Preface

Perhaps the simplest motivation for algebraic topology is the following basic question:

If m and n are distinct positive integers, is $\mathbf{R}^m$ ever homeomorphic to $\mathbf{R}^n$?

Results from point set topology imply the answer is **NO** if one of m and n is equal to 1. If a homeomorphism $h : \mathbf{R}^m \to \mathbf{R}$ existed then for each $\mathbf{x} \in \mathbf{R}^m$ we could conclude that $\mathbf{R}^n - \{\mathbf{x}\}$ is homeomorphic to $\mathbf{R} - \{h(\mathbf{x})\}$. Since $\mathbf{R}^m - \{\mathbf{x}\}$ is connected for all $\mathbf{x} \in \mathbf{R}$ if $m > 1$ while $\mathbf{R} - \{t\}$ is not connected for any choice of $t \in \mathbf{R}$, it follows that $\mathbf{R}^m - \{\mathbf{x}\}$ is never homeomorphic to $\mathbf{R} - \{t\}$ if $m > 1$ and hence $\mathbf{R}^m$ cannot be homeomorphic to $\mathbf{R}$. Similarly, results on fundamental groups imply that for all relevant choices of $\mathbf{x}$ the set $\mathbf{R}^m - \{\mathbf{x}\}$ is simply connected if $m > 2$ while $\mathbf{R}^2 - \{\mathbf{x}\}$ has an infinite cyclic fundamental group, so we also know that $\mathbf{R}^m$ is not homeomorphic to $\mathbf{R}^2$ if $m > 2$. One basic goal of an introductory course in algebraic topology is to show that $\mathbf{R}^m$ is never homeomorphic to $\mathbf{R}^n$ if $m \neq m$.

The idea behind proving such results is to define certain abelian groups which give an **algebraic picture** of a given topological space; in particular, if two topological spaces are homeomorphic, then their associated groups will be algebraically isomorphic. Unfortunately, the definitions for these **homology groups** are less straightforward than the definition of the fundamental group, and much of the work in this course involves the construction of such groups and the proofs that they have good formal properties.

In analogy with standard results for fundamental groups, the homology groups of two spaces will be isomorphic if the spaces satisfy a condition that is somewhat weaker than the existence of a homeomorphism between them; namely, an the groups are isomorphic if the two spaces have the same *homotopy type* as defined on page 363 of the book by Munkres cited below.

Since the constructions for the associated groups are somewhat complicated, it is natural to expect that they should be useful for more than simply answering the homeomorphism question for Euclidean spaces. In particular, one might ask if these groups (and a course in algebraic topology) can shed new light on some questions left open in undergraduate or beginning graduate courses in mathematics.

1. The material in introductory graduate level courses does not really give much insight into the popular characterization of topology as a "rubber sheet geometry." In other words, topology is generally viewed as the study of properties that do not change under various sorts of bending or stretching operations. Some aspects of this already appear in the study of fundamental groups, and one objective of this course is to develop these ideas much further.

2. As a refinement of the problem at the beginning of this preface, one can ask if there is some topological criterion which characterizes the algebraic notion of n–dimensionality, at least for spaces that are relatively well-behaved.

3. An algebraic topology course should also yield better insight into several issues that arise in undergraduate courses, including (a) the Fundamental Theorem of Algebra, (b) various facts about planar and nonplanar networks, (c) insides and outsides of plane curves and closed surfaces in 3–dimensional space, and (d) Euler's Formula for "nice" polyhedra in $\mathbf{R}^3$; namely, if P is a polyhedron bounding a convex body in $\mathbf{R}^3$, then the numbers V, E and F of vertices, edges and faces satisfy the equation $E + 2 = V + F$.

4. If time permits, another goal will be to give a unified approach to certain results in multivariable calculus involving the ∇ operator, Green's Theorem, Stokes' Theorem and the Divergence Theorem, and to formulate analogs for higher dimensions.

Throughout the course we shall use the following book as a reference for many topics and definitions:

> **J. R. Munkres.** *Topology* (Second Edition), *Prentice-Hall, Saddle River NJ*, 2000. ISBN: 0–13–181629–2.

The official text for this course is the following book:

> **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`

One important feature of homology groups is that if $f : X \to Y$ is a continuous mapping of topological spaces, then there is an associated homomorphism f_* from the homology groups of X to the homology groups of Y; this is again similar to the situation for fundamental groups of pointed spaces, and it plays an important role in addressing the issues listed above. In fact, algebraic topology turns out to be an effective means for analyzing the following central problem:

> Given two "reasonably well-behaved" spaces X and Y, describe the homotopy classes of continuous mappings from X to Y.

In general, the descriptions of the homotopy classes can be3 quite complicated, and only a few cases of such problems can be handled using the methods of a first course, but we shall mention a few special cases at various points in the course.

Many of the basic properties of homology groups and homomorphisms are best stated using the formalisms of **Category Theory**, and many of the constructions and theorems in algebraic topology are best stated within the framework of **Homological Algebra**. We shall develop these subjects in the course to the extent that we need them.

Prerequisites

The name "algebraic topology" suggests that the subject uses input from both algebra and topology, and this is in fact the case; since topology began as a branch of geometry, it is also reasonable to expect that some geometric input is also required. Our purpose here is to summarize the main points from prerequisite courses that will be needed. Additional background material which is usually not covered explicitly in the prerequisites will be described in the first unit of these notes.

Set theory

Everything we shall need from set theory is contained in the following online directory:

`http://math.ucr.edu/~res/math144`

In particular, a fairly complete treatment is contained in the documents `setsnotes`n`.*`, where $1 \leq n \leq 8$ and the file type `*` is one of `doc`, `ps` or `pdf`. In most cases the **pdf** versions are the most convenient to use, but parts of the **doc** files are in color rather than black and white.

There are two features of the preceding that are somewhat nonstandard. The first is the definition of a function from a set A to another set B. Generally this is given formally by the graph, which is a subset $G \subset A \times B$ such that for each $a \in A$ there is a unique $b \in B$ such that $(a, b) \in G$. Our definition of function will be a **triple** $f = (A, G, B)$, where $G \subset A \times B$ satisfies the condition in the preceding sentence. The reason for this is that we must specify the target set or **codomain** of the function explicitly; in fact, the need to specify the codomain has already arisen at least implicitly in prerequisite graduate topology courses, specifically in the definition of the fundamental group. A second nonstandard feature is the concept of **disjoint union** or **sum** of an indexed family $\{ X_\alpha \}$ of sets. The important features of the disjoint sum, which is written $\amalg_\alpha X_\alpha$, are that it is a union of subsets Y_α which are canonically in 1–1 correspondence with the sets X_α and that $Y_\alpha \cap Y_\beta = \emptyset$ if $\alpha \neq \beta$. Another source of information on such objects is Unit V of the online notes for Mathematics 205A which are cited below.

Topology

This course assumes familiarity with the basic material in graduate level topology courses through the theory of fundamental groups and covering spaces (in other words, the material in Mathematics 205A and 205B). Everything we need from the first of these courses can be found in the following online directory:

`http://math.ucr.edu/~res/math205A`

In particular, the files `gentopnotes2005.*` contain a fairly complete set of lecture notes for the course. This material is based upon the textbook by Munkres cited in the Preface. Two major differences between the notes and Munkres appear in Unit V. The discussion of quotient topologies is somewhat different from that of Munkres, and in analogy with the previously mentioned discussion of set-theoretic disjoint sums there is a corresponding construction of disjoint sum for an indexed family of topological spaces.

The necessary material on fundamental groups and covering spaces appears in the following sections of Munkres:

51 – 56

58 – 59

61 – 64

67 – 71

79 – 82

Supplementary exercises for Chapter 13.

At many points of these notes we shall rely heavily on the contents of these sections.

Algebra

As in the later parts of Munkres, we shall assume some familiarity with certain topics in group theory. Nearly everything we need is in Sections 67 – 69 of Munkres, but we shall also need the following basic result:

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.

A proof of this fundamental algebraic result may be found in Sections II.1 and II.2 of the following standard graduate algebra textbook:

> Hungerford, Thomas W. Algebra. Reprint of the 1974 original. (Graduate Texts in Mathematics, 73.) *Springer-Verlag, New York–Berlin–etc.*, 1980. ISBN: 0-387-90518-9.

Material from standard undergraduate linear algebra courses will also be used as needed.

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.

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

I. Foundational and Geometric Background

Aside from the formal prerequisites, algebraic topology relies on some background material from other subjects that is generally not covered in prerequisites. In particular, two concepts from the foundations of mathematics, namely **categories** and **functors**, play a central role in formulating the basic concepts of algebraic topology. Furthermore, since algebraic topology places heavy emphasis on spaces that can be constructed from certain fundamental building blocks, some relatively elementary but fairly detailed properties of the latter are indispensable. The purpose of this unit is to develop enough of category theory so that we can use it to formulate things efficiently and to describe the topological and geometric properties of a class of well-behaved spaces called **polyhedra** that will be needed in the course.

I.1 : Categories and functors

(Hatcher, § 2.3)

If mathematics is the study of abstract systems, then category theory may be viewed as an abstract formal setting for working with such systems. In fact, the theory was originally developed by S. Eilenberg and S. MacLane in the 1940s to provide an effective conceptual framework for handling various constructions and phenomena related to algebraic topology. The formal definition may be viewed as a generalization of familiar properties of ordinary set-theoretic functions.

Definition. A **CATEGORY** is a system $\mathbf{C}$ consisting of

(a) a class $\mathrm{Obj}\,(\mathbf{C})$ of sets called the **objects** of $\mathbf{C}$,

(b) for each ordered pair of objects X and Y a set $\mathrm{Morph}_{\mathbf{C}}(X,Y)$ called the **morphisms** from X to Y,

(c) for each ordered triple of objects X, Y and Z a *composition pairing* $\mathrm{Morph}_{\mathbf{C}}(X,Y) \times \mathrm{Morph}_{\mathbf{C}}(Y,Z) \longrightarrow \mathrm{Morph}_{\mathbf{C}}(X,Z)$, whose value for (f,g) is generally written $g \circ f$, such that the following hold:

(1) The sets $\mathrm{Morph}_{\mathbf{C}}(X,Y)$ and $\mathrm{Morph}_{\mathbf{C}}(Z,W)$ are disjoint unless $X = Z$ and $Y = W$.

(2) For each object X there is an identity morphism $1_X = \mathrm{id}_X \in \mathrm{Morph}_{\mathbf{C}}(X,X)$ such that for each $f \in \mathrm{Morph}_{\mathbf{C}}(X,Y)$ and each $g \in \mathrm{Morph}_{\mathbf{C}}(Z,X)$ we have $f \circ 1_X = f$ and $1_X \circ g = g$.

(3) The composition pairings satisfy an associative law; in other words, if $f \in \mathrm{Morph}_{\mathbf{C}}(X,Y)$, $g \in \mathrm{Morph}_{\mathbf{C}}(Y,Z)$, and $h \in \mathrm{Morph}_{\mathbf{C}}(Z,W)$, then $(h \circ g) \circ f = h \circ (g \circ f)$.

By the assumptions, for each $f \in \mathrm{Morph}_{\mathbf{C}}(X,Y)$ the objects X and Y are uniquely determined, and they are called the **domain** and **codomain** of f respectively. When working within a given category we generally use familiar notation like $f : X \to Y$ to indicate that $f \in \mathrm{Morph}_{\mathbf{C}}(X,Y)$.

As in set theory, at some points one must take care to avoid difficulties with classes that are "too large" to be sets (for example, we cannot discuss the set of all sets), but in practice it is always possible to circumvent such problems by careful choices of definitions and wordings (for example, using the theory of *Grothendieck universes*), so we shall generally not dwell on such points.

By the remarks preceding the definition of a category, it is clear that we have a category **SETS** whose objects are given by all sets, whose morphisms are set-theoretic functions from one set to another (with the conventions mentioned in the Prerequisites!), and whose composition is merely ordinary composition of mappings. Here are some further examples:

1. Given a field $\mathbf{F}$, there is the category $\mathbf{VEC_F}$ whose objects are vector spaces, whose morphisms are $\mathbf{F}$-linear transformations, and whose composition is ordinary composition. The important facts here are that the identity on a vector space is a linear transformation, and the composite of two linear transformations is a linear transformation.

2. There is also a category **GRP** whose objects are groups and whose morphisms are group homomorphisms (with the usual composition). Once again, the crucial properties needed to check the axioms for a category are that identity maps are homomorphisms and the composite of two homomorphisms is a homomorphism.

3. Within the preceding example, there is the subcategory **ABGRP** whose objects are abelian groups, with the same morphisms and compositions. In this category, the set of morphisms from one object to another has a natural abelian group structure given by pointwise addition of functions, and the resulting abelian group of homomorphisms is generally denoted by $\mathrm{Hom}(X, Y)$.

4. If P is a partially ordered set with ordering relation $\leq$, then one has an associated category whose objects are the elements of P and such that $\mathrm{Morph}\,(x, y)$ consists of a single point if $x \leq y$ and is empty otherwise. This is an example of a **small** category in which the class of objects is a set.

5. One can also use partially ordered sets to define a category **POSETS** whose objects are partially ordered sets and whose morphisms are monotonically nondecreasing functions from one partially ordered set to another; as in most other cases, composition has its usual meaning.

6. If G is a group, then G also defines a small category as follows: There is exactly one object, the morphisms of this object to itself are given by the elements of G, and composition is given by the multiplication in G.

7. There is a category **TOP** whose objects are topological spaces, whose morphisms are continuous maps between topological spaces, and whose composition is the usual notion. Again, the crucial properties needed to verify the axioms for a category are that identity maps are continuous and composites of continuous maps are also continuous.

8. There are also categories whose objects are topological spaces and whose morphisms are **open** maps or **closed** maps. The categories with various types of morphisms are distinct.

9. One also has a category **MET–UNIF** whose objects are metric spaces and whose morphisms are **uniformly continuous** mappings (with the usual composition).

10. Given an arbitrary category $\mathbf{C}$, one has the **dual** or **opposite** category $\mathbf{D} = \mathbf{C}^{\mathbf{OP}}$ with the same objects as $\mathbf{C}$, but with $\mathrm{Morph}_{\,\mathbf{D}}(X, Y) = \mathrm{Morph}_{\,\mathbf{C}}(Y, X)$ and composition $*$ defined by $g * f = f \circ g$. Note that if $\mathbf{D} = \mathbf{C}^{\mathbf{OP}}$ then $\mathbf{C} = \mathbf{D}^{\mathbf{OP}}$.

In most of the preceding examples of categories, there is a fundamental notion of **isomorphism**, and in fact one can formulate this abstractly for an arbitrary category:

Definition. Let $\mathbf{C}$ be a category, and let X and Y be objects of $\mathbf{C}$. A morphism $f : X \to Y$ is an *isomorphism* if there is a morphism $g : Y \to X$ (an inverse) such that $g \circ f = 1_X$ and $f \circ g = 1_Y$.

This generalizes notions like an invertible linear transformation, a group isomorphism, and a homeomorphism of topological spaces.

PROPOSITION 1. *Suppose that $f : X \to Y$ is an isomorphism in a category $\mathbf{C}$ and g and h are inverses to f. Then $h = g$.*

Proof. Consider the threefold composite $h \circ f \circ g$. Since $h \circ f = 1_X$, this is equal to g, and since $f \circ g = 1_Y$, it is also equal to h.■

Functors

The examples of categories illustrate a basic principle in modern mathematics: Whenever one defines a type of mathematical system, there is usually a corresponding type of morphism for such systems (and in some cases there are several reasonable choices for morphisms). Since a category is an example of a mathematical system, it is natural to ask whether there is a corresponding notion of morphisms in this case too. In fact, there are two concepts of morphism that turn out to be important and useful. We shall start with the simpler one.

Definition. Let $\mathbf{C}$ and $\mathbf{D}$ be categories. A *covariant functor* assigns (i) to each object X of $\mathbf{C}$ an object $T(X)$ of $\mathbf{D}$, (ii) to each morphism $f : X \to Y$ in $\mathbf{C}$ a morphism $T(f) : T(X) \to T(Y)$ in $\mathbf{D}$ such that the following hold:

(1) For each object X in $\mathbf{C}$ we have $T(1_X) = 1_{T(X)}$.

(2) For each pair of morphisms f and g in $\mathbf{C}$ such that $g \circ f$ is defined, we have $T(g \circ f) = T(g) \circ T(f)$.

HISTORICAL TRIVIA. Eilenberg and MacLane "borrowed" the word **category** from the philosophical writings of the 18^{th} century German philosopher I. Kant and the word **functor** from the philosophical writings of the 20^{th} century German-American philosopher R. Carnap, who was strongly influenced by Kant's writings on the philosophy of science.

Examples of covariant functors

Numerous constructions from undergraduate and elementary graduate courses can be interpreted as functors; in many cases this does not shed much additional light on the objects constructed, but in other cases the concept does turn out to be extremely useful.

1. Given a category $\mathbf{C}$, there is the identity functor from $\mathbf{C}$ to itself, which takes all objects and morphisms to themselves.

2. Given a category $\mathbf{C}$ and another nonempty category $\mathbf{D}$, for each object A of $\mathbf{D}$ there is a constant functor k_A from $\mathbf{C}$ to $\mathbf{D}$ which sends every object of $\mathbf{C}$ to A and every morphism to the identity morphism 1_A.

3. In categories where the objects are given by sets with some extra structure and the morphisms are ordinary functions with additional properties, there are **forgetful functors** which take objects to the underlying sets and morphisms to the underlying mappings of sets. For example, there are forgetful functors from $\mathbf{VEC_F}$, **GRP**, **POSETS**, and **TOP** to **SETS**. Likewise, there is an obvious forgetful functor from **MET–UNIF** to **TOP** which takes a metric space to its underlying topological space and simply views a uniformly continuous mapping as a continuous mapping.

