* \circ f_*$.

(iii) If f is as above and $h : (X, p) \to (Y, q)$ is basepoint preserving homotopic to f, then $f_* = h_*$.

(iv) If f is a constant map then f_* is the trivial homomorphism.

Proof. The existence of the mapping f_* on homotopy classes is a special case of results in Section I.3 because $\pi_1(X, p) = [(S^1, 1), (X, p)]$, and the elementary identity

$$f \circ (\alpha_1 + \alpha_2) \quad = \quad f \circ \alpha_1 \; + \; f \circ \alpha_2$$

(whose verification is left to the reader) implies that f_* is a homomorphism with respect to the group operation. Properties (i) and (ii) also follow from the more general results in Section I.3, while Property (iii) follows because $f \simeq h$ implies $f \circ \alpha \simeq h \circ \alpha$ for all α. Finally, Property (iv) follows because $f \circ \alpha$ is a constant curve if f is a constant mapping.∎

COROLLARY 9. *If K is a convex subset of $\mathbb{R}^n$ for some n and $p \in K$, then $\pi_1(K, p)$ is trivial.*

Proof. The hypothesis implies that the identity on K is basepoint preserving homotopic to the constant map whose value everywhere is p. By the properties in the theorem, this means that the identity map on $\pi_1(K, p)$ is equal to the trivial homomorphism, and the only groups satisfying this condition are trivial groups consisting of only one element.∎

The next result implies that the fundamental group of a pointed space only contains information about the arc component of the basepoint.

PROPOSITION 10. *If X is a topological space with $p \in X$ and A is the arc component of X containing p with inclusion mapping i, then the homomorphism $i_* : \pi_1(A, p) \to \pi_1(X, p)$ is an isomorphism.*

Proof. We first prove that i_* is onto. If α is a closed curve in X which starts and ends at p, then the image of α is arcwise connected and hence lies in the arc component of X which contains p, and this arc component is A. Therefore we may write $\alpha = i \circ \alpha'$ for some closed curve α', so that $[\alpha] - i_*([\alpha'])$. Next we prove that i_* is 1–1. Given two closed curves in A which start and end at p and are endpoint preserving homotopic in H, the image of the homotopy lies in A because it is

also an arcwise connected subset containing p, and this means that the closed curves must in fact be end point preserving homotopic in A.∎

One thing we have not done up to this point is give examples of spaces whose fundamental groups are nontrivial; this will be done in the next section. However, we shall prove the following result, which provides a simple way of constructing new spaces with nontrivial fundamental groups out of old ones.

THEOREM 11. *If (X, x) and (Y, y) are pointed spaces, then the (basepoint preserving) coordinate projections $p_X : (X \times Y, (x, y)\,) \to (X, x)$ and $p_Y : (X \times Y, (x, y)\,) \to (Y, y)$ define a group isomorphism*

$$(p_{X*}, p_{Y*}) : \pi_1(X \times Y, (x, y)\,) \;\longrightarrow\; \pi_1(X, x) \times \pi_1(Y, y)\ .$$

Recall that the group operation on a direct product of two groups is defined coordinatewise.

Proof. Since a continuous mapping into a product space is completely determined by its coordinate projections onto the factors and likewise for a homotopy of continuous mappings, it follows that the map (p_{X*}, p_{Y*}) defines an isomorphism of pointed sets. This map is also a group homomorphism because the theorem on fundamental groups implies that p_{X*} and p_{Y*} are group homomorphisms.∎

The final result of this section states that the fundamental group of an arbitrary topological space is completely determined by the fundamental groups of its compact subsets and the homomorphisms of such groups that are induced by inclusions of one compact subset in a larger one.

PROPOSITION 12. (Compact Generation Property) *Let (X, x) be a pointed space.*

(i) If $u \in \pi_1(X, x)$, then there is a compact subset $K \subset X$ such that $x \in K$ and u lies in the image of the homomorphism from $\pi_1(K, x)$ to $\pi_1(X, x)$ induced by inclusion.

(ii) If $K_j \subset X$ is a compact subset of X containing x for $j = 1, 2$ and we have classes $u_j \in \pi_1(K_j, x)$ with the same image in $\pi_1(X, x)$, then there is some compact subset of $L \subset X$ such that L contains $K_1 \cup K_2$ and the classes u_j have the same image in $\pi_1(L, x)$.

Sometimes this result is summarized by the statement, "the fundamental group of a space is compactly supported."

Proof. *(i)* If $u \in \pi_1(X, x)$ is represented by the closed curve α then we can take K to be the image of α, which is a compact set containing x.

(ii) If $u_j \in \pi_1(K_j, x)$, suppose that u_j is represented by α_j, let $H : [0, 1] \times [0, 1] \to X$ be a homotopy from $i_1 \circ \alpha_1$ to $i_2 \circ \alpha_2$, and let L_0 denote the image of H. Then L_0 is compact, and $L - K_1 \cup K_2 \cup L_0$ is a compact set such that the images of α_1 and α_2 are endpoint preserving homotopic in L.∎

There is a useful variant of the preceding result in one of the exercises.

VIII.2 : Important special cases

(Munkres, 53–54, 65, 73; Hatcher, 1.1)

At this point we urgently need is to verify that $\pi_1(X, x)$ is nontrivial for some examples of pointed spaces (X, x), and in fact we want to give simple examples of spaces which have already arisen in these notes. In particular, we need to show that the fundamental group of the unit circle S^1 is nontrivial, and the objective of this section is to do so. This will require new techniques, and these techniques turn out to yield fundamentally important results in geometry and topology; for example, the machinery plays an indispensable role in the theory of functions of one complex variable.

Here is the main result:

THEOREM 1. *The group $\pi_1(S^1, 1)$ is infinite cyclic, and a generator is given by the simple counterclockwise parametrization of the circle: $\alpha(t) = \exp 2\pi\, i\, t$ (where $0 \leq t \leq 1$)*

We shall follow the standard approach of deriving Theorem 1 from the following two results:

THEOREM 2. (Path Lifting Property) *Let $p : \mathbb{R} \to S^1$ be the map*

$$p(t) \quad = \quad \exp(2\pi\, i\, t)$$

which wraps the real line around the circle in the counterclockwise direction. If $\gamma : [0, 1] \to S^1$ is a continuous curve and $t_0 \in \mathbb{R}$ satisfies $p(t_0) = \gamma(0)$, then there is a unique lifting *$\widetilde{\gamma} : [0, 1] \to \mathbb{R}$ such that $p \circ \widetilde{\gamma} = \gamma$ and $\widetilde{\gamma}(0) = t_0$.*

Geometrically speaking, this theorem implies that if γ is a curve in $\mathbb{C} - \{0\}$, then we can unambiguously define the angle that the ray $[0 \to \gamma(t)$ makes with the x-axis if we are given the angle's value $\gamma(0)$ (in general, the possibilities for this value have the form $\theta_0 + 2k\pi$, where $\theta \in [0, 2\pi)$ and k is an integer, so the initial value corresponds to a specific choice of k.

The second theorem we need is analogous to the first one.

THEOREM 3. (Covering Homotopy Property) *Suppose that γ_0 and γ_1 are homotopic continuous curves on S^1, let $h : [0, 1] \times [0, 1] \to S^1$ be a homotopy from γ_0 to γ_1, and choose $t_0 \in \mathbb{R}$ such that $p(t_0) = h(0, 0)$. Then there is a unique* covering homotopy *$H : [0, 1] \times [0, 1] \to \mathbb{R}$ such that $p \circ H = h$ and $H(0, 0) = t_0$.*

The proofs of these results are based in turn upon the following property of the mapping p:

THEOREM 4. *Let p be as above, and let $z_0 \in S^1$. Then there is an open neighborhood U of z_0 in S^1 with the following properties:*

(i) The inverse image $p^{-1}[U]$ is a union of disjoint intervals of the form $(a + k, b + k)$ where k runs through all the integers and $b - a < 1$ (the latter implies that the intervals in question are pairwise disjoint for different choices of k).

(ii) The restriction of p to each interval $(a + k, b + k)$ defines a homeomorphism from the latter to U.

We are now ready to work backwards, first proving Theorem 4, then using this result to derive Theorems 2 and 3, and finally using the latter to derive Theorem 1.

Proof of Theorem 4. We shall first prove the result when $z_0 = 1 \in S^1$, and then we shall use the rotational symmetry of the circle and the exponential law $p(t_1 + t_2) \ = \ p(t_1) \cdot p(t_2)$ to retrieve the general case.

THE CASE $z_0 = 1$. — Let $U \subset S^1$ be the set of all points $z = x + yi \in S^1$ such that $x > 0$, so that U consists of all z such that the angle which $[0 \to z$ makes with the x-axis is between $\pm \frac{1}{2}\,\pi$. We then have

$$p^{-1}[U] \quad = \quad \bigcup_{n \in \mathbb{Z}} \, \left(n - \tfrac{1}{4}, n + \tfrac{1}{4}\right)$$

and hence the inverse image is a disjoint union of open intervals, and on each one the restriction of p is continuous, 1–1 and onto. In order to verify that the restrictions map these intervals homeomorphically onto their common image, we need to show that their inverses are also continuous. This follows because the inverses are given explicitly by the functions

$$\sigma_n(x + yi) \quad = \quad n \ + \ \frac{\arcsin y}{2\pi}$$

where we take the range of the inverse sine function to be the interval from $-\frac{1}{2}\,\pi$ to $\frac{1}{2}\,\pi$.

THE GENERAL CASE — Suppose that $z_0 \neq 1$ and choose $t_0 \in (0,1)$ such that $p(t_0) = z_0$ (there is always a unique t_0 for which this is true). Let $M(z_0)$ denote the homeomorphism of S^1 to itself sending z to $z_0 z$ for all $z \in S^1$, and let $A(t_0)$ be the homeomorphism of $\mathbb{R}$ with itself sending t to $t_0 + t$. Then the inverse image of the open neighborhood $V = M(z_0)[U]$ of z_0 is a union of pairwise disjoint intervals

$$\left(t_0 + n - \tfrac{1}{4}, t_0 + n + \tfrac{1}{4}\right)$$

and p maps each interval to V by a mapping which is continuous, 1–1 and onto. The continuity of the inverse mappings then follows because the inverses are given by the composites $A(t_0) \circ \sigma_n \circ M(z_0)^{-1}$, where σ_n is given as in the proof of the special case.∎

Proof of Theorem 2. Let $V(z_0) \subset S^1$ be the open neighborhood of z_0 defined in the preceding proof, let $\mathcal{W}$ be the open covering of $[0,1]$ given by the inverse images $\gamma^{-1}[V(z_0)]$. By the Lebesgue Covering Lemma (see `gentopnotes2014.pdf` near the end of Section III.1) there is some $\eta > 0$ such that if $|u - v| < \eta$ then u and v lie inside an open subset taken from the covering $\mathcal{W}$. Therefore, if n is a positive integer such that $\frac{1}{n} < \eta$, then γ maps each interval

$$\left[\frac{j-1}{n}, \, \frac{j}{n}\right] \qquad (1 \leq j \leq n)$$

into an element of $\mathcal{W}$.

We shall now use Theorem 4 to construct liftings of γ over the intervals

$$\left[\frac{0}{n}, \, \frac{j}{n}\right]$$

inductively with respect to j, where $1 \leq j \leq n$.

Suppose first that $j = 1$. By the preceding discussion there is some open semicircular arc V_1 such that γ maps $[0, \frac{1}{n}]$ into V_1. There is a unique component U_1 of $p*{-}1[V_1]$ containing t_0; let σ_1 be defined on V_1 as the composite of $(p|U_1)^{-1}$ with the inclusion of $U_1 \subset \mathbb{R}$, and define $\widetilde{\gamma}$ on $[0, \frac{1}{n}]$ to be the composite of γ with σ_1. — Now assume that $\widetilde{\gamma}$ has been constructed over the subinterval from

0 to $(j-1)/n$, and let t_{j-1} be the ending point of the lifted curve. We need to lift $\gamma|[(j-1)/n,\ j/n]$ to $\mathbb{R}$ such that the value at the left hand endpoint is equal to $t_{j-1} = \widetilde{\gamma}(\,(j-1)/n)$, and this can be done by an argument similar to the one employed when $j = 1$. Specifically, the image of $\gamma|[(j-1)/n,\ j/n]$ is contained in some semicircular arc V_j, and there is a unique component U_j of $p * -1[V_j]$ containing t_{j-1}; let σ_j be defined on V_j as the composite of $(p|U_j)^{-1}$ with the inclusion of $U_j \subset \mathbb{R}$, and define $\widetilde{\gamma}$ on $[(j-1)/n,\ j/n]$ to be the composite of γ with σ_j. By construction the value of of this lifting at the left hand endpoint agrees with the value of the known lifting at its right hand endpoint, so we can piece the two definitions together and obtain the desired lifting of γ over $[0, j/n]$. This completes the inductive argument, and hence it also completes the proof that a lifting of γ exists (and satisfies the given initial value condition).

To conclude the proof, we need to show that the uniqueness of a lifting with the desired properties. Suppose that α and β are such liftings, so that $\alpha(0) = \beta(0) = t_0$. Let E be the (nonempty) set of points where $\alpha = \beta$. Since the codomains of the curves are a metric (hence Hausdorff) space, we can use a result in Section II.4 (in the subsection *Products and the Hausdorff Separation Property*) to conclude that E is closed in $[0, 1]$.

On the other hand, we also claim that E is open; since $[0, 1]$ is connected and E is nonempty, this will imply that $E = [0, 1]$ and hence $\alpha = \beta$. Suppose that $\alpha(x) = \beta(x) = y$, and let J be an open connected subset of $[0, 1]$ such that $x \in J$ and γ maps J into some semicircular arc V_α. Since the continuous image of an arcwise connected space is arcwise connected, it follows that both α and β must map J into the open subset $U \subset p^{-1}[V_\alpha]$ such that $y \in U$ and p maps U homeomorphically onto V_α. Since the restriction of p to U is 1–1 and we know that $p \circ \alpha = p \circ \beta$, it follows that the restrictions of α and β to the open subset J must be equal. Therefore, we have shown that if $x \in E$ then x has an open neighborhood J in $[0, 1]$ such that $J \subset E$, and this implies that E is an open subset of $[0, 1]$.∎

Proof of Theorem 3. This argument is similar to the preceding one, the main differences being that we must cut the square $[0, 1] \times [0, 1]$ up into small squares such that each is mapped into some semicircular arc and we need to describe a linear ordering on the squares in order to carry out the inductive construction.

Let h be the homotopy given in the statement of the theorem. As in the proof of Theorem 2, there is some positive integer n such that h maps each of the squares

$$C_{i,j} \quad = \quad \left[\frac{i-1}{n}, \frac{i}{n}\right] \times \left[\frac{j-1}{n}, \frac{j}{n}\right] \qquad (1 \quad \leq \quad i, j \quad \leq n)$$

into a semicircular arc. Order these squares lexicographically.

As in the proof of Theorem 2, we shall construct the lifting H inductively over the union B_k of the first k squares, where $1 \leq k \leq n^2$. If $k = 1$ then the image of the first square lies in some semicircular arc, and since we know the desired value at $(0, 0)$ we can construct the lifting over the first square by the same method employed in the proof of Theorem 2. Assume now that the lifting has been constructed on the union of the first $k-1$ squares, and let $C_{i,j}$ be the k^{th} square. Once again, we know that h maps the square into a semicircular arc V_{k-1}. If D_k is the intersection of $C_{i,j}$ with B_{k-1} for $k \geq 2$, then we know that D_k consists of one or two edges in $C_{i,j}$ which contain the lower left hand corner point

$$\left(\frac{i-1}{n}, \frac{j-1}{n}\right)$$

and therefore we need to construct a lifting of $h|C_{i,j}$ which agrees with H on D_k. But D_k is arcwise connected, and consequently there is a unique component U_{k-1} of $p^{-1}[V_{k-1}]$ such that H maps D_k

into U_{k-1}. As before, we can extend $H|D_k$ to a continuous mapping from $C_{i,j}$ to U_{k-1}, and this yields an extension of the lifting H to all of B_k. This completes the inductive argument and hence also the existence proof. The uniqueness proof is similar to that of Theorem 2, the main difference being that J will now be an open subset of the square $[0,1] \times [0,1]$ which contains a given point (x_1, x_2).∎

Proof of Theorem 1. Let γ be a closed curve in S^1 such that $\gamma(0) = 1 = \gamma(1)$. Theorem 2 implies that there is a unique lifting $\widetilde{\gamma}$ of γ which starts at 0; since $p\widetilde{\gamma}(1) = \gamma(1) = 1$, it follows that $\widetilde{\gamma}(1)$ must be an integer, which we shall call the **degree** of γ and denote by d_γ. We shall prove that the map sending γ to its degree yields an isomorphism from $\pi_1(S^1, 1)$ to $\mathbb{Z}$.

The first step is to show that the degree only depends upon the class of γ in the fundamental group. Suppose that α and β are closed curves in S^1 which start and end at 1 and are endpoint preserving homotopic, and let h be an endpoint preserving homotopy from α to β. By Theorem 3 there is a unique covering homotopy H such that $h = p \circ H$ and $H(0,0) = 0$. The crucial step is to describe the restrictions of H to the four edges of the square.

> The restriction of H to $[0,1] \times \{0\}$ is a lifting of α which starts at 0, so it is the unique lifting given by Theorem 2.
>
> The restrictions of H to the vertical edges $\{0\} \times [0,1]$ and $\{1\} \times [0,1]$ are liftings of the constant curve whose image is $\{1\}$, and hence these images lie in $\mathbb{Z}$. Since the images of the liftings are arcwise connected, it follows that these lifings must be constant. In particular, $H(0,s) = 1$ for all s and $H(s,1) = d_\alpha$ for all s.
>
> Since $H(0,1) = 1$, the restriction of H to $[0,1] \times \{1\}$ is a lifting of β which starts at 0, so it is the unique lifting given by Theorem 2.

Since the ending point of the lifting for β is d_β by definition, the preceding observations imply that $d_\beta = d_\alpha$.

Next, we shall prove that the mapping $d : \pi_1(S^1, 1) \to \mathbb{Z}$ determined by the argument up to this point onto, and γ is basepoint preserving homotopic to a constant mapping if and only if $d_\gamma = 0$. The proof that d is onto can be done by an explicit construction. Let $\theta(t) = n \cdot t$ where n is an integer; then θ is the unique lifting of $\gamma = p \circ \theta$ starting at 0, and therefore we have $d_{p\theta} = n$. We must now verify the that d is nullhomotopic if and only if $d_\gamma = 0$. If the latter holds, then $\widetilde{\gamma}$ is a closed curve in $\mathbb{R}$ and at the fundamental group level we have $[\gamma] = p_*([\widetilde{\gamma}])$. Since the domain of p_* is $\pi_1(\mathbb{R}, 0)$ and the latter is trivial by convexity, it follows that $[\gamma]$ is the trivial element of $\pi_1(S^1, 1)$. Conversely, if $[\gamma]$ is trivial then $d_\gamma = d_C$ where C is the constant curve in S^1. Since the unique lifting of C is the constant curve in $\mathbb{R}$, it follows that $d_C = 0$ and hence $d_\gamma = 0$.

Finally, we need to prove that d is a homomorphism of groups. Suppose that we are given basepoint preserving closed curves α and β in S^1; let A and B be the unique liftings of these curves which start at 0. For each positive integer d let T_d be the translation map on $\mathbb{R}$ which sends t to $t + d$. Then it follows that the unique lifting of $\alpha + \beta$ starting at 0 must be the curve

$$\eta \ + \ A \ + \ T_a \circ B$$

where $a = d_\alpha$. By construction we have

$$d(\alpha + \beta) \ = \ T_a \circ B(1) \ = \ T_a(d_\beta) \ = \ d_\alpha + d_\beta$$

and this proves that the mapping d is a homomorphism. By the preceding paragraphs it also follows that d must define a group isomorphism.∎

More examples of nontrivial fundamental groups

Now that we have one space with a nontrivial fundamental group, we can construct an infinite family of examples such that their fundamental groups are pairwise nonisomorphic.

THEOREM 5. *For each positive integer n there is an arcwise connected space T^n whose fundamental group is isomorphic to $\mathbb{Z}^n$.*

Proof. The example is the $n-$**torus** T^n, which is the product of n copies of S^1 with itself. By Theorem VIII.1.11 and the results of this section we have

$$\pi_1\left(T^n, (1, \cdots, 1)\right) \quad \cong \quad \mathbb{Z}^n \ . \blacksquare$$

Clearly one can ask the following more general question:

Realization problem for fundamental groups. *Given a group G, is there a pointed space (X, x) such that $\pi_1(X, x) \cong G$?*

The answer is that all groups can be realized, and a construction is outlined in Example 1B.7 of Hatcher. Furthermore, if $\varphi : G \to H$ is a group homomorphism, then there are arcwise connected pointed spaces (X, x) and (Y, y) with canonical isomorphisms $\pi_1(X, x) \to G$ and $\pi_1(Y, y) \to H$ and also a continuous mapping $f : (X, x) \to (Y, y)$ such that f_* corresponds to φ. The proof involves things not covered in this course, but for the record one reference is Proposition 1.B.9 on page 30 of Hatcher.

Before proceeding, we note one simple but useful property of the degree, which follows from the fact that d is a group homomorphism: $d_{-\alpha} = -d_\alpha$

Change of basepoints

It is natural to ask what happens to the fundamental group of an arcwise connected space X if we change the basepoint from $p \in X$ to $q \in X$. The answer contains some good news and some bad news. We begin with the good news:

THEOREM 6. *Let X be an arcwise connected space with $p.q \in X$. Then the groups $\pi_1(X, p)$ and $\pi_1(X, q)$ are isomorphic.*

NOTATIONAL CONVENTION. Although concatenation of curves is not associative, it is associative up to endpoint preserving homotopy, and therefore the endpoint preserving homotopy class of an iterated concatenation for a curve sequence $\gamma_1, \cdots, \gamma_r$ is independent of the bracketry of the terms; for example, if we have four curves then some possibilities for inserting brackets are

$$((12)(34)) \ , \quad (((12)3)4) \ , ((1(23))4) \ , \quad (1(2(34))) \ , \quad (1((23)4))$$

and the homotopy classes corresponding to the various classes are the same. Therefore we shall often write

$$[\gamma_1 \ + \ \cdots \ + \ \gamma_r]$$

to denote the common endpoint preserving homotopy classes of these concatenations.

Proof of Theorem 6. Let γ be a curve joining p to q, and define $\gamma^* : \pi_1(X, p) \to \pi_1(X, q)$ such that the class of a closed curve α is sent to the class of the concatenation $(-\gamma) + (\alpha + \gamma)$. One can check directly that this does not depend upon the endpoint preserving homotopy classes of γ

or α. Furthermore, it follows that γ^* is the identity if γ is the constant curve whose value is p, γ^* is a group homomorphism, and an inverse map to γ^* is given by $(-\gamma)^*$. The first assertion follows because

$$C^*([\alpha]) \quad = \quad [C + \alpha + C] \quad = \quad [\alpha]$$

and the second follows because

$$\gamma^*([\alpha][\beta]) \quad = \quad \gamma^*([\alpha + \beta]) \quad = \quad [(-\gamma) + \alpha + \beta + \gamma] \quad =$$

$$[(-\gamma) + \alpha + \text{constant} + \beta + \gamma] \quad = \quad [(-\gamma) + \alpha + \gamma + (-\gamma) + \gamma + \beta + \gamma] \quad = \quad \gamma^*([\alpha]) \cdot \gamma^*([\beta]) \ .$$

In order to prove the third assertion, we need an identity to describe maps of the form $(\gamma_1 + \gamma_2)^*$:

$$(\gamma_1 + \gamma_2)^*([\alpha]) \quad = \quad [-(\gamma_1 + \gamma_2) + \alpha + \gamma_1 + \gamma_2] \quad = \quad [(-\gamma_2) + (-\gamma_1) + \alpha + \gamma_1 + \gamma_2] \quad =$$

$$\gamma_2^*[(-\gamma_1) + \alpha + \gamma_1] \quad = \quad \gamma_2^* \left(\gamma_1^*([\alpha])\right) \quad = \quad \gamma_2^* \circ \gamma_1^*([\alpha])$$

If $\gamma_2 = -\gamma_1$ it follows that γ_2^* and γ_1^* are inverse to each other, which is what we wanted to prove.∎

The bad news related to Theorem 6 is that the isomorphism frequently depends upon the endpoint preserving homotopy class of the choice of path joining p to q. This can be seen immediately in the case where $p = q$, in which case we have

$$\gamma^*([\alpha]) \quad = \quad [\gamma]^{-1} \cdot [\alpha] \cdot [\gamma]$$

and if $\pi_1(S^1, 1)$ is nonabelian then one can choose γ so that this mapping is not the identity. More generally, if the fundamental group is not abelian then one exercise for this section shows that the isomorphisms γ^* in Theorem 6 will depend upon the choice of γ.

Free homotopy classes of closed curves

Another question about the role of basepoints concerns the need or usefuless for them in the discussion of homotopy classes of closed curves. If X is an arcwise connected space and $p \in X$, then there is a canonical "forgetful mapping" from $\pi_1(X, p)$ to the set $[S^1, X]$ of *free homotopy classes* of closed curves in which one makes no assumptions about preserving basepoints for the closed curves or for homotopies of closed curves. The result is fairly simple to state algebraically.

THEOREM 7. *If X is an arcwise connected space and $p \in X$, then the forgetful mapping ε from $\pi_1(X, p)$ to $[S^1, X]$ is onto, and two elements $a, b \in \pi_1(X, p)$ go to the same class in $[S^1, X]$ if and only if they are conjugate in the fundamental group; in other words, there is some $c \in \pi_1(X, p)$ such that $b = c^{-1}ac$.*

This gives one important reason for dealing with point spaces. The fundamental group functor is easier to manipulate algebraically than the functor defined by free homotopy classes of closed curves, and one can retrieve the latter from the fundamental group very directly.

Our proof of Theorem 7 will use the following observation:

PROPOSITION 8. *Let X be a topological space, let p and q be points of X, and let $\alpha : [0, 1] \to X$ be a continuous curve such that $\alpha(0) = p$ and $\alpha(1) = q$. Then there is a homotopy $H : [0, 1] \times [0, 1] \to X$ such that $H(1, s) = H(t, 1) = q$ for all s and t, with $H(t, 0) = \alpha(t)$ and $H(0, s) = \alpha(s)$ for all s and t.*

One can rewrite the conclusion to state that α is homotopic to the constant curve C_q by a homotopy which is constant on the right hand endpoint and looks like α on the left hand endpoint.

Proof of Proposition 8. The idea is that at each times x the curve H_s is determined by the restriction of α to $[1-s, 1]$. One way of doing this is to set $H(t,s) = \alpha(s+t)$ if $s+t \leq 1$ and $H(t,s) = q$ if $s+q \geq 1$. This mapping is continuous because the two definitions agree on the overlapping set where there is a potential ambiguity (if $s+t=1$ then both formulas yield the value q), and it is routine to check that H has all the properties which appear in the statement of the proposition.∎

Proof of Theorem 7. The argument has three parts:

(*i*) Verifying that the forgetful map ε is onto.

(*ii*) Showing that conjugate elements have the same image under ε.

(*iii*) Showing that $\varepsilon(a) = \varepsilon(b)$ implies that a and b are conjugate.

Proof of (i): Let $\gamma : S^1 \to X$ be a closed curve with $\gamma(1) = q$, and let let α be a curve joining p to q. We shall prove that $\beta = (-\alpha) + \gamma) + \alpha$ is freely homotopic to $\beta = (C_q + \gamma) + C_q$; since the latter is endpoint preserving homotopic to γ, it will follow that β and α are freely homotopic. Furthermore, it also follows that the class $[\gamma]$ of γ in the free homotopy classes is equal to the class $[\beta]$, which by construction lies in the image of $\pi_1(X,p)$.

We can use Proposition 8 to construct a free homotopy K as follows:

On $[0, \frac{1}{4}] \times [0,1]$, we have $K(t,s) = H(4t,s)$.

On $[\frac{1}{4}, \frac{1}{2}] \times [0,1]$, we have $K(t,s) = \gamma(4t-1)$.

On $[\frac{1}{2}, 1] \times [0,1]$, we have $K(t,s) = H(2t-1,s)$.

On the overlapping sets where $t = \frac{1}{4}$ or $t = \frac{1}{3}$ each of the three defining formulas yields the same value q, and therefore K is a well-defined homotopy between the given two mappings. Furthermore, it passes to a homotopy on $S^1 \times [0,1]$ because $K(0,s) = K(1,s) = \alpha(s)$ for all s. This completes the proof of the fist statement given above.

Proof of (ii): This follows from the same considerations as in (i), specialized to the case $q = p$. If this holds, then the arguments in (i) imply that for all choices of γ and α, the fundamental group classes $[\gamma]$ and $[\alpha]^{-1}[\gamma][\alpha]$ map to the same element of $[S^1, X]$.

Proof of (iii): Let α and β be basepoint preserving closed curves, and suppose that $H : S^1 \times [0,1] \to X$ be a free homotopy from α to β. If $p : [0,1] \to S^1$ is the usual mapping $p(t) = \exp 2\pi\, i\, t$ let $K = H \circ (p \times \mathrm{id}_{[0,1]})$.

Next, let $L : S^1 \to [0,1] \times [0,1]$ be the broken line closed curve which goes around the boundary of the square once in the counterclockwise sense. If ω is the closed curve $K|\{0\} \times [0,1]$, then the homotopy class $[K \circ L]$ in $\pi_1(X,p)$ is equal to the class of the concatenation

$$[\alpha \ + \ \omega \ + \ (-\beta) \ + \ (-\omega)] \ .$$

If we denote the classes of α, β, ω in $\pi_1(X,p)$ by a, b, c respectively, then we may rewrite the preceding observation in the form $[K \circ L] = acb{-}1c^{-1}$. On the other hand, the factorization of $K \circ L$ as a composite

$$S^1 \ \longrightarrow \ [0,1] \times [0,1] \ \longrightarrow \ S^1 \times [0,1] \ \longrightarrow \ X$$

shows that $[K \circ L]$ lies in the image of $\pi_1([0,1] \times [0,1])$, which is trivial. If we combine the preceding two sentences, we see that $acb{-}1c^{-1} = 1$, which is equivalent to the conjugacy equations $a = cbc^{-1}$ and $b = c^{-1}ac$.∎

The same methods yield the following generalization which will be needed in Section 6:

PROPOSITION 9. *In the setting of the preceding results, let α and β be freely homotopic closed curves in X such that the endpoints of α are p and the endpoints of β are q. Then there is a curve ω joining u to v such that the endpoint preserving homotopy class of the iterated concatenation*

$$[\alpha \ + \ \omega \ + \ (-\beta) \ + \ (-\omega)]$$

is the trivial element of $\pi_1(X, u)$.∎

Finally, the following special case of Theorem 7 is important enough to be stated formally:

COROLLARY 10. *If X is arcwise connected and $p \in X$, the forgetful map from $\pi_1(X, p)$ to $[S^1, X]$ is 1– and onto if and only if $\pi_1(X, p)$ is abelian. In particular, the forgetful map from $\pi_1(S^1, 1)$ to $[S^1, S^1]$ is an isomorphism.*

Proof. By the Theorem 7, the mapping is 1–1 if and only if each conjugacy class in the fundamental group consists of a single element. This is true if and only if for all a and c in the fundamental group we have $a = c^{-1}ac$, and the latter holds if and only if the fundamental group is abelian, proving the first assertion. The second assertion follows because $\pi_1(S^1, 1)$ is abelian.∎

VIII.3 : Covering spaces

(Munkres, 53; Hatcher, 1.3)

> Like most good theorems, its conclusion has become a definition.
>
> J. L. Kelley, *General Topology*, p. 135.

The techniques which we developed to compute $\pi_1(S^1, 1)$ have far-reaching generalizations which play important roles in topology, geometry and other subjects (*e.g.*, complex variables). In this section we shall formulate an abstract setting in which the conclusion of Theorem 2.4 is valid and the proofs of Theorems 2.2 and 2.3 can be generalized with only minor changes. The discussion relies heavily on the default hypotheses of Section I.0; namely, all spaces are assumed to be Hausdorff and locally arcwise connected unless there is an explicit statement that they are not.

The main concepts

We begin by abstracting the conclusion to Theorem 2.4.

Definition. Let $p : X \to Y$ be a continuous onto mapping of topological spaces. Then p is said to be a *covering map* or *covering space projection* if for each $y \in Y$ there is an open neighborhood V of Y such that the following condition holds:

> *V is* **evenly covered:** There is a discrete set F and a homeomorphism $\varphi : V \times F \to p^{-1}[V]$ such that $p \circ \varphi(v, f) = v$ for all $v \in V$.

Before proceeding, we shall explain the significance of being evenly covered. If W_f is the image of $V \times \{f\}$, then the sets W_f are open and pairwise disjoint subsets of X, and p maps each subset

W_f homeomorphically onto V; an explicit continuous inverse is given by $\sigma_f(v) = \varphi(v, f)$. The sets W_f are often called the *sheets* of the open covering over V.

Frequently one uses variations on the term "covering space projection" to describe the situation described above. Examples include the phrases *X is a covering space of Y*, *$p : X \to Y$ is a covering space*, and so on.

Primary example. Theorem 2.4 implies that the mapping $p : \mathbb{R} \to S^1$ defined by $p(t) = \exp(2\pi\, i\, t)$ is a covering space projection.

It follows immediately that if $p : X \to Y$ is a covering space projection, then X and Y have the same properties locally; for example, if Y is locally connected or locally arcwise connected, then so is X, and the converse also holds.

IMPORTANT. Even though the word "covering" appears in the phrases "open covering" and "covering spaces," there is no direct connection between the two usages; however, in practice this ambiguity usually does not cause any difficulties.

Before proving the main results, we shall state and prove two properties of covering space projections that are important to know but are sometimes not mentioned in textbooks.

PROPOSITION 1. *Suppose that $p : E \to B$ is a covering space projection, and assume that the default hypotheses are satisfied.*

(i) If B_0 is an arc component of B and $E_0 = p^{-1}[B_0]$, then the restriction of p defines a covering space projection $p_0 : E_0 \to B_0$.

(ii) A covering space projection is an open mapping.

(iii) If B is nonempty and (arcwise) connected, then the cardinality of $p^{-1}[\{b\}]$ is the same for all $b \in B$.

