Algebraic Topology: A Computational Approach©

T. Kaczynski
Université de Sherbrooke

K. Mischaikow
Georgia Institute of Technology

M. Mrozek
Jagellonian University

April 18, 2000

Contents

Chapter 1

Introduction

The subject of Topology grew out of the foundations of calculus and more generally analysis. If you took a typical calculus sequence, then you began by learning about functions of the real line. The focus was on differentiable functions and how they can be best approximated locally by linear functions (the derivative). Along the way you learned about continuous functions. Again, the emphasis was on local properties such as limits; a notable exception is the intermediate value theorem. Later on these concepts were generalized to functions of more than one variable, i.e. functions from $\mathbf{R}^n$ to $\mathbf{R}^m$. Topology incorporates further generalizations. In particular, it allows one to study the local and global properties of continuous functions between general spaces.

To read this book you do not need to have studied general topology. This introductory chapter summarizes the elementary topology which we will need.

As was mentioned above one of the powers of the calculus is that through differentiation differentiable functions are locally approximated by linear functions. Linear functions are, of course, much easier to work with. Furthermore, linear functions can be studied algebraically as you learned in your linear algebra course. As an example of the advantage gained by this process of algebratization consider the following question. Is the function $f : \mathbf{R}^2 \to \mathbf{R}^2$ given by

$$f(x, y) = (x^2 - 3xy + y + 2, xy - 2y^2 - 4x - 1)$$

invertible near the point $(1, 2)$? Trying to find an explicit inverse is difficult. However, calculus gives us a simpler way to answer the question. Differentiating f gives

$$Df(x,y) = \begin{bmatrix} 2x-3y & -3x+1 \\ y-4 & x-4y \end{bmatrix}.$$

Evaluating this at $(1,2)$ we get

$$Df(1,2) = \begin{bmatrix} -4 & -2 \\ -2 & -7 \end{bmatrix}.$$

Since the determinant of this matrix does not equal zero, f is invertible in a neighborhood of $(1,2)$. Of course this is just a special case of the following theorem.

Theorem 1.1 [Inverse Function Theorem] *Let U be an open set in $\mathbf{R}^n$ and let $f : U \to \mathbf{R}^n$ be a differentiable function. Let $x \in U$. If $Df(x)$, the derivative of f at x, is an invertible matrix, then there is an open neighborhood $V \subset U$ containing x such that $f : V \to f(V)$ is invertible with a differentiable inverse.*

The important point of this example is that through calculus we have reduced an analytic problem to an algebraic problem. In fact, this method allows us to develop an algorithmic approach to answering this question. For example using the computer package MAPLE we can solve this problem as follows.

```
with(linalg):
f1 := (x,y) -> x^2 -3*x*y +y +2:
f2 := (x,y) -> x*y - 2*y^2 -4*x-1:
f  := (x,y) -> (f1(x,y),f2(x,y)):
Df := (x,y) -> array(1..2, 1..2,[[D[1](f1)(x,y),D[2](f1)(x,y)],
[D[1](f2)(x,y),D[2](f2)(x,y)]]):
'f(x,y)'=f(x,y);
'Df(x,y)'=Df(x,y);
'Df(1,2)'=Df(1,2);
'Det(Df(1,2))'=det(Df(1,2));
```

On a superficial level we might say that calculus, through the derivative, provides us with a way to transform the study of local properties of differentiable functions to problems in linear algebra. Furthermore, since for elementary functions many of the operations used in calculus can be implemented as algorithms and since linear algebra is also amenable to algorithmic implementation, many problems can be reduced to simple symbolic computations as described above. As will be shown in this book algebraic topology provides a means by which one can transform the study of the global properties of topological spaces and continuous functions to problems in algebra, or more precisely group theory (don't worry about what a group is at this moment - it will be introduced when the time comes). There are several different algebraic structures that can be assigned to topological spaces, the one we will study is called homology. Our focus will be on developing an algorithmic approach to homology theory which allows us to use the computer to solve topological problems.

1.1 Basic Notions from Topology

It was stated above that knowledge of general topology is not a prerequisite for this book. While this is correct, familiarity with the basic ideas of topology is worthwhile for at least two reasons. First, it is hoped that after finishing this book you will be motivated to continue your study of topology, and therefore, you may as well begin using the language of topology at this point. Second, as in the case of all important mathematics, the abstraction helps clarify the essential ideas.

1.1.1 Topological Spaces

The most fundamental definition is that of a topological space.

Definition 1.2 A *topology* on a set X is a collection $\mathcal{T}$ of subsets of X with the following properties:

1. $\emptyset$ and X are in $\mathcal{T}$.

2. Any union of elements of $\mathcal{T}$ is in $\mathcal{T}$.

3. Any *finite* intersection of elements of $\mathcal{T}$ is in $\mathcal{T}$.

The elements of the topology $\mathcal{T}$ are called *open sets*. A set X for which a topology $\mathcal{T}$ has been specified is called a *topological space*.

This is a fairly abstract definition - fortunately we don't need to work at this level of generality. In fact in everything we do we will always assume that the set X is a subset of $\mathbf{R}^n$ and that X inherits the standard topology from $\mathbf{R}^n$. To explain what we mean by this recall the following ideas from analytic geometry.

Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. The *Euclidean norm* of x is given by

$$||x||_2 := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Given a point $x \in \mathbf{R}^n$, the *ball of radius* $r > 0$ *centered at* x is given by

$$B_2(x, r) := \{y \in \mathbf{R}^n \mid ||x - y||_2 < r\}.$$

The topology on $\mathbf{R}^n$ is typically defined in terms of the Euclidean norm. Since a topology is nothing but a collection of sets that satisfies the conditions of Definition 1.2, another way of saying this is that the open sets in $\mathbf{R}^n$ can be defined in terms of the Euclidean norm.

Definition 1.3 A set $U \subset \mathbf{R}^n$ is *open* if and only if for every point $x \in U$ there exists an $\epsilon > 0$ such that $B_2(x, \epsilon) \subset U$.

The reader should check that this definition of an open set is consistent with the definition of a topology (see Exercise 1.1). This topology is called the *standard topology* on $\mathbf{R}^n$. Unless it is explicitly stated otherwise $\mathbf{R}^n$ will always be chosen to be the topological space specified by the standard topology.

Example 1.4 The interval $(-1, 2) \subset \mathbf{R}$ is an open set in the standard topology on $\mathbf{R}$. To prove this let $x \in (-1, 2)$. This is equivalent to the conditions $-1 < x$ and $x < 2$. Choose $r_0 = (x + 1)/2$ and $r_1 = (2 - x)/2$. Then, both $r_0 > 0$ and $r_1 > 0$. Let $\epsilon = \min\{r_0, r_1\}$. Thus, $B_2(x, \epsilon) \subset (-1, 2)$. Since this is true for any $x \in (-1, 2)$, we have shown that $(-1, 2)$ is an open set in the standard topology on $\mathbf{R}$.

Generalizing this argument leads to the following result.

Proposition 1.5 *Any interval of the form* (a, b), (a, ∞) *or* $(-\infty, b)$ *is open in* $\mathbf{R}$.

Proof: See Exercise 1.3. ■

From Definition 1.2.2, it follows that the arbitrary union of intervals is open, e.g. $(a,b) \cup (c,d)$ is an open set.

Example 1.6 The *unit n-ball*

$$D^n := \{x \in \mathbf{R}^n \mid ||x||_2 < 1\}$$

is an open set in the standard topology on $\mathbf{R}^n$. Observe that if $x \in D^n$ then $||x||_2 < 1$. Therefore, $0 < 1 - ||x||_2$. Let $r = \frac{1-||x||_2}{2}$. Then, $B_2(x,r) \subset D^n$.

Example 1.7 Of course not every set is open. As an example consider $(0,1] \subset \mathbf{R}$. $1 \in (0,1]$, but given any $\epsilon > 0$, $B_2(1,\epsilon) \not\subset (0,1]$. Therefore, $(0,1]$ is not open in the standard topology on $\mathbf{R}$. The same argument shows that any interval of the form $(a,b]$, $[a,b)$ or $[a,b]$ is not open in the standard topology on $\mathbf{R}$.

Since open sets play such an important role in topology it is useful to be able to refer to the largest open set contained by a set.

Definition 1.8 The *interior* of a set A is the union of all open sets contained in A. The interior of A is denoted by

$$\mathrm{int}\,(A)$$

Since the arbitrary union of open sets is open, $\mathrm{int}\,(A)$ is an open set.

One of the advantages of the abstract definition of a topology is that it does not explicitly involve a particular norm or distance. In fact, there are other norms that can be put on $\mathbf{R}^n$ which give rise to the same topology. For our purposes the *supremum norm* which is defined by

$$||x|| := \sup_{1\leq i\leq n} |x_i| \qquad \text{for } x = (x_1, \ldots, x_n) \in \mathbf{R}^n$$

is particularly convenient. Given a point $x \in \mathbf{R}^n$, the *ϵ-cube centered at x* is

$$B(x,\epsilon) := \{y \in \mathbf{R}^n \mid ||x-y|| < \epsilon\}.$$

Since the supremum norm represents a different way of measuring distance an ϵ-cube is different from an ϵ-ball (see Figure 1.1)

As before we can use this norm to define a collection of sets.

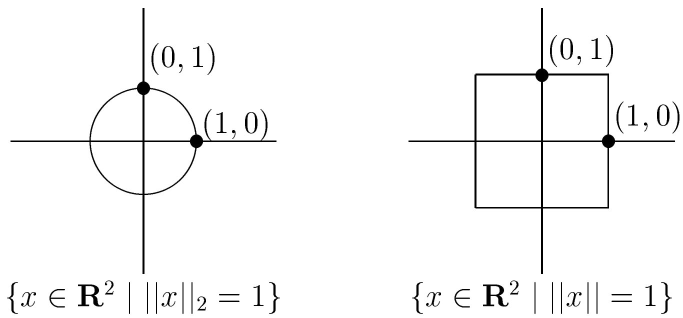


Figure 1.1: The unit distance from the origin in the Euclidean norm and the unit distance from the origin in the Supremum norm.

Definition 1.9 Let $V \in \mathcal{T}_{\text{sup}}$ if and only if for every point $x \in V$ there exists $\epsilon > 0$ such that $B(x, \epsilon) \subset V$.

Again, the reader should check that $\mathcal{T}_{\text{sup}}$ defines a topology on $\mathbf{R}^n$ (see Exercise 1.2).

Proposition 1.10 *$\mathcal{T}_{\text{sup}}$ is the same as the standard topology on $\mathbf{R}^n$.*

Proof: To prove this result it needs to be shown that every set $V \in \mathcal{T}_{\text{sup}}$ is in the standard topology and every set in the standard topology is in $\mathcal{T}_{\text{sup}}$.

Let $V \in \mathcal{T}_{\text{sup}}$. Let $x \in V$. Then there exists $\epsilon > 0$ such that $B(x, \epsilon) \subset V$. Observe that $B_2(x, \epsilon) \subset B(x, \epsilon)$. Therefore, V satisfies Definition 1.3 which means that V is in the standard topology.

Let U be an open set in the standard topology. Let $x \in U$. Then there exists $\epsilon > 0$ such that $B_2(x, \epsilon) \subset U$. One can check that $B(x, \frac{\epsilon}{\sqrt{n}}) \subset B_2(x, \epsilon)$. Thus $U \in \mathcal{T}_{\text{sup}}$. ∎

As important as an open set is the notion of a closed set.

Definition 1.11 A subset K of a topological space X is *closed* if its complement

$$X \setminus K := \{x \in X \mid x \notin K\}$$

is open.

Example 1.12 The interval $[a, b]$ is a closed subset of $\mathbf{R}$. This is straightforward to see since its complement $\mathbf{R} \setminus [a, b] = (-\infty, a) \cup (b, \infty)$ is open. Similarly, $[a, \infty)$ and $(-\infty, b] \subset \mathbf{R}$ are closed.

Example 1.13 The set $C^n := \{x \in \mathbf{R}^n \mid ||x|| \leq 1\}$ is closed. This is equivalent to claiming that $R^n \setminus C^n$ is open, i.e. that $\{x \in \mathbf{R}^n \mid ||x|| > 1\}$ is open. Observe that $||x|| > 1$ is equivalent to $\max_{i=1,\dots,n}\{|x_i|\} > 1$. Thus, there exists at least one coordinate, say the j-th coordinate, such that $|x_j| > 1$. Then

$$B(x, \frac{|x_j| - 1}{2}) \subset R^n \setminus C^n.$$

Remark 1.14 The reader should take care not to get lulled into the idea that a set is either open or closed. Many sets are *neither*. For example, the interval $(0, 1] \subset \mathbf{R}$ is neither open nor closed. As was observed in Example 1.7, it is not open. Similarly, it is not closed since its complement is $(-\infty, 0] \cup (1, \infty)$ which is not open.

Theorem 1.15 *Let X be a topological space. Then the following statements are true.*

1. *$\emptyset$ and X are closed sets.*
2. *Arbitrary intersections of closed sets are closed.*
3. *Finite unions of closed sets are closed.*

Proof: (1) $\emptyset = X \setminus X$ and $X = X \setminus \emptyset$.

(2) Let $\{K_\alpha\}_{\alpha \in \mathcal{A}}$ be an arbitrary collection of closed sets. Then

$$X \setminus \bigcap_{\alpha \in \mathcal{A}} K_\alpha = \bigcup_{\alpha \in \mathcal{A}} (X \setminus K_\alpha).$$

Since, by definition $X \setminus K_\alpha$ is open for each $\alpha \in \mathcal{A}$ and the arbitrary union of open sets is open, $X \setminus \bigcap_{\alpha \in \mathcal{A}} K_\alpha$ is open. Therefore, $\bigcap_{\alpha \in \mathcal{A}} K_\alpha$ is closed.

(3) See Exercise 1.8. ∎

Definition 1.16 Let X be a topological space and let $A \subset X$. The *closure* of A in X is the intersection of all closed sets in X containing A. The closure of A is denoted by $\operatorname{cl} A$ (many authors also use the notation $\bar{A}$.)

By Theorem 1.15 the arbitrary intersection of closed sets is closed, therefore the closure of an arbitrary set is a closed set. Also, observe that $A \subset \mathrm{cl}\, A$ and therefore $\mathrm{cl}\, A$ is the smallest closed set which contains A.

Example 1.17 Consider $[0, 1) \subset \mathbf{R}$. Then $\mathrm{cl}\,[0, 1) = [0, 1]$. This is not too difficult to prove. First one needs to check that $[0, 1)$ is not closed. This follows from the fact that $[1, \infty)$ is not open. Then one shows that $[0, 1]$ is closed by showing that $(-\infty, 0) \cup (1, \infty)$ is an open set in $\mathbf{R}$. Finally one observes that any closed set that contains $[0, 1)$ must contain $[0, 1]$.

Similar argument shows that

$$\mathrm{cl}\,(0, 1) = \mathrm{cl}\,[0, 1) = \mathrm{cl}\,(0, 1] = \mathrm{cl}\,[0, 1] = [0, 1].$$

Definition 1.18 Let X be a topological space and let $A \subset X$. The *boundary* of A is defined to be

$$\mathrm{bd}\, A := \mathrm{cl}\, A \cap \mathrm{cl}\,(X \setminus A).$$

Example 1.19 Consider $[0, 1] \subset \mathbf{R}$. From Example 1.17, $\mathrm{cl}\,[0, 1] = [0, 1]$. Observe that $\mathrm{cl}\,((-\infty, 0) \cup (1, \infty)) = (-\infty, 0] \cup [1, \infty)$. Therefore,

$$\mathrm{bd}\,[0, 1] = \{0\} \cup \{1\}$$

The following proposition gives a nice characterization of points that lie in the boundary of a set.

Proposition 1.20 *Let $A \subset X$. A point $x \in \mathrm{bd}\, A$ if and only if for every open set $U \subset X$ containing x, $U \cap A \neq \emptyset$ and $u \cap (X \setminus A) \neq \emptyset$.*

Proof: ■

Up to this point, the only topological spaces that have been considered are those of $\mathbf{R}^n$ for different values of n. The abstract definition of a topology only requires that one begin with a set X. So consider $X \subset \mathbf{R}^n$. Is there a natural way to specify a topology for X in such a way that it matches as closely as possible the topology on $\mathbf{R}^n$? The answer is yes, but we begin with a more general definition.

Definition 1.21 Let Z be a topological space with topology $\mathcal{T}$. Let $X \subset Z$. The *subspace topology* on X is the collection of sets

$$\mathcal{T}_X := \{X \cap U \mid U \in \mathcal{T}\}.$$

Before this definition can be accepted, the following proposition needs to be proved.

Proposition 1.22 $\mathcal{T}_X$ *defines a topology on* X.

Proof: The three conditions of Definition 1.2 need to be checked.

First, observe that $\emptyset \in \mathcal{T}_X$ since $\emptyset = X \cap \emptyset$. Similarly, $X \in \mathcal{T}_X$ since $X = X \cap Z$.

The intersection and union properties follow from the following equalities:

$$\begin{aligned} \bigcap_{i=1}^{n} (X \cap U_i) &= X \cap \left(\bigcap_{i=1}^{n} U_i \right) \\ \bigcup_{i \in \mathcal{I}} (X \cap U_i) &= X \cap \left(\bigcup_{i \in \mathcal{I}} U_i \right) \end{aligned}$$

for any indexing set $\mathcal{I}$. ∎

Using this definition of the subspace topology, any set $X \subset \mathbf{R}^n$ can be treated as a topological space.

It is important to notice that while open sets in the subspace topology are defined in terms of open sets in the ambient space, the sets themselves may "look" different.

Example 1.23 Consider the interval $[-1, 1] \subset \mathbf{R}$ with the subspace topology induced by the standard topology on $\mathbf{R}$. $(0, 2)$ is an open set in $\mathbf{R}$, hence

$$(0, 1] = (0, 2) \cap [-1, 1]$$

is an open set in $[-1, 1]$. We leave it to the reader to check that any interval of the form $[-1, a)$ and $(a, 1]$ where $-1 < a < 1$ is an open set in $[-1, 1]$.

Example 1.24 Let $X = [-1, 0) \cup (0, 1]$. Observe that $[-1, 0) = (-2, 0) \cap X$ and $(0, 1] = (0, 2) \cap X$, thus both are open sets. However, $[-1, 0) = X \setminus (0, 1]$ and $(0, 1] = X \setminus [-1, 0)$ so both are also closed sets. This shows that for general topological spaces one can have nontrivial sets that are both open and closed.

Exercises

1.1 Prove that Definition 1.3 defines a topology for $\mathbf{R}^2$.

1.2 Prove that $\mathcal{T}_{\text{sup}}$ defines a topology for $\mathbf{R}^2$.

1.3 Prove Proposition 1.5.

1.4 Prove that any set consisting of a single point is closed in $\mathbf{R}^n$.

1.5 Prove that $B(x, \frac{\epsilon}{\sqrt{n}}) \subset B_2(x, \epsilon)$.

1.6 Let

$$Q^n := \{x \in \mathbf{R}^n \mid 0 \leq x_i \leq 1\} \subset \mathbf{R}^n.$$

Let

$$\kappa^{n-1} := \operatorname{bd} Q^n.$$

Prove the following:

1. $Q^n \subset \mathbf{R}^n$ is closed.
2. $\kappa^{n-1} = \{x \in C^n \mid x_i \in \{0, 1\} \text{ for some } i = 1, \ldots, n\}$.

1.7 Let Z be a topological space with topology $\mathcal{T}$. Let $Y \subset X \subset Z$. Let $\mathcal{T}_X$ be the subspace topology obtained from viewing $X \subset Z$. Let $\mathcal{T}_Y$ be the subspace topology obtained from viewing $Y \subset Z$. Let $\mathcal{S}_Y$ be the subspace topology obtained from viewing $Y \subset X$ where X has the topology $\mathcal{T}_X$. Prove that $\mathcal{S}_Y = \mathcal{T}_Y$.

1.8 Prove that the finite intersection of closed sets is closed.

1.9 Let $Q = [k_1, k_1 + 1] \times [k_2, k_2 + 1] \times [k_3, k_3 + 1] \subset \mathbf{R}^3$ where $k_i \in \mathbf{Z}$ for $i = 1, 2, 3$. Prove that Q is a closed set.

1.1.2 Continuous Maps

With the notion of subspace topology we have at our disposal a multitude of different topological spaces, in particular we can topologize any subset of $\mathbf{R}^n$. A natural question is which topological spaces are "equivalent" and which are "different." The quotation marks are included because these terms need to be defined before an answer can be given.

Example 1.25 The square $X := [0,1] \times [0,1] \subset \mathbf{R}^2$ and a portion of the closed unit disk $Y := \{x \in \mathbf{R}^2 \mid ||x|| \leq 1,\ x_1 \geq 0,\ x_2 \geq 0\} \subset \mathbf{R}^2$ are clearly different from the geometric point of view: the first one is a polyhedron, the second one is not. However, we would like to think of them as being "equivalent" in a topological sense, since they can be transformed from one to the other and back by simply stretching or contracting the spaces.

To be more precise, observe that any element of Y has the form $y = (r\cos\theta, r\sin\theta)$ where $0 \leq r \leq 1$ and $0 \leq \theta \leq \pi/2$. Define $f : Y \to X$ by

$$f(r\cos\theta, r\sin\theta) := \begin{cases} (r, r\tan\theta) & \text{if } 0 \leq \theta \leq \pi/4, \\ (r\cot\theta, r) & \text{if } \pi/4 \leq \theta \leq \pi/2. \end{cases}$$

Observe that this map just expands Y by moving points out along the rays emanating from the origin.

One can also write down a map $g : X \to Y$ which shrinks X onto Y along the same rays (see Exercise 1.10).

You have already seen maps of the form of f in the previous example in your calculus class under the label of a continuous functions. Since we introduced the notion of topology on an abstract level, we need to define continuous functions in an equally abstract way.

Recall that a topological space consists of two objects, the set X and the topology $\mathcal{T}$. Therefore, to compare two different topological spaces one needs to make a comparison of both the elements of the sets - this is done using functions - and one needs to compare the open sets that make up the two topologies.

Definition 1.26 Let X and Y be topological spaces with topologies $\mathcal{T}_X$ and $\mathcal{T}_Y$, respectively. A function $f : X \to Y$ is *continuous* if and only if for every open set $V \in \mathcal{T}_Y$ its preimage under f is open in X, i.e.

$$f^{-1}(V) \in \mathcal{T}_X.$$

Even in this very general setting we can check that some maps are continuous.

Proposition 1.27 *Let X and Y be topological spaces.*

(i) The identity map $\mathbf{1}_X : X \to X$ is continuous.

(ii) *Let* $y_0 \in Y$. *The constant map* $f : X \to Y$ *given by* $f(x) = y_0$ *is continuous.*

Proof: (i) Since $\mathbf{1}_X$ is the identity map, $\mathbf{1}_X^{-1}(U) = U$ for every set $U \subset X$. Thus, if U is open, its preimage under $\mathbf{1}_X$ is open.

(ii) Let $V \subset Y$ be an open set. If $y_0 \in V$ then $f^{-1}(V) = X$ which is open. If $y_0 \notin V$, then $f^{-1}(V) = \emptyset$ which is also open. ■

Proposition 1.28 *If* $f : X \to Y$ *and* $g : Y \to Z$ *are continuous maps, then* $g \circ f : X \to Z$ *is continuous.*

Proof: Let W be an open set in Z. To show that $g \circ f$ is continuous we need to show that $(g \circ f)^{-1}(W)$ is an open set. However, $(g \circ f)^{-1}(W) = g^{-1}(f^{-1}(W))$. Since f is continuous, $f^{-1}(W)$ is open and since g is continuous $g^{-1}(f^{-1}(W))$ is open. ■

This definition tells us how we will compare topological spaces. Therefore, to say that two topological spaces are equivalent it seems natural to require that both objects, the sets and the topologies, be equivalent. On the level of set theory the equivalence of sets is usually taken to be the existence of a bijection. To be more precise, let X and Y be sets. A function $f : X \to Y$ is an *injection* if for any two points $x, z \in X$, $f(x) = f(z)$ implies that $x = z$. f is a *surjection* if for any $y \in Y$ there exists $x \in X$ such that $f(x) = y$. If f is both an injection and a surjection then it is a *bijection*. If f is a bijection then one can define an inverse map $f^{-1} : Y \to X$.

Definition 1.29 Let X and Y be topological spaces with topologies $\mathcal{T}_X$ and $\mathcal{T}_Y$, respectively. A bijection $f : X \to Y$ is a *homeomorphism* if and only if both f and f^{-1} are continuous.

Proposition 1.30 *Homeomorphism defines an equivalence relation on topological spaces.*

Proof: Recall (see A.2) that to show that homeomorphism defines an equivalence relation we need to show that it is reflexive, symmetric and transitive.

To see that it is reflexive, observe that given any topological space X the identity map $\mathbf{1}_X : X \to X$ is a homeomorphism from X to X.

Assume that X is homeomorphic to Y. By definition this implies that there exists a homeomophism $f : X \to Y$. Observe that $f^{-1} : Y \to X$ is also a homeomorphism and hence Y is homeomorphic to X. Thus, homeomorphism is a symmetric relation.

Finally, Proposition 1.28 shows that homeomorphism is a transitive relation, that is if X is homeomorphic to Y and Y is homeomorphic to Z, then X is homeomorphic to Z. ■

As before, we have introduced the notion of continuous function on a level of generality much greater than we need. The following result indicates that this abstract definition matches that learned in calculus.

Theorem 1.31 *Let $f : \mathbf{R} \to \mathbf{R}$. Then, f is continuous if and only if for every $x \in \mathbf{R}$ and any $\epsilon > 0$, there exists a $\delta > 0$ such that if $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$.*

Proof: ($\Rightarrow$) Let $f : \mathbf{R} \to \mathbf{R}$ be continuous. Consider $x \in \mathbf{R}$ and $\epsilon > 0$. Observe that the interval $B(f(x), \epsilon) = (f(x) - \epsilon, f(x) + \epsilon)$ is an open set in the range of f. Since f is continuous, $f^{-1}(B(f(x), \epsilon))$ is an open set in $\mathbf{R}$. Obviously $x \in f^{-1}(B(f(x), \epsilon))$. Hence, by the definition of an open set in the standard topology on $\mathbf{R}$, there exists $\delta > 0$ such that

$$B_(x, \delta) = (x - \delta, x + \delta) \subset f^{-1}(B(f(x), \epsilon)).$$

We will now check that this is the desired δ. If $y \in \mathbf{R}$ such that $|x - y| < \delta$, then $y \in (x - \delta, x + \delta)$ and hence $f(y) \in B(f(x), \epsilon)$. Therefore, $|f(x) - f(y)| < \epsilon$.

($\Leftarrow$) This direction is a bit more difficult since we have to check that for every open set $V \subset \mathbf{R}$, $f^{-1}(V)$ is open. With this in mind, let V be an arbitrary open set in $\mathbf{R}$. By definition for each $z \in V$ there exists $\epsilon_z > 0$ such that $B(z, \epsilon_z) \subset V$. Observe that

$$V = \bigcup_{z \in V} B(z, \epsilon_z).$$

Assume for the moment that we can prove that for every $z \in V$, $f^{-1}(B(z, \epsilon_z))$ is open. Then we are done, since

$$f^{-1}(V) = \bigcup_{z \in V} f^{-1}(B(z, \epsilon_z))$$

and the arbitrary union of open sets is open.