4. There is a **power set functor** P_* on the category **SETS** defined as follows: The set $P_*(X)$ is just the set of all subsets (also known as the power set), and if $f : X \to Y$ is a set-theoretic function, then $P_*(f) : P_*(X) \to P_*(Y)$ takes an element $A \in P(X)$ — which by definition is just a subset of X — to its image $f[A] \subset Y$. A short argument is needed to verify this construction actually defines a covariant functor, but it is elementary. First, we need to check that for every set X we have $P_*(1_X) = 1_{P(X)}$; this follows because $1_X[A] = A$ for all $A \subset X$. Next, we must check that $P_*(g \circ f) = P_*(g) \circ P_*(f)$ for all composable f and g. But this is a consequence of the elementary identity $g[\,f[A]\,] \ = \ g \circ f[A]$.

5. If we are given two partially ordered sets and a mapping f from the first to the second such that $u \leq v$ implies $f(u) \leq f(v)$, then f may be interpreted as a covariant functor on the associated categories.

6. If we are given two groups and a homomorphism f from the first to the second, then f may be interpreted as a covariant functor on the associated categories.

7. Finally, we shall give a more substantial example that played a central role in mathematics 205B. Define a new category $\mathbf{TOP}_*$ of *pointed topological spaces* whose objects are pairs (X, y), where X is a topological space and $y \in X$; the point y is said to be the *basepoint* of the pointed space. A morphism $f : (X, y) \to (Z, w)$ in this category will be a continuous mapping from X to Z (usually also denoted by f) which maps y to w (*i.e.*, a **basepoint preserving** continuous mapping). The fundamental group $\pi_1(X, y)$ then has a natural interpretation as a covariant functor, for if f is a morphism of pointed spaces, then then one has an associated homomorphism f_* from $\pi_1(X, y)$ to $\pi_1(Z, w)$, and these have the required properties that $1_{(X,y)*}$ is the identity and $(g \circ f)_* = g_* \circ f_*$.

Contravariant functors and examples

Experience shows there are many instances in which it is useful to work with functors that **reverse** the order of function composition; such objects are called *contravariant functors*.

Definition. Let **C** and **D** be categories. A *contravariant functor* assigns (i) to each object X of **C** an object $U(X)$ of **D**, (ii) to each morphism $f : X \to Y$ in **C** a morphism $U(f) : U(Y) \to U(X)$ in **D** (note that the domain and codomain are the opposites of those in the covariant case!) such that the following hold:

(1) For each object X in **C** we have $U(1_X) = 1_{U(X)}$.

(2) For each pair of morphisms f and g in **C** such that $g \circ f$ is defined, we have $U(g \circ f) = U(f) \circ U(g)$.

The simplest examples of contravariant functors are given by the *pseudo-identity functors*, which map the objects and morphisms in the category **C** to their obvious counterparts in the opposite category $\mathbf{C^{OP}}$. In fact, there is an obvious correspondence between contravariant functors from **C** to **D** and covariant functors from **C** to $\mathbf{D^{OP}}$, or equivalently covariant functors from $\mathbf{C^{OP}}$ to **D**. The best way to motivate the definition is to give some less trivial examples.

1. Let $\mathbf{C}$ be the category of all vector spaces over some field, and consider the construction which associates to each vector space its *dual space* V^* of linear mappings from V to the scalar field F. There is a simple way of defining a corresponding construction for morphisms; if $L : V \to W$ is a linear transformation, consider the linear transformation $L^* : W^* \to V^*$ whose value on a linear functional $h : W \to F$ is given by $L^*(h) = h \circ L$, which is a linear functional on V. Standard results in linear algebra show that L^* is a linear transformation, that L^* is an identity map if L is an identity map, and if L is a composite $L_1 \circ L_2$, then we have $L^* = L_2^* \circ L_1^*$.

2. There is a contravariant power set functor P^* on the category **SETS** defined as follows: As before, the set $P^*(X)$ is just the set of all subsets, but now if $f : X \to Y$ is a set-theoretic function, then $P^*(f) : P^*(Y) \to P^*(X)$ takes an element $B \in P(Y)$ — which by definition is just a subset of Y — to its **inverse image** $f^{-1}[B] \subset X$. As in the case of P_*, a short elementary argument is needed to verify this construction actually defines a contravariant functor. The construction preserves identity maps because $1_X^{-1}[B] = B$ for all $B \subset X$, and the identity $P_*(g \circ f) = P^*(f) \circ P^*(g)$ is essentially a restatement of the elementary identity $f^{-1}[\,g^{-1}[B]\,] \;=\; (g \circ f^{-1}[B]$.

3. The preceding example actually yields a little more. Define a *Boolean algebra* to be a set with two binary operations $\cap$ and $\cup$, a unary operation $x \to x'$, and special elements 0 and 1 such that the system satisfies the usual properties for unions, intersections, and complementation for the algebra $P(X)$ of subsets of a set X, where 0 corresponds to the empty set and 1 corresponds to X. One then has an associated category **BOOL–ALG** whose objects are Boolean algebras and whose morphisms preserve unions, intersection, complementation, and the special elements. Obviously each power set $P(X)$ is a Boolean algebra, and in fact P^* defines a contravariant functor from **SETS** to **BOOL–ALG**. — In contrast, the covariant functor P_* does NOT define such a functor because $P_*(f)$ does not preserves intersections even though it does preserve unions (for example, we can have $f[A] \cap f[B] \neq \emptyset$ when $A \cap B = \emptyset$).

4. The desirability of having both contravariant and covariant functors is illustrated by the following examples. Given a category $\mathbf{C}$, modulo foundational questions we can informally view the set $\mathrm{Morph}_{\mathbf{C}}(X, Y)$ of morphisms from X to Y as a function of two variables on $\mathbf{C}$. What happens if we hold one of these variables constant to get a single variable construction? — Suppose first that we hold X constant and set $A_X(Y) = \mathrm{Morph}_{\mathbf{C}}(X, Y)$. Then we can make A_X into a covariant functor aas follows: Given a morphism $g : Y \to Z$, let $A_X(g)$ take $f \in A_X(Y) = \mathrm{Morph}_{\mathbf{C}}(X, Y)$ to the composite $g \circ f$. The axioms for a category then imply that $A_X(1_Y)$ is the identity and that $A_X(h \circ g) = A_X(h) \circ A_X(g)$ if g and h are composable. — Now suppose that we hold Y constant and set $B_Y(X) = \mathrm{Morph}_{\mathbf{C}}(X, Y)$. Then we can make B_Y into a **contravariant** functor as follows: Given a morphism $k : W \to X$, let $B_Y(g)$ take $f \in B_Y(X) = \mathrm{Morph}_{\mathbf{C}}(X, Y)$ to the composite $f \circ k$. The axioms for a category then imply that $B_Y(1_X)$ is the identity and that $B_Y(k \circ h) = B_Y(h) \circ B_Y(k)$ if h and k are composable.

5. In the preceding example, suppose that $\mathbf{C}$ is the category of topological spaces and continuous mappings, and let Y be the real numbers with the usual topology. In this case the contravariant functor B_Y has the algebraic structure of a commutative ring with unit given by pointwise multiplication of continuous real valued functions, and if $f : W \to X$ is continuous then $B_Y(f)$ is in fact a homomorphism of commutative rings with unit. Therefore, if we define a category of continuous rings with unit (whose morphisms are unit preserving homomorphisms), it follows that B_Y defines a functor from topological

spaces and continuous mappings to commutative rings with unit. — In contrast, there is no comparable structure for the covariant functor B_X if X is the real numbers.

Properties of functors

One of the most important properties of functors is that they send isomorphic objects in one one category to isomorphic objects in the other.

PROPOSITION 2. *Let* $\mathbf{C}$ *and* $\mathbf{D}$ *be categories, let* $T : \mathbf{C} \to \mathbf{D}$ *be a (covariant or contravariant) functor, and let* $f : X \to Y$ *be an isomorphism in* $\mathbf{C}$. *Then* $T(f)$ *is an isomorphism in* $\mathbf{D}$. *Furthermore, if* g *is the inverse to* f, *then* $T(g)$ *is the inverse to* $T(f)$.

Proof. CASE 1. Suppose the functors are covariant. Then we have

$$1_{T(X)} = T(1_X) = T(g \circ f) = T(g) \circ T(f)$$

$$1_{T(Y)} = T(1_Y) = T(f \circ g) = T(f) \circ T(g)$$

and hence $T(g)$ is inverse to $T(f)$. — *CASE 2.* Suppose that the functors are contravariant. Then we have

$$1_{T(X)} = T(1_X) = T(g \circ f) = T(f) \circ T(g)$$

$$1_{T(Y)} = T(1_Y) = T(f \circ g) = T(g) \circ T(f)$$

and hence $T(g)$ is inverse to $T(f)$.■

The next result states that a composite of two functors is also a functor.

PROPOSITION 3. *Suppose that* $\mathbf{C}$, $\mathbf{D}$ *and* $\mathbf{E}$ *are categories and that* $F : \mathbf{C} \to \mathbf{D}$ *and* $G : \mathbf{D} \to \mathbf{E}$ *are functors (in each case, the functor may be covariant or contravariant). Then the composite* $G \circ F$ *also defines a functor; this functor is covariant if* F *and* G *are both covariant or contravariant, and it is contravariant if one of* F, G *is covariant and the other is contravariant.*■

This result has a curious implication:

COROLLARY 4. *There is a "category of small categories"* **SMCAT** *whose objects are small categories and whose morphisms are covariant functors from one small category to another.*■

SEMANTIC TRIVIA. (For readers who are familiar with contravariant and covariant tensors.) In the applications of linear algebra to differential geometry and topology, one often sees objects called *contravariant tensors* and *covariant tensors*, and for finite-dimensional vector spaces these are given by finitely iterated tensor products $V \otimes \cdots \otimes V$ of V with itself in the contravariant case and similar objects involving V^* in the covariant case; for our purposes it will suffice to say that if U and W are vector spaces with bases $\{ \mathbf{u}_i \}$ and $\{ \mathbf{w}_j \}$ respectively, then their tensor product $U \otimes W$ is a vector space having a basis of the form $\{ \mathbf{u}_i \otimes \mathbf{w}_j \}$ where i and j are allowed to vary independently (hence the dimension of $U \otimes W$ is $[\dim U] \cdot [\dim W]$). Since the identity functor on the category of vector spaces is covariant and the dual space functor is covariant, at first it might seem that something is the opposite of what it should be. However, the classical tensor notation refers to the manner in which the **coordinates** transform; now coordinates for a vector space may be viewed linear functionals on that space, or equivalently as elements of the dual space, which is contravariant. Therefore individual coordinates on $V \otimes \cdots \otimes V$ correspond to elements of the *dual space* of the latter, and in fact the construction which associates the space $(V \otimes \cdots \otimes V)^*$ to V defines a *contravariant* functor on the category of finite-dimensional vector spaces over the given scalars; likewise, the construction which associates the space $(V^* \otimes \cdots \otimes V^*)^*$ to V defines a **covariant** functor on the category of finite-dimensional vector spaces over the given scalars.

Natural transformations

The final concept in category theory to be considered here is the notion of **natural transformation** from one functor to another. In fact, the motivation for category theory in the work of Eilenberg and MacLane was a need to discuss "natural mappings" in a mathematically precise manner. There are actually two definitions, depending whether both functors under consideration are covariant or contravariant.

Definition. Let $\mathbf{C}$ and $\mathbf{D}$ be categories, and let F and G be covariant functors from $\mathbf{C}$ to $\mathbf{D}$. A *natural transformation* θ from F to G associates to each object X in $\mathbf{C}$ a morphism $\theta_X : F(X) \to G(X)$ such that for each morphism $f : X \to Y$ we have $\theta_Y \circ F(f) = G(f) \circ \theta_X$.

The morphism identity is often expressed graphically by saying the the diagram

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ \downarrow{\theta_X} & & \downarrow{\theta_Y} \\ G(X) & \xrightarrow{G(f)} & G(Y) \end{array}$$

is a **commutative diagram**. The idea is that all paths of arrows from one object-vertex to another yield the same function.

The definition of a natural transformation of contravariant functors is similar.

Definition. Let $\mathbf{C}$ and $\mathbf{D}$ be categories, and let T and U be contravariant functors from $\mathbf{C}$ to $\mathbf{D}$. A *natural transformation* θ from F to G associates to each object X in $\mathbf{C}$ a morphism $\theta_X : T(X) \to U(X)$ such that for each morphism $f : X \to Y$ we have $\theta_X \circ T(f) = U(f) \circ \theta_Y$.

Here is the corresponding commutative diagram:

$$\begin{array}{ccc} T(Y) & \xrightarrow{T(f)} & T(X) \\ \downarrow{\theta_Y} & & \downarrow{\theta_X} \\ U(Y) & \xrightarrow{U(f)} & U(X) \end{array}$$

Once again we need to give some decent examples

1. Given any functor $T : \mathbf{C} \to \mathbf{D}$, there is an obvious identity transformation j^T from T to itself; specifically, j^T_X is the identity map on $T(X)$.

2. Let $\mathbf{C}$ be one of the categories as above for which a diagonal functor can be defined. Then there is a natural diagonal transformation Δ from the identity to the diagonal functor such that for each object X the mapping $\Delta_X : X \to X \times X$ is the diagonal map.

3. On the category of vector spaces over some field F, one can iterate the dual space functor to obtain a covariant *double dual space functor* $(V^*)^*$. There is a natural transformation $e_V : V \to (V^*)^*$ defined as follows: For each $\mathbf{v} \in V$, let $e_V(\mathbf{v}) : V^* \to F$ be the linear function given by evaluation at v; in other words, the value of $e_V(\mathbf{v})$ on a linear functional f is given by $f(\mathbf{v})$. If V is finite-dimensional, this map is an isomorphism.

Note that if V is finite-dimensional then V and its dual space V^* are isomorphic, but the isomorphism depends upon some additional data such as the choice of a basis or an inner product. In contrast, *the natural isomorphism e_V does not depend upon any such choices*.

4. In the category of sets or topological spaces and continuous mappings, let A be an arbitrary object and define functors L_A and R_A such that $L_A(X) = A \times X$ and $R_A(X) = X \times A$. One can make these into covariant functors by sending the morphism $f : X \to Y$ to $L_A(X) = 1_A \times f$ and $R_A(f) = f \times 1_A$. There is an obvious natural transformation $t : L + A \to R_A$ such that $t_A(X) : A \times X \to X \times A$ sends (a, x) to (x, a) for all $a \in A$ and $x \in X$, and it is an elementary exercise to verify that this is a natural transformation such that each map $t_A(X)$ is an *isomorphism*; in other words, t_A is a natural isomorphism from the functor L_A to the functor R_A.

5. For the morphism examples A_X and B_Y discussed previously, if $h : W \to X$ is a morphism in the category, then it defines a natural transformation $h^* : A_X \to A_W$ which sends $f \in A_X(Y)$ to $f \circ h \in A_W(Y)$; the naturality condition follows from associativity of composition. Similarly, if $g : Y \to Z$ is a morphism then there is a natural transformation $g_* : B_Y \to B_Z$ sending f to $g \circ f$; once again, the key naturality condition follows from the associativity of composition. Furthermore, h^* is a natural isomorphism if h is an isomorphism and g_* is a natural isomorphism if g_* is an isomorphism,

A basic exercise in category theory is to prove the following:

PROPOSITION 5. *There are $1 - 1$ correspondences between natural transformations from A_X to A_W and morphisms from W to X and between natural transformations from B_Y to B_Z and morphisms from Y to Z.*

Sketch of proof. The main point is to retrieve the function from the natural transformation. Given $\theta : A_X \to A_W$, one does this by considering the image of 1_X, and given $\varphi : B_Y \to B_Z$, one does this by considering the image of 1_Y.∎

Finally, we have the following result on natural isomorphisms (*i.e.*, natural transformations θ such that each map θ_X is an isomorphism:

PROPOSITION 6. *Let $\theta : F \to G$ be a natural transformation such that for each object X the map θ_X is an isomorphism. The there is a natural transformation $\varphi_X : G \to F$ such that for each X the map φ_X is inverse to θ_X.*

Proof. The main thing to check is that the relevant diagrams are commutative; we shall only do the case where F and G are covariant, leaving the other case to the reader. Since $\theta_X \circ \varphi_X$ is the identity on $G(X)$ and $\varphi_X \circ \theta_X$ is the identity on $F(X)$, we have

$$\theta_Y \circ \varphi_Y \circ G(f) \quad = \quad G(f) \quad = \quad g(f) \circ \theta_X \circ \varphi_X \quad = \quad \theta_Y \circ F(f) \circ \varphi_X$$

and if we compose with the inverse θ_X on the left of these expressions we obtain

$$\varphi_Y \circ G(f) \quad = \quad F(f) \circ \varphi_X$$

which is the naturality condition.∎

We say that two functors are *naturally isomorphic* if there is a natural isomorphism from one to the other.

Equivalences of categories

One can obviously define an isomorphism of categories to be a covariant functor $T : \mathbf{C} \to \mathbf{D}$ for which there is an inverse covariant functor $U : \mathbf{D} \to \mathbf{C}$ such that the composites $T \circ U$ and $U \circ T$ are the identities on $\mathbf{C}$ and $\mathbf{D}$ respectively. However, there is a less rigid notion of category equivalence that suffices for most purposes.