Proof. *(i)* If $x \in B_0$, then x has an arcwise connected and evenly open neighborhood U_x in B, and this neighborhood is contained in the maximal arcwise connected subset containing x; namely, B_0. Therefore $p_{-1}[U_x]$ is contained in E_0.∎

(ii) Let Ω be an open subset of E, and let $\mathcal{U}$ be an open covering of B by arcwise connected sets which are evenly covered. Then for each $U_\alpha \in \mathcal{U}$ we know that $p^{-1}[U_\alpha]$ is a union of open subsets $V_{\alpha,\beta}$ such that the restriction of p to each of these subsets is 1–1, continuous and open. If $\mathcal{V}$ is the family of all $V_{\alpha,\beta}$'s, then $\mathcal{V}$ is an open covering of E. Therefore we have $\Omega = \cup_{\alpha,\beta}\, \Omega \cap V_{\alpha,\beta}$, so that

$$p[\Omega] \;=\; \bigcup_{\alpha,\beta}\, p[\Omega \cap V_{\alpha,\beta}]\ .$$

Since each intersection $\Omega \cap V_{\alpha,\beta}$ is an open subset and the restriction of p to each open set $V_{\alpha,\beta}$, it follows that $p[\Omega]$ must be open.∎

(iii) Define an equivalence relation on B by $x \sim y$ if and only if $p^{-1}[\{x\}]$ and $p^{-1}[\{y\}]$ have the same cardinality. By the definition of covering space, the equivalence classes of this relation are all open, and since they are pairwise disjoint it follows that they are also closed. Since B is connected, there can only be a single equivalence class.∎

The Path Lifting and Covering Homotopy Properties

The generalizations of Theorems 2.2 and 2.3 are easy to formulate.

THEOREM 2. (Path Lifting Property) *Let $p : E \to B$ be a covering space projection. If $\gamma : [0,1] \to B$ is a continuous curve and $e_0 \in E$ satisfies $p(e_0) = \gamma(0)$, then there is a unique* lifting *$\widetilde{\gamma} : [0,1] \to E$ such that $p \circ \widetilde{\gamma} = \gamma$ and $\widetilde{\gamma}(0) = e_0$.*

THEOREM 3. (Covering Homotopy Property) *In the setting of the previous theorem, suppose that γ_0 and γ_1 are homotopic continuous curves on B, let $h : [0,1] \times [0,1] \to B$ be a homotopy from γ_0 to γ_1, and choose $e_0 \in E$ such that $p(t_0) = h(0,0)$. Then there is a unique* covering homotopy *$H : [0,1] \times [0,1] \to E$ such that $p \circ H = h$ and $H(0,0) = e_0$.*

The proofs of Theorems 2.2 and 2.3 go through almost word for word, the main difference being the need to replace the phrase "semicircular arc" with the more general phrase "evenly covered open subset." A second modification is that expressions like $\cup_n \ (a+n, b+n)$ must be replaced by $\cup_f \ W_f$. The details of making these changes are left to the reader.∎

Examples of covering spaces

We shall now give some additional examples of covering space projections.

THE REAL PROJECTIVE PLANE. This space is denoted by $\mathbb{RP}^2$, and two equivalent constructions of it as a quotient space are described in Unit V and the file `gentopexercises2014.pdf`; the reasons for considering this quotient are discussed at great length in the online document

`http://math.ucr.edu/~res/progeom/pg-all.pdf`

and accompanying files in the directory `http://math.ucr.edu/~res/progeom`. For our purposes here, it is convenient to think of $\mathbb{RP}^2$ as the quotient of S^2 by the equivalence relation which identifies $\mathbf{x}$ and $\mathbf{y}$ if and only if one of these unit vectors is ± 1 times the other. We claim that the quotient map from S^2 to $\mathbb{RP}^2$ is a covering space projection. It is possible to prove this directly (see Theorem 60.3 on page 372 of Munkres), but it will ultimately be more efficient to prove a general result which will yield larger classes of examples.

Unfortunately, we shall need to introduce some notation. The notion of a group action on a topological space is defined in Exercise 31.8 on page 199 of Munkres. For our purposes it will suffice to take a group and to view it as topological groups with respect to the discrete topology. If G is such a group and X is a topological space, the group action itself is given by a continuous mapping $\Phi : G \times X \to X$, with $\Phi(g,x)$ usually abbreviated to $g \cdot x$ or gx, such that $1 \cdot x = x$ for all x and $(gh) \cdot x \ = \ g \cdot (h \cdot x)$ for all g, h and x. One can then define an equivalence relation on X by stipulating that $y \sim x$ if and only if $y = g \cdot x$ for some $g \in G$, and the quotient space with respect to this relation is called the *orbit space* of the group action and written X/G. By the cited exercise in Munkres, this space is Hausdorff if X is.

If we are given a group action as above and A is a subset of X, then for a given $g \in G$ it is customary to let $g \cdot A$ (the translate of A by g) be the set $\Phi[\,\{g\} \times A\,]$; this is the set of all points expressible as $g \cdot a$ for the fixed g and some $a \in A$.

Definition. We shall say that a group action Φ as above is a **free action** (or G acts freely) if for every $x \in X$ the only solution to the equation $g \cdot x = x$ is the trivial solutions for which $g = 1$. — If $X = S^2$ as above and G is the order two subgroup $\{\pm 1\}$ of the real numbers (with respect to multiplication), then scalar multiplication defines a free action of G on S^2, and the quotient space is just $\mathbb{RP}^2$. Of course, there are also similar examples for which 2 is replaced by an arbitrary positive integer n, and in this case the quotient space $S^n/\{\pm 1\}$ is called *real projective n-space*.

The next result implies that the orbit space projections $S^n \to \mathbb{RP}^n$ are covering space projections.

THEOREM 4. *Let G be a finite group which acts freely on the Hausdorff topological space X, and let $\pi : X \to X/G$ denote the orbit space projection. Then π is a covering space projection.*

Proof. Let $x \in X$ be arbitrary, and let $g \neq 1$ in G. Then there are open neighborhoods $U_0(g)$ of x and $V_0(g)$ of $g \cdot x$ that are disjoint. If we let $W(g) = U(g) \cap g^{-1} \cdot V(g)$ is another open set containing x, while $g \cdot W(g)$ is an open set containing $g \cdot x$, and we have $W(g) \cap g \cdot W(g) = \emptyset$. Let

$$W \;=\; \bigcap_{h \neq 1} W(h)$$

so that W is an open set containing x.

We claim that if $g_1 \neq g_2$, then $g_1 \cdot W \cap g_2 \cdot W \;=\; \emptyset$. If we know this, then it will follow immediately that $\pi[W]$ is an open set in X/G whose inverse image is the open subset of X given by $\cup_g \, g \cdot W$. This and the definition of the quotient topology imply that $\pi[W]$ is an evenly covered open neighborhood of x, and therefore it will follow that π is a covering space projection.

Thus it remains to prove the statement in the first sentence of the preceding paragraph. Note first that it will suffice to prove this in the special case where $g_1 = 1$; assuming we know this, in the general case we then have

$$g_1 \cdot W \cap g_2 \cdot W \;=\; g_1 \left(W \cap (g_1^{-1} g_2) \cdot W \right)$$

and the coefficient of g_1 on the right hand side is empty by the special case when $g_1 = 1$ and the fact that $g_1 \neq g_2$ implies $1 \neq g_1^{-1} \cdot g_2$. — But if $g \neq 1$ then we have $W \cap g \cdot W \subset W(g) \cap g \cdot W(g)$, and we know that the latter is empty by construction. Therefore $W \cap g \cdot W = \emptyset$, and as noted before this completes the proof.∎

ANOTHER EXAMPLE. Define an action of the finite group $\mathbb{Z}_2$ on the torus $T^2 = S^1 \times S^1$ so that the nontrivial element $T \in \mathbb{Z}_2$ satisfies $T \cdot (z, w) = (-z, \overline{w})$ where S^1 is viewed as the set of unit complex numbers and the bar denotes conjugation. This is a free action because $T(z, w) = (z, w)$ would imply $z = -z$, and we know this is impossible over the complex numbers. In this case the quotient space is the **Klein bottle**.

STILL MORE EXAMPLES. Let $\mathbb{D}$ denote either the complex numbers or the quaternions, let d be the dimension of $\mathbb{D}$ as a real vector space, and let G be a finite subgroup of the group S^{dm-1} of elements of $\mathbb{D}$ with unit length. For example, if $\mathbb{D} = \mathbb{C}$ (the complex numbers), then G can be a cyclic group of arbitrary order, while if $\mathbb{D}$ is the quaternions then one also has some nonabelian possibilities, most notably the quaternion group of order 8 whose elements are given by ± 1, $\pm \mathbf{i}$, $\pm \mathbf{j}$, and $\pm \mathbf{k}$. If $\mathbb{D} = \mathbb{C}$ and $m > 1$, then the quotient spaces $S^{2m-1}/\mathbb{Z}_q$ (for $q > 1$) are the objects known as (simple) **lens spaces** (sometimes the case $q = 2$ is excluded because that quotient is the previously described real projective space); the reason for assuming $m > 1$ is that the corresponding quotient space for S^1 is homeomorphic to S^1. If $\mathbb{D}$ is the quaternions, G is the nonabelian quaternion group of order 8 described above and $m = 1$, then the space S^3/G is called the 3-*dimensional quaternionic space form* associated to the group G.

In all examples of this type, for each point y in X/G the inverse image of $\{y\}$ in X consists of $|G|$ points, where $|G|$ is the order of G.

We shall compute the fundamental groups of the preceding examples in Section 5 of this unit.

Composites of covering space projections

Exercise 54.4 (Munkres, p. 341) shows that under suitable restrictions the composite of two covering space projections is also a covering space projections. However, in general this is not necessarily true, and here is an example: Let X be a connected, locally arcwise connected space, and let $p : E \to X$ be a connected covering map that is nontrivial (not a homeomorphism). Let $Y = X \times X \times X \times ...$ be the countably infinite product of X with itself (with the product topology), let E^n denote the product of n copies of E with itself, and for an arbitrary space Y let $Y_n = E^n \times Y$. Define $p_n : Y_n \to Y$ by

$$(e_1, ... e_n; x_1, x_2, ...) \to (p(e_1), ... p(e_n); x_1, x_2, ...)$$

Then each map p_n is a covering map. Next, let $\widetilde{Z} = \amalg_{n \geq 1} Y_n$ and let Z be the countably infinite sum of Y with itself. Let $q = \amalg_n p_n : \widetilde{Z} \to Z$ and let $r : Z \to Y$ be the obvious projection. Then r and q are covering maps but that the composite rq is not a covering map.

The proof is only moderately difficult, but it is also a bit lengthy and requires input involving the product topology for infinite products, and therefore the proof will be left as an exercise [*Hint:* It suffices to show that basic open sets in the product topology are not evenly covered]. Another property involving covering spaces and composites appears in the first additional exercise for this section, and in a subsequent section we shall give yet another exercise with a sufficient condition under which the composite of two covering space projections is also a covering space projection.

VIII.4 : Fundamental groups of spheres

(Munkres, 59)

The methods developed thus far also allow us to compute the fundamental groups of higher dimensional spheres.

THEOREM 1. *If $n \geq 2$ and $x \in S^n$, then $\pi_1(S^n, x)$ is trivial.*

Before proving this result, we shall prove a general result about the fundamental group of an open subset in $\mathbb{R}^k$ where k is a positive integer. Some of the techniques employed in the proof of this result are also used in the proof of Theorem 1.

THEOREM 2. *If U is an open connected subset of $\mathbb{R}^n$ and $x \in U$ has rational coordinates, then $\pi_1(U, x)$ is finite or countably infinite.*

The bound on the cardinality is the best possible general estimate because S^1 is a strong deformation retract of $\mathbb{R}^2 - \{\mathbf{0}\}$, and hence the fundamental group of the latter is infinite cyclic.

The condition on the basepoint is not a serious restriction for two reasons: Since $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$ the open set U always contains a point with rational coordinates, and by the results of Section

VIII.2 the isomorphism type of the fundamental group of an arcwise connected space is the same for all choices of basepoint. At the end of this section we shall use Theorem 2 to describe a compact subset of $\mathbb{R}^2$ which is arcwise connected but is not homotopy equivalent to an open subset of some $\mathbb{R}^n$.

Proof of Theorem 2. It suffices to consider **proper** open subsets of $\mathbb{R}^n$ because the latter is convex and hence we already know its fundamental group is trivial.

Let γ be a continuous closed curve in U. Since the image of γ is compact we know that the continuous function $\mathbf{d}(\gamma(t), \mathbb{R}^n - U)$ has a minimum value, which we shall call h. Thus for every value of t the open disk of radius h centered at $\gamma(t)$ is contained in U, and by uniform continuity we know that there is some positive integer m such that γ maps every subinterval of length $1/m$ into a disk of radius $h/3$. For each integer j such that $0 \leq j \leq m$ choose $p_j \in U \cap \mathbb{Q}^n$ such that $p_0 = p_m = u$ and $|p_j - \gamma(j/m)| < h/3$ otherwise (recall that $\gamma(0) = \gamma(1) = u$).

We claim that for each $j > 0$ the closed line segment joining p_{j-1} to p_j is contained in the neighborhood of radius h centered at $\gamma(j/m)$; by convexity it suffices to prove that the two endpoints are contained in this neighborhood. By construction we have $|p_j - \gamma(j/m)| < h/3$ and we also have $|p_{j-1} - \gamma(j/m)| \leq |p_{j-1} - \gamma((j-1)/m)| + |\gamma((j-1)/m) + \gamma(j/m)| < \frac{1}{3}\, h + \frac{1}{3}\, h < h$, so the endpoints do lie in this neighborhood as claimed.

Let β be the broken line curve defined on $[(j-1)/m, j/m]$ by taking the closed line segment from p_{j-1} to p_j, which is a curve from $[0,1]$ into U, and composing it with the unique increasing linear function $[(j-1)/m, j/m] \to [0,1]$. This yields a well-defined continuous function because the values at the endpoints were chosen to be compatible. Since the restrictions of γ and β to $[(j-1)/m, j/m]$ map into a convex open disk of radius h in U, it follows that these restrictions are homotopic by a straight line homotopy which lies inside of the h-neighborhood and hence also lies inside U. Once again, these homotopies are consistently defined on the overlapping pieces $\{k/m\} \times [0,1]$, and since $u = \gamma(0) = \gamma(1) = p_0 = p_m$ it follows that the straight line homotopy defines a basepoint preserving homotopy in U from γ to the broken line curve formed by connecting the points p_j.

The preceding discussion shows that each element of $\pi_1(U, u)$ is represented by a broken line curve which is determined by the sequence of points $u = p_0, p_1, \ \cdots \ p_{m-1}, p_m = u$, where each p_j lies in U and has rational coordinates. There are only countably many such sequences, so we have shown that the elements of $\pi_1(U, u)$ determined by these broken lines are the entire fundamental group, and hence it follows that $\pi_1(U, u)$ is either finite or countably infinite.∎

Proof of Theorem 1. Take the standard unit vector $\mathbf{e}_1 = (1, 0, \ \cdots, 0)$ as a basepoint for S^n and $\mathbb{R}^{n+1} - \{\mathbf{0}\}$, so that the inclusion $j : S^n \subset \mathbb{R}^{n+1} - \{\mathbf{0}\}$ is basepoint preserving. Since S^1 is a strong deformation retract of $\mathbb{R}^{n+1} - \{\mathbf{0}\}$, the induced mapping in fundamental groups j_* is an isomorphism, and likewise for the standard homotopy inverse ρ which goes in the opposite direction. Let γ be a closed curve in S^n which starts and ends at $\mathbf{e}_1$. Using the method employed to prove Theorem 2, choose a basepoint preserving broken line curve β in $\mathbb{R}^{n+1} - \{\mathbf{0}\}$ which is basepoint preserving homotopic to $j \circ \gamma$ and consists of line segments indexed by the sequence of points p_j with rational coordinates.

We claim that there is a line through the origin which is disjoint from the image of β. To see this, note that the image of β is contained in the union of the linear subspaces W_j, where W_j is spanned by p_j and p_{j-1}. Each of these is a proper subspace of dimension $\leq 2 < n + 1$, and the assertion in the first sentence of this paragraph follows because this finite union of proper linear subspaces is a proper subset of $\mathbb{R}^{n+1} - \{\mathbf{0}\}$. If we compose everything with ρ we see that the closed curves $\gamma = \rho \circ j \circ \gamma$ and $\rho \circ \beta$ represent the same element of $\pi_1(S^n, \mathbf{e}_1)$. However, since $n \geq 2$ we also

know that the image of $\rho \circ \beta$ is a proper subset of S^n. It follows that the homotopy class $[\rho \circ \beta]$ lies in the image of the homomorphism $\pi_1(S^n - \{\text{point}\}, \mathbf{e}_1) \to \pi_1(S^n, \mathbf{e}_1)$ induced by inclusion. Since $S^n - \{\text{point}\}$ is homeomorphic to $\mathbb{R}^n$, the domain of this homomorphism is trivial, and therefore the class $[\gamma] = [\rho \circ \beta]$ must also be trivial. Since $[\gamma]$ was arbitrary, this means that $\pi_1(S^n, \mathbf{e}_1)$ must be trivial.∎

Spaces with large fundamental groups

Compact subsets of the plane with uncountable fundamental groups. In one of the exercises we noted that the Cantor Set does not have the homotopy type of an open subset in some Euclidean space because it has uncountably many components. Since we also know that the fundamental group of an open subset in some Euclidean space is countable, it is natural to ask if one can also construct a compact subset of, say, the plane whose fundamental group is uncountable. An example of this sort (the shrinking wedge of circles, sometimes also known as the *Hawaiian earring* or *Hawaiian necklace*) is described in Chapter 1 of Hatcher (see Example 1.25 on pp. 49–50).

Open subsets of the plane whose fundamental groups are not finitely generated. We have seen that the fundamental group of $\mathbb{R}^2 - \{\mathbf{0}\}$ is infinite cyclic and that every finitely generated free abelian group can be realized as the fundamental group of a product of circles, and similarly we can realize every such group as the fundamental group of a some open subset in some Euclidean space. An example of an open subset in $\mathbb{R}^2$ with an infinitely generated fundamental group is given by taking the complement U of the set of all negative integers $\{-1, -2, \cdots\}$.

Here is one way of proving that $\pi_1(U, 1)$ is not finitely generated: View U as a subset of the complex plane, and let α_k denote the closed curve in U given by the counterclockwise circle with radius $(2k+1)/4$ and center $(3-2k)/4$, so that α_k meets the real axis at the points 1 and $(1-4k)/2$. Therefore each α_k defines an element a_k of the fundamental group of $\pi_1(U, 1)$. For each positive integer j let U_j denote the complement of $\{-j\}$, and let φ_j denote the map of fundamental groups determined by the inclusion of U in U_j followed by an isomorphism from $\pi_1(U_j, 1)$ to the integers $\mathbf{Z}$. It follows that $\varphi_j(a_k)$ is a generator if $j \leq k$ and trivial if $j > k$; this is true because the point $-j$ is inside the circle α_k if $j \leq k$ and outside the circle if $j > k$ (see the additional exercise for Section 56).

We may combine the preceding homomorphisms to define a homomorphism Φ from $\pi_1(U, 1)$ to a product $\prod^\infty \mathbb{Z}$ of countably infinitely many copies of $\mathbb{Z}$ (with addition defined coordinatewise); specifically, for all j, the j^{th} coordinate of Φ is φ_j. Since the homomorphic image of a finitely generated group is finitely generated, it will suffice to show that the image of Φ is not finitely generated. This final step is purely algebraic, and it depends upon the standard structure theorems for finitely generated abelian groups.

It is a routine exercise to check that if $\{G_\alpha\}$ is a family of groups such that the only elements of finite order are the identities, then the product $\prod_\alpha G_\alpha$ also has this property. Similarly, the product is abelian if each factor is abelian. Both of these properties carry over to subgroups, and in particular they apply to the image of Φ. Therefore, if the image of Φ is finitely generated, the structure theorem for finitely generated abelian groups implies that it is a direct sum of infinite cyclic groups, and as such it has some finite rank r. For each positive integer m, let H_m denote the subgroup generated by the classes $\varphi_j(a_j)$ for $j \leq m$. Then H_m is the subgroup of $\prod^\infty \mathbb{Z}$ consisting of all elements for which the p^{th} coordinate is zero for all $p > m$. This group has rank m;

it follows that for each positive integer m the image of Φ contains free abelian subgroups of rank m. Since the image of Φ has no nontrivial elements of finite order, our finite generation assumption on the image of Φ implies that the latter is free abelian and has some fixed finite rank q. General considerations involving finitely generated abelian groups imply that every subgroup of the image of Φ is also free abelian and has rank at most q. This contradicts the final sentence in the preceding paragraph. The ultimate source of this contradiction is our assumption that the image of Φ is finitely generated. As noted before, this suffices to show that the fundamental group of U is not finitely generated.∎

In fact, it is possible to show that every countably generated group can be realized as the fundamental group of an open subset in $\mathbb{R}^n$ if $n \geq 4$, but proving this result would require methods and results which are outside the scope of the course.

VIII.5 : Simply connected spaces

(Munkres, 53)

Definition. A (nonempty) topological space X is said to be *simply connected* if it is arcwise connected and $\pi_1(X, x_0)$ is trivial for some $x_0 \in X$. By the result on change of basepoint, the latter condition is equivalent to the triviality of $\pi_1(X, x)$ for all $x \in X$.

PROPOSITION 1. *Let X be a nonempty topological space. Then the following hold:*

(i) The space X is arcwise connected if and only if every continuous mapping $S^0 \to X$ extends to a continuous mapping from S^1 to X.

(ii) The space X is simply connected if and only if every continous mapping $S^1 \to X$ extends to a continuous mapping from D^2 to X.

For the record, we note that our proof of this result does not depend upon the default hypothesis in Section I.0.

Proof. (i) This follows from arcwise connectedness and the fact that the linear mapping $f(t) = \frac{1}{2}(t+1)$ defines a homeomorphism from $D^1 = [-1, 1]$ to $[0, 1]$ such that the respective left and right hand endpoints correspond to each other.∎

(ii) $(\Rightarrow)$ Suppose that the condition on extending continuous mappings is satisfied, let $\gamma : S^1 \to X$ be a basepoint preserving closed curve, and let σ be an extension of γ to D^2. Then $[\gamma] = \gamma_*([\mathrm{id}(S^1)]) = \sigma_* \circ j^*([\mathrm{id}(S^1)])$, where $j : S^1 \to D^2$ is inclusion. Since D^2 is a convex subset of $\mathbb{R}^2$, its fundamental group is trivial, and therefore $j*^*$ is the trivial homomorphism. If we combine the statements in the preceding two sentences, we see that $[\gamma]$ must be trivial; since $[\gamma]$ was arbitrary, this means $\pi_1(X, x)$ must be trivial.

$(\Leftarrow)$ Suppose now that the fundamental group of X is trivial and $f : S^1 \to X$ is continuous. Choose $x = f(1)$ to be the basepoint for X. Then there is a continuous mapping $H : S^1 \times [0, 1] \to X$ such that $H|S^1 \times \{0\} = f$ and H is constant on $\{1\} \times [0, 1] \cup S^1 \times \{1\}$. The latter condition implies that H has a factorization

$$S^1 \times [0, 1] \longrightarrow K \longrightarrow X$$

where K is the quotient space of $S^1 \times [0, 1]$ whose equivalence classes are one point subsets of $S^1 \times [0, 1)$ and $S^1 \times \{1\}$ (*i.e.*, the closed subset $S^1 \times \{1\}$ is collapsed to a point). The conclusion

will follow if there is a homeomorphism from K to D^2 such that $S^1 \times \{0\}$ corresponds to $S^1 \subset D^2$. To construct this homomorphism, consider the mapping from $S^1 \times [0,1]$ to D^2 sending $\mathbf{v}$ to $(1-|\mathbf{v}|)\cdot\mathbf{v}$. This passes to a continuous mapping from K to D^2 which is 1–1 and onto, and this mapping also sends $S^1 \times \{0\}$ to $S^1 \subset D^2$. By construction, K is compact (it is the quotient of a compact space), and since D^2 is Hausdorff it follows that the map $K \to D^2$ must be a homeomorphism. As indicated earlier in the paragraph, this suffices to complete the proof.∎

Simply connected spaces play an important role in the study of covering spaces. Our first objective is to show that a simply connected space has no nontrivial covering spaces (see Corollary 4 below); this will be a consequence of the next two results:

THEOREM 2. *If (E, e_0) and (B, b_0) are connected spaces satisfying the Default Hypothesis and $p : (E, e_0) \to (B, b_0)$ is a covering space projection, then $p_* : \pi_1(E, e_0) \to \pi_1(B, b_0)$ is a 1–1 homomorphism.*

IMPORTANT REMARK. The preceding discussion shows that there is generally no relation between the injectivity or surjectivity of a continuous map f and the analogous properties for the associated homomorphism f_* of fundamental groups. In particular, the covering space example shows that f_* can be 1–1 but not onto when f is onto but not 1–1, and the example $(S^1, 1) \subset (D^2, 1)$ shows that f_* can be onto but not 1–1 when f is 1–1 but not onto.

Proof of Proposition 2. Assume that γ is a basepoint preserving closed curve in E and the corresponding element $[\gamma] \in \pi_1(E, e_0)$ satisfies $p_*([\gamma]) = 1$. Then there is a homotopy $h : [0,1] \times [0,1] \to B$ such that $p \circ \gamma = h|[0,1] \times \{0\}$ and h is constant on the other three edges of the square $[0,1] \times [0,1]$. Let $H : [0,1] \times [0,1] \to E$ be the unique covering homotopy such that $H|[0,1] \times \{0\} = \gamma$. If $F \subset E$ is the inverse image of b_0, then $e_0 \in F$ and F is discrete in the subspace topology. Since $h = p \circ H$, it follows that H maps the union B of the three edges

$$\{0\} \times [0,1] \ \cup \ [0,1] \times \{0\} \ \cup \ \{1\} \times [0,1]$$

into F; in fact, since B is connected and F is discrete, it follows that $H|B$ is constant and hence $H(s,t) = H(0,0) = e_0$ at all points of B. The latter in turn implies that H is a basepoint preserving homotopy from γ to a constant map, which means that $[\gamma]$ is trivial in $\pi_1(E, e_0)$.∎

It turns out that the cosets of $p_* [\pi_1(E, e_0)]$ in $\pi_1(B, b_0)$ also have a topological interpretation which is analogous to the proof that the fundamental group of the circle is $\mathbb{Z}$.

THEOREM 3. *Let $p : (E, e_0) \to (B, b_0)$ be a covering space projection, where E and B are connected spaces satisfying the Default Hypothesis, and let $F \subset E$ denote the inverse image $p^{-1}[\{b_0\}]$ (the* fiber *of the basepoint; note that $e_0 \in F$). Then there is a* right action *of the group $\pi_1(B, b_0)$ on F; in other words, a set-theoretic mapping $\Phi : F \times \pi_1(B, b_0) \to F$ such that*

$$\Phi(x, 1) \ = \ x \ , \quad \Phi(x, gh) \ = \ \Phi\left(\Phi(x, g), h\right)$$

for all $x \in F$ and $g, h \in \pi_1(B, b_0)$. Every point in F has the form $\Phi(e_0, g)$ for some g, and $\Phi(e_0, h) = e_0$ if and only if h lies in the image of $\pi_1(E, e_0)$.

Usually we simplify notation and denote $\Phi(x, g)$ by $x \cdot g$ or simply xg. One advantage of this convention is that we can rewrite the defining identities as $x \cdot 1 = x$ and $(xg)h = x(gh)$.

COROLLARY 4. *Let $p : (E, e_0) \to (B, b_0)$ be a covering space projection, where E and B are connected spaces satisfying the Default Hypothesis, and suppose that B is simply connected. Then p is a homeomorphism.*

Proof of Corollary 4. (Assuming Theorem 3.) — By Theorem 3 we know that the fiber F is in 1–1 correspondence with the cosets of $p_*\,[\pi_1(E, e_0)]$ in $\pi_1(B, b_0)$, and this set of cosets has exactly one element since B has a trivial fundamental group. Every covering space projection is continuous and open, and since $F = \{e_0\}$ we can use Proposition 3.1 to conclude that for each $b \in B$ the subset $p^{-1}[\{b\}]$ also consists of a single point. Therefore our continuous, open and onto mapping p is also 1–1, and this means that p must be a homeomorphism.∎

Proof of Theorem 3. We construct Φ using the Path Lifting and Covering Homotopy Properties. If $x \in F$ and $g \in \pi_1(B, b_0)$, choose a basepoint preserving closed curve $\gamma : S^1 \to B$ representing g, let $\widetilde{\gamma}$ be the unique lifting to a curve $[0,1] \to E$ such that $\widetilde{\gamma}(0) = x$ given by the Path Lifting Property, and provisionally take xg to be $\widetilde{\gamma}(1)$. In order for this provisional construction to be well-defined, we need to check that we get the same end point for all choices of γ, so suppose that h is a basepoint preserving homotopy from γ to another closed curve β, and let H be a covering homotopy for h which starts at γ.

Then $p \circ H = h$ is constant on the vertical edges $\{0,1\} \times [0,1]$ of the square $[0,1] \times [0,1]$, and as in the proof of Theorem 2 it follows that H must also be constant on these edges. Since $H(0,0) = e_0$ this means that $H(0,1) = e_0$, which implies that $H|[0,1] \times \{1\}$ is the unique lifting $\widetilde{\beta}$ for β which starts at e_0. Similarly, we also have $H(1,s) = H(1,0) = \widetilde{\gamma}(1)$ for all s, so that $\widetilde{\beta}(1) = H(1,1) = \widetilde{\gamma}(1)$ and hence $xg = x \cdot g$ is well-defined.

The right group action identities follow immediately (for $x\,1 = x$ the lifting of the constant curve at b_0 is the constant curve at x, while for $(xg)h = x(gh)$ one chooses representatives α and β for g and h, so that $\alpha + \beta$ is a representative for gh and the appropriate lifting of the latter curve ends at $(xg) \cdot h$, which verifies the property in question). Given x in F, there is a curve θ from e_0 to x in E, and if g represents the class of $p \circ \theta$ in the fundamental group, then by construction we have $x = e_0 g$. Finally, if $e_0 g = e_0$ then in the previous notation we have $\widetilde{\gamma}(1) = e_0$ so that $\widetilde{\gamma}$ is a closed curve representing some $g' \in \pi_1(E, e_0)$ such that $p_*(g') = g$; conversely, if g satisfies this condition and we take β to be a closed curve in E_0 representing g', then β is the associated lifting of $p \circ \gamma$, and this implies that $e_0 \cdot p_*(g') = e_0$ as required.∎

Applications to computing fundamental groups

We shall use the preceding results to determine the fundamental groups of some orbit spaces X/G considered in Section 3; more precisely, we shall do this for some examples where X is simply connected.

THEOREM 5. *Suppose that G is a finite group which acts freely on the simply connected space X. Then the fundamental group of X/G is* anti-isomorphic *to G; in other words there is a 1–1 correspondence φ between groups such that $\varphi(ab) = \varphi(b) \cdot \varphi(a)$.*

The reason that the groups are anti-isomorphic is that the action of G on X is a left action but the action of $\pi_1(X/G)$ on the inverse image of a basepoint is a right action. One can equally well say that the groups are isomorphic because every group G is anti-isomorphic to itself by the map sending $g \in G$ to g^{-1}.

Examples. Since S^n is simply connected if $n \geq 2$, it follows that $\pi_1(\mathbb{RP}^n, p) \cong \mathbb{Z}_2$ for all $n \geq 2$, and if $m, q \geq 2$ then the fundamental group of the lens space $S^{2m-1}/\mathbb{Z}_q$ is isomorphic to $\mathbb{Z}_q$. If we combine this with the formula for the fundamental group of a product and the structure theorem for finitely generated abelian groups (which states that every such group is a product of cyclic groups), we have shown how to realize every finitely generated abelian group as the fundamental group of some compact metric space.

Theorem 5 also shows that at least one nonabelian finite group can be realized. Specifically, if G is the nonabelian quaternion group of order 8 described above and $m = 1$, then it follows that the space S^3/G has a nonabelian fundamental group.

Before proving Theorem 5, we shall state and prove an elementary fact which will be needed in the course of the argument.

LEMMA 6. *Suppose that X is a simply connected space and $p, q \in X$. Then there is exactly one endpoint preserving homotopy class of continuous curves joining p to q.*

Proof of Lemma 6. If α and β are curves joining p to q, then

$$[\alpha] \;=\; [\alpha \;+\; \text{constant}] \;=\; [\alpha \;+\; (-\beta) \;+\; \beta]$$

where $\alpha + (-\beta)$ is a closed curve and thus is nullhomotopic. Therefore the right hand side is endpoint preserving homotopic to constant $+ \beta$, and since the latter is endpoint preserving homotopic to β it follows that $[\alpha] = [\beta]$.∎

Proof of Theorem 5. Let $y \in X$, and let $z \in X/G$ denote its image under the orbit space projection p. Define a map

$$\theta : G \;\longrightarrow\; \pi_1(X/G, z)$$

as follows: Take α to be a curve joining y to $g \cdot y$. The preceding results imply that $g \cdot y = z \cdot b$ for some b in $\pi_1(X/G, z)$, so we want to set $\theta(g) = b$. By construction, b is equal to the class $[p \circ \alpha]$ in $\pi_1(X/G, z)$. In order to show this assignment is well-defined, we need to know that b does not depend upon the choice of α. To see this, suppose that we have two curves α_1 and α_2 joining p to q; then Lemma 6 implies that α_1 and α_2 are endpoint preserving homotopic, and the latter implies that $p \circ \alpha_1$ and $p \circ \alpha_2$ are basepoint preserving homotopic.

Our results to this point imply that θ is well-defined and 1–1 onto, so the only thing remaining is to describe $\theta(g_1 g_2)$. Since the fundamental group acts freely on $F = p^{-1}[\{z\}]$ when X is simply connected, it will suffice to prove that $y \cdot \theta(g_1 g_2) = (y \cdot \theta(g_2)) \cdot \theta(g_1)$ for all g_1 and g_2.

In order to find $\theta(g_1 g_2)$ we need to find a curve in X joining y to $(g_1 g_2) \cdot y = g_1 \cdot (g_2 \cdot y)$. If α_i is a curve joining y to $g_i y$ for $i = 1, 2$ then such a curve is given by

$$\alpha_2 \;+\; g_2 \cdot \alpha_1$$

and hence $\theta(g_1 g_2)$ is the element of the fundamental group given by $c = [p \circ \alpha_2 + p \circ g_2 \alpha_1]$, and since $p(g_2 \cdot x) = p(x)$ for all $x \in X$ we may rewrite c as $[p \circ \alpha_2 + p \circ \alpha_1] = [p \circ \alpha_2] \cdot [p \circ \alpha_1]$. By construction the latter is $\theta(g_2) \cdot \theta(g_1)$, and thus we have shown that the 1–1 correspondence θ is an anti-isomorphism,∎

Fundamental group of the Klein Bottle

The Klein bottle turns out to be an example of a space whose fundamental group is infinite and nonabelian.

Recall that we have constructed the Klein bottle as the base space of a 2-sheeted covering space projection $T^2 \to K$. If we compose this with the standard covering space projection from $\mathbb{R}^2$ to T^2, we obtain an infinitely sheeted covering space projection φ from $\mathbb{R}^2$ to K by Exercise 53.4 on page 341 of Munkres.