Thus, all that we need to prove is that given $z \in V$ and $\epsilon_z > 0$, but sufficiently small, then $f^{-1}(B(z, \epsilon_z))$ is open.

With this in mind observe that it is possible that $f^{-1}(B(z, \epsilon_z)) = \emptyset$. This is okay since $\emptyset$ is an open set. So assume that $f^{-1}(B(z, \epsilon_z)) \neq \emptyset$. Then there exists $w \in f^{-1}(B(z, \epsilon_z))$. This implies that $f(w) \in B(z, \epsilon_z) = (z - \epsilon_z, z + \epsilon_z)$. Let $\mu = \frac{1}{2}\min\{f(w) - z + \epsilon_z, z + \epsilon_z - f(w)\}$. Then, $B(f(w), \mu) \subset B(z, \epsilon_z)$

We are finally ready to use the definition of continuity from calculus. Let $\epsilon = \mu$, then there exists $\delta > 0$ such that $|w - y| < \delta$ implies $|f(x) - f(y)| < \mu$. Another way of saying this is that

$$f(B(w, \delta)) \subset B(f(w), \mu) \subset B(z, \epsilon_z).$$

This implies that $B(w, \delta) \subset f^{-1}(B(z, \epsilon_z))$. Since w was an arbitrary element of $f^{-1}(B(z, \epsilon_z))$, $f^{-1}(B(z, \epsilon_z))$ is open. ∎

A straightforward generalization of this proof gives the following theorem

Theorem 1.32 *Let $f : \mathbf{R}^n \to \mathbf{R}^m$. Then, f is continuous if and only if for every $x \in \mathbf{R}^n$ and any $\epsilon > 0$, there exists a $\delta > 0$ such that if $||x - y|| < \delta$ then $||f(x) - f(y)|| < \epsilon$.*

Thus, using Theorem 1.31 we can easily show that a variety of simple topological spaces are homeomorphic.

Proposition 1.33 *The following topological spaces are homeomorphic:*

(i) $\mathbf{R}$,

(ii) (a, ∞) *for any* $a \in \mathbf{R}$,

(iii) $(-\infty, a)$ *for any* $a \in \mathbf{R}$,

(iv) (a, b) *for any* $-\infty < a < b < \infty$.

Proof: We begin by proving that $\mathbf{R}$ and (a, ∞) are homeomorphic. Let $f : \mathbf{R} \to (a, \infty)$ be defined by

$$f(x) = a + e^x.$$

This is clearly continuous. Furthermore, $f^{-1}(x) = \ln(x - a)$ is also continuous.

Observe that $f : (a, \infty) \to (-\infty, -a)$ given by $f(x) = -x$ is a homeomorphism. Thus, any iterval of the form $(-\infty, b)$ is homeomorphic to $(-b, \infty)$ and hence to $\mathbf{R}$.

Finally, to see that (a, b) is homeomorphic to $\mathbf{R}$ observe that $f : (a, b) \to \mathbf{R}$ given by

$$f(x) = \ln\left(\frac{x-a}{b-x}\right)$$

is continuous and has a continuous inverse given by $f^{-1}(x) = (be^y + a)/(1 + e^y)$. ■

Proposition 1.34 *The following topological spaces are homeomorphic:*

1. $[-1, 1]$,
2. $[a, b]$ *for any* $-\infty < a < b < \infty$.

Proof: See Exercise 1.11. ■

Another useful way to characterize continuous functions is as follows.

Proposition 1.35 *Let* $f : X \to Y$. f *is continuous if and only if for every closed set* $K \subset Y$, $f^{-1}(K)$ *is a closed subset of* X.

Proof: ($\Rightarrow$) Let $K \subset Y$ be an a closed set. Then $Y \setminus K$ is an open set. Since f is continuous, $f^{-1}(Y \setminus K)$ is an open subset of X. Hence $X \setminus f^{-1}(Y \setminus X)$ is closed in X. Thus, it only needs to be shown that $X \setminus f^{-1}(Y \setminus K) = f^{-1}(K)$. Let $x \in X \setminus f^{-1}(Y \setminus K)$. Then $f(x) \in Y$ and $f(x) \notin Y \setminus K$. Therefore, $f(x) \in K$ or equivalently $x \in f^{-1}(K)$. Thus, $X \setminus f^{-1}(Y \setminus K) \subset f^{-1}(X)$. Now assume $x \in f^{-1}(K)$. Then, $x \notin f^{-1}(Y \setminus K)$ and hence $x \in X \setminus f^{-1}(Y \setminus K)$.

($\Leftarrow$) Let $U \subset Y$ be an open set. Then $Y \setminus U$ is a closed subset. By hypothesis, $f^{-1}(Y \setminus U)$ is closed. Thus $X \setminus f^{-1}(Y \setminus U)$ is open. But $X \setminus f^{-1}(Y \setminus U) = f^{-1}(U)$. ■

Exercises

1.10 Refering to Example 1.25:

(a) Write down the inverse function for f.

(b) Prove that f is a continuous function.

1.11 Prove Proposition 1.34.

1.1.3 Connectedness

One of the most fundamental global properties of a topological space is whether or not it can be broken into two distinct open subsets. The following definition makes this precise.

Definition 1.36 Let X be a topological space. X is *connected* if the only subsets of X that are both open and closed are $\emptyset$ and X. If X is not connected then it is *disconnected*.

Example 1.37 Let $X = [-1, 0) \cup (0, 1] \subset \mathbf{R}$. Then X is a disconnected space since by Example 1.24 $[-1, 0)$ and $(0, 1]$ are both open and closed in the subspace topology.

While it is easy to produce examples of disconnected spaces proving that a space is connected is more difficult. Even the following intuitively obvious result is fairly difficult to prove.

Theorem 1.38 *Any interval in* $\mathbf{R}$ *is connected.*

Hints as to how to prove this theorem can be found in Exercise 1.12 or the reader can consult [2]).

A very useful theorem is the following.

Theorem 1.39 *Let* $f : X \to Y$ *be a continuous function. If* X *is connected, then so is* $f(X) \subset Y$.

Proof: Let $Z = f(X)$. Suppose that Z is disconnected. Then there exists an set $A \subset Z$, where $A \neq \emptyset, Z$, that is both open and closed. Since f is continuous, $f^{-1}(A)$ is both open and closed. But $f^{-1}(A) \neq \emptyset, X$ which contradicts the assumption that X is connected. ■

We can now prove one of the more fundamental theorems from topology that you made use of in your calculus class.

Theorem 1.40 [Intermediate Value Theorem] *If* $f : [a, b] \to \mathbf{R}$ *is a continuous function and if* $f(a) > 0$ *and* $f(b) < 0$, *then there exists* $c \in [a, b]$ *such that* $f(c) = 0$.

Proof: The proof is by contradiction. Assume that there is no $c \in [a, b]$ such that $f(c) = 0$. Then

$$f([a, b]) \subset (-\infty, 0) \cup (0, \infty).$$

Let $U = (-\infty, 0) \cap f([a, b])$ and $V = (0, \infty) \cap f([a, b])$. Using the subspace topology, U and V are open sets and $f([a, b]) = U \cup V$. Since $f(a) > 0$ and $f(b) < 0$, U and V are not trivial. Therefore, $f([a, b])$ is disconnected contradicting Theorems 1.38 and 1.39. ■

Example 1.41 The half-closed interval $(0, 1]$ is not homeomorphic to the open interval $(0, 1)$. We will argue by contradiction. Suppose that $f : (0, 1] \to (0, 1)$ is a homeomorphism and let $t := f(1)$. Then the restriction of f to $(0, 1)$ is a homeomorphism of $(0, 1)$ onto the set $(0, t) \cup (t, 1)$. That is impossible since the first set is connected and the second is not, contradicting Theorem 1.39.

Exercises

1.12 This exercise leads to a proof that the interval $[0, 1]$ is a connected set. With this in mind, let A and B be two disjoint nonempty open sets in $I = [0, 1]$. The following arguments will show that $I \neq A \cup B$.

Let $a \in A$ and $b \in B$, then either $a < b$ or $a > b$. Assume without loss of generality that $a < b$.

(a) Show that the interval $[a, b] \subset I$.

Let $A_0 := A \cap [a, b]$ and $B_0 := B \cap [a, b]$.

(b) Show that A_0 and B_0 are open in $[a, b]$ under the subspace topology.

Let c be the least upper bound for A_0, i.e.

$$c := \inf\{x \in \mathbf{R} \mid x > y \text{ for all } y \in A_0\}.$$

(c) Show that $c \in [a, b]$.

(d) Show $c \notin B_0$. Use the fact that c is the least upper bound for A_0 and that B_0 is open.

(e) Show that $c \notin A_0$. Again use the fact that c is the least upper bound for A_0 and that A_0 is open.

Finally, observe that $c \in I$, but $c \notin A_0 \cup B_0$ and therefore, that $I \neq A_0 \cup B_0$.

1.13 Let A and B be connected sets. Assume that $A \cap B \neq \emptyset$. Prove that $A \cup B$ is connected.

1.14 Show that S^1 is connected.

1.15 We say that a topological space X has the *fixed point property* if every continuous map $f : X \to X$ has a fixed point, i.e. a point $x \in X$ such that $f(x) = x$.
a) Show that the fixed point property is a topological property, i.e. that it is invariant under a homeomorphism.
b) Show that any closed bounded interval $[a, b]$ has the fixed point property.
Hint: Apply the Intermediate Values Theorem to the function $f(x) - x$.

1.16 Show that the unit circle $S^1 = \{x \in \mathbf{R}^2 \mid \|x\| = 1\}$ is not homeomorphic to an interval (whether it is closed, open or neither).
Hint: Use an argument similar to that in Example 1.41.

1.17 * A *simple closed curve* in $\mathbf{R}^n$ is an image of an interval $[a, b]$ under a continuous map $\sigma : [a, b] \to \mathbf{R}^n$ (called a *path*) such that $\sigma(s) = \sigma(t)$ for any $s < t, s, t \in [a, b]$ if and only if $s = a$ and $t = b$. Prove that any simple closed curve is homeomorphic to a unit circle.

1.2 Linear Algebra

Homology theory (what we will learn in this book) provides an excellent geometric way to proceed from linear algebra to more abstract algebraic structures. As was indicated earlier, we do assume that you are familiar with the most basic ideas from linear algebra. We shall review them, but as in the previous section we shall present these ideas in a fairly general framework. If the words feel unfamiliar don't worry they will be repeated many times throughout this text.

1.2.1 Fields

Let us begin with the real numbers $\mathbf{R}$. In the previous section we were concerned with $\mathbf{R}$ as a topological space. In this section we will consider it to be a purely algebraic object. Let's review its properties in this context.

Recall that there are two operations addition $+ : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ and multiplication $\cdot : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ defined on $\mathbf{R}$. We usually write the operations as $x+y$ and $x \cdot y$ or simply xy. The operations satisfy the following conditions.

1. Addition is *commutative*,

$$x + y = y + x$$

for all $x, y \in \mathbf{R}$.

2. Addition is *associative*,

$$x + (y + z) = (x + y) + z$$

for all $x, y, z \in \mathbf{R}$.

3. There is a unique element 0 (zero) in $\mathbf{R}$ such that $x + 0 = x$ for all $x \in \mathbf{R}$. 0 is the *identity element* for addition.

4. For each $x \in \mathbf{R}$ there exists a unique element $-x \in R$ such that $x + (-x) = 0$. $-x$ is the *additive inverse* of the element x.

5. Multiplication is *commutative*,

$$x \cdot y = y \cdot x$$

for all $x, y \in \mathbf{R}$.

6. Multiplication is *associative*,

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$

for all $x, y, z \in \mathbf{R}$.

7. There is a unique element 1 (one) in $\mathbf{R}$ such that $x \cdot 1 = x$ for all $x \in \mathbf{R}$. 1 is the identity element for multiplication.

8. For each $x \in \mathbf{R} \setminus \{0\}$ there exists a unique element $x^{-1} \in R$ such that $x \cdot x^{-1} = 1$. x^{-1} is the multiplicative inverse of the element x.

9. Multiplication *distributes* over addition; that is

$$x \cdot (y + z) = x \cdot y + x \cdot z$$

for all $x, y, z \in \mathbf{R}$.

These properties can be abstracted which leads to the notion of a field.

Definition 1.42 A *field* is a set F along with two operations, addition $+ : F \times F \to F$ and multiplication $\cdot : F \times F \to F$, that satisfy properties (1) - (9).

Typically we simplify the expression of multiplication and write xy instead of $x \cdot y$.

Example 1.43 The set of complex numbers $\mathbf{C}$ and the set of rational numbers $\mathbf{Q}$ are fields.

Example 1.44 The integers $\mathbf{Z}$ do not form a field. In particular, $2 \in \mathbf{Z}$, but $2^{-1} = \frac{1}{2} \notin \mathbf{Z}$.

Example 1.45 A very useful field is $\mathbf{Z}_2$, the set of integers module 2. The rules for addition and multiplication are as follows:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

We leave it to the reader to check that properties (1)-(9) of a field are satisfied.

Example 1.46 Another field is $\mathbf{Z}_3$, the set of integers module 3. The rules for addition and multiplication are as follows:

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Again we leave it to the reader to check that properties (1)-(9) of a field are satisfied. However, we note that $-1 = 2$ and $2^{-1} = 2$.

Example 1.47 $\mathbf{Z}_4$, the set of integers module 4 is not a field. The rules for addition and multiplication are as follows:

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Observe that the element $2^{-1} \notin \mathbf{Z}_4$.

Exercises

1.18 Prove that the set of rational numbers $\mathbf{Q}$ is a field.

1.19 Let $\mathbf{Z}_n$ denote the set of integers modulo n. For which n is $\mathbf{Z}_n$ a field?

1.2.2 Vector Spaces

In your linear algebra course you learned about vector spaces, most probably the real vector spaces $\mathbf{R}^n$. As before let us think about this in an abstract manner. The first time through you should read the following definition substituting $\mathbf{R}$ for the field F and $\mathbf{R}^n$ for the vector space V.

Definition 1.48 A *vector space* over a field F is a set V with two operations, vector addition $+ : V \times V \to V$ and scalar multiplication $F \times V \to V$. Furthermore, if $u, v \in V$ then $u + v \in V$ and given $\alpha \in F$ and $v \in V$, $\alpha v \in V$. Vector addition satisfies the following conditions.

1. Vector addition is commutative,

$$v + u = u + v$$

for all vectors $u, v \in V$.

2. Vector addition is associative,

$$u + (v + w) = (u + v) + w$$

for all vectors $u, v, w \in V$.

3. There exists a unique zero vector $\mathbf{0} \in V$ such that $v + \mathbf{0} = v$ for all $v \in V$.

4. For each vector $v \in V$ there exists a unique vector $-v \in V$ such that $v + (-v) = \mathbf{0}$.

The scalar multiplication satisfies the following rules:

1. For every $v \in V$, 1 times v equals v where $1 \in F$ is the unique element one in the field.

2. For every $v \in V$ and $\alpha, \beta \in F$

$$\alpha(\beta v) = (\alpha\beta)v$$

3. For every $\alpha \in F$ and all $u, v \in V$,

$$\alpha(u + v) = \alpha u + \alpha v.$$

4. For all $\alpha, \beta \in F$ and every $v \in V$

$$(\alpha + \beta)v = \alpha v + \beta v.$$

Definition 1.49 *Let V and W be vector spaces over a field F. W is a subspace of V, if $W \subset V$.*

This definition of a vector space may look formidable, however, ignoring the formality for a moment, this is the way most calculus textbooks introduce vectors. Typically to describe the vector space $\mathbf{R}^3$ one is told that the symbols $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ represent basis vectors pointing in the x, y and z directions. They can be scaled by multiplying by a real number, e.g. $2\mathbf{i}$ or $\sqrt{3}\mathbf{j}$. Of course, $1\mathbf{i} = \mathbf{i}$ and $0\mathbf{i} = \mathbf{0}$ is the zero vector. Finally, an arbitrary vector is just a sum of these vectors, e.g.

$$v = \alpha\mathbf{i} + \beta\mathbf{j} + \gamma\mathbf{k} \tag{1.1}$$

where $\alpha, \beta, \gamma \in \mathbf{R}$.

An equivalent but different formalism is the use of column vectors. In $\mathbf{R}^3$; one lets

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

and then (1.1) becomes

$$v = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}. \tag{1.2}$$

Depending on the context we will use both formalisms in this book.

The advantage of the abstract definition of a vector space is that it allows us to talk about many different types of vector spaces.

Example 1.50 Let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ represent basis vectors for a vector space over the field $\mathbf{Z}_2$. This vector space is denoted by $\mathbf{Z}_2^3$ and the typical vector has the form

$$v = \alpha\mathbf{i} + \beta\mathbf{j} + \gamma\mathbf{k}$$

where $\alpha, \beta, \gamma \in \mathbf{Z}_2$. If we choose to write v as a column vector, then we would have

$$v = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}.$$

Since $\mathbf{Z}_2$ has only two elements we can actually write out all the vectors in the vector space $\mathbf{Z}_2^3$. Using both sets of notation they are:

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{i}+\mathbf{j} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{i}+\mathbf{k} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \mathbf{j}+\mathbf{k} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{i}+\mathbf{j}+\mathbf{k} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Observe that in this vector space each vector is its own additive inverse.

Example 1.51 One can try to do the same construction over the integers. Since $\mathbf{Z}$ is not a field we will not, by definition, get a vector space. On the other hand we can mimic what has been done before and define an algebraic object which we will denote by $\mathbf{Z}^3$. Let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ be basis elements, then it makes perfectly good sense to talk about linear combinations of these elements,

$$v = \alpha\mathbf{i} + \beta\mathbf{j} + \gamma\mathbf{k}$$

where $\alpha, \beta, \gamma \in \mathbf{Z}$. This addition is clearly associative and commutative. The zero vector is given by

$$0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}.$$

and $-v$ is given by $-\alpha\mathbf{i} + (-\beta)\mathbf{j} + (-\gamma)\mathbf{k}$. Similarly, properties 1-4 of scalar multiplication also hold. Nevertheless, since $\mathbf{Z}$ is not a field, $\mathbf{Z}^3$ is *not* a vector space. The importance of this last statement will become clear in Chapter 3.

To make it clear why in the definition of a vector space we insist that the scalars form a field we need to recall some of the most fundamental ideas from linear algebra.

Definition 1.52 Let V be a vector space. A set of vectors $S \subset V$ is *linearly independent* if for any finite set of vectors $\{v_1, \ldots, v_n\} \subset S$ the only solution to the equation

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$$

is $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. The set S *spans* V if every element $v \in V$ can be written as a finite sum of multiples of elements in S, i.e.

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

for some collection $\{v_1, \ldots, v_n\} \subset S$ and $\{\alpha_1, \ldots \alpha_n\} \subset F$. A *basis* for V is a linearly independent set of vectors in V which spans V. V is a *finite-dimensional* vector space if it has a finite basis.

One of the most important results concerning finite dimensional vector spaces is that it has a dimension.

Theorem 1.53 *If V is a finite dimensional vector space, then any two bases of V have the same number of elements.*

This theorem allows us to make the following defintion.

Definition 1.54 The *dimension* of a vector space is the number of elements in a basis.

A very closely related result is the following.

Proposition 1.55 *Let S be a linearly independent subset of a vector space V. Suppose w is a vector in V which is not in the subspace spanned by S. Then the set obtained by adjoining w to S is linearly independent.*

Proof: The proof is by contradiction. Assume that by adjoining w to S, linear independence is lost. This means that there are distinct vectors $v_1, \ldots, v_n$ S and nonzero scalars $\alpha_1, \ldots, \alpha_n, \beta$ in the field F such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n + \beta w = 0. \tag{1.3}$$

Since F is a field, $\beta^{-1} \in F$. Thus we can rewrite (1.3) as

$$w = \beta^{-1}(\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n)$$

which contradicts the assumption that w is not in the subspace spanned by S. ■

Remark 1.56 In the proof of Proposition 1.55 we made crucial use of the fact that F was a field. If we return to Example 1.51 then we can see that Proposition 1.55 need not hold in $\mathbf{Z}^3$. Let

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \right\} \quad \text{and} \quad w = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Observe that w is not in the span of S since $2^{-1} \notin \mathbf{Z}$, but $S \cup \{w\}$ is not a linearly independent set.

The previous remark may seem somewhat trivial and esoteric, but as we shall soon see it has a profound effect on the homology groups of topological spaces.

Exercises

1.20 Let $\mathbf{Z}_3^3$ denote the three dimensional vector space over the field $\mathbf{Z}_3$. Write down all the elements of $\mathbf{Z}_3^3$.

1.2.3 Linear Maps

We now turn to a brief discussion of maps between vector spaces.

Definition 1.57 Let V and W be vector spaces over a field F. A *linear map* or *linear operator* from V to W is a function $L : V \to W$ such that

$$L(\alpha v + u) = \alpha(Lv) + Lu$$

for all $u, v \in V$ and all scalars $\alpha \in F$. L is an *isomorphism* if L is invertible. The vector spaces V and W are said to be *isomorphic* if there exists an isomorphism $L : V \to W$.

A fundamental result from linear algebra is the following.

Theorem 1.58 *Let V and W be finite dimensional vector spaces over a field F. Then, V and W are isomorphic if and only if* $\dim V = \dim W$.

Definition 1.59 Let $L : V \to W$ be a linear map. The *kernel* of L is

$$\ker L := \{v \in V \mid Lv = 0\}$$

and the *image* of L is

$$\operatorname{image} L := \{w \in W \mid Lv = w \text{ for some } v \in V\}.$$

Proposition 1.60 *If $L : V \to W$ be a linear map, then* $\ker L$ *is a subspace of V and* $\operatorname{image} L$ *is a subspace of W.*

Proof: See Exercise 1.21. ■

Exercises

1.21 Prove Proposition 1.60.

1.22 Let $L : \mathbf{R}^2 \to \mathbf{R}^2$ be given by

$$L = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Compute $\ker L$ and $\operatorname{image} L$. Draw them as subspaces of $\mathbf{R}^2$.

1.2.4 Quotient Spaces

As will become clear in the next chapter, the notion of a quotient space is absolutely fundamental in algebraic topology. We will return to this type of construction over and over again.

Consider V and W, vector spaces over a field F, with W a subspace of V. Let us set

$$v \sim u \quad \text{if and only if} \quad v - u \in W.$$

Proposition 1.61 *$\sim$ defines an equivalence relation on elements of V.*

Proof: To prove that $\sim$ is an equivalence relation we need to verify the following three properties:

1. $v \sim v$ for all $v \in V$ since $v - v = 0 \in W$.
2. $v \sim u$ if and only if $u \sim v$ since $v - u \in W$ if and only if $u - v \in W$.
3. $v \sim u$ and $u \sim x$ implies $v \sim x$ since $v - u \in W$ and $u - x \in W$ implies that $v - u + u - x = v - x \in W$.

■

Because these equivalence classes are so important we will give them a special notation. Given $v \in V$ let $[v]$ denote the equivalence class of v under this equivalence relation, i.e.

$$[v] := \{u \in V \mid u - v \in W\}.$$

Observe that if $w \in W$, then $w \sim 0$ and hence $[w] = [0]$.

Definition 1.62 The *quotient space* V/W is the vector space over F consisting of the set of equivalence classes defined above. Vector addition is defined by

$$[v] + [u] := [v + u] \quad \text{for all } u, v \in V$$

and scalar multiplication is given by

$$\alpha[v] := [\alpha v] \quad \text{for all } \alpha \in F,\ v \in V.$$

We leave it to the reader to check that this does indeed define a vector space (see Problem 1.23). A little intuition as to what this represents may be in order. Consider the vector space $V = \mathbf{R}^2$. Then a typical element of V has the form

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Let us now assume that we don't care about the value of the second coordinate. This means that as far as we are concerned

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

since they agree in the first coordinate and we don't care about the value of the second coordinate. We can still add vectors, multiply by scalars and all the rest but it seems a bit inefficient to carry around the second coordinate since we are ignoring it. How can we use quotient spaces to resolve this? Let

$$W := \{w \in V \mid w = \begin{bmatrix} 0 \\ w_2 \end{bmatrix}\}.$$

Observe that W is a subspace of V and in the induced equivalence class

$$\begin{bmatrix} a \\ b \end{bmatrix} \sim \begin{bmatrix} a \\ c \end{bmatrix}$$

We can now consider the vector space V/W whose elements are the equivalence classes. This vector space is a 1-dimensional vector space, i.e. we can represent an element of V/W by a single number x, since we can easily recover the corresponding equivalence class by considering the set of vectors $\begin{bmatrix} x \\ v_2 \end{bmatrix} \subset V$.

Of course the best way to compare two different vector spaces is through linear transformations from one to the other. Consider the linear map $\pi : V \to V/W$ given by the matrix $\pi = [1\ 0]$. Then

$$\pi \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = [1\ 0] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = [v_1],$$

i.e. the second coordinate is ignored. Observe that π is surjective, i.e. every element of V/W is in the image of π. Finally, notice that $\ker \pi = W$. Thus, for this example the process of creating a quotient space is equivalent to the existence of a particular linear map. As will be made clear in Chapter 3, this is not a coincidence.

Exercises

1.23 Prove that V/W as defined in Definition 1.62 is a vector space over F. In particular, prove that vector addition and scalar multiplication are well defined operations.

1.24 Let W be the subspace of $\mathbf{R}^2$ spanned by the vector

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Draw a picture indicating the equivalence classes in $\mathbf{R}^2/W$. What is the dimension of $\mathbf{R}^2/W$?

Chapter 2

Motivating Examples

Why study algebraic topology? This chapter contains a description of problems where algebraic topological methods have proven useful. These problems have their origins in topology (not surprising), computer graphics, dynamical systems, parallel computing, and numerics. Obviously for such a broad set of issues a single chapter cannot do any of the topics justice. They are included solely for the purpose of motivating the formidable algebraic machinery we are about to start developing. This chapter is meant to be enjoyed in the sense of an entertaining story. Don't sweat the details - try to get a feeling for the big picture. We will return to these topics throughout the rest of this book.

2.1 Topology

The importance in linear algebra of the dimension of a vector space is that any two finite dimensional vector spaces (over the same field) of the same dimension are isomorphic. In other words from the point of view of linear algebra they are indistinguishable. Said yet another way, the set of finite dimensional vector space can be classified according to a single natural number.

Algebraic topology is an attempt to do a similar thing, but in the context of topological spaces. Since topological spaces are more varied than vector spaces, the classification is done in terms of algebraic objects rather than the natural numbers. As pertains to this book the goal is as follows. Given a topological space X we want to define an algebraic object $H_*(X)$, called the homology of X, which is a *topologically invariant*; that is, if X and Y are homeomorphic then $H_*(X)$ and $H_*(Y)$ are isomorphic.

2.1.1 Homotopy

Notice that we did not claim that homology classifies spaces up to homeomorphism. It is *not* true that if two spaces have the same homology, then they are homeomorphic. Unfortunately, the classification problem in topology is too difficult for any purely algebraic classification. In fact, this problem is so difficult, that mathematicians have pretty much given up trying to classify arbitrary topological spaces up to homeomorphism. Instead they study the weaker equivalence relation known as homotopy type. Before giving the definition let us consider a motivating example.

We begin by recalling the intermediate value theorem which we proved earlier (Theorem 1.40).