Definition. A covariant functor $T : \mathbf{C} \to \mathbf{D}$ is a *category equivalence* (or equivalence of categories) if there is a covariant functor $U : \mathbf{D} \to \mathbf{C}$ such that the composites $T \circ U$ and $U \circ T$ are *naturally isomorphic to* the identities on $\mathbf{C}$ and $\mathbf{D}$ respectively.

In particular, if T and U define an equivalence of categories, then every object in $\mathbf{D}$ is isomorphic to an object of the form $T(X)$, and conversely every object in $\mathbf{C}$ is isomorphic to an object of the form $U(A)$.

I.2 : Barycentric coordinates and polyhedra

(Hatcher, § 2.1)

Drawings to illustrate many of the concepts in this and other sections of the notes can be found in the following document(s):

`http://math.ucr.edu/~res/math246A/algtopfigures.*`

Here the suffix $*$ is one of `doc`, `ps` or `pdf`.

A more leisurely and detailed discussion of barycentric coordinates, and more generally the use of linear algebra to study geometric problems, is contained in Section I.4 of the following online document, in which $*$ is one of the options in the preceding paragraph:

`http://math.ucr.edu/~res/math133/geomnotes1.*`

The files `math133exercises1.*`, `math133solutions1.*` and `solvedproblems`n`.*`, where $n =$ 1 or 2, in the directory

`http://math.ucr.edu/~res/math133`

contain further material on these topics.

Affine independence and barycentric coordinates

The crucial algebraic information is contained in the following result.

PROPOSITION 1. *Suppose that the ordered set of vectors $\mathbf{v}_0, \cdots, \mathbf{v}_n$ lie in some vector space V. Then the vectors $\mathbf{v}_1 - \mathbf{v}_0, \cdots, \mathbf{v}_n - \mathbf{v}_n$ are linearly independent if and only if every vector $\mathbf{x} \in V$ has at most one expansion of the form $t_0\mathbf{v}_0 + \cdots + t_n\mathbf{v}_n$ such that $t_0 + \cdots + t_n = 1$.*

A finite ordered set of vectors satisfying either (hence both) conditions is said to be *affinely independent*. Note that since the second condition does not depend upon the choice of ordering, a set of vectors is affinely independent if and only if for some arbitrary j the vectors $\mathbf{v}_i - \mathbf{v}_j$ (where $i \neq j$) is linearly independent. A linear combination in which the coefficients add up to 1 is called an *affine combination*.

Sketch of proof. To show the first statement implies the second, use the fact that $\mathbf{x} - \mathbf{v}_0$ has at most one expansion as a linear combination of $\mathbf{v}_1 - \mathbf{v_0}, \cdots, \mathbf{v}_n - \mathbf{v}_n$. To prove the reverse implication, show that if $\mathbf{x} - \mathbf{v}_0$ has more than one expansion as a linear combination of $\mathbf{v}_1 - \mathbf{v_0}, \cdots, \mathbf{v}_n - \mathbf{v}_n$, then $\mathbf{x}$ has more than one expansion as an affine combination of $\mathbf{v}_0, \cdots, \mathbf{v}_n$.■

COROLLARY 2. *If $S = \{\mathbf{v}_0, \cdots, \mathbf{v}_n\}$ is affinely independent, then every nonempty subset of S is affinely independent.*

This follows immediately from the uniqueness of expansions of vectors as affine combinations of vectors in S.■

The coefficients t_i are called **barycentric coordinates**. If we put physical weights of t_i units at the respective vertices $\mathbf{v}_i$, then the center of gravity for the system will be at the point $t_0\mathbf{v}_0 + \cdots + t_n\mathbf{v}_n$. If, say, $n = 2$, then this center of gravity will be inside the triangle with the given three vertices if and only if each t_i is positive, and it will be on the triangle defined by these vertices if and only if each t_i is nonnegative and at least one is equal to zero.

More generally, if $\mathbf{v}_0, \cdots, \mathbf{v}_n$ are affinely independent then the **n-simplex** with vertices $\mathbf{v}_0, \cdots, \mathbf{v}_n$ is the set of all points expressible as affine combinations such that each coefficient is nonnegative (*i.e.*, **convex combinations**).

Frequently the n-simplex described above will be denoted by $\mathbf{v}_0 \cdots \mathbf{v}_n$. Note that if $n = 0$, then a 0-simplex consists of a single point, while a 1-simplex is a closed line segment, a 2-simplex is given by a triangle and the points that lie "inside" the triangle (also called a *solid triangle*), and a 3-simplex is given by a pyramid with a triangular base (*i.e.*, a *tetrahedron*) together with the points inside this pyramid (also called a *solid tetrahedron*).

The following definition will also play an important role in our discussions.

Definition. If $\mathbf{v}_0, \cdots, \mathbf{v}_n$ form the vertices of a simplex $\mathbf{v}_0 \cdots \mathbf{v}_n$, then the **faces** of this simples are the simplices whose vertices are given by proper subsets of $\{\mathbf{v}_0, \cdots, \mathbf{v}_n\}$; note that such proper subsets are affinely independent by Corollary 2. If a proper subset $T \subset S$ has $k+1$ elements, then we shall say that the simplex $\Delta(T)$ whose vertices are given by T is a **k-face** of the original n-simplex, which in this notation is equal to $\Delta(S)$.

Sets with simplicial decompositions

In calculus textbooks, the derivation of Green's Theorem is often completed only for special sorts of closed regions such as the simplex whose vertices are $(0,0)$, $(1,0)$ and $(1,1)$. One then finds discussions indicating how the general case can be retrieved from special cases by splitting a general region into pieces that are nicely homeomorphic to closed regions of the special type; in particular, there is one such discussion on page 523 of the text by Marsden and Tromba, and it is taken further in the online document with figures for these notes (see Figure I.2.8 in the document `algtopfigures.pdf`).

Here are the formal descriptions.

Definition. A subset $P \subset \mathbf{R}^m$ is a *polyhedron* if
(i) P is a finite union of simplices $A_1, \cdots, A_q$,
(ii) For each pair of indices $i \neq j$, the intersection $A_i \cap A_j$ is a common face.

The simplices $A_1, \cdots, A_q$ are said to form a *simplicial decomposition* of P, and if $\mathbf{K}$ is the collection of simplices given by the A_j and all their faces, then the ordered pair $(P, \mathbf{K})$ is called a (finite) *simplicial complex*.

If X is an arbitrary topological space, then a (finite) *triangulation* of X consists of a simplicial complex $(P, \mathbf{K})$ and a homeomorphism $t : P \to X$.

With these definitions, we can say that Green's Theorem holds for "decent" closed plane regions because Such regions have nice triangulations.

SIMPLE EXAMPLE. Consider the solid rectangle in the plane given by $[a, b] \times [c, d]$, where $a < b$ and $c < d$. Everyday geometrical experience shows this can be split into two 2-simplices along a diagonal, and in fact it is the union of two 2-simplices, one with vertices (a, c), (a, d) and (b, d), and the other with vertices (a, c), (b, c) and (b, d). A point (x, y) which lies in the solid rectangle will be in the first simplex if and only if

$$(y - c)(b - a) \quad \leq \quad (d - c)(x - a)$$

and this point will be in the second simplex if and only if

$$(y - c)(b - a) \quad \geq \quad (d - c)(x - a)$$

Generalizations of this example will play an important role in the standard approach to algebraic topology.

If $(P, \mathbf{K})$ is a simplicial complex, then a subset $\mathbf{L} \subset \mathbf{K}$ is said to be a *subcomplex* if $\sigma \in \mathbf{L}$ implies that every face of σ also lies in $\mathbf{L}$. The union of the simplices in $\mathbf{L}$ is a closed subspace of P which is denoted by $|\mathbf{L}|$. With this notation we have $P = |\mathbf{K}|$.

Decompositions of prisms

The rectangular example has the following important generalization:

PROPOSITION 3. *Suppose that $A \subset \mathbf{R}^m$ is a simplex with vertices $\mathbf{v}_0, \cdots, \mathbf{v}_n$. Then $A \times [0, 1] \subset \mathbf{R}^{m+1}$ has a simplicial decomposition with exactly $n + 1$ simplices of dimension $n + 1$.*

Proof. For each i between 0 and n let $\mathbf{x}_i = (\mathbf{v}_i, 0)$ and $\mathbf{y}_i = (\mathbf{v}_i, 1)$. We claim that the vectors

$$\mathbf{x}_0, \cdots, \mathbf{x}_i, \mathbf{y}_i \cdots, \mathbf{y}_n$$

are affinely independent and the corresponding simplices

$$\mathbf{x}_0 \cdots \mathbf{x}_i \mathbf{y}_i \cdots \mathbf{y}_n$$

(where $0 \leq i \leq n$) form a simplicial decomposition of $A \times [0, 1]$.

An illustration for the case $n = 2$ is given in Figure I.2.11 of `algtopfigures.pdf`).

To prove affine independence, take a fixed value of i and suppose we have

$$\sum_{j<i} t_j \mathbf{x}_j \quad + \quad a \mathbf{x}_i \ + \ b \mathbf{y}_i \quad + \quad \sum_{j>i} t_j \mathbf{y}_j \quad =$$

$$\sum_{j<i} t'_j \mathbf{x}_j \quad + \quad a' \mathbf{x}_i \ + \ b' \mathbf{y}_i \quad + \quad \sum_{j>i} t'_j \mathbf{y}_j$$

where the coefficients in each expression add up to 1; the summation will be taken to be zero if the limits reduce to $j < 0$ or $j > n$. If we view $\mathbf{R}^{m+1}$ as $\mathbf{R}^m \times \mathbf{R}$ and project down to $\mathbf{R}^m$ we obtain the equation

$$\sum_{j<i} t_j\,\mathbf{v}_j \;+\; (a+b)\,\mathbf{x}_i \;+\; \sum_{j>i} t_j\,\mathbf{v}_j \;=\; \sum_{j<i} t'_j\,\mathbf{v}_j \;+\; (a'+b')\,\mathbf{v}_i \;+\; \sum_{j>i} t'_j\,\mathbf{v}_j$$

and by the affine independence of the vectors $\mathbf{v}_k$ it follows that $t_j = t'_j$ if $j \neq i$ and also that $a + b = a' + b'$. On the other hand, if we project down to the second coordinate (the copy of $\mathbf{R}$), then we obtain

$$b \;+\; \sum_{j>i} t_j \;=\; b' \;+\; \sum_{j>i} t'_j$$

and since $t_j = t'_j$ for all j it follows that $b = b'$. Finally, since the sum of all the coefficients is equal to 1, the preceding observations imply that $1 - a = 1 - a'$, and therefore we also have $a = a'$. Therefore the vectors

$$\mathbf{x}_0,\; \cdots\;, \mathbf{x}_i, \mathbf{y}_i\; \cdots\;, \mathbf{y}_n$$

are affinely independent.

We shall next check that every point $(\mathbf{z}, u) \in A \times [0,1]$ lies in one of the simplices

$$\mathbf{x}_0\;\cdots\;\mathbf{x}_i\mathbf{y}_i\;\cdots\;\mathbf{y}_n$$

listed above. Write $\mathbf{z} = \sum_j\, t_j\,\mathbf{v}_j$ where $t_j \geq 0$ for all j and $\sum\, t_j = 1$. It follows that $u \leq 1 = \sum_{j\geq 0}\, t_j$; let $i \leq n$ be the largest nonnegative integer such that $u \leq \sum_{j\geq i}\, t_j$. We claim that $(\mathbf{z}, u)$ lies in the simplex $\mathbf{x}_0\;\cdots\;\mathbf{x}_i\mathbf{y}_i\;\cdots\;\mathbf{y}_n$. Let $b = \sum_{j\geq i}\, t_j \;-\; u$, and let $a = u - \sum_{j>i}\, t_j \;=\; t_i - b$. Then we have $a,\; b \geq 0$, and

$$(\mathbf{z}, u) \;=\; \sum_{j<i} t_j\,\mathbf{x}_j \;+\; a\,\mathbf{x}_i \;+\; b\,\mathbf{y}_i \;+\; \sum_{j>i} t_j\,\mathbf{y}_j$$

where all the coefficients are nonnegative and add up to 1.

To conclude the proof, we need to show that the intersection of two simplices as above is a common face. Suppose that $k < i$ and

$$(\mathbf{z}, u) \;\in\; \Big(\mathbf{x}_0\;\cdots\;\mathbf{x}_i\mathbf{y}_i\;\cdots\;\mathbf{y}_n\Big) \;\cap\; \Big(\mathbf{x}_0\;\cdots\;\mathbf{x}_k\mathbf{y}_k\;\cdots\;\mathbf{y}_n\Big)\;.$$

Then we must have

$$\sum_{j\leq i} p_j\,\mathbf{x}_j \;+\; \sum_{j\geq i} q_j\,\mathbf{y}_j \;=\; \sum_{j\leq k} p'_j\,\mathbf{x}_j \;+\; \sum_{j\geq k} q'_j\,\mathbf{y}_j$$

where all the coefficients are nonnegative and the coefficients on each side of the equation add up to 1. If we project down to $\mathbf{R}^m$ we obtain $p_j + q_j \;=\; p'_j + q'_j$ for all j (by convention, we take a coefficient to be zero if it does not lie in the corresponding summation as above). It follows immediately that $p_j = p'_j$ if $j < k$, while $p_j = q'_j$ if $k < j < i$ and $q_j = q'_j$ if $j > i$. Furthermore, if we project down to the last coordinate we see that

$$u \;=\; \sum_{j\geq i} q_j \;=\; \sum_{j\geq k} q'_k\;.$$

Since $q_j = q'_j$ if $j > i$, it follows that

$$q_i \;\; = \sum_{k \le j \le i} q'_j$$

and since all the coefficients are nonnegative, it follows that $q_i \ge q'_i$. On the other hand, we also have $q'_i = p'_i + q'_i = p_i + q_i$, and hence we conclude that $q_i = q'_i$ and $p_i = 0$. Applying the first of these, we see that

$$0 \;\; = \sum_{k \le j < i} q'_j$$

and hence the nonnegativity of the coefficients implies that $q'_j = 0$ for all j such that $k \le j < i$. We also know that $p'_j = 0$ for $j > k$, and therefore it follows that $p'_j + q'_j = 0$ when $k < j < i$ The equations $p_j + q_j = p'_j + q'_j$ and the nonnegativity of all terms now imply that $p_j = q_j = 0$ when $k < j < i$.

The conclusions of the preceding paragraph imply that the point $(\mathbf{z}, u)$ actually lies on the simplex

$$\mathbf{x}_0 \;\cdots\; \mathbf{x}_k \mathbf{y}_i \;\cdots\; \mathbf{y}_n$$

and since the latter is a common face of $\mathbf{x}_0 \;\cdots\; \mathbf{x}_i \mathbf{y}_i \;\cdots\; \mathbf{y}_n$ and $\mathbf{x}_0 \;\cdots\; \mathbf{x}_k \mathbf{y}_k \;\cdots\; \mathbf{y}_n$ it follows that the $(n+1)$-simplices

$$\mathbf{x}_0 \;\cdots\; \mathbf{x}_i \mathbf{y}_i \;\cdots\; \mathbf{y}_n$$

(where $0 \le i \le n$) form a simplicial decomposition of $A \times [0,1]$.■

COROLLARY 4. *If $P \subset \mathbf{R}^m$ is a polyhedron, then $A \times [0,1] \subset \mathbf{R}^{m+1}$ is also a polyhedron.*

Before discussing the proof of this we note one important special case.

COROLLARY 5. *For each positive integer m, the hypercube $[0,1]^m \subset \mathbf{R}^m$ is a polyhedron.*

Proof of Corollary 5. If $m = 1$ this follows because the unit interval is a 1-simplex; by Corollary 4, if the result is true for $m = k$ then it is also true for $m = k+1$. Therefore the result is true for all m by induction.■

Proof of Corollary 4. Let $\mathbf{K}$ be a simplicial decomposition for P, and let $\mathbf{K}^*$ be obtained from $\mathbf{K}$ by including all the faces of simplices in $\mathbf{K}$. Choose a linear ordering of the vertices in $\mathbf{K}^*$ (note there are finitely many). For each vertex $\mathbf{v}$ of $\mathbf{K}^*$, as before let $\mathbf{x} = (\mathbf{v}, 0)$ and $\mathbf{y} = (\mathbf{v}, 1)$. Then $P \times [0,1]$ is the union of all simplices of the form

$$\mathbf{x}_0 \;\cdots\; \mathbf{x}_i \mathbf{y}_i \;\cdots\; \mathbf{y}_n$$

where $\mathbf{v}_i < \mathbf{v}_{i+1}$ with respect to the given linear ordering of the vertices in $\mathbf{K}^*$ **and also** the vertices $\mathbf{v}_i$ are the vertices of a simplex in $\mathbf{K}^*$. The set $P \times [0,1]$ is the union of these simplices by Proposition 3 and the fact that P is the union of the simplices $\mathbf{v}_0 \;\cdots\; \mathbf{v}_n$. The fact that these simplices form a simplicial decomposition will follow from the construction and the next result.

LEMMA 6. *Suppose that we have two polyhedra P_1 and P_2 in $\mathbf{R^q}$, and there exist simplicial decompositions $\mathbf{K}_1$ and $\mathbf{K}_2$ such that the following hold:*

(i) Both $\mathbf{K}_1$ and $\mathbf{K}_2$ are taking faces of simplices.

(ii) The set $\mathbf{L}_1$ of all simplices in $\mathbf{K}_1$ contained in $P_1 \cap P_2$ equals the set $\mathbf{L}_2$ of all simplices in $\mathbf{K}_2$, and this collection determines a simplicial decomposition of $P_1 \cap P_2$.