Definition. If $p : E \to B$ is a covering space projection, then a homeomorphism $T : E \to E$ is said to be a *covering space transformation* of $p : E \to B$ if $p \circ T = p$. It follows immediately that the set $\Gamma(p)$ of all covering space transformations of $p : E \to B$ is a group with respect to composition of mappings.

We shall need the following strengthening of Theorem 5:

COMPLEMENT 5A. *Let $p : E \to B$ be a covering space transformation where E and B are connected, suppose that there is a subgroup π of $\Gamma(p)$ such that Γ acts freely and transitively on the inverse image of $F \subset E$ of some point $b_0 \in B$, and assume further that E is simply connected. Then $\pi_1(B, b_0)$ is (anti-)isomorphic to π.*

The proof is basically the same as the argument for Theorem 5. Note that in the setting of Theorem 5 the group action maps $x \to g \cdot x$ are covering space transformations which are transitive on F; in the result stated above, the finiteness hypothesis regarding π or G is dropped, but the assumption that the group acts by covering transformations is stronger than the corresponding hypothesis in the finite case.∎

By Complement 5A, it is only necessary to observe that there is a subgroup of covering transformations π for the covering space projection $\mathbb{R}^2 \to K$ which is infinite, transitive and nonabelian.

Let $e \in K$ be the image of $(1, 1) \in T^2$, and view $\mathbb{R}^2$ as the complex numbers $\mathbb{C}$. Then $\varphi^{-1}[\,\{e\}\,]$ consists of all complex numbers having the form $\frac{1}{2}m \, + \, n\mathbf{i}$ where m and n are integers. If we let χ denote complex conjugation, then covering transformations for φ are given by

$$X(z) \;\; = \;\; \chi(z) \; + \; \tfrac{1}{2} \, , \qquad Y(z) \;\; = \;\; z \; + \; n\mathbf{i}$$

and one can check directly that the subgroup generated by these transformations is transitive on $\varphi^{-1}[\,\{e\}\,]$ Note that X^2 and Y generate the group of covering transformations for the torus covering $\mathbb{R}^2 \to T^2$, which is isomorphic to $\mathbb{Z}^2$, and this subgroup has index 2 in the group generated by X and Y.

It follows that the group G generated by X and Y is the fundamental group of the Klein bottle. A routine computation shows that

$$Y \, X \, Y^{-1} X^{-1} \;\; = \;\; Y^2$$

and from this we can conclude that $\pi_1(K, e)$ is infinite and not abelian.∎

Note on spaces with finite fundamental groups

We have already seen that a simply connected space can only have 1-sheeted connected covering spaces (Corollary 4). For spaces with finite fundamental groups, we have the following weaker result, which is a nontrivial restriction on the number of sheets in a connected covering space.

PROPOSITION 7. *Let $p : (E, e_0) \to (B, b_0)$ be a covering space projection, where E and B are connected spaces satisfying the Default Hypothesis, and suppose that $\pi_1(B, b_0)$ is finite. Then $F = p^{-1}[\{b_0\}]$ is finitem and the number of elements in F divides the order of $\pi_1(B, b_0)$.*

Proof. By Theorem 3 we know that the fundamental group $\pi_1(B, b_0)$ acts trantively as a permutation group on F, and the subgroup of elements which fix the basepoint e_0 is equal to the subgroup $p_* \, [\pi_1(E, e_0)]$. Therefore standard considerations from group theory imply that the permutation action on F is equivalent to the action of $\pi_1(B, b_0)$ on the set of cosets

$$\pi_1(B, b_0) \; / \; p_* \, [\pi_1(E, e_0)]$$

and hence the cardinality of F is equal to the cardinality of this set of cosets. By Lagrange's Theorem on cosets we know that the latter cardinality divides the order of $\pi_1(B, b_0)$, and therefore the same conclusion must be true for F.∎

Simply connected regions in $\mathbb{R}^2$

Simply connected open subsets of the complex plane (equivalently, $\mathbb{R}^2$) play an important role in the theory of (analytic) functions of one complex variable. One reason for this involves the implications of simple connectivity for path independence of line integrals; we shall cover this explicitly in the next section of the notes. For the time being, we shall merely state a fundamental result on such regions known as the **Riemann Mapping Theorem**:

THEOREM 8. (Riemann Mapping Theorem) *Let $U \subset \mathbb{C} = \mathbb{R}^2$ be a simply connected (hence arcwise connected) open set. Then U is homeomorphic to the complex plane. In fact, if U is a proper subset of the complex plane, then there is a homeomorphism h from U to the open unit disk*

$$N_1(0) \;=\; \{\, z \in \mathbb{C} \mid |z| \;<\; 1 \,\}$$

such that both h and h^{-1} are complex analytic functions.

The proof of this result requires input from complex analysis, and one standard reference for a proof is Section 6.1 of Ahlfors, *Complex Analysis* (Third Edition). We shall not need the statement of this result or its proof in any of the subsequent material for this course.

For the record, we should also note that there are infinitely many homeomorphism types of simply connected open subsets in $\mathbb{R}^n$ for each $n \geq 3$. In particular, if S and T are finite subsets of $\mathbb{R}^n$ with different numbers of points, then $\mathbb{R}^n - S$ and $\mathbb{R}^n - T$ are not homeomorphic. The course directory file `openRn.pdf` gives a fairly short proof of this result using techniques developed in Mathematics 205B.

VIII.6 : Homotopy of paths and line integrals

(Munkres, 56)

Although the material in this section is not part of the official course coverage from the viewpoint of midterm, final or qualifying examinations, the subject matter (path dependence of line integrals and a proof of the Fundamental Theorem of Algebra) seems worthwhile.

For the most part, we shall work with curves $\Gamma : [0,1] \to U$, where U is open in $\mathbb{R}^n$ and Γ is a *piecewise regular smooth curve*; *i.e.*, there is a partition of $[0,1]$ into subintervals $J_1, \cdots, J_m$ such that the restriction of Γ to each J_α has a continuous derivative (= tangent vector) which is never zero. The boundary curve of a square in the counterclockwise sense is a typical example. One important point to note about these curves is that if z is a common endpoint of two subintervals J_α and $J_{\alpha+1}$, then the tangent vectors at z coming from the two subintervals need not be equal.

Previously we mentioned a few identities involving line integrals and operations such as concatenation. For the sake of convenience we shall now summarize them more formally (as above,

assume we are working inside the open connected subset U, the functions P and Q have continuous partial derivatives, and all curves which appear in the statements are piecewise smooth):

(1) If the curve γ is obtained by concatenating α and β, then the line integral of $P\,dx + Q\,dy$ over γ is equal to the sum of the corresponding line integrals over α and β.

(2) If the curve $-\gamma$ is obtained by reversing the direction of γ as in Section VIII.2, then the line integral of $P\,dx + Q\,dy$ over $-\gamma$ is equal to the negative of the corresponding line integral over γ.

(3) If C is the constant curve, then the line integral of $P\,dx + Q\,dy$ over C is zero.

Path dependence of line integrals

Let U be an open connected subset of $\mathbb{R}^2$, and let P and Q be real valued functions on U with continuous partial derivatives. In multivariable calculus one learns that certain choices of P and Q the line integrals

$$\int_\Gamma P\,dx \;+\; Q\,dy$$

depend only on the endpoints of Γ. The simplest examples are those for which the integrands satisfy

$$P \;=\; \frac{\partial f}{\partial x}\,, \qquad Q \;=\; \frac{\partial f}{\partial y}$$

for some smooth real valued function f defined on U. In such cases one can use the Fundamental Theorem of Calculus, the chain rule for partial differentiation, and the definition of a line integral to conclude that

$$\int_\Gamma P\,dx \;+\; Q\,dy \;=\; f\circ\Gamma(1) \;-\; f\circ\Gamma(0)\;.$$

More generally, if

$$\frac{\partial P}{\partial y} \;=\; \frac{\partial Q}{\partial x}$$

and P and Q have continuous partial derivatives on U, then the path dependence of this integral is a nontrivial issue in multivariable calculus; although the integral may depend upon path, examples show that the integral is the same for large families of closely related curves.

Standard example. Consider the line integral

$$\int_\Gamma \frac{x\,dy - y\,dx}{x^2 + y^2}$$

over the counterclockwise unit circle $(\cos t, \sin t)$ for $0 \leq t \leq 2\pi$. Direct computation shows that the value obtained is 2π, but if we consider the corresponding line integral over the counterclockwise circle of radius $\frac{1}{3}$ centered at $(\frac{2}{3}, 0)$ with parametrization

$$x(t) \;=\; \frac{2}{3} \;+\; \frac{1}{3}\cos t, \qquad y(t) \;=\; \frac{1}{3}\sin t \qquad (0 \leq t \leq 2\pi)$$

then direct computation shows that the integral's value is zero. On the other hand, further study shows that one obtains the same value of 2π for all circular curves in $\mathbb{R}^2 - \{\mathbf{0}\}$ which contain $\mathbf{0}$ in

their interior, and one obtains the same value of 0 for all curves which lie in the open half-plane defined by $x > 0$.

It is natural to ask the extent to which the line integral varies with the choice of path, and the basic results in this direction are sometimes stated without proof (or even complete definitions) in some multivariable calculus texts. In fact, it turns out that a precise formulation and proof involve homotopy classes of curves, so in this section we shall state and prove the basic results on this topic for open subsets in $\mathbb{R}^2$. Similar results also hold in $\mathbb{R}^n$ for $n \geq 3$, but formulating and proving them would require the development of additional background material; for the sake of completeness, Section V.6 in `advancednotes2014.pdf` summarizes how this can be done using more sophisticated techniques.

For our purposes it will be convenient to center the exposition around the following version of the main results:

THEOREM 1. *Let U be a connected open subset of $\mathbb{R}^2$, and let P and Q be smooth functions on U with continuous partial derivatives which satisfy the condition*

$$\frac{\partial P}{\partial y} \;=\; \frac{\partial Q}{\partial x}$$

at all points of U. If Γ is a piecewise smooth closed curve which starts and ends at $\mathbf{p}_0 \in U$ which is basepoint preserving homotopic to a constant in U, then

$$\int_\Gamma P\,dx + Q\,dy \;=\; 0.$$

Before proving this result, we shall state a few alternate versions and derive them from Theorem 1.

THEOREM 2. *Let U, P, Q be given as in Theorem 1, but suppose now that Γ and Γ' are two piecewise smooth curves in U with the same endpoints $\mathbf{p}$ and $\mathbf{q}$ such that Γ and Γ' are homotopic by an endpoint preserving homotopy. Then*

$$\int_\Gamma P\,dx + Q\,dy \;=\; \int_{\Gamma'} P\,dx + Q\,dy\ .$$

Proof that Theorem 1 implies Theorem 2. In the setting of Theorem 2 the curve $\Gamma' + (-\Gamma)$ is a closed piecewise smooth curve that is homotopic to a constant because $\Gamma \simeq \Gamma'$ implies $[\Gamma' + (-\Gamma)] = [\Gamma + (-\Gamma)] = [\text{constant}]$. Therefore Theorem 1 implies that the line integral over this curve is zero. On the other hand, by the three properties of line integrals listed above, the line integral over $\Gamma' + (-\Gamma)$ is equal to the difference of the line integrals over Γ' and Γ. Combining these observations, we see that the line integrals over Γ' and Γ must be equal.∎

The next result is often also found in multivariable calculus texts.

COROLLARY 3. *If in the setting preceding theorems we also know that the region U is* simply connected, *then the following hold:*

(i) for every piecewise smooth closed curve Γ in U we have

$$\int_\Gamma P\,dx + Q\,dy \;=\; 0$$

(*ii*) *for every pair of piecwise smooth curves* Γ, Γ' *with the same endpoints we have*

$$\int_{\Gamma} P\,dx + Q\,dy \quad = \quad \int_{\Gamma'} P\,dx + Q\,dy.$$

The first part of the corollary follows from the triviality of the fundamental group of U, the conclusion of Theorem 1, and the triviality of line integrals over constant curve. The second part follows formally from the first in the same way that the Theorem 2 follows from Theorem 1.∎

Finally, we have the following result, which plays a fundamental role in the theory of functions of one complex variable.

THEOREM 4. *Let* U, P, Q *be given as in Theorems* 1 *and* 2, *but suppose now that* Γ *and* Γ' *are two piecewise smooth closed curves in* U *such that* Γ *and* Γ' *are* freely *homotopic. Then*

$$\int_{\Gamma} P\,dx + Q\,dy \quad = \quad \int_{\Gamma'} P\,dx + Q\,dy.$$

This proof will require some additional input, so the argument will be postponed until after the proof of Theorem 1 is completed.

Background from multivariable calculus

The following result can be found in many multivariable calculus textbooks.

THEOREM 5. *Let* U *be a rectangular open subset of the coordinate plane of the form* $(a_1, b_1) \times (a_2, b_2)$ *where each factor is an open interval in the real line, let* P *and* Q *be functions on* U *with continuous partial derivatives on* U *such that*

$$\frac{\partial P}{\partial y} \quad = \quad \frac{\partial Q}{\partial x}$$

and let Γ *and* Γ' *be two piecewise smooth curves in* U *with the same endpoints. Then*

$$\int_{\Gamma} P\,dx + Q\,dy \quad = \quad \int_{\Gamma'} P\,dx + Q\,dy.$$

Sketch of proof. The underlying idea behind the proof is to construct a function f such that $\nabla f = (P, Q)$; if this can be done then as before the result will follow from the Fundamental Theorem of Calculus and the chain rule for partial differentiation. We start with an arbitrary point (x_0, y_0) in U; given $(x, y) \in U$, consider the following two broken line curves in U:

(HV) Take the horizontal line segment curve from (x_0, y_0) to (x, y_0) and concatenate it with the vertical line segment from (x, y_0) to (x, y). If either $x_0 = x$ or $y_0 = y$ then the corresponding line segment curve is a constant curve.

(VH) Take the the vertical line segment from (x_0, y_0) to (x_0, y) and concatenate it with the horizontal line segment from (x_0, y) to (x, y). If either $x_0 = x$ or $y_0 = y$ then the corresponding line segment curve is a constant curve.

The curve VH+(−HV) traces the boundary of a solid rectangle contained in U, and thus we can use the condition on partial derivatives along with Green's Theorem to conclude that the line integral along this curve is zero. This means that

$$\int_{\mathrm{VH}} P\,dx + Q\,dy \;\; = \;\; \int_{\mathrm{HV}} P\,dx + Q\,dy \;.$$

Define $f(x,y)$ to be the common value of these two integrals. One can now use Green's Theorem to derive the identity $\nabla f(x,y) = (\,P(x,y),\,Q(x,y)\,)$ for $(x,y) \in U$.∎

Broken line inscriptions

We begin with some standard definitions. Given two points $\mathbf{p} = (p_1, p_2)$ and $\mathbf{q} = (q_1, q_2)$ in $\mathbb{R}^n$, the *closed straight line segment* joining them is the curve $[\mathbf{p}, \mathbf{q}]$ defined by $(1-t)\mathbf{p} \,+\, t\mathbf{q}$ over the interval $[0,1]$.

A broken line curve corresponding to an ordered sequence of points

$$\mathbf{p}_0,\;\mathbf{p}_1,\;\cdots\;\mathbf{p}_m$$

is obtained by joining $\mathbf{p}_0$ to $\mathbf{p}_1$ by a straight line segment $[\mathbf{p}_0\mathbf{p}_1]$, then joining $\mathbf{p}_1$ to $\mathbf{p}_2$ by a straight line segment $[\mathbf{p}_1\mathbf{p}_2]$, and so on. The points $\mathbf{p}_0$, $\mathbf{p}_1$, $\mathbf{p}_2$, *etc.* are called the *vertices* of the broken line curve. One technical problem with this involves the choices of linear parametrizations for the pieces. However, since line integrals for such curves do not depend upon such parametrizations and in fact we have

$$\int_C P\,dx + Q\,dy \;\; = \;\; \sum \int_{[\mathbf{p}_{i-1},\mathbf{p}_i]} P\,dx + Q\,dy$$

we shall not worry about the specific choice of parametrization. Filling in the details will be left as an exercise to a reader who is interested in doing so; this is basically elementary but tedious.

We shall be considering broken line approximations to a piecewise smooth curve, and this requires a few more definitions. A *partition* of the interval $[a,b]$ is a sequence of points

$$\Delta : a \;\; = \;\; t_0 \;<\; t_1 < \;\cdots\; < t_m \;\; = \;\; b$$

and the *mesh* of Δ, written $|\Delta|$, is the maximum of the differences $t_i - t_{i-1}$ for $1 \leq i \leq m$. Given a piecewise smooth curve Γ defined on $[a,b]$, the *broken line inscription* $\mathsf{Lin}\,(\Gamma, \Delta)$ is the broken line curve with vertices

$$\Gamma(a) \;=\; \Gamma(t_0),\;\Gamma(t_1),\;\cdots\;\Gamma(t_m) \;=\; \Gamma(b)\;.$$

We are now ready to prove one of the key technical steps of the proof of the main result.

LEMMA 6. *Let U, P, Q, Γ be as usual, where Γ is defined on $[a,b]$ and P and Q satisfy the condition*

$$\frac{\partial P}{\partial y} \;\; = \;\; \frac{\partial Q}{\partial x}\;.$$

Then there is a positive constant $\delta > 0$ such that the following hold for all partitions Δ of $[a,b]$ with $|\Delta| < \delta$:

(i) The curves Γ and $\mathsf{Lin}\,(\Gamma, \Delta)$ are endpoint preserving homotopic.

(*ii*) *We have a line integral equation*

$$\int_\Gamma P\,dx + Q\,dy \;\; = \;\; \int_{\mathsf{Lin}\,(\Gamma,\Delta)} P\,dx + Q\,dy \;.$$

Proof. If K is the image of Γ then K is a compact subset of the open set U, and therefore there is an $\varepsilon > 0$ so that if $\mathbf{x} \in \mathbb{R}^2$ satisfies $|\mathbf{x} - \mathbf{v}| < \varepsilon$ for some $\mathbf{v} \in K$ then $\mathbf{x} \in U$. It follows that if $\mathbf{v} \in K$ then the inner region for the square centered at $\mathbf{v}$ with sides parallel to the coordinate axes of length $\varepsilon\sqrt{2}$ lies entirely in U.

By uniform continuity there is a $\delta > 0$ so that if $s, t \in [a, b]$ satisfy $|s - t| < \delta$ then

$$|\Gamma(s) - \Gamma(t)| \;\; < \;\; \frac{\varepsilon\sqrt{2}}{2} \;.$$

Let Δ be a partition of $[a, b]$ whose mesh is less than δ. Then for all i the restriction of Γ to $[t_{i-1}, t_i]$ lies in the open disk of radius $\frac{1}{2}\varepsilon\sqrt{2}$. It follows that both this restriction and the closed straight line segment joining $\Gamma(t_{i-1})$ to $\Gamma(t_i)$ lie in the open square region centered at $\Gamma(t_{i-1})$ with sides parallel to the coordinate axes of length of length $\varepsilon\sqrt{2}$. The first conclusion of the lemma now follows because the image of the straight line homotopy from Γ to $\mathsf{Lin}\,(\Gamma, \Delta)$ will be contained in U.

We shall now prove the second conclusion in the lemma. Since the open squares lie entirely in U. it follows that P and Q are defined on these square regions. Therefore, by Theorem 5 we have

$$\int_{\Gamma|[t_{i-1},t_i]} P\,dx + Q\,dy \;\; = \;\; \int_{[\Gamma(t_{i-1}),\Gamma(t_i)]} P\,dx + Q\,dy$$

for each i. But the line integral over Γ is the sum of the line integrals over the curves $\Gamma|[t_{i-1}, t_i]$, and the line integral over the broken line inscription is the sum of the line integrals over the line segments $[\Gamma(t_{i-1}), \Gamma(t_i)]$, and therefore it follows that the line integral over Γ is equal to the line integral over the broken line inscription, as required.∎

We shall also need a version of Lemma 6 with weaker hypotheses on Γ.

LEMMA 7. *Let U, P, Q, Γ be as in Lemma 6, but now assume only that Γ is a continuous curve. Then there is a positive constant $\delta > 0$ such that the following hold for all partitions Δ of $[a, b]$ with $|\Delta| < \delta$:*

(*i*) *The curves Γ and $\mathsf{Lin}\,(\Gamma, \Delta)$ are endpoint preserving homotopic.*

(*ii*) *The values of the line integral*

$$\int_{\mathsf{Lin}\,(\Gamma,\Delta)} P\,dx + Q\,dy$$

are equal for all choices of Δ such that $|\Delta| < \delta$.

This lemma can be used to define a **formal value** for the line integral

$$\int_\Gamma P\,dx + Q\,dy$$

even if Γ is not a rectifiable curve, provided P and Q satisfy the condition on partial derivatives; namely, we can take the line integral of some broken line inscription satisfying the condition in Lemma 7.

Proof of Lemma 7. We can prove the first part exactly as in Lemma 6, so it is only necessary to show that the line integrals over all the broken line approximations $\mathsf{Lin}\,(\Gamma, \Delta)$ are equal if $|Delta| < \delta$.

We shall first prove this if one partition is a refinement of the other; as usual, a partition Δ' is said to be a *refinement* of Δ if every partition point in Δ is also a partition point of Δ'; it follows immediately that the mesh of Δ' is no greater than the mesh of Δ. Every refinement can be viewed as the composite of a sequence of elementary refinements

$$\Delta \;=\; \Delta_0 \;<\; \Delta_1 \;<\; \cdots \;<\; \Delta_m \;=\; \Delta'$$

such that Δ_j is obtained from Δ_{j-1} by adding a single point, and therefore by an inductive argument it suffices to prove that the line integrals over $\mathsf{Lin}\,(\Gamma, \Delta)$ and $\mathsf{Lin}\,(\Gamma, \Delta')$ are equal if Δ' is obtained from Δ by adding a single point.

The additional partition point u lies between t_{j-1} and t_j for some j, and it follows that the difference between the line integral over $\mathsf{Lin}\,(\Gamma, \Delta)$ and $\mathsf{Lin}\,(\Gamma, \Delta')$ is given by

$$D \;=\; \int_{[\Gamma(t_{j-1})\Gamma(t_j)]} \Omega \;-\; \int_{[\Gamma(t_{j-1})\Gamma(u)]} \Omega \;-\; \int_{[\Gamma(u)\Gamma(t_j)]} \Omega$$

where $\Omega = P\,dx + Q\,dy$. Since the mesh of Δ is small, it follows that the image of $[t_{j-1}t_j]$ under Γ is a small open square. Therefore Theorem 5 implies the path independence identity

$$\int_{[\Gamma(t_{j-1})\Gamma(t_j)]} \Omega \;=\; \int_{[\Gamma(t_{j-1})\Gamma(u)]} \Omega \;+\; \int_{[\Gamma(u)\Gamma(t_j)]} \Omega\;.$$

If we combine these observations, we see that $D = 0$, and as noted above this implies that the line integrals over $\mathsf{Lin}\,(\Gamma, \Delta)$ and $\mathsf{Lin}\,(\Gamma, \Delta')$ are equal.

In general, if we are given two partitions Δ and Δ' there is a third partition Δ^* which is a refinement of both; it suffices to take the partition whose partition points are the union of the partition points for Δ and Δ'. By the preceding paragraph, we then know that the line integrals over both $\mathsf{Lin}\,(\Gamma, \Delta)$ and $\mathsf{Lin}\,(\Gamma, \Delta')$ are equal to the line integral over $\mathsf{Lin}\,(\Gamma, \Delta^*)$, and hence it follows that the first two line integrals are equal, which is what we wanted to prove.∎

Proof of Theorems 1 *and* 4

We shall prove these in order.

Proof of Theorem 1. Let $H : [0,1] \times [0,1] \to U$ be a continuous map such that $H(s,0) \;=\; \Gamma(s)$ for all s and H is constant on both $[0,1] \times \{1\}$ and $\{0,1\} \times [0,1]$.

If L is the image of H then L is a compact subset of the open set U, and as in the proof of the lemma there is an $\varepsilon' > 0$ so that if $\mathbf{x} \in \mathbb{R}^2$ satisfies $|\mathbf{x} - \mathbf{v}| < \varepsilon'$ for some $\mathbf{v} \in L$ then $\mathbf{x} \in U$. It follows that if $\mathbf{v} \in L$ then the inner region for the square centered at $\mathbf{v}$ with sides parallel to the coordinate axes of length $\varepsilon'\sqrt{2}$ lies entirely in U.

By uniform continuity there is a $\delta' > 0$ so that if $\mathbf{s}, \mathbf{t} \in [0,1] \times [0,1]$ satisfy $|\mathbf{s} - \mathbf{t}| < \delta'$ then

$$|H(\mathbf{s}) - H(\mathbf{t})| \quad < \quad \frac{\varepsilon' \sqrt{2}}{2} \ .$$

Without loss of generality we may assume that δ' is no greater than the δ in the previous lemma. Let Δ be a partition of $[a,b]$ whose mesh is less than $\frac{1}{2}\delta'\sqrt{2}$, and choose a positive integer N such that

$$\frac{1}{N} \quad < \quad \frac{\delta' \sqrt{2}}{2}.$$

Then for all i such that $1 \leq i \leq m$ and all j such that $1 \leq j \leq N$ the restriction of H to $[t_{i-1}, t_i] \times [\frac{j-1}{N}, \frac{j}{N}]$ lies in an open disk of radius $\frac{1}{2}\varepsilon'\sqrt{2}$.

A special case. To motivate the remainder of the argument, we shall first specialize to the case where H extends to a map on an open set containing the square $[0,1] \times [0,1]$ and has continuous partial derivatives on this open set. For each i such that $0 \leq i \leq m$ and each j such that $1 \leq j \leq N$ let $A(i,j)$ be the broken line curve in the square with vertices

$$(0, \tfrac{j-1}{N}),\ \ldots\ (t_i, \tfrac{j-1}{N}),\ (t_i, \tfrac{j}{N}),\ \ldots\ (1, \tfrac{j}{N})\ .$$

In other words, this curve is formed by starting with a horizontal line segment from $(0, \frac{j-1}{N})$ to $(t_i, \frac{j-1}{N})$, then concatenating with a vertical line segment from $(t_i, \frac{j-1}{N})$ to $(t_i, \frac{j}{N})$, and finally concatenating with a horizontal line segment from $(t_i, \frac{j}{N})$ to $(1, \frac{j}{N})$. If $W(i,j)$ denotes the composite $H \circ A(i,j)$, then it follows that $W(i,j)$ is a piecewise smooth closed curve in U. Furthermore, $W(m,1)$ is just the concatenation of Γ with a constant curve and $W(0,N)$ is just a constant curve, so the proof of the main result reduces to showing that the line integrals of the expression $P\,dx + Q\,dy$ over the curves $W(m,1)$ and $W(0,N)$ are equal. We claim this will be established if we can show the following hold for all i and j:

(1) The corresponding line integrals over the curves $W(0, j-1)$ and $W(m,j)$ are equal.

(2) The corresponding line integrals over the curves $W(i-1, j)$ and $W(i,j)$ are equal.

To prove the claim, first note that (2) implies that the value of the line integral over $W(i,j)$ is a constant z_j that depends only on j, and then note that (1) implies $z_{j-1} = z_j$ for all j. Thus the two assertions combine to show that the line integrals over all the curves $W(i,j)$ have the same value.

We begin by verifying (1). Since H is constant on $\{0,1\} \times [0,1]$, it follows that $W(m,j)$ is formed by concatenating $H|[0,1] \times \{\frac{j}{m}\}$ and a constant curve (in that order), while $W(0, j-1)$ is formed by concatenating a constant curve and $H|[0,1] \times \{\frac{j}{m}\}$ (again in the given order). Thus the line integrals over both $W(0,j-1)$ and $W(m,j)$ are equal to the line integral over $H|[0,1] \times \{\frac{j}{m}\}$, proving (1).

Turning to (2), since the broken line curves $A(i,j)$ and $A(i-1,j)$ differ only by one vertex, it follows that the difference

$$\int_{W(i,j)} P\,dx + Q\,dy \ - \ \int_{W(i-1,j)} P\,dx + Q\,dy$$

is equal to

$$\int_{V(i,j)} P\,dx + Q\,dy \ - \ \int_{V'(i,j)} P\,dx + Q\,dy$$

where $V(i,j)$ is the composite of H with the broken line curve with vertices

$$\left(t_{i-1}, \tfrac{j-1}{N}\right), \quad \left(t_i, \tfrac{j-1}{N}\right), \quad \left(t_i, \tfrac{j}{N}\right)$$

and $V'(i,j)$ is the composite of H with the broken line curve with vertices

$$\left(t_{i-1}, \tfrac{j-1}{N}\right), \quad \left(t_{i-1}, \tfrac{j}{N}\right), \quad \left(t_i, \tfrac{j}{N}\right).$$

Our hypotheses imply that both of these curves lie in an open disk of radius $\frac{1}{2}\varepsilon'\sqrt{2}$ and thus also in the open square centered at $\mathbf{v}$ with sides parallel to the coordinate axes of length $\varepsilon'\sqrt{2}$; by construction the latter lies entirely in U. Therefore by the previously quoted result from multivariable calculus we have

$$\int_{V(i,j)} P\,dx + Q\,dy \quad = \quad \int_{V'(i,j)} P\,dx + Q\,dy$$

for each i and j, so that the difference of the line integrals vanishes. Since this difference is also the difference between the line integrals over $W(i,j)$ and $W(i-1,j)$, it follows that the line integrals over the latter two curves must be equal.

The general case. If H is an arbitrary continuous function the preceding proof breaks down because we do not know if the continuous curves $W(i,j)$ are well enough behaved to define line integrals. We shall circumvent this by using broken line approximations to these curves and appealing to the previous lemma to relate the value of the line integrals over these approximations to the value on the original curve. Since the proof is formally analogous to that for the special case we shall concentrate on the changes that are required.

Let $X(i,j)$ denote the broken line curve with vertices

$$H(0, \tfrac{j-1}{N}),\ \ldots\ H(t_i, \tfrac{j-1}{N}),\ H(t_i, \tfrac{j}{N}),\ \ldots\ H(1, \tfrac{j}{N}).$$

By our choice of Δ these broken lines all lie in U, and the constituent segments all lie in suitably small open disks inside U.

We claim that it will suffice to prove that the line integrals over the curves $X(0, j-1)$ and $X(m,j)$ are equal for all j and for each j the corresponding line integrals over the curves $X(i-1,j)$ and $X(i,j)$ are equal. As before it will follow that the line integrals over all the broken line curves $X(i,j)$ have the same value. But $X(m,N)$ is a constant curve, so this value is zero. On the other hand, by construction the curve $X(m,1)$ is formed by concatenating $\mathsf{Lin}\,(\Gamma,\Delta)$ and a constant curve, so this value is also the value of the line integral over $\mathsf{Lin}\,(\Gamma,\Delta)$. But now the Lemma implies that the values of the corresponding line integrals over Γ and $\mathsf{Lin}\,(\Gamma,\Delta)$ are equal, and therefore the value of the line integral over the original curve Γ must also be equal to zero.

The first set of equalities follow from the same sort argument used previously for $W(0,j-1)$ and $W(m,j)$ with the restriction of Γ replaced by the broken line curve with vertices

$$H(0, \tfrac{j}{N}),\ \ldots\ H(1, \tfrac{j}{N}).$$

To verify the second set of equalities, note that the difference between the values of the line integrals over $X(i,j)$ and $X(i-1,j)$ is given by

$$\int_{C(i,j)} P\,dx + Q\,dy \ - \ \int_{C'(i,j)} P\,dx + Q\,dy$$

where $C(i,j)$ is the broken line curve with vertices

$$H\left(t_{i-1}, \tfrac{j-1}{N}\right), \quad H\left(t_i, \tfrac{j-1}{N}\right), \quad H\left(t_i, \tfrac{j}{N}\right)$$

and $C'(i,j)$ is the broken line curve with vertices

$$H\left(t_{i-1}, \tfrac{j-1}{N}\right), \quad H\left(t_{i-1}, \tfrac{j}{N}\right), \quad H\left(t_i, \tfrac{j}{N}\right).$$

By the previously quoted result from multivariable calculus we have

$$\int_{C(i,j)} P\,dx + Q\,dy \;\; = \;\; \int_{C'(i,j)} P\,dx + Q\,dy$$

for each i and j, and therefore the difference between the values of the line integrals must be zero. Therefore the difference between the values of the line integrals over $X(i,j)$ and $X(i-1,j)$ must also be zero, as required. This completes the proof.∎

Before proceeding to the proof of Theorem 4, we shall note the following consequence of the results obtained thus far:

PROPOSITION 8. *Let U be an open connected subset of $\mathbb{R}^2$, let $u_0 \in U$, and let P and Q be functions on U with continuous partial derivatives on U such that*

$$\frac{\partial P}{\partial y} \;\; = \;\; \frac{\partial Q}{\partial x}\,.$$

Then there is a group homomorphism $\Sigma : \pi_1(U, u_0) \to \mathbb{R}$ (with the additive group structure on $\mathbb{R}$) such that if γ is a basepoint preserving piecewise smooth curve in U which starts and ends at u_0, then

$$\Sigma([\gamma]) \;\; = \;\; \int_\gamma \; P\,dx \; + \; O\,dy\,.$$

Proof. Lemma 7 shows that every basepoint preserving homotpy class of closed curves x has an representative γ which is piecewise smooth (in fact, one can find a broken line approximation to a given continuous curve). Define $\Sigma(x)$ to be the line integral of $\Omega = P\,dx + Q\,dy$ over γ. By Theorem 1 of this section, the value of the line integral does not depend upon the choice of representative, so the mapping is well-defined. The additivity of Σ follows from the general properties of line integrals at the beginning of this unit.∎

Proof of Theorem 4. Suppose that α and β are freely homotopic curves in U, and suppose that they start and end at p and q respectively. By Proposition 2.9 there is a curve ω joining p to q such that the class $[\alpha + \omega + (-\beta) + (-\omega)]$ in $\pi_1(U,p)$ is trivial. By Lemmas 6 and 7 there are broken line inscriptions $L(\alpha),\ L(\beta),\ L(\omega)$ of $\alpha,\ \beta,\ \omega$ such that $L(\xi)$ is endpoint preserving homotopic to ξ for $\xi = \alpha, \beta, \omega$. Combining these observations, we see that

$$[\alpha + \omega + (-\beta) + (-\omega)] \;\; = \;\; [L(\alpha) + L(\omega) + (-L(\beta)) + (-L(\omega))] \;\; = \;\; 1 \;\; \in \;\; \pi_1(U,p)$$

and therefore the basic algebraic properties of line integrals and Theorem 1 imply that

$$0 \;\; = \;\; \int_{L(\alpha)+L(\omega)+(-L(\beta))+(-L(\omega))} P\,dx \; + \; Q\,dy \;\; =$$

$$\int_{L(\alpha)} P\,dx \;+\; Q\,dy \;+\; \int_{L(\omega)} P\,dx \;+\; Q\,dy \;-\; \int_{L(\beta)} P\,dx \;+\; Q\,dy \;-\; \int_{L(\omega)} P\,dx \;+\; Q\,dy \;=$$

$$\int_{L(\alpha)} P\,dx \;-\; \int_{L(\beta)} P\,dx \;+\; Q\,dy$$

and hence the line integrals over $L(\alpha)$ and $L(\beta)$ are equal. But Theorem 1 implies that the line integrals of $L(\xi)$ and ξ are equal for $\xi = \alpha, \beta$, and if we combine this with the preceding sentence we see that the line integrals over α and β are equal.■

Line integrals of complex analytic functions

For the sake of completeness, we recall that the *Fundamental Theorem of Algebra* states that every nonconstant polynomial $p(z)$ over the complex numbers has a root; *i.e.*, there is some complex number c such that $p(c) = 0$.