Theorem 2.1 *If $f : [a, b] \to \mathbf{R}$ is a continuous function and if $f(a) > 0$ and $f(b) < 0$, then there exists $c \in [a, b]$ such that $f(c) = 0$.*

This is a model topological theorem. The function is only assumed to be continous, global rather than local information is assumed, i.e. the values of the end points are given, and yet one is still able to draw a conclusion concerning the behavior of the function on its domain.

Homology provides us with a variety of algebraic tools for determining if there exists a point c such that $f(c) = 0$. But this process of going from topology to algebra loses information. This should not be surprising. Think back to calculus where one uses the derivative to obtain a linear approximation of the differentiable function. Many different functions can have the same derivative at a point. To get a better approximation one has to use Taylor polynomials. In fact only analytic functions can be approximated exactly by their derivatives.

What families of spaces or maps will give us the same algebraic toplogical information? To answer this consider again the intermediate value theorem. The only important points are the endpoints so let $f, g : [a, b] \to \mathbf{R}$ be different continuous functions with $f(a) > 0$ and $g(a) > 0$, and $f(b) < 0$ and $g(b) < 0$. Now consider the family of functions $F : [a, b] \times [0, 1] \to \mathbf{R}$ defined by

$$F(x, s) = (1 - s)f(x) + sg(x).$$

Observe that $F(\cdot, 0) = f(\cdot)$ and $F(\cdot, 1) = g(\cdot)$. For any fixed value of $s \in [0, 1]$ we have yet another function $F(\cdot, s) : [a, b] \to \mathbf{R}$. Observe that

$$F(a, s) = (1 - s)f(a) + sg(a) > 0$$

and

$$F(b, s) = (1 - s)f(b) + sg(b) < 0$$

Thus, we can apply the intermediate value theorem to $F(\cdot, s)$ for any $s \in [0, 1]$. This family of functions is a special case of what is known as a homotopy.

Definition 2.2 Let X and Y be topological spaces. Let $f, g : X \to Y$ be continuous functions. f is *homotopic* to g if there exists a continuous map $F : X \times [0, 1] \to Y$ such that

$$F(x, 0) = f(x) \quad \text{and} \quad F(x, 1) = g(x)$$

for each $x \in X$. The map F is called a *homotopy* between f and g. f homotopic to g is denoted by $f \sim g$.

It is fairly straight forward to check that homotopy is an equivalence relation (see Excercise 2.1). How does this help us with the classification problem in topology? Since homotopy is an equivalence relation it can be used to define an equivalence between topological spaces.

Definition 2.3 Two topological spaces X and Y are *homotopic* if there exist continuous functions $f : X \to Y$ and $g : Y \to X$ such that

$$g \circ f \sim 1_X \quad \text{and} \quad f \circ g \sim 1_Y$$

where 1_X and 1_Y denote the identity maps. X homotopic to Y is denoted by $X \sim Y$.

Example 2.4 Two topological spaces can appear to be quite different and still be homotopic. For example it is clear that $\mathbf{R}^n$ is not homeomorphic to the point $\{0\}$. On the other hand these two spaces are homotopic. To see this let $f : \mathbf{R}^n \to \{0\}$ be defined by $f(x) = 0$ and let $g : \{0\} \to \mathbf{R}^n$ be defined by $g(0) = 0$. Observe that $f \circ g = 1_{\{0\}}$ and hence $f \circ g \sim 1_{\{0\}}$. To show that $g \circ f \sim 1_{\mathbf{R}^n}$ consider the function $F : \mathbf{R}^n \times [0, 1] \to \mathbf{R}^n$ defined by

$$F(x, s) = (1 - s)x.$$

Clearly, $F(x, 0) = x = 1_{\mathbf{R}^n}$ and $F(x, 1) = 0$.

A special case of homotopy is that of a deformation retract.

Definition 2.5 Let $A \subset X$. A *deformation retraction* of X onto A is a continuous map $F : X \times [0,1] \to X$ such that

$$\begin{aligned} F(x,0) = x &\quad \text{for} \quad x \in X \\ F(x,1) \in A &\quad \text{for} \quad x \in X \\ F(a,s) = a &\quad \text{for} \quad a \in A. \end{aligned}$$

If such an F exists then A is called a *deformation retract* of X. It is easy to check that if A is a deformation retract of X and B is a deformation retract of A, then B is a deformation retract of X.

Example 2.6 $\{0\}$ is a deformation retract of $[0,1]$. Define $F : [0,1] \to \{0\}$ by $F(x,s) = (1-s)x$.

Homology has the property that if two spaces are homotopic then their homologies are the same. On the other hand, there are spaces with the same homologies which are not homotopic. Thus, the algebraic invariants that we will develop in this book are extremely crude measurements of the topology of the space. Still there are interesting problems to which one can apply homology theory.

Example 2.7 Let

$$\Gamma^n := \{x \in \mathbf{R}^{n+1} \mid ||x|| = 1\}.$$

There is no deformation retraction of Γ^n to a point. We include this example at this point to try to indicate that this is a nontrivial problem. In particular, we encourage you to try to find a proof of this fact. As motivation for the study of this subject we assure you that once you know homology theory, this example will become a triviality.

Exercises

2.1 Prove that homotopy is an equivalence relation.

2.2 Let $f, g : X \to Y$ be continuous maps. Under the following assumptions on X and Y prove that $f \sim g$.

- $X = Y = [0,1]$
- $X = \Gamma^1$ and $Y = [0,1]$
- X is any topological space and $y \in Y$ is a deformation retract of Y.

Obviously, if you prove the last case, then you have proven the first two.

2.3 Prove that $\mathbf{R}^n \setminus \{0\}$ is homotopic to S^{n-1}.

2.1.2 Graphs

Up to now we have given no indication how one moves from the topology to the algebra. To motivate the ideas and build some intuition before beginning with the formal definitions it is useful to have a simple but large class of topological spaces.

Definition 2.8 A *finite graph* G consists of a finite collection of points in $\mathbf{R}^3$ $\{v_1, \ldots v_n\}$, called *vertices*, together with straight line segments $\{e_1, \ldots, e_m\}$, joining vertices, called *edges* which satisfy the following intersection conditions:

1. if two edges intersect nontrivially, then they intersect at a unique vertex, and
2. if an edge and a vertex intersect, then the vertex is an endpoint of the edge.

A *loop* L in the graph is a union of edges $e_1, e_2, \ldots, e_k$ such that $e_j \cap e_{j+1} \neq \emptyset$ for $j = 1, \ldots, k-1$, and $e_k \cap e_1 \neq \emptyset$. A graph which is connected and has no loops is called a *tree*.

One of the important properties of homology is that it can be determined from combinatorial information. With this in mind we present the following definition which indicates there is a natural reduction of a finite graph to a combinatorial object.

Definition 2.9 An *abstract finite graph* is a pair $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a finite set whose elements are called *vertices* and $\mathcal{E}$ is a collection of pairs of distinct elements of $\mathcal{V}$ called *edges*.

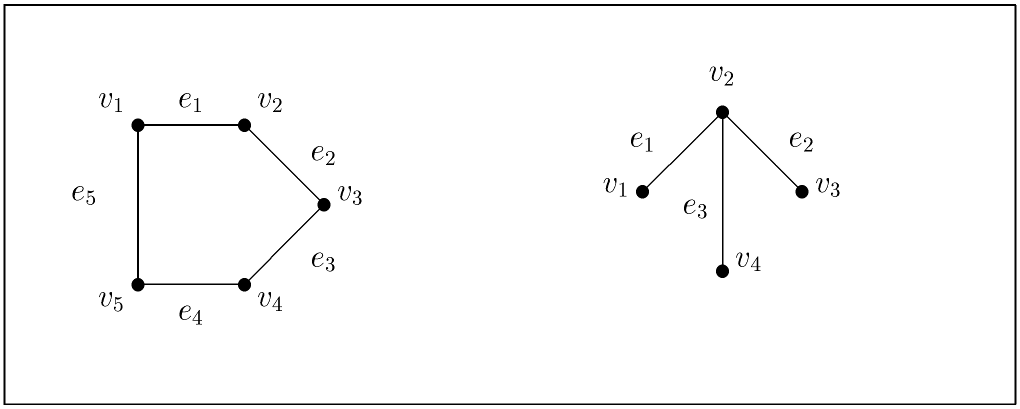


Figure 2.1: A loop and a tree.

Before turning to the algebra we want to consider the topology of trees. In particular, we will show that any tree is homotopically equivalent to a single point.

A vertex which only intersects a single edge is called a *free vertex.*

Proposition 2.10 *Every tree which contains at least one free vertex.*

Proof: Assume not. Then there exists a tree T with 0 free vertices. Let n be the number of edges in T. Let e_1 be an edge in T. Label its vertices by v_1^- and v_1^+. Since T has no free vertices, there is an edge e_2 with vertices $v_2^\pm$ such that $v_1^+ = v_2^-$. Continuing in this manner we can label the edges by e_i and the vertices by $v_i^\pm$ where $v_i^- = v_{i-1}^+$. Note since there are only a finite number of vertices, at some point $v_i^+ = v_j^-$ for some $i > j \geq 1$. Then $\{e_j, e_{j+1}, \ldots, e_i\}$ forms a loop. This is a contradition. ■

Lemma 2.11 *Every edge is homotopic to a point.*

Proof: Let e be an edge with vertices v^- and v^+. Since e is a line segment it is homeomorphic to $[0,1]$. Let $h : [0,1] \to e$ be such a homeomorphism with the property that $h(0) = v^-$ and $h(1) = v^+$. Define $F_e : e \times [0,1] \to e$ by $F_e(x,s) = h(sh^{-1}(x))$. Observe that $F(x,1) = h(h^{-1}(x)) = x$ and hence is the identity. $F_e(x,0) = h(0 \cdot h^{-1}(x)) = h(0) = v^-$. Therefore, F_e defines a retract of e to v^-. ■

Proposition 2.12 *Every tree T contains a vertex v such that there exists a deformation retraction of T onto v.*

Proof: The proof is by induction on the number of edges in the tree.

The simplest tree consists of a single edge. By Lemma 2.11 this is homotopic to a point. The homotopy is the deformation retraction.

Assume that the result is true for all trees with n edges or less. We need to show it is true for a tree with $n+1$ edges. Let T be a tree with $n+1$ edges. By Proposition 2.10 T has a free vertex v^+. Let e be the edge which contains the vertex v^+. Let the other vertex of e be denoted by v^-. Let T' be the tree obtained from T by removing the edge e and the vertex v^+. Now define $G : T \times [0,1] \to T$ by

$$G(x,s) = \begin{cases} x & \text{if } x \in T' \\ F_e(x,s) & \text{if } x \in e \end{cases}$$

where F_e is defined as in Lemma 2.11. This shows that G is a deformation retraction of T onto T'.

The result now follows by induction. ∎

Exercises

2.4 Up to homotopy how many different planar graphs are there with 5 edges?

2.1.3 A Preview of Homology

In Example 2.6 we showed that an interval is homotopic to a point. In fact by Proposition 2.12 every tree is homotopic to a point. Since a point is the simplest nontrivial topological space, up to homotopy trees must be fairly simple topological spaces. In Example 2.7 it was stated that Γ^1 is not homotopic to a point. We will use this contrast to motivate how homology can be used to measure the difference in the complexity of these two topological spaces. However, we need to begin with a word of caution. The proof that homology is a topological invariant is fairly complicated. As such it will be dealt with much later.

To make clear at the outset why working with a graph is not sufficient to establish the topological invariance of homology, observe that given our definition, a finite graph is a fixed subset of $\mathbf{R}^3$. However, as the following example indicates different graphs can give rise to the same subset.

Example 2.13 Consider the family of graphs defined by

$$G_n = \{\{j/n\} \mid j = 0, \dots n\} \cup \{[j/n, (j+1)/n] \mid j = 0, \dots, n-1\}$$

Observe that as a subset of $\mathbf{R}$ each graph describes $[0,1]$. Thus, the same topological space has many different representations as a graph. In our motivation of homology we will use abstract graphs. Thus to prove topological invariance we would have to show that given any two abstract graphs that are associated to finite graphs that in turn represent homeomorphic spaces the corresponding homology is the same. This is not trivial.

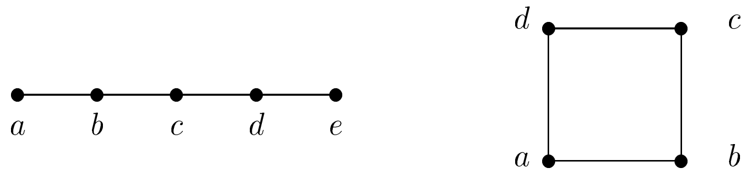
Having made these explicit disclaimers we now take the liberty of using language in which we are implicitly assuming that we are working with a topological invariant. With this in mind we begin by asking the question how can we show that $[0,1]$ and Γ^1 are topologically different.

It is worth making an observation at this point. Locally, $(-1,1)$ and Γ^1 are indistinguishable. More precisely given points $x \in (-1,1)$, $y \in \Gamma^1$, and sufficiently small neighborhoods, U_x and V_y of these points, then there exists a homeomorphism between U_x and V_y. Locally the only difference between $[0,1]$ and Γ^1 are the boundary points $\{0,1\}$ of $[0,1]$. We shall try to measure this distinction algebraically.

A word of caution is needed before we go further. The notion of topological boundary is ambiguous here because it depends on the outer space the graph is imbedded to. For instance, let a, b be two distinct vertices in $\mathbf{R}^2$. Then $\operatorname{bd}[0,1] = \emptyset$ in the topology of $[a,b]$, $\operatorname{bd}[a,b] = \{a,b\}$ in the topology of the line passing through a and b, and $\operatorname{bd}[a,b] = [a,b]$ in the topolgy of $\mathbf{R}^2$. But no matter what is the outer space, the points a, b are clearly distinct from the other points of $[a,b]$ in the sense that they are extreme points of the interval. That distinction is exhibited by the following definition.

Definition 2.14 A point x of a graph G is called a *regular point* of G if a sufficiently small ball in G around x is homeomorphic to an open interval. A point which is not a regular point is called an *extreme point* of G. The set of all extreme points of G is called the *geometric boundary* of G and denoted by $\operatorname{bd} G$.

Let us now think of $[0,1]$ and Γ^1 as graphs. To be precise consider the graphs indicated in Figure 2.2. $[0,1]$ is represented by a graph consisting of four intervals $[a,b]$, $[b,c]$, $[c,d]$ and $[d,e]$.



$\mathcal{E} = \{[a,b],[b,c],[c,d],[d,e]\}$

$\mathcal{V} = \{a,b,c,d,e\ \}$

$\mathcal{E}' = \{[a,b],[b,c],[c,d],[d,a]\}$

$\mathcal{V}' = \{\{a\},\{b\},\{c\},\{d\}\ \}$

Figure 2.2: Finite graphs and corresponding abstract finite graphs for $[0,1]$ and Γ^1

We mentioned earlier that the boundary points of $[0,1]$ are where we can see a difference in local topology. To keep our computations local we indicated in the left hand column of Table 2.1 the topological boundaries of each of the edges. In the right hand column are what for the moment can be considered fictional algebraic quantities derived from the corresponding elements of the abstract finite graph.

Topology		Algebra
$\text{bd}\,[a,b] = \{a\} \cup \{b\}$		$\partial[\mathbf{a},\mathbf{b}] = \mathbf{a}+\mathbf{b}$
$\text{bd}\,[b,c] = \{b\} \cup \{c\}$		$\partial[\mathbf{b},\mathbf{c}] = \mathbf{b}+\mathbf{c}$
$\text{bd}\,[c,d] = \{c\} \cup \{d\}$		$\partial[\mathbf{c},\mathbf{d}] = \mathbf{c}+\mathbf{d}$
$\text{bd}\,[d,e] = \{d\} \cup \{e\}$		$\partial[\mathbf{d},\mathbf{e}] = \mathbf{d}+\mathbf{e}$

Table 2.1: Topological and algebric boundaries in $[0,1]$.

On the topological level addition and subtraction of edges and points is not an obvious concept. On our fictional algebraic level, however, we will allow ourselves this luxury. Recalling the discussion in the previous chapter where we described vector spaces, we write the algebraic objects in bold and allow ourselves to formally add them. For example $\{a\}$ becomes $\mathbf{a}$. What should we use for the scalars? A possible idea is $\mathbf{Z}_2$ - this way, if we make ∂

a linear operator, we can match the topological expression

$$\operatorname{bd}([a,b] \cup [b,c]) = \{a\} \cup \{c\}$$

with the algebraic expression

$$\begin{aligned} \partial([\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}]) &= \partial([\mathbf{a},\mathbf{b}]) + \partial([\mathbf{b},\mathbf{c}]) \\ &= \mathbf{a} + \mathbf{b} + \mathbf{b} + \mathbf{c} \\ &= \mathbf{a} + 2\mathbf{b} + \mathbf{c} \\ &= \mathbf{a} + \mathbf{c}. \end{aligned}$$

Continuing in this way we have that

$$\begin{aligned} \partial([\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{e}]) &= \mathbf{a} + \mathbf{b} + \mathbf{b} + \mathbf{c} + \mathbf{c} + \mathbf{d} + \mathbf{d} + \mathbf{e} \\ &= \mathbf{a} + \mathbf{e}. \end{aligned}$$

As an indication that we are not too far off track observe that on the topological level $\operatorname{bd}[0,1] = \{a\} \cup \{e\}$.

Doing the same for the graph and abstract graph representing Γ^1 we get Table 2.2. Adding up the algebraic boundaries we have

$$\partial([\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{a}]) = 0. \tag{2.1}$$

Topology		Algebra
$\operatorname{bd}[a,b] = \{a\} \cup \{b\}$		$\partial[\mathbf{a},\mathbf{b}] = \mathbf{a} + \mathbf{b}$
$\operatorname{bd}[b,c] = \{b\} \cup \{c\}$		$\partial[\mathbf{b},\mathbf{c}] = \mathbf{b} + \mathbf{c}$
$\operatorname{bd}[c,d] = \{c\} \cup \{d\}$		$\partial[\mathbf{c},\mathbf{d}] = \mathbf{c} + \mathbf{d}$
$\operatorname{bd}[d,a] = \{d\} \cup \{a\}$		$\partial[\mathbf{d},\mathbf{a}] = \mathbf{d} + \mathbf{a}$

Table 2.2: Topology and algebra of boundaries in Γ^1.

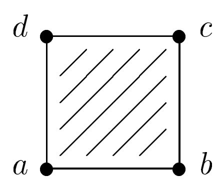
Based on these two examples one might make the extravagent claim that spaces with *cycles*, i.e. algebraic objects whose boundaries add up to zero, are topologically nontrivial. This is almost true.

To see how this fails, observe that $\Gamma^1 \subset C^2$, and in fact $\Gamma^1 = \operatorname{bd} C^2$. Since there exists a deformation retract of C^2 to a point we need to understand

Topology	Algebra
$\operatorname{bd} C^2 = [a,b] \cup [b,c] \cup [c,d] \cup [d,a]$	$\partial \mathbf{C^2} = [\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{a}]$
$\operatorname{bd}[a,b] = \{a\} \cup \{b\}$	$\partial[\mathbf{a},\mathbf{b}] = \mathbf{a} + \mathbf{b}$
$\operatorname{bd}[b,c] = \{b\} \cup \{c\}$	$\partial[\mathbf{b},\mathbf{c}] = \mathbf{b} + \mathbf{c}$
$\operatorname{bd}[c,d] = \{c\} \cup \{d\}$	$\partial[\mathbf{c},\mathbf{d}] = \mathbf{c} + \mathbf{d}$
$\operatorname{bd}[d,a] = \{d\} \cup \{a\}$	$\partial[\mathbf{d},\mathbf{a}] = \mathbf{d} + \mathbf{a}$

Table 2.3: Topology and algebra of boundaries in C^2.

how the nontrivial algebra in Γ^1 becomes trivialized. To do this we need to go beyond graphs into cubical complexes which will be defined later. For the moment consider the picture and collection of sets in Figure 2.3. The new aspect is the square C^2. This is coded in the combinatorial information as the element $\{C^2\}$.



$$\{C^2\}$$

$$\mathcal{E}' = \{[a,b],[b,c],[c,d],[d,a]\}$$

$$\mathcal{V}' = \{a,b,c,d\}$$

Figure 2.3: Simplicial complex and corresponding abstract simplicial complex for C^2.

Table 2.3 contains the topological boundary information and fictional algebra that we are associating to it for C^2.

Since $\Gamma^1 \subset C^2$, one should expect to see the contents of Table 2.2 contained in Table 2.3. Now observe that

$$\partial \mathbf{C^2} = [\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{a}].$$

Equation (2.1) indicated that the cycle $[\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{a}]$ was the interesting algebraic aspect of Γ^1. In C^2 it appears as the boundary of an object. The observation that we will make is that cycles which are boundaries should be considered trivial.

Restating this purely algebraically what we are looking for are cycles, i.e. elements of the kernel of some operator. Let us denote this operator by ∂ to remind us that it should be related to taking the boundary of a topological space. Furthermore, if this cycle is a boundary, i.e. the image of this operator, then we wish to ignore it. In other words we are interest in an algebraic quantity which takes the form

$$\text{kernel of } \partial/\text{image of } \partial.$$

We have by now introduced many vague and complicated notions. If you feel things are spinning out of control - don't worry, be happy! Admittedly, there are a lot of loose ends that we need to tie up and we will begin to do so in the next chapter. The process of developing new mathematics typically involves developing new intuitions and finding new patterns - in this case we have the advantage of knowing that it will all work out in the end. For now lets just enjoy trying to match topology and algebra.

In fact, lets do it again. Recall that earlier we asked the question what should be use for scalars? We chose $\mathbf{Z}_2$ last time. Are there other choices that make sense? Consider Figure 2.4 which looks alot like Figure 2.2 except that we have added arrows to our graphs to suggest a direction (the fancy word is orientation) through which we traverse the interval. Similarly, we have indicated a direction through which we can traverse the loop Γ^1. We could argue that $\mathbf{Z}$ is a natural choice since it is not clear what a fractional amount of a vertex or an edge of an abstract graph should represent. Furthermore, using the integers we can assign a plus or a minus sign to the edge or vertex depending on whether we traverse it following the assigned direction or not.

So let us declare that

$$\begin{aligned} \partial([\mathbf{a},\mathbf{b}]+[\mathbf{b},\mathbf{c}]+[\mathbf{c},\mathbf{d}]+[\mathbf{d},\mathbf{e}]) &= \mathbf{b}-\mathbf{a}+\mathbf{c}-\mathbf{b}+\mathbf{d}-\mathbf{c}+\mathbf{e}-\mathbf{d} \\ &= \mathbf{e}-\mathbf{a}. \end{aligned}$$

Again we see that there is consistency between the algebra and the topology since $\operatorname{bd}[0,1]=\{e\}\cup\{a\}$ and the arrows suggest traversing from a to e.

Doing the same for the graph and abstract graph representing Γ^1 gives rise to Table 2.4

Again, we see that the algebra that corresponds to the interesting topology is a cycle - a sum of algebraic objects whose boundaries add up to zero.

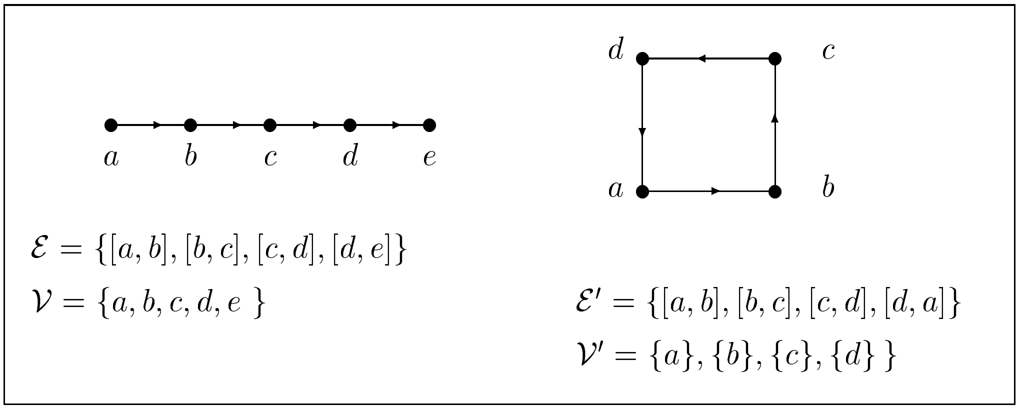


Figure 2.4: Finite graphs and corresponding abstract finite graphs for $[0,1]$ and Γ^1 with a sense of direction.

Topology		Algebra
$\mathrm{bd}\,[a,b] = \{a\} \cup \{b\}$		$\partial\!<\mathbf{a},\mathbf{b}> = \mathbf{b} - \mathbf{a}$
$\mathrm{bd}\,[b,c] = \{b\} \cup \{c\}$		$\partial\!<\mathbf{b},\mathbf{c}> = \mathbf{c} - \mathbf{b}$
$\mathrm{bd}\,[c,d] = \{c\} \cup \{d\}$		$\partial\!<\mathbf{c},\mathbf{d}> = \mathbf{d} - \mathbf{c}$
$\mathrm{bd}\,[d,a] = \{d\} \cup \{a\}$		$\partial\!<\mathbf{d},\mathbf{a}> = \mathbf{a} - \mathbf{d}$

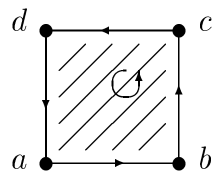
Table 2.4: Topology and algebra of boundaries in Γ^1 using $\mathbf{Z}$ coefficients.

More precisely we again arrive at equation (2.1). We still need to understand what happens to this algebra when we consider $\Gamma^1 \subset C^2$. Consider Figure 2.5. Table 2.5 contains the topological boundary information and fictional algebra that we are associating to it for C^2.

Since $\Gamma^1 \subset C^2$, we again see the contents of Table 2.4 contained in Table 2.5. As before

$$\partial \mathbf{C^2} = [\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{a}].$$

Equation (2.1) indicated that the cycle $[\mathbf{a},\mathbf{b}] + [\mathbf{b},\mathbf{c}] + [\mathbf{c},\mathbf{d}] + [\mathbf{d},\mathbf{a}]$ was the interesting algebraic aspect of Σ^1. In C^2 it appears as the boundary of an object. Again, the observation that we will make is: cycles which are boundaries should be considered trivial.



$\{C^2\}$

$\mathcal{E}' = \{[a, b], [b, c], [c, d], [d, a]\}$

$\mathcal{V}' = \{a, b, c, d\}$

Figure 2.5: Simplicial complex and corresponding abstract simplicial complex for C^2.

Topology	Algebra
$\operatorname{bd} C^2 = \Gamma^1 = [a, b] \cup [b, c] \cup [c, d] \cup [d, a]$	$\partial \mathbf{C^2} = [\mathbf{a}, \mathbf{b}] + [\mathbf{b}, \mathbf{c}] + [\mathbf{c}, \mathbf{d}] + [\mathbf{d}, \mathbf{a}]$
$\operatorname{bd} [a, b] = \{a\} \cup \{b\}$	$\partial[\mathbf{a}, \mathbf{b}] = \mathbf{b} - \mathbf{a}$
$\operatorname{bd} [b, c] = \{b\} \cup \{c\}$	$\partial[\mathbf{b}, \mathbf{c}] = \mathbf{c} - \mathbf{b}$
$\operatorname{bd} [c, d] = \{c\} \cup \{d\}$	$\partial[\mathbf{c}, \mathbf{d}] = \mathbf{d} - \mathbf{c}$
$\operatorname{bd} [d, a] = \{d\} \cup \{a\}$	$\partial[\mathbf{d}, \mathbf{a}] = \mathbf{a} - \mathbf{d}$

Table 2.5: Topology and algebra of boundaries in C^2.