Then $\mathbf{K}_1 \cup \mathbf{K}_2$ determines a simplicial decomposition of $P_1 \cup P_2$.

The hypothesis clearly applies to the construction in Proposition 3, so Corollary 4 indeed follows once we prove Lemma 6.∎

Proof of Lemma 6. It follows immediately that $P_1 \cup P_2$ is the union of the points of the simplices in $\mathbf{K}_1 \cup \mathbf{K}_2$. Suppose now that we are given an intersection of two simplices in the latter. This intersection will be a common face if both simplices lie in either $\mathbf{K}_1$ or $\mathbf{K}_2$, so the only remaining cases are those where one simplex α lies in $\mathbf{K}_1$ and the other simplex β lies in $\mathbf{K}_2$.

We know that $\alpha \cap \beta$ us convex. Furthermore, by the hypotheses we know that $\alpha \cap \beta$ must be a union of simplices that are faces of both α and β. Therefore it follows that every point in $\alpha \cap \beta$ is a convex combination of the vertices which lie in $\alpha \cap \beta$, and consequently $\alpha \cap \beta$ is the common face determined by all vertices which lie in $\alpha \cap \beta$.∎

DEFAULT HYPOTHESIS. Unless specifically indicated otherwise, we shall assume that the set of simplices in a simplicial decomposition $\mathbf{K}$ is closed under taking faces. In order to justify this, we need to know that if $\mathbf{K}^*$ is obtained from $\mathbf{K}$ by adding all the faces of simplices in the latter, then the intersection of two simplices in $\mathbf{K}^*$ is a (possibly empty) common face. — To see this, suppose that α and β are simplices in $\mathbf{K}^*$, where α and β are faces of the simplices α' and β' in $\mathbf{K}$. If $\mathbf{x} \in \alpha \cap \beta$, then $\mathbf{x}$ is a convex combination of vertices in $\alpha' \cap \beta'$, and in fact these vertices must lie in both α and β. Since $\alpha \cap \beta$ is convex, it follows that $\alpha \cap \beta$ must be the simplex whose vertices lie in α and in β.∎

I.3 : Subdivisions

(Hatcher, § 2.1)

For many purposes it is convenient or necessary to replace a simplicial decomposition $\mathbf{K}$ of a polyhedron P with another decomposition $\mathbf{L}$ with smaller simplices. More precisely, we would like the smaller simplices in $\mathbf{L}$ to determine simplicial decompositions for each of the simplices in $\mathbf{K}$.

Simple examples

1. If P is a 1-simplex with vertices $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{K}$ is the standard decomposition given by P and the endpoints, then there is a subdivision $\mathbf{L}$ given by trisecting P; specifically, the vertices are given by $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} = \frac{2}{3}\mathbf{x} + \frac{1}{3}\mathbf{y}$, and $\mathbf{z} = \frac{1}{3}\mathbf{x} + \frac{2}{3}\mathbf{y}$, and the 1-simplices are $\mathbf{xw}$, $\mathbf{wz}$ and $\mathbf{zy}$. This is illustrated as Figure I.3.1 in the file `algtopfigures`.

2. Similarly, if $[a, b]$ is a closed interval in the real line and we are given a finite sequence $a = t_0 < \cdots < t_m = b$, then these points and the intervals $[t_{j-1}, t_j]$, where $1 \leq j \leq n$, form a subdivision of the standard decomposition of $[a, b]$.

3. If P is the 2-simplex with vertices $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, and $\mathbf{K}$ is the standard decomposition given by P and its faces, then there is an obvious decomposition $\mathbf{L}$ which splits P into two simplices $\mathbf{xyz}$ and $\mathbf{xyw}$, where $\mathbf{w} = \frac{1}{2}\mathbf{y} + \frac{1}{2}\mathbf{z}$ is the midpoint of the 1-simplex $\mathbf{yz}$. Similar eamples exist if we take $\mathbf{z} = a\mathbf{y} + (1 - a)\mathbf{z}$, where a is an arbitrary number such that $0 < a < 1$ (see Figure I.3.2 in the file `algtopfigures`).

Definition of subdivisions

Each of the preceding examples is consistent with the following general concept.

Definition. Let $(P, \mathbf{K})$ be a simplicial complex, and let $\mathbf{L}$ be a simplicial decomposition of P. Then $\mathbf{L}$ is called a (linear) *subdivision* of $\mathbf{K}$ if every simplex of $\mathbf{L}$ is contained in a simplex of $\mathbf{K}$.

The following observation is very elementary, but we shall need it in the discussion below.

PROPOSITION 0. *Suppose P is a polyhedron with simplicial decompositions $\mathbf{K}$, $\mathbf{L}$ and $\mathbf{M}$ such that $\mathbf{L}$ is a subdivision of $\mathbf{K}$ and $\mathbf{M}$ is a subdivision of $\mathbf{L}$. Then $\mathbf{M}$ is also a subdivision of $\mathbf{K}$.*■

Figure I.3.3 in `algtopfigures` depicts two subdivisions of a 2-simplex that are different from the one in Example 3 above. As indicated by Figure I.3.4 in the same document, in general if we have two simplicial decompositions of a polyhedron then neither is a subdivision of the other. However, it is possible to prove that if $\mathbf{K}$ and $\mathbf{L}$ are simplicial decompositions of the same polyhedron P, then there is a third decomposition which is a subdivision of both $\mathbf{K}$ and $\mathbf{L}$. Proving this requires more machinery than we need for other purposes, and since we shall not need the existence of such subdivisions in this course we shall simply note that one can prove this result using methods from the second part of the following book, which we shall call [Munkres2]:

> **J. R. Munkres**. Elementary differential topology. Lectures given at Massachusetts Institute of Technology, Fall, 1961. Revised edition. Annals of Mathematics Studies, No. 54. Princeton University Press, Princeton, N. J., 1966. ISBN: 0-691-09093-9.

SUBDIVISION AND SUBCOMPLEXES. These two concepts are related by the following elementary results.

PROPOSITION 1. *Suppose that $(P, \mathbf{K})$ is a simplicial complex and that $(P_1, \mathbf{K}_1)$ is a subcomplex of $(P, \mathbf{K})$. If $\mathbf{L}$ is a subdivision of $\mathbf{K}$ and $\mathbf{L}_1$ is the set of all simplices in $\mathbf{L}$ which are contained in P_1, then $(P_1, \mathbf{L}_1)$ is a subcomplex of $(P, \mathbf{L})$.*■

Recall our Default Hypothesis (at the end of Section I.2) that all simplicial decompositions should be closed under taking faces unless specifically stated otherwise.

COROLLARY 2. *Let P, $\mathbf{K}$ and $\mathbf{L}$ be as above, and let $A \subset P$ be a simplex of $\mathbf{K}$. Then $\mathbf{L}$ determines a simplicial decomposition of A.*■

Barycentric subdivisions

We are particularly interested in describing a systematic construction for subdivisions that works for all simplicial complexes and allows one to form decompositions for which the diameters of all the simplices are very small. This will generalize a standard method for partitioning an interval $[a, b]$ into small intervals by first splitting the interval in half at the midpoint, then splitting the two subintervals in half similarly, and so on. If this is done n times, the length of each interval in the subdivision is equal to $(b - a)/2^n$, and if $\varepsilon > 0$ is arbitrary then for sufficiently large values of n the lengths of the subintervales will all be less than ε.

The generalization of this to higher dimensions is called the **barycentric subdivision**.

Definition. Given an n-simplex $A \subset \mathbf{R}^m$ with vertices $\mathbf{v}_0, \cdots, \mathbf{v}_n$, the *barycenter* $\mathbf{b}_A$ of A is given by

$$\mathbf{b}_A = \frac{1}{n+1} \sum_{i=0}^{n} \mathbf{v}_i \ .$$

If $n \leq m \leq 3$, this corresponds to the physical center of mass for A, assuming the density in A is uniform.

Definition. If $P \subset \mathbf{R}^m$ is a polyhedron and $(P, \mathbf{K})$ is a simplicial complex, then the *barycentric subdivision* $\mathbf{B(K)}$ consists of all simplices having the form $\mathbf{b}_0 \cdots \mathbf{b}_k$, where (i) each $\mathbf{b}_j$ is the barycenter of a simplex $A_j \in \mathbf{K}$, (ii) for each $j > 0$ the simplex A_{j-1} is a face of A_j.

In order to justify this definition, we need to prove the following result:

PROPOSITION 3. *Let A be an n-simplex, suppose that we are given simplices $A_j \subset A$ such that $Aj - 1$ is a face of A_j for each j, and let $\mathbf{b}_j$ be the barycenter of A_j. Then the set of vertices $\{\mathbf{b}_0, \cdots, \mathbf{b}_q\}$ is affinely independent.*

Proof. We can extend the sequence of simplices $\{A_j\}$ to obtain a new sequence $C_0 \subset \cdots \subset C_n = A$ such that each C_k is obtained from the preceding one C_{k-1} by adding a single vertex, and it suffices to prove the result for the corresponding sequence of barycenters. Therefore we shall assume henceforth in this proof that each A_j is obtained from its predecessor by adding a single vertex and that A is the last simplex in the list.

It suffices to show that the vectors $\mathbf{b}_j - \mathbf{b}_0$ are linearly independent. For each j let $\mathbf{v}_{j_i}$ be the vertex in A_j that is not in its predecessor. Then for each $j > 0$ we have

$$\mathbf{b}_j - \mathbf{b}_0 \quad = \quad \left(\frac{1}{j+1} \sum_{k \leq j} \mathbf{v}_{i_k}\right) \; - \; \mathbf{v}_0 \quad = \quad \frac{1}{j+1} \sum_{k \leq j} \mathbf{v}_{i_k} - \mathbf{v}_{i_0} \; .$$

which is a linear combination of the linearly independent vectors $\mathbf{v}_{i_1} - \mathbf{v}_{i_0}, \cdots, \mathbf{v}_{i_j} - \mathbf{v}_{i_0}$ such that the coefficient of the last vector in the set is nonzero.

If we let $\mathbf{u}_k = \mathbf{v}_{i_k} - \mathbf{v}_{i_0}$, then it follows that for all $k > 0$ we have $\mathbf{b}_k - \mathbf{b}_0 = a_k \mathbf{u}_k + \mathbf{y}_k$, where $\mathbf{y}_k$ is a linear combination of $\mathbf{u}_1, \cdots, \mathbf{u}_{k-1}$ and $a_k \neq 0$. Since the vectors $\mathbf{u}_j$ are linearly independent, it follows that the vectors $\mathbf{b}_k - \mathbf{b}_0$ (where $0 < k \leq n$) are linearly independent and hence the vectors $\mathbf{b}_0, \cdots, \mathbf{b}_n$ are affinely independent.■

The simplest nontrivial examples of barycentric subdivisions are given by 2-simplices, and Figure I.3.6 in `algtopfigures` gives a typical example. We shall enumerate the simplices in such a barycentric subdivision using the definition. For the sake of definiteness, we shall call the simplex P and the vertices $\mathbf{v}_0$, $\mathbf{v}_1$ and $\mathbf{v}_2$.

(i) The 0-simplices are merely the barycenters $\mathbf{b}_A$, where A runs through all the nonemtpy faces of P and P itself. There are 7 such simplices and hence 7 vertices in $\mathbf{B(K)}$.

(ii) The 1-simplices have the form $\mathbf{b}_A \mathbf{b}_C$, where A is a face of C. There are three possible choices for the ordered pair $(\dim A, \dim C)$; namely, $(0,1)$, $(0,2)$ and $(1,2)$. The number of pairs $\{A,\ C\}$ for the case $(0,1)$ is equal to 6, the number for the case $(0,2)$ is equal to 3, and the number for the case $(0,1)$ is also equal to 3, so there are 12 different 1-simplices in $\mathbf{B(K)}$.

(iii) The 2-simplices have the form $\mathbf{b}_A \mathbf{b}_C \mathbf{b}_E$, where A is a face of C and C is a face of E. There are 6 possible choices for $\{A,\ C,\ E\}$.

Obviously one could carry out a similar analysis for a 3-simplex but the analysis would be more complicated.

Of course, it is absolutely essential to verify that barycentric subdivision actually defines simplicial decompositions.

THEOREM 4. *If $(P, \mathbf{K})$ is a simplicial complex and $\mathbf{B(K)}$ is the barcentric subdivision of $\mathbf{K}$, then $(P, \mathbf{B(K)}\,)$ is also a simplical complex (in other words, the collection $\mathbf{B(K)}$ determines a simplicial decomposition of P).*

Proof. We shall concentrate on the special case where P is a simplex. The general case can be recovered from the special case and Lemma I.2.6.

Suppose now that P is a simplex with vertices vertices $\mathbf{v}_0, \cdots, \mathbf{v}_n$. We first show that P is the union of the simplices in $\mathbf{B(K)}$. Given $\mathbf{x} \in P$, write $\mathbf{x}$ as a convex combination $\sum_j t_j \mathbf{v_j}$, and rearrange the scalars into a sequence

$$t_{k_0} \;\geq\; t_{k_1} \;\cdots\; \geq\; t_{k_n}$$

(this is not necessarily unique, and in particular it is not so if $t_u = t_v$ for $u \neq v$). For each i between 0 and n, let A_i be the simplex whose vertices are $\mathbf{v}_{k_0}, \cdots, \mathbf{v}_{k_i}$. We CLAIM that $x \in \mathbf{b}_0 \cdots \mathbf{b}_n$, where $\mathbf{b}_i$ is the barycenter of A_i.

Let $s_i = t_{k_i} - t_{k_{i+1}}$ for $0 \leq i \leq n-1$ and set $s_n = t_{k_n}$. Then $s_i \geq 0$ for all i, and it is elementary to verify that

$$\mathbf{x} \;=\; \sum_{i=0}^{n} (i+1)\, s_i\, \mathbf{b}_i \;, \qquad \text{where} \qquad \sum_{1=0}^{n} (i+1)\, s_i \;=\; \sum_{i=0}^{n} t_{k_i} \;=\; 1\;.$$

Therefore $\mathbf{x} \in \mathbf{b}_0 \cdots \mathbf{b}_n$, so that every point in A lies on one of the simplices in the barycentric subdivision.

To conclude the proof, we must show that the intersection of two simplices in $\mathbf{B(K)}$ is a common face. First of all, it suffices to show this for a pair of n-dimensional simplices; this follows from the argument following the Default Hypothesis at the end of Section I.2.

Suppose now that α and γ are n-simplices in $\mathbf{B(K)}$. Then the vertices of α are barycenters of simplices $A_0, \cdots, A_n$ where A_j has one more vertex than A_{j-1} for each j, and the vertices of γ are barycenters of simplices $C_0, \cdots, C_n$ where C_j has one more vertex than C_{j-1} for each j. Label the vertices of the original simplex as $\mathbf{v}_{i_0}, \cdots, \mathbf{v}_{i_n}$ where $A_j = \mathbf{v}_{i_0} \cdots \mathbf{v}_{i_j}$ and also as $\mathbf{v}_{k_0}, \cdots, \mathbf{v}_{k_n}$ where $C_j = \mathbf{v}_{k_0} \cdots \mathbf{v}_{k_j}$. The key point is to determine how $(i_0, \cdots, i_n)$ and $(k_0, \cdots, k_n)$ are related.

If $\mathbf{x}$ lies on the original simplex and $\mathbf{x}$ is written as a convex combination $\sum_j t_j \mathbf{v}_j$, then we have shown that $\mathbf{x} \in A$ if $t_{i_0} \leq \cdots \leq t_{i_n}$. In fact, we can reverse the steps in that argument to show that if $\mathbf{x} \in A$ then conversely we have $t_{i_0} \leq \cdots \leq t_{i_n}$. Similarly, if $\mathbf{x} \in C$ then $t_{k_0} \leq \cdots \leq t_{k_n}$. Therefore if $\mathbf{x} \in A \cap C$ then $t_{i_j} = t_{k_j}$ for all j. Choose $m_0, \cdots, m_q \in \{0, \cdots, n\}$ such that $t_{m_j} > t_{m_{j+1}}$, with the convention that $t_{n+1} = 0$, and split $\{0, \cdots, {}_n\}$ into equivalence classes $\mathcal{M}_0, \cdots, \mathcal{M}_q$ such that $\mathcal{M}_j$ is the set of all u such that $t_u = t_{m_j}$. It follows that $\mathbf{x}$ lies on the simplex $\mathbf{z}_0 \cdots \mathbf{z}_q$, where $\mathbf{z}_j$ is the barycenter of the simplex whose vertices are $\mathcal{M}_0 \cup \cdots \cup \mathcal{M}_j$. The vertices of this simplex are vertices of both A and C. Since $A \cap C$ is convex, this implies that it is the simplex whose vertices are those which lie in $A \cap C$, and thus $A \cap C$ is a face of both A and C.■

Terminology. Frequently the complex $(P, \mathbf{B(K)})$ is called the *derived complex* of $(P, \mathbf{K})$. The barycentric subdivision construction can be iterated, and thus one obtains a sequence of decompositions $\mathbf{B}^r(\mathbf{K})$. The latter is often called the r^{th} barycentric subdivision of $\mathbf{K}$ and $(P, \mathbf{B}^r(\mathbf{K}))$ is often called the r^{th} derived complex of $(P, \mathbf{K})$.

Given a metric space $(X, \mathbf{d})$, its *diameter* is the least upper bound of the distances $\mathbf{d}(y, z)$, where $y, z \in X$; if the set of distances is unbounded, we shall follow standard usage and say that the diameter is infinite or equal to ∞.