As noted on pp. 353–354 of Munkres, there are many proofs of the Fundamental Theorem of Algebra, and ultimately they all require some input that is intrinsically nonalgebraic and involves the geometry or topology of the complex plane (so this is really a theorem **about** algebra and not a theorem **of** algebra. In particular, a standard approach using the theory of functions of one complex variable is mentioned at the top of page 354. If one looks carefully at the proofs of the Fundamental Theorem of Algebra in many complex variables texts, issues about the completeness of the arguments often arise. Usually these concern path independence properties of line integrals. A logically rigorous approach to these issues normally requires some information about homotopy classes of closed curves in open subsets of the plane (the same input which appears explicitly in Munkres' proof). Therefore we shall discuss some issues involving line integrals of complex analytic functions over rectifiable curves before looking at the proof of the Fundamental Theorem of Algebra.

One immediate complication involves the definition of an analytic function; in some references it is defined as a complex valued function f defined on an open subset $U \subset \mathbb{C}$ such that f' exists and is continuous on U, and in other references it is taken to be a function f for which f' exists, with no *a priori* assumption of its continuity. In fact, the two notions are equivalent, for the existence of f' guarantees its continuity, but this is a nontrivial result. We shall consider both cases here, beginning with the easier one in which f' is assumed to be continuous.

Suppose we know that f' exists and is continuous. Suppose that we are given a piecewise smooth (or, more generally, a rectifiable continuous) curve γ. Write the function f in the form $f = u + v\mathbf{i}$, where u and v are functions with continuous partial derivatives satisfying the Cauchy-Riemann equations. Then the line integral $\int_\gamma f(z)\,d(z)$ is equal to

$$\int_\gamma u\,dx \;-\; v\,dy \;+\; \mathbf{i}\cdot\int_\gamma v\,dx \;+\; u\,dy\;.$$

Assume now that the region U in the complex plane is rectangular with sides parallel to the coordinate axes (all $x + y\,i$ such that $a \leq x \leq b$ and $c \leq y \leq d$). We claim that the given line integral depends only upon the initial and final points of γ. This is shown using corresponding results from multivariable calculus about path independence. By Green's Theorem, a line integral $\int_\gamma P\,dx + Q\,dy$ over a rectangular region is path independent if we have

$$\frac{\partial Q}{\partial x} \;=\; \frac{\partial P}{\partial Y}$$

and using the Cauchy-Riemann equations $u_x = v_y$, $u_y = -v_x$, we see that the displayed relation holds for the integrands in the real and imaginary parts of $\int_\gamma \ f(z)\,dz$. The material in this section then leads to the following basic result:

PROPOSITION 9. *Let f be an analytic function on the open set $U \subset \mathbb{C}$ in the stronger sense (f' is continuous), and let α and β be continuous rectifiable curves in U such that α and β are freely homotopic. Then $\int_\alpha \ f(z)\,dz = \int_\beta \ f(z)\,dz$.*∎

Now suppose we know that f' exists but we are not given any information regarding its continuity (the weaker definition of analytic function that is found in many texts). We can use the preceding approach PROVIDED we can show that if U is a rectangular region then $\int_\gamma \ f(z)\,dz$ only depends upon its endpoints. This is done in many complex variables books; for example, it appears on pp. 109–115 of the book by Ahlfors, Section 9.2 of the book by Curtiss, and Section 2.3 of the book by Fisher, all of which are listed below:

L. V. Ahlfors. *Complex Analysis* (3rd Ed.), *McGraw-Hill, New York*, 1979.

J. H. Curtiss. *Introduction to Functions of a Complex Variable* (Pure and Applied Math., Vol. 44). Marcel Dekker, New York, 1978.

S. D. Fisher. *Complex Variables* (2nd Ed.), *Dover, New York*, 1990.

The notion of homotopy also leads to a definitive version of the Cauchy Integral Formula for an analytic function f defined near the complex number a:

$$f(a) \quad = \quad \frac{1}{2\pi\, i} \cdot \int_\gamma \ \frac{f(z)}{z-a}\, dz$$

The point is that we can give and explicit description of the type of curve γ for which the formula is valid; namely, if f is defined on the open set U and $a \in U$, then we can take γ to be an arbitrary continuous rectifiable curve in $U - \{a\}$ which is homotopic to a counterclockwise circle of arbitrary radius centered at a.

Application to the Fundamental Theorem of Algebra

We shall conclude this section by showing deriving the Fundamental Theorem of Algebra. The first step is a familiar limit formula.

LEMMA 10. *If $p(z)$ is a nonconstant monic polynomial in the complex plane then*

$$\lim_{z \to \infty} |p(z)| = \infty \ ,$$

Sketch of proof. Use the identity

$$p(z) \quad = \quad z^n \cdot \Bigl(1 + \frac{c_{n-1}}{z} + \frac{c_{n-2}}{z^2} + \cdots + \frac{c_1}{z^{n-1}} + \frac{c_0}{z^n} \Bigr)$$

and the fact that the limit of the term inside the parentheses is zero.∎

COROLLARY 11. *In the setting above there is an $r > 0$ such that $R \geq r$ implies that $p(z)$ is never zero on a circle C_R of radius R about the origin.*∎

Now let $\Gamma(p, R)$ be the closed curve given by $p(R \cdot \exp(2\pi\, i\, t))$ for $0 \leq t \leq 1$, so that $\Gamma(p, R)$ just describes the behavior of the polynomial p on the circle of radius R about the origin. Consider the so-called *winding number integral*

$$\int_{\Gamma(p,R)} \frac{x\, dy - y\, dx}{x^2 + y^2}.$$

The proof of the Fundamental Theorem of Algebra has two remaining steps.

(1) *If $p(z) \neq 0$ for all z satisfying $|z| \leq R$, then the winding number integral is zero.*

(2) *If p has degree n then the winding number integral is equal to n.*

Proof of first statement. By construction $\Gamma(p, R)$ lies in the punctured plane $\mathbb{C} - \{0\}$. Since p has no zero points it follows that p also defines a map into the punctured plane, and the restriction of p to the solid disk of radius R defines a basepoint preserving homotopy from $\Gamma(p, R)$ to the constant curve. Therefore by homotopy invariance we know that the corresponding line integrals over $\Gamma(p, R)$ and the constant curve are equal. Since the latter integral is zero it follows that the original winding number integral is also zero.

Proof of second statement. First of all, if $p(z) = z^n$ then it follows that the winding number is n by direct calculation. It will suffice to show that for R sufficiently large the closed curves $\Gamma(p, R)$ and z^n are homotopic, for then we can use the modified form of the main result to show that the line integrals associated to the two polynomials are equal.

To prove this we use the identity

$$p(z) \quad = \quad z^n \cdot \left(1 + \frac{c_{n-1}}{z} + \frac{c_{n-2}}{z^2} + \cdots + \frac{c_1}{z^{n-1}} + \frac{c_0}{z^n}\right)$$

to conclude that

$$\lim_{z \to \infty} \frac{p(z)}{z^n} \quad = \quad 1.$$

In particular, there is an $S > 0$ such that $R > S$ implies that

$$\left| \frac{p(z)}{z^n} - 1 \right| \quad < \quad \frac{1}{2}.$$

This in turn implies that if $|z| = R$ then the line segment joining 1 to

$$\frac{p(z)}{z^n}$$

lies entirely in the punctured plane. If $h(z, t)$ is this straight line homotopy on the circle $|z| = R$ then $z^n h(z, t)$ defines a homotopy between $\Gamma(p, R)$ and the closed curve defined by the restriction of z^n to the circle of radius R. As noted before, this completes the proof of the second statement and of the Fundamental Theorem of Algebra.■

IX. Computing fundamental groups

To a great extent, the value of an algebraic construction depends upon how well one can do computations, and the fundamental group is no exception to this principle. Usually quantitative computations are necessary or desirable, but in some cases qualitative information is enlightening or useful (for example, the countability of $\pi_1(U, u)$ if U is an open subset of $\mathbb{R}^n$ for some n). In this unit we shall describe some fundamentally important techniques for computing fundamental groups, with particular emphasis on the following question: *If we are given a space X as a union of two arcwise connected subsets U and V such that $U \cap V$ is connected, can we recover the fundamental group of X from the data associated to the fundamental groups of U, V and $U \cap V$?*

Analysis of this question will require some general input from group theory, and the relevant topics are covered in Sections 1 and 2. The associated definitions and results contrast sharply with most of the material on group theory in entry level graduate level algebra courses, where the emphasis quickly turns to finite groups. We shall need concepts and results which are mainly useful for infinite groups; this dichotomy reflects a basic theme which runs through much of group theory:

Methods and examples from topology and geometry often play a crucial role in the study of problems involving infinite groups, and many topics involving such groups arise in connection with topological or geometrical examples.

Here is an example which can be stated simply: There is a large, topologically significant class of infinite groups for which the only element of finite order is the identity, and for some purposes it may be worthwhile to study groups with this property from a purely algebraic viewpoint.

The main theorem in this unit (the **Seifert-van Kampen Theorem**) is stated and proved in Section 3, and it gives a purely algebraic answer to the basic question which appears at the end of the first paragraph. In the final section we shall use the techniques of the first three sections to carry out computations of fundamental groups in some specific and important special cases.

IX.1 : Free groups

(Munkres, 67–69; Hatcher, 1.2)

> Informally, a free object over a set A can be thought as being a "generic" algebraic structure over A: the only equations that hold between [*sic*] elements of the free object are those that follow from the defining axioms of the algebraic structure.
>
> `http://en.wikipedia.org/wiki/Free_object`

If R is an associative ring with unit and A is some set, then the concept of a *free R-module generated by A* appears in entry level graduate algebra courses (*e.g.*, see Section IV.2 in Hungerford). The notion of a *free abelian group generated by A* (Hungerford, Section II.1) can be viewed as a special case in which R is the integers. If M is isomorphic to a free R-module generated by

$A \subset M$, we often say that M is freely generated by A. By Theorem IV.2.1 in Hungerford, the notion of a free module on a set A is characterized abstractly by the following mapping property:

> *If M is freely generated by A and we are given a second R- module N together with some set-theretic mapping $f : A \to N$, then there is a unique R-module homomorphism $F : M \to N$ such that $F|A = f$.*

We shall use this characterization to give an axiomatic definition for the free group and free monoid generated by a set A; recall that a *monoid* is a set with a binary operation which is associate and has a unit element. The use of monoids will be helpful for constructing free groups.

Definition for monoids. Let M be a monoid, and suppose that $X \subset M$. Then M is said to be a *free monoid on X* (or *X freely generates M as a monoid*) provided (i) X generates M (*i.e.*, the only submonoid containing X is M itself), (ii) if N is a monoid and $f : X \to N$ is a map of sets, then there is a unique extension of f to a monoid homomorphism $F : M \to N$.

Definition for groups. Let G be a group, and suppose that $X \subset G$. Then G is said to be a *free group on X* (or *X freely generates G as a group*) provided (i) X generates G (*i.e.*, the only subgroup containing X is G itself), (ii) if H is a group and $f : X \to H$ is a map of sets, then there is a unique extension of f to a group homomorphism $F : G \to H$.

Sometimes we stretch the language in these definitions a little, replacing the condition that X be a subset of the given algebraic object with an assumption that we hve a 1–1 map of sets from X to M or G. In both cases, the second property in the definition is an example of a **Universal Mapping Property**, and sometimes the extension of f is said to be determined by that property.

The first result is a uniqueness theorem for free groups and monoids.

THEOREM 1. *Let X be a set, for $i = 1, 2$ let F_i be a group or monoid, and let $j_i : X \to F_i$ be $1 - 1$ mappings. If each F_i is free on $j[X]$, then there is a unique isomorphism (of groups or monoids) $\varphi : F_1 \to F_2$ such that $j_2 = \varphi \circ j_1$.*

Proof. We first note that if F is free on $j[X]$ then the only homomorphism $h : F \to F$ such that $h \circ j = j$ is the identity; this is true because the identity on F has this property, and therefore the uniqueness condition in the definition of a free group or monoid implies that h must be equal to the identity.

Suppose now that we have a setting with in which the hypotheses of the proposition are satisfied. The universal mapping properties imply that there are homomorphisms $g_1 : F_1 \to F_2$ and $g_2 : F_2 \to F_1$ such that $g_1 \circ j_1 = g_2$ and $g_2 \circ j_2 = g_1$. Therefore we have

$$(g_2 \circ g_1) \circ j_1 \;=\; g_2 \circ (g_1 \circ j_1) \;=\; g_1 \circ j_2 \;=\; j_1 \quad \text{and} \quad (g_1 \circ g_2) \circ j_2 \;=\; g_1 \circ (g_2 \circ j_2) \;=\; g_2 \circ j_1 \;=\; j_2$$

and hence we can apply the conclusion in the first paragraph to conclude that $g_2 \circ g_1 = \mathrm{id}(\mathrm{F}_1)$ and $g_1 \circ g_2 = \mathrm{id}(\mathrm{F}_2)$. Therefore F_1 is an isomorphism and F_2 is its inverse.∎

For most but not all choices of the ring R, one can prove that if a module is free on two sets of generators A and B, then we must have $|A| = |B|$; in words, A and B have the same cardinality. For example, if R is a commutative ring with unit, then this conclusion is true (see Hungerford, page 184, for further information on such questions and a reference to an example where the conclusion is false). We shall state and prove an analogous result for free groups; namely, any two sets of free generators must have the same cardinality.

THEOREM 2. *Let G be a group, and let A and B be subsets of G such that G is freely generated by each of A and B. Then $|A| = |B|$.*

A similar result holds for monoids, and the proof for groups has a direct extension to such objects; this argument is left to the reader as an exercise.

Proof of Theorem 2. There are three basic steps.

(1) Show that A and B are both finite or both infinite.

(2) Show that if A and B are both finite. then $|A| = |B|$.

(3) Show that if A and B are both infinite. then $|A| = |B|$.

Proof of (1). If a group F is free on the set X, we claim that there are $2^{|X|}$ homomorphisms from F to $\mathbb{Z}_2$. To see this, note that every homomorphism h determines a subset of X; namely, the set of all $x \in X$ such that $h(x)$ is nontrivial. Conversely, if Y is a subset of X and $\chi : X \to \mathbb{Z}_2$ is the characteristic function of Y (1 at points of Y, and 0 at points of $X - Y$), then by the definition of a free group there is a unique group homomorphism $F \to \mathbb{Z}_2$ whose restriction to X is χ. Hence homomorphisms $F \to \mathbb{Z}_2$ are in 1–1 correspondence with subsets of X, and this means that the cardinality of the set of all such homomorphisms is $2^{|X|}$.

To prove (1), note that $2^{|X|}$ is finite if and only if $|X|$ is finite. Therefore it follows that X is finite if and only if the set of homomorphisms $F \to \mathbb{Z}_2$ is finite. Since the cardinality of the latter set does not depend upon the free generating sets A and B, it follows that the finiteness of the set of homomorphisms from F to $\mathbb{Z}_2$ is equivalent to the finiteness of A and also to the finiteness of B. Thus A is finite if and only if B is finite.∎

Proof of (2). This is just a continuation of the argument in the previous part. If F is free on the finite sets A and B, then the preceding discussion implies that $2^{|A|} = 2^{|B|}$, and since both $|A|$ and $|B|$ are finite the latter implies that $|A| = |B|$.

Proof of (3). We still ge the conclusion $2^{|A|} = 2^{|B|}$ if A and B are both infinite, but the results of W. Easton show that there are models of set theory in which this exponential equation does not imply $|A| = |B|$ (see Section I.3 at the end of the first part of the course notes), so we need to develop another approach.

CLAIM. *If X is an infinite set and F is free on X, then $|F| = |X|$.*

If this statement is true and F is free on the infinite sets A and B, then it will follow that $|A| = |F| = |B|$, and the proof of (3) will be complete. — To prove the claim, note first that $X \subset F$ implies that $|X| \leq |F|$, so it suffices to verify the reverse inequality. We shall do so using the fact that every element of F is a (finite) monomial with factors of the form x_α^ε, where $x_\alpha \in X$ and $\varepsilon = \pm 1$.

Formally, define the set of *signed monomials* $\mathcal{M}^{\pm}(X)$ in elements of X to be all infinite sequences $(a_1, a_2, \cdots)$ where each a_j is either (i) an ordered of the form (x_α, ε) as in the preceding paragraph or (ii) the term a_j is equal to 1, **AND** the following conditions are satisfied:

(A) If $a_j = 1$ for some j, then $a_k = 1$ for all $k \geq j$.

(B) There is some (first) j such that $a_j = 1$.

It follows that every signed monomial sequence satisfies $a_j = 1$ for all but finitely many j, and if $L + 1 \geq 1$ is the first value of j such that $a_j = 1$, then the first L terms of the sequence are nontrivial (*i.e.*, they are terms of the first type, which we shall write henceforth in the form x_α^ε). Note that we can have $L = 0$, in which case the sequence consists entirely of ones. More generally, if the first L terms of a monomial are nontrivial but the remaining terms are trivial, we shall say

that L is the *length* of the monomial. Each monomial has a nonnegative length, and we can use the length to partition $\mathcal{M}^{\pm}(X)$ into subsets $\mathcal{M}^{\pm}[n]$, where n runs over all the nonnegative integers.

By associativity there is a well-defined multiplication homomorphism

$$\Pi : \mathcal{M}^{\pm}(X) \longrightarrow F$$

which is formed by multiplying together the first $L+1$ elements of the sequence in the given order, and this mapping is onto since X generates F. Since Π is onto, the desired inequality $|F| \leq |X|$ will follow if we can show that the domain of Π has cardinality $\leq |X|$, and since $|X|$ is infinite it will suffice to show that each of the subsets $\mathcal{M}^{\pm}[n]$ has cardinality $\leq |X|$. But $\mathcal{M}^{\pm}[0]$ consists of a single point, and if $n > 0$ we know that

$$\mathcal{M}^{\pm}[n] \;=\; \prod^{n} (X \times \{\pm 1\}) \; .$$

But since X is infinite we know that

$$|X \times \{\pm 1\}|^n \;=\; |X|^n \;=\; |X|$$

and therefore $|\mathcal{M}^{\pm}[n]| = |X|$ if $n \geq 1$; as noted above, this and $|\mathcal{M}^{\pm}[0]| = 1$ suffice to show that $|F| \leq |X|$ and hence $|F| = |X|$.∎

Existence of free groups

In order to justify the preceding discussion, we need to prove the following result:

THEOREM 3. *Let X be a set. Then there is a monoid M and a group G, together with $1-1$ mappings $\alpha : X \to M$ and $\beta : X \to G$, such that M is freely generated by $\alpha[X]$ and G is freely generated by $\beta[X]$.*

Proof. We begin by proving the result for monoids. As in the final part of the proof of Theorem 2, define the set $\mathcal{W}(X)$ of *words* in X to be the set of all sequences $(a_1, a_2, \cdots)$ such that either $a_j \in X$ or else $a_j = 1$ for a suitable object $1 \notin X$, and conditions (A) and (B) in that proof are satisfied. Once again we can talk about words of length n for every nonnegative integer n, agree that the word $(1, 1, \cdots)$ has length zero, and write $\mathcal{W}(X)$ as a union of the pairwise disjoint subsets of words $\mathcal{W}[n]$ of length n. Given a word of length n, take its *standard truncation* to be the sequence of length n obtained by eliminating the infinite string of 1's beginning with term $n+1$. Define multiplication on $\mathcal{W}(X)$ as follows:

If $a \in \mathcal{W}[p]$ and $b \in \mathcal{W}[q]$ for $p, q > 0$, define $a \cdot b \in \mathcal{W}[p+q]$ by concatenating the standard truncation of a with the standard truncation of b, and filling in 1's at all higher values of the sequence.

If $a \in \mathcal{W}[p]$ and $b \in \mathcal{W}[q]$ where at least one of p, q is zero, define $a \cdot 1 = a$ if $q > 0$ and $1 \cdot b = b$ if $p > 0$. Note that if $p = q = 0$, then both of these constructions yield the identity $1 \cdot 1 = 1$.

One can now verify that this binary operation is associative and the word of length zero is a two-sided identity. Therefore we have a monoid with an inclusion mapping $X \cong \mathcal{W}[1] \subset \mathcal{W}(X)$, and by construction we know that X generates this monoid. In order to prove that the image of X freely generates $\mathcal{W}(X)$, we need to verify the mapping property. Suppose now that S is a monoid and we have a map of sets $j : X \to S$. Extend j to a homomorphism g on $\mathcal{W}(X)$ as follows: By associativity we have well-defined mappings from each set $\mathcal{W}[n]$ to S (where $n > 0$) given by taking

products of the coordinates (in order), and by construction this is a homomorphism of monoids. This completes the proof in the case of monoids.

The first step in constructing a free **group** on a set X is to consider the free monoid

$$\mathcal{W}\left(X \amalg X^{-1}\right)$$

on the set $X \amalg X^{-1} = X \times \{pm\, 1\}$. One then defines an equivalence relation $\mathcal{A}$ on this monoid which is generated by stipulating that a monomial

$$x_1^{\alpha(1)} \ \cdots \ x_r^{\alpha(r)} x_{r+1}^{\alpha(r+1)} \ \cdots \ x_n^{\alpha(n)}$$

(where $\alpha(j) = \pm 1$) is equivalent to the contracted monomial formed by deleting the terms $x_r^{\alpha(r)}$ and $x_{r+1}^{\alpha(r+1)}$ if $x_r = x_{r+1}$ and $\alpha(r) = -\alpha(r+1)$. Note that the cases $r = 1$ and $r + 1 = n$ are allowed. It follows that if $b \ \mathcal{A} \ b_1$ and $c \ \mathcal{A} \ c_1$ in $\mathcal{W}\left(X \amalg X^{-1}\right)$ then we also have $bc \ \mathcal{A} \ b_1 c_1$, for if we have a sequences of contractions and their opposites going from b to b_1 then we obtain a similar sequence from bc to $b_1 c$, and similarly if we have a sequence from c to c_1 we get a sequence from $b_1 c$ to $b_1 c_1$. Therefore we obtain a binary operation on the set of equivalence classes $\mathcal{G}(X)$ such that the equivalence class projection

$$\mathcal{W}\left(X \amalg X^{-1}\right) \ \longrightarrow \ \mathcal{G}(X)$$

is a monoid homomorphism. To see that this class is a group, we need to find two-sided inverses, and this is easy because if we are given a typical monomial $x_1^{\alpha(1)} \ \cdots \ x_n^{\alpha(n)}$ then an inverse for its image in $\mathcal{G}(X)$ is given by $x_n^{-\alpha(n)} \ \cdots \ x_1^{-\alpha(1)}$.

Clearly the image of X generates $\mathcal{G}(X)$ as a group, so all that remains is to prove the Universal Mapping Property. Suppose that $h : X \to H$ is a map of sets and K is a group. Since H is a monoid, the Universal Mapping Property for free monoids implies that there is a unique homomorphism of monoids H_0 from $\mathcal{W}\left(X \amalg X^{-1}\right)$ to K such that $H_0(x) = h(x)$ and $H_0(x^{-1}) = h(x)^{-1}$ for all $x \in X$. We claim that if $b \ \mathcal{A} \ c$ then $H_0(b) = H_0(c)$, so that H_0 passes to a homomorphism H from the quotient set $\mathcal{G}(X)$ to K; the proof of the claim reduces to considering pairs b, c such that one is obtained by a simple contraction of the other, and clearly if b and c satisfy this condition then $H_0(b) = H_0(c)$ by construction. Finally, this homomorphism must be unique because X generates $\mathcal{G}(X)$ as a group.∎

If A is a set and M is a module over the ring R which is freely generated, then every element of M has a canonical and unique expansion as a linear combination $\sum r_a\, a$, where a runs through the elements of A and only finitely many coefficients r_1 are nonzero. We would like to have analogous expansions for the elements of a free group on a set X which are canonical and unique; these will be described in the next section as special cases of more general results.

Quotients of free groups

The following result is simple but fundamentally important:

PROPOSITION 4. *Every group G is isomorphic to a quotient of a free group F.*

Proof. Let F be the free group on G. Then the identity mapping from G to itself has a unique extension to a homomorphism $\varphi : F \to G$. Since the restriction of φ to G is the identity, it follows that φ is onto. Therefore $G \cong F/H$ if we take H to be the kernel of φ.∎

Using this result we can formulate a nonabelian analog of finitely generated abelian groups.

Definition. A group G is said to be *finitely presented* if it is isomorphic to a quotient F/R, where F is a finitely generated free group and R is a subgroup which is normally generated by some finite subset.

The condition on R means that there is a finite subset E such that R is generated by E together with its conjugates in F. We are not assuming that R is algebraically generated as a group by a finite subset.

Examples. 1. If G is a finite group, then G is finitely presented, for we can let F be the free group on G and take R to be normally generated by all monomials xyz^{-1}, where $x, y \in G$ and z is the product of x and y in G (the proof of this is left as an exercise).

2. If G is a finitely generated abelian group, then G is finitely presented. — The quickest way to do this is to use the fact that G is isomorphic to a product of cyclic groups. Specifically, assume that the direct summands are generated by elements a_j, and let m_j be equal to the order of a_j if a_j has finite order. Then G is isomorphic to the quotient of the free group F with free generators a_j by the subgroup R which is normally generated by the elements $a_j^{m_j}$ such that a_j has finite order m_j in G together with the commutators of the form $a_i a_j a_i^{-1} a_j^{-1}$ where a_i and a_j run through all the elements of A.

3. The fundamental group of a Klein bottle is also finitely presented. It is a quotient F/R, where F is free on two generators x, y and R is normally generated by $xyx^{-1}y$.

4. A frequently cited paper of G. Baumslag (*Mathematische Zeitschrift* **75** (1960/1961). pp. 22–28) gives standard examples of finitely generated groups which are not finitely presented.

Subgroups of free groups. A fundamental result in graduate algebra courses states that a subgroup A of a free abelian group F is free, and in fact the cardinality of a set of free generators for A is less than or equal to the cardinality of a set of free generators for F. There are similar but not quite identical results for free groups. In fact, topological methods provide a very illuminating proof that a subgroup of a free group is free, and this is worked out in Section 85 of Munkres (it uses material not covered in this course). There is also an analogous result on the cardinality of a set of free generators, *but it is only valid for* infinite *sets of free generators*. In fact, Theorem 85.3 on page 515 of Munkres shows that if F is a free group on a finite set of k generators and S is a subgroup of F, then S may be free on a larger number of generators (one can also show that if F is a free group on 2 or more generators, then it can have subgroups with infinitely many generators, but this set must be countable). Further information on some of these statements is given in the exercises.

IX.2 : Sums and pushouts of groups

(Munkres, 68; Hatcher, 1.2)

As noted at the beginning of Section 68 in Munkres, a free product of two groups should be an object which is somehow analogous to the direct sum of two abelian groups, or more generally the direct sum of two modules over some ring. However, there is also one substantial difference between a free product of two groups and a direct sum of two modules. The latter objects can

also be viewed algebraically as direct products of two modules, but a direct product and a free product of two groups turn out to be much different from each other. There is a better analogy with constructions on sets; the free product of two groups is to their direct product much like the disjoint union of two sets is to their cartesian product. We shall elaborate upon this when we state our formal definition of free products.

Another algebraic reference for material in this section is Section I.9 in Hungerford.

More generally, in analogy with the concept of direct products, it is often necessary to consider free products of arbitrary indexed families of groups. However, unlike the case of direct products, the explicit construction is awkward and tedious in some respects, and therefore it is convenient to split things into two parts:

(1) Formulating the mapping properties of free products which characterize them up to isomorphism.

(2) Constructing — by brute force if necessary — an algebraic system which has these mapping properties.

In fact, our construction will exhibit some important facts about free products, including a unique factorization result for the elements of such groups (compare the issue regarding free groups that was mentioned in the preceding section).

Definition and uniqueness properties

Definition. Let $\{G_\alpha \mid \alpha \in A\}$ be an indexed family of groups. The data for a *free product* of the indexed family consist of a group G and homomorphisms $i_\alpha : G_\alpha \to G$ (for each $\alpha \in A$) such that the following **Universal Mapping Property** is satisfied:

> *If H is a group and $f_\alpha : G_\alpha \to H$ is an indexed family of homomorphisms, then* there is a unique group homomorphism *$F : G \to H$ such that $F \circ i_\alpha = f_\alpha$ for all $\alpha \in A$.*

The characterization of free products can be viewed as "dual" to the following characterization of direct products:

PROPOSITION 0. *Let be a family of groups, let $P = \prod_\alpha G_\alpha$ be their direct product, and for each α let $\pi_\alpha : P \to P_\alpha$ denote the appropriate coordinate projection. Assume also that we are given data consisting of a group Q and homomorphisms $q_\alpha : Q \to G_\alpha$ with the following universal mapping property:*

> *If H is a group and $h_\alpha : H \to G_\alpha$ is an indexed family of homomorphisms, then there is a unique group homomorphism $\Gamma : H \to Q$ such that $q_\alpha \circ \Gamma = h_\alpha$ for all $\alpha \in A$.*

THEN *there is a unique isomorphism $\Phi : Q \to P$ such that $q_\alpha \circ \Phi = \pi_\alpha$ for all α.*

Proof. We begin by noting that the data given by P and the homomorphisms π_α satisfy the universal mapping property in the proposition, for a maps into products are uniquely determined by their coordinate projections (and conversely, a family of maps into the factors determines a map into the product whose coordinate functions are given by the original family).

Since the data involving P satisfy the universal mapping property, there is a unique mapping $\Phi : Q \to P$ such that $q_\alpha = \pi_\alpha \circ \Phi$ for all α,and since the data involving Q also satisfy this property, we obtain a unique mapping $\Psi : P \to Q$ such that $\pi_\alpha = q_\alpha \circ \Psi$ for all α. The composites of these mappings satisfy

$$\pi_\alpha \circ \Phi \circ \Psi \;=\; \pi_\alpha \;, \qquad q_\alpha \circ \Psi \circ \Phi \;=\; q_\alpha$$

for all α, so by the uniqueness part of the universal mapping property we must have $\Psi \circ \Phi = \mathrm{id}_Q$ and $\Phi \circ \Psi = \mathrm{id}_P$. These combine to show that Φ is an isomorphism which satisfies the required identities.∎

Note that the last part of the preceding proof was very similar to the proof of Theorem 1.1 in this unit, and in fact both are special cases of more general uniqueness results for data which have suitable universal mapping properties (For the sake of completeness, we note that such data generally determine initial or terminal objects of some category, and this is the source of the unique isomorphisms because there is a unique isomorphism between two initial or terminal objects in a category; further information and background can be found on pages 42 and 50–51 of the following survey article, which is out of date in many places, but still gives a clear account of many basic concepts in the first few sections.).

> **Mac Lane, S.** Categorical algebra. *Bulletin of the American Mathematical Society* **71** (1965), 40–106. (This article is freely available online)

Although we shall not need to understand the general phenomena in these notes, but it still instructive to note the parallels between the proof of Proposition 0 and our first result on free products; the wording in one result is essentially the mirror image of the wording in the other.

THEOREM 1. *Let $\{G_\alpha \mid \alpha \in A\}$ be an indexed family of groups, and assume that we are given two sets of data*

$$i_\alpha : G_\alpha \longrightarrow S \ , \qquad j_\alpha : G_\alpha \longrightarrow T$$

which satisfy the conditions for a free product. Then there is a unique isomorphism $F : S \to T$ such that $j_\alpha = F \circ i_\alpha$ for all α.

Proof. Since both sets of data involving S satisfy the universal mapping property, there is a unique mapping $H : T \to S$ such that $H \circ j_\alpha = i_\alpha$ for all α,and since the data involving T also satisfy this property, we obtain a unique mapping $F : S \to T$ such that $F \circ i_\alpha = j_\alpha$ for all α. The composites of these mappings satisfy

$$H \circ F \circ i_\alpha \ = \ j_\alpha \ , \qquad F \circ H \circ j_\alpha = i_\alpha$$

for all α, so by the uniqueness part of the universal mapping property we must have $H \circ F = \mathrm{id}_S$ and $F \circ H = \mathrm{id}_T$. These combine to show that F is an isomorphism which satisfies the required identities.∎

The usefulness of the preceding material depends upon showing that every family of groups has a free product. We shall begin by proving that a free group is a free product of infinite cyclic groups, with one for each element in a set of free generators for the group.

THEOREM 2. *Let G be a group, let $X \subset G$ be a nonempty subset such that G is a free group on X, and for each $x \in X$ let X denote the cyclic group generated by x. Then G together with the inclusions $i_x : H_x \subset G$ present G as a free product of the subgroups H_x.*

Proof. We first show that each subgroup H_x is infinite cyclic. Consider the map from X to $\mathbb{Z}$ which sends x to 1 and all other elements of X to 0; by the definition of a free group, this map extends to a homomorphism $h_x : G \to \mathbb{Z}$. Since $h_x(x)$ has infinite order and homomorphisms preserve elements of finite order, it follows that x itself has infinite order, so that H_x is an infinite cyclic group.

Suppose now that Γ is a group and we have homomorphisms $f_x : H_x \to \Gamma$. Since G is free on X there is a homomorphism $F : G \to \Gamma$ such that $F(x) = f_x(x)$ for each $x \in X$. To complete the

proof, we need to verify that $F \circ i_x(y) = F(y) = f_x(y)$ for each $y \in H_x$ (the first equation reflects the fact that i_x denotes the inclusion of a subgroup).