Exercises

2.5 Repeat the above computations for a graph which represents a triangle in the plane.

2.1.4 $\mathbf{Z}_2$ Homology of Graphs

We have done the same example twice using different scalars but the conclusion was the same. We should look for a linear operator that somehow algebraically mimics what is done by taking the topological boundary. Then, having found this operator we should look for cycles (elements of the kernel) but ignore boundaries (elements of the image). This is still pretty fuzzy so lets do it again; a little slower and more formally, but in the general setting of graphs using the algebra of vector spaces.

Let G be an abstract graph. Let G_0 denote the set of vertices of G and let G_1 denote the set of edges of G. We will construct two vector spaces $C_0(G;\mathbf{Z}_2)$ and $C_1(G;\mathbf{Z}_2)$ as follows. Declare the set of vertices G_0 to be the set of basis elements of $C_0(G;\mathbf{Z}_2)$ and let the scalar field be $\mathbf{Z}_2$. Thus, if $G_0 = \{v_1,\dots,v_n\}$, then the collection $\{\mathbf{v}_i \mid i = 1,\dots,n\}$ is a basis for $C_0(G;\mathbf{Z}_2)$ and the typical element of $C_0(G;\mathbf{Z}_2)$ takes the form

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$$

where $\alpha_i \in \mathbf{Z}_2$.

Similarly, let the set of edges G_1 be the set of basis elements of $C_1(G;\mathbf{Z}_2)$ and again let the scalar field be $\mathbf{Z}_2$. If $G_1 = \{e_1,\dots,e_k\}$, then the collection $\{\mathbf{e}_i \mid i = 1,\dots,k\}$ is a basis for $C_1(G;\mathbf{Z}_2)$ and the typical element of $C_1(G;\mathbf{Z}_2)$ takes the form

$$\mathbf{e} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \cdots + \alpha_k\mathbf{e}_k$$

where $\alpha_i \in \mathbf{Z}_2$. The vector spaces $C_i(G;\mathbf{Z}_2)$ are called the *i-chains* for G.

It is convenient to introduce two more vector spaces $C_2(G;\mathbf{Z}_2)$ and $C_{-1}(G;\mathbf{Z}_2)$. We will always take $C_{-1}(G;\mathbf{Z}_2)$ to be the trivial vector space, i.e. the vector space consisting of exactly one element $\mathbf{0}$. For graphs we will also set $C_2(G;\mathbf{Z}_2)$ to be the trivial vector space. As we will see later for more complicated spaces this need not be the case.

We now need to formally define the boundary operators that were alluded to earlier. Let

$$\begin{aligned}\partial_0 &: C_0(G;\mathbf{Z}_2) \to C_{-1}(G;\mathbf{Z}_2)\\ \partial_1 &: C_1(G;\mathbf{Z}_2) \to C_0(G;\mathbf{Z}_2)\\ \partial_2 &: C_2(G;\mathbf{Z}_2) \to C_1(G;\mathbf{Z}_2)\end{aligned}$$

be *linear maps*. Since we have chosen bases for these vector spaces, we can think of ∂_0, ∂_1 and ∂_2 as matrices. Since $C_{-1}(G;\mathbf{Z}_2) = \mathbf{0}$, it is clear that ∂_0 must be the matrix with all zeros. Similarly, ∂_2 is the zero matrix. The entries of the matrix ∂_1 are determined by how ∂_1 acts on the basis elements, i.e. the edges $\mathbf{e}_i$. In line with the previous discussion we make the following definition. Let the edge e_i have vertices v_j and v_k. Define

$$\partial_1\mathbf{e}_i := \mathbf{v}_j + \mathbf{v}_k.$$

In our earlier example we were interested in cycles, i.e. elements of the kernel of the boundary operator. So define

$$Z_0(G;\mathbf{Z}_2) := \ker \partial_0 = \{v \in C_0(G;\mathbf{Z}_2) \mid \partial_0 v = 0\}$$
$$Z_1(G;\mathbf{Z}_2) := \ker \partial_1 = \{v \in C_1(G;\mathbf{Z}_2) \mid \partial_1 v = 0\}$$

Since $\partial_0 = 0$ it is obvious that $Z_0(G;\mathbf{Z}_2) = C_0(G;\mathbf{Z}_2)$.

We also observed that cycles which are boundaries are not interesting. To formally state this, define the set of boundaries to be

$$B_0(G;\mathbf{Z}_2) := \operatorname{im} \partial_1 = \{v \in C_0(G;\mathbf{Z}_2) \mid \exists\, e \in C_1(G;\mathbf{Z}_2) \text{ such that } \partial_1 e = v\}$$
$$B_1(G;\mathbf{Z}_2) := \operatorname{im} \partial_1 = \{0 \in C_0(G;\mathbf{Z}_2)\}$$

Observe that $B_0(G;\mathbf{Z}_2) \subset C_0(G;\mathbf{Z}_2) = Z_0(G;\mathbf{Z}_2)$. Since we have not yet defined ∂_2 we shall for the moment declare $B_1(G;\mathbf{Z}_2) = \mathbf{0}$. We can finally define homology in this rather special setting. For $i = 0, 1$ the i-th homology with $\mathbf{Z}_2$ coefficients is defined to be the quotient space

$$H_i(G;\mathbf{Z}_2) := Z_i(G;\mathbf{Z}_2)/B_i(G;\mathbf{Z}_2).$$

Observe that since this is a quotient space of vector spaces, homology with $\mathbf{Z}_2$ coefficients is a vector space.

Let us compute the homology for the graphs in Figure 2.2.

Example 2.15 Let G be the graph representing $[0, 1]$. Then,

$$\begin{aligned} G_0 &= \{a, b, c, d, e\} \\ G_1 &= \{[a, b], [b, c][c, d], [d, e]\} \end{aligned}$$

Since G_0 and G_1 are the bases for the 0-chains and 1-chains we have that

$$\begin{aligned} C_0(G;\mathbf{Z}_2) &\approx \mathbf{Z}_2^5 \\ C_1(G;\mathbf{Z}_2) &\approx \mathbf{Z}_2^4. \end{aligned}$$

To do the computations it is convenient to use a column vector notation. So let

$$\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{e} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

and

$$[\mathbf{a}, \mathbf{b}] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ [\mathbf{b}, \mathbf{c}] = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ [\mathbf{c}, \mathbf{d}] = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ [\mathbf{d}, \mathbf{e}] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

With this convention, ∂_1 becomes the 5×4 matrix

$$\partial_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Lets do a quick check. For example

$$\partial_1[\mathbf{b}, \mathbf{c}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{b} + \mathbf{c}.$$

Now consider $Z_1(G; \mathbf{Z}_2) := \ker \partial_1$. Observe that the vector $v \in C_1(G; \mathbf{Z}_2)$ is in $Z_1(G; \mathbf{Z}_2)$ if and only if $\partial_1 v = \mathbf{0}$. If we write

$$v = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$$

then this is equivalent to solving the equation

$$\partial_1 \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_1 + \alpha_2 \\ \alpha_2 + \alpha_3 \\ \alpha_3 + \alpha_4 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

which implies that $\alpha_i = 0$ for $i = 1, \ldots 4$. Thus, the only element in $Z_1(G; \mathbf{Z}_2)$ is $\mathbf{0}$ and hence $Z_1(G; \mathbf{Z}_2) = \mathbf{0}$, the 0-dimensional vector space. By definition $B_1(G; \mathbf{Z}_2) = 0$. So

$$H_1(G; \mathbf{Z}_2) := Z_1(G; \mathbf{Z}_2)/B_1(G; \mathbf{Z}_2) = \mathbf{0}/\mathbf{0} = \mathbf{0}.$$

We still need to compute $H_0(G;\mathbf{Z}_2)$. We know that $C_0(G;\mathbf{Z}_2) \approx \mathbf{Z}_2^5$. Furthermore, since $\ker \partial_1 = \mathbf{0}$, $\partial_1(C_1(G;\mathbf{Z}_2)) \approx \mathbf{Z}_2^4$. Thus,

$$H_0(G;\mathbf{Z}_2) := Z_0(G;\mathbf{Z}_2)/B_0(G;\mathbf{Z}_2) \approx \mathbf{Z}_2^5/\mathbf{Z}_2^4 \approx \mathbf{Z}_2.$$

Example 2.16 Let G be the graph representing Γ^1. Then

$$\begin{aligned} G_0 &= \{a,b,c,d\} \\ G_1 &= \{[a,b],[b,c][c,d],[d,a]\} \end{aligned}$$

Since G_0 and G_1 are the bases for the 0-chains and 1-chains we have that

$$\begin{aligned} C_0(G;\mathbf{Z}_2) &\approx \mathbf{Z}_2^4 \\ C_1(G;\mathbf{Z}_2) &\approx \mathbf{Z}_2^4. \end{aligned}$$

To do the computations it is convenient to use a column vector notation. So let

$$\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix},\ \mathbf{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

and

$$[\mathbf{a},\mathbf{b}] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ [\mathbf{b},\mathbf{c}] = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ [\mathbf{c},\mathbf{d}] = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix},\ [\mathbf{d},\mathbf{e}] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

With this convention, ∂_1 becomes the 4×4 matrix

$$\partial_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Now consider $Z_1(G;\mathbf{Z}_2) := \ker \partial_1$. So we need to solve the equation

$$\partial_1 \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} \alpha_1 + \alpha_4 \\ \alpha_1 + \alpha_2 \\ \alpha_2 + \alpha_3 \\ \alpha_3 + \alpha_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Observe that since we are using $\mathbf{Z}_2$ coefficients,

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$$

is a solution. In particular, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$ is a non-trivial solution. Thus, $Z_1(G; \mathbf{Z}_2) \approx \mathbf{Z}_2$. By definition $B_1(G; \mathbf{Z}_2) = \mathbf{0}$. So

$$H_1(G; \mathbf{Z}_2) := Z_1(G; \mathbf{Z}_2)/B_1(G; \mathbf{Z}_2) \approx \mathbf{Z}_2.$$

We still need to compute $H_0(G; \mathbf{Z}_2)$. We know that $C_0(G; \mathbf{Z}_2) \approx \mathbf{Z}_2^4$. Furthermore, since $\ker \partial_1 \approx \mathbf{Z}_2$, $\partial_1(C_1(G; \mathbf{Z}_2)) \approx \mathbf{Z}_2^3$. Thus,

$$H_0(G; \mathbf{Z}_2) := Z_0(G; \mathbf{Z}_2)/B_0(G; \mathbf{Z}_2) \approx \mathbf{Z}_2^4/\mathbf{Z}_2^3 \approx \mathbf{Z}_2.$$

Exercises

2.6 Compute $H_*(G; \mathbf{Z}_2)$ where G is a graph for the following figures:

2.7 Prove that if G_1 and G_2 are disjoint graphs, then

$$H_*(G_1 \cup G_2; \mathbf{Z}_2) \approx H_*(G_1; \mathbf{Z}_2) \oplus H_*(G_2; \mathbf{Z}_2).$$

2.8 * Let G be a graph with a free vertex v^+ that lies on edge e. Let G' be the graph obtained by removing e and v^+ from G. Prove that

$$H_*(G; \mathbf{Z}_2) \approx H_*(G'; \mathbf{Z}_2).$$

2.9 * Prove that if T is a tree, then

$$H_0(T; \mathbf{Z}_2) \approx \mathbf{Z}_2, \quad H_1(T; \mathbf{Z}_2) = 0.$$

In light of Proposition 2.12 this is suppose to help you believe that homology might be a topological invariant. Of course this is not a proof of that.

2.2 Approximation of Maps

The purpose of the last section was to motivate the homology of topological spaces. The process which we adopted can be summarized as follows. We began with a topological space $G \subset \mathbf{R}^3$ which for the sake of simplicity we took to be a graph. We then observed that graphs could be represented combinatorially and finally we used this combinatorics to produce an algebraic quantity $H_*(G)$ which we call the homology of G. Now assume that we have two topological spaces X and Y and a continous map $f : X \to Y$. In this section we will mimic this process in such a way that we obtain a linear map $f_* : H_*(X) \to H_*(Y)$.

2.2.1 Approximating Maps on an Interval

To keep the technicalities to an absolute minimum, we begin our discussion with maps of the form $f : [a, b] \to [c, d]$. We do this for two reasons. First, each interval can be represented by a graph and so using the types of arguments employed in the previous section we can compute the homology. Second, we can actually draw pictures of the functions. This latter point is to help us develop our intuition, in practice we will want to apply these ideas to problems where it is not feasible to visualize the maps, either because the map is too complicated or because the dimension is too high.

With this in mind let $X = [-2, 2] \subset \mathbf{R}$, $Y = [-2, 4] \subset \mathbf{R}$ and let $f : X \to Y$ be defined by $f(x) = (x - \sqrt{2})(x + 1)$. Thus, we have two topological spaces and a continous map between them. To treat these combinatorially we think of the spaces as abstract graphs. As was indicated in Example 2.13 there is no unique representation of these intervals as graphs, so we have the freedom to choose. Let us begin with the representations given in Table 2.6

The question we now face is how do we go from the continuous map f, to a map which takes the combinatorial data $\mathcal{E}(X)$ and $\mathcal{V}(X)$ to $\mathcal{E}(Y)$ and $\mathcal{V}(Y)$? Three issues need to be considered in constructing the map.

1. We want to make sure that after we have completed all our calculations we have the correct answer.

2. Because we want to use the computer we can only do a finite number of evaluations of the function f.

Edges of X	$\mathcal{E}(X) \;=\; \{[-2,-1],\ [-1,0],\ [0,1],\ [1,2]\}$
Vertices of X	$\mathcal{V}(X) \;=\; \{[-2],\ [-1],\ [0],\ [1],\ [2]\}$
Edges of Y	$\mathcal{E}(X) \;=\; \{[-2,-1],\ [-1,0],\ [0,1],\ [1,2],\ [2,3],\ [3,4]\}$
Vertices of Y	$\mathcal{V}(X) \;=\; \{[-2],\ [-1],\ [0],\ [1],\ [2],\ [3],\ [4]\}$

Table 2.6: Edges and Vertices for the graphs of $X = [-2, 2]$ and $Y = [-2, 4]$.

3. In the end we are only interested in computing an object $f_* : H_*(X) \to H_*(Y)$. We have stated that homology is a homotopical invariant, so we should not need to have a very precise understanding of f but rather an approximation up to homotopy.

Let us begin with this last point. In Figure 2.6 we show two functions f and g which are homotopic. Recall from Exercise 2.2, that any two functions from one interval to another are homotopic. We include the figure to emphasize the fact that two homotopic functions can behave very differently locally, e.g. the derivatives of these functions are very different. If we move to more complicated spaces, then it is not true that all functions are homotopic (this is a non-trivial result). However, as will be made clear later for reasonable spaces if for every $x \in X$, the distance between $f(x)$ and $g(x)$ is sufficiently small, then f and g are homotopic.

The second point was that we only wanted to do a finite number of calculations. Since we want to develop algorithms that will allows us to do these computations, we want to have a systematic method for choosing which calculations to perform. There are, of course, many different approaches that we could pursue, however we will adopt the following. Observe that

$$X = [-2, -1] \cup [-1, 0] \cup [0, 1] \cup [1, 2].$$

Therefore, we will do our computations in terms of edges. From the combi-

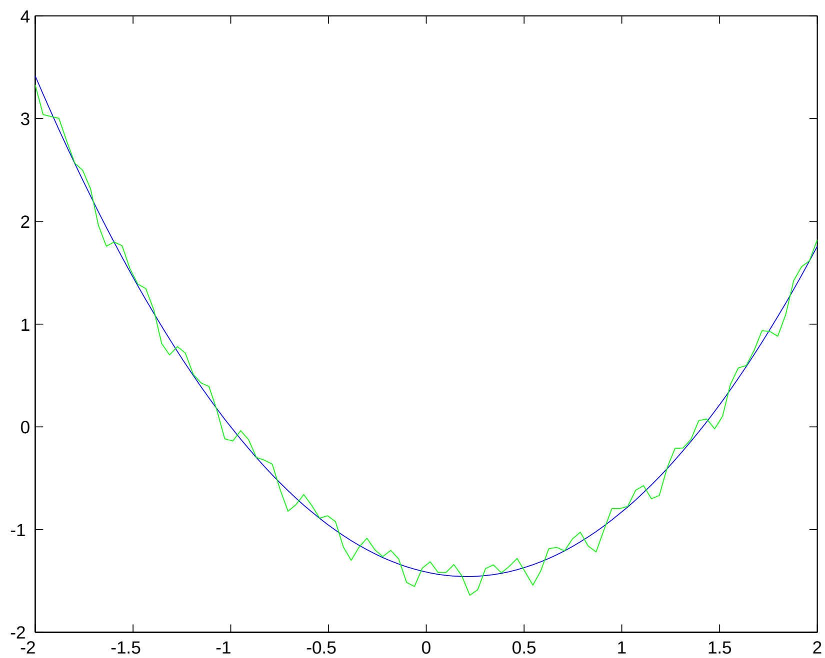


Figure 2.6: The function $f(x) = (x - \sqrt{2})(x + 1)$ and a homotopic function g.

natorial point of view, this suggests trying to map edges to edges. Since $f(-2) = 3.41421356\ldots$, $f(-1) = 0$, and f is monotone over the edge $[-2, -1]$, it is clear that

$$f([-2, -1]) \subset [0, 4] = [0, 1] \cup [1, 2] \cup [2, 3] \cup [3, 4].$$

Thus we could think of defining a map that takes the edge $[0, 1]$ to the collection of edges $\{[0, 1],\ [1, 2],\ [2, 3],\ [3, 4]\}$. Of course, this strategy of looking at the endpoints does not work for the edge $[0, 1]$ since f is not monotone here.

To deal with this problem let us go back to calculus to develop a method for getting good estimates on the function.

Theorem 2.17 [Taylor's Theorem] *Let f be a function that is n-times differentiable. Then,*

$$f(x) = f(a) + \sum_{i=1}^{n-1} \frac{f^{(i)}(a)}{i!}(x - a)^i \ + \int_a^x \frac{(x - t)^{n-1}}{(n - 1)!} f^{(n)}(t)\, dt$$

To apply this to our problem observe that $f''(x) = 2$ and so we can obtain the inequality

$$\begin{aligned} f(x) &= f(a) + f'(a)(x-a) + \int_a^x (x-t)2\,dt \\ &= f(a) + f'(a)(x-a) + (x-a)^2 \\ f(x) - f(a) &= f'(a)(x-a) + (x-a)^2 \\ |f(x) - f(a)| &\leq |f'(a)||x-a| + (x-a)^2. \end{aligned}$$

For our purposes it is more convenient to write this last inequality as

$$f(a) - |f'(a)||x-a| - (x-a)^2 < f(x) < f(a) + |f'(a)||x-a| + (x-a)^2. \quad (2.2)$$

Returning to the interval $[0,1]$, let $a = \frac{1}{2}$. Then, for any $x \in [0,1]$ the inequality (2.2) implies that

$$\begin{aligned} f(\frac{1}{2}) - |f'(\frac{1}{2})||x - \frac{1}{2}| - (x - \frac{1}{2})^2 &< f(x) < f(\frac{1}{2}) + |f'(\frac{1}{2})||x - \frac{1}{2}| + (x - \frac{1}{2})^2 \\ -1.3713 - 0.5858 \cdot \frac{1}{2} - 0.25 &< f(x) < -1.3713 + 0.5858 \cdot \frac{1}{2} + 0.25 \\ -1.914 &< f(x) < -0.8284 \end{aligned}$$

We can use this inequality to determine where to map the edge $[0,1]$:

$$f([0,1]) \subset [-2,0] = [-2,-1] \cup [-1,0].$$

In Table 2.7 we have applied the relationship (2.2) to the midpoints of all the intervals in X and from that derived the mappings of the edges. Observe that since each interval has length 1 (2.2) reduces to

$$f(a) - 0.5|f'(a)| - 0.25 < f(x) < f(a) + 0.5|f'(a)| + 0.25.$$

We can think of Table 2.7 as defining a map from edges to sets of edges. For example

$$[0,1] \mapsto [-2,-1] \cup [-1,0]$$

and we can represent this graphically by means of the rectangle

$$[0,1] \times [-2,0] \subset [-2,2] \times [-2,4] = X \times Y.$$

Doing this for all the edges in the domain gives the the region shown in Figure 2.7. Observe that the graph of $f : X \to Y$ is a subset of this region and therefore we can think of the region as representing an outerbound on the function f.

We would like to make clearer this idea of mapping edges to sets of edges.

Edge of X	Bounds on the image	Image Edges
$[-2,-1]$	$-0.5 < f(x) < 3.5$	$\{[-1,0],\ [0,1],\ [1,2],\ [2,3],\ [3,4]\}$
$[-1,0]$	$-1.92 < f(x) < 0.1$	$\{[-2,-1],\ [-1,0],\ [0,1]\}$
$[0,1]$	$-1.92 < f(x) < -0.83$	$\{[-2,-1],\ [-1,0]\}$
$[1,2]$	$-1.33 < f(x) < 1.76$	$\{[-2,-1],\ [-1,0],\ [0,1],\ [1,2]\}$

Table 2.7: Edges and Vertices for the graphs of $X = [-2,2]$ and $Y = [-2,4]$.

Definition 2.18 Let X and Y be sets. A *multivalued map* $\mathcal{F} : X \overrightarrow{\to} Y$ is a function from X to subsets of Y, i.e. for every $x \in X$, $\mathcal{F}(x) \subset Y$.

Using this language we can view our edge mapping as a multivalued map $\mathcal{F} : [-2,2] \overrightarrow{\to} [-2,4]$ defined by

$$\mathcal{F}(x) := \begin{cases} [-1,4] & \text{if } x = -2 \\ [-1,4] & \text{if } x \in (-2,-1) \\ [-1,1] & \text{if } x = -1 \\ [-2,1] & \text{if } x \in (-1,0) \\ [-2,0] & \text{if } x = 0 \\ [-2,0] & \text{if } x \in (0,1) \\ [-2,0] & \text{if } x = 1 \\ [-2,2] & \text{if } x \in (1,2) \\ [-2,2] & \text{if } x = 2 \end{cases}$$

There are three observations to be made at this point. First, observe that $\mathcal{F}$ is defined in terms of the vertices and the interior of the edges, i.e. the edges without its endpoints. Since we will used this idea later let us introduce some notation and a definition.

Definition 2.19 Let e be and edge with endpoints $v^{\pm}$. The corresponding open edge is

$$\overset{\circ}{e} := e \setminus \{v^{\pm}\}.$$

The second observation, is that we used the edges to define the images of the vertices. In particular, we used the formula that if v is a vertex that lies in edge e_1 and e_2, then

$$\mathcal{F}(v) = \mathcal{F}(\overset{\circ}{e}_1) \cap \mathcal{F}(\overset{\circ}{e}_2). \tag{2.3}$$

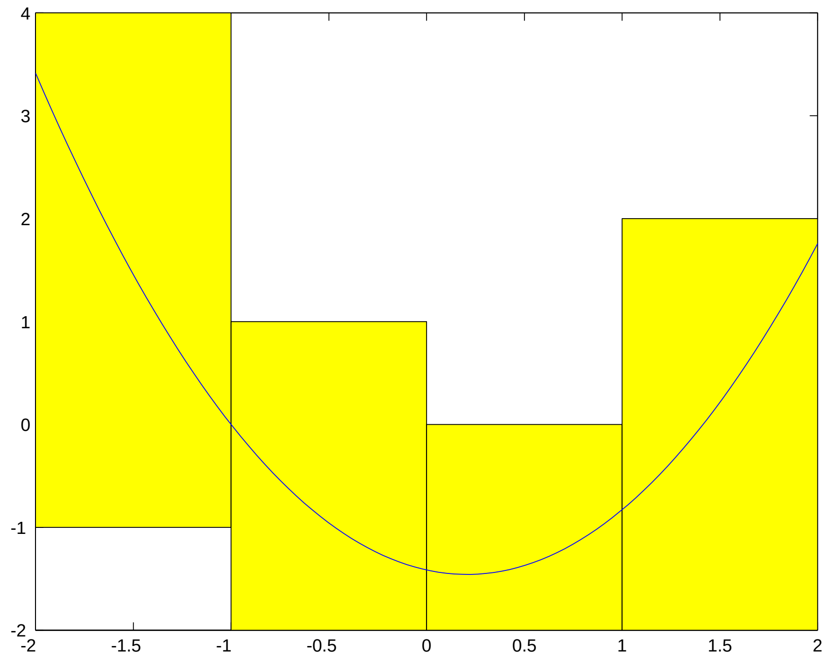


Figure 2.7: The graph of the map produced by sending edges to sets of edges. Observe that the graph of the function $f(x) = (x - \sqrt{2})(x+1)$ lies inside the graph of this edge map.

The final point is that even though $\mathcal{F} : X \rightrightarrows Y$ is a map that is defined on uncountably many points, it is completely characterized by its values on the four edges that make up X. Thus, $\mathcal{F}$ is a finitely representable map. This is important because it means that it can be stored and manipulated by the computer.

The multivalued map $\mathcal{F}$ that we constructed above is fairly coarse. If we want a better approximation, then one approach is to use finer graphs to describe X and Y. For example let us write

$$X = \bigcup_{i=0}^{8} [-2 + \frac{i}{2}, -1.5 + \frac{i}{2}] \quad \text{and} \quad Y = \bigcup_{i=0}^{12} [-2 + \frac{i}{2}, -1.5 + \frac{i}{2}]$$

Using the same approximation (2.2) as above we obtain the data described in Table 2.8. The graph of the corresponding multivalued map is shown in Figure 2.8. Observe that this is a better approximation to the function than

Edge of X	Bounds on the image	Image Edges
$[-2,-1.5]$	$1.30 < f(x) < 3.42$	$\{[1,1.5],\ [1.5,2],\ [2,2.5],\ [2.5,3],\ [3,3.5]\}$
$[-1.5,-1]$	$-0.12 < f(x) < 1.46$	$\{[-0.5,0],\ [0,0.5],\ [0.5,1],\ [1,1.5]\}$
$[-1,-0.5]$	$-1.53 < f(x) < 0.01$	$\{[-2,-1.5],\ [-1.5,-1],\ [-1,0],\ [0,1.5]\}$
$[-0.5,0]$	$-1.52 < f(x) < -0.95$	$\{[-2,-1.5],\ [-1.5,-1],\ [-1,-0.5]\}$
$[0,0.5]$	$-1.52 < f(x) < -1.37$	$\{[1,1.5],\ [1.5,2],\ [2,2.5],\ [2.5,3],\ [3,3.5]\}$
$[0.5,1]$	$-1.49 < f(x) < -0.83$	$\{[-0.5,0],\ [0,0.5],\ [0.5,1],\ [1,1.5]\}$
$[1,1.5]$	$-0.95 < f(x) < 0.22$	$\{[-2,-1.5],\ [-1.5,-1],\ [-1,0],\ [0,1.5]\}$
$[1.5,2]$	$0.08 < f(x) < 1.76$	$\{[0,0.5],\ [0.5,1],\ [1,1.5],\ [1.5,2]\}$

Table 2.8: Edges and Vertices for the graphs of $X = [-2,2]$ and $Y = [-2,4]$.

what was obtained with intervals of unit length. In fact, one can obtain as good an approximation as one likes by choosing the edge lengths sufficiently small. In Figure 2.9 one sees the graph of the multivalued map when the lengths of the edges is 0.1.