PROPOSITION 5. *Let $A \subset \mathbf{R}^n$ be an n-simplex with vertices $\mathbf{v}_0, \cdots, \mathbf{v}_n$. Then the diameter of A is the maximum of the distances $|\mathbf{v}_i - \mathbf{v}_j|$, where $0 \leq i, j \leq n$.*

Proof. Let $\mathbf{x}, \mathbf{y} \in A$, and write these as convex combinations $\mathbf{x} = \sum_j t_j \mathbf{v}_j$ and $\mathbf{y} = \sum_j s_j \mathbf{v}_j$. Then

$$\mathbf{x} - \mathbf{y} \quad = \quad \left(\sum_i s_i \right) \mathbf{x} \; - \; \left(\sum_j t_j \right) \mathbf{y} \quad = \quad \sum_{i,j} s_i t_j \, \mathbf{v}_j \; - \; \sum_{i,j} s_i t_j \, \mathbf{v}_i \; .$$

Therefore we have

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) \quad + \quad |\mathbf{x} - \mathbf{y}| \quad \leq \quad \Big| \sum_{i,j} s_i t_j \, (\mathbf{x}_j - \mathbf{x}_i) \Big| \quad \leq$$

$$\sum_{i,j} s_i t_j \, |\mathbf{v}_i - \mathbf{v}_j| \quad \leq \quad \sum_{i,j} \max |\mathbf{v}_k - \mathbf{v}_\ell| \quad = \quad \max |\mathbf{v}_k - \mathbf{v}_\ell|$$

as required.■

Definition. If $\mathbf{K}$ is a simplicial decomposition of a polyhedron P, then the *mesh* of $\mathbf{K}$, written $\mu(\mathbf{K})$, is the maximum diameter of the simplices in $\mathbf{K}$.

PROPOSITION 6. *In the preceding notation, the mesh of $\mathbf{K}$ is the maximum distance $|\mathbf{v} - \mathbf{w}|$, where $\mathbf{v}$ and $\mathbf{w}$ are vertices of some simplex in $\mathbf{K}$.*■

The main result in this discussion is a comparison of the mesh of $\mathbf{K}$ with the mesh of $\mathbf{B(K)}$.

PROPOSITION 7. *Suppose that $(P, \mathbf{K})$ be a simplicial complex and all simplices of $\mathbf{K}$ have dimension $\leq n$. Then*

$$\mu(\,\mathbf{B}(\mathbf{K})\,) \quad \leq \quad \frac{n}{n+1} \cdot \mu(\mathbf{K}) \; .$$

Before proving this result, we shall derive some of its consequences.

COROLLARY 8. *In the preceding notation, if $r \geq 1$ then*

$$\mu(\,\mathbf{B}^r(\mathbf{K})\,) \quad \leq \quad \left(\frac{n}{n+1} \right)^r \cdot \mu(\mathbf{K}) \; .■$$

COROLLARY 9. *In the preceding notation, if $\varepsilon > 0$ then there exists an r_0 such that $r \geq r_0$ implies $\mu(\,\mathbf{B}^r(\mathbf{K})\,) \; < \; \varepsilon$.*

Corollary 9 follows from Corollary 8 and the fact that $\lim_{r \to \infty} \frac{n}{n+1} \; = \; 0$.■

Proof of Proposition 7. By Proposition 5 and the definition of barycentric subdivision we know that $\mu(\,\mathbf{B}(\mathbf{K})\,)$ is the maximum of all distances $|\mathbf{b} - \mathbf{b}_C|$, where $\mathbf{b}_A$ and $\mathbf{b}_C$ are barycenters of simplices $A, \; C \in \mathbf{K}$ such that $A \subset C$. Suppose that A is an a-simplex and C is a c-simplex, so that $0 \leq a < c \leq n$. We then have

$$|\mathbf{b}_A - \mathbf{b}_C| \quad = \quad \left| \frac{1}{a+1} \sum_{\mathbf{v} \in A} \mathbf{v} \; - \; \frac{1}{c+1} \sum_{\mathbf{w} \in C} \mathbf{w} \right|$$

and as in the proof of Proposition 5 we have

$$\frac{1}{a+1}\sum_{\mathbf{v}\in A} \mathbf{v} \;-\; \frac{1}{c+1}\sum_{\mathbf{w}\in C} \mathbf{w} \;=\; \frac{1}{(a+1)(c+1)}\sum_{\mathbf{v},\mathbf{w}} (\mathbf{v}-\mathbf{w}) \;.$$

There are $(a+1)$ terms in this summation which vanish (namely, those for which $\mathbf{w} = \mathbf{v}$), and therefore we have

$$|\mathbf{b}_A - \mathbf{b}_C| \;=\; \left|\frac{1}{(a+1)(c+1)}\sum_{\mathbf{v}\neq\mathbf{w}} (\mathbf{v}-\mathbf{w})\right| \;\leq\; \frac{1}{(a+1)(c+1)}\sum_{\mathbf{v}\neq\mathbf{w}} |\mathbf{v}-\mathbf{w}| \;\leq$$

$$\frac{1}{(a+1)(c+1)} \cdot \max|\mathbf{v}-\mathbf{w}| \cdot \Big[\,(a+1)(c+1) - (a+1)\,\Big] \;=$$

$$\Big(\max|\mathbf{v}-\mathbf{w}|\Big)\cdot\left(1 \;-\; \frac{1}{c+1}\right) \;\leq\; \left(1 \;-\; \frac{1}{n+1}\right) \;.$$

At the last step we use $c \leq n$ and the fact that the function $1-(x/n)$ is an increasing function of x if $x > 0$. The inequality in the corolary follows directly from the precedng chain of inequalities.■

One further consequence of Proposition 7 will be important for our purposes.

COROLLARY 10. *Let $(P, \mathbf{K})$ be a simplicial complex, and let $\mathcal{W}$ be an open covering of P. Then there is a positive integer r_0 such that $r \geq r_0$ implies that every simplex of $\mu(\,\mathbf{B}^r(\mathbf{K})\,)$ is contained in an element of $\mathcal{W}$.*

Proof. By construction, P is a compact subset of a the metric space $\mathbf{R}^m$. Therefore the Lebesgue Covering Lemma implies the existence of a real number $\eta > 0$ such that every subset of diameter $< \eta$ is contained in an element of $\mathcal{W}$. If we choose $r_0 > 0$ such that $r \geq r_0$ implies $\mu(\,\mathbf{B}^r(\mathbf{K})\,) \;<\; \eta$, then $\mathbf{B}^r(\mathbf{K})$ will have the required properties.■

II. Homotopy and cell complexes

The notion of homotopy is introduced in Mathematics 205B, and it is central to both algebraic and geometric topology as well as many of the applications of topology to algebra and analysis. Part of the material is a review of topics from the second part of Munkres' book; some of the revies topics and most of the new material are also covered in Chapters 0 and 1 of Hatcher.

The new material covers two related topics. The first (in Section 3) is to describe generalizations of simplicial complexes called **cell complexes** that are more convenient for many purposes of algebraic topology, and the second (in Section 4) provides a fundamental illustration of the usefulness of such objects. One objective is an important result on the following central problem:

EXTENSION QUESTION. *Suppose that X and Y are topological spaces, that A is a subspace of X, and $g : A \to Y$ is continuous. Is there an extension of g to a continuous mapping $f : X \to Y$ (in other words, a continuous mapping f such that the restriction $f|A$ is equal to g?*

One of the main results in Section 4 provides an extremely useful answer to this question in terms of the main concepts of this unit: *If X is a cell complex and A is a subcomplex, then g has a continuous extension to X if and only if some mapping homotopic to g has such an extension.*

This and subsequent units of the notes will be less self-contained that Unit I, and there will be numerous references to Munkres or Hatcher for details.

II.1 : Homotopic mappings

(Hatcher, Ch. 0; Munkres, §§ 51, 58)

The general notion of homotopy for (continuous) mappings is defined on page 323 of Munkres and page 3 of Hatcher. Following standard practice we shall write $f \simeq g$ to indicate that f is homotopic to g. We shall state some basic properties of homotopic mappings that are particularly important for our purposes.

PROPOSITION 1. (Munkres, Lemma 51.1, p. 324.) *The binary relation $\simeq$ of homotopy on the set of continuous mappings from one topological space X to a second topological space Y is an equivalence relation.*■

In the proposition above, we allow the possibility that $X = Y$. The set of homotopy classes of continuous mappings from X to Y is generally denoted by $[X, Y]$.

PROPOSITION 2. (Munkres, Exercise 1, p. 330.) *If we are given continuous maps $f_0 \simeq f_1 : X \to Y$ and $g_0 \simeq g_1 : Y \to Z$, then $g_0 \circ f_0 \simeq g_1 \circ f_1$.*■

COROLLARY 3. *There is a category* **HTOP** *(the* homotopy category*) whose objects are topological spaces and whose morphisms are given by $[X, Y]$ such that if $u \in [X, Y]$ is represented by f and $v \in [X, Y]$ is represented by g, then $v \circ u = [g \circ f]$.*■

Not surprisingly, the identity morphism in $[X, X]$ is the homotopy class of the identity on X.

Given a continuous mapping $f : X \to Y$, then f represents an isomorphism in **HTOP** if and only if there is a mapping $g : Y \to X$ such that $g \circ f \simeq 1_X$ and $f \circ g \simeq 1_Y$. A mapping f which satisfies these properties is said to be a *homotopy equivalence.* — Since every map is homotopic to itself, it follows immediately that *every homeomorphism is a homotopy equivalence.*

Definition. Two topological spaces X and Y are *homotopy equivalent* if there is a homotopy equivalence from X to Y (in which case there is also a homotopy equivalence from Y to X). Note that the relation "X is homotopy equivalent to Y" is reflexive, symmetric and transitive. Frequently one also says that *X and Y have the same homotopy type.*

Special types of homotopy equivalences

We shall begin with a homotopy between to basic types of continuous mappings.

Definition. A *contracting homotopy* of a topological space X is a mapping $H : X \times [0,1] \to X$ such that $H(x,0) = x$ for all $x \in X$ and $H|X \times \{1\}$ is a constant mapping.

We shall say that a topological space is **contractible** if it admits a contracting homotopy.

An arbitrary topological space X is not necessarily contractible, and in some sense most spaces are not. For example, if X is the circle S^1 this is not the case because in $[S^1, S^1] \cong \pi_1(S^1, 1)$ the identity map and the constant map determine different homotopy classes. In fact, one can manufacture many similar examples using the following lemma.

PROPOSITION 4. *If A, B and C are topological spaces, then there is an isomorphism*

$$\theta : [A, B \times C] \quad \cong \quad [A,B] \times [A,C]$$

sending a homotopy class $[f]$ to the ordered pair $([p_B \circ f], [p_C \circ f])$, where $p_B : B \times C \to B$ and $p_C : B \times C \to C$ are the coordinate projections.

Sketch of proof. The mapping θ is well-defined by the preceding two results. It is onto, for if we are given an ordered pair of homotopy classes $([g],[h])$, then this class is $\theta([f])$, where $f : A \to B \times C$ is the unique continuous mapping such that $p_B \circ f = g$ and $p_C \circ f = h$. To see it is also 1–1, suppose $\theta([f]) = \theta([f'])$. Then there are homotopies $K : p_B \circ f \simeq p_B \circ f'$ and $L : p_C \circ f \simeq p_C \circ f'$, and if we take the map H whose projections onto B and C are K and L respectively, then H defines a homotopy from f to f'.■

COROLLARY 5. *If X is a nonempty topological space, then $X \times S^1$ is* NOT *contractible.*

The proof of this result is relatively simple and formal, but it is important to understand it because the argument reflects the viewpoint underlying much of algebraic topology.

Proof. It will suffice to show that the identity map on $X \times S^1$ is not homotopic to a constant map. Let $q : X \times S^1$ to S^1 be projection onto the second coordinate, let $j : S^1 \to X \times S^1$ project to the constant map on the first factor and to the identity on the second, and let k be a constant map from $X \times S^1$ to itself. If the identity on $X \times S^1$ is homotopic to a constant map, then we have

$$[\mathrm{id}(S^1)] \quad = \quad [q \circ j] \quad = \quad [q] \circ [j] \quad = \quad [q] \circ [\mathrm{id}(X \times S^1)] \circ [j] \quad =$$

$$[q] \circ [k] \circ [j] \quad = \quad [q \circ k \circ j] \quad = \quad [\text{constant}]$$

which contradicts the fact that the identity on S^1 is not homotopic to a constant. Therefore the identity on $X \times S^1$ cannot be homotopic to a constant map.■

One can clearly "leverage" this result to construct further examples; in particular, if T^k is the product of k copies of S^1, then an inductive argument combined with the preceding corollary shows that $X \times T^k$ is not contractible.■

Example. If K is a convex subset of $\mathbf{R}^n$, then K is contractible by a so-called *straight line homotopy*: Take an arbitrary point $\mathbf{y} \in K$ and set

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

so that H shrinks K down to $\{\mathbf{y}\}$ along the straight lines joining points $\mathbf{x} \in K$ to the chosen point $\mathbf{y}$.■

In the preceding example, the inclusion of $\{\mathbf{y}\}$ in K is a special case of the following general concept.

Definition. Let X be a topological space, and let $A \subset X$ with inclusion mapping i_A. Then A is said to be a *deformation retract* of X if there is a map $r : X \to A$ such that $r|A$ is the identity and $i_A \circ r_A$ is homotopic to the identity on X. — If there is a homotopy $H : i_a \circ r_A \simeq 1_X$ such that $H(a,t) = a$ for all $(a,t) \in A \times [0,1]$ (*i.e.*, the homotopy is fixed on A), we say that A is a *strong deformation retract of* X.

More generally, in a category $\mathbf{C}$, a morphism $f : X \to Y$ is said to be a *retract* if there is a morphism $g : Y \to X$ such that $g \circ f = 1_X$, and a morphism $h : A \to B$ is said to be a *retraction* if there is a morphism $k : B \to A$ such that $k \circ h = 1_B$. — If A is a deformation retract of X, then the inclusion i_A is a retract and the mapping r is a retraction.

Example. The sphere S^n is a strong deformation retract of $\mathbf{R}^{n+1} - \{0\}$. The standard choice of r in this case is given by $r(\mathbf{x}) = |\mathbf{x}|^{-1} \cdot \mathbf{x}$ and $i \circ r$ is homotopic to the identity by the straight line homotopy sending $(\mathbf{x},\, t)$ to $t\mathbf{x} + (1-t)r(\mathbf{x})$.

Counting homotopy classes

We shall conclude this section by proving a result mentioned earlier.

THEOREM 6. *If K is a compact subset of $\mathbf{R}^n$ for some n and U is an open subset of $\mathbf{R}^m$ for some m, then $[K, U]$ is countable.*

One major step in the proof is the following result of independent interest:

LEMMA 7. *Let X and U be as above, and let $f : K \to U$ be continuous. Then there is some $\delta > 0$ such that if $g : K \to U$ is another continuous function satisfying $\mathbf{d}(\, f(\mathbf{x}),\, g(\mathbf{x})\,) < \delta$ for all $\mathbf{x}$, then g is homotopic to f as mappings from X to U.*

Sketch of proof of Lemma 7. We can define a continuous function $h : K \to \mathbf{R}$ by $h(\mathbf{x}) = \mathbf{d}(\, f(\mathbf{x}), \mathbf{R}^m - U\,)$. In fact, this function is positive valued because f maps K into U, and by the compactness of K it takes a minimum value δ. Therefore, if $\mathbf{x}$ is an arbitrary point in K and $\mathbf{d}(\, f(\mathbf{x}),\, \mathbf{v}\,) < \delta$, then the closed line segment joining $f(\mathbf{x})$ to $\mathbf{v}$ lies entirely in U. Consequently, if g satisfies the condition in the lemma for this choice of δ, the image of the straight line homotopy from f to g lies entirely in U.■

NOTE AND EXAMPLE. The preceding lemma reflects one reason for including the codomain as an extra piece of data in our definition of a function. Given any two functions f and g as above, they are always homotopic as maps into $\mathbf{R}^m$ by a straight line homotopy. The crucial point in the lemma is that the image of the homotopy is contained in U. — Without the constraint involving a positive constant δ, the result is false. To see this, let $K = S^1$ and $U = \mathbf{R}^2 - \{\mathbf{0}\}$, and take f to be the usual inclusion. Then f is not homotopic to a constant map, for if $r : U \to K$ is the retraction described above, then $r \circ f$ is not homotopic to a constant, but if f were homotopic to a constant map k, then we would have

$$\mathrm{id}(S^1) \quad \simeq \quad r \circ f \quad \simeq \quad r \circ k \quad = \quad \text{constant}$$

and we know this is not the case.

The observations of the previous paragraph have the following positive implication: *If $H : S^1 \times [0,1] \to \mathbf{R}^2$ is a homotopy from the inclusion map to the constant map, then there is some $(\mathbf{x}_0,\, t_0\,) \in S^1 \times [0,1]$ such that $H(\mathbf{x}_0,\, t_0\,) = \mathbf{0}$.*■

A major objective of the course is to develop tools that will yield generalizations of the preceding observation to mappings from $S^n \times [0,1] \to \mathbf{R}^{n+1}$.

Sketch of proof of Theorem 6. Suppose that $f : K \to U$ as above is continuous, and let $\delta > 0$ be given as in Lemma 7. Denote the coordinate projections of f by f_i, where $1 \leq i \leq m$.

By the Stone-Weierstrass Approximation Theorem, there are polynomial functions p_i on $K \subset \mathbf{R}^n$ such that

$$|\,(p_i|K) - f_i\,| \quad < \quad \frac{\delta}{2\sqrt{n}}$$

for each i, and in fact we can also find polynomials g_i with rational coefficients such that

$$|\,(p_i|K) - (g_i|K)\,| \quad < \quad \frac{\delta}{2\sqrt{n}}\ .$$

If we let $g : \mathbf{R}^n \to \mathbf{R}^n$ be the function whose coordinates are given by the polynomials g_i, it follows that $|\,f - (g|K)\,| < \delta$.