If $y \in H_x$, then by the first paragraph we have $y = x^m$ for some unique integer m. Since homomorphisms take m^{th} powers to m^{th} powers, it follows that

$$F \circ i_x(y) \;=\; F(y) \;=\; F(x^m) \;=\; [F(x)]^m \;=\; [f_x(x)]^m \;=\; f_x(x^m) \;=\; f_x(y)$$

and therefore $F \circ i_x = f_x$, which is what we needed to verify in order to complete the proof.∎

In an ordinary direct sum, the images of the various submodules are isomorphic to the original objects and for every pair of distinct summands, the intersection of their images is the trivial subgroup. It will be helpful to have an analog of these facts for free products:

PROPOSITION 3. *Let $\{G_\alpha \mid \alpha \in A\}$ be an indexed family of groups, and assume that we are given data*

$$i_\alpha : G_\alpha \;\longrightarrow\; G \;, \qquad j_\alpha : G_\alpha \;\longrightarrow\; T$$

which satisfy the conditions for a free product. Then each homomorphism i_α is injective, and for each pair of distinct indices $\beta \neq \gamma$ we have $i_\beta[G_\beta] \cap i_\gamma[G_\gamma] \;=\; \{1\}$.

Proof. To prove that each i_α is injective, consider the family of homomorphisms $f_\beta : G_\beta \to G_\alpha$ such that f_α is the identity but f_β is the trivial homomorphism if $\beta \neq \alpha$. Let $F : G \to G_\alpha$ be a homomorphism such that $f_\beta = F \circ i_\beta$ for all β. Then $F \circ i_\alpha$ is the identity on G_α, and hence this homomorphism is 1–1 on G_α; but this means that i_α must also be 1–1.

To prove that the intersections of the various subgroups are independent, start with the setting of the preceding paragraph and let $\beta \neq \alpha$. If $x \in i_\beta[G_\beta] \cap i_\alpha[G_\alpha]$, we may write it in the form $x = i_\beta(x_\beta) = i_\alpha(x_\alpha)$. Consider what happens if we apply F. Since $F \circ i_\beta$ is the trivial homomorphism we have $F(x) = 1$, but as above we also have $x_\alpha = F(x)$. Therefore $x_\alpha = 1$, which means that $x = i_\alpha(1) = 1$.∎

Construction of free products

Here is a relatively short existence proof:

THEOREM 4. *If $\{G_\alpha \mid \alpha \in A\}$ is a nonempty indexed family of groups, then there exist data consisting of a group S and homomorphisms $i_\alpha : G_\alpha \to S$ which present S as a free product of the groups G_α.*

Proof. Let M denote the free monoid on the set

$$X \;=\; \amalg_\alpha \, G_\alpha$$

so that X is a union of pairwise disjoint copies of the groups G_α. Define $\mathcal{A}$ to be the equivalence relation on M generated by the following two relations:

(*i*) Contraction to remove an identity map from a sequence or word in M. Symbolically, this can be written in the form $(\cdots, 1_\alpha, \cdots) \;\mathcal{A}\; (\cdots, , \cdots)$, where 1_α is the identity for G_α.

(*ii*) Contraction to multiply consecutive terms if they lie in the same group G_α. Symbolically, this can be written in the form $(\cdots, a, b, \cdots) \;\mathcal{A}\; (\cdots, ab, \cdots)$, where $a, b \in G_\alpha$ for some α.

Let S be the set of $\mathcal{A}$-equivalences in M. As in the proof of Theorem 1.3 we can verify directly that if $b, c \in M$ and $b_1, c_1 \in M$ are such that either b_1 is obtained from b by a contraction of

either type as above or c_1 is obtained from c in this fashion, then then b_1c_1 is similarly obtained from bc. Since these contraction relations generate $\mathcal{A}$, it follows that the quotient set S inherits a monoid operation such that the quotient projection $\rho : M \to S$ is a homomorphism of monoids. The existence of inverses follows because if $(a_1, \ \cdots \ , a_n) \in S$ then we have

$$\rho(a_1, \ \cdots \ , a_n)^{-1} \quad = \quad \rho\left(a_n^{-1}, \ \cdots \ , a_1^{-1}\right)$$

as in the proof of Theorem 1.3.

Let ε_α denote the embedding of G_α in M as the set of length 1 sequences whose nontrivial term lies in G_α, and define j_α to be the composite $G_\alpha \to M \to S$. Then by construction the mapping j_α is a homomorphism (this is slightly less trivial than it looks; see the exercises for details). We claim that the data S and $\{j_\alpha \mid \alpha \in A\}$ present S as a free product.

Suppose that we are given group homomorphisms $f_\alpha : G_\alpha \to \Gamma$ for some group Γ. Since M is a free monoid, there is a unique monoid homomorphism $\varphi : M \to \Gamma$ such that $\varphi \circ \varepsilon_\alpha = f_\alpha$. It now follows that φ can be factored through the quotient map ρ; *i.e.*, there is a homomorphism $f : S \to \Gamma$ such that $f \circ \rho = \varphi$ and $f \circ j_\alpha = f \circ \rho \circ \varepsilon_\alpha = \varphi \circ \varepsilon_\alpha = f_\alpha$.

To complete the proof, we need to show that the only homomorphism h such that $h \circ j_\alpha = f_\alpha$ for all α is equal to the mapping f described in the preceding paragraph. Clearly the latter will hold if the images of the homomorphisms j_α generate S as a group. But this follows from the facts that the images of the embeddings ε_α generate S together with the identities $j_\alpha = \rho \circ \varepsilon_\alpha$.∎

Notation. The universal mapping properties for free products of groups are completely analogous to the corresponding properties of disjoint unions of sets in topological spaces which appeared in Unit V. They dualize the mapping properties for direct products in the sense that the directions of all arrows representing functions are reversed, and from the viewpoint of category theory they are examples of **coproducts**. Such constructions are frequently denoted by symbolism like

$$\coprod_{\alpha \in A} G_\alpha$$

where the coproduct symbol $\coprod$ is merely the product symbol $\prod$ turned upside down. Free products for finite indexed sets of groups are sometimes also denoted by symbolism like

$$G_1 \ * \ \cdots \ * \ G_n \qquad \text{or} \qquad *_{i=1}^n \ G_i \ .$$

Before proceeding, we note that the construction yields another basic property of free products.

COROLLARY 5. *Let $\{G_\alpha \mid \alpha \in A\}$ be an indexed family of groups, and assume that we are given free product data $i_\alpha : G_\alpha \ \to S$. Then S is generated by the union of the image $\cup_\alpha \ i_\alpha[G_\alpha]$.*

Proof. By the uniqueness of free group data up to isomorphism, it suffices to verify this for an explicit construction, and this property follows immediately from the construction in the theorem.∎

Unique factorization in free products. The proofs of Theorem 4 in both Munkres and Hungerford are designed to yield stronger conclusions; specifically, one of their objectives is to prove the following unique decomposition (or factorization) result for elements of a free product. In Hungerford the approach is more direct but the argument is only sketched (see Theorem 9.1 on page 65), while the less direct approach in Munkres is mathematically complete (see page 418, beginning with line 3). We shall simply state the conclusion without proof in these notes.

THEOREM 6. (Unique factorization in free products) *Suppose we have a group G and data $\{i_\alpha : G_\alpha \to G \mid \alpha \in A\}$ which present G as a free product of the groups G_α. Then every nontrivial element in G has a unique factorization of the form $a_1 \cdot \cdots \cdot a_n$, where each a_j is nontrivial and lies in the image of some $G_{\alpha(j)}$ for each j, but two consecutive factors a_k and a_{k+1} always lie in the images of different subgroups.*∎

This result yields the unique factorization result for free groups which was mentioned in the preceding section.

COROLLARY 7. (Unique factorization in free groups) *Suppose that the group G is free on the set X. Then every nontrivial element of G has a unique factorization of the form $\prod_\alpha x_j^{m_j}$ where each $x_j \in X$, each exponent m_j is a nonnegative integer, and $x_j \neq x_{j+1}$ for all j.*

Proof. By Theorem 2 we know that G is a free product of the infinite cyclic groups H_x, where H_x is generated by x. Since each nontrivial element of H_x has the form x^m for some unique $m \neq 0$, this reduces the corollary to a special case of Theorem 5.∎

A curious example

Page 22 of Hatcher describes a property of the free product $\mathbb{Z}_2 * \mathbb{Z}_2$ that is not immediately obvious but turns out to be quite important in certain contexts. Namely, this group is isomorphic to the *infinite dihedral group* D_∞ which has two generators x and y such that $x^2 = 1$ and $xyx^{-1} = y^{-1}$. More explicitly, if we take u_1 and u_2 to be the nontrivial elements in the first and second copies of $\mathbb{Z}_2$, then we can let $x = u_1$ and $y = u_1 u_2$. The element y generates an infinite cyclic subgroup which has index 2 and (hence) is normal. Every element of this group has a unique description as a product $x^\varepsilon y^n$, where n is an integer and $\varepsilon = 0$ or ± 1. The reason for the name involves the ordinary dihedral groups of order $2n$, where $n \geq 3$ is an integer. These groups are the subgroups of the group $\mathbf{O}(2)$ of 2×2 orthogonal matrices which send the standard regular n-gon — which has vertices $\exp(2\pi \mathbf{i} k/n)$, where $1 \leq k \leq n$ — to itself. Generators for this group are given by the matrix A which acts by counterclockwise rotation through an angle of $360/n$ degrees, and the matrix B which acts by reflection with respect to the x-axis. These two matrices satisfy the relation $BAB^{-1} = A^{-1}$, and in fact the dihedral group of order $2n$ is isomorphic to a quotient of D_∞ via the map sending x to B and y to A (its kernel is the subgroup generated by y^n). Note that if $n = 2$ the analogous group is just the Klein Four Group $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Remark. A striking and fundamental result of A. G. Kurosh gives an elegant description of the subgroups of a free product. One proof of this result (using topological constructions as in this course!) is given on pages 392–393 of J. Rotman, *An Introduction to the Theory of Groups* (Fourth Ed., Springer-Verlag, 1995). A purely algebraic proof is presented on pages 315–319 of M. Hall, *Theory of Groups*, Macmillan, 1959).

Pushout diagrams

As in Unit V, there are many situations in which one wants to construct a space X which is the union of two subspaces A and B. If we wish to analyze the fundamental group of such a space, it is natural to ask whether the fundamental group of X can be built in some similar fashion using the fundamental groups of the pieces. In fact, the main result of Section 3 vindicates this idea if the relevant subspaces are always open and arcwise connected. We shall conclude this section by formulating a general method for approaching the underlying group-theoretic question, where the

pieces are subgroups and we have some information on how they intersect. If the intersection is the trivial group, then everything will reduce to the free product construction.

We shall begin by reformulating a topological space decomposition $X = A \cup B$ in terms of commutative squares and universal mapping properties. If A and B are both open subsets or both closed subsets of X, then we know that a a pair continuous functions $f_A : A \to Y$ and $f_B : B \to Y$ can be pieced together to form a continuous function on X if and only if $f_A|A \cap B = f_B|A \cap B$. This can be reformulated in terms of morphism diagrams and universal mapping properties as indicated below; none of this yields anything new for topological spaces, but it provides a key for formulating the corresponding problem involving groups.

Universal Mapping Property for $X = A \cup B$. *Let X be a space which is presented as a union of two subspaces A and B, where both are either open in X or closed in X. Then we have a commutative diagram*

$$\begin{array}{ccc} A \cap B & \xrightarrow{i_A} & A \\ \downarrow{\scriptstyle i_B} & & \downarrow{\scriptstyle j_A} \\ B & \xrightarrow{j_B} & X \end{array}$$

in which all the mappings are subspace inclusions. Furthermore, if we are given a pair of continuous mappings $f_A : A \to Y$ and $f_B : B \to Y$ such that the diagram

$$\begin{array}{ccc} A \cap B & \xrightarrow{i_A} & A \\ \downarrow{\scriptstyle i_B} & & \downarrow{\scriptstyle f_A} \\ B & \xrightarrow{f_B} & Y \end{array}$$

commutes (in other words, $f_A \circ i_A = f_B \circ i_B$), then there is a unique $f : X \to Y$ such that $f \circ j_A = f_A$ and $f \circ j_B = f_B$.∎

In the language of category theory, one says that X is a *pushout* of the diagram

$$B \longleftarrow A \cap B \to A$$

and is the universal way of constructing a commutative square as above.

Definition. Let $\mathcal{C}$ be a category (if this is two abstract, take the category of groups and homomorphisms), and suppose that we are given a commutative square

$$\begin{array}{ccc} D & \xrightarrow{i_A} & A \\ \downarrow{\scriptstyle i_B} & & \downarrow{\scriptstyle j_A} \\ B & \xrightarrow{j_B} & X \end{array}$$

in $\mathcal{C}$. This diagram is said to be a *pushout diagram* if for every commutative square in $\mathcal{C}$ of the form

$$\begin{array}{ccc} D & \xrightarrow{i_A} & A \\ \downarrow{\scriptstyle i_B} & & \downarrow{\scriptstyle f_A} \\ B & \xrightarrow{f_B} & Y \end{array}$$

(in other words, $f_A \circ i_A = f_B \circ i_B$), there is a unique $f : X \to Y$ such that $f \circ j_A = f_A$ and $f \circ j_B = f_B$.

As in previous discussions, if the two squares are identical, the universal mapping property implies that the identity is the only map $\Phi : X \to X$ such that $\Phi \circ j_A = j_A$ and $\Phi \circ j_B = j_B$. Similarly, we have the following uniqueness result for pushout diagrams.

PROPOSITION 8. *If we are given two pushout diagrams for $B \leftarrow D \to A$ of the form*

$$\begin{array}{ccc} D & \xrightarrow{i_A} & A \\ \downarrow{\scriptstyle i_B} & & \downarrow{\scriptstyle j_A} \\ B & \xrightarrow{j_B} & X \end{array} \qquad\qquad \begin{array}{ccc} D & \xrightarrow{i_A} & A \\ \downarrow{\scriptstyle i_B} & & \downarrow{\scriptstyle k_A} \\ B & \xrightarrow{k_B} & Y \end{array}$$

then there is a unique isomorphism $H : X \to Y$ such that $k_A = H \circ j_A$ and $k_B = H \circ j_B$.

Exactly as in previous arguments involving universal mapping properties, the mapping H and its inverse are given by the Universal Mapping Property for pushout diagrams.∎

Examples. We shall describe some special cases of pushouts in the category of groups. In each case we assume that X is a pushout for $B \leftarrow D \to A$. Verifications are left to the reader as exercises.

1. If $D \to A$ and $D \to B$ are trivial homomorphisms, then X is trivial.

2. If D is the trivial group then X is isomorphic the free product of A and B.

3. If A is trivial then X is isomorphic the quotient of B and by the subgroup which is normally generated by the image of D.

4. If $D \to A$ is an isomorphism then so is $B \to X$.

There are additional exercises which consider more specific types of examples.

Construction of group pushouts

We have noted that free products are special cases of pushouts, and the existence of arbitrary pushouts for groups will be shown using the existence of free products.

THEOREM 9. *Every diagram of group homomorphisms*

$$\begin{array}{ccc} H & \xrightarrow{h_1} & G_1 \\ \downarrow{\scriptstyle h_2} & & \\ G_2 & & \end{array}$$

has a pushout.

Proof. Let $G_1 * G_2$ be a free product of G_1 and G_2, and let $i_1 : G_1 \to G_1 * G_2$ and $i_2 : G_2 \to G_1 * G_2$ denote the inclusion mappings which are part of the free product structure. Take N to be the subgroup of $G_1 * G_2$ which is normally generated by all elements of the form

$$k(x) \;\; = \;\; i_1 h_1(x) \cdot (i_2 h_2(x)\,)^{-1}$$

where x runs through all the elements of H, and let $q : G_1 * G_2 \to (G_1 * G_2)/N$ denote the quotient group projection; it follows immediately that

$$q \circ i_1 \circ j_1 \;\; = \;\; q \circ i_2 \circ j_2 \; .$$

This means that we have a commutative square

$$\begin{array}{ccc} H & \xrightarrow{h_1} & G_1 \\ \big\downarrow h_2 & & \big\downarrow j_1 \\ G_2 & \xrightarrow{j_2} & (G_1 * G_2)/N \end{array}$$

in which $j_t = q \circ i_t$ for $t = 1, 2$. We claim that it is a pushout diagram of groups.

To prove that we have a pushout, suppose that we are given homomorphisms $f_t : G_t \to \Gamma$ (where $t = 1, 2$) such that $f_1 \circ h_1 = f_2 \circ h_2$. Then the existence of a suitable homomorphism $F : (G_1 * G_2)/N \to \Gamma$ can be shown as follows: There is a homomorphism F_0 from the free product $G_1 * G_2$ to Γ such that $F_0 \circ i_t = f_t$ for $t = 1, 2$. We claim that N is contained in the kernel of F_0. It suffices to show that F_0 is trivial on a set that normally generates N, and so the proof of the claim reduces to showing that $F_0 \circ k(x)$ is trivial, where $k(x)$ is defined as above for $x \in H$. Explicit computation yields the chain of equalities

$$F_0 \circ k(x) \;=\; F_0 \left(i_1 h_1(x) \cdot (i_2 h_2(x)\,)^{-1} \right) \;=$$

$$F_0\,(i_1 h_1(x)) \cdot F_0\,(i_2 h_2(x)\,)^{-1} \;=\; f_1 h_1(x) \cdot f_2 h_2(x)^{-1} \;=\; 1$$

where the next to last equation follows from $F_0 \circ i_t = f_t$ and the last equation follows because $f_1 \circ h_1 = f_2 \circ h_2$. This means that we have a factorization $F_0 = F \circ q$, and to complete the proof of existence we need to show that $F \circ j_t = f_t$ for $t = 1, 2$. But this follows from the chain of equations

$$F \circ j_t \;=\; F \circ q \circ i_t \;=\; F_0 \circ i_t \;=\; f_t$$

and completes the proof that a suitable homomorphism exists.

Finally, to prove uniqueness, note that if Φ is any homomorphism such that $\Phi \circ j_t = f_t$ for $t = 1, 2$, Then Φ and F agree on the images of G_1 and G_2 in $(G_1 * G_2)/N$, so it will suffice to verify that the images of the first two groups generate the third group. But this follows because G_1 and G_2 generate the free product by Corollary 5 and a quotient projection is onto by construction.∎

Amalgamation along a common subgroup

Frequently one is interested in pushout diagrams for $G_1 \leftarrow H \rightarrow G_2$ such that both maps are injective, and in these cases one often says that the pushout is an *amalgamated product of* G_1 *and* G_2 *with respect to* H and denotes the pushout group by $G_1 *_H G_2$. In such cases the associated map

$$H \;\longrightarrow\; G_i \;\longrightarrow\; G_1 *_H G_2$$

(which is the same mapping for $i = 1$ or 2) is injective, and in fact one has the following canonical description of classes in the amalgamated product:

THEOREM. *Assume that we are given an amalgamated product as above, and choose elements* $z_\alpha \in G_1 \amalg G_2$ *(the disjoint union of the groups, which is just a set) such that* $z_\alpha \neq 1$ *for all* α*, the cosets* Hz_α *exhaust* G_1 *and* G_2*, and no two of these cosets are equal. Then every element of* $G_1 *_H G_2$ *has a unique factorization of the form* $hz_1 \cdots z_t$*, where* $h \in H$ *and each* z_j *lies in the set* $Z = \{z_\alpha \mid \alpha \in A\}$*.*

A proof of this result is described on page 314 of the previously cited book by Hall. As noted in that text, the proof is similar to the earlier unique factorization results but lengthier (see also pages 401–406 of the previously cited book by Rotman). Since the result is not needed in the course, we shall not present a proof in these notes.

IX.3 : The Seifert-van Kampen Theorem

(Munkres, 68; Hatcher, 1.2)

REMARK ON THE DEFAULT HYPOTHESIS. In this section we do not make any assumption that the spaces under consideration are Hausdorff or locally arcwise connected.

The main result of this section shows that, in some special but fundamentally important cases, a pushout of topological spaces induces a pushout diagram of fundamental groups. A more precise version of the main result is given below. In the literature this result is also frequently known simply as van Kampen's Theorem.

THEOREM 1. (Standard version of the Seifert-van Kampen Theorem) *Suppose that X is the union of two open subsets $U \cup V$, and assume that U, V and $U \cap V$ are all arcwise connected and contain the point p. Then the induced commutative diagram of fundamental groups*

$$\begin{array}{ccc} \pi_1(U \cap V, p) & \xrightarrow{i_{1*}} & \pi_1(U, p) \\ \downarrow i_{2*} & & \downarrow j_{1*} \\ \pi_1(V, p) & \xrightarrow{j_{2*}} & \pi_1(X, p) \end{array}$$

is a pushout diagram.

Remarks on possible extensions of this result. 1. This theorem fails systematically if $U \cap V$ is not arcwise connected. The simplest counterexample is $X = S^1$ with $U = S^1 - \{1\}$ and $V = S^1 - \{-1\}$ (recall that each of the latter is homeomorphic to $\mathbb{R}$), in which case both $\pi_1(U)$ and $\pi_1(V)$ are trivial but $\pi_1(X)$ is infinite cyclic. More generally, by Proposition III.1.4 in the more advanced document

`http://math.ucr.edu/~res/math246A/advancednotes.pdf`

(see page 61) if the intersection is not arcwise connected then *the subgroup of $\pi_1(X)$ generated by $\pi_1(U)$ and $\pi_1(V)$ always has* infinite index *in $\pi_1(X)$.* — In particular, if $U \cap V$ has $k+1$ connected components, then the proposition implies that there is a homomorphism from $\pi_1(X)$ onto $\mathbb{Z}^k$.

2. The document `polishcircleB.pdf` describes an example showing that the analog of the Seifert-van Kampen Theorem for closed subspaces is false; in fact, we can take X to be a subset of the plane which is homeomorphic to $S^1 \times [0, 1]$. On the other hand, there are many important examples in which X is a union of closed subspaces A and B such that A, B and $A \cap B$ are all arcwise connected and the analog of the Seifert-van Kampen Theorem is valid. In the most frequently seen cases, one knows that there are open neighborhoods U and V of A and B such that (*i*) A is a deformation retract of U, (*ii*) B is a deformation retract of V, and (*iii*) $A \cap B$ is a deformation retract of $U \cap V$. This generalization follows from the validity of the result for the

fundamental groups of groups of X, U, V and $U \cap V$ together with the fact that the inclusion maps $A \subset U$, $B \subset V$ and $A \cap B \subset U \cap V$ are all homotopy equivalences.

3. Section 1.2 of Hatcher contains a more general result in which X is a union of more than two arcwise connected open subsets, but we shall only consider a union of two subsets because the proof for two subsets is already fairly complicated and this simpler case is already strong enough to yield many far-reaching implications.

4. At the end of this section we shall comment on generalizations of the Seifert-van Kampen Theorem to situations where the intersection is not arcwise connected.

Setting up the proof of Theorem 1. Let Γ be the pushout of the diagram

$$\pi_1(U,p) \;\longleftarrow\; \pi_1(U \cap V, p) \;\longrightarrow\; \pi_1(V,p) \; .$$

Since the fundamental group is part of a covariant functor on pointed spaces, it follows that there is a unique homomorphism $\varphi : \Gamma \to \pi_1(X,p)$ such that the composites

$$\pi_1(U) \;\longrightarrow\; \Gamma \;\longrightarrow\; \pi_1(X) \; , \qquad \pi_1(V) \;\longrightarrow\; \Gamma \;\longrightarrow\; \pi_1(X)$$

are the homomorphisms induced by the inclusions j_U and j_V. The first and easier part of the proof is to show that φ is onto, and the second part of the proof is to show that φ is 1–1; the latter is considerably less transparent.

Proof of surjectivity

It will be useful to introduce some notation. If $u, v \in [0,1]$ are such that $u < v$, let $\lambda_{u,v}$ denote the unique increasing linear mapping of the real line which sends $[0,1]$ to $[u,v]$.

For each point $x \in X$, choose a curve α_x joining the basepoint p to x such that the following conditions are satisfied:

(1) If $x = p$, then α_x is the constant curve.

(2) If $x \in U \cap V$, then α_x is a curve in $U \cap V$.

(3) If $x \in U$ but not in $U \cap V$, then α_x is a curve in U.

(4) If $x \in V$ but not in $U \cap V$, then α_x is a curve in V.

We can find these curves in each case because U, V and $U \cap V$ are all assumed to be arcwise connected.

Let γ be a closed curve in X. Since $\{U, V\}$ is an open covering of X, by a Lebesgue number argument there is some positive integer n such that γ maps each interval $[(k-1)/n, k/n]$ into either U or V for $k = 1, \cdots, n$. Define γ_k to be the composite of $\gamma \,|\, [k/n, (k+1)/n]$ with the linear mapping $\lambda_{k/n,(k+1)/n}$, and set $q_k = \gamma_{k/n}$; note that $q_0 = q_n = p$. We then have a chain of equations involving basepoint and endpoint preserving homotopy classes as follows, in which C_x denotes the constant curve for the point x (see also the drawing in the file `svk-fig1.pdf`):

$$[\gamma] \;=\; \left[\sum_k \gamma_k\right] \;=\; \left[C_p \;+\; \left(\sum_k \gamma_k + C_{q_k}\right)\right] \;=$$

$$\left[\alpha_p \ + \ \left(\sum_k \ \gamma_k + (-\alpha_{q_k} + \alpha_{q_k})\right)\right] \ = \ \left[\left(\sum_k \ \alpha_{q_{k-1}} + \gamma_k + (-\alpha_{q_k})\right) \ + \ \alpha_p\right]$$

Each of the terms in the summation defines a closed curve in either U or V by the preceding discussions and the conditions on the curves α_q, and from this it follows that the images of $\pi_1(U)$ and $\pi_1(V)$, which generate the pushout by construction, will also generate $\pi_1(X)$. This means that the induced homomorphism from Γ to $\pi_1(X)$ is onto.∎

Proof of injectivity

Let θ_U and θ_V denote the homomorphisms from $\pi_1(U)$ and $\pi_1(V)$ into the pushout Γ which are part of the basic pushout structure. We shall follow the approach in

`http://www.math.jhu.edu/~jmb/note/vanK.pdf`

and define a homomorphism σ from $\pi_1(X)$ into Γ. More precisely, given a closed curve γ in X which starts and ends at p, we shall use the ideas in the surjectivity proof to define a class $S(\gamma) \in \Gamma$ as follows:

There are many partitions of the unit interval

$$0 \ = \ t_0 \ < \ \cdots \ < \ t_n \ = \ 1$$

with break points t_k such that γ maps each subinterval $[t_{k-1}, t_k]$ into either U or V; in particular, a Lebesgue number argument implies this is true if each interval has length less than some positive constant δ, but we are not insisting that this stronger condition be satisfied. Define the curves α_k joining the basepoint p to $\gamma(t_k)$ as in the proof of surjectivity, and take γ_k to be the curve obtained by restricting γ to the subinterval $[t_{k-1}, t_k]$. Now define a mapping which sends

$$\gamma \ \simeq \ \left(\sum_k \ \alpha_{q_{k-1}} + \gamma_k + (-\alpha_{q_k})\right)$$

to the product element $S(\gamma) \in \Gamma$ given by

$$\prod_k \ \theta_{W(k)} \left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right]$$

where $\theta_{W(k)}$ is taken to be θ_U or θ_V depending upon whether the image of γ_k lies in U or V (if the image lies in both image sets, we have to show that the class in question does not depend upon which one we choose). As the parenthetical remark indicates, this construction sometimes involves choosing $W(k)$ to be U or V, and it also involves choosing the set of curves α_x and the partition of $[0, 1]$ into subintervals. We need to show that the resulting element of Γ does not depend upon these choices, and furthermore the value of the product depends only on the class of γ in $\pi_1(X)$.

More systematically, we claim this construction has the following properties:

(0) If $\gamma = \zeta + \zeta'$ where ζ and ζ' are also closed curves which start and end at the basepoint, then we can make all choices so that $S(\gamma) = S(\zeta) \cdot S(\zeta')$.

(1) If the image of γ lies in U or V, then we can make all choices so that $S(\gamma)$ lies in the image of θ_U or θ_V respectively.

(2) In cases where one or more curves of the form γ_k are mapped into $U \cap V$, and both the curve family $\{\alpha_x\}$ and the partition are held fixed, the resulting curve $S(\gamma)$ does not depend upon whether we choose $W(k)$ to be U or V.

(3) If we replace the curve family $\{\alpha_x\}$ by another family $\{\beta_x\}$ with the same properties, then $S([\gamma])$ is the same for both choices.

(4) The class $S(\gamma)$ does not depend upon the partition of $[0,1]$, so long as it is fine enough for the condition on subintervals to be fulfilled.

(5) (*The main objective.*) The class $S(\gamma)$ only depends upon the class of γ in $\pi_1(X)$.

Assuming these properties, we shall show how to derive injectivity from them. The final property implies that S passes to a homomorphism σ from $\pi_1(X,p)$ to Γ, and Property (1) implies that $S \circ j_{U*} = \theta_U$ and $S \circ j_{V*} = \theta_V$. Since the universal mapping property implies that the map $\varphi : \Gamma \to \pi_1(X)$ satisfies $\varphi \circ \theta_U = j_{U*}$ and $\varphi \circ \theta_V = j_{V*}$, it follows that $\varphi \circ \sigma$ agrees with the identity on the images of j_{U*} and j_{U*} in $\pi_1(X)$. By the surjectivity proof we know that these images generate the entire fundamental group, and therefore it follows that $\varphi \circ \sigma$ is the identity on $\pi_1(X)$. In particular, this means that σ is injective. On the other hand, since the images of θ_U and θ_V generate Γ, it follows that σ must also be onto, so that σ is an isomorphism and its inverse is φ.∎

All that remains now is to verify (0) – (5).

Verification of (0). Use the same family of curves α_x for ζ and ζ', and concatenate the chosen partitions, for which the images of all the curves ζ_k and ζ'_m are all contained in either U or V. This yields the formula

$$S(\zeta + \zeta') \quad = \quad \prod_k \theta_{W(k)} \left[\alpha_{k-1} + \zeta_k + (-\alpha_k)\right] \cdot \prod_k \theta_{W'(m)} \left[\alpha_{m-1} + \zeta'_m + (-\alpha_m)\right]$$

which is equal to $S(\zeta) \cdot S(\zeta')$.∎

Verification of (1). For these examples, we can take the trivial partition of $[0,1]$ into a single interval (namely, itself), so that $S(\gamma) = \theta_W([\gamma])$ by definition.∎

Verification of (2). There is an ambiguity about the choice of W only if the image of γ_k lies in $U \cap V$. In this case $\alpha_{k-1} + \gamma_k + (\alpha_k)$ is a closed curve in $U \cap V$, so we can write the corresponding classes in $\pi_1(U)$ and $\pi_1(V)$ as $i_{U*}(c)$ and $i_{V*}(c)$ respectively for some $c \in \pi_1(U \cap V)$. But by the construction of pushouts this means that

$$\theta_U \circ i_{U*}(c) \quad = \quad \theta_V \circ i_{V*}(c)$$

and hence we get the same factor regardless of whether we take W to be U or V.∎

Verification of (3). This is more complicated than the preceding arguments. Suppose that we are given a second system of curves $\{\beta_x\}$ with the same properties as the system $\{\alpha_x\}$. We need to show that

$$\prod_k \theta_{W(k)} \left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right] \quad = \quad \prod_k \theta_{W(k)} \left[\beta_{k-1} + \gamma_k + (-\beta_k)\right] \ .$$

Consider the closed curves $\mu_k = \beta_k + (-\alpha_k)$, which start and end at the basepoint p; by construction the images of these curves are contained in the open sets $W(k) \cap W(k-1)$, and $\mu_0 = \mu_n$ is the constant curve. It follows that

$$\left[\beta_{k-1} + \gamma_k + (-\beta_k)\right] \quad = \quad \left[\mu_{k-1}\right] \cdot \left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right] \cdot \left[-\mu_k\right]$$

and since $[-\mu_k] = [\mu_k]^{-1}$ this yields

$$\prod_k \theta_{W(k)}\left[\beta_{k-1} + \gamma_k + (-\beta_k)\right] \quad =$$

$$\prod_k \theta_{W(k)}[\mu_{k-1}] \cdot \theta_{W(k)}\left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right] \cdot \theta_{W(k)}[\mu_k]^{-1} \quad = \quad \prod_k \theta_{W(k)}\left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right]$$

since $\theta_{W(k)}[\mu_k] = \theta_{W(k-1)}[\mu_k]$ by Property (2) and both $[\mu_0]$ and $[\mu_n]$ are the trivial homotopy class.∎

Verification of (4). This is a standard sort of argument involving partitions and follows the same pattern as a result from Section VIII.6 on invariance under partitions. Two partitions of and interval always have a common refinement obtained by taking the all the break points in both partitions, so it suffices to show invariance under partitions when one is a refinement of the other. Since a refinement is given by a finite sequence of operations which add one break point at a time, by induction it suffices to show invariance if one partition is obtained from the other by inserting a single point.

Assume we have this situation, and suppose that the extra point r lies between the partition points t_{k-1} and t_k. Let γ be a closed curve, and let γ_k^- and γ_k^+ be obtained from the restrictions of γ to $[t_{k-1}, r]$ and $[r, t_k]$ respectively. Also, let α_r denote the α-curve joining the basepoint to $\gamma(r)$. Then the difference between the product expressions for $S(\gamma)$ with respect to the smaller and larger partitions is that the term $\theta_{W(k)}\left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right]$ is replaced by the product

$$\theta_{W(k)}\left(\left[\alpha_{k-1} + \gamma_k^- + (-\alpha_r)\right] \cdot \left[\alpha_r + \gamma_k^+ + (-\alpha_k)\right]\right)$$

and since

$$\left[\alpha_{k-1} + \gamma_k + (-\alpha_k)\right] \quad = \quad \left[\alpha_{k-1} + \gamma_k^- + \gamma_k^+ + (-\alpha_k)\right] \quad =$$

$$\left[\alpha_{k-1} + \gamma_k^- + (-\alpha_r) + \alpha_r + \gamma_k^+ + (-\alpha_k)\right] \quad = \quad \left[\alpha_{k-1} + \gamma_k^- + (-\alpha_r)\right] \cdot \left[\alpha_r + \gamma_k^+ + (-\alpha_k)\right]$$

the replacement is equal to the original term.∎

Verification of (5). Before proceeding, we note that Properties (1) – (4) imply $S(\gamma)$ is well-defined (the homotopy class does not depend upon the choice of partition or the system of curves joining the basepoint to the points of X).

Suppose now that H is a basepoint preserving homotopy from one closed curve γ_0 to another closed curve γ_1. Another Lebesgue number argument shows that for some $n > 0$ we can decompose $[0,1] \times [0,1]$ into n^2 nonoverlapping solid squares whose edges have length $1/n$ such that H maps each small solid square into either U or V. To be more precise, the small solid squares are given by

$$R_{i,j} \quad = \quad \left[\frac{i-1}{n}, \frac{i}{n}\right] \quad \times \quad \left[\frac{j-1}{n}, \frac{j}{n}\right]$$

where $1 \leq i, j \leq n$.