2.2.2 Constructing Chain Maps

In the previous section we considered the problem of approximating maps from one interval to another. Of course the goal of this course is to use such an approximation to reduce the analytic problem to an algebraic problem. So in this section we begin with the question: How can we use the information in Figure 2.7 to construct a map $f_* : H_*([-2,2]) \to H_*([-2,4])$?

Let us begin by emphasizing that this is not an obvious task. Recall that homology is by definition a quotient of cycles by boundaries, which in turn belong to subspace of the set of chains. Thus, it seems that the first place to begin is on the level of chains. Furthermore, in order to be able to use our intuition from linear algebra we will consider homology with $\mathbf{Z}_2$ coefficients. In keeping with Figure 2.7 we will consider $[-2,2]$ and $[-2,4]$ to be the graphs made up of the edges with vertices having integer values.

In defining the approximation, we started on the level of edges. In trying to generate the algebra we will start with the vertices. Recall that $C_0([-2,2])$ is the vector space over $\mathbf{Z}_2$ whose basis is given by the set of

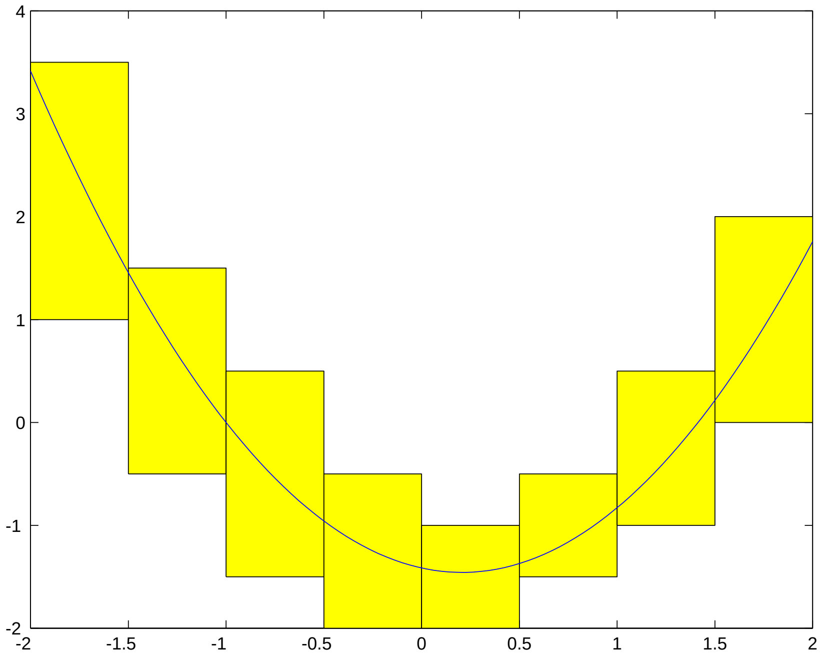


Figure 2.8: The graph of the multivalued approximation to $f(x) = (x - \sqrt{2})(x+1)$ with edges of length 0.5.

vertices $\{\{-2\},\{-1\},\{0\},\{1\},\{2\}\}$ and that $C_0([-2,4])$ is generated by the vertices $\{\{-2\},\{-1\},\{0\},\{1\},\{2\},\{3\},\{4\}\}$. We will begin by defining a linear map

$$f_{\#0} : C_0([-2,2]) \to C_0([-2,4]).$$

Of course, to define a linear map it is sufficient to define how it acts on the basis elements. For lack of a better idea lets define $f_{\#0}(v) := \max \mathcal{F}(v)$. If we order the basis elements of $C_0([-2,2])$ and $C_0([-2,4])$ according to the obvious ordering of the vertices then

$$f_{\#0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

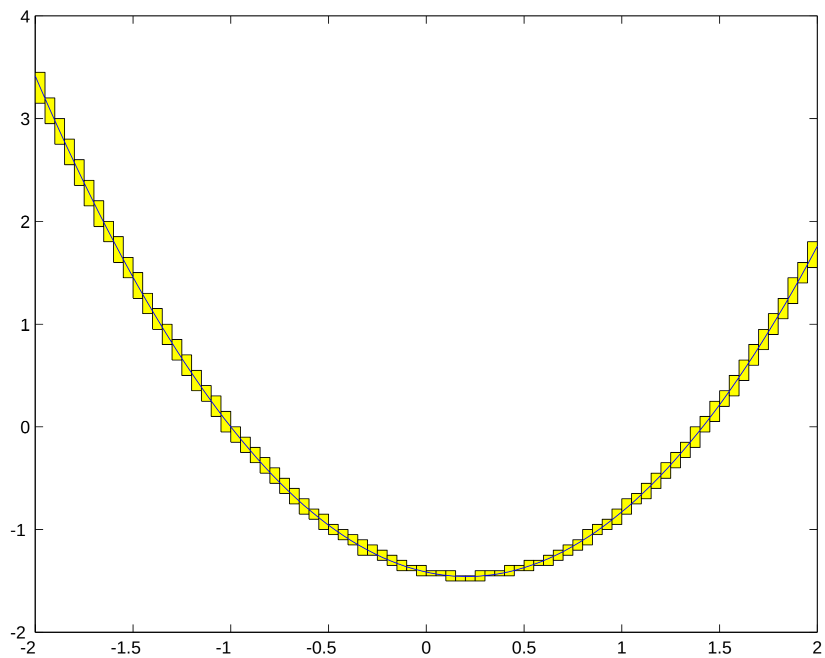


Figure 2.9: The graph of the multivalued approximation to $f(x) = (x - \sqrt{2})(x+1)$ with edges of length 0.1.

We have now defined a linear map between the 0-chains of the two spaces. The next step is to "lift" the definition of $f_{\#0}$ to obtain a linear map $f_{\#1} : C_1([-2,2]) \to C_1([-2,4])$. Of course the basis of these spaces are given by the intervals. So consider the interval $[-2,-1] \subset [-2,2]$. How should we define $f_{\#1}([-2,-1])$? We know that $f_{\#0}(\{-2\}) = \{4\}$ and $f_{\#0}(\{-1\}) = \{1\}$ so it seems reasonable to define $f_{\#1}([-2,-1]) = [1,2] + [2,3] + [3,4]$. Similarly, $f_{\#1}([-1,-0]) = [0,1]$. But what about $f_{\#1}([0,1])$ where $f_{\#1}(\{0\}) = f_{\#1}(\{1\}) = \{0\}$? Since the two endpoints are the same, let us just declare that $f_{\#1}([0,1])$ does not map to any intervals, i.e. that $f_{\#1}([0,1]) = 0$. Again ordering the intervals of $[-2,2]$ and $[-2,4$ in the obvious way an apply these rules to each of the intervals we obtain the following matrix

$$f_{\#1} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

In figuring out how to define $f_{\#1}$ we used the phrase "it seems reasonable to define" but this does not mean we should not define it a different way. Given our choice for $f_{\#0}$ are there any restrictions on the way we define $f_{\#1}$? The answer is an emphatic yes. Recall that our goal is to use $f_{\#}$ to obtain a map on homology, i.e. $f_* : H_*([-2,2]; \mathbf{Z}_2) \to H_*([-2,4]; \mathbf{Z}_2)$. Thus, our real interest is in cycles rather than arbitrary chains. After all elements of homology are equivalence classes of cycles which are very special chains.

Let c be a cycle, by definition $\partial c = 0$. Now if $f_{\#}$ is supposed to generate a map on homology, it is important that $f_{\#}$ map cycles to cycles. Thus $f_{\#}(c)$ should be a cycle which again by definition means that $\partial f_{\#}(c) = 0$. Notice that since $f_{\#}$ is a linear map this leads to the following interesting equation

$$\partial f_{\#}(c) = 0 = f_{\#}(\partial c).$$

Again, let c be a cycle, but this time assume that it is also a boundary, i.e. $c = \partial b$ for some chain b. This means that in homology c is in the equivalence class of 0, i.e. in homology $[c] = 0$. But, we want the homology map f_* to be linear, so $f_*(0) = 0$ and hence $f_*([c]) = 0$.

What does this mean on the level of cycles. If $f_{\#}$ takes cycles to cycles, then $f_{\#}(c)$ is a cycle. But as we just noted we want $f_*([c]) = 0$ and so the simplest condition to require is that $f_{\#}(c)$ be a boundary which means that in homology $f_{\#}(c)$ is in the same equivalence class as 0. How can this be guarenteed? In other words, what kind of constraint on $f_{\#}$ will guarentee that cycles which are boundaries go to boundaries?

To answer this lets repeat what we have said. c is a boundary so we can write $c = \partial b$ for some chain b. Thus $f_{\#}(c) = f_{\#}(\partial b)$. But we want $f_{\#}(c)$ to be the boundary of some chain. What chain? The only one we have at our disposal is b, so the easiest constraint is to ask that $f_{\#}(c) = \partial f_{\#}(b)$. Notice that once again we are led to the interesting equation

$$\partial f_{\#}(b) = f_{\#}(c) = f_{\#}(\partial b).$$

As one might have guessed from the time spent discussing it this relationship is extremely important and linear maps on the set of chains that satisfy

$$\partial f_{\#} = f_{\#}\partial$$

are called *chain maps.*

Let us now check whether the linear maps $f_{\#0}$ and $f_{\#1}$ are chain maps, i.e. that they satisfy the relation $\partial f_{\#} = f_{\#}\partial$. We were sloppy about the subscripts in our discussion above so now we need to be a bit more careful.

First we have two sets of boundary maps

$$\begin{aligned}
\partial_2^{[-2,2]} : C_2([-2,2];\mathbf{Z}_2) &\rightarrow C_1([-2,2];\mathbf{Z}_2)\\
\partial_1^{[-2,2]} : C_1([-2,2];\mathbf{Z}_2) &\rightarrow C_0([-2,2];\mathbf{Z}_2)\\
\partial_0^{[-2,2]} : C_0([-2,2];\mathbf{Z}_2) &\rightarrow 0
\end{aligned}$$

and

$$\begin{aligned}
\partial_2^{[-2,4]} : C_2([-2,2];\mathbf{Z}_2) &\rightarrow C_1([-2,2];\mathbf{Z}_2)\\
\partial_1^{[-2,4]} : C_1([-2,2];\mathbf{Z}_2) &\rightarrow C_0([-2,2];\mathbf{Z}_2)\\
\partial_0^{[-2,4]} : C_0([-2,2];\mathbf{Z}_2) &\rightarrow 0.
\end{aligned}$$

Using this notation we see that the relation $\partial f_{\#} = f_{\#}\partial$ should be written as

$$f_{\#0}\partial_1^{[-2,2]} = \partial_1^{[-2,4]} f_{\#1}. \tag{2.4}$$

In the matrix form this equation becomes

$$\begin{bmatrix} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&1&1&0\\ 0&1&0&0&0\\ 0&0&0&0&1\\ 0&0&0&0&0\\ 1&0&0&0&0 \end{bmatrix}
\begin{bmatrix} 1&0&0&0\\ 1&1&0&0\\ 0&1&1&0\\ 0&0&1&1\\ 0&0&0&1 \end{bmatrix}
=
\begin{bmatrix} 1&0&0&0&0&0\\ 1&1&0&0&0&0\\ 0&1&1&0&0&0\\ 0&0&1&1&0&0\\ 0&0&0&1&1&0\\ 0&0&0&0&1&1\\ 0&0&0&0&0&1 \end{bmatrix}
\begin{bmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&1&0&1\\ 1&0&0&1\\ 1&0&0&0\\ 1&0&0&0 \end{bmatrix}$$

and it is left to the reader to check that this is an equality. Thus the maps $f_{\#0}$ and $f_{\#1}$ are chain maps.

Recall that the constraint of being a chain map was imposed in order to guarentee that $f_{\#}$ would generate a map on homology, $f_* : H_*([-2,2];\mathbf{Z}_2) \rightarrow H_*([-2,4];\mathbf{Z}_2)$. From Section 2.1.3 we know that

$$H_0([-2,2];\mathbf{Z}_2) \cong \mathbf{Z}_2 \quad \text{and} \quad H_1([-2,2];\mathbf{Z}_2) = 0$$

and similarly

$$H_0([-2,4];\mathbf{Z}_2) \cong \mathbf{Z}_2 \quad \text{and} \quad H_1([-2,4];\mathbf{Z}_2) = 0$$

Thus, the only interesting map is

$$f_0 : H_0([-2,2];\mathbf{Z}_2) \to H_0([-2,4];\mathbf{Z}_2).$$

How should we define the map f_0? By definition the elements of $H_0([-2,2];\mathbf{Z}_2)$ are equivalence classes of the cycles $Z_0([-2,2];\mathbf{Z}_2)$. But $\partial_0^{[-2,2]} = 0$ so any 0-chain is a 0-cycle, i.e. $C_0([-2,2];\mathbf{Z}_2) = Z_0([-2,2];\mathbf{Z}_2)$. By looking at the matrix which represents $\partial_1^{[-2,2]}$ it is possible to check that the vertex $\{-2\}$ is not in the image of $\partial_1^{[-2,2]}$, i.e. there is no 1-chain w such that $\partial_1^{[-2,2]} w = \{-2\}$.

Thus, we can take the equivalence class which contains the vertex $\{-2\}$ as a generator for $H_0([-2,2];\mathbf{Z}_2)$. Since the field $\mathbf{Z}_2$ consists of two elements 0 and 1, $H_0([-2,2];\mathbf{Z}_2)$ consists of two vectors which we will write as 0 and 1. Since the equivalence class of the cycle $\{-2\}$ generates $H_0([-2,2];\mathbf{Z}_2)$, we can write

$$[\{-2\}] = 1 \in H_0([-2,2];\mathbf{Z}_2).$$

Returning to our map on homology, to define f_0 we need to determine $f_0(1)$. Of course we want to use the chain map $f_{\#0}$ to do this. 1 is a homology class so $f_{\#0}(1)$ is not defined. However, as was mentioned above $\{-2\}$ is a generator for 0 and $f_{\#0}(\{-2\})$ is a cycle so we can define $f_0(1)$ to be the equivalence class which contains the cycle $f_{\#0}(\{-2\})$, i.e.

$$f_0(1) := [f_{\#0}(\{-2\})] = [\{4\}].$$

The same arguments that led to $[\{-2\}] = 1 \in H_0([-2,2];\mathbf{Z}_2)$, also show that $[\{4\}] = 1 \in H_0([-2,4];\mathbf{Z}_2)$. Thus

$$f_0(1) = 1.$$

In other words, $f_0 : H_0([-2,2];\mathbf{Z}_2) \to H_0([-2,4];\mathbf{Z}_2)$ is the linear map given by multiplication by 1.

This is probably a good place to restate the caveat that we are motivating the ideas behind homology at this point. If you do not find these definitions and constructions completely rigorous that is good, they are not. We will fill in the details later. For the moment we are just trying to get a feel for how we can relate algebraic quantities to topological objects.

Exercises

2.10 Equation (2.4) involves the boundary operators on the level of 1-chains, i.e. $\partial_1^{[-2,2]}$ and $\partial_1^{[-2,4]}$. Discuss how to make sense of this relation as it pertains to the boundary operators on the levels of 0-chains and 2-chains.

2.11 Show that $f_0 : H_0([-2,2];\mathbf{Z}_2) \to H_0([-2,4];\mathbf{Z}_2)$ is well defined.

2.2.3 Maps of the Circle

Up to now we have considered maps from one interval to another. Since the homology of an interval is fairly simple it is not surprising that the maps on homology are equally trivial. So let us consider a space with non-trivial homology such as Γ^1 of Section 2.1.3. Unfortunately, it is rather difficult to draw the graph of a function $f : \Gamma^1 \to \Gamma^1$. In order to draw simple pictures we will think of Γ^1 as the unit interval $[0,1]$ but where the endpoints are identified, i.e. $0 = 1$. In fact we will go a step further and think of Γ^1 as the real line where we make the identification $x = x + 1$ for every $x \in \mathbf{R}$, e.g. $0.5 = 1.5 = 2.5$.

To see how this works in practice consider the function $f : [0,1] \to \mathbf{R}$ given by $f(x) = 2x$. We want to think of f as a map from $\Gamma^1 \to \Gamma^1$ and do this via the identification of $y = y + 1$ (see Figure 2.10).

While this process allows us to draw nice figures it must be kept in mind that what we are really interested in is the f as a continuous mapping from Γ^1 to Γ^1. How should we interpret the drawing in Figure 2.10(b)? Observe that as we move across the interval $[0, 0.5]$ the graph of f covers all of $[0,1]$. So going half way around Γ^1 in the domain corresponds to going once around Γ^1 in the image. Thus, going all the way around Γ^1 in the domain results in going twice around Γ^1 in the image. In other words, f wraps Γ^1 around itself twice. In Figure 2.11 we show a variety of different maps and indicate how many times they wrap Γ^1 around itself. Our goal in this section is to see if we can detect the differences in these maps algebraically.

Recall that

$$H_0(\Gamma^1;\mathbf{Z}_2) = \mathbf{Z}_2 \quad \text{and} \quad H_1(\Gamma^1;\mathbf{Z}_2) \cong \mathbf{Z}_2.$$

We will focus our attention on $f_1 : H_1(\Gamma^1;\mathbf{Z}_2) \to H_1(\Gamma^1;\mathbf{Z}_2)$.

Let us begin by considering the map $f : \Gamma^1 \to \Gamma^1$ given by $f(x) = 2x(1-x)$ which is drawn in Figure 2.11(a). The first step is to view Γ^1 as a graph. So we divide it into the intervals $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$, and $[0.75, 1]$. Of

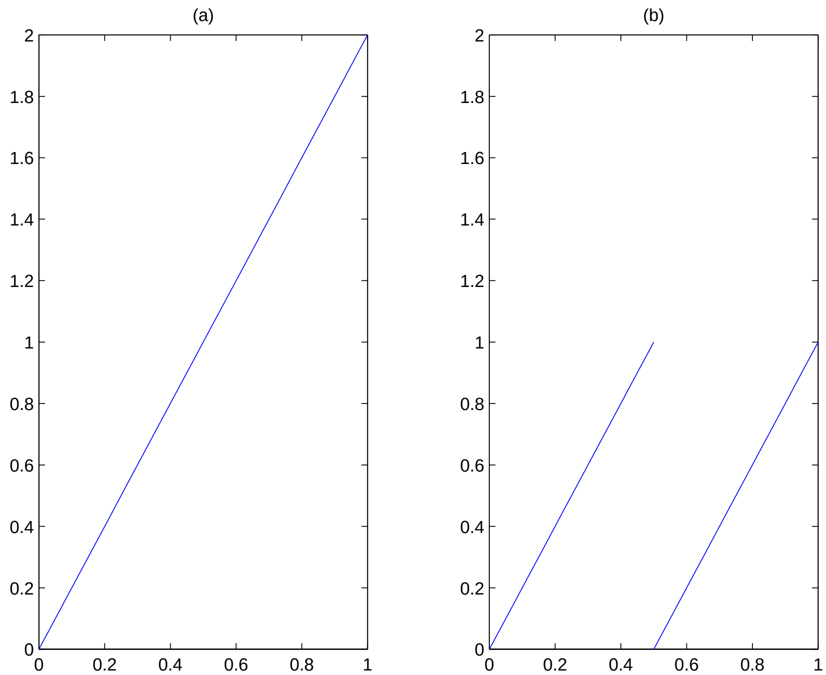


Figure 2.10: Two versions of the graph of $f(x) = 2x$. The left hand drawing indicates $f : [0,1] \to \mathbf{R}$. In the right hand drawing we have made the identification of $y = y+1$ and so can view $f : [0,1] \to [0,1]$. It is important to keep in mind that on both the x and y axis we make the identification of $0 = 1$. Thus $f(0) = 0 = 1 = f(1)$.

course $0 = 1$ so this decomposition of Γ^1 into an abstract graph is exactly the same as that used in Section 2.1.3.

The next step is to obtain an approximation for f. We do this using the Taylor approximation. Since $f''(x) = 4$ equation (2.2) becomes

$$f(a) - |f'(a)||x-a| - 2(x-a)^2 < f(x) < f(a) + |f'(a)||x-a| + 2(x-a)^2.$$

In Figure 2.12(a) we indicate the resulting multivalued map $\mathcal{F}$ that is an outer approximation for f. Of course, it is easier to understand what is happening if we can view these bounds in the unit square. Using the identification $y = y+1$ we obtain Figure 2.12(b). Recall that we defined the images of vertices via equation (2.3). This implies that

$$\mathcal{F}(\{0.25\}) \quad = \quad \mathcal{F}([0, 0.25]) \cap \mathcal{F}([0.25, 0.5])$$

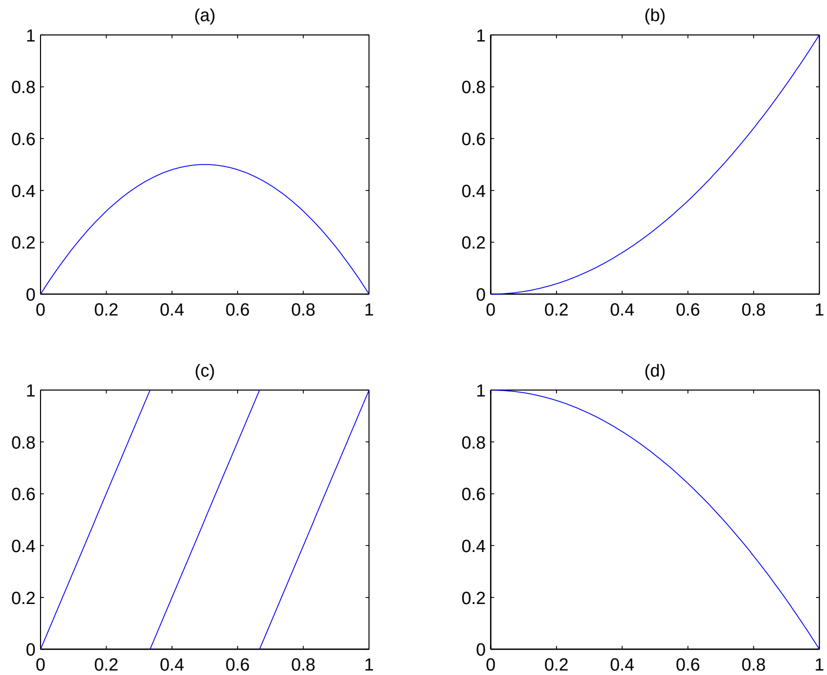


Figure 2.11: Four different maps $f : \Gamma^1 \to \Gamma^1$. How do these different f's wraps Γ^1 around Γ^1? (a) f wraps the interval $[0, 0.5]$ half way around Γ^1 and then over the interval $[0.5, 1]$ f unwraps it. Thus, we could say that the total amount of wrapping is 0. (b) f wraps Γ^1 once around Γ^1. (c) f wraps Γ^1 three times around Γ^1. (d) f wraps Γ^1 once around Γ^1, but in the opposite direction from the example in (b).

$$= \quad [0.25, 0.5] \cup \{0.75\}.$$

This is troubling. What we are saying is that using this procedure the outer approximation of a point is the union of two disjoint sets. It doesn't seems right that a connected set needs to be approximated by a disconnected set. We have two possibilities at this point. One we could redefine our multivalued map $\mathcal{F}$ or two we can try to make a finer approximation of Γ^1.

Since we do not know of a more efficient way of defining $\mathcal{F}$ we will adopt the approach of refining our approximation of Γ^1. This means representing Γ^1 in terms of shorter edges. So let us consider

$$\Gamma^1 = [0, 0.2] \cup [0.2, 0.4] \cup [0.4, 0.6] \cup [0.6, 0.8] \cup [0.8, 1.0]. \tag{2.5}$$

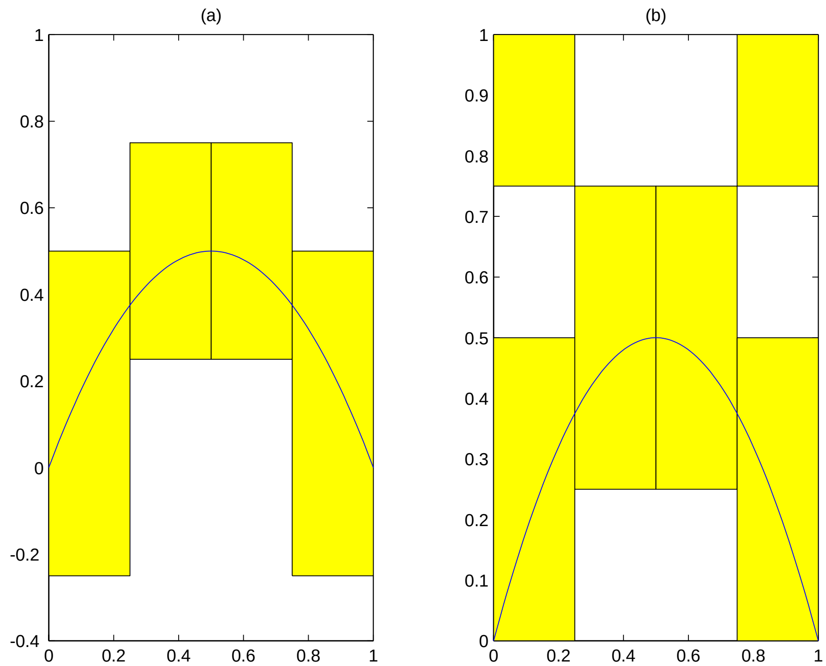


Figure 2.12: The outer approximation for the map $f(x) = 2x(1-x)$.

If we repeat the approximation scheme described above for this representation of Γ^1 we get the outer approximation described in Figure 2.13. Using this approximation $\mathcal{F}(v)$ is an interval for every vertex v.

Using the same rules as before we end up with the multivalued map

$$\mathcal{F}(x) = \begin{cases} [0, 0.4] \cup [0.8, 1] & \text{if } x = 0 \\ [0, 0.4] \cup [0.8, 1] & \text{if } x \in (0, 0.2) \\ [0.2, 0.4] & \text{if } x = 0.2 \\ [0.2, 0.6] & \text{if } x \in (0.2, 0.4) \\ [0.4, 0.6] & \text{if } x = 0.4 \\ [0.4, 0.6] & \text{if } x \in (0.4, 0.6) \\ [0.4, 0.6] & \text{if } x = 0.6 \\ [0.2, 0.6] & \text{if } x \in (0.6, 0.8) \\ [0.2, 0.4] & \text{if } x = 0.8 \\ [0, 0.4] \cup [0.8, 1] & \text{if } x \in (0.8, 1) \\ [0, 0.4] \cup [0.8, 1] & \text{if } x = 1 \end{cases}$$

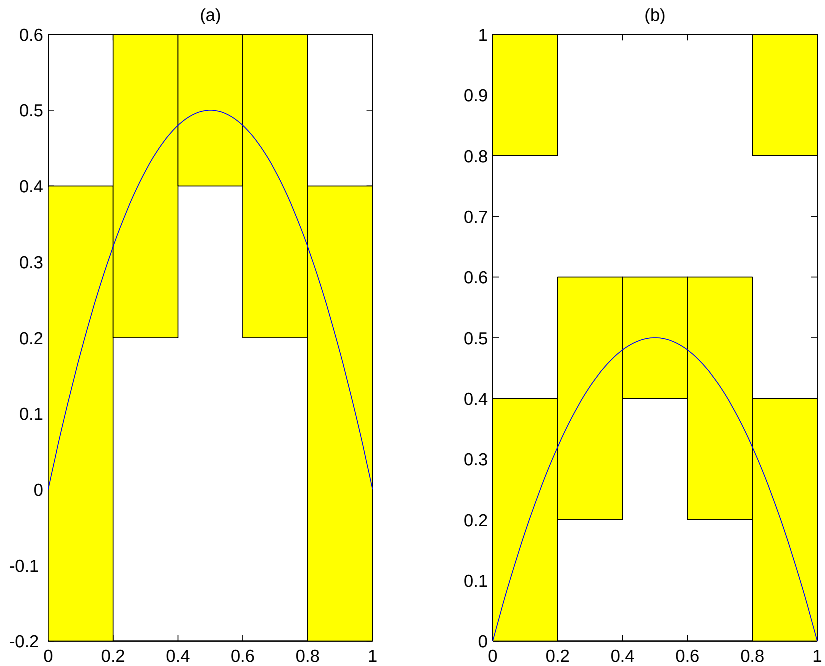


Figure 2.13: The outer approximation for the map $f(x) = 2x(1-x)$ based on edges of length 0.2.