Standard set-theoretic computations show that there are only countably many polynomials in n variables with rational coefficients, and it follows that there are only countably many choices for g.

Combining the preceding two paragraphs with Lemma 7, we conclude that f is homotopic to one of the countable family of continuous functions whose coordinates are given by polynomials in n variables with rational coefficients, and therefore the set $[K, U]$ is countable.■

Using the fact that the inclusion of S^1 in $\mathbf{R}^2 - \{\mathbf{0}\}$ is a homotopy equivalence, one can show that

$$\mathbf{Z} \quad \cong \quad [S^1, S^1] \quad \cong \quad [S^1, \mathbf{R}^2 - \{\mathbf{0}\}]$$

(see the exercises for this section) and therefore the cardinality bound of $\aleph_0$ on $[K, U]$ is the best possible general result.■

Important standard notation

Unless stated otherwise, in the remainder of these notes the symbol I will denot the closed unit interval $[0, 1]$.

II.2 : The fundamental group

(Hatcher, §§ 1.1 – 1.3, 1.A – 1.B; Munkres, §§ 52, 54)

This subject was treated in Mathematics 205B, and it might be useful to review this material before proceeding.

Section 1.B of Hatcher is devoted to proving a fundamental result in topology which has numerous uses in geometry and complex variables:

THEOREM 1. *Let G be an arbitrary group. Then there is an arcwise connected, locally arcwise connected, and locally simply connected Hausdorff space BG such that $\pi_1(BG, \mathrm{pt.})$ is isomorphic to G and the universal covering space of G is contractible. Furthermore, if X and Y are spaces which have these properties, then X is homotopy equivalent to Y.*

The existence argument is contained in Example 1.B.7 of Hatcher, while the uniqueness up to homotopy type is stated as Theorem 1.B.8 and established by the argument in Proposition 1.B.9.

Definition. A topological space X is (strongly) **aspherical** if it is arcwise connected and it has a contractible covering space.

As noted in Hatcher, the torus T^k is aspherical because its universal covering space is $\mathbf{R}^k$, and the covering space projection is given by $p(x_1, \cdots, x_k) = \big(\exp(2\pi\, i\, x_1), \cdots, \exp(2\pi\, i\, x_k)\big)$. Also, as noted in Hatcher, all compact connected surfaces except S^2 and $\mathbf{RP}^2$ are aspherical.

Generalization. (For students who have taken Mathematics 205C or are familiar with the notion of sectional curvature in a riemannian manifold.) There is an important generalization of all these facts due to J. Hadamard: *If M is a compact smooth n-manifold which has a riemannian metric whose sectional curvature is everywhere nonpositive, then the universal covering of M is diffeomorphic to $\mathbf{R}^n$.* — We shall not use this result at any future point in the course.

II.3 : Abstract cell complexes

(Hatcher, Ch. 0)

One possible way to view a polyhedron is to think of it as an object that is constructible in a finite number of steps as follows:

(0) Start with the finite set P_0 of vertices,

(n) If P_{n-1} is the partial polyhedron constructed at Step $n-1$, at Step n one adds finitely many simplices S_j, identifying each face of each simplex S_j with a simplex in P_{n-1}.

In fact, one can do this in order of increasing dimension, attaching all 1-simplices to the vertices at Step 1, then attaching 2-simplices along the boundary faces at Step 2, and so on. It is often useful in topology to consider objects that are generalizations of this procedure that are more flexible in certain key respects. The objects used these days in algebraic topology are known as **cell complexes**.

One immediate difference between cell complexes and simplicial complexes is that the former use the closed unit disk $D^n \subset \mathbf{R}^n$ and its boundary S^{n-1} in place of an n-simplex Δ and its boundary $\partial\Delta$, which is the union of its faces. In order to reconcile this replacement, one needs the following basic result:

THEOREM 1. *Let A be an n-simplex with boundary ∂A. Then there is a homeomorphism from A to D^n such that ∂A is mapped to S^{n-1} and the barycenter of A is mapped to $\mathbf{0} \in D^n$.*

Definition. The *standard n-simplex* Δ_n is the set of all points $(t_0, \cdots, t_n) \in \mathbf{R}^{n+1}$ such that $t_j \geq 0$ for all j and $\sum_j t_j = 1$. Note that the set of unit vectors $\{ \mathbf{e}_0, \cdots, \mathbf{e}_n \}$ is affinely independent because the set $\{ \mathbf{e}_1 - \mathbf{e}_0, \cdots, \mathbf{e}_{n+1} - \mathbf{e}_0 \}$ is linearly independent.

Proof. The first two steps are simple adjustments. If A is an arbitrary n-simplex, we claim there is a homeomorphism h from Δ_n to A which sends $\partial\Delta_n$ to ∂A and sends the barycenter of Δ_n to the barycenter of A. Specifically, if the vertices of A are given by $\mathbf{v}_0, \cdots, \mathbf{v}_n$, then the homeomorphism h sends $(t_0, \cdots, t_n) \in \Delta_n$ to $\sum_j t_j\mathbf{e}_j \in A$. Because of this, it will suffice to prove the theorem when $A = \Delta_n$; if $g : \Delta_n \to D^n$ is the homeomorphism with the desired properties and A is an arbitrary n-simplex, then $g \circ h^{-1}$ is a homeomorphism A to D^n with the desired properties.

The hyperplane $H \subset \mathbf{R}^{n+1}$ defined by the linear equation $\sum_j t_j = 1$ is isometric to $\mathbf{R}^n$; to see this, let $\mathbf{b}$ be the barycenter of Δ_n, and let $S : \mathbf{R}^{n+1} \to \mathbf{R}^{n+1}$ be the isometry $S(\mathbf{y}) = \mathbf{y} - \mathbf{b}$, so that S maps H to the vector subspace W defined by the linear homogeneous equation $\sum_j t_j = 0$. Standard results on finite-dimensional inner product spaces imply that W is isometric to $\mathbf{R}^n$, and thus H is also isometric to $\mathbf{R}^n$. Since an arbitrary closed disk D in $\mathbf{R}^n$ is homeomorphic to the standard unit disk by a map which sends centers to centers and boundary spheres to boundary spheres, it will suffice to prove there is a homeomorphism from some closed disk $B \subset H$ with center $\mathbf{b}$ to Δ_n such that $\mathbf{b}$ is sent to itself and the boundary of B maps to the boundary of Δ_n.

The next step is to find a suitable disk B satisfying the conditions of the previous paragraph. As before, let $\mathbf{b}$ denote the barycenter of Δ_n. Then all the coordinates of $\mathbf{b}$ are positive, and thus there is some $\delta > 0$ such that if $\mathbf{x} \in H$ and $|\mathbf{x} - \mathbf{b}| \leq \delta$, then all the barycentric coordinates of $\mathbf{x}$ are also positive, so that the closed disk $B \subset H$ with center $\mathbf{b}$ and radius δ is contained in Δ_n (in fact, it is contained in $\Delta_n - \partial\Delta_n$).

We shall now construct a *radial projection homeomorphism* ρ from B to Δ_n such that $rho(\mathbf{b}) = \mathbf{b}$ and ρ sends the boundary sphere of B to $\partial\Delta_n$. Let ∂B denote the boundary sphere of B, let $\mathbf{x} \in \partial B$, and let Γ denote the ray consisting of all points of the form $t\,\mathbf{x} + (1-t)\mathbf{b}$ where $t \geq 0$. Then $Gamma$ is a convex set, and therefore the set $K = \Gamma \cap \Delta_n$ is also convex; furthermore, it is closed since Γ and Δ_n are closed, and it is bounded because Δ_n is also bounded. Furthermore, it contains all points of the form $t\,\mathbf{x} + (1-t)\mathbf{b}$ where $0 \leq t \leq 1$. Therefore it follows that there is some $u(\mathbf{x}) > 0$ such that K consists of all points of the form $t\,\mathbf{x} + (1-t)\mathbf{b}$ where $0 \leq t \leq u(\mathbf{x})$. The radial projection mapping ρ will be defined by

$$\rho(t\,\mathbf{x} + (1-t)\mathbf{b}) \quad = \quad ut\,\mathbf{x} + (1-ut)\mathbf{b})\ .$$

We need to prove that this mapping is well-defined and determines a homeomorphism from B to Δ_n with all the required properties. However, before beginning the proof, we need to verify the following CLAIM : *If $\mathbf{x} \in \partial B$, then $\rho(\mathbf{x}) \in \partial\Delta_n$.*

To prove the claim, suppose that the conclusion is false; then by the definition of $\partial\Delta_n$ we know that all the barycentric coordinates of $\rho(x) = u\,\mathbf{x} + (1-u)\mathbf{b})$ are positive. Since the barycentric coordinates are just the usual linear coordinates if we view $\Delta_n \subset H \subset \mathbf{R}^{n+1}$, it follows that there is some $\varepsilon > 0$ such that $0 < v - u < \varepsilon$ implies that the barycentric coordinates of $v\,\mathbf{x} + (1-v)\mathbf{b})$ are positive. This means that the latter point must also lie in Δ_n. However, since $v > u$ we know that this point does not lie in Δ_n, and this contradiction implies that $\rho(\mathbf{x})$ must lie on $\partial\Delta_n$.

To see that ρ is well-defined, note that every point in $H - \{\mathbf{b}\}$ can be written as $c\,\mathbf{x}$ for unique choices of $\mathbf{x} \in \partial B$ and $c > 0$. This suffices to show that ρ is well-defined at all points except possibly $\mathbf{b}$. However, if we write $\mathbf{b} = c\,\mathbf{x}$, where $c \geq 0$ and $\mathbf{x} \in \partial B$, then the formula yields $\rho(0\,\mathbf{x}) = \mathbf{b}$ for all choices of $\mathbf{x}$. Therefore the function ρ is well-defined.

To see that ρ is 1–1, suppose that

$$\rho(t\,\mathbf{x} + (1-t)\mathbf{b}) \quad = \quad \rho(s\,\mathbf{y} + (1-s)\mathbf{b})$$

where $\mathbf{x}$ and $\mathbf{y}$ lie on ∂B and $s, t \geq 0$. The preceding equation implies that

$$t\,u(\mathbf{x})\,(\mathbf{x} - \mathbf{b}) \quad = \quad s\,u(\mathbf{y})\,(\mathbf{y} - \mathbf{b})$$

and since the vectors on both sides of the equation are nonzero and the function u takes only positive values, it follows that $s = 0$ if and only if $t = 0$, in which case we have

$$t\,\mathbf{x} + (1-t)\mathbf{b} \quad = \quad s\,\mathbf{y} + (1-s)\mathbf{b} \quad = \quad \mathbf{b}\ .$$

Therefore we have shown that if two distinct points

$$t\,\mathbf{x} + (1-t)\mathbf{b}\ , \qquad s\,\mathbf{y} + (1-s)\mathbf{b}$$

have the same image under ρ then neither is equal to $\mathbf{b}$, which is equivalent to saying that both s and t are nonzero.

By construction, both vectors $\mathbf{x}$ and $\mathbf{y}$ lie on ∂B, and the previously derived equation

$$t\,u(\mathbf{x})\,(\mathbf{x} - \mathbf{b}) \quad = \quad s\,u(\mathbf{y})\,(\mathbf{y} - \mathbf{b})$$

implies that $\mathbf{x} - \mathbf{b}$ and $\mathbf{y} - \mathbf{b}$ are positive multiples of each other. But the condition $\mathbf{x},\ \mathbf{y} \in \partial B$ implies that $|\mathbf{x} - \mathbf{b}| = |\mathbf{y} - \mathbf{b}|$ and thus it follows that $\mathbf{x} - \mathbf{b} = \mathbf{y} - \mathbf{b}$, so that we have $\mathbf{x} = \mathbf{y}$. Therefore, if neither $\mathbf{x}$ nor $\mathbf{y}$ is equal to $\mathbf{b}$ but

$$\rho(t\,\mathbf{x} + (1-t)\mathbf{b}) \quad = \quad \rho(s\,\mathbf{y} + (1-s)\mathbf{b})$$

then we must have

$$t\,u(\mathbf{x})\,(\mathbf{x} - \mathbf{b}) \quad = \quad s\,u(\mathbf{x})\,(\mathbf{x} - \mathbf{b})$$

and the latter immediately implies $s = t$. This completes the verification that ρ is 1–1.

We shall next prove that ρ is onto. Since ρ sends $\mathbf{b}$ to itself, it will suffice to show that every point $\mathbf{y}$ in $\Delta_n - \{\mathbf{b}\}$ lies in the image of ρ. The crucial observation is that $\mathbf{y}$ can be written in the form $(1-s)\mathbf{z} + s\,\mathbf{b}$, where $0 \leq s < 1$ and $\mathbf{z}$ lies on the boundary of Δ_n.

If $\mathbf{y}$ lies on the boundary, the statement in the preceding sentence is true for trivial reasons, so we shall now consider the case where $\mathbf{y}$ is not a boundary point. This means that all the barycentric coordinates of $\mathbf{y}$ are positive numbers. Let t^* be the minimum barycentric coordinate of $\mathbf{y}$; since the minimum of finitely many real valued continuous functions is continuous, it follows that t^* is a positive valued continuous function of $\mathbf{y}$. Furthermore, since we are assuming that $\mathbf{y} \neq \mathbf{b}$ we know that at least one barycentric coordinate is strictly less than $1/(n+1)$ and hence the same is true for t^*. It follows immediately that if $\mathbf{y} \in \Delta_n - \{\mathbf{b}\}$ does not lie on the boundary then we may rewrite the barycentric coordinate expansion $\mathbf{y} = \sum_j \ t_j \, \mathbf{e}_j$ as

$$(n+1)t^*\mathbf{b} \ + \ \sum_j \, (t_j - t^*)\mathbf{e}_j \quad = \quad (n+1)t^*\mathbf{b} \ + \ (1 - (n+1)t^*) \, \sum_j \, \frac{t_j - t^*}{1 - (n+1)t^*} \, \mathbf{e}_j \ .$$

Since $t^* = t_i$ for some choice of i, it follows that the i^{th} term in the summation

$$\mathbf{z} \quad = \quad \sum_j \, \frac{t_j - t^*}{1 - (n+1)t^*} \, \mathbf{e}_j$$

is zero; in fact, the coefficients of the vectors $\mathbf{e}_j$ are all nonnegative (the numerator of each fraction is nonnegative and the denominator is positive), and direct calculation shows that they add up to 1, so that $\mathbf{z}$ represents a point of $\partial\Delta_n$. Therefore we have $(1-s)\mathbf{z} + s\,\mathbf{b}$, where $s = (n+1)t^*$, so that $0 < s < 1$ (recall that $\mathbf{y}$ is not equal to $\mathbf{b}$ and does not lie on the boundary).

It now follows that $\mathbf{y}$ is the image under ρ of the point

$$\mathbf{b} \ + \ \frac{\delta\,|\mathbf{y} - \mathbf{b}|}{|\mathbf{z} - \mathbf{b}|} \cdot \big(\mathbf{z} - \mathbf{b}\big)$$

and therefore we hae shown that ρ maps B onto Δ_n. Closer inspection of the formulas above shows that ρ maps points of ∂B to $\partial\Delta_n$ and points of $B - \partial B$ to $\Delta_n - \partial\Delta_n$.

Finally, we must show that ρ is continuous; once we know that ρ is a 1–1 onto continuous mapping from one compact Hausdorff space to another, general considerations will imply that ρ must be a homeomorphism. We shall first show that ρ is continuous on $B - \{\mathbf{b}\}$ and then we shall verify continuity at $\mathbf{b}$ by a separate argument. Suppose that $\mathbf{x} \in B - \{\mathbf{b}\}$, write $\mathbf{x}$ as a convex combination $\sum_j \ v_j \, \mathbf{e}_j$, and use previous arguments to show that the ray starting at $\mathbf{b}$ and passing through $\mathbf{x}$ meets $\partial\Delta_n$ at the point

$$\mathbf{z}(\mathbf{x}) \quad = \quad \sum_j \, \frac{s_j - s^*}{1 - (n+1)s^*} \, \mathbf{e}_j$$

(where s^* is the least barycentric coordinate) and hence $\mathbf{z}$ is a continuous function of $\mathbf{x}$. Since

$$\rho(\mathbf{x}) \quad = \quad \mathbf{b} \ + \ \frac{|\mathbf{z} - \mathbf{b}|}{\delta} \cdot \big(\mathbf{x} - \mathbf{b}\big)$$

the continuity of ρ away from $\mathbf{b}$ follows from the continuity of $\mathbf{z}$.

To show the continuity at zero, first observe that $\mathbf{x} \in \Delta_n$ implies $|\mathbf{x} - \mathbf{b}| < 1$; of course, this also means that $\delta < 1$. If $0 < \theta \leq \delta$ and $B(\theta)$ denotes the closed disk of radius θ in H with center $\mathbf{b}$, then it follows from the definitions that ρ maps $B(\theta)$ to the simplex with vertices $\mathbf{b} + \theta\delta^{-1}(\mathbf{e}_j - \mathbf{b})$; note that these points are affinely independent because the difference vectors

$$\big(\mathbf{b} + \theta\delta^{-1}(\mathbf{e}_j - \mathbf{b})\,\big) \ - \ \big(\mathbf{b} + \theta\delta^{-1}(\mathbf{e}_0 - \mathbf{b})\,\big) \quad = \quad \theta\delta^{-1}(\mathbf{e}_j - \mathbf{e}_0)$$

(where $1 \leq j \leq n$) are linearly independent.