If $\gamma_{j/n}$ denotes the restriction of the homotopy H to the horizontal line segment $[0,1] \times \{j/n\}$ for $0 \leq j \leq n$, then by an inductive argument it will suffice to show that $S(\gamma_{(j-1)/n}) = S(\gamma_{j/n})$ for all $j \geq 1$. For the remainder of this argument we shall let j be fixed, and we shall denote $\gamma_{(j-1)/n}$ and $\gamma_{j/n}$ by ξ and η respectively; in this notation, the verification reduces to showing that $S(\xi) = S(\eta)$.

We shall prove that $S(\xi) = S(\eta)$ using a sequence of intermediate curves formed from pieces of ξ and η and other curves derived from the homotopy H. If $1 \leq k \leq n$, let ξ_k and η_k denote the curves formed by the restrictions of ξ and η to $[(k-1)/n, k/n]$. Furthermore, let ω_k be the curve formed by restricting the homotopy H to the vertical line segment $\{k/n\} \times [(j-1)/n, j/n]$ (see `svk-fig2.pdf` for a drawing depicting these curves); note that ω_0 and ω_n are constant curves which map the entire interval to the basepoint p. For each k, the closed curve formed by concatenating $-\omega_{k-1}$, η_k, ω_k, and $-\xi_k$ corresponds to the image under H of the boundary for the solid square $R_{k,j}$ parametrized in the counterclockwise sense. The restriction of H to this square yields a nullhomotopy of this concatenation, and hence it follows that the curves $-\omega_{k-1} + \eta_k$ and $\xi_k + (-\omega_k)$ are endpoint preserving homotopic.

We need just one more piece of notation. The α-curve joining the basepoint to $H(i/n, (j-1)/n)$ will be denoted by α_j^-, and the α-curve joining the basepoint to $H(i/n, (j)/n)$ will be denoted by α_j^+.

We shall now use the conclusions and notation in the preceding paragraphs to prove that $S(\xi) = S(\eta)$. Since ω_0 and ω_n are constant curves, we know that $S(\xi) = S(\xi + \omega_n)$ and $S(\eta) = S(\omega_0 + \eta)$, and hence it will be enough to show that $S(\omega_0 + \eta) = S(\xi + \omega_n)$. In fact, we shall prove that for each $k \geq 0$ that the group elements $g_k \in \Gamma$ given by

$$\prod_{i=1}^{k} \theta_{W(i)} \left[-\alpha_{i-1}^+ + \xi_i + (-\alpha_i^+)\right] \cdot \theta_{W(k)} \left[\alpha_k^+ + (-\omega_k) + (\alpha_k^-)\right] \cdot \prod_{i=k+1}^{n} \theta_{W(i)} \left[-\alpha_{i-1}^- + \eta_i + (-\alpha_i^-)\right]$$

are equal, with the convention that the first product does not appear if $k = 0$ and the second does not appear if $k = n$. By the previous observations we have $g_0 = [\eta]$ and $g_n = [\xi]$, so the chain of equations $[g_k] = [g_{k-1}]$ will yield (5).

If $k \geq 1$, then the difference between g_{k-1} and g_k is that the consecutive pair of factors

$$[-\alpha_{k-1}^+ + (-\omega_{k-1}) + (-\alpha_{k-1}^-)] \cdot [\alpha_{k-1}^- + \eta_k + (-\alpha_k^-)]$$

is replaced by

$$[-\alpha_{k-1}^+ + \xi_k + (-\alpha_k^+)] \cdot [\alpha_k^+ + (-\omega_k) + (-\alpha_k^-)]$$

(see `svk-fig2.pdf` for a drawing). These expressions can be simplified to

$$[-\alpha_{k-1}^+ + (-\omega_{k-1}) + \eta_k + (-\alpha_k^-)] \qquad \text{and} \qquad [-\alpha_{k-1}^+ + \xi_k + (-\omega_k) + (-\alpha_k^-)]$$

respectively, and they are equal because we have seen that the middle terms $(-\omega_{k-1}) + \eta_k$ and $\xi_k + (-\omega_k)$ are endpoint preserving homotopic. ∎

As noted previously, the verification of Properties (0) – (5) completes the proof of Theorem 1.

Generalization to disconnected intersections

It is natural to ask if the Seifert-van Kampen Theorem can be extended to situations with weaker hypotheses. In particular, one can ask for information about a space X which is a union of two arcwise connected open subsets U and V where $U \cap V$ is not necessarily arcwise connected. We have already noted that the theorem itself is systematically false if the intersection is disconnected. However, there is a generalization which states that the *fundamental groupoid* of X (see

the discussion following the statement of Corollary VIII.2.7) is a suitably defined pushout of the diagram of fundamental groupoids

$$\Pi(U) \longleftarrow \Pi(U \cap V) \longrightarrow \Pi(V)$$

where the arrows are induced by inclusions of subspaces (a *groupoid* is defined to be a category in which all morphisms are isomorphisms, and a group can be viewed as a groupoid which has only one object).

One reference for this highly abstract version of the Seifert-van Kampen Theorem is the following book:

R. (= Ronald) **Brown.** *Elements of Modern Topology.* McGraw-Hill, New York, 1968.

Extensively revised versions of this book also exist (one published by Ellis Horwood in 1988, and another by BookSurge in 2006). The following online document (by the same author) gives a completely self-contained proof of a slightly more general result:

`http://pages.bangor.ac.uk/~mas010/pdffiles/vKT-proof.pdf`

In particular, the theorem in this document includes the special case where X is a union of two arcwise connected open subsets (with a nonempty intersection).

Note on disambiguation. Since there is more than one R. Brown who has worked in algebraic topology during the past few decades, we note that the home page for the author of the book and online paper is

`http://www.bangor.ac.uk/~mas010/welcome.html`

and the home page of one other topologist with a similar name (Robert F. Brown) is

`http://www.math.ucla.edu/~rfb/` .

Both have written topology books of potential interest to graduate students.

IX.4 : Examples and computations

(Munkres, 59, 71–72; Hatcher, 0, 1.2)

In this unit we shall describe a few ways in which the Seifert-van Kampen Theorem can be applied to analyze the fundamental groups of certain spaces. The first result gives the most basic examples of spaces whose fundamental groups are free on a finite number of generators.

PROPOSITION 1. *Let X be a compact Hausdorff space which is a union of n closed subsets C_i such that each C_i is homeomorphic to S^1, and suppose that there is some point $p \in X$ such that if $i \neq j$ then $C_i \cap C_j = \{p\}$ (we assume n is finite). Then $\pi_1(X, p)$ is a free group on n generators.*

Proof. We shall do this by induction on the number n of circles; then we know the result is true when $n = 1$. Assume the result is known for a union of k circles satisfying the intersection condition, where $k \geq 1$.

Assume now that we have a union of $n = k + 1$ circles as in the statement of the proposition. For each i, let $h_i : S^1 \to C_i$ be a homeomorphism; composing h with a rotation of the circle if

necessary, we shall assume that $h_i(1) = p$. Let $q_i = h_i(-1)$ for $1 \leq i \leq k$. Then $\{p\}$ is a strong deformation retract of each punctured circle $S^1 - \{q_i\}$. We can now piece together the homotopy inverses on the individual punctured circles and show that $\{p\}$ is a strong deformation retract of $X - \{q_1, \cdots, q_{k+1}\}$ because the circles overlap at only one point and the values of all maps at that point are equal to p. Similarly, if we remove $q_1, \cdots, q_k$, the complement U is an open subset which is a strong deformation retract of the circle C_{k+1}, while if we remove q_{k+1} the complement V is a open set which is a deformation retract of the union of the first k circles $C_1 \cup \cdots \cup C_k$. We have already seen that $U \cap V$, which is the complement of all the points q_i, has the homotopy type of a point, and therefore the Seifert-van Kampen Theorem implies that $\pi_1(X)$ is the pushout of the diagram

$$\pi_1(C_{k+1}) \leftarrow \{1\} \rightarrow \pi_1(C_1 \cup \cdots \cup C_k) \ .$$

By induction we know that the group on the right is free on k generators, so $\pi_1(X)$ is a free product of $\mathbb{Z}$ with a free group on k generators. Since this free product is a free group on $k+1$ generators, this completes the proof.∎

There is a similar result showing that if two groups G_1 and G_2 can be realized as fundamental groups of two spaces X_1 and X_2, then there is a space which is built out of pieces closely resembling the latter whose fundamental group is isomorphic to the free product $G_1 * G_2$. This is left as an exercise.

Regular attachment of a cell

Frequently in topology and other subjects, one is given a space X which is a union of two closed subspaces $A \cup B$, where B is homeomorphic to a k-disk D^k such that $A \cap B$ corresponds to the boundary sphere S^{k-1}. If A (and hence X) is arcwise connected, then the Seifert-van Kampen Theorem shows that the fundamental groups of A and X are very closely related.

PROPOSITION 2. *Let $X = A \cup B$ satisfy the conditions described above, and suppose that the basepoint p lies in $A \cap B$.*

(i) If $k = 2$, let γ be a closed curve corresponding to some basepoint preserving homeomorphism $A \cap B \cong S^1$. Then $\pi_1(X, p)$ is isomorphic to the quotient of $\pi_1(A, p)$ by the normal subgroup (normally) generated by the class $[\gamma] \in \pi_1(A, p)$.

(ii) If $k \geq 3$, then the inclusion mapping induces an isomorphism from $\pi_1(A, p)$ to $\pi_1(X, p)$.

Proof. Choose a homeomorphism $h : D^k \rightarrow B$; the hypotheses imply that $h[S^{k-1}] = C$. Let $z = h(0)$, let e be the unique point of D^k such that $h(e) = p$, and let $q = h\left(\frac{1}{2}\, e\right)$ be the image of the midpoint of 0 and e. Define open subsets $U = X - \{z\}$ and $V = X - A$. Then V is homeomorphic to $D^k - S^{k-1}$ and $U \cap V$ is homeomorphic to $S^{k-1} \times (0.1)$. Since S^{k-1} is a strong deformation retract of $D^k - \{0\}$, the same is true for $C \subset B - \{z\}$ and hence also for $A = C \cup A \subset (B - \{z\}) \cup A = X - \{z\}$. We can now use the Seifert-van Kampen Theorem to conclude that we have the following pushout diagram:

$$\begin{array}{ccc} \pi_1(U \cap V, q) & \xrightarrow{\ i_{1*}\ } & \pi_1(U, q) \\ \Big\downarrow i_{2*} & & \Big\downarrow j_{1*} \\ \{1\} \ = \ \pi_1(V, q) & \xrightarrow{\ j_{2*}\ } & \pi_1(X, q) \end{array}$$

Note that we have changed basepoints from p to q because $p \notin V$. At this point the argument splits into cases depending upon whether $k \geq 3$ or $k = 2$.

(*ii*) Suppose that $k \geq 3$. Then $\pi_1(U \cap V)$ is trivial and therefore we know that $\pi_1(U, q) \to \pi_1(X, q)$ is an isomorphism. But we also know that A is a strong deformation retract of U and hence $\pi_1(A, p) \to \pi_1(U, p)$ is an isomorphism. To see that $\pi_1(A, p) \to \pi_1(X, p)$ is an isomorphism, consider the following commutative diagram in which γ is a curve in U joining q to p and γ^* denotes an associated change of basepoints isomorphism.

$$\begin{array}{ccccc} \pi_1(A,p) & \xrightarrow{a_*} & \pi_1(U,p) & \xrightarrow{i_{U*}} & \pi_1(X,p) \\ & & \downarrow \gamma_U^* & & \downarrow \gamma_X^* \\ & & \pi_1(U,q) & \xrightarrow{I_*} & \pi_1(X,q) \end{array}$$

Since I_* is an isomorphism and the change of basepoint maps γ_U^* and γ_X^* are isomorphisms, it follows that i_{U*} is also an isomorphism, and therefore $\pi_1(A, p) \to \pi_1(X, p)$ is also an isomorphism.∎

(*i*) Suppose now that $k = 2$. Then the pushout diagram implies that $\pi_1(X, q)$ is isomorphic to the quotient of $\pi_1(U, q)$ by the normal subgroup generated by $\pi_1(U \cap V, q)$. We need to translate this into statements about fundamental groups whose basepoints are p. Consider the following commutative diagram:

$$\begin{array}{ccccccc} \pi_1(C,p) & \xrightarrow{=} & \pi_1(C,p) & \longrightarrow & \pi_1(A,p) & & \\ \downarrow \cong & & & & \downarrow \cong & & \\ \pi_1(B-\{z\},p) & & & & \pi_1(U,p) & \xrightarrow{i_{U*}} & \pi_1(X,p) \\ \downarrow \gamma^* & & & & \downarrow \gamma^* & & \downarrow \gamma^* \\ \pi_1(B-\{z\},q) & \xleftarrow{\cong} & \pi_1(U\cap V,q) & \longrightarrow & \pi_1(U,q) & \xrightarrow{I_*} & \pi_1(X,q) \end{array}$$

In this diagram the lower left horizontal arrow is an isomorphism because $(B - \{z\}, U \cap V)$ is homeomorphic to the pair $\big(D^k - \{0\}, D^k - (\{0\} \cup S^{k-1})\big)$, and the upper left vertical arrow is an isomorphism because $(B - \{z\}, C)$ is homeomorphic to the pair $(D^k - \{0\}, S^{k-1})$.

Our analysis of the diagram now proceeds as follows: The Seifert-van Kampen Theorem implies that I_* is onto and its kernel is generated by $\pi_1(U \cap V, q)$. Furthermore, under the isomorphism from $\pi_1(A, p)$ to $\pi_1(U, q)$, the diagram shows that the image of $\pi_1(U \cap V, q) \cong \mathbb{Z}$ in $\pi_1(U, q)$ corresponds to the image of $\pi_1(C, p) \cong \mathbb{Z}$ in $\pi_1(A, p)$. Therefore the commutative diagram implies that $\pi_1(X, p)$ is the quotient of $\pi_1(A, p)$ modulo the normal subgroup generated by $\pi_1(C, p) \cong \mathbb{Z}$.∎

Realizing finitely presented groups as fundamental groups

We shall now use the preceding result to construct reasonably small and well-behaved spaces whose isomorphism classes of fundamental groups run through all finitely presented groups.

THEOREM 3. *Let G be a finitely presented group. Then there is a compact arcwise connected space X such that X is a subspace of some $\mathbb{R}^n$ and $\pi_1(X, p) \cong G$, where $p \in X$ is arbitrary.*

The change of basepoint results imply that the isomorphism type of the fundamental group is the same regardless of which basepoint is chosen. As Hatcher notes in his book, a result of S. Shelah (*Can the fundamental group of a space be the rationals?*, Proceedings of the American Mathematical Society, Vol. **103** (1988), pp. 627 — 632) implies that if X is an arcwise connected compact metric space then its fundamental group is either finitely presented or uncountable, and

we have already mentioned that Hatcher constructs the "Hawaiian earring" example for which the fundamental group is uncountable.

It will convenient to isolate two steps in the proof of Theorem 3 and formulate them as lemmas. The first is very close to issues we have previously considered.

LEMMA 4. *The disk D^k is homeomorphic to the quotient of $S^{k-1} \times [0,1]$ modulo the equivalence relation $\mathcal{A}$ whose equivalence classes are the one point subsets of $S^{k-1} \times [0,1)$ and the closed subset $S^{k-1} \times \{1\}$.*

Proof of Lemma 4. Since a 1–1 onto continuous map from a compact space to a Hausdorff space is a homeomorphism, it will suffice to construct a continuous mapping from $S^{k-1} \times [0,1]$ to D^k which passes to a 1–1 correspondence from the quotient $(S^{k-1} \times [0,1])/\mathcal{A}$ to D^k. One mapping with this property is $f(x,t) = (1-t)\,x$.∎

LEMMA 5. *Let $A \subset \mathbb{R}^n$ be a compact subset, and suppose that $C \subset A$ is homeomorphic to S^{k-1} for some $k \geq 2$. Then there is a compact subset $X \subset \mathbb{R}^{n+1}$ such that $X = A \times \{0\} \cup B$, where B is homeomorphic to D^k and $A \times \{0\} \cap B = C \times \{0\}$.*

Proof of Lemma 5. As suggested in the statement of the lemma, view $\mathbb{R}^n$ as the subset $\mathbb{R}^n \times \{0\}$ of $\mathbb{R}^{n+1}$, and likewise for all subsets. Fix a homeomorphism $h : S^{k-1} \to C$, and define a mapping $g : S^{k-1} \times [0,1] \to \mathbb{R}^{n+1}$ by $g(x,t) = (t \cdot h(v), (1-t)\,v)$. If $\mathcal{A}$ is defined as in Lemma 4, then g passes to a map g' on $(S^{k-1} \times [0,1])/\mathcal{A} \cong D^k$ which is continuous and 1–1, and therefore the image B of g and g' must be homeomorphic to D^k. By construction, this map extends the original homeomorphism h on S^{k-1}.∎

Proof of Theorem 3. Suppose that the group G is given by generators $g_1, \cdots, g_k$ and relations $r_1, \cdots, r_m$ (note that the latter list might be empty).

If in fact there are no relations, then G is a free group on the given set of generators, and in this case we take $X_0 \subset \mathbb{R}^2$ to be the union of the circles C_a whose centers are the points $(a,0)$ for integers a such that $1 \leq a \leq k$, and whose radii are equal to a. Then $C_a \cap C_b = \{(0,0)\}$ if $a \neq b$, and therefore the set $X_0 = \cup C_a$ satisfies the hypotheses of Proposition 1; the latter then implies that $\pi_1(X_0)$ is a free group on k generators.

The inductive step. Suppose that the result is known for groups given by k generators and $r-1$ relations, and let G be given by the generators $g_1, \cdots, g_k$ and relations $r_1, \cdots, r_m$. Define $\widetilde{G}$ to be the finitely presented group with the same generators but with relations $r_1, \cdots, r_{m-1}$. By construction, G is a quotient of $\widetilde{G}$ formed by factoring out the normal subgroup generated by the image ρ of r_m.

By the inductive hypothesis there is a compact subset X_{m-1} in some $\mathbb{R}^n$ such that $\pi_1(X_{m-1}) \cong \widetilde{G}$. Let $f : S^1 \to X_{m-1}$ represent the image ρ of the relation r_m. We do not know if f is a 1–1 mapping, but we can make it so up to homotopy if we replace X_{m-1} by $A_m = X_{m-1} \times D^2 \subset \mathbb{R}^{n+2}$ and replace f by the map $h : S^1 \to A_m$ such that $h(z) = (f(z), z)$. Then Lemma 5 applies to A_m and $C = h[S^1]$, and by the conclusion of that lemma we can form a space $X_m = A_m \times \{0\} \cup B$, where $B \cong D^k$ and $A_m \times \{0\} \cap B = C \times \{0\}$. We can now apply Proposition 2 to conclude that $\pi_1(X_m)$ is isomorphic to the quotient of $\widetilde{G}$ modulo the normal subgroup generated by ρ. Since this quotient is isomorphic to G, we have shown that $\pi_1(X_m) \cong G$.∎

COROLLARY 6. *If G is a finite group, then there is a compact subset X in some $\mathbb{R}^n$ such that $\pi_1(X) \cong G$.*∎

In our construction we have very little control over the value of n such that $X \subset \mathbb{R}^n$. However, for each group G we can find an example such that $n = 5$; I have not checked whether this is an

optimal result, but I suspect that $n = 4$ is lowest possible dimension such that one has examples for all finitely presented groups.

Isomorphisms between finitely presented groups

In the final two paragraphs of Section 69 in Munkres (see p. 425), the *isomorphism decision problem* for finitely presented groups is mentioned. This problem asks whether there is some uniform, totally systematic procedure for determining whether two finitely presented groups are isomorphic. The criteria for such a procedure are that it should lead to a computer program which could, after a finite amount of time, determine whether or not two finite presentations (of generators and relations) define isomorphic groups, and at each step there is never any doubt about what to do next. As noted in Munkres, one can prove that no such procedure exists. This is one of several decision problems about groups that were shown to be unsolvable during the nineteen fifties. Further information on such questions appears in Chapter 12 of the previously cited book by Rotman, and the unsolvability of the isomorphism question appears as Corollary 12.34 on page 469 of that reference. The following Wikipedia reference discusses the central question (the *Word Problem*) starting from first principles:

`http://en.wikipedia.org/wiki/Word_problem_for_groups`

The following book contains more detailed information on this area of group theory:

> **C. F. Miller, III.** On group-theoretic decision problems and their classification. Annals of Mathematics Studies, No. 68. *Princeton University Press, Princeton NJ*, 1971.

In contrast to the preceding negative information, we should note that there is a decision process for deciding whether two finitely generated **abelian** groups are isomorphic, and it can be constructed using the standard techniques in a first year graduate algebra course which yield the structure theorem at the beginning of these notes.

Although the answer to the isomorphism problem for finitely presented groups is negative, there are positive results on whether two finite presentations yield isomorphic groups. Specifically, there is a family of operations on finite presentations known as *Tietze transformations* such that two finite presentations yield the same group if and only if one can be obtained from the other by a finite sequence of Tietze transformations. From this perspective, the negative answer to the isomorphism problem means that one cannot describe a systematic way of determining whether there is such a sequence or, if it exists, finding an explicit recursive description of it. Further information on Tietze transformations appears in Section I.5 of the following book:

> **W. Magnus, A. Karrass, and D. Solitar.** Combinatorial group theory. Presentations of groups in terms of generators and relations (Reprint of the 1976 second edition). *Dover Publications, Mineola NY*, 2004.

Fundamental groups of surfaces

Unless a topology textbook is strictly limited to topics in point set topology, it usually contains material on surfaces. The latter are defined to be topological spaces which satisfy the following conditions:

(1) The Hausdorff Separation Property.

(2) Second countability (this might not be assumed at first, but generally it is assumed at some point in the discussion).

(3) Every point in the space has an open neighborhood which is homeomorphic to an open subset in $\mathbb{R}^2$.

Properties (1) — (3) imply that a surface is locally compact (and hence regular by the results of Section VI.4), and in fact the Urysohn Metrization Theorem implies that a surface is metrizable.

In particular, Chapter 12 in Munkres and numerous passages in Hatcher (scattered throughout the book but summarized in the Index) discuss some fundamental results, especially for the special case of compact surfaces.

Surfaces play an important role in many branches of mathematics, and one further reason for studying them is the way in which the topic combines topological methods with more geometrical considerations. However, one crucial issue is that many parts of the intuitive geometrical content are at best difficult to formulate in a completely rigorous manner at the beginning graduate level.

The treatment in Chapter 12 of Munkres is thorough and self-contained, and the techniques can accurately be described as "elementary" (but one must be warned that "elementary" arguments can often be extremely lengthy, complicated and tedious). In particular, Munkres formally states and proves some key results which are often at worst omitted from beginning graduate level treatments of surface theory, and at best are frequently stated informally and motivated by informal, intuitive discussions. One particularly important example in this direction is Theorem 77.8 on page 469, which states that a compact triangulated surface can be viewed as a polygonal region in the plane modulo identifying various edges on its boundary. However, Munkres sidesteps the question of whether an arbitrary surface has a triangulation — a theorem which was first proved by T. Radó in the 1920s — with a reference to two sources for proofs at the top of page 472 (although both are good, accessible references).

> *Remarks on terminology.* Although the exposition in Munkres is extremely good, there are some points on terminology that might cause confusion. His use of the term "homology" differs from the standard description of such groups; in Munkres, one only has 1-dimensional homology groups for arcwise connected spaces, and this group is defined to be the abelianizations of the fundamental group (a basic theorem going back to H. Poincaré, at least in some form, implies that Munkres' definition agrees with the usual one given in courses like 205B). Also, the choice of the word "scheme" to describe certain data is highly nonstandard; in algebraic geometry a *scheme* is a fundamental concept which has a totally unrelated meaning.

In this subsection we shall give a simplified approach to some results in surface theory which yield new and nontrivial fundamental group computations; for example, we shall only give a complete discussion for one compact surface (the *double torus*) which is not homeomorphic to the sphere S^2 or the torus T^2, and we shall only mention the results in the general case with informal descriptions of proofs which resemble the arguments given here.

Our construction and analysis of the double torus will be easier if we use the following construction of T^2.

PROPOSITION 7. *The torus T^2 is homeomorphic to the quotient of the disk D^2 modulo the equivalence relation $\mathcal{A}$ generated as follows:*

(1) *If a point p lies in the interior of the disk, then its equivalence class $[p]$ of p is equal to $\{p\}$.*

(2) *If a point $p = (x, y)$ lies on the unit circle S^1 and $|x| < |y|$, then the equivalence class $[p]$ of p consists of the two points $p = (x, y)$ and $(-x, y)$.*

(3) *If a point $p = (x, y)$ lies on the unit circle S^1 and $|x| > |y|$, then the equivalence class $[p]$ of p consists of the two points $p = (x, y)$ and $(x, -y)$.*

(4) *If a point $p = (x, y)$ lies on the unit circle S^1 and $|x| = |y|$, then the equivalence class $[p]$ of p consists of the four points $p = (x, y)$, $(-x, y)$, $(-x, -y)$ and $(x, -y)$.*

The composite of the quotient map $D^2 \to D^2/\mathcal{A}$ and the homeomorphism $D^2/\mathcal{A} \to T^2$ will be denoted by Φ.

Proof of Proposition 7. We start with the standard model of T^2 as the quotient of $[0,1] \times [0,1]$ modulo the identification of $\{0\} \times [0,1]$ with $\{1\} \times [0,1]$ and of $[0,1] \times \{0\}$ with $[0,1] \times \{1\}$, and then we apply the homeomorphism from $[0,1] \times [0,1]$ to D^2 given by the main result in Appendix A. Direct examination shows that the equivalence relation on $[0,1] \times [0,1]$ translates into the displayed equivalence relation on D^2.∎

Intuitively, we can think of the double torus as follows: Take two disjoint copies of T^2, in each copy find a closed subset D homeomorphic to D^2, and remove the subset of D which corresponds to the open disk from each torus, obtaining two spaces E_1 and E_2 which contain the unit circles C_1 and C_2 of the embedded disks. The double torus can be viewed as the space obtained from the disjoint union of E_1, E_2 and $A = S^1 \times [0,1]$ by identifying C_1 with $S^1 \times \{0\}$ and C_2 with $S^1 \times \{1\}$. However, we must be more formal about this construction in order to ensure that it has all the properties that are needed to compute the fundamental group of the space using the Seifert-van Kampen Theorem.

We begin with a special case of Theorem III.1.8.

LEMMA 8. *Let S be a compact Hausdorff topological space, and let $\mathcal{E}$ be an equivalence relation $\mathcal{E}$ on S such that the quotient space S is metrizable. If T is a compact metric space and $\mathcal{E}_T$ is the equivalence relation on $S \times T$ whose equivalence classes have the form $[s] \times \{t\}$, where $[s]$ is an equivalence class of some $x \in S$ and $t \in T$, then there is a canonical homeomorphism from $(S \times T)/\mathcal{E}_T$ to $(S/\mathcal{E}) \times T$ which sends the equivalence class of $(s,t) \in S \times T$ to $([s], t)$ for all s and t.*

We are primarily interested in the special case where T is the unit interval. This lemma may seem trivial and indeed its proof is not difficult, but the conclusion does not necessarily hold unless one imposes suitable hypotheses on S and T (one can get by with weaker hypotheses, but the assumptions in the lemma make the proof very simple, and they are adequate for our purposes). Quotient topologies can be very troublesome to work with.

Proof of Lemma 8. Let $F : S \times T \to (S/\mathcal{E}) \times T$ be the continuous mapping sending (s,t) to $([s], t)$. Since $(s,t)\ \mathcal{E}_T\ (s',t')$ if and only if $s\ \mathcal{E}\ s'$ and $t = t'$, the map F is constant on $\mathcal{E}_T$ equivalence classes and hence passes to a continuous mapping $f : (S \times T)/\mathcal{E}_T \to (S/\mathcal{E}) \times T$. By construction this mapping is 1–1 and onto. Since both S and T are compact Hausdorff and $S/\mathcal{E}$ is metrizable, we know that $(S \times T)/\mathcal{E}_T$ is compact and $S/\mathcal{E}) \times T$ is Hausdorff. Therefore by Theorem III.1.8 we know that f must be a homeomorphism.∎

We might as well also prove the following fact about quotient topologies that is needed. Once again, the conclusion may look obvious, but we have already mentioned that quotient topologies can sometimes behave unpredictably.

LEMMA 9. *Let S be a topological space, let $\mathcal{E}$ be an equivalence relation on S, let $A \subset S$ be a union of equivalence classes, and let $\mathcal{E}|A$ denote the equivalence relation on A induced by $\mathcal{E}$. Then*

there is a well-defined $1-1$ *mapping*

$$j : A/(\mathcal{E}|A) \;\longrightarrow\; S/\mathcal{E}$$

such that j *sends the equivalence class of* a *in* A *to the equivalence class of* a *in* S*, and* if the quotient projection $S \to S/\mathcal{E}$ is a closed mapping *then the subspace topology determined by* j *is equal to the quotient topology on* $A/(\mathcal{E}|A)$*, where the latter is defined by starting with the subspace topology on* $A \subset S$*.* (One can paraphrase the conclusion to say that "the quotient topology for the subspace is the subspace topology for the quotient.")

Note that if S is a compact metric space and $S/\mathcal{E}$ is metrizable then by Theorem III.1.8 the quotient space projection $S \to S/\mathcal{E}$ is automatically a closed mapping.

Proof. Let π and π_A denote the associated equivalence class projections on S and A respectively, and let $J : A \to X$ be the inclusion map. The map j exists because the composite $\pi \circ J$ is constant on equivalence classes. To see that it is 1–1, observe that $j([a]) = j[a']$ implies $\pi \circ J(a) = \pi \circ J(a')$, so that a and a' determine the same equivalence class in S, and the equivalence classes of a and a' in S coincide with their respective equivalence classes in A. Furthermore, since the subspace topology with respect to j is the smallest topology on $A/(\mathcal{E}|A)$ such that j is continuous, it follows that the quotient topology on this space contains the subspace topology.

Assume now that π is a closed mapping. We need to show that the quotient topology is contained in the subspace topology. Suppose that $C \subset A/(\mathcal{E}|A)$ is closed in the quotient topology. Then $C' = \pi_A^{-1}[C]$ is closed in A with respect to the subspace topology, and hence $C' = D \cap A$ for some closed subset D of S. Since π is closed it follows that $\pi[D]$ is also closed and hence $D' = \pi^{-1}\left[\pi[D]\right]$ is a closed subset of S containing D.

We claim that $C' = D' \cap A$. In one direction we have $C' = D \cap A \subset D' \cap A$. Conversely, if $x \in D' \cap A$ then $x \in D'$ implies that $x \;\mathcal{E}\; y$ for some $y \in D$, and since $x \in A$ implies that the entire equivalence class of x lies in A we also know that $y \in A$. In other words, we have $x \;\mathcal{E}\; y$ for some $y \in C' = D \cap A$, so that $\pi(x) = \pi(y)$. However, the set C' is also a union of equivalence classes, so this means that $x \in C'$, completing the proof that $C' = D' \cap A$.

To conclude the proof, note that C', D and A are all unions of equivalence classes, and therefore we have

$$C \;=\; \pi[C'] \;=\; \pi[D] \;\cap\; \pi[A] \;=\; \pi[D] \;\cap\; \mathrm{Image}\, j$$

which shows that C is closed with respect to the subspace topology on $A/(\mathcal{E}|A)$.∎

With these preliminary observations in hand, we can proceed to the formal discussion of the double torus. As suggested above, we start with Proposition, which in the terminology of Munkres states that $\Phi : D^2 \to T^2$ is a quotient map. Note that if $A \subset D^2$ contains S^1, then A is a union of equivalence classes determined by Φ (which are the inverse images of one point subsets). We now define

$E \subset T^2$ to be the image of the set $\{\; z \in D^2 \mid |z| \geq \frac{2}{3} \;\}$,

F to be the image of the set $\{\; z \in D^2 \mid |z| \geq \frac{1}{3} \;\}$,

V to be the image of the set $\{\; z \in D^2 \mid |z| > \frac{1}{3} \;\}$, and

B to be the image of S^1,

so that $B \subset E \subset V \subset F \subset T^2$ with V is open in T^2 and the remaining subsets are closed. Lemma 9 implies that each of these subsets is a quotient of its inverse image under Φ. Note that B is a union of two circles which have a single point in common.

The next result will play a crucial role in the computation of the fundamental group of the double torus:

PROPOSITION 10. *In the setting described above, the subspace B is a strong deformation retract of E, V and F.*

Proof. Let S be one of the subspaces E, V or F. By Lemma 9 the map

$$\Phi_S \;=\; \Phi|\Phi^{-1}[S] \;:\; \Phi^{-1}[S] \;\longrightarrow\; S$$

is a quotient map. By construction, B is a subspace of S. To construct the deformation retract data, first define

$$\widetilde{H} : \Phi^{-1}[S] \times [0,1] \;\longrightarrow\; \Phi^{-1}[S]$$

by the formula

$$\widetilde{H}(z,t) \;\;=\;\; \Phi\left((1-t)\, z \;+\; t\,|z|^{-1}\, z \right)$$

and notice that $\widetilde{H}$ is constant on Φ-equivalence classes, so that Lemma 8, Lemma 9 and the basic properties of quotient topologies yield a continuous mapping $H : S \times [0,1] \to S$ such that $\Phi_S \circ \widetilde{H} = H \circ \left(\Phi_S \times \mathrm{id}_{[0,1]}\right)$. By construction, the restriction of this homotopy to $S \times \{0\}$ is the identity on S, the restriction to $S \times \{1\}$ is given by a continuous mapping $\rho_S : S \to B$ such that $\rho|S$ is the identity, and for each $t \in [0,1]$ the restriction to $B \times \{t\}$ is just the inclusion of B in S. These properties imply that B is a strong deformation retract of S, and the mapping H defines the homotopy from the identity to $i_{B \subset S} \circ \rho_S$, where $i_{B \subset S}$ denotes the inclusion of B in S.∎

We are particularly interested in the following consequence involving fundamental groups:

PROPOSITION 11. *The fundamental group of V is isomorphic to a free group on two generators. In fact, if we take take $\omega(t)$ to be the counterclockwise circle $\frac{1}{2}\exp\left(2\pi\, i\left(t + \frac{1}{8}\right)\right)$, then there are free generators θ_1 and θ_2 for $\pi_1(V)$ such that the class of ω is equal to the commutator $[\theta_1, \theta_2] = \theta_1 \theta_2 \theta_1^{-1} \theta_2^{-1}$.*

Proof. Since B is a strong deformation retract of V, the fundamental groups of these spaces are isomorphic. By definition $B \subset T^2$ is the subset $S^1 \times \{1\} \cup \{1\} \times S^1$, and therefore by Proposition IX.1. this group is a free group on two generators. Furthermore, if $\Delta(t)$ is the closed curve in B given by $\Phi(\exp\left(2\pi\, i\left(t + \frac{1}{8}\right)\right)$, then by Proposition [quotient] the class of Δ in the fundamental group of B is the commutator $[g_1, g_2]$ of two standard generators for $\pi_1(B)$.