Of course, we have not computed the homology of the graph representing Γ^1 given by (2.5). The reader is encouraged to check that in this case the homology of Γ^1 does not change. However, what should be clear is that it would be nice to have a general theorem that says that if one has the homology of a space does not depend on the approximation used in the computation. Again, we will address these issues later. For the moment we will just assert that the 1-chain given by the sum of all the intervals generates $H_1(\Gamma^1; \mathbf{Z}_2)$, i.e.

$$[[0, 0.2] + [0.2, 0.4] + [0.4, 0.6] + [0.6, 0.8] + [0.8, 1.0]] = 1 \in H_1(\Gamma^1; \mathbf{Z}_2).$$

Having determined the multivalued map $\mathcal{F}$ for this approximation we will construct the chain map $f_{\#0} : C_0(\Gamma^1; \mathbf{Z}_2) \to C_0(\Gamma^1; \mathbf{Z}_2)$ in the same manner as in Section 5.1. Set $f_{\#0}(v) = \max \mathcal{F}(v)$ for any vertex v. Thus for example, $f_{\#0}(\{0\}) = \{1\}$ and $f_{\#0}(\{0.2\}) = \{0.4\}$. Having defined $f_{\#0}$, the construction of $f_{\#1} : C_1(\Gamma^1; \mathbf{Z}_2) \to C_1(\Gamma^1; \mathbf{Z}_2)$ also follows as in Section 5.1.

Using the natural ordering of the intervals which are a basis for $C_1(\Gamma^1; \mathbf{Z}_2)$ we can write

$$f_{\#1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In order to understand the induced map on $H_1(\Gamma^1; \mathbf{Z}_2)$ we need to see how $f_{\#1}$ acts on the generator of $H_1(\Gamma^1; \mathbf{Z}_2)$.

In vector notation as an element of $C_1(\Gamma^1; \mathbf{Z}_2)$, we have

$$[0, 0.2] + [0.2, 0.4] + [0.4, 0.6] + [0.6, 0.8] + [0.8, 1.0] = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Recall that we are using $\mathbf{Z}_2$ coefficients hence $f_{\#1}([0, 0.2]+[0.2, 0.4]+[0.4, 0.6]+[0.6, 0.8] + [0.8, 1.0])$ is given by

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Therefore, $f_1 : H_1(\Gamma^1; \mathbf{Z}_2) \to H_1(\Gamma^1; \mathbf{Z}_2)$ is given by multiplication by 0. Notice that this corresponds to the number of times that f wraps Γ^1 around its.

Lets do this again for the map $f(x) = x^2$. We proceed exactly as before. Again we need estimates on the approximation. Since $f''(x) = 2$ we can use equation (2.2). Figure 2.14 shows the resulting multivalued map. To obtain an appropriate multivalued map we have chosen to represent Γ^1 as follows

$$\begin{aligned} \Gamma^1 &= [0, 0.125] \cup [0.125, 0.25] \cup [0.25, 0.375] \cup [0.375, 0.5] \\ &\quad \cup [0.5, 0.625] \cup [0.625, 0.75] \cup [0.75, 0.875] \cup [0.875, 1] \end{aligned}$$

As before it is the sum of all these intervals which generates $H_1(\Gamma^1; \mathbf{Z}_2)$.

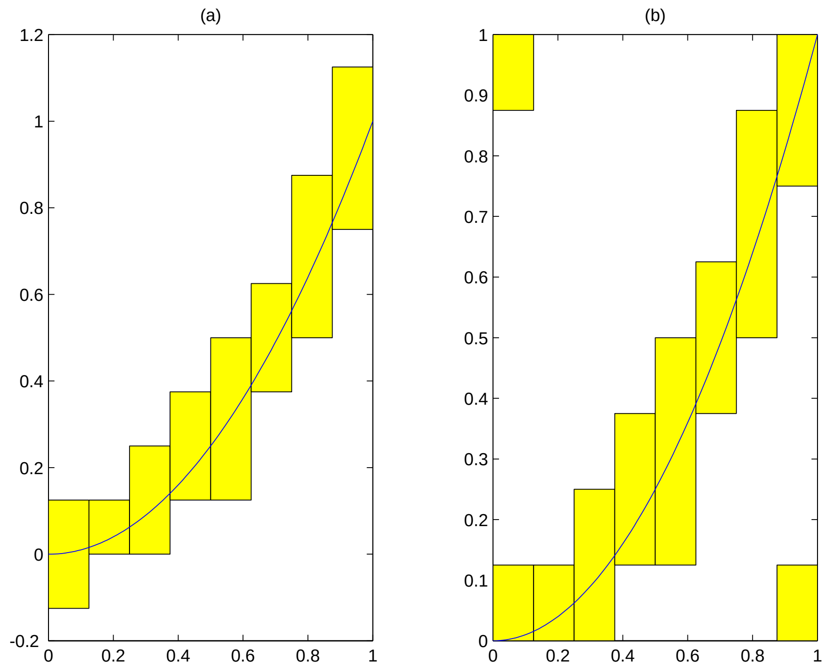


Figure 2.14: The outer approximation for the map $f(x) = x^2$.

Constructing $f_\#$ as before and using the natural ordering of the intervals which are a basis for $C_1(\Gamma^1; \mathbf{Z}_2)$ we can write

$$f_{\#1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

If we let $f_{\#1}$ act on the 1-chain which generates $H_1(\Gamma^1; \mathbf{Z}_2)$, then we are performing the following computation

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Thus, $f_1 : H_1(\Gamma^1; \mathbf{Z}_2) \to H_1(\Gamma^1; \mathbf{Z}_2)$ is given by $f_1(1) = 1$, i.e. it is multiplication by 1. Observe that this again is the same as the number of times that $f(x) = x^2$ wraps Γ^1 around itself.

We shall do one more example, that of $f(x) = 2x$. Figure 2.15 shows the multivalued map that acts as an outer approximation when the represention of Γ^1 is given by

$$\begin{aligned} \Gamma^1 &= [0, 0.125] \cup [0.125, 0.25] \cup [0.25, 0.375] \cup [0.375, 0.5] \\ &\quad \cup [0.5, 0.625] \cup [0.625, 0.75] \cup [0.75, 0.875] \cup [0.875, 1]. \end{aligned}$$

Following exactly the same process as in the case of $f(x) = x^2$ we obtain

$$f_{\#1} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Again viewing how this acts on the generator of $H_1(\Gamma^1; \mathbf{Z}_2)$ we have

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

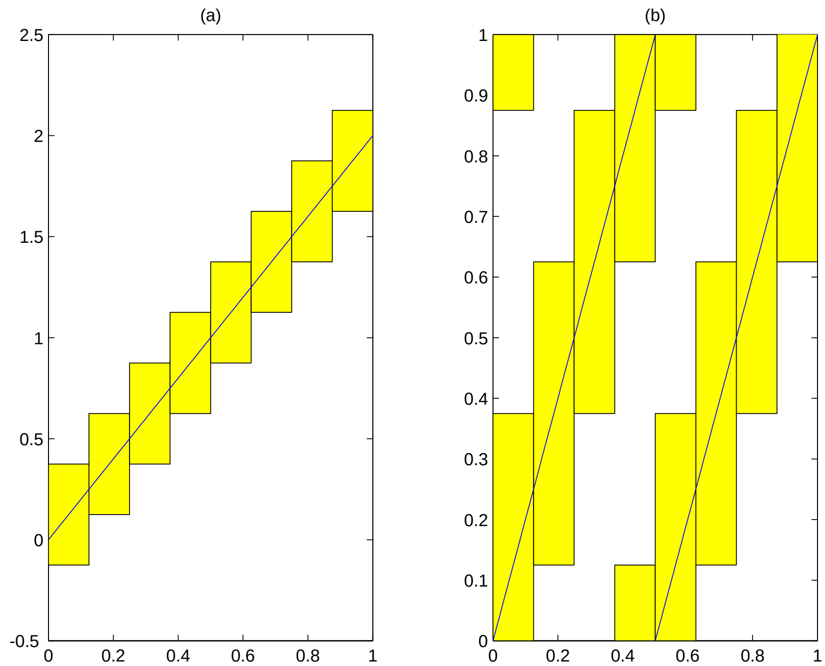


Figure 2.15: The outer approximation for the map $f(x) = 2x$.

In this case we end up with $f_1(1) = 0$, i.e. the homology map on the first level is multiplication by 0. This does not match our geomtrical observation that $f(x) = 2x$ wraps Γ^1 around itself twice. On the other hand, it is clear that $f_1(1) = 0$ precisely because we are using $\mathbf{Z}_2$ coefficients. If we had been using integers we might expect to obtain that f_1 is multiplication by 2. Unfortunately, using the integers as a scalar does not lead to a vector space. With this in mind we will spend the next chapter studying the algebra needed to be able to rigorously do homology over the integers.

Exercises

2.12 Compute $f_1 : H_1(\Gamma^1; \mathbf{Z}_2) \to H_1(\Gamma^1; \mathbf{Z}_2)$ for $f(x) = 3x$.

Chapter 3

Abelian Groups

In Chapter 2 we computed homology groups using linear algebra. As was pointed out in our analysis of maps on the circle it would be nice if we could move beyond linear algebra. In this chapter we will introduce the abelian group theory that lies at the basis of homological algebra.

3.1 Groups

A *binary operation* on a set G is any mapping $q : G \times G \to G$. Rather than writing the operation in this functional form, e.g. q(a,b), one typically uses a notation such as $a + b$ or ab.

Definition 3.1 An *abelian group* is a set G, together with a binary operation $+$ defined on G and satisfying the following four axioms:

1. For all $a, b, c \in G$,

$$a + (b + c) = (a + b) + c \qquad \text{(associativity)}$$

2. There exist an *identity* element $0 \in G$ such that for all $a \in G$

$$a + 0 = 0 + a = a.$$

3. For each $a \in G$ there exists an *inverse* $-a \in G$ such that

$$a + -a = b + -b = 0.$$

4. For all $a, b \in G$,

$$a + b = b + a \qquad \text{(commutativity)}$$

It follows from the axioms (see Exercise 3.1) that the identity element 0 is unique and that given any $a \in G$ its inverse element $-a$ is also unique.

Example 3.2 We denote the set of integers by $\mathbf{Z}$, the rationals by $\mathbf{Q}$, the real numbers by $\mathbf{R}$ and the complex numers by $\mathbf{C}$. All these sets are abelian groups under addition.

Example 3.3 Recall that the set $\mathbf{N}$ of *natural numbers* is the same as the set of nonnegative integers. Addition is a binary operation on N. Furthermore, it is commutative, associative and $0 \in \mathbf{N}$. However, N is not an abelian group since its elements have no inverses under addition. For example, $1 \in \mathbf{N}$, but $-1 \notin \mathbf{N}$.

Example 3.4 The vector space $\mathbf{R}^n$ is an abelian group under coordinatewise addition with the identity element $\mathbf{0} = (0, 0, \ldots, 0)$.

Example 3.5 Given a positive integer n, let $\mathbf{Z}_n := \{0, 1, 2, \ldots n-1\}$ with the addition defined by $(a, b) \to (a + b) \bmod n$, where $(a + b) \bmod n$ is the remainder of $a+b \in \mathbf{Z}$ in the division by n, i.e. the smallest integer $c \geq 0$ such that $a+b-c$ is divisible by n. We shall abandon the $\bmod n$ notation when it will be clear that we mean the addition in $\mathbf{Z}_n$ and not in $\mathbf{Z}$. It is convenient to describe finite groups such as $\mathbf{Z}_n$ by giving their table of addition, here is one for $\mathbf{Z}_3$:

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Definition 3.6 Let G be a group with the binary operation $+$. A nonempty subset $H \subset G$ is a *subgroup* of G if:

1. $0 \in H$,

2. for every $a \in H$ its inverse $-a \in H$,

3. H is *closed* under $+$, i.e. given $a, b \in H$, $a + b \in H$.

Proposition 3.7 *Let H be a subset of G with the property that for any $a, b \in H$, implies $a - b \in H$. Then H is a subgroup of G.*

The proof of this proposition is left as an exercise.

Given $a \in G$ and $n \in \mathbf{Z}$, we use the notation

$$na := \underbrace{a + a + \cdots + a}_{n\ terms}$$

to denote the sum of a with itself n times. If n is a negative integer, then this should be interpreted as the n-fold sum of $-a$.

Definition 3.8 Given a group G, a set of elements $\{g_j\}_{j \in J} \subset G$ *generates* G if any $a \in G$ can be written as a finite sum

$$a = \sum a_j g_j \tag{3.1}$$

where $a_j \in \mathbf{Z}$. By the finiteness of the above sum we mean that $a_j = 0$ for all but finitely many j. The elements of $\{g_j\}_{j \in J}$ are called *generators*. If there is a finite set of generators, then G is a *finitely generated group.*

Observe that the concept of a generating set for a group is similar to that of a spanning set in linear algebra. What makes vector spaces so nice is that they have bases which one can use to uniquely represent any vector in the vector space.

Definition 3.9 A family $\{g_j\}_{j \in J}$ of generators is called a *basis* of G if for any $a \in G$ there is a unique set of integers a_j such that

$$a = \sum a_j g_j\ . \tag{3.2}$$

A group is *free* if it has a basis.

Example 3.10 The group of integers $\mathbf{Z}$ is a free group generated by a single element basis: either $\{1\}$ or $\{-1\}$. Observe that for any $k \in \mathbf{Z} \setminus \{0\}$, $\{k\}$ is a maximal linearly independent set. However, if $k \neq \pm 1$, then $\{k\}$ does not generate $\mathbf{Z}$. This is easily seen by noting that if $\{k\}$ did generate, then there would be an integer n such that $nk = 1$.

Observe that the uniqueness condition implies that a set of generators $\{g_j\}_{j\in J}$ of a group G is a basis of G if and only if it is *linearly independent*, i.e.

$$0 = \sum_{j\in J} a_j g_j \quad \Leftrightarrow \quad a_j = 0 \text{ for all } j \in J.$$

Every vector space has a basis, this is *not* true for groups.

Example 3.11 The group of rational numbers $\mathbf{Q}$ is not free. To see this assume that $\{g_j\}_{j\in J}$ formed a basis for $\mathbf{Q}$. Recall that any element $a \in Q$ can be written in the form $a = p/q$ where p and q are relatively prime integers. Assume that the basis consists of a unique element $g = p/q$. Then $g/2 \in Q$, but it is impossible to solve the equation $ng = g/2$ for some integer n. Therefore, the basis must contain more than one element. In particular, there exists p_1/q_1 and p_2/q_2 in the basis. Now observe that

$$p_1 p_2 = (p_2 q_1) p_1/q_1 = (p_1 q_2) p_2/q_2$$

which violates the uniqueness condition.

Theorem 3.12 *Any two bases of a finitely generated free abelian group G have the same number of elements. This number is called the* rank *of G.*

Proof: The proof is by contradiction. Let $\{g_1, g_2, \ldots, g_n\}$ and $\{h_1, h_2, \ldots, h_m\}$ be two bases of G with $n < m$. Then each element of one basis can be expressed as a linear combiation of the elements of the other basis with integer coefficients. By using matrix notation,

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix} = A \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \text{ and } \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} = B \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix},$$

where $A = (a_{ij})$ and $B = (b_{ij})$ are, respectively, $m \times n$ and $n \times m$ matrices with integer coefficients. Thus

$$\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix} = AB \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix}.$$

By the uniqueness of the expansion, $AB = \mathbf{1}_{m\times m}$ (the identity $m\times m$ matrix) which contradicts that $n < m$: Indeed, the ranks of A and B are at most n, thus the rank of AB is at most n. But the rank of $\mathbf{1}_{m\times m}$ is $m > n$. ∎

Example 3.13 Consider the group $\mathbf{Z}^2$. Set $\mathbf{i} := (1,0)$ and $\mathbf{j} := (0,1)$. Then, $\{\mathbf{i},\mathbf{j}\}$ is a basis for $\mathbf{Z}^2$ and so the rank of $\mathbf{Z}^2$ is 2. Another choice of basis is $\{\mathbf{i},\mathbf{j}-\mathbf{i}\}$. But $\{2\mathbf{i},3\mathbf{j}\}$ is not a basis for $\mathbf{Z}^2$ even though it is a maximal linearly independet set in $\mathbf{Z}^2$. This set is a basis for $2\mathbf{Z}\times 3\mathbf{Z}$ which is a proper subgroup of $\mathbf{Z}^2$ of the same rank 2. We will learn more about product groups in the next section.

A group G generated by a single element a is called *cyclic* and is denoted by $\langle a\rangle$. In general, if $a \in G$ then $\langle a\rangle$ is a cyclic subgroup of G. The *order* of G denoted by $|G|$ is the number of elements of G. Thus $|\mathbf{Z}| = \infty$ and $|\mathbf{Z}_n| = n$. The *order of an element* $a \in G$ denoted by $o(a)$ is the smallest positive integer n such that $na = 0$, if it exists, and ∞ if not. Observe that $|\langle a\rangle| = o(a)$. Of course, a group which has a cyclic element of finite order other than zero cannot be free. The set of all elements in G with finite order is a subgroup called the *torsion subgroup* of G. Observe that a free group is torsion free, i.e. it has no elements of finite order. The converse is not true (see exercises). If a is of infinite order, the cyclic group $\langle a\rangle$ is a free abelian group which may also be denoted by $\mathbf{Z}a$ or by $a\mathbf{Z}$.

Example 3.14 The addition table for $\mathbf{Z}_6$ is as follows:

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Using the table it is easy to check that: 0 has order 1, 1 and 5 have order 6 thus each of them generates the whole group, 2 has order 3 and 3 has order 2. Note the relation between the divisors of 6 and orders of elements of $\mathbf{Z}_6$.

We end this section with the following observation

Lemma 3.15 *Any subgroup of a cyclic group is cyclic.*

Proof: Let G be a cyclic group generated by a and let $H \neq 0$ be a subroup of G. Let k be the smallest positive integer such that $ka \in H$. We show that ka generates H. Clearly, $nka \in H$ for all integers n and we need to show that all elements of H are of that form. Indeed, if not, there exists $h \in H$ of the form $h = (nk + r)a$ where $0 < r < k$. Since $nka \in H$, we get $ra \in H$, which contradicts the minimality of k. ∎

Exercises

3.1 Let G be a group.

(a) Prove that the identity element 0 is unique.

(b) Prove that, given any $a \in G$, the inverse $-a$ of a is unique.

3.2 (a) Write down the tables of addition and multiplication for $\mathbf{Z}_5, \mathbf{Z}_6, \mathbf{Z}_8$.

(b) If $\mathbf{Z}'_n := \mathbf{Z}_n \setminus 0$, show that $\mathbf{Z}'_5$ is a multiplicative group but $\mathbf{Z}'_6, \mathbf{Z}'_8$ are not.

(c) Let now $\mathbf{Z}^*_n := \{k \in \mathbf{Z}_n : k \text{ and } n \text{ are relatively prime}\}$. Show that $\mathbf{Z}^*_n$ is a multiplicative group for any positive integer n.

3.3 (a) Determine the orders of all elements of $\mathbf{Z}_5, \mathbf{Z}_6, \mathbf{Z}_8$

(b) Determine the orders of all elements of $\mathbf{Z}^*_5, \mathbf{Z}^*_6, \mathbf{Z}^*_8$, where $\mathbf{Z}^*_n$ is defined in the preceeding exercise and the order of a in a multiplicative group is the least positive integer n such that $a^n = 1$

3.4 Prove Proposition 3.7

3.5 Let G be an abelian group.

(a) Let $H := \{a \in G \mid o(a) < \infty\} \cup \{0\}$. Prove that H is a subgroup of G.

(a) Show that if G is free then it is torsion-free.

(b) Show that the additive group Q is torsion-free.

(c) Show that if G is finitely generated and torsion free then it is free.

3.2 Products and Sums

Let $G_1, G_2, \ldots, G_n$ be a family of groups and let

$$G = \prod_{i=1}^{n} G_i = G_1 \times G_2 \times \cdots \times G_n \tag{3.3}$$

be the cartesian product of $G_1, G_2, \ldots, G_n$. G becomes a group with the coordinate-wise addition

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

called the *direct product* of $G_1, G_2, \ldots, G_n$. The direct product of n copies of a group G is simply denoted by G^n. There is an obvious analogy between the addition and scalar multiplication in the vector space $\mathbf{R}^n$ and in the direct product of groups: the difference is that in the direct product of groups we are only allowed to multiply by integer scalars from $\mathbf{Z}$.

Let A and B be subgroups of G. We define their *sum* by

$$A + B := \{c \in G : c = a + b \text{ for some } a \in A, b \in B\}\ . \tag{3.4}$$

We say that G is a *direct sum* of A and B and write

$$G := A \oplus B$$

if $G = A + B$ and the decomposition $c = a + b$ of any $c \in G$ is unique. We have the following simple criterion for a direct sum.

Proposition 3.16 *Let G be the sum of its subgroups A and B. Then $G = A \oplus B$ if and only if $A \cap B = \{0\}$.*

Proof: Suppose that $A \cap B = \{0\}$ and that $c = a_1 + b_1 = a_2 + b_2$ are two decompositions of $c \in G$, $a_1, a_2 \in A$ and $b_1, b_2 \in B$. Then $a_1 - a_2 = b_2 - b_1 \in A \cap B = \{0\}$ hence $a_1 = a_2$ and $b_1 = b_2$. Hence the decomposition is unique. Conversely, let $A \cap B \neq \{0\}$ and let $c \in A \cap B$, $c \neq 0$. Then c can be decomposed as $c = a + b$ in at least two ways: by posing $a := c$, $b := 0$ or $a := 0$, $b := c$. ∎

In a similar way one defines the sum and direct sum of any family $G_1, G_2, \ldots, G_n$ of subgroups of a given group G. G is the *direct sum* of

$G_1, G_2, \ldots, G_n$ if every $g \in G$ can be uniquelly written as $a = \sum_{i=1}^n g_i$, where $g_i \in G$ for all $i = 1, 2, \ldots, n$. We write

$$G = \bigoplus_{i=1}^n G_i = G_1 \oplus G_2 \oplus \cdots \oplus G_n \ . \tag{3.5}$$

The criterion analogous to that in Proposition 2.1 for a sum to be a direct sum is

$$G_i \cap G_j = \{0\} \text{ if } i \neq j \ .$$

There is a close relation between direct products and direct sums. Let $G = G_1 \times G_2 \times \cdots \times G_n$. We may identify each G_i with the subgroup

$$j_i G_i := \{0\} \times \cdots \times \{0\} \times \underbrace{G_i}_{\text{i'th place}} \times \{0\} \times \cdots \times \{0\} \ .$$

Then $G = j_1 G_1 \oplus j_2 G_2 \oplus \cdots \oplus j_n G_n$ and, for the simplicity of notation, we may write $G = G_1 \oplus G_2 \oplus \cdots \oplus G_n$. This identification of direct products and sums will become more formal when we talk about isomorphisms of groups in the next section. When infinite families of groups are considered, their direct sum may only be identified with a subgroup of the direct product consisting of sequences which have zeros in all but finitely many places. In this text, however, we shall not need to study infinite sums and products.

Example 3.17

Let G be a free abelian group with a basis $\{g_1, g_2, \ldots, g_n\}$. By the definition of a basis,

$$G = \mathbf{Z}g_1 \oplus \mathbf{Z}g_2 \oplus \cdots \oplus \mathbf{Z}g_n \ .$$

Example 3.18 Consider the group $\mathbf{Z}^2 = \mathbf{Z} \times \mathbf{Z}$. Then $\mathbf{Z}^2 = \mathbf{Zi} \oplus \mathbf{Zj}$, hence we may write $\mathbf{Z}^2 = \mathbf{Z} \oplus \mathbf{Z}$. This decomposition of $\mathbf{Z}^2$ to a direct sum is related to a particular choice of basis $\{\mathbf{i}, \mathbf{j}\}$ called the *canonical basis* of $\mathbf{Z}^2$. As for vector spaces, there may be many bases, and hence, many direct sum decompositions, e.g. $\mathbf{Z}^2 = \mathbf{Zi} \oplus \mathbf{Z}(\mathbf{j} - \mathbf{i})$.

The same consideration applies to $\mathbf{Z}^n$ with the *canonical basis* $\{e^1, e^2, \ldots, e^n\}$, where the coordinates of e^i are given by

$$(e^i)_j = \begin{cases} 1 & \text{if } \ i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Example 3.19 In the group ${\mathbf{Z}_2}^2 = \mathbf{Z}_2 \times \mathbf{Z}_2$ of order 4, all 3 nonzero elements $(0,1)$, $(1,0)$, and $(1,1)$ have order 2. Thus this is not a cyclic group.

Consider the group $\mathbf{Z}_2 \times \mathbf{Z}_3$. Here are the orders of its elements:

$$o(0) = 1,\ o((1,0)) = 2,\ o((0,1)) = o((0,2)) = 3,\ o((1,1)) = o((1,2)) = 6\ .$$

Thus $\mathbf{Z}_2 \times \mathbf{Z}_3$ is cyclic of order 6, generated by $(1,1)$ and by $(1,2)$. The notion of isomorphism introduced in the next section will permit to identify this group with $\mathbf{Z}_6$. The same consideration applies to $\mathbf{Z}_n \times \mathbf{Z}_m$ where n and m are relatively prime (see exercises).