Suppose now that $\varepsilon > 0$; without loss of generality, we may assume $\varepsilon < \delta$. As the preceding paragraph, we know that the simplex $\Delta[\varepsilon]$ with vertices $\mathbf{b} + \varepsilon(\mathbf{e}_j - \mathbf{b})$ is contained in $B(\varepsilon)$.

Furthermore, the observations of the preceding paragraph also show that ρ maps $B(\varepsilon \cdot \delta)$ to $\Delta[\varepsilon]$. Combining these observations, we see that $|\mathbf{x} - \mathbf{b}| < \delta \cdot \varepsilon$ implies $|\rho(\mathbf{x}) - \rho(\mathbf{b})| < \varepsilon$, and therefore it follows that ρ is continuous at $\mathbf{b}$ as required.■

Adjoining cells to a space

We shall now give the basic step in the construction of cell complexes. The discussion below relies heavily on the material in Unit V of the online Mathematics 205A notes that were previously cited.

Definition. Let X be a compact Hausdorff space and let A be a closed subset of X. If k is a nonnegative integer, we shall say that *the space X is obtained from A by adjoining finitely many k-cells* if there are continuous mappings $f_i : S^{k-1} \to A$ for $i = 1, \cdot, N$ such that X is homeomorphic to the quotient space of the topological disjoint union

$$A \amalg \{1, \cdots, N\} \times D^k$$

modulo the equivalence relation generated by identifying $(j, \mathbf{x}) \in \{j\} \times S^{k-1}$ with $f_j(\mathbf{x}) \in A$, where the homeomorphism maps $A \subset X$ to the image of A in the quotient by the canonical mapping.

By construction, there is a 1–1 correspondence of sets between X and

$$A \amalg \{1, \cdots, N\} \times \mathbf{open}(D^k)$$

where $\mathbf{open}(D^k) \subset D^k$ is the complement of the boundary sphere. The set $E_j \subset X$ corresponding to the image of $\{j\} \times D^k$ in the quotient is called a *(closed) k-cell*, and the subset $E_j^{\mathbf{O}}$ corresponding to the image of $\{j\} \times \mathbf{open}(D^k)$ in the quotient is called an *open k-cell.* One can then restate the observation in the first sentence of the paragraph to say that X is a union of A and the open k-cells, and these subsets are pairwise disjoint.

Before discussing some topological properties of a space obtained by adjoining k-cells, we shall consider some EXAMPLES.

1. Let $(P, \mathbf{K})$ be a simplicial complex,let P_k be the union of all k-simplices in $\mathbf{K}$, and let P_{k-1} be defined similarly. Then the whole point of stating and proving Theorem 1 was to justify an assertion that P_k is obtained from P_{k-1} by attaching k-cells, one for each k-simplex in $\mathbf{K}$. Specifically, for each k-simplex A the map f_A is given by the composite of the homeomorphism $S^{k-1} \to \partial A$ with the inclusion $\partial A \subset P_{k-1}$. The homeomorphism from the quotient of the disjoint union to P_k is given by starting with the composite

$$P_{k-1} \amalg \{1, \cdots, N\} \times D^k \quad \longrightarrow \quad P_{k-1} \amalg_A A \quad \longrightarrow \quad P_k$$

where $\amalg_A$ runs over all the k-simplices of $\mathbf{K}$, the first map is a disjoint union of homeomorphisms on the pieces where the maps of Theorem 1 are used to define the homeomorphisms $\{j\} \times D^k \cong A$, and the second map is inclusion on each disjoint summand. This composite passes to a map of the quotient of the space on the left modulo the equivalence relation described above, and it is straightforward to show this map is 1–1 onto and hence a homeomorphism (all relevant spaces are compact Hausdorff).

2. (GRAPHS) One may define a finite (vertex-edge) graph to be a space obtained from a finite discrete space by adjoining 1-cells. Frequently there is an added condition that the attaching maps for the boundaries should be 1–1 (so that each 1-cell has two endpoints), and the weaker notion introduced here (and in Hatcher) is then called a *pseudograph*. The graph corresponds to a simplicial decomposition of a simplicial complex if and only if different 1-cells have different endpoints. The simplest example of a graph structure that is not a pseudograph and does not come from a simplicial complex is given by taking $X = S^1$ and $A = S^0$ with two 1-cells corresponding to the upper and lower semicircles $E^1_\pm$ in the complex plane. The attaching maps are defined to map the endpoints of $D^1 = [-1, 1]$ bijectively to $-1, 1$. — Another example that is historically noteworthy is the Königsberg Bridge Graph, in which the vertices correspond to four land masses in the city of Königsberg (now Kaliningrad, Russia) and the 1-cells (or *edges*) correspond to the bridges which joined pairs of land masses in the 18^{th} century (see the `figures` file for a drawing). This is another example of a graph that does not come from a simplicial complex but is not a pseudograph; if there are two bridges joining the same pairs of land masses, then the graph has two deges with the same boundary points. — In the next unit we shall see how Euler's analysis of this graph may be stated in terms of algebraic topology.

We shall encounter further examples after we define the main concept of this section. For the time being, we mention a few simple properties of spaces obtained by attaching k-cells for some k

PROPOSITION 2. *If X is obtained from A by attaching 0-cells, then X is homeomorphic to the disjoint union of A with a finite discrete space.*

This is true because the 0-disk D^0 has an empty unit sphere, so there are no attaching maps and the equivalence relation on the $A \amalg \{1, \cdot, N\}$ is the equality relation.■

PROPOSITION 3. *If X is obtained from A by attaching k-cells, then each open cell $E_j^{\mathbf{O}}$ is an open subset of X, and each such open cell is homeomorphic to* $\mathbf{open}(D^k)$.

Proof. Each closed cell is compact because it is a continuous image of D^k, and hence each such subset is closed in X. By the set-theoretic description given above, the open cell $E_j^{\mathbf{O}}$ is just the complement of the closed set

$$A \cup \bigcup_{i \neq j} E_i$$

and hence it is open in X. Since the quotient space map from the disjoint union to X defines a 1–1 onto continuous mapping from $\mathbf{open}(D^k)$ to $E_j^{\mathbf{O}}$, it suffices to show that an open subset of $\mathbf{open}(D^k)$ is sent to an open subset of $E_j^{\mathbf{O}}$. Let

$$\varphi : A \ \amalg \ \{1, \ \cdots \ , N\} \times D^k \ \longrightarrow \ X$$

be the continuous onto quotient map corresponding to the cell attachments, and suppose that U is open in $\{j\} \times \mathbf{open}(D^k)$. By construction we then have

$$U \ = \ \varphi^{-1}\big[\,\varphi[U]\,\big]$$

and thus $\varphi[U]$ is open in X by the definition of the quotient topology.■

The last result in this subsection implies that the inclusion of A in X is homotopically well-behaved if X is obtained from A by adjoining k-cells.

PROPOSITION 4. *If X is obtained from A by attaching k-cells and U is an open subset of X containing A, then there is an open subset V such that*

$$A \ \subset \ V \subset \ \overline{V} \ \subset U$$

and A is a strong deformation retract of both V and $\overline{V}$.

Proof. As in the preceding argument, take

$$\varphi : A \amalg \{1, \cdots, N\} \times D^k \longrightarrow X$$

to be the continuous onto map corresponding to the k-cell attachments.

Let $F = X - U$, and let $F_0 = \varphi^{-1}[F]$, so that F_0 corresponds to a disjoint union $\amalg_j F_j$, where each F_j is a compact subset of $\mathbf{open}(D^k)$; compactness follows because the image of each F_j in X is a closed subset of the compact k-cell E_j. Therefore we can find constants c_j such that $0 < c_j < 1$ and F_j is contained in the open disk of radius c_j about the origin in $\{j\} \times D^k$; let c be the maximum of the numbers c_j, and let $V \subset X$ be the image under φ of the set

$$W \quad = \quad A \amalg \bigcup_j \{j\} \times \{ \mathbf{x} \in D^k \mid c < |\mathbf{x}| \leq 1 \} .$$

Then V is open because it is the complement of a compact set, and it follows that $\overline{V}$ is the image of

$$Y \quad = \quad A \amalg \bigcup_j \{j\} \times \{ \mathbf{x} \in D^k \mid c \leq |\mathbf{x}| \leq 1 \} .$$

Each of the sets W and Y is a strong deformation retract of

$$B \quad = \quad A \amalg \bigcup_j \{j\} \times S^{k-1} .$$

Specifically, the homotopies deforming W and Y into B are the identity on A and map each of the sets $\{ c < |\mathbf{x}| \leq 1 \}$, $\{ c \leq |\mathbf{x}| \leq 1 \}$ to S^{k-1} by sending a (necessarily nonzero) vector $\mathbf{y}$ to $|\mathbf{y}|^{-1}\mathbf{y}$ and taking a staight line homotopy to join these two points. A direct check of the equivalence relation defining φ shows that the associated maps and homotopies $W \to B \to W$ and $Y \to B \to Y$ pass to the quotients $V \to A \to V$ and $\overline{V} \to A \to \overline{V}$, and these quotient maps display A as a strong deformation retract of both V and $\overline{V}$.■

Cell complex structures

By the preceding discussion, a simplicial complex $(P, \mathbf{K})$ has a finite, linearly ordered chain of closed subspaces

$$\emptyset \quad = \quad P_{-1} \quad \subset \quad P_0 \quad \subset \quad \cdots \quad \subset \quad P_m \quad = \quad P$$

such that for each k satisfying $0 \leq k \leq m$, the subspace P_k is obtained from P_{k-1} by attaching finitely many k-cells. We shall generalize this property into a definition for arbitrary cell complex structures.

Definition. Let X be a topological space. A *finite cell complex structure* (or finite CW structure) on X is a chain $\mathcal{E}$ of closed subspaces

$$\emptyset \quad = \quad X_{-1} \quad \subset \quad X_0 \quad \subset \quad \cdots \quad \subset \quad X_m \quad = \quad X$$

such that for each k satisfying $0 \leq k \leq m$, the subspace X_k is obtained from X_{k-1} by attaching finitely many k-cells. The subspace X_k is called the *k-skeleton* of X, or more correctly the k-skeleton of $(X, \mathcal{E})$

At this level of abstraction, the notion of cell complex structure is due to J. H. C. Whitehead; his definition extended to infinite cell complex structures and the letters CW were described as abbreviations for two properties of the infinite complexes that are explained in the Appendix of Hatcher's book, but one should also note the coincidence(?) that the letters also represent Whitehead's last two initials.

It follows immediately that simplicial complexes are examples of cell complexes. Numerous further examples appear on pages 5–8 of Hatcher. Furthermore, the Δ-complexes discussed on pages 102–104 are also examples of cell complexes. In analogy with (edge-vertex) graphs, the main difference between Δ-complexes and simplicial complexes is that two k-simplices in a Δ-complex may have the same faces, but two k-simplices in a simplicial complex have at most a single $(k-1)$-face in common.

Because of the following result, one often describes a cell complex structure as a *cellular decomposition* of X.

PROPOSITION 5. *If X is a space and $\mathcal{E}$ is a cell decomposition of X, then every point of X lies on exactly one open cell of X.*

Proof. Since $X = \cup_k \ (X_k - X_{k-1})$, it follows that every point $y \in X$ lies in a exactly subset of the form $X_k - X_{k-1}$. Therefore there is at most one value of k such that x can lie on an open k-cell. Furthermore, since $X_k - X_{k-1}$ is a union of the opne k-cells and the latter are pairwise disjoint, it follows that x lies on exactly one of these open k-cells.■

NOTE. If a cell complex has an n-cell for some $n > 0$ and $0 < m < n$, the cell complex might not have any m-cells (in contrast to the situation for, say, simplicial complexes); see Example 0.3 on page 6 of Hatcher.

Finally, we shall give a slightly different definition of subcomplex than the one in Hatcher.

Definition. If $(X, \mathcal{E})$ is a cell complex, we say that a closed subspace $A \subset X$ determines a *cell subcomplex* if for each $k \geq 0$ the set $A_k = X_k \cap A$ is obtained from A_{k-1} by attaching k-cells such that the every k-cell for A is also a k-cell for X.

There is an simple relationship between this notion of cell subcomplex and the previous definition of subcomplex for a simplicial complex; the proof is straightforward.

PROPOSITION 6. *If $(P, \mathbf{K})$ is a simplicial complex and $(P_1, \mathbf{K}_1)$ is a simplicial subcomplex, then P_1 also determines a cell subcomplex.*■

Finally, here are two further observations regarding subcomplexes. Again, the proofs are straightforward.

PROPOSITION 7. *If X is a cell complex such that $A \subset X$ determines a subcomplex of X and $B \subset A$ determines a subcomplex of A, then B also determines a subcomplex of A. Likewise, if B determines a subcomplex of X then B determines a subcomplex of A.*■

PROPOSITION 8. *If X is a cell complex such that $A \subset X$ determines a subcomplex of X, then for each $k \geq 0$ the set $X_k \cup A$ determines a subcomplex of X.*■

II.4 : The Homotopy Extension Property

(Hatcher, Ch. 0, § 2.1)

In this section we shall bring together several concepts from the preceding sections. The basis is the following central Extension Question stated at the beginning of this unit, and our first result describes a condition under which this question always has an affirmative answer.

PROPSITION 1. *Suppose that X and Y are topological spaces, that $A \subset X$ is a retract, and that $g : A \to Y$ is continuous. Then there is an extension of g to a continuous mapping $f : X \to Y$.*

Proof. Let $r : X \to A$ be a continuous function such that $r|A$ is the identity, and define $f = g \circ r$. Then if $a \in A$ we have $f(a) = g \circ r(a) = g(\,r(a)\,)$, and the latter is equal to $g(a)$ since $r|A$ is the identity.■

The hypothesis of the proposition is fairly rigid, but the result itself is a key step in proving a general result on the Extension Question.

THEOREM 2. (HOMOTOPY EXTENSION PROPERTY) *Let $(X, \mathcal{E})$ be a cell complex, and suppose that A determines a subcomplex. Suppose that Y is a topological space, that $g : A \to Y$ is a continuous map, and $f : X \to Y$ is a continuous map such that $f|A$ is homotopic to g. Then there is a continuous map $G : X \to Y$ such that $G|A = g$.*

COROLLARY 3. *Suppose that X and A are as above and that $g : A \to Y$ is homotopic to a constant map. Then g extends to a continuous function from X to Y.*

COROLLARY 4. *Suppose that X and A are as above and that $g : A \to X$ is homotopic to the inclusion map. Then g extends to a continuous function from X to itself.*

Corollary 3 follows because every constant map from A to Y extends to the analogous constant map from X to Y, and Corollary 4 follows because the inclusion of A in X extends continuously to the identity map from X to itself.■

One important step in the proof of the Homotopy Extension Property relies upon the following result:

PROPOSITION 5. *For all $k > 0$ the set $D^k \times \{0\} \;\cup\; S^{k-1} \times [0,1]$ is a strong deformation retract of $D^k \times [0,1]$.*

Proof. This argument is outlined in Proposition 0.16 on page 15 of Hatcher, and there is a drawing to illustrate the proof in the `figures` document.

The retraction $r : D^k \times [0,1] \to D^k \times \{0\} \;\cup\; S^{k-1} \times [0,1]$ is defined by a radial projection with center $(\mathbf{0}, 2) \in D^k \times \mathbf{R}$. As indicated by the drawing, the formula for r depends upon whether $2|\mathbf{x}| + t \geq 2$ or $2|\mathbf{x}| + t \leq 2$. Specifically, if $2|\mathbf{x}| + t \geq 2$ then

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

while if $2|\mathbf{x}| + t \leq 2$ then we have

$$r(\mathbf{x},\, t) \quad = \quad \frac{1}{2}\big(\,(2 - t)\mathbf{x},\, 0\,\big)$$

and these are consistent when $2|\mathbf{x}|+t=2$ then both formulas yield the value $|\mathbf{x}|^{-1}(\mathbf{x},\,0)$. Elementary but slightly tedious calculation then implies that $r(\mathbf{x},\,t)$ always lies in $D^k\times[0,1]$, and likewise that r is the identity on $D^k\times\{0\}\ \cup\ S^{k-1}\times[0,1]$. The homotopy from inclusion$\circ r$ to the identity is then the straight line homotopy

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

and this completes the proof of the proposition.■

Proof of Theorem 2. In the course of the proof we shall need the following basic fact: *If A and B are compact Hausdorff spaces and $\varphi:A\to B$ is a quotient map in the sense of Munkres' book, then for each compact Hausdorff space C the product map $\varphi\times 1_C:A\times C\to B\times C$ is also a quotient map.* This follows because $\varphi\times 1_C$ is closed, continuous and surjective.