If we take $\Delta(0) = \omega(0)$ to be the basepoint for V, then the map

$$H(t,s) \;\;=\;\; \Phi\left(\tfrac{s+1}{2} \cdot \exp\left(2\pi\, i\left(t + \tfrac{1}{8}\right)\right)\right)$$

defines a free homotopy from ω to Δ. Note that these two curves have different values at the basepoint; namely, v_0 and v_1. We can now use the reasoning of Proposition VIII.2.9 to conclude that the class $[\omega] \in \pi_1(V, v_0)$ is equal to $\gamma^*([g_1, g_2])$, where $\gamma(s) = H(0,s)$ and γ^* is the associated group isomorphism from $\pi_1(V, v_1)$ to $\pi_1(V, v_0)$. Thus it suffices to let $\theta_i = \gamma^*(g_i)$ for $i = 1, 2$.∎

Now we are finally ready to describe the double torus formally.

Construction of the double torus. This space, which we shall denote by $\Sigma^2(2)$ (the superscript indicates the dimension, and the number in parentheses reflects an alternate name for this space which appears in Munkres; namely, it is often called a surface of genus two). As a preliminary step, we define $G \subset T^2$ to be the subset

$$\Phi\left[\, \{\; z \in D^2 \;|\; \tfrac{1}{3} \leq |z| \leq \tfrac{2}{3} \;\}\, \right] \;;$$

Since Φ is 1–1 on the set described inside the brackets and the latter is homeomorphic to $S^1 \times \left[\frac{1}{3}, \frac{2}{3}\right]$, it follows that G is also homeomorphic to $S^1 \times \left[\frac{1}{3}, \frac{2}{3}\right]$. We then define $\Sigma^2(2)$ to be the quotient space $(F \amalg F)/\mathcal{E}$, where $\mathcal{E}$ is an equivalence relation with the following equivalence classes:

> **Type 1 equivalence classes.** The equivalence class of a point x which is **NOT** in $(G \amalg G) \subset (F \amalg F)$ is the one point set $\{x\}$.

To simplify the next statement, we shall use the definition of a disjoint union $A \amalg B = (A \times \{1\}) \cup (B \times \{2\})$ to describe points in the two pieces of $A \amalg B$.

> **Type 2 equivalence classes.** Under the identification of G with $S^1 \times \left[\frac{1}{3}, \frac{2}{3}\right]$ via Φ, the equivalence classes of points in $G \amalg G$ are the two point sets consisting of $(z, t; 1)$ and $(z, 1-t; 2)$, where z and t run through all the elements of S^1 and $\left[\frac{1}{3}, \frac{2}{3}\right]$ respectively.

The correspondence between the two copies of G is a homeomorphism, and we shall denote this map by $h : G \times \{1\} \to G \times \{2\}$. We shall also use Ψ to denote the quotient space projection from $F \amalg F$ to $\Sigma^2(g)$.

The computation of $\pi_1(\Sigma^2(2))$ will use the following properties of our construction for this space:

PROPOSITION 12. *In the setting described above, if we write $F \amalg F$ as $F \times \{1, 2\}$ and take S_i to be the image of $S \times \{i\} \subset F \times \{i\}$ in $\Sigma^2(2)$ under the quotient space projection for $S = F$, V or E, then the construction for $\Sigma^2(2)$ has the following properties:*

(1) *If K is a closed subset of $F \times \{i\}$, where $i = 1$ or 2, then $\Psi[K \times \{i\}]$ is closed in $\Sigma^2(2)$.*

(2) *If $W \times \{i\}$ is an open subset of $V \times \{i\}$, where $i = 1$ or 2, then $\Psi[W \times \{i\}]$ is open in $\Sigma^2(2)$.*

(3) *If $i = 1$ or 2, then the mappings $\Psi | F \times \{i\}$ and $\Psi | V \times \{i\}$ are both $1 - 1$. Furthermore, $\Psi | F \times \{i\}$ is a closed mapping and and $\Psi | V \times \{i\}$ is an open mapping; consequently, Ψ maps the subsets $F \times \{i\}$ and $V \times \{i\}$ homeomorphically onto their images.*

(4) *$\Sigma^2(2)$ is a surface which is compact and second countable (in fact, it is also metrizable).*

(5) *$\Sigma^2(2)$ is arcwise connected.*

None of these properties are surprising or difficult to prove, but proofs are still needed.

Proof. (1) By the construction of the quotient topology this i equivalent to proving that $\Psi^{-1}[\,\Psi[K_i]\,]$ is closed in $F \amalg F$. If $i = 1$ then the space in question is equal to $K \amalg h[K \cap G]$; but $h[K \cap G]$ is closed in F because G is closed in F (hence $K \cap G$ is also closed) and h is a homeomorphism, and therefore $\Psi^{-1}[\,\Psi[K_1]\,]$ is closed in $F \amalg F$. Similarly, if $i = 2$ then we have $\Psi^{-1}[\,\Psi[K_2]\,] = h^{-1}[G \cap K] \amalg K$, and for similar reasons the latter is also closed in $F \amalg F$.∎

(2) Let $W \subset V$ be the image of W_i under the map $V \times \{i\} \to V$ which is projection onto the first coordinate. Since Ψ is onto we have $\Sigma^2(2) - \Psi[V_i] = \Psi[(F \amalg F) - W_i]$, so it suffices to show that the latter subspace is closed in $\Sigma^2(2)$. By construction we have

$$(F \amalg F) - W_i \;\; = \;\; (F - W) \amalg F \quad \text{or} \quad F \amalg (F - W)$$

depending upon whether $i = 1$ or 2; in either case the set in question has the form $A \cup B$ where both A and B are closed in F. Since $\Psi[A \amalg B] = \Psi[A \times \{1\}] \cup \Psi[B \times \{2\}]$, the conclusions of (1) show that such a set $\Psi[A \amalg B]$ must be closed in $\Sigma^2(2)$, and hence its complement, which is $\Psi[W \times \{i\}]$, must be open in $\Sigma^2(2)$.∎

(3) The preceding two properties imply that the restrictions of Ψ to the subspaces F_i are closed, and the restrictions of Ψ to the subspaces V_i are open. By construction these restrictions are continuous functions, and since every equivalence class in $F \amalg F$ contains at most one point from F_i (and hence at most one point from $V_i \subset F_i$) it follows that each restriction mapping is 1–1. Since each restriction mapping is continuous, 1–1 and either closed or open, it follows that each such restriction maps F_i and V_i homeomorphically onto its image.∎

(4) By (3) we know that F_1 and F_2 are homeomorphic to F and each is closed in $\Sigma^2(2)$; similarly, we know that V_1 and V_2 are homeomorphic to V and each is open in $\Sigma^2(2)$. By construction $\Sigma^2(2)$ is a continuous image of the compact space $F \amalg F$, and hence $\Sigma^2(2)$ is automatically compact. We can now derive the other properties of $\Sigma^2(2)$ as follows:

> Since $F_i \cong F$, each F_i is a closed Hausdorff subset of $\Sigma^2(2)$. The Hausdorff property for the latter is then a special case of the following statement, whose proof is left to the reader: *If a space X is a union of two* closed *Hausdorff subspaces A and B, then $X = A \cup B$ is also Hausdorff.*
>
> Since $V_i \cong V$ and V is homeomorphic to an open subset of the compact metric space T^2, it follows that each V_i is second countable. The second countability property for the latter is then a special case of the following statement, whose proof is left to the reader: *If a space X is a union of two* open *second countable subspaces U and W, then $X = U \cup W$ is also second countable.*
>
> If X is a surface and W is an open subset of X, then it follows immediately that W is a surface (the proof is left to the reader again). Therefore each open subset $V_i \subset \Sigma^2(2)$ is a surface. If $x \in \Sigma^2(2)$ then either $X \in V_1$ or $x \in V_2$; since the arguments in both cases are similar, we shall only prove the case $x \in V_1$ here. By the preceding discussion, we know that X has an open neighborhood $U \subset V_1$ such that U is homeomorphic to an open subset in $\mathbb{R}^2$. Since U open in V_1 and V_1 open in $\Sigma^2(2)$ implies that U is open in $\Sigma^2(2)$, it follows that x has an open neighborhood in $\Sigma^2(2)$ — namely, U — such that U is homeomorphic to an open subset in $\mathbb{R}^2$. We have already noted that similar considerations apply if $x \in V_2$, and thus it follows that $\Sigma^2(2)$ satisfies the conditions in the definition of a surface as given above.∎

(5) Since $\Sigma^2(2) = F_1 \cup F_2$ and $F_1 \cap F_2$ is nonempty, by (3) it suffices to show that F is arcwise connected. The latter follows because F is the image of the arcwise connected set $\{\ z \in D^2 \mid |z| \geq \frac{1}{3}\ \}$, which is homeomorphic to $S^1 \times [0,1]$.∎

We shall also need the following algebraic input.

PROPOSITION 13. *Suppose that we are given the following pushout diagram of groups:*

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{\ i_1\ } & G_1 \\ \downarrow i_2 & & \downarrow j_1 \\ G_2 & \xrightarrow{\ j_2\ } & P \end{array}$$

If g is a generator for $\mathbb{Z}$, then P is isomorphic to the quotient of the free product $G_1 * G_2$ by the normal subgroup generated K by the single element $i_1(g) \cdot i_2(g)^{-1}$.

Proof. Let $a = i_1(g)$ and $b = i_2(g)$. By construction the pushout P is the quotient of the free product $G_1 * G_2$ by the normal subgroup N generated by all elements of the form $a^m b^{-m}$, where m runs over all integers, so it suffices to check that $N = K$. Since a normal generating set for K

is contained in a normal generating set for N we clearly have $K \subset N$. Let q_K denote the quotient projection from $G_1 * G_2$ to $G_1 * G_2/K$, and identify a and b with their images in G_1 and G_2 respectively. Then we have $q_K(a) = q_K(b)$ by the definition of K, and since q_K is a homomorphism we also have

$$q_N(a^m b^{-m}) \;=\; q_N(a)^m \cdot q_N(b)^{-m} \;=\; q_N(a)^m \cdot q_N(a)^{-m} \;=\; 1$$

for all $m \in \mathbb{Z}$, so that every element of the form $a^m b^{-m}$ also lies in the kernel of q_K, which is K. Therefore we have shown that K contains a normal generating set for N, which means that N is also a subgroup of K; since the reverse inclusion has already been established, it follows that $K = N$.∎

Finally, here is the result which computes $\pi_1(\Sigma^2(2), p)$; since $\Sigma^2(2)$ is arcwise connected, the isomorphism type of this group does not depend upon the choice of p.

THEOREM 14. *For an arbitrary basepoint p the group $\pi_1(\Sigma^2(2), p)$ isomorphic to the quotient of a free group on four generators a, b, c, d modulo the normal subgroup generated by the commutator product $[a, b]\,[c, d] = aba^{-1}b^{-1}cdc^{-1}d^{-1}$. This group is nonabelian, and in fact it has a quotient group which is isomorphic to the a free abelian group on four generators.*

Proof. We apply the Seifert-van Kampen Theorem to the decomposition

$$\Sigma^2(2) \;=\; V_1 \,\cup\, V_2$$

in which the sets V_i are open in the surface. Each V_i is arcwise connected because it is homeomorphic to the continuous image of $\{\; z \in D^2 \mid |z| \geq \frac{1}{3} \;\} \cong S^1 \times (0, 1]$ under the continuous quotient mapping Φ. Furthermore, by Proposition 12 the intersection $V_1 \cap V_2$ is homeomorphic to the arcwise connected set $S^1 \times \left(\frac{1}{3}, \frac{2}{3}\right)$. Therefore the Seifert-van Kampen Theorem and Proposition 11 imply that $\pi_1(\Sigma^2(2), p)$ is given by a pushout diagram

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{\;i_1\;} & F_1 \\ \big\downarrow i_2 & & \big\downarrow j_1 \\ F_2 & \xrightarrow{\;j_2\;} & \pi_1(\Sigma^2(2), p) \end{array}$$

where F_1 is free on two generators a and b, F_2 is free on two generators c and d, the map i_1 sends a generator $g \in \mathbb{Z}$ to $[a, b]$, and the map i_2 sends the same generator $g \in \mathbb{Z}$ to $[d, c]$. Therefore Lemma 13 implies that $\pi_1(\Sigma^2(2), p)$ is isomorphic to the quotient of the free group on a, b, c, d modulo the normal subgroup generated by $[a, b]\,[d, c]^{-1}$. Since $[c, d] = [d, c]^{-1}$ (verify this!), it follows that $\pi_1(\Sigma^2(2), p)$ is given by the description in the theorem.

To prove the nontriviality statement, let D be the normal subgroup of the free group $F(a, b, c, d)$ which is generated by all commutators $[x, y]$ in that group. Then it is a straightforward exercise to verify that $F(a, b, c, d)/D$ is a free abelian group on the images of a, b, c, d. By the preceding arguments we know that $\pi_1(\Sigma^2(2), p)$ is the quotient of $F(a, b, c, d)$ by the normal subgroup N generated by one element of D, and therefore we know that $N \subset D$, so that there is a surjective homomorphism

$$\pi_1(\Sigma^2(2)) \;\cong\; F(a, b, c, d)/N \;\longrightarrow F(a, b, c, d)/D \;.$$

This proves that the fundamental group of the surface has a quotient group which is isomorphic to a free abelian group on four generators.∎

The corresponding results for fundamental groups of other surfaces are given in Chapter 12 of Munkres.

Appendix A : Topological equivalence of disks and hypercubes

In Section IX.3 we need to know that a solid square is homeomorphic to a solid 2-dimensional disk. The following result gives us everything we need.

THEOREM. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two points in* $\mathbb{R}^n$ *(where* $n \geq 2$*) with coordinate expressions* $(x_1, \cdots, x_n)$ *and* $(y_1, \cdots, y_n)$ *respectively, and let* r_1 *and* r_2 *be two real numbers (possibly equal). Then the closed disk of radius* r_1 *centered at* $\mathbf{x}$ *with respect to the usual Euclidean metric is homeomorphic to the hypercube* $\prod_j [y_j - r_2, y_j + r_2]$ *such that the respective boundaries correspond.*

Note that the standard hypercube $[0,1]^n$ is a special case of the second construction where $\mathbf{y}$ has all coordinates equal to $\frac{1}{2}$ and $r_2 = \frac{1}{2}$.

There are two main ideas behind the proof:

(1) A hypercube in $\mathbb{R}^n$ with sides of length $2a$ is the set of all vectors of length $\leq a$ for some norm $|\cdots|_\infty$ on $\mathbb{R}^n$.
(2) If $|cdots|_a$ and $|cdots|_b$ are two norms on $\mathbb{R}^n$, then for each $r > 0$ there is a norm-preserving homeomorphism from $(\mathbb{R}^n, |\cdots|_a)$ to $(\mathbb{R}^n, |\cdots|_a)$. In particular, the set of points with $a-$length ≤ 1 corresponds to the set of points with $b-$length ≤ 1 such that the respective sets of points with length strictly equal to 1 correspond to each other.

We do not need the full generality of (2), so we shall only prove the case where one of the norms $|\cdots|_2$ is the usual norm and $|\mathbf{v}|_\infty$ is the maximum of the scalars $|v|_i$, where the coordinates of $\mathbf{v}$ are given by $(v_1, \cdots, v_n)$.

Proof. We begin with the easiest pair; namely, the disk and the hypercube $[-1,1]^n$. Given a vector $\mathbf{x} \in \mathbb{R}^n$, let $|\mathbf{x}|_2$ denote its length with respect to the usual inner product and let $|\mathbf{x}|_\infty$ be the maximum of the absolute values of the coordinates ($= \max_i |x_i|$). Both of these define norms on $\mathbb{R}^n$, and the unit disks with respect to these norms are D^n and $[-1,1]^n$ respectively. If one defines a map f of $\mathbb{R}^n$ to itself by $f(\mathbf{0}) = \mathbf{0}$ and by

$$f(\mathbf{x}) \quad = \quad \frac{|\mathbf{x}|_\infty}{|\mathbf{x}|_2} \cdot \mathbf{x}$$

if $\mathbf{x} \neq \mathbf{0}$, then it follows that f is 1–1 onto and a homeomorphism except possibly at $\mathbf{0}$, and that for each $r > 0$ the map f sends points satisfying $|\mathbf{x}|_2 = r$ to points satisfying $|\mathbf{x}|_\infty = r$; one can check continuity of f and its inverse at $\mathbf{0}$ using the elementary inequalities

$$|\mathbf{x}|_\infty \quad \leq \quad |\mathbf{x}|_2 \quad \leq \quad n \cdot |\mathbf{x}|_\infty \ .$$

It follows that f defines a homeomorphism from D^n to $[-1,1]^n$.

To prove the general case, we first map the disk $|\mathbf{v} - \mathbf{x}|_2 \leq r_1$ to the unit disk centered at the origin by the affine map $r_1^{-1}(\mathbf{v} - \mathbf{x})$, then we apply f, after which we multiply by r_2 and translate by q:

$$h(\mathbf{v}) \quad = \quad r_2 \cdot f\left(r_1^{-1}(\mathbf{v} - \mathbf{x})\right) \ + \ q$$

One can check directly that this mapping sends the closed disk of radius r_1 centered at $\mathbf{x}$ with respect to the usual Euclidean metric is homeomorphic to the hypercube $\prod_j [y_j - r_2, y_j + r_2]$ such that the disk boundary (all points with $|\mathbf{v} - \mathbf{x}|_2 = r_1$) goes to the hypercube boundary (all points with $|\mathbf{w} - \mathbf{y}|_\infty = r_2$).∎

Appendix B : Topological manifolds

The concept of a surface, as defined in Section IX.4, is the 2-dimensional case of a fundamentally important mathematical structure:

Definition. A topological space X is said to be a **topological n-manifold** if it is Hausdorff and each point $x \in X$ has an open neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$.

The term "manifold" has evolved from G. F. B. Riemann's description of n-manifolds as n-fold extended magnitudes (roughly speaking, *manifold = many + fold*) in his highly influential lecture/essay, "Über die Hypothesen, welche der Geometrie zugrunde liegen." An English translation of this article by W. K. Clifford (*On the Hypotheses which lie at the Bases of Geometry*) was published on pages 14–17, and 36–37 of *Nature*, Volume 8 (1873), and it is available online at the following site:

`http://www.maths.tcd.ie/pub/HistMath/People/Riemann/Geom/WKCGeom.html`

One would like to say that the integer n is the *dimension* of the topological manifold, but in order to do so one must dispose of the following question about potential ambiguities:

> If m and n are positive integers and M is both a topological m-manifold and n-manifold, does it follow that $m = n$?

In fact, the answer to this question is YES by classical results from algebraic topology. Specifically, a positive answer follows from Theorem VII.1.7 in the following notes for 205B:

`http://math.ucr.edu/~res/math205B-2012/algtop-notes.pdf`

Since we shall not need the result here and the proof uses material not covered in this course, we shall not attempt to discuss the theorem any further. There will be several other points at which input from 205B is needed; in these cases we shall often refer to the notes cited above and abbreviate the reference to `algtop-notes.pdf`.

The definition of a topological manifold has numerous implications. For example, the following result is an immediate consequence of the definitions:

PROPOSITION 1. *Let X be a Hausdorff topological space. Then X is a topological n-manifold if and only if every point has a neighborhood base of open neighborhoods that are homeomorphic to (open balls/disks in) $\mathbb{R}^n$.*

Proof. The ($\Longleftarrow$) implication follows immediately from the definition because $\mathbb{R}^n$ is open in itself, so we now turn to the ($\Longrightarrow$) direction.

More generally, we have the following elementary observation: *If $x \in X$ has an open neighborhood homeomorphic to U such that x corresponds to $y \in U$ and $\{W_\alpha\}$ is a neighborhood base at y, then x has a neighborhood base consisting of sets homeomorphic to the sets in $\{W_\alpha\}$.* Specializing to the case where X is a topological n-manifold, we know that an arbitrary point x has an open neighborhood homeomorphic to an open subset $U \subset \mathbf{R}^n$, so it is only necessary to show that every point in U has a neighborhood base of the type described. But this follows immediately; an arbitrary point $y \in U$ has an open neighborhood base of sets $N_{1/k}(y)$ where k is a sufficiently large positive integer, so everything reduces to show that each of these neighborhoods is homeomorphic to $\mathbf{R}^n$. Note first that $N_\varepsilon(y)$ is homeomorphic to $N_1(0)$ by the map

$$h(u) \;\; = \;\; \left(\frac{1}{\varepsilon}\right) \, \cdot \, (u - y) \; .$$

Finally, note that the $N_1(0)$ is homeomorphic to $\mathbf{R}^n$ by the map

$$k(v) \;\; = \;\; \left(\frac{1}{1 - |v|} \right) \cdot v \; .$$

This completes the proof.∎

Example. Given that we have added the Hausdorff condition in the definition of a topological manifold, one might suspect that there are spaces that satisfy the main condition in the definition (locally Euclidean) but are not Hausdorff. The **Forked Line**, or something homeomorphic to it, is the standard example in the 1-dimensional case. Similar examples exist for all dimensions ≥ 1.

Here is the basic construction. Let X be the quotient space of $\mathbb{R} \times \{0,\, 1\}$ modulo the equivalence relation whose equivalence classes are given by the two point sets

$$\{ \; (y, 0), \, (y, 1) \; \}$$

for $y \neq 0$ and the one point sets given by $(0, 0)$ and $(0, 1)$, By construction, X is locally Euclidean of dimension 1. However, we claim that *the images of* $(0, 0)$ *and* $(0, 1)$ *do not have disjoint open neighborhoods*, or equivalently if we are given open neighborhoods U_0 and U_1 of these respective points then $U_0 \cap U_1 \neq \emptyset$.

Let U_0 and U_1 be open neighborhoods in X for the equivalence classes determined by $(0, 0)$ and $(0, 1)$ respectively, and let

$$q : \mathbf{R} \times \{0,\, 1\} \longrightarrow X$$

be the quotient space projection. Then $q^{-1}(U_0)$ and $q^{-1}(U_1)$ are open subsets of $\mathbf{R} \times \{0,\, 1\}$ that are unions of equivalence classes and contain $(0, 0)$ and $(0, 1)$ respectively. Since $q^{-1}(U_0)$ is an open subset containing $(0, 0)$ it must contain an open interval of the form $(-a, a) \times \{0\}$, and since it is also a union of equivalence classes it must also contain the interval

$$\Big(\; (-a, 0) \; \cup \; (0, a) \; \Big) \; \times \; \{1\} \; .$$

Similarly, we must have

$$q^{-1}(U_1) \;\; \supset \;\; \{ \; (0, 1) \; \} \;\; \cup \;\; \Big(\; (-b, 0) \; \cup \; (0, b) \; \Big) \; \times \; \{0,\, 1\}$$

for some $b > 0$. Therefore, if c denotes the smaller of a and b, then we know that $U_0 \cap U_1 \;\; \supset \;\; q(J_c)$, where

$$J_c \;\; = \;\; \Big(\; (-c, 0) \; \cup \; (0, c) \; \Big) \; \times \; \{0,\, 1\} \; .$$

In particular, U_0 and U_1 cannot be disjoint, and therefore X cannot be a Hausdorff space.

Finally, we note that although X is not Hausdorff, it is not difficult to verify that X is a $\mathbf{T_1}$ space.∎

The online *MathWorld* encyclopedia entry

`http://mathworld.wolfram.com/TopologicalManifold.html`

gives a reference to Hawking and Ellis for uses of non-Hausdorff locally Euclidean spaces in theoretical physics; however, we shall not need such objects subsequently in this course aside from

perhaps a few exercises. For the sake of completeness, here is a detailed bibliographic description of the book mentioned above:

> **S. W. Hawking and G. F. R. Ellis.** The Large Scale Structure of Space-Time. (Cambridge Monographs on Mathaemtical Physics.) *New York: Cambridge University Press, New York, NY*, 1975.

(**Note.** The trade and reader reviews of this book on `www.amazon.com` are definitely worth reading!)

Separation properties

The preceding discussion of the Hausdorff spaces leads naturally to questions about other separation properties of topological manifolds. We begin with some immediate consequences of the definitions and standard results in point set topology.

PROPOSITION 2. *If X is a topological n-manifold, then the following hold:*

(i) The space X is locally compact and Hausdorff, and hence X is also $\mathbf{T}_3$; in fact, it is also completely regular.

(ii) The space X is locally arcwise connected.

(iii) Every point $x \in X$ has a simply connected open neighborhood.

Verification of this result is left to the reader.∎

COROLLARY 3. *If X is a topological n-manifold, then the connected components are the same as the path components, and these are open sets ($\Longrightarrow$ topological n-manifolds themselves).*

Everything except the statement in parentheses follows from local arcwise connectedness, and the statement in parentheses follows by combining the prior portion of the conclusion with the first part of the previous proposition.∎

Another fundamental question along these lines is to determined necessary and sufficient conditions for a topological n-manifold to be metrizable. We shall discuss this later.

Examples and nonexamples

Example 0. Every open subset U of $\mathbf{R}^n$ is a topological n-manifold. — For each $x \in X$ one can take the "nice" neighborhood to be U itself.∎

Example 1. More generally, if U is an open subset of X and X is a topological n-manifold, then U is a topological n-manifold. — The proof of this is left as an exercise.∎

Example 2. We already mentioned that the standard 2-dimensional sphere is a topological 2-manifold. More generally, the following argument shows that the standard n-dimensional sphere S^n is a topological n-manifold:

By definition the standard n-dimensional unit sphere S^n is the set of all points x in $\mathbf{R}^n$ such that $|x|^2 = 1$, where $|v|$ denotes the length of v as a vector in $\mathbf{R}^n$. If $n = 1$ or 2 these definitions yield the standard circle in sphere in $\mathbf{R}^2$ and $\mathbf{R}^3$ respectively. — By construction the space S^n is Hausdorff, so we need to prove it is locally Euclidean. For $\sigma = \pm$, let U_j^σ be the set of points on S^n such that the j^{th} coordinate is positive for $\sigma = +$ and negative for $\sigma = -$. Now every

point on S^n must have at least one nonzero coordinate, and since this coordinate is either positive or negative it follows that every point lies in (at least) one of the sets U_j^σ. Furthermore, each of the sets U_j^σ is open because it is the intersection of S^n with the open set in $\mathbf{R}^n$ consisting of all points whose j^{th} coordinates lie in either $(-\infty, 0)$ or $(0, +\infty)$ depending upon the choice of σ. To complete the verification that S^n is a topological manifold, it will suffice to prove that each set U_j^σ is homeomorphic to $N_1(0) \subset \mathbf{R}^n$. Let $Q_j : \mathbf{R}^{n+1} \to \mathbf{R}^n$ be the linear transformation whose value on the standard unit vector $\mathbf{e}_i$ (with a one in the i^{th} coordinate and zeros elsewhere) is equal to $\mathbf{e}_i$ if $i < j$, zero if $i = j$, and $\mathbf{e}_{i-1}$ if $i > j$, and let k_j^σ be the restriction of Q_j to U_j^σ. We claim these maps define homeomorphisms from the sets U_j^σ onto $N_1(0)$. In fact, we shall construct explicit inverse mappings h_j^σ as follows: Let $S_j : \mathbf{R}^n \to \mathbf{R}^{n+1}$ be the linear transformation whose value on the standard unit vector $\mathbf{e}_i$ is $\mathbf{e}_i$ if $i < j$ and $\mathbf{e}_{i+1}$ if $i \geq j$. Then elementary calculations show that the continuous map

$$h_j^\sigma(x) \;\; = \;\; S_j(x) \; + \; \sigma \sqrt{1 - |x|^2}\, \mathbf{e}_j$$

is an inverse to k_j^σ.

Since the formulas for k_j^σ and h_j^σ are given in relatively concise form, the following descriptions of the functions when $n = j = 2$ might be helpful:

$$k_j^\sigma(x,\, y,\, z) \;\; = \;\; (x,\, z)$$

$$h_j^\sigma(u,\, v) \;\; = \;\; (u,\, \sigma \sqrt{1 - u^2 - v^2},\, v) \; \blacksquare$$

Remark. One important feature of the n-sphere is that it is a **compact** topological manifold, in contrast to nonempty open subsets of $\mathbf{R}^n$ which are always noncompact [PROOF: If U is a compact open subset of $\mathbf{R}^n$ then it is closed, so by the connectedness of $\mathbf{R}^n$ we either have $U = \emptyset$ or $U = \mathbf{R}^n$. Since $\mathbf{R}^n$ is not compact, it follows that U must be empty.]

Example 3. Another example of a compact topological 2-manifold is the 2-**torus**, which by definition is equal to $S^1 \times S^1$. More generally, if X is a topological n-manifold and Y is a topological m-manifold, then $X \times Y$ is a topological $(m + n)$-manifold. — The proof of this is also left as an exercise.

Example 4. Still more generally, if for each j such that $1 \leq j \leq m$ the space X_j is a topological n_j-manifold, then the product $\prod_j X_j$ is a topological d-manifold, where $d = \sum_j \; n_j$. in the special case where $X_j = S^1$ for each j, this product is known as the n-**torus** and denoted by T^n.$\blacksquare$

Remark. Various considerations in topology and geometry lead to a converse question: *If X and Y are spaces such that $X \times Y$ is a topological manifold for some n, are X and Y topological manifolds?* — There are many examples showing that the answer to this question is no. Here are two classical references:

[1] R. H. Bing, *The cartesian product of a certain nonmanifold and a line is E^4*. Ann. of Math. (2) **70** (1959), 399–412.

[2] R. M. Fox, *On a problem of S. Ulam concerning Cartesian products*, Fund. Math. **34** (1947), 278–287.

[3] J. Glimm, *Two Cartesian products which are Euclidean spaces*, Bull. Soc. Math. France **88** (1960), 131–135.

[4] K. W. Kwun, *Products of Euclidean spaces modulo an arc*, Ann. of Math. **79** (1964), 104–108.

Example 5. A Hausdorff space is a topological 0-manifold if and only if it is discrete. This is also left as an exercise (in fact the proof is almost trivial).∎

Example 6. If E and X are connected Hausdorff spaces and $p : E \to X$ is a covering space projection, then E is a topological n-manifold if and only if X is. — Once again the proof is left as an exercise.∎

Before proceeding to give examples of spaces that are not topological manifolds, we shall note one simple but important consequence of the preceding observation:

PROPOSITION 4. *If X is a connected topological n-manifold, then X has a simply connected covering space Y that is also a topological n-manifold.*

Proof. We have already observed that every point in a topological manifold has a neighborhood base of simply connected open subsets, and for such spaces the existence of a simply connected covering space follows from the main theorem in Section 82 of Munkres. To see that this covering space Y is a topological manifold, note first that every point $y \in Y$ has an open neighborhood V_0 which is homeomorphic to an open neighborhood U_0 of some point $x \in X$ (we assume y maps to x under the covering space projection). Since X is a topological manifold, we can find an open subneighborhood $U \subset U_0$ such that $x \in U$ and U is homeomorphic to an open subset of $\mathbb{R}^n$. If $V \subset V_0$ corresponds to U under the homeomorphism $V_0 \cong U_0$, then it follows that V is an open neighborhood of y which is homeomorphic to an open subset of $\mathbb{R}^n$.∎

Example 7. A Figure 8 curve is an example of a Hausdorff space that is not a topological manifold of any dimension. One specific example of such a curve is given by the parametric equations

$$\gamma(t) \quad = \quad (\sin 2t,\, \sin t)$$

where t lies in some open interval containing $[0, 2\pi]$. Detailed discussions of an equivalent curve (x- and y-coordinates switched, one axis compressed by a factor of $\frac{1}{2}$) may be found at the following online sites:

`http://www-gap.dcs.st-and.ac.uk/~history/Curves/Eight.html`

`http://www.xahlee.org/SpecialPlaneCurves_dir/LemniscateofGerono_dir/` [*continue*]
`lemniscateofGerono.html`

If one simply looks at the character

$$8$$

it seems likely that the crossing point in the center cannot have a neighborhood that is homeomorphic to an open n-disk in any Euclidean space. This follows from the results of Chapter VII in `algtop-notes.pdf`, and there is a detailed discussion in the course directory document `nonmanifolds.pdf`. **Example 8.** In the reference cited above, there is also a proof that the set $\mathbb{R}^n_+$ of all points in $\mathbb{R}^n$ with a nonnegative last coordinate is not a topological manifold.

Example 9. The *Hilbert cube* $[0, 1]^\infty$, which is defined to be a cartesian product of $\aleph_0$ copies of the unit interval $[0, 1]$, is not a topological n-manifold for any n. A proof of this fact, which uses input from `algtop-notes.pdf` and Section 50 of Munkres (either in the text itself or in the accompanying exercises) is given in `hilbert-cube.pdf`.

Homogeneity of connected manifolds

In Appendix C to `gentopnotes2014.pdf` we showed that if U is an open connected subset of $\mathbb{R}^n$, then for every pair of points $x, y \in U$ there is a homeomorphism h from U to itself such that $h(x) = y$. This generalizes directly to connected topological manifolds.

HOMOGENEITY THEOREM. *If M is a connected topological n-manifold, then then for every pair of points $x, y \in M$ there is a homeomorphism h from M to itself such that $h(x) = y$.*

Proof. The key step in the proof is the first proposition from Appendix C, which we restate for the sake of convenience:

> *Let D^n be the solid unit disk in $\mathbb{R}^n$, 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 $f(0) = v$.*

As in Appendix C, define an equivalence relation on the points of M by $x \sim y$ if and only if there is a homeomorphism h from M to itself such that $h(x) = y$.

We claim that the equivalence classes of this relation are open. Given $x \in M$, let V be an open neighborhood of x which is homeomorphic to an open n-disk $V_0 \subset \mathbb{R}^n$ centered at $\mathbf{0}$ such that x corresponds to $\mathbf{0}$. Now let $D_0 \subset V_0$ be a closed subdisk of smaller radius centered at $\mathbf{0}$, denote the images of D_0 and its boundary sphere in V_0 by D and S respectively, and let U be the open set $D - S$. If $y \in U$, then we can use the cited proposition to construct a homeomorphism f of D which is the identity on S and sends x to y. Since the $M = D \cup (M - U)$ and $S = D \cap (M - U)$, we can extend f to a homeomorphism h from M to itself by taking it to be the identity on $M - U$. By construction we have $h(x) = y$, and therefore we have shown that if $y \in U$ then $x \sim y$. This implies that the equivalence class of a point $x \in M$ is open.