Example 3.17 will now be approached in a different way. Let $S = \{s_1, s_2, \ldots, s_n\}$ be any finite set of objects. What the objects are does not matter. For example, S may be a class of mathematics students, or as is more relevant to this course, a set of edges or vertices in a graph. With the discussion of Chapter 2 in mind, the goal is to give meaning to the sum

$$a_1 s_1 + a_2 s_2 + \cdots + a_n s_n\ ,$$

where $a_1, a_2, \ldots, a_n$ are integers. For this purpose, let us go back to the definition of cartesian product in (3.3). The cartesian product G^n of n copies of G formally is the set of all functions φ from the finite set $\{1, 2, \ldots, n\}$ to G. Thus a point $(x, y, z) \in G^3$ formally is a function $\varphi : \{1, 2, 3\} \to G$ given by $\varphi(1) = x, \varphi(2) = y, \varphi(3) = z$. The group structure is given by pointwise addition: $(\varphi + \psi)(i) := \varphi(i) + \psi(i)$. With the understanding of this we may now define the *free abelian group* $\mathbf{Z}^S$ *generated by* S as the set of all functions $\varphi : S \to \mathbf{Z}$, with the pointwise addition

$$(\varphi + \psi)(s_i) := \varphi(s_i) + \psi(s_i), i = 1, 2, ...n\ .$$

Why is this a free group? Consider the functions $\hat{s}_i : S \to \mathbf{Z},\ i = 1, 2, \ldots, n$ defined by

$$\hat{s}_i(s_j) := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

It is easily verified that $\hat{S} := \{\hat{s_1}, \hat{s_2}, \ldots, \hat{s_n}\}$ is a basis for $\mathbf{Z}^S$. It is called *the canonical basis* and it may be identified with S. Note that if $S = \{1, 2, ...n\}$ we recover $\mathbf{Z}^S = \mathbf{Z}^n$ with the canonical basis e^i defined in Example 3.18.

Exercises

3.6 (a) Let m, n be relatively prime. Show that $\mathbf{Z}_m \oplus \mathbf{Z}_n$ is cyclic of order mn.

(b) Let $G = \mathbf{Z}_{12} \oplus \mathbf{Z}_{36}$. Express G as a direct sum of cyclic groups whose orders are powers of primes.

3.7 (a) Prove that a group of prime order has no proper subgroup.

(b) Prove that if G is a cyclic group and p is a prime dividing $|G|$, then G contains an element of order p.

3.8 Prove the following statements.

(a) If G is a finite multiplicative group and $a \in G$, then $a^{|G|} = 1$.
(**Hint:** Use Proposition 2.4 with $H = \langle a \rangle$)

(b) (Fermat's Little Theorem) If p is a prime and and p does not divide $a \in \mathbf{Z}$ then $a^{p-1} \equiv 1 \pmod{p}$.
(**Hint:** Recall Exercice 2(c) Section 1)

(c) If p is a prime then $b^p \equiv b \pmod{p}$ for all $b \in \mathbf{Z}$.

3.3 Quotients

In Chapter 2, in the setting of vector spaces we defined homology as a quotient of chains by boundaries. We need to extend this idea to the setting of groups.

Let H be a subgroup of G and $a \in G$. The set

$$a + H := \{a + h \; : \; h \in H\}$$

is called a *coset* of H in G. The element a is called its *representative*. Typically a coset will have many different representatives. For example, let $h_0 \in H$, $a \in G$ and $b = a + h_0$, then a and b are representatives for the same coset. The following proposition makes this precise.

Proposition 3.20 *Let H be a subgroup of G and $a, b \in G$. Then*

(a) The cosets $a + H$ and $b + H$ are either equal or disjoint.

(b) $a + H = b + H$ if and only if $b - a \in H$.

Proof: (a) Suppose that $(a + H) \cap (b + H) \neq \emptyset$. Then there exist $h_1, \; h_2$ such that $a + h_1 = b + h_2$. Hence, for any $h \in H$, $b + h = a + h_1 - h_2 + h \in a + H$ so $b + H \subset a + H$. The reverse inclusion holds by the symmetric argument.

(b) Let $a + H = b + H$ and let $h_1, \; h_2$ be as in (a). Then $b - a = h_2 - h_1 \in H$. Conversely, if $b - a \in H$ then $b + 0 = a + (b - a) \in (b + H) \cap (a + H)$, thus the concusion follows from (a). ∎

Writing cosets in the form of $a + H$ is a bit cumbersome, so we shorten it to $[a] := a + H$. Notice that to use this notation it is essential that we know the subgroup H that is being used to form the cosets. We can define a binary operation on the set of cosets by setting

$$[a] + [b] = [a + b]. \tag{3.6}$$

Observe that $[0] + [a] = [0 + a] = [a]$ so $[0]$ acts like an identity element. Furthermore, $[a] + [-a] = [a + -a] = [0]$, so there are inverse elements. It is also easy to check that this operation is associative and commutative. The only serious issue is whether this new operation is well defined, in other words does it depend on which representative we use.

Proposition 3.21 *The formula (3.6) does not dependent on the choice of coset representative used, and therefore, defines a group structure on* $\{a + H\}_{a \in G}$.

Proof: If $a' + H = a + H$ and $b' + H = b + H$ then, by Proposition 3.20, $a' - a \in H$, $b' - b \in H$ and so $(a' + b') - (a + b) = (a' - a) + (b' - b) \in H$. Hence $a' + b' + H = a + b + H$. ■

Definition 3.22 The group of cosets described by Proposition 2.3 is called the *quotient group* of G by H and denoted by G/H.

An alternative way of introducing the quotient group is in terms of an equivalence relation. Define the relation $a \sim b$ if and only if $b - a \in H$. Note that this is an *equivalence relation* in G, i.e.

i) $a \sim a$, for all $a \in G$;

ii) $a \sim b \Leftrightarrow b \sim a$, for all $a, b \in G$;

iii) $a \sim b$ and $b \sim c \Rightarrow a \sim c$, for all $a, b, c \in G$.

The *equivalence class* of $a \in G$ is the set of all $b \in G$ such that $b \sim a$. Thus, by Proposition 3.20 the group of cosets exactly is the group of equivalence classes of $a \in G$.

Proposition 3.23 *Let G be a finite group and H its subgroup. Then each coset $a + H$ has the same number of alements. Consequently,*

$$|G| = |G/H| \cdot |H| .$$

Proof: The first conclusion is an obvious consequence of the cancellation law for the group addition: $a + h_1 = a + h_2 \Leftrightarrow h_1 = h_2$. The second conclusion is an immediate conseqence of the first one and the Proposition 2.2(a). ■

Example 3.24 Let $G = \mathbf{Z}$ and $H = k\mathbf{Z}$ for some $k \in \mathbf{Z}$, $k \neq 0$, the group $G/H = \mathbf{Z}/k\mathbf{Z}$ has k elements $[0], [1], \ldots, [k-1]$. Since the coset $[a + b]$ is also represented by the remainder of the division of $a + b$ by k, this group may be identified with Z_k discussed in the previous section. What "identification" means, will become clear in the next section, when we talk about isomorphisms.

Example 3.25 Let $G = \mathbf{Z}^2$ and $H = \mathbf{Z}(\mathbf{j} - \mathbf{i}) = \{(-n, n) \ : \ n \in \mathbf{Z}\}$. We may choose coset representatives of the form $m\mathbf{i} = (m, 0)$, $m \in \mathbf{Z}$. Since any element $(m, n) \in \mathbf{Z}^2$ can be written as $(m+n)\mathbf{i} + n(\mathbf{j} - \mathbf{i}) \in (m+n)\mathbf{i} + H$, we have $G/H = \{[m\mathbf{i}]\}_{m \in \mathbf{Z}}$. It is easily seen that $[k\mathbf{i}] \neq [m\mathbf{i}]$ whenever $k \neq m$, thus there is a bijection between G/H and $\mathbf{Z}$.

Example 3.26 Consider $\mathbf{Z}$ as a subgroup of $\mathbf{R}$ and the quotient $\mathbf{R}/\mathbf{Z}$. Since any real number is an integer translation of a number in the interval $[0, 1)$, $\mathbf{R}/\mathbf{Z}$ is represented by the points of that interval. Moreover there is a bijection between $\mathbf{R}/\mathbf{Z}$ and $[0, 1)$, since no two numbers in that interval may differ by an integer. For any $\alpha, \beta \in [0, 1)$, the coset $[\alpha + \beta]$ is represented in $[0, 1)$ by the fractional part of $\alpha + \beta$. Since $1 \sim 0$, $\mathbf{R}/\mathbf{Z}$ may be visualised as a circle obtained from the interval $[0, 1]$ by gluing 1 to 0.

A very similar example explaining the concept of polar coordinates is the quotient group $\mathbf{R}/2\pi\mathbf{Z}$. The equivalence relation is now $\alpha \sim \beta \Leftrightarrow \beta - \alpha = 2n\pi, n \in \mathbf{Z}$ and the representatives may be sarched, for example, in the interval $[0, 2\pi)$. Thus the elements of $\mathbf{R}/2\pi\mathbf{Z}$ may be identified with the points on the circle $x^2 + y^2 = 1$ in the plane, via the polar coordinate θ in $x = \cos\theta,\ y = \sin\theta$.

3.4 Homomorphisms

Let G and G' be two abelian groups. If we wish to compare them then we need to be able to talk about functions between them. Of course these functions need to preserve the group structure, in other words they need to respect the binary operation. This leads to the following definition.

Definition 3.27 A map $f : G \to G'$ is called a *homomorphism* if

$$f(a + b) = f(a) + f(b)$$

for all $a, b \in G$.

There are some immediate consequences of this definition. For example, as the following argument shows, homomorphisms map the identity element to the identity element.

$$\begin{aligned} f(0) &= f(0+0) = f(0) + f(0) \\ f(0) - f(0) &= f(0) \\ 0 &= f(0) \end{aligned}$$

A similarly trivial argument shows that

$$f(na) = nf(a)$$

for all $n \in \mathbf{Z}$ and $a \in G$.

Proposition 3.28 *Let $f : G \to G'$ be a homomorphism. Then*

(a) for any subgroup H of G, its image $f(H)$ is a subgroup of G';

(b) for any subgroup H' of G', its inverse image $f^{-1}(H)$ is a subgroup of G;

(c) if f is bijective (i.e. one-to-one and onto) then its inverse $f^{-1} : G' \to G$ also is a bijective homomorphism.

Proof: (a) We must show that $f(H)$ satisfies the group axioms. Since $f(H) \subset G'$, the binary operation on $f(H)$ is the same as that of G' and therefore is associative and commutative. Since $f(0) = 0$, $0 \in H$. Let $b \in H$, then there exists $a \in G$ such that $b = f(a)$. Now observe that $0 = f(a + -a) = f(a) + f(-a)$. Therefore, $f(a) = -f(a)$. Finally, we need to show that $f(H)$ is closed under the operation $+$. If $b, b' \in f(H)$, then there exist $a, a' \in H$ such that $f(a) = b$ and $f(a') = b'$. Furthermore, $b + b' = f(a) + f(a') = f(a + a') \in f(H)$.

(b) and (c) follow from similar types of arguments and are left to the reader. ■

Definition 3.29 The set $\operatorname{im} f := f(G)$ is called the *image* or *range* of f in G and, by the previous proposition is a subgroup of G'. The set

$$\ker f := f^{-1}(0) = \{a \in G \mid f(a) = 0\}$$

is called the *kernel* of f and is a subgroup of G.

Definition 3.30 A homomorphism $f : G \to G'$ is called an *epimorphism* if it is surjective (or onto) i.e. $\operatorname{im} f = G'$ and a *monomorphism* if it is injective (or 1-1), i.e. for any $a \neq b$ in G, $f(a) \neq f(b)$. This condition obviously is equivalent to the condition $\ker f = 0$. Finally, f is called an *isomorphism* if it is both a monomorphism and an epimorphism.

The last definition requires some discussion since the word isomorphism takes different meanings in different branches of mathematics. Let X, Y be any sets and $f : X \to Y$ any map. Then f is called *invertible* if there exists a map $g : Y \to X$, called *inverse* of f with the property

$$gf = \mathbf{1}_X \text{ and } fg = \mathbf{1}_Y \tag{3.7}$$

where $\mathbf{1}_X$ and $\mathbf{1}_Y$ denote the identity maps on X and Y respectively. It is easy to show that f is invertible if and only if it is bijective. If this is the case, g is uniquelly determined and denoted by f^{-1}. When we speak about a particular class of maps, by an invertible map or an isomorphism we mean a map which has an inverse in the same class of maps. For example, if continuous maps are of concern, an isomorphism would be a continuous map which has a continuous inverse: The continuity of a bijective map does not guarantee, in general, the continuity of its inverse. Proposition 3.1(c) guarantees that this problem does not occur in the class of homomorphisms. Thus, a homomorphism is an isomorphism if and only if it is invertible in the class of homomorphisms.

When $G = G'$, a homomorphism $f : G \to G$ may be also be called an *endomorphism* and an isomorphism $f : G \to G$ may be called an *automorphism*.

Groups G and G' are called *isomorphic*, notation $G \cong G'$, if there exists an isomorphism $f : G \to G'$, we may then write $f : G \stackrel{\sim}{\to} G'$ or $G \stackrel{f}{\cong} G'$. It is easy to see that $G \cong G'$ is an equivalence relation. We shall often permit ourselves to identify isomorphic groups, unless an additional structure that is not preserved by isomorphisms is involved.

Example 3.31 $\mathbf{Z}_6 \cong \mathbf{Z}_2 \times \mathbf{Z}_3$.

Example 3.32 Let A, B be subgroups of G such that $G = A \oplus B$. Then the map $f : A \times B \to G$ defined by $f(a, b) = a + b$ is an isomorphism with the inverse defined by $f^{-1}(c) = (a, b)$ where $c = a+b$ is the unique decomposition of $c \in G$ with $a \in A$ and $b \in B$. This can be generalised to direct sums and products of any finite number of groups.

Example 3.33 Let G be a cyclic group of infinite order generated by a. Then $f : \mathbf{Z} \to G$ defined by $f(n) = na$ is an isomorphism with the inverse defined by $f^{-1}(na) = n$. By the same argument, any cyclic group of order k is isomorphic to $\mathbf{Z}_k$.

Example 3.34 Let G be a free abelian group generated by $\{s_1, s_2, \ldots, s_n\}$ discussed in the previous section. Then $G \cong \mathbf{Z}^n$. Indeed, the map $f : \mathbf{Z}^n \to G$ defined on the elements of the canonical basis by $f(e^i) := s_i$ and extended by linearity is a well defined isomorphism.

Example 3.35 Let $f : \mathbf{Z} \to \mathbf{Z}$ be any homomorphism. By linearity, f is completely defined by its values on 1. If $f(1) = k$ then $f(n) = nk$ for all n. If $k = 0$, f is trivial and $\ker f = \mathbf{Z}$. Otherwise $\ker f = 0$ and $\operatorname{im} f = k\mathbf{Z}$. Since $k\mathbf{Z} = \mathbf{Z}$ if and only if $k = \pm 1$, the only automorphisms of $\mathbf{Z}$ are $\mathbf{1}_{\mathbf{Z}}$ and $-\mathbf{1}_{\mathbf{Z}}$.

Example 3.36 Let A, B, and G be as above. The inclusion map $i : A \to G$ is a monomorphism and the projection map $p : G \to A$ defined by $p(c) = a$ where $c = a+b$ with $a \in A$ and $b \in B$, is an epimorphism. Note that $pi = \mathbf{1}_A$ hence p may be called a *left inverse* of i and i a *right inverse* of p. Note that a left inverse is not necessarily unique. Indeed, take subgroups $A = \mathbf{Z}\mathbf{i}$, $B = \mathbf{Z}\mathbf{j}$ of $\mathbf{Z}^2$. Another choice of a left inverse of i is $p'(n\mathbf{i}+m\mathbf{j}) = (n+m)\mathbf{i}$ (a "slant" projection).

Example 3.37 Let H be a subgroup of G and define $q : G \to G/H$ by the formula $q(a) := a+H$. It is easy to see that q is an epimorphism and its kernel is precisely H. This map is called the *canonical quotient homomorphism.*

Let now $f : G \to G'$ be a homomorphism and $H = \ker f$. Then, for any $a \in G$ and $h \in H$, we have $f(a + h) = f(a)$. Hence the image of any coset $a + H$ under f is

$$f(a + H) = \{f(a)\} .$$

Moreover, that image is independent on the choice of a representative of a coset $a + H$. Indeed, if $a + H = b + H$ then $b - a \in H$ thus $f(b) = f(a)$. We may now state the following

Theorem 3.38 *Let $f : G \to G'$ be a homomorphism and $H = \ker f$. Then the map*

$$\bar{f} : G/H \to \operatorname{im} f$$

defined by $\bar{f}(a + H) = f(a)$ is an isomorphism, called the quotient isomorphism.

Proof: By the preceeding discussion, the formula for f is independent of the choice of coset representatives, thus $\bar{f}$ is well defined. Since,

$$\bar{f}((a+H)+(b+H)) = \bar{f}(a+b+H) = f(a+b) = f(a)+f(b)$$

it is a homomorphism. $\bar{f}$ is a monomorphism since $f(a+H) = f(a) = 0$ which is equivalent to $\ker f = H$.

Finally, $\bar{f}$ is, also, an epimorphism since $\operatorname{im}\bar{f} = \operatorname{im} f$. ∎

Example 3.39 Let $q : G \to G/H$ be the canonical homomorphism from Example 3.37. Then $\bar{q} = \mathbf{1}_{G/H}$, so this is the trivial case of Theorem 3.1.

Example 3.40 Let $f : \mathbf{Z} \to \mathbf{Z}_n$ be given by $f(a) = a \bmod n$ (the remainder of a in the division by n). Then f is a well defined epimorphism with $\ker f = k\mathbf{Z}$. Thus $\bar{f} : \mathbf{Z}/k\mathbf{Z} \overset{\simeq}{\rightrightarrows} \mathbf{Z}_k$.

Example 3.41 Let's go back to p' in Example 3.36. $\operatorname{im} p' = \mathbf{Zi} = A$ and $\ker p' = \mathbf{Z}(\mathbf{j}-\mathbf{i})$ Thus $\bar{f} : \mathbf{Z}^2/\mathbf{Z}(\mathbf{j}-\mathbf{i}) \overset{\simeq}{\rightrightarrows} \mathbf{Zi}$. Note that $\mathbf{Z}^2 = \mathbf{Zi} \oplus \mathbf{Z}(\mathbf{j}-\mathbf{i}) = \operatorname{im} p' \oplus \ker p'$. This observation will be later generalized.

Example 3.42 Consider Example 3.26 in terms of the quotient isomorphism. Let S^1 be the unit circle in the complex plane, i.e. the set defined by $|z| = 1$, $z = x + iy \in \mathbf{C}$, i the primitive square root of -1. Then S^1 is a multiplicative group with the complex number multiplication and the unity $1 = 1 + i0$. We define $\varphi : \mathbf{R} \to S^1$ by $\varphi(\theta) = e^{i\theta} = \cos\theta + i\sin\theta$. Then φ is a homomorphism from the additive group of $\mathbf{R}$ to the multiplicative group S^1. It is an epimorphism with the kernel $\ker\varphi = 2\pi\mathbf{Z}$. Thus $\bar{\varphi} : \mathbf{R}/2\pi\mathbf{Z} \overset{\simeq}{\rightrightarrows} S^1$.

Exercises

3.9 If m and n are relatively prime, show that $\mathbf{Z}_m \oplus \mathbf{Z}_n \simeq \mathbf{Z}_{mn}$ (see Exercise 3.2).

3.10 Let $f : G \to F$ be a homomorphism of abelian groups.

(a) If F is free, show that there exists a subgroup G' of G such that $G = \ker f \oplus G'$. Conclude that $G' \simeq F$.

(b) Give an example showing that if F is not free than the conlusion may be wrong.

3.11 Let $g : H \to G$ be a monomorphism, $f : G \to F$ an epimorphism and suppose that $\ker f = \operatorname{im} g$. If F is free, show that $G \simeq H \oplus F$.

3.5 Matrix Algebra over Z and Normal Form

A basic technique in the study of linear maps of vector spaces is the row and column reduction of matrices. In this section we discuss the analogy of this technique in the study of homomorphisms of free abelian groups. Many results of elementary matrix algebra have straightforward extensions to our case but there is one subtlety; our matrices have integer coefficients and division is not allowed. For example, the operation of multiplying the i-th row of a matrix by a number a is an *elementary row operation over* $\mathbf{Z}$ if and only if $a = \pm 1$, otherwise it is not invertible.

Let G and G' be finitely generated free abelian groups with bases, respectively, $\{g_1, g_2, \ldots, g_n\}$ and $\{g'_1, g'_2, \ldots, g'_m\}$. If $f : G \to G'$ is any homomorphism, then it is determined by its action on the basis elements of G. Even more, there are unique $a_{ij} \in \mathbf{Z}$, $i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n$ such that

$$f(g_j) = \sum_{i=1}^{m} a_{ij} g'_i \,. \tag{3.8}$$

Conversely, if $A = (a_{ij})$ is any $n \times m$ matrix with integer coefficients, then the formula (3.8) extends by linearity to a unique homomorphism $f : G \to G'$. Thus f may be identified with the matrix A called the *matrix of* f *with respect to the given bases* on G and G'.

Due to the isomorphism in Example 3.34 associating any basis in G and G' to the canonical bases in $\mathbf{Z}^n$ and $\mathbf{Z}^m$, we may suppose that $G = \mathbf{Z}^n$ and $G' = \mathbf{Z}^m$. Then $f : \mathbf{Z}^n \to \mathbf{Z}^m$ is represented by the matrix multiplication $y = f(x) = Ax$ or, more explicitely, by

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} . \tag{3.9}$$

Recall that the columns of A generate the image $\operatorname{im} A := \operatorname{im} f$. In particular, if $n = m$ and Equation (3.9) is a change of coordinates, the columns of A are elements of the new basis for $\mathbf{Z}^m$ expressed in terms of the canonical basis of $\mathbf{Z}^m$.

For a fixed matrix A, denote by $R_1, R_2, \ldots, R_m$ its rows and by $C_1, C_2, \ldots, C_n$ its columns. Here are the three types of *elementary row operations over* $\mathbf{Z}$:

(r1) Exchange rows R_i and R_k ;

(r2) Multiply R_i by -1 ;

(r3) Replace R_i by $R_i + qR_k$, where $q \in \mathbf{Z}$.

Note that these operations are invertible over $\mathbf{Z}$. Indeed, (r1) and (r2) are self-inverses and the inverse of (r3) is replacing R_i by $R_i - qR_k$. Each operation can be expressed in terms of matrix multiplication: new matrix B is obtained by multiplying A on the left by an *elementaty matrix* E which is obtained by performing the same operation on the identity $m \times m$ matrix $I_{m\times m}$.

Example 3.43 Let A be a 5×3 matrix. If we wish to exchange the second and third column, this can be done by the elementary matrix

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

since

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{13} & a_{14} & a_{15} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{13} & a_{12} & a_{14} & a_{15} \\ a_{21} & a_{23} & a_{22} & a_{24} & a_{25} \\ a_{31} & a_{33} & a_{32} & a_{34} & a_{35} \end{bmatrix}.$$

The same applies to *elementary column operations over* $\mathbf{Z}$:

(c1) Exchange columns C_j and C_l ;

(c2) Multiply C_j by -1 ;

(c3) Replace C_j by $C_j + qC_l$, where $q \in \mathbf{Z}$,

which are, in fact, row operations on the transposed matrix A^T. The elementary column operations correspond to the right multiplication of A by elementary matrices D obtained by performing the same operation on the identity $n \times n$ matrix $I_{n\times n}$.

Each row operation corresponds to a change of basis in the range space $\mathbf{Z}^m$. Indeed, if $B = EA$ where E is an elementary matrix, then the equation $y = Ax$ is equivalent to $\bar{y} = Bx$, where $\bar{y} := Ey$. Since E is invertible, $y = E^{-1}\bar{y}$, and the columns of E^{-1} are the new basic vectors in $\mathbf{Z}^m$. Similarly, each column operation corresponds to a change of basis in the domain $\mathbf{Z}^n$. If $C = AE$, where E is an elementary matrix, then the equation $y = Ax$ is equivalent to $y = C\bar{x}$ where $\bar{x} := E^{-1}x$, or $x = E\bar{x}$. Thus the columns of E represent the new basic vectors in $\mathbf{Z}^n$. The following propositions are straightforward analogies of elementary linear algebra results.

Proposition 3.44 *Let A be an $n \times m$ matrix with integer coefficients.*
(a) The elementary row operations over $\mathbf{Z}$ preserve the subgroups $\ker A$ and $\operatorname{coim} A := \operatorname{im} A^T$ of $\mathbf{Z}^n$.
(b) The elementary column operations over $\mathbf{Z}$ preserve the subgroups $\operatorname{im} A$ and $\operatorname{coker} A := \ker A^T$ of $\mathbf{Z}^m$.

The group $\operatorname{coim} A$ is traditionally called *the row space of A* and $\operatorname{im} A$ *the column space of A*. This terminology is justified by the above remark that the columns of A generate $\operatorname{im} A$.

Definition 3.45 A matrix A is in *row echelon form* if the following property is satisfied. Let a_{ij} be the first non-zero entry in its row R_i, then $a_{kj} = 0$ for all $k > j$.

Proposition 3.46 *Suppose that A is in row echelon form, then the non-zero rows of A are linearly independent, and thus they form a basis for $\operatorname{coim} A$.*

Example 3.47 We show that the elements $(3,2), (2,0)$ and $(0,3)$ of $\mathbf{Z}^2$ generate the whole group $\mathbf{Z}^2$, although no two of them do. Indeed, row operations over $\mathbf{Z}$ give

$$\begin{bmatrix} 3 & 2 \\ 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{matrix} R_1 - R_2 \\ \longrightarrow \\ \longrightarrow \end{matrix} \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{matrix} \longrightarrow \\ R_2 - 2R_1 + R_3 \\ \longrightarrow \end{matrix} \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 0 & 3 \end{bmatrix}$$

$$\begin{matrix} R_1 + 2R_2 \\ (-1)R_2 \\ R_3 - 3R_2 \end{matrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

hence the first two rows $(1,0), (0,1)$ generate the row space of the initial matrix.

Example 3.48 Let $A : \mathbf{Z}^3 \to \mathbf{Z}^4$ be given by

$$A = \begin{bmatrix} 0 & 2 & 2 \\ 1 & 0 & -1 \\ 3 & 4 & 1 \\ 5 & 3 & -2 \end{bmatrix} .$$

We will find bases for $\ker A$ and $\operatorname{im} A$. The simultanuous row operations over $\mathbf{Z}$ of the identity matrix $I_{3\times 3}$ and A^T give

$$[I|A^T] = \left[\begin{array}{ccc|cccc} 1 & 0 & 0 & 0 & 1 & 3 & 5 \\ 0 & 1 & 0 & 2 & 0 & 4 & 3 \\ 0 & 0 & 1 & 2 & -1 & 1 & -2 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|cccc} 0 & 1 & 0 & 2 & 0 & 4 & 3 \\ 1 & 0 & 0 & 0 & 1 & 3 & 5 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 \end{array}\right] = [P^T|C^T] ,$$

where $C = AP$. Since the matrix C^T is in a row echelon form, its first two rows $(2,0,4,3)$ and $(0,1,3,5)$ form a basis for $\operatorname{im} C = \operatorname{im} A$. The third row $(1,-1,1)$ of P^T generates $\ker A$.

The following two theorems show that the method presented in the above examples may be applied to any integer matrix. Their proofs are constructive and may be used to obtain formal algorithms.

Theorem 3.49 *Let A be an $n \times m$ matrix with integer coefficients. Then A can be brought to a row echelon form by means of elementary row operations over* $\mathbf{Z}$.

Proof: The proof is by induction on the number m of rows of A.

If $m = 1$, then

$$A = [a_{11}\ a_{12}\ a_{13}\ \cdots\ a_{1n}]$$

which is in row echelon form.

From now on assume $m > 1$.