Since the homotopy relation on continuous functions is transitive, a standard recursive argument reduces the proof of the theorem to the special cases subcomplex inclusions

$$X_{k-1}\ \cup\ A\quad\subset\quad X_k\ \cup\ A\ .$$

In other words, *it will suffice to prove the theorem when X is obtained from A by attaching k-cells.*

We now assume the condition in the preceding sentence. Let $h:A\times[0,1]\to Y$ be the homotopy from f (when $t=0$) to g (when $t=1$). If we can show that the inclusion

$$A\times[0,1]\ \cup\ X\times\{0\}\quad\subset\quad X\times[0,1]$$

is a retract, then we can use Proposition 1 to find an extension of the map

$$\theta\quad=\quad\text{“}h\cup f\text{''}:A\times[0,1]\ \cup X\times\{0\}\quad\longrightarrow\quad Y$$

to $X\times[0,1]$, and the restriction of this extension to $X\times\{1\}$ will be a continuous extension of g. — In fact, we shall show that *the space $A\times[0,1]\cup X\times\{0\}$ is a strong deformation retract of $X\times[0,1]$.*

As in earlier discussions let

$$\varphi:A\ \amalg\ \{1,\ \cdots\ ,N\}\times(D^k)\ \longrightarrow\ X$$

be the topological quotient map which exists by the definition of attaching k-cells. By Proposition 5 we know that the space

$$A\times[0,1]\ \amalg\ \{1,\ \cdots\ ,N\}\times\left(S^{k-1}\times[0,1]\cup D^k\times\{0\}\right)$$

is a strong deformation retract of

$$\left(\ A\ \amalg\ \{1,\ \cdots\ ,N\}\times D^k\ \right)\times[0,1]$$

because we can the mappings piecewise using the identity on $A\times[0,1]$ and the functions from Proposition 5 on each of the pieces $\{j\}\times D^k\times[0,1]$. Let

$$r':\left(\ A\ \amalg\ \{1,\ \cdots\ ,N\}\times D^k\ \right)\times[0,1]\ \longrightarrow\ A\times[0,1]\ \amalg\ \{1,\ \cdots\ ,N\}\times\left(S^{k-1}\times[0,1]\cup D^k\times\{0\}\right)$$

be the retraction obtained in this fashion, and let

$$H' : \Big(\big(A \amalg \{1, \cdots, N\} \times D^k \big) \times [0,1] \Big) \times [0,1] \longrightarrow \big(A \amalg \{1, \cdots, N\} \times D^k \big) \times [0,1]$$

be defined similarly. It will suffice to show that these pass to continuous mappings of quotient spaces; in other words, we want to show there are (continuous) mappings r and H such that the following diagrams are commutative, in which ψ is the mapping whose values are given by φ:

$$\begin{array}{ccc} (A \amalg \cdots) \times [0,1] & \xrightarrow{r'} & A \times [0,1] \amalg \big(\{1,...,N\} \times [\cdots]\big) \\ \big\downarrow \varphi \times 1 & & \big\downarrow \psi \\ X \times [0,1] & \xrightarrow{r} & A \times [0,1] \cup X \times \{0\} \end{array}$$

$$\begin{array}{ccc} \big((A \amalg \cdots) \times [0,1]\big) \times [0,1] & \xrightarrow{H'} & (A \amalg \cdots) \times [0,1] \\ \big\downarrow \varphi \times 1 \times 1 & & \big\downarrow \phi \times 1 \\ \big(X \times [0,1]\big) \times [0,1] & \xrightarrow{H} & X \times [0,1] \end{array}$$

Standard results on factoring maps through quotient spaces imply that such commutative diagrams exist if and only if (i) if two points map to the same point under $\psi \circ r'$, then they map to the same point under $\varphi \times 1$, (ii) if two points map to the same point under $\phi \times 1 \circ H'$, then they map to the same point under $\varphi \times 1 \times 1$. It is a routine exercise to check both of these statements are true.■

COROLLARY 6. *Suppose that X and Y are as in the theorem and Y is contractible. Then every continuous mapping $f : X \to Y$ has a continuous extension to X.*

Proof. It will suffice to prove that an arbitrary continuous mapping $f : A \to Y$ is homotopic to a constant. We know that 1_Y is homotopic to a constant map k, and therefore $f = 1_Y \circ f$ is homotopic to the constant map $k \circ f$.■

III. Simplicial homology

The goal of this unit is to define a sequence of abelian groups associated to a simplicial complex $(P, \mathbf{K})$ which are called **homology groups** and denoted by $H_n(P, \mathbf{K})$, where n runs through all the integers but the groups are all zero if n is negative. These groups may be interpreted as furnishing an "algebraic picture" of the underlying topological space P. In order to develop the important properties of these groups it will be necessary to introduce some basic concepts and results from *homological algebra*, but efforts will be made to keep this to a minimum.

We have stated that the groups provide information about the underlying space P rather than the simplicial complex $(P, \mathbf{K})$ because *these groups turn out to depend only upon P itself*. This fact will drop out of the more general constructions in the next unit, where homology groups are defined for an arbitrary topological space and shown to agree with the groups of this unit if the space P has a simplicial decomposition.

Some motivation from vector analysis

Suppose that U is an open subset of $\mathbf{R}^3$ and Σ is some sort of compact oriented surface in U (for our purposes, it suffices to think of Σ as having a continuously defined unit normal vector at every point). Then the boundary of Σ is some union of closed curves Γ_i, where the sense of Γ_i is chosen such that for each point of such a curve the ordered triple of vectors given by

> the chosen unit normal vector to the surface at the point,
>
> the unit tangent vector to the curve at the point,
>
> the unit vector which is tangent to the surface at the point, but perpendicular to the curve's tangent vector and directed **into** the surface

will form a right handed triad (see the illustration in the `figures` document); we shall not try to make everything rigorous here because the goal is to provide some intuition. In such a situation one sometimes says that the formal sum $\sum_i \Gamma_i$ of the sensed curves Γ_i is homologous to zero in U, and by Stokes' Theorem we have the following:

If $\sum_i \Gamma_i$ is homologous to zero in U and $\mathbf{F}$ is a smooth vector field defined on U such that $\nabla \times \mathbf{F} = \mathbf{0}$, then

$$\sum_i \int_{\Gamma_i} \mathbf{F} \cdot d\mathbf{x} \quad = \quad 0 \text{ .}\blacksquare$$

It is important to note that if V is an open subset of U and $\sum_i \Gamma_i$ is homologous to zero in U, then $\sum_i \Gamma_i$ *is not necessarily homologous to zero in V*. The standard example for this involves the ordinary unit circle Γ in $\mathbf{R}^2 \subset \mathbf{R}^3$ whose center is the origin and whose radius is 1. This curve is homologous to zero in $\mathbf{R}^3$ because it bounds the closed unit disk. To see it is not homologous to zero in $V = (\mathbf{R}^2 - \{\mathbf{0}\}$, consider the vector field given by

$$\mathbf{F}(u, v) \quad = \quad \left(\frac{v}{u^2 + v^2}, \frac{-u}{u^2 + v^2}, 0 \right)$$

and note that $\nabla \times \mathbf{F} = \mathbf{0}$ and the standard computation

$$\int_{\Gamma} \mathbf{F} \cdot d\mathbf{x} \quad = \quad 2\pi$$

imply that Γ cannot be homologous to zero in V.

Suppose now that we have a union of pairwise disjoint closed oriented surface Σ_j in our open set U; the term "closed" means that the surfaces have no boundary curves, just like the unit sphere defined by $u^2 + v^2 + w^2 = 1$. We shall say that the formal sum $\Sigma_1 + \cdots + \Sigma_j$ is homologous to zero in U if $\cup_j \Sigma_j$ bounds a region $W \subset U$ such that the closure of W is equal to the union of W and $\cup_j \Sigma_j$ and the normal directions to Σ are all outward pointing. — For example, the unit sphere is homologous to zero in $\mathbf{R}^3$ because it bounds a unit disk, and if Σ_r denotes the sphere of radius r in $\mathbf{R}^3$, then $\Sigma_1 \cup \Sigma_2$ is homologous to zero if we orient the pieces so that the normal vectors on Σ_2 point outward (away from the origin) and the normal vectors on Σ_1 point inward (towards the origin). The Divergence Theorem from vector analysis then has the following implication:

If $\Sigma_1 + \cdots + \Sigma_n$ is homologous to zero in U and $\mathbf{F}$ is a smooth vector field defined on U such that $\nabla \cdot \mathbf{F} = 0$, then

$$\sum_i \iint_{\Sigma_i} \mathbf{F} \cdot d\mathbf{\Sigma} \quad = \quad 0 \ . \blacksquare$$

We can now show that Σ_1 is not homologous to zero in $\mathbf{R}^3 - \{\mathbf{0}\}$ by an argument similar to the preceding one. Let $\mathbf{F}$ be the vector field on $\mathbf{R}^3 - \{\mathbf{0}\}$ defined by $\mathbf{F}(\mathbf{x}) = |\mathbf{x}|^{-1}\mathbf{x}|$; then it is a routine exercise to prove that $\nabla \cdot \mathbf{F} = 0$ but direct computation shows that

$$\iint_{\Sigma_1} \mathbf{F} \cdot d\mathbf{\Sigma} \quad = \quad 4\pi \ .$$

Homology theory provides an organized algebraic framework for studying such phenomena.

III.1 : Exact sequences and chain complexes

(Hatcher, § 2.1)

This section is basically algebraic, and at first the need for formally introducing the concepts may be unclear. However, the notions described here arise repeatedly in algebraic topology and other subjects.

Definition. Suppose we are given a diagram of the form

$$A \xrightarrow{\quad f \quad} B \xrightarrow{g} C$$

in which the objects are abelian groups (possibly with some additional structure) and the morphisms are abelian group homomorphisms (possibly preserving the extra structure). We shall say that the diagram is *exact* at B if the kernel of g is equal to the image of f.

More generally, if we are given a linear diagram such as

$$\cdots \longrightarrow Z \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow D \longrightarrow \cdots$$

we shall say that it is an *exact sequence* if it is exact at every object which is the domain of one morphism and the codomain of another.

Examples

There are many standard exact sequences in elementary algebra.

1. A *short exact sequence* is one having the form $0 \to A \to B \to C \to 0$. Exactness at A means that the kernel of $A \to B$ is the image of $0 \to A$, which is equivalent to saying that the map is injective. Similarly, exactness at C means that the kernel of $C \to 0$ is the image of $B \to C$, whic is equivalent to saying that the map is surjective. The short exact sequence property is then equivalent to saying that $A \to B$ is injective, and C is isomorphic to the quotient of B by the image of A.

2. The *cokernel* of a homomorphism $f : A \to B$ is defined to be the quotient group $B/f[A]$. Given an arbitrary homomorphism $f : A \to B$, one then has the following *kernel – cokernel exact sequence*:

$$0 \longrightarrow \mathrm{Ker}(f) \longrightarrow A \longrightarrow B \longrightarrow \mathrm{Coker}(f) \longrightarrow 0$$

3. Let U be a connected open subset of $\mathbf{R}^2$, let $\mathbf{C}^\infty(U)$ denote the infinitely differentiable real valued functions on U, and let let $\mathbf{V}(U)$ denote the infinitely differentiable (2-dimensional) vector fields on U in the sense of vector analysis. If we let $\mathbf{R} \to \mathbf{C}^\infty(U)$ denote the inclusion of the constant functions and take the gradient map from $\mathbf{C}^\infty(U)$ to $\mathbf{V}(U)$, then it follows that the sequence $\mathbf{R} \to \mathbf{C}^\infty(U) \to \mathbf{V}(U)$ is exact. Furthermore, if we take the map $\mathbf{V}(U) \to \mathbf{C}^\infty(U)$ which sends a vector field $\mathbf{F} = (P, Q)$ to its "scalar curl" $Q_1 - P_2$, then the sequence $\mathbf{C}^\infty(U) \to \mathbf{V}(U) \to \mathbf{C}^\infty(U)$ will be exact **provided** U is convex (or more generally star-shaped). — On the other hand, the second sequence is not exact if $U = \mathbf{R}^2 - \{\mathbf{0}\}$, for the previously described vector field on U with coordinate functions v/r and $-u/r$ has zero scalar curl but is not the gradient of any smooth function on U; this follows from Green's Theorem and the previous line integral calculation.

We can extend the preceding if U is a connected open set in $\mathbf{R}^3$ by considering the following sequence:

$$\mathbf{R} \xrightarrow{\text{constants}} \mathbf{C}^\infty(U) \xrightarrow{\text{grad}} \mathbf{V}(U) \xrightarrow{\text{curl}} \mathbf{V}(U) \xrightarrow{\text{div}} \mathbf{C}^\infty(U)$$

This is again exact at the left hand object $\mathbf{C}^\infty(U)$, and standard results in vector analysis imply that the kernel of the curl is contained in the image of the gradient, while the kernel of the divergence is contained in the image of the curl. If U is convex, then one can also show that the sequence is exact, but in general this is not true. Our previous examples give a vector field on $\mathbf{R}^2 - \{\mathbf{0}\}) \times \mathbf{R}$ whose curl is zero but cannot be expressed as a gradient over U, and a vector field on $\mathbf{R}^3 - \{\mathbf{0}\}$ whose divergence is zero but cannot be expressed as the curl of another vector field over U.

Graded objects

The next concept is simple but indispensable.

Definition. Let A be a set, and let $\mathbf{C}$ be a category. A *graded object* over $\mathbf{C}$ with grading set A is a function X from A to the objects of $\mathbf{C}$. The object corresponding to a is generally denoted by X_a

For example, one can construct a graded vector space over the reals with grading set the integers $\mathbf{Z}$ by taking $V_n = \mathbf{R}^n$ for $n \geq 0$ and setting V_n equal to the zero space if $n < 0$.

Another example is obtainable from an algebra of polynomials $\mathbf{R}[x_1, \cdots, x_n]$ in finitely many indeterminates. Here we can take V_n to be the set of all homogeneous polynomials of degree n together with the zero polynomial.

In this course we shall mainly be interested in *nonnegatively graded objects*, where the indexing set is $\mathbf{Z}$ and the object X_n is a suitable zero object if $n < 0$. For the categories of abelian groups or modules over some associative ring with unit, the meaning of zero object is obvious, and these categories are the only ones to be considered here.

Definition. If X and Y are nonnegatively graded objects over a category $\mathbf{C}$, then a *graded morphism of degree zero* or *grade preserving morphism* is a function f which assigns to each $n \in \mathbf{Z}$ a morphism $f_n : X_n \to Y_n$ in the category $\mathbf{C}$.

In the polynomial example, one can define a grade preserving homomorphism by sending the homogeneous polynomial $p(x_1, x_2, \cdots x_n)$ to the homogeneous polynomial $q(x_1, x_2, \cdots x_n) = p(x_1, x_1 + x_2, \cdots x_n)$. Obviously there are many other maps of this type.

The following observation is immediate:

PROPOSITION 1. *Given a category* $\mathbf{C}$*, the* $\mathbf{Z}$*-graded objects over* $\mathbf{C}$ *and graded morphisms of degree zero form a category.*■

In fact, this category has many structural properties that are direct analogs of properties that hold for $\mathbf{C}$ (for example, subobjects, quotient objects, direct products. and so on).

Chain complexes

The following concept is absolutely fundamental.

Definition. Let $\mathbf{C}$ be the category of abelian groups and homomorphisms or a category of unital modules over an associative ring with unit R. A **chain complex** over $\mathbf{C}$ is a pair (A, d) consisting of a graded object A over $\mathbf{C}$ indexed by the integers together with morphisms $d_j : A_j \to A_{j-1}$ such that $d_{j-1} \circ d_j = 0$ for all j.

Here are a few simple examples.

1. Given an arbitrary graded module A, one can make it into a chain complex by taking $d_j = 0$ for all j. More generally, given a sequence of homomorphisms $f_{2j} : A_{2j} \to A_{2j-1}$, one can define a chain complex whose graded module is A with $d_{2j} = f_{2j}$ and $d_{2j-1} = 0$.

2. Suppose we are given three modules B, H, and B'. The we can define a chain complex with $A_2 = B$, $A_1 = B \oplus H \oplus B'$, and $A_0 = B'$ and $A_j = 0$ otherwise such that d_2 is injection into the first summand, d_1 is projection onto the third summand, and all other maps d_j must be zero (since either their domain or codomain is zero).

3. If U is open in $\mathbf{R}^2$, then one can obtain a chain complex from the previous sequence involving $\mathbf{C}^\infty(U)$ and $\mathbf{V}(U)$, if one takes A_3 to be the reals, A_2 and A_0 to be the smooth functions, A_0 to be the vector fields, with morphisms given by inclusion of constants from A_3 to A_2, gradient from A_2 to A_1, scalar curl from A_1 to A_0, and with all other real vector spaces and morphisms equal to zero. Similarly, if U is open in $\mathbf{R}^3$ one has a system with A_4 equal to the reals, A_3 and A_0 equal to the smooth functions, A_2 and A_1 equal to the vector fields, with morphisms given by inclusion of constants from A_4 to A_3, gradient from A_3 to A_2, curl from A_2 to A_1, divergence from A_1 to A_0, and with all other real vector spaces and morphisms equal to zero.

The mapping d is often called a differential; the motivation is related to the preceding examples where the maps are given by some form of differentiation.

Definition. Given two chain complexes (A, d) and (B, e) a **chain map** $f : A \to B$ is a graded morphism such that for all integers j we have $e_j \circ f_j = f_{j-1} \circ d_j$. In other words, the following diagram is commutative:

$$\begin{array}{ccc} A_j & \xrightarrow{f_j} & B_j \\ \downarrow d_j & & \downarrow e_j \\ A_{j-1} & \xrightarrow{f_{j-1}} & B_{j-1} \end{array}$$

If the differential in a chain complex (A, d) is unambiguous from the context we shall frequently write A instead of (A, d).

The following consequences of the definitions are elementary but important.

PROPOSITION 2. *Given a category* $\mathbf{C}$*, the chain complexes over over* $\mathbf{C}$ *and chain complex morphisms form a category.*

PROPOSITION 3. *If* (A, d) *and* (B, e) *are chain complexes over* $\mathbf{C}$ *and* $f : (A, d) \to (B, e)$ *is a morphism of chain complex such that the mappings* f_j *are all isomorphisms, then the map* f^{-1} *of graded modules defined by* $(f^{-1})_j = f_j^{-1}$ *is also a chain map.*

TO BE CONTINUED