The rest of the argument follows familiar lines: Since the equivalence classes are open and pairwise disjoint subsets, the complement of a single equivalence class — which is the union of all the other equivalence classes — is also open and hence a single equivalence class is both open and closed. Since M is connected, there must be only one equivalence class, and the Homogeneity Theorem follows from this fact and the definition of the equivalence relation.∎

The topological classification problem

In many parts of theoretical mathematics, it is interesting and important to study *classification problems*. For example, in the theory of finite groups, one natural question is to describe all groups of a fixed order n up to isomorphism, and every undergraduate course in abstract algebra answers this question if n is prime (all finite groups of prime order are cyclic). The corresponding problem for many other values of n arises frequently in graduate abstract algebra courses, and it is answered completely if the prime factorization of n is not too complicated (for example, if n is a square of a prime or twice and odd prime). For topological manifolds, or subclasses satisfying suitable restrictions, the corresponding question involves classification up to homeomorphism:

Classification Problem for Manifolds. *Let $\mathcal{A}$ be a class of topological manifolds. Find an explicitly describable subclass $\mathcal{A}_0 \subset \mathcal{A}$ such that every space in $\mathcal{A}$ is homeomorphic to a unique space in $\mathcal{A}_0$.*

Munkres gives a fairly detailed (and nearly mathematically complete) account of the important special case where $\mathcal{A}$ is the class of all compact topological 2-manifolds (also known as *compact* or

closed surfaces). Specific references are Sections 74–78 in Munkres, and particularly Section 77. Of course, one can also pose similar questions about classifying topological n-manifolds for other values of n, and we shall conclude this section by describing known results in these cases.

If $n = 0$ the classification is completely trivial because a topological 0-manifold is a discrete space; therefore, if $\mathcal{A}$ is the class of all second countable topological 0-manifolds, then one can take $\mathcal{A}_0$ to consist of one discrete space of each cardinality up to and including $\aleph_0$. If $n = 1$ and $\mathcal{A}$ is the class of all connected topological 1-manifolds, then one can use the methods of point set topology to prove that every such manifold is homeomorphic to the real line or the circle (and of course the latter are not homeomorphic because the second is compact and the first is not). The main ideas behind the proof of this result appear in the texts and online reference listed below; specifically, the reference in Hocking and Young is Section 2–5 on pages 52–55 with background material in the preceding section, and the reference in Christensen and Voxman is 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.

J. G. Hocking and G. S. Young. Topology. (Second edition.) *Dover, New York NY*, 1988.

C. O. Christenson and W. L. Voxman. Aspects of Topology. [FIRST EDITION.] (Pure and applied Mathematics, Vol. 39.) *Marcel Dekker, New York-Basel*, 1977.

`http://wolfweb.unr.edu/homepage/jabuka/Classes/2006_spring/···`
`···topology/Notes/08%20-%20One%20dimensional%20manifolds.pdf`

If $n = 3$ and $\mathcal{A}$ is the class of all compact topological 3-manifolds, then the answer to the classification question is considerably more difficult. The first significant advances took place near the end of the 19^{th} century, and subsequent work on the problem had a profound impact on geometric topology for most of the 20^{th} century. The ultimate classification scheme is related to a question called the 3-dimensional *Geometrization Conjecture*, and during the first decade of the 21^{st} century the validity of this conjecture (and the resulting classification result for compact 3-manifolds) was confirmed. This classification is summarized in the *Wikipedia* article cited below, and the cited book contains more detailed information on the proof of the Geometrization Conjecture.

`http://en.wikipedia.org/wiki/Geometrization_conjecture`

L. Bessieres, G. Besson, M. Boileau, S. Maillot, J. Porti. *Geometrisation of 3-manifolds*. European Mathematical Society Tracts in Mathematics, Vol. 13. European Mathematical Society, Zurich, Switzerland, 2010.

If $n = 4$ and $\mathcal{A}$ is some reasonable class of all compact topological 4-manifolds, then the answer to the classification question is beyond being unknown: It turns out to be mathematically unsolvable. The reasons are essentially algebraic and depend upon the recursive unsolvability of certain group-theoretic questions; one example is the impossibility of finding uniform criteria to decide whether two "reasonable" groups are isomorphic. One uses the fundamental group to reduce the topological questions to group-theoretic ones. A detailed discussion may be found in the book by Miller cited in Section IX.4.

Despite this negative result in dimensions greater than 3, there are classification results for certain classes of manifolds with restricted fundamental groups, and the most noteworthy case consists of simply connected manifolds. In particular, the topological classification of compact simply connected 4-dimensional manifolds was worked out in two stages, the first by J. Milnor in the 1950s and the second by M. H. Freedman and F. S. Quinn in the 1980s. A topological classification of compact simply connected 5-manifolds follows from results of D. Barden from the

1960s; although the latter were not originally stated as a topological classification, subsequent work of R. C. Kirby and L. C. Siebenmann (from later in the 1960s) shows that these results yield a topological classification. In principle it appears that one can use known techniques to derive classifications in higher dimensions, but things quickly become forbiddingly complicated even in the 6-dimensional case.

Metrization criteria

A standard topological counterexample called the **Long Line** shows that a topological 1-manifold in the sense of these notes is not necessarily metrizable (in fact, not necessarily $\mathbf{T}_4$). This example requires a considerable amount of background about well-ordered sets that we shall not otherwise need, and therefore the construction and proofs have been placed into the course directory file `longline.pdf`. Clearly we can propagate this example to higher dimensions by taking its product with $\mathbb{R}^k$. Another example of a nonmetrizable topological 2-manifold (the *Prüfer manifold*) is described in the following *Wikipedia* article:

`http://en.wikipedia.org/wiki/Pr%C3%BCfer_manifold`

For our purposes the Long Line's significance is that it leads directly to the following question: *Under what conditions on the topology of a topological manifold X is the latter metrizable?*

The first result is merely a special case of a result from Unit VI.

THEOREM 5. *If X is a topological n-manifold, then X is metrizable if and only if X is paracompact.*

Recall that the definition of paracompactness is given on page 253 of Munkres and in Section VI.5 of `gentopnotes2014.pdf`.

By Proposition B.2.(ii) above, every topological manifold splits into a disjoint union of its components (equivalently, its arc components), each of which is a topological manifold, and therefore the metrization question for connected topological manifolds is of particular interest. For these examples, we have the following result:

THEOREM 6. *If X is a connected topological n-manifold that is metrizable, then X is second countable.*

Since an arbitrary manifold is a disjoint union of its components, we can generalize the preceding theorem to arbitrary topological manifolds as follows:

THEOREM 7. *If X is a topological n-manifold, then X is metrizable if and only if each component of X is second countable.*

The rest of this appendix is devoted proving these three results. Note that the third result implies that all compact topological n-manifolds are metrizable. A more direct proof of this special case appears as Theorem 36.2 on pages 326–327 of Munkres.

Metrizability and second countability

Theorem 6 states that metrizability and second countability are equivalent for connected topological manifolds, and its proof splits naturally into two parts — proving that second countability implies metrizability and vice versa.

Proof of one implication in Theorem 6: *Second countability $\Longrightarrow$ metrizability.* By Corollary 2, a topological manifold is automatically $\mathbf{T_3}$, and by the Urysohn Metrization Theorem we know that a space is metrizable if it is $\mathbf{T_3}$ and second countable. Therefore a second countable topological manifold is metrizable.∎

The proof of the reverse implication depends upon the following result:

THEOREM 8. (The Basic σ-Compactness Theorem.) *Let X be a space that is paracompact $\mathbf{T_2}$, locally compact and connected. Then there is a countable family of compact subsets $K_n \subset X$ such that $X = \cup_n K_n$.*

The following definition motivates the name for this result:

Definition. A topological space X is said to be σ-compact if there is a countable family of compact subsets $K_n \subset X$ such that $X = \cup_n K_n$.

Examples. Every closed subset X of $\mathbb{R}^k$ is σ-compact, for we can take K_n to be the closed and bounded (hence compact) intersection of X with the closed unit disk of radius n centered at the origin. Another class of examples are locally compact $\mathbf{T_2}$ spaces which are second countable (for such a space, there is an countable open neighborhood base consisting of subsets with compact closures).

The proof of Theorem 8 requires the following auxiliary result.

LEMMA 9. *Let X be a topological space, let K be a compact subset of X, and let $\{U_\alpha\}$ be a locally finite open covering of X. Then there are only finitely many open sets U_β in the open covering such that $K \cap U_\beta = \emptyset$.*

Proof Lemma 9. For each $x \in K$ there is an open neighborhood V_x whose intersection with all but finitely many of the sets U_α is empty. By compactness K is contained in a finite union of the form

$$V_{x_1} \cup \ \cdots \ \cup V_{k_m}$$

and the intersection of this finite union with U_α is empty for all but finitely many α. Therefore the intersection of K with U_α is also empty for all but finitely many α.∎

Proof of Theorem 8. Let $\{U_\alpha\}$ be an open covering of X by subsets whose closures are compact. Such a covering exists because X is locally compact. Since every open covering has a locally finite refinement, we may as well assume that $\{U_\alpha\}$ itself is locally finite (note that the condition about compact closures is true for refinements of an open covering if it is true for the covering itself).

Choose W_0 to be an arbitrary nonempty set U_β from the open covering. Define a sequence of subspaces $\{W_n\}$ recursively by

$$W_n = \cup_\alpha \ \{U_\alpha \mid U_\alpha \cap W_{n-1} \neq \emptyset\}.$$

By construction this is an increasing sequence of open subsets. We claim that $X = \cup_k W_n$. Since the right hand side is nonempty and open, it suffices to show that $\cup_k W_n$ is closed. Suppose that x lies in the closure of $\cup_k W_n$. Then $x \in U_\alpha$ for some α, and since the closure of a set is the union of that set and its limit points it follows that

$$U_\alpha \ \bigcap \ (\cup_k W_n) \neq \emptyset.$$

The latter in turn implies that $U_\alpha \cap W_{n_0} \neq \emptyset$ for some n_0. But this implies that $x \in W_{n_0+1}$. Therefore all points in the closure $\overline{\cup_k \, W_n}$ in fact lie in $\cup_n W_n$, and hence the latter is closed.

We shall now show that the sets $\overline{W_n}$ is compact by induction on n; if $k = 0$ this holds because $\overline{U_\beta}$ is compact. If $\overline{W_n}$ compact, then by the lemma there are only finitely many U_α such that $U_\alpha \cap \overline{W_n} \neq \emptyset$; call these $U_{\alpha_1}, \cdots, U_{\alpha_p}$. It then follows that

$$\overline{W_{n+1}} \;\; = \;\; \overline{U_{\alpha_1}} \, \cup \, \cdots \, \cup \, \overline{U_{\alpha_p}} \; .$$

Since each of the closures on the right hand side is compact, it follows that the left hand side is a finite union of compact subsets and therefore is compact. Therefore if we set $K_n = \overline{W_n}$ then we know that K_n is compact, $K_n \subset K_{n+1}$ for all n, and $X = \cup_n \, K_n$.∎

Proof of the other implication in Theorem 6: *Metrizable and connected $\Longrightarrow$ second countable.* Assume that X is a topological manifold with these properties. As noted above, a space X is paracompact if it is metrizable. Therefore by the Theorem 8 we know that $X = \cup_n \, K_n$ where each K_n is compact. Furthermore, since X is metrizable each K_n is also metrizable. The latter implies that each K_n has a countable dense subset D_n, and therefore the countable subset $\cup D_n$ is dense in X. Since X is metric, this means that it is also second countable.∎

Finally, we shall prove Theorem 7. One key step in the proof is contained in the next result:

PROPOSITION 10. *Suppose that X is a Hausdorff topological space whose connected components are all open. Then X is metrizable if and only if each component is metrizable.*

Proof. The ($\Longrightarrow$) implication follows because subspaces of metrizable spaces are metrizable, so we focus on the other direction for the rest of the proof. Write $X = \cup_\alpha \, X_\alpha$ where each X_α is an open connected metrizable subset. It follows immediately that V is open in X if and only if $V = V_\alpha$ where V_α is open in X_α for each α.

Given a metric space, one can always find another metric with diameter ≤ 1 that defines the same topology; therefore for each α we can find a metric d_α for X_α with diameter ≤ 1. Using these metrics, we shall define a candidate for a metric on X that will define the same topology on X as the original one. Specifically, for $p.q \in X$ define $d(p, q)$ to be $d_\alpha(p, q)$ if there is a (necessarily unique) α such that both p and$q$ belong to X_α and set $d(p, q) = 2$ otherwise. It is a routine exercise to verify that d defines a metric on the set X; verification of the Triangle Inequality is the least trivial part, and this is done on a case by case basis depending upon whether points lie in the same or different components of X.

We must now show that the d-metric topology is equal to the original topology, which we shall call $\mathbf{U}$. First of all we claim that every ε disk for d belongs to $\mathbf{U}$. If $\varepsilon \leq 2$ and $p \in X_\alpha$, then the d-disk about p in X and the $_\alpha$-disks about p in X_α are identical; since the later is open in X_α and hence X, this proves the result when $\varepsilon \leq 2$. On the other hand, if $varepsilon > 2$ then the d-disk about p of radius ε in X is equal to X. This implies that the d-topology is contained in $\mathbf{U}$. Conversely, if $V \in \mathbf{U}$ then $V = V_\alpha$ where V_α is open in X_α for each α. By choice of the metrics d_α we know that V_α is also d_α-open in X_α for each α. Furthermore, by construction the d-open and d_α-open subsets of X_α are equal. Therefore each X_α is d-open in X_α, and by the openness of the latter in X it also follows that each V_α is d-open in X. Thus we have shown that every $\mathbf{U}$-open subset is d-open. Since we have already shown the converse, it follows that the d-topology is equal to $\mathbf{U}$.∎

Proof of Theorem 7. If X is a topological manifold, then it satisfies the hypotheses in Proposition 10, so X is metrizable if and only if each of its connected components is metrizable. By Theorem 6, the components of X are all metrizable if and only if they are all second countable, and Theorem 7 follows by combining this with Proposition 10.∎

Appendix C : Fiber spaces and fundamental groups

In Section VIII.5 we computed the fundamental groups of the real projective spaces $\mathbb{RP}^n$ for $N \geq 2$ using the standard double covering map $S^n \to \mathbb{RP}^n$. There is also a related family of objects known as *complex projective spaces*, and one goal of this appendix is to describe an analogous method which yields the fundamental groups of these spaces. Two secondary goals are to introduce a generalization of covering spaces known as *locally trivial fibrations*, which generalize covering space projections, and to show that they satisfy a weak version of the Path Lifting Property. Locally trivial fibrations and their generalizations ultimately play fundamentally important roles in both topology and geometry, and they are also used extensively in some branches of real and complex analysis, and they are also fundamental to several areas of mathematical physics (*e.g.*, gauge theory) which have been studied very actively and successfully for several deaces.

We shall closely follow the exposition in Section 3.4 of the following book:

E. L. Lima, *Fundamental groups and covering spaces* (Transl. by J. Gomes). AK Peters, Natick, MA, 2003.

For the most part, we shall refer to this book for proofs of the various results which are quoted here, but we shall also include some examples to illustrate a few important points.

Construction of complex projective space. We shall base our approach on `pg-all.pdf` for the standard algebraic and geometric constructions of projective n-space over a field, and we shall refer to `projspaces.pdf` for some key results on the topological properties of projective spaces. Formally, if $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ then the *n-dimensional coordinate projective space $\mathbb{FP}^n$ over $\mathbb{F}$* is defined to be the quotient of $\mathbb{F}^{n+1} - \{\mathbf{0}\}$ modulo the equivalence relation given by $\mathbf{x} \sim \mathbf{y}$ if and only if $\mathbf{x}$ and $\mathbf{y}$ are nonzero scalar multiples of each other. The algebraic and geometric reasons for considering this object are covered in `pg-all.pdf`, and it seems best simply to refer the reader to those notes.

> Generations of mathematicians are growing up who are on the whole splendidly trained, but suddenly find that, after all, they do need to know what a projective plane is.
>
> I. Kaplansky, *Linear Algebra and Geometry: A Second Course*, p. *vii*.

Eventually we shall need to work with the constructions in `projspaces.pdf` directly to obtain further information about the quotient projection $\Pi : \mathbb{F}^{n+1} - \{\mathbf{0}\} \to \mathbb{FP}^n$. However, for the time being the main implication from that file is that if $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ then $\mathbb{FP}^n$ is a compact Hausdorff topological dn-manifold, where $d = 1$ if $\mathbb{F} = \mathbb{R}$ and $d = 2$ if $\mathbb{F} = \mathbb{C}$.

For our purposes, the crucial property of Π is that it resembles a covering space projection in the following weak sense:

Definition. A continuous map $p : E \to B$ is said to be a *locally trivial fibration* (or a *fiber/fibre bundle*) if for each point $b \in B$ there is an open neighborhood V and a homeomorphism $h : V \times F \to p^{-1}[V]$ (for some space F) such that $p_V \circ h(x, y) = x$ for all $(x, y) \in V \times F$, where p_V denotes the restriction of p to $p^{-1}[V]$.

In the setting of the definition, the space E is called the **total space** (from the French *espace total*), the space B is called the **base space**, and for each $x \in F$ the subspace $F_x = p^{-1}[\{x\}]$ is

called the **fiber** over x (the spelling **fibre** is also found frequently, even in writings using U. S. writing conventions).

Definition. Two fiber bundles $p_1 : E_1 \to B$ and $p_2 : E_2 \to B$ (same base!) are *topologically equivalent* if there is a homeomorphism $\Phi : E_1 \to E_2$ such that $p_2 \circ \Phi = p_1$. This definition has several elementary but significant consequences:

(1) Topological equivalence of fiber bundles defines an equivalence relation on a set $\mathcal{A}$ of fiber bundles over B; if $\aleph$ is a cardinal number, it turns out that one can find a set $\mathcal{U}(\aleph)$ of fiber bundles over B such that every fiber bundle over B whose total space has cardinality $\leq \aleph$ is equivalent to a fiber bundle in $\mathcal{U}(\aleph)$, but we shall not need this fact.

(2) If $\Phi : E_1 \to E_2$ defines a topological equivalence of fiber bundles, then for each $x \in B$ the map Φ sends the E_1-fiber over x homeomorphically to the E_2-fiber over x.

(3) If $p : E \to B$ is a fiber bundle over a $\mathbf{T}_1$ space and $x \in B$, then the fiber F_x is closed because it is the inverse image of $\{x\}$.

(4) If $p : E \to B$ is a fiber bundle, then p is an open mapping (the proof is similar to the proof for covering space projections).

(5) Similarly, if $p : E \to B$ is a fiber bundle and B is connected, then for each x and y in E the fibers F_x and F_y are homeomorphic (the definition implies that the homeomorphism type is locally constant, and this means one can write a proof analogous to the argument showing that the number of sheets over a point in a covering space projection is the same for all points, provided B is connected).

Examples.

1. The simplest examples are *trivial fiber bundles* given by the maps $\pi_B : B \times F \to B$ which project onto the B coordinate.

2. Covering space projections are special cases of fiber bundles in which the fibers are all discrete sets. We shall see that some, but not all, basic properties of covering space projections can be generalized to fiber bundles. Covering spaces also provide examples of fiber bundles $p : E \to B$ where E is not homeomorphic to a product of B with the fiber over some point.

3. Another example of a nontrivial fiber bundle is given by the Klein bottle, viewed as the quotient of T^2 modulo identifying (z, w) with $(-z, \overline{w})$ for each z and w, with the projection map first sending the equivalence class $[z, w]$ to the class $[z]$ in S^1 modulo the equivalence relation $\mathcal{R}$ identifying z with $-z$ and then composing with the standard homeomorphism $S^1/\mathcal{R} \cong S^1$ sending the class $[\exp \pi\, i\, t]$ (where $t \in [0, \pi]$) to $\exp 2\pi\, i\, t$. If U is an open arc in S^1 of length $\lambda < \pi$, then its inverse image in the Klein bottle is given by taking a subset of the form $U' \times S^1$ — where U' is an open arc of length $\frac{1}{2}\,\lambda$ — and projecting it down to the Klein bottle using the double covering $T^2 \to K$. One can check directly that the restriction of the quotient map to such a set is 1–1. In this case the fiber at every point is homeomorphic to S^1 and the base is equal to S^1, but the total space is not a product because $\pi_1(K)$ is a nonabelian group and $\pi_1(T^2)$ is abelian.

4. Here is a simpler example like the previous one, but without complete proofs: The open Möbius strip can be viewed as the quotient of $S^1 \times (-1, 1)$ modulo identifying (z, t) with $(-z, -t)$, which is equivalent to the usual description as $[0, 1] \times (-1, 1)$ modulo identifying $(0, t)$ with $(1, -t)$. As in the previous example, if M is the open Möbius strip then we can view it as the total space of a fiber bundle over S^1 whose fibers are homeomorphic to (-1.1). However, this is also not equivalent to a trivial bundle, for the one point compactification of M turns out to be $\mathbb{RP}^2$ and the one point compactification of $S^1 \times (-1, 1)$ turns out to have an infinite fundamental group (take the

inclusion of the one point compactification S^1 of $\{1\} \times (-1,1)$ into the one point compactification of $S^1 \times (-1,1)$ and check that this map is a retract).

We are particularly interested in another class of fiber bundles:

THEOREM 6. *Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, let $d = 1$ or 2 depending upon whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and let n be a positive integer. Then the homogeneous coordinate projection map*

$$\Pi : \mathbb{F}^{n+1} - \{\mathbf{0}\} \longrightarrow \mathbb{FP}^n$$

is a fiber bundle whose fibers are homeomorphic to $\mathbb{F} - \{0\}$.

In contrast to Propositions 3.6 and 3.11 of Lima (pages 63 and 69–70), Theorem 6 realizes $\mathbb{FP}^n$ as the base space of a fiber bundle where the total space and fibers are noncompact. However, it turns out that the fibrations in Lima and this appendix are homotopically equivalent fibrations over B in an appropriately defined sense. Our proof is designed to reflect the classical algebraic approach to projective spaces, in which they are highly symmetric objects defined over arbitrary fields, and it relies heavily on `projspaces.pdf`.

Proof of Theorem 6. Let $j : \mathbb{F}^n \to \mathbb{FP}^n$ be the standard inclusion of affine n-space over $\mathbb{F}$ into projective n-space over $\mathbb{F}$ which sends $(t_1, \cdots, t_n)$ to $\Pi(t_1, \cdots, t_n, 1)$. By Proposition 6 in `projspaces.pdf` this map is a homeomorphism onto an open subset W_{n+1} of $\mathbb{FP}^n$, and it follows immediately that $\Pi^{-1}[W_{n+1}] = \mathbb{F}^n \times (\mathbb{F} - \{0\})$. If $A : W_{n+1} \to \mathbb{F}^n$ is the inverse map, so that $A(j(\mathbf{v})) = \mathbf{v}$ for all $\mathbf{v}$ and $h : W_{n+1} \times (\mathbb{F} - \{0\}) \to \Pi^{-1}[W_{n+1}]$ is defined by

$$h(x,t) \;\; = \;\; \big(t^{-1}A(x), t\big)$$

then h is a homeomorphism such that $\Pi \circ h(x,t) = x$ for all (x,t), and therefore Π satisfies the defining condition for a fiber bundle at every point of $W_{n+1} \subset \mathbb{FP}^n$.

Suppose now that $x \in \mathbb{FP}^n$ is not in the image of j and $\xi = (x_0, \cdots, x_{n+1})$ is a set of homogeneous coordinates for x. Then $x_{n+1} = 0$ but there is some $k < n+1$ such that $x_k \neq 0$. Let C_k be the projective collineation of $\mathbb{FP}^n$ arising from the invertible linear transformation $\mathbb{F}^{n+1}$ which switches the k^{th} and last coordinates (and leaves the remaining coordinates untouched), and let B_k be the homeomorphism induced by this invertible linear transformation on $\mathbb{F}^{n+1} - \{\mathbf{0}\}$; note that B_k and C_k are their own inverses because they are defined by a permutation which switches two numbers. It follows that x will lies in the image W_k of $j_k = C_k \circ j$; since C_k is a homeomorphism (by Theorem 5 in `projspaces.pdf`) and j is an open mapping, it also follows that W_k is a open neighborhood of x. Therefore it suffices to show that the defining condition for a fiber bundle is satisfied over all points of W_k.

By construction we have a commutative diagram of the form

$$\begin{array}{ccc} \mathbb{F}^{n+1} - \{\mathbf{0}\} & \xrightarrow{B_k} & \mathbb{F}^{n+1} - \{\mathbf{0}\} \\ \big\downarrow \Pi & & \big\downarrow \Pi \\ \mathbb{FP}^n & \xrightarrow{C_k} & \mathbb{FP}^n \end{array}$$

such that C_k maps W_{n+1} onto W_k and B_k maps $\Pi^{-1}[W_{n+1}]$ onto $\Pi^{-1}[W_k]$.

Finally, let

$$h_k(x,t) \;\; = \;\; B_k\,(\,h(C_k(x),t)\,)$$

where h is defined as at the beginning of the proof; it follows that h maps $W_k \times (\mathbb{F}^{n_1} - \{\mathbf{0}\})$ homeomorphically onto $\pi^{-1}[W_k]$, and furthermore we have

$$\Pi \circ B_k\,(\,h(C_k(x),t)\,) \;=\; C_k \circ \Pi\,(\,h(C_k(x),t)\,) \;=\; C_k\,(\,C_k(x)\,)$$

and the latter is merely x because C_k is equal to its own inverse. Therefore Π satisfies the defining condition for a fiber bundle over U_k for all k, and this proves that $\Pi : \mathbb{F}^{n+1} - \{\mathbf{0}\} \longrightarrow \mathbb{FP}^n$ is indeed a fiber bundle.∎

We have said that fiber bundles are generalizations of covering space projections, so the next task is to derive analogs of some basic results on covering spaces.

THEOREM 7. (Path Lifting Property) *Suppose that $p : E \to B$ is a fiber bundle, let $\gamma : [0,1] \to B$ be a continuous curve and let $y_0 \in E$ be a point such that $p(y_0) = \gamma(0)$. Then there is a continuous lifting of γ to a curve $\alpha : [0,1] \to E$ such that $p \circ \alpha = \gamma$ and $\alpha(0) = y_0$.*

In other words, the existence half of the Path Lifting Property for covering spaces is also valid for fiber bundles. However, such liftings are far from unique. For example, if we consider the trivial fiber bundle $B \times F \to B$ and $y_0 = (b_0, f_0)$ projects to b_0, then the liftings of γ which satisfy the given properties are all curves of the form $(\gamma(t), \theta(t))$ where θ is a continuous curve such that $\theta(0) = f_0$. If, say, F is an arcwise connected metric space with infinitely many points, then there are infinitely many different choices for θ and hence infinitely many choices of liftings.

Sketch of proof for Theorem 7. (See Lima, Proposition 3.12 and its proof on pages 70–71, for more details.) A Lebesgue number argument shows that there is a partition of $[0,1]$ into finitely many intervals such that γ maps each subinterval into an open set over which the bundle is equivalent to a product bundle. For the sake of definiteness write the partition in the form

$$0 \;=\; t_0 \;<\; t_1 \;<\; \cdots \;<\; t_m = 1\;.$$

By the preceding discussion involving liftings for products, we can find a lifting α_1 of $\gamma|[0,t_1]$ to E such that $\alpha_1(0) = y_0$. Assume by induction that we have a lifting α_k of $\gamma|[0,t_k]$ such that $\alpha_1(0) = y_0$; if $k < m$ we want to extend α_k to a similar lifting α_{k+1} of $\gamma|[0,t_{k+1}]$. We can do this because γ maps $[t_k, t_{k+1}]$ into an open subset U such that $p^{-1}[U] \cong U \times F$ by first defining a lifting β of $\gamma|[t_k, t_{k+1}]$ such that $\beta(t_k) = \alpha_k(t_k)$ (using the fact that the bundle is a product over U) and then gluing β and α_k together at their common endpoint to obtain a continuous lifting α_{k+1}. When we complete this step for $k+1 = m$, then we have the desired lifting of the entire curve.∎

We also have the following result relating the fundamental groups of the total and base spaces of a fiber bundle. Note that there is a relation between the fundamental groups, but the conclusion is totally different from the result for covering space projections.

THEOREM 8. *Let $p : E \to B$ be a fiber bundle where B and F are arcwise connected and $p(e_0) = b_0$. Then the E is arcwise connected and the homomorphism $p_* : \pi_1(E, e_0) \to \pi_1(B, b_0)$ is onto.*

Sketch of proof. (See Lima, Corollary 3.6 and Proposition 3.13 and their proofs on page 77, for more details.)

We first prove that E is arcwise connected. Let $y_0, y_1 \in E$, and let $x_0, x_1 \in B$ denote their images under p. If γ is a continuous curve joining x_0 to x_1, let α be a lifting of γ such that $\alpha(0) = y_0$. Then $\alpha(1)$ and y_1 both lie in the fiber of x_1. Our hypotheses imply that the fibers over all points are arcwise connected, and therefore we can also join $\alpha(1)$ to y_1 by a continuous curve

in the fiber over x_1. If we concatenate these curves, we obtain a continuous curve joining y_0 to y_1, and this implies that E is arcwise connected.

We shall now prove that the map in fundamental groups is onto. Let γ be a basepoint preserving closed curve in B, and let α be a continuous lifting such that $\alpha(0) = e_0$. There is no reason to expect that $\alpha(1) = \alpha(0)$, but we do know that $\alpha(1)$ lies in the same fiber as $\alpha(0)$. Since the fibers are arcwise connected, we can join $\alpha(1)$ to $\alpha(0)$ by a curve β which lies completely in the fiber. Then the concatenation $\alpha + \beta$ is a closed curve in E, and its projection onto B is merely the concatenation $\gamma +$constant of two closed curves, and we know that this curve is basepoint preseving homotopic to γ. Therefore, given a class $[\gamma]$ in the fundamental group of B, we have constructed a closed curve $\alpha + \beta$ in E such that $[\gamma] = p_*([\alpha + \beta])$.∎

Theorem 8 immediately yields a computation for the fundamental group of $\mathbb{CP}^n$, where $n \geq 1$ is arbitrary.

THEOREM 9. *For each $n \geq 1$ and each basepoint x, the group $\pi_1(\mathbb{CP}^n, x)$ is trivial.*

Proof. We shall use the fiber bundle $\Pi : (\mathbb{C}^{n+1} - \{\mathbf{0}\}) \to \mathbb{CP}^n$, for which the fibers are all homeomorphic to the arcwise connected space $\mathbb{C} - \{0\} \cong S^1 \times \mathbb{R}$. Since $\mathbb{C}^{n+1} - \{\mathbf{0}\} \cong S^{2n+1} \times \mathbb{R}$ and $n \geq 1$ we know that the total space is simply connected. By Theorem 8 we also know that the fundamental group of $\mathbb{CP}^n$ is isomorphic to a quotient group of $\pi_1\left(\mathbb{C}^{n+1} - \{\mathbf{0}\}\right)$, and since the latter is trivial we know that $\pi_1(\mathbb{C}^n)$ must also be trivial.∎

Summary of Files in the Course Directory

The course directory is ∼res/math205A-2014 on the math.ucr.edu network. This summary only lists files which are relevant to the second part of the course (the material covered in these notes). There is not much overlap with the list at the end of gentopnotes2014.pdf.

affine+convex.pdf Some basic properties of a class of homeomorphisms from $\mathbb{R}^n$ to itself which are known as *affine* (AFF-fine) *transformations*; these include all isometries and similarities of $\mathbb{R}^n$ (see also metgeom.pdf).

algtop-notes.pdf

Course notes for Mathematics 205B, the next course in the sequence.

advancednotes2014.pdf

Course notes for Mathematics 246A, which is a sequel to Mathematics 205B.

beyond205A.pdf

Remarks about subsequent courses in the sequence and some areas of mathematics which are closeley related to this course, including information on the continuation of the material from Part II of Munkres at the beginning of Mathematics 205B.

categories2014.pdf

A brief survey of category theory; the material will be used extensively in this part of the course.

concat.pdf

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.

fundgp-notes.pdf

This document, which is the second part of the course notes.

fundgpsubheadings.pdf

A more detailed table of contents which lists all the subheadings in the preceding document.

hilbert-cube.pdf

Some properties of the Hilbert Cube, which is a countably infinite product of copies of the unit interval; this is only needed in one of the appendices to the class notes.

knots.pdf

Remarks concerning simple closed curves in $\mathbb{R}^3$ and the fundamental groups of their complements.

longline.pdf

An example of a nonmetrizable Hausdorff space in which every point has an open neighborhood homeomorphic to an open interval (*i.e.*, a *topological* 1*-manifold*).

`metgeom.pdf`

This an appendix to the course notes which discusses isometries and similarity maps of Euclidean spaces, including a proof that partial isometries and similarities on subsets of Euclidean spaces extend to global isometries and similarities of the appropriate Euclidean spaces. The file `affine+convex.pdf` contains further information on some of the topics discussed here.

`nicecurves.pdf`

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.

`openRn.pdf`

This document describes a proof that if $n \geq 3$ and X_k is a subset of $\mathbb{R}^n$ with k points, then $\mathbb{R}^n - X_p$ and $\mathbb{R}^n - X_q$ are simply connected open sets in $\mathbb{R}^n$ which are not homeomorphic; the proof uses input from Mathematics 205B.

`outline205A2.pdf`

An outline of sections in the class notes which will be covered in course and qualifying examinations.

`pg-all.pdf`

A fairly detailed introduction to projective geometry and projective spaces, written for undergraduate mathematics majors with a firm understanding of linear algebra.

`projspaces.pdf`

This document states and proves results about the topological properties of real and complex projective spaces.

`polishcircle.pdf`

An example of a subset of the plane which looks like a circle but is not locally connected, and some unusual topological properties of this space.

`polishcircleA.pdf`

Constructions of some well-behaved approximations to the space studied in the preceding document.

`polishcircleB.pdf`

An example to show that the Seifert-van Kampen Theorem does not extend to spaces presented as a union of two closed (as opposed to open) subspaces (both of which are arcwise connected, as is their intersection). The exposition relies on some material from the preceding two documents.

`projspaces.pdf`

This document states and proves results about the topological properties of real and complex projective spaces.

`radproj.pdf`

A drawing illustrating one way of constructing a homeomorphism from a closed disk to itself which is the identity on the boundary sphere and sends the center to some other point in the open disk.

`secVII4-addendum.pdf`

A continuation of Section VII.4 in the notes, with details and drawings for an example discussed very briefly in Munkres; this example is similar to the example from Munkres at the end of the indicated section of the notes.

`straightline.pdf`

A simple example to illustrate the need to be careful when working with straight line homotopies.

`svk-fig1.pdf`

A drawing illustrating the first part of the proof of the Seifert-van Kampen Theorem; namely, the fundamental group of the entire space is generated by the fundamental groups of the two open subsets.

`svk-fig2.pdf`

A drawing illustrating one step in the second part of the proof of the Seifert-van Kampen Theorem; namely, the inductive step to prove that two curves which are constructed out of a homotopy, and differ by a small amount, determine the same element in the pushout group.