Case 1. *The first column C_1 of A has at most one nonzero entry.*

Assume that C_1 has one nonzero entry a_{k1}. Apply the row operation **r1** to exchange rows 1 and k. Then the new matrix has the form

$$\left[\begin{array}{c|ccc} a_{k1} & a_{k2} & \cdots & a_{kn} \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right]$$

If all entries in C_1 were zero, then the matrix would have a similar form even without exchanging any rows. Observe that A' is an $n-1 \times m-1$ matrix and so by the induction arguement can be reduced using row operations to row echelon form. Thus, A can be reduced to row echelon form.

Hence from now on it is assumed that C_1 has multiple nonzero entries. Let

$$\alpha = \alpha(A) := \min\{|a_{i1}| \mid a_{i1} \neq 0,\ i = 1, 2, \ldots m\}$$

Let $|a_{k1}| = \alpha$. Without loss of generality we may assume that $a_{k1} = \alpha$. If not, we could use the row operation **r2** to change the sign of a_{k1}. There are two cases to consider.

Case 2. *α divides all entries of C_1.*

The assumption that α divides all entries of C_1 is equivalent to the statement that for each $a_{i1} \neq 0$ there exists $q_i \in \mathbf{Z}$ such that $\alpha q_i := a_{i1}$. For each $a_{i1} \neq 0$ apply the row operation **r3** to replace R_i by $R_i - q_i R_k$. This results in a new first column all of whos entries are zero except a_{k1}. Thus the problem is reduced to Case 1.

Case 3. *α fails to divide some entry a_{i1}, $i \neq k$ of the first column.*

If α does not divide the entry a_{i1}, then $a_{i1} = q_i\alpha + r_i$, where $q_i, r_i \in \mathbf{Z}$ and $O < |r_i| < \alpha$. Let $A_1 := A$ and let A_2 be the matrix obtained by replacing R_i by $R_i - q_i R_k$. The first entry of the new row R_i is r_i. Returning to the definition of α observe that

$$\alpha(A_2) = |r_i| < \alpha(A_1).$$

If A_2 satisfies Case 2, then we are done. If it does not, then applying the argument of Case 3 using $\alpha(A_2)$ results in a matrix A_3. Applying Case 3 mulitple times results in a series of matrices $A_1, A_2, A_3 \ldots$ where

$$\alpha(A_1) > \alpha(A_2) > \alpha(A_3) > \ldots .$$

Since any strictly decreasing sequence of positive integers is finite, there is a matrix A_l which falls into Case 2. ■

Theorem 3.50 *Let $A \neq 0$ be an $n \times m$ matrix with integer coefficients. By means of elementary row and column operations over $\mathbf{Z}$, it is possible to bring*

A to the form

$$B = \left[\begin{array}{ccccc|c} b_1 & & & & & \\ & b_2 & & 0 & & \\ & & \cdot & & & \\ & & & \cdot & & 0 \\ & 0 & & & \cdot & \\ & & & & b_s & \\ \hline & & 0 & & & 0 \end{array}\right], \tag{3.10}$$

where b_i are postive integers and b_i divides b_{i+1} for all i.

Proof: The proof is essentially an a more elaborated version of the arguments of the previous proof. The induction is now on the totatl number of entries of the matrix nm.

If $nm = 1$, then A is trivially in normal form.

From now on assume that $nm > 1$. Let

$$\alpha = \alpha(A) := \min\{|a_{ij}| a_{ij} \neq 0 \; i = 1, 2, \ldots m, \; j = 1, 2, \ldots n\}.$$

Let $|a_{kl}| = \alpha$. As before, we may assume that $a_{kl} = \alpha$ since otherwise we multiply R_k by -1. There are three cases to consider.

Case 1. *α divides all entries of A.*

The following simple observation is crucial.

Observation: *If an integer α divides all entries of A and a matrix B is obtained from A by elementary row and column operations over* **Z***, then α divides all entries of B.*

By row and column exchanges we get $\alpha = a_{11}$. By the arguments of the previous proof we get a matrix whose first column is $[\alpha, 0, 0, \ldots, 0]^T$ and using those arguments for A^T gives the first row $[\alpha, 0, 0, \ldots, 0]$. We put $b_1 := \alpha$ and use the induction hypothesis for the matrix A' obtained by removing the first row and first column. By the above observation, b_1 divides b_i for all $i > 1$.

Case 2. *$\alpha = a_{kl}$ fails to divide some entry of its row R_k or its column C_l.*

Then we apply the same arguments as in the previous proof to reduce the problem to Case 1.

Case 3. *$\alpha = a_{kl}$ divides all entries in its row and column but it fails to divide some entry a_{ij} with $i \neq k$ and $j \neq l$.*

Let $q = a_{il}\alpha^{-1} \in \mathbf{Z}$. We first replace R_i by $R_i - qR_k$ so to get a new i'th row R_i' whose l'th entry is 0 and j'th entry is $a_{ij} - qa_{kj}$. Then we replace R_k by $R_k' = R_k + R_i'$. The first entry of R_k' is α and the j'th entry is $a_{kj}' = (1-q)a_{kj} + a_{ij}$. By the hypothesis, α does not divide a_{kj}', so the problem is reduced to Case 2. ■

The matrix B given by Theorem 4.2 is called the *normal form* of A. Due to the relation between elementary row and column operations over $\mathbf{Z}$ and changes of bases discussed at the begenning of this section, we reach the following

Corollary 3.51 *Let $f : G \to G'$ be a homomorphism of finitely generated free abelian groups. Then there are bases of G and G' such that the matrix of f with respect to those bases is in the normal form (4).*

It should be emphasized that the problem of reducing a matrix to the normal form (4) should be well distinguished from a more difficult problem of diagonalizing an $n \times n$ real matrix A. In the second case, the problem is to find one basis, the same one for $\mathbf{R}^n$ viewed as the domain and as the range of A.

Exercises

3.12 For each matrix A specified below, find its normal form B and two integer matrices P and Q, invertible over $\mathbf{Z}$, such that $QB = AP$. Use the information provided by P and Q for presenting bases with respect to which the normal form is assumed, a basis for $\ker A$, and a basis for $\operatorname{im} A$.

(a) $A = \begin{bmatrix} 6 & 4 \\ 4 & 0 \\ 0 & 6 \end{bmatrix}$

(b) The matrix A in Example **??**.

(c) $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 9 \end{bmatrix}$

3.6 Decomposition Theorem for Abelian Groups

The goal of this section is to prove the following decomposition theorem for finitely generated free abelian groups.

Theorem 3.52 *Let G be a finitely generated abelian group. Then G can be decomposed as a direct sum of cyclic groups. More explicitely, there exist generators $g_1, g_2, \ldots, g_n$ of G and an integer $0 \leq r \leq n$ such that*

1. $$G = \bigoplus_{i=1}^{n} \langle g_i \rangle \ ,$$

2. *If $r > 0$, $g_1, g_2 \ldots, g_m$ are of infinite order,*

3. *If $k = n - r > 0$ then $g_{r+1}, g_{r+2}, \ldots, g_{m+k}$ have finite orders $t_1, t_2, \ldots, t_k$, respectively and $1 < t_1 | t_2 | \ldots | t_k$*

The numbers m and $t_1, t_2, \ldots, t_k$ are uniquely determined by G, although generators $g_1, g_2, \ldots, g_n$ are not.

The above theorem allows us to write G as $G = F \oplus T$ where

$$F = \bigoplus_{i=1}^{r} \mathbf{Z} g_i \quad T = \bigoplus_{i=1}^{k} \langle g_{\beta+i} \rangle \ .$$

T is the torsion subgroup of G mentioned in Section 1 and F is a maximal free subgroup of G. The number r is the rank of F and it is called the *betti number* of G and the numbers $t_1, t_2, \ldots, t_k$ are called the *torsion coefficients* of G.

By Example 3.17, we get the following

Corollary 3.53 *Let G be a finitely generated abelian group. Then G is isomorphic to*

$$\mathbf{Z}^r \oplus \mathbf{Z}/t_1 \oplus \mathbf{Z}/t_2 \oplus \ldots \oplus \mathbf{Z}/t_k$$

where r and $t_1, t_2, \ldots, t_k$ are as in Theorem 4.1.

By Exercise 3.9, if m, n are relatively prime, then $\mathbf{Z}_{mn} \simeq \mathbf{Z}_m \oplus \mathbf{Z}_n$. Thus, by decomposing the numbers $t_1, t_2, \ldots, t_k$ to products of primes we get the following

Corollary 3.54 *Any finitely generated abelian group G is isomorphic to*

$$\mathbf{Z}^r \oplus \mathbf{Z}/p_1^{m_1} \oplus \mathbf{Z}/p_2^{m_2} \oplus \ldots \oplus \mathbf{Z}/p_s^{m_s}$$

where $p_1, p_2, \ldots, p_s$ are prime numbers.

To prove Theorem 3.52 requires the following results.

Proposition 3.55 *Let F be a finitely generated free abelian group. Let H be a subgroup of F, then H is finitely generated.*

Proof: Since F is a finitely generated free abelian group there is an integer n such that $F \cong \mathbf{Z}^n$. Using this isomorphism we shall identify F with $\mathbf{Z}^n$ and think of H as a subgroup of $\mathbf{Z}^n$.

To show that H is finitely generated, it is sufficient to find a finite collection $\{h_1, h_2, \ldots, h_n\}$ of elements of $\mathbf{Z}^n$ which generate H. Let $\pi_i : \mathbf{Z}^n \to \mathbf{Z}$ be the canonical projection that sends $(a_1, a_2, \ldots, a_n) \mapsto a_i$. Define

$$H_m := \{b \in H \mid \pi_i(b) = 0 \text{ if } i > m\}.$$

Observe that an element of H_m is of the form $(b_1, b_2, \ldots, b_m, 0, \ldots, 0)$. From this it is easy to check that for all $m \leq n$, H_m is a subgroup of H and $H_n = H$.

For $m = 1, \ldots n$ consider $\pi_m(H_m)$. We will use this group to define the above mentioned generator h_m.

If $\pi_m(H_m) = 0$, then define $h_m = 0$.

If $\pi_m(H_m) \neq 0$, then $\pi_m(H_m)$ is a nontrivial subgroup of $\mathbf{Z}$, and therefore cyclic. This means that there exists $k_m \in \mathbf{Z}$ such that $< k_m >= \pi_m(H_m)$. Define h_m by $\pi_m(h_m) = k_m$.

We need to show that the set $\{h_1, h_2, \ldots, h_n\}$ generates H. This will be done by induction on m. If $m = 1$, then

$$\langle \pi_1(h_1) \rangle = \pi_1(H_1)$$

which implies that $\langle h_1 \rangle = H_1$ or that $H_1 = 0$.

Now assume that $\{h_1, h_2, \ldots, h_{m-1}\}$ generates H_{m-1}. Let $h \in H_m$. Then $\pi_m(h) = k\pi_m(h_m)$ for some integer j. This implies that $\pi_m(h - jk_m) = 0$, and hence $h - ih_m \in H_{m-1}$. Thus

$$h = ih_m + i_1h_1 + i_2h_2 + \cdots + i_{m-1}h_{m-1}$$

and the conclusion follows.

It is left as exercice to prove that tha non-zero elements of $\{h_1, h_2, \ldots, h_n\}$ are linearly independent, hence they form a basis for H. ■

Proposition 3.56 *Let F be a finitely generated free abelian group. Then any subgroup H of F is free of rank $r(H) \leq r(F)$.*

Proof: Since F is a finitely generated free abelian group there is an integer n such that $F \cong \mathbf{Z}^n$. Using this isomorphism we shall identify F with $\mathbf{Z}^n$ and think of H as a subgroup of $\mathbf{Z}^n$. By Proposition 3.55 there exist $h_1, h_2, \ldots, h_m \in \mathbf{Z}^n$ generators of H. Consider a matrix A whose i'th row is the vector h_i. Then H is the row space of A. By Theorem 3.49, A may be reduced over $\mathbf{Z}$ to a row echelon form. The non-zero rows of the reduced matrix are linearly independent and hence they form a basis for H. Of course, the number of non-zero rows of an echelon matrix is less or equal than the number n of columns, thus $r(H) \leq r(F)$. ■

Proof of Theorem 3.52: Let $S := \{s_1, \ldots, s_m\}$ be a set of generators for G. Consider the free abelian group $\mathbf{Z}^S$. Recall we defined the functions $\hat{s}_i : S \to \mathbf{Z}$, $i = 1, \ldots m$ by

$$\hat{s}_i(s_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise,} \end{cases}$$

which form a basis for $\mathbf{Z}^S$.

Define $f : \mathbf{Z}^S \to G$ by $f(\hat{s}_i) = s_i$. This is a group homomorphism and so $H := \ker f$ is a subgroup of $\mathbf{Z}^S$. By Theorem 3.38,

$$\bar{f} : \mathbf{Z}^S/H \to G$$

is an isomorphism. Thus to prove the theorem it is sufficient to obtain the desired decomposition for the group $\mathbf{Z}^S/H$.

Since $\mathbf{Z}^S$ is a finitely generated free abelian group, by Proposition **??** H is a free group and $r := \operatorname{rank} H \leq m$.

Let $j : H \to \mathbf{Z}^S$ be the inclusion homomorphism. Then by Theorem 3.50 there exist bases $\{h_1, h_2, \ldots, h_r\}$ for H and $\{z_1, z_2, \ldots z_m\}$ for $\mathbf{Z}^S$ such that the matrix for j has the form

$$\left[\begin{array}{ccc} b_1 & & 0 \\ & \ddots & \\ 0 & & b_r \\ \hline & & \\ & 0 & \\ & & \end{array}\right]$$

where $b_1 \geq 1$ and $b_i | b_{i+1}$. Since, j is a monomorphism each $b_i \neq 0$.

Observe that the basis for H as a subset of $\mathbf{Z}^S$ is $\{b_1 z_1, b_2 z_2, \ldots, b_r z_r\}$. It is now easy to see that

$$\mathbf{Z}^S / H \cong \mathbf{Z} z_1 / \mathbf{Z} b_1 z_1 \oplus \cdots \oplus \mathbf{Z} z_r / \mathbf{Z} b_r z_r \oplus \mathbf{Z} z_{r+1} \oplus \cdots \oplus \mathbf{Z} z_m .$$

If $b_1, \ldots, b_s = 1$, then for $i = 1, \ldots, s$,

$$\mathbf{Z} z_i / \mathbf{Z} b_i z_i \cong 0.$$

If $b_{s+1}, \ldots, b_r > 1$, then for $j = s+1, \ldots, r$,

$$\mathbf{Z} z_i / \mathbf{Z} b_i z_i \cong \mathbf{Z}_{b_j} .$$

Therefore,

$$\mathbf{Z}^S / H \cong \mathbf{Z}_{b_{s+1}} \oplus \cdots \oplus \mathbf{Z}_{b_r} \oplus \mathbf{Z}^{m-r} .$$

∎

3.7 Homology Groups

We now turn to a purely algebraic description of Homology groups. Recall that in Chapter 2 we were forced to deal with Homology groups in the context of vector spaces, with what we have learned in this Chapter we can now handle the general case, at least in the purely algebraic setting.

Definition 3.57 A *chain complex* $\mathcal{C} = \{C_n, \partial_n\}_{n\in\mathbf{Z}}$ consists of abelian groups C_n, called *chains*, and homomorphisms $\partial_n : C_n \to C_{n-1}$, called *boundary operators*, such that

$$\partial_n \circ \partial_{n+1} = 0 \tag{3.11}$$

$\mathcal{C}$ is a *free chain complex* if C_n is free for all $n \in \mathbf{Z}$. The *cycles* of $\mathcal{C}$ is the subgroup

$$Z_n := \ker \partial_n$$

while the *boundaries* are the subgroups

$$B_n := \operatorname{im} \partial_{n+1}.$$

Observe that (3.11) implies that

$$\operatorname{im} \partial_{n+1} \subset \ker \partial_n$$

and hence the following definition makes sense.

Definition 3.58 The n-th *homology group* of the chain complex $\mathcal{C}$ is

$$H_n(\mathcal{C}) := \text{cycles/boundaries} = \ker \partial_n / \operatorname{im} \partial_{n+1}.$$

Observe that this is a purely algebraic definition.

Definition 3.59 $\mathcal{C}$ is a *finite chain complex* if:

1. each C_n is a finitely generated free abelian group,
2. there exists an $N \geq 0$ such that $C_n = 0$ for all $n > N$ and $n < 0$.

We will only be concerned with free finite chain complexes in this book.

Theorem 3.60 (Standard Basis for Free Chain Complexes) *Let $\mathcal{C} = \{C_n, \partial_n\}$ be a free finite chain complex. Then, for every $n \in \mathbf{Z}$ there exist subgroups U_n, V_n, and W_n of C_n such that*

$$C_n = U_n \oplus V_n \oplus W_n$$

where

$$\partial_n(U_n) \subset W_{n-1}, \quad \partial_n(V_n) = 0 \quad \partial_n(W_n) = 0.$$

Furthermore, there are bases for U_n and W_{n-1} for which the matrix of ∂_n takes the form

$$\partial_n = \begin{bmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_l \end{bmatrix} \quad b_i \geq 1, \; b_i \mid b_{i+1}.$$

Proof: Let $Z_n := \ker \partial_n$. These are the cycles introduced in Chapter 2. Similarly, the boundaries are $B_n : \operatorname{im} \partial_{n+1}$. Define

$$W_n := \{c \in C_n \mid \exists\, k \in \mathbf{Z} \setminus \{0\} \text{ such that } kc \in B_n\}.$$

Lemma 3.61 *W_n is a subgroup of C_n.*

Proof: $0 \in W_n$ since $0 \in B_n$. If $w \in W_n$, then $kw \in B_n$ for some integer $k \neq 0$. However, B_n is a group so $-kw \in B_n$ which implies that $-w \in B_n$.

Finally, if $w, w' \in W_n$, then there exist nonzero integers k and k' such that $kw, k'w' \in B_n$. Since B_n is a group, $k'kw, kk'w \in B_n$ and hence $k'kw + kk'w' \in B_n$ which implies that $k'k(w + w') \in B_n$. Therefore, $(w + w') \in W_n$. ■

W_n is called the group of *weak boundaries.*

Lemma 3.62 *$W_n \subset Z_n$.*

Proof: If $w \in W_n$, then $kw \in B_n$ for some $k \in \mathbf{Z} \setminus \{0\}$. But $B_n \subset Z_n$ hence $0 = \partial_n kw = k\partial_n w = 0$. However, C_{n-1} is free, and hence, $\partial_n w = 0$. ■

$H_n(\mathcal{C})$ is a finitely generated abelian group and hence

$$H_n(\mathcal{C}) \cong \mathbf{Z}^k \oplus T_n(\mathcal{C})$$

where $T_n(\mathcal{C})$ is the torsion subgroup of $H_n(\mathcal{C})$. Consider the projection

$$p : H_n(\mathcal{C}) \to H_n(\mathcal{C})/T_n(\mathcal{C}) \cong \mathbf{Z}^k.$$

Lemma 3.63 $\ker p \cong W_n$ *and hence* $Z_n/W_n \cong H_n(\mathcal{C})/T_n(\mathcal{C})$.

Proof: By definition $H_n(\mathcal{C}) \cong Z_n/B_n$. So cosets in $H_n(\mathcal{C})$ have the form $c + B_n$. If $kc \in B_n$ for some integer $k \neq 0$, then

$$k[c + B_n] = [kc + B_n] = [B_n] = 0,$$

i.e. $[c+B_n] \in T_n(\mathcal{C})$. On the other hand, if for all nonzero integers k, $kc \notin B_n$, then $k[c + B_n] \neq 0$ for all $k \in \mathbf{Z} \setminus \{0\}$. Thus, $\langle [c + B_n] \rangle \cong \mathbf{Z}$. In conclusion then $[c + B_n] \in T_n(\mathcal{C})$ if and only if $c \in W_n$. ■

Let $\{c_1, \ldots, c_k\}$ be a basis for Z_n/W_n. Let $\{d_1, \ldots, d_l\}$ be a basis for W_n. Then $Z_n \cong V_n \oplus W_n$ where $V_n = \langle c_1, \ldots, c_k \rangle$.

Let $\{e_1, \ldots, e_j\}$ be a basis for C_n and let $\{e'_1, \ldots, e'_m\}$ be a basis for C_{n-1} such that $\partial_n : C_n \to C_{n-1}$ has the form

$$\partial_n = \left[\begin{array}{ccc|c} b_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & b_l & \\ \hline & 0 & & 0 \end{array}\right]$$

The following three observations follow directly from the form of this matrix:

1. $\{e_{l+1}, \ldots, e_n\}$ is a basis for Z_n.
2. $\{b_1 e'_1, \ldots, b_l e'_l\}$ is a basis for B_{n-1}.
3. $\{e'_1, \ldots, e'_l\}$ is a basis for W_{n-1}.

The proof of the theorem is finished once we define $U_n = \langle e_1, \ldots e_l \rangle$. Then $C_n = U_n \oplus Z_n = U_n \oplus V_n \oplus W_n$ where V_n and W_n are defined as above. ■

Theorem 3.64 *The homology groups of a finite free chain complex* $\mathcal{C} = \{\mathbf{C}_n, \partial_n\}$ *are computable.*

Proof: By the previous theorem there exists a standard basis for the free chain complex. Furthermore, this standard basis can be computed using the row and column reductions described in Theorem 3.50. In this basis we can identify the subgroups U_n, V_n and W_n. $B_n = \operatorname{im} U_{n+1}$ and

$$H_n(\mathcal{C}) \cong V_n \oplus W_n/B_n.$$

■

Before ending this section we will introduce yet another construction that leads to homology groups.

Definition 3.65 Let $\mathcal{C} = \{C_n, \partial_n\}$ be a chain complex. A chain complex $\mathcal{D} = \{D_n, \partial'_n\}$ is a *subchain complex* of $\mathcal{C}$ if:

1. D_n is a subgroup of C_n for all $n \in \mathbf{Z}$.
2. $\partial'_n = \partial_n \mid_{D_n}$.

The condition that $\partial'_n = \partial_n \mid_{D_n}$ indicates the boundary operator of a subchain complex is just the boundary operator of the larger complex restricted in its domain. For this reason and to simplify the notation we shall let $\partial' = \partial$.

Let $\mathcal{C} = \{C_n, \partial_n\}$ be a chain complex and let $\mathcal{D} = \{D_n, \partial'_n\}$ be a subchain complex. We can create a new chain complex called the *relative chain complex* whose chains consist of the groups C_n/D_n and whose boundary operators are the induced maps

$$\bar{\partial}_n : C_n/D_n \to C_{n+1}/D_{n+1}$$

given by

$$[c + D_n] \mapsto [\partial_n c + D_{n-1}].$$

$\bar{\partial}_n$ is well defined since $\bar{\partial}_n(D_n) \subset D_{n-1}$. Furthermore,

$$\begin{aligned} \bar{\partial}_n \circ \bar{\partial}_{n+1}[c + D_{n+1}] &= \bar{\partial}_n[\partial_{n+1} c + D_n] \\ &= [\partial_n \circ \partial_{n+1} c + D_{n-1}] \\ &= [0 + D_{n-1}] \\ &= 0. \end{aligned}$$

Definition 3.66 The relative *n-cycles* are $Z_n(\mathcal{C}, \mathcal{D}) := \ker \bar{\partial}_n$. The relative *n-boundaries* are $B_n(\mathcal{C}, \mathcal{D}) := \ker \bar{\partial}_{n+1}$. The *relative homology groups* are

$$H_n((\mathcal{C}, \mathcal{D}) := Z_n(\mathcal{C}, \mathcal{D})/B_n(\mathcal{C}, \mathcal{D}).$$

Chapter 4

Cubical Homology

In Sections **??** we suggested what were the important elements in Homology. In particular, we used the edges and vertices of a graph to generate algebraic objects that measured the nontriviality of the topology of the graph. In this chapter we shall formally defined cubical homology. However, the first step is to generalize the combinatorics of graphs to higher dimensional spaces.

There are several ways to extract combinatoric and algebra information from a set in $\mathbf{R}^n$. The classical approach is by means of triangulations of the space. For example if $n = 2$ that means subdividing the space into triangles so that any two triangles are either disjoint, intersect at a common edge, or at a vertex. The algebra of triangulations is the Simplicial Homology Theory.

An approach arising naturally from numerical computations and graphics is by means of cubical grids which subdivide the space to cubes with vertices in an integer lattice. Look for example at Figure 4.1 The picture seems to be composed of curves which do not look like polygonal curves. But, like any picture produced by a computer, there is only a finite amount of information involved. If we blow up a section of the figure we will see in Figure 4.2 a chain of small squares called in computer graphics *pixels*. Note that any two pixels are either disjoint, intersect at a common edge or at a vertex. The classical Simplicial Homology Theory would require from us subdividing each pixel to a union of at least two triangles in order to compute homology. But that seems to be very artificial: what we see does already have a nice combinatoric structure and we should be able to extract algebra out of it. This approach is the Cubical Homology Theory presented here. At the end of this chapter we shall give a brief overview of the Simplicial Homology and compare the two theories, empasizing strong and weak points of each approach.

Figure 4.1: A typical computer graphics picture.

Figure 4.2: A blow up of the previous figure.

In numerical and graphical analysis one needs to consider very fine cubical grids. The size of cubes of a grid cannot be arbitrarily small because of the computer's capacity. From a theoretical point of view, the size of a grid is just a question of choice of units. With appropriate units we may assume in this chapter that each cube is unitary i.e. it has sides of length 1 and vertices with integer coordinates. Later on we will investigate what happens with the algebra extracted from a cubical grid when we change units.

4.1 Cubical Sets

4.1.1 Elementary Cubes

Definition 4.1 An *elementary interval* is a closed interval $I \subset \mathbf{R}$ of the form

$$I = [l, l+1] \quad \text{or} \quad I = [l, l]$$

for some $k \in \mathbf{Z}$. To simplify the notation we will use the notation

$$[l] = [l, l]$$

for an interval that contains only one point. Elementary intervals that consist of a single point are *degenerate*. Elementary intervals of length one are *nondegenerate*.

Example 4.2 The intervals $[2, 3]$, $[-15, -14]$, and $[7]$ are all examples of elementary intervals. On the other hand, $[\frac{1}{2}, \frac{3}{2}]$ is not an elementary interval since the boundary points are not integers. Similarly, $[1, 3]$ is not an elementary interval since the length of the interval is greater than 1.

Definition 4.3 An *elementary cube* Q is a finite product of elementary intervals, i.e.

$$Q = I_1 \times I_2 \times \cdots \times I_n \subset \mathbf{R}^n$$

where each I_i is an elementary interval. The set of all elementary cubes in $\mathbf{R}^n$ is denoted by $\mathcal{K}^n$. The set of all elementary cubes is denoted by $\mathcal{K}$, i.e.

$$\mathcal{K} := \bigcup_{n=1}^{\infty} \mathcal{K}^n.$$

Figure 4.3 indicates a variety of elementary cubes. Observe that the cube $[1,2] \subset \mathbf{R}$ is different from the cube $[1,2] \times [0] \subset \mathbf{R}^2$ since they are subsets of different spaces. Of course using the inclusion map $\iota : \mathbf{R} \to \mathbf{R}^2$ given by $\iota(x) = (x, 0)$ we can identify these two elementary cubes. However, we will take great care in this book to explicitly state this identification if we make it. Thus, if the identification is not clearly stated, then they should be treated as distinct sets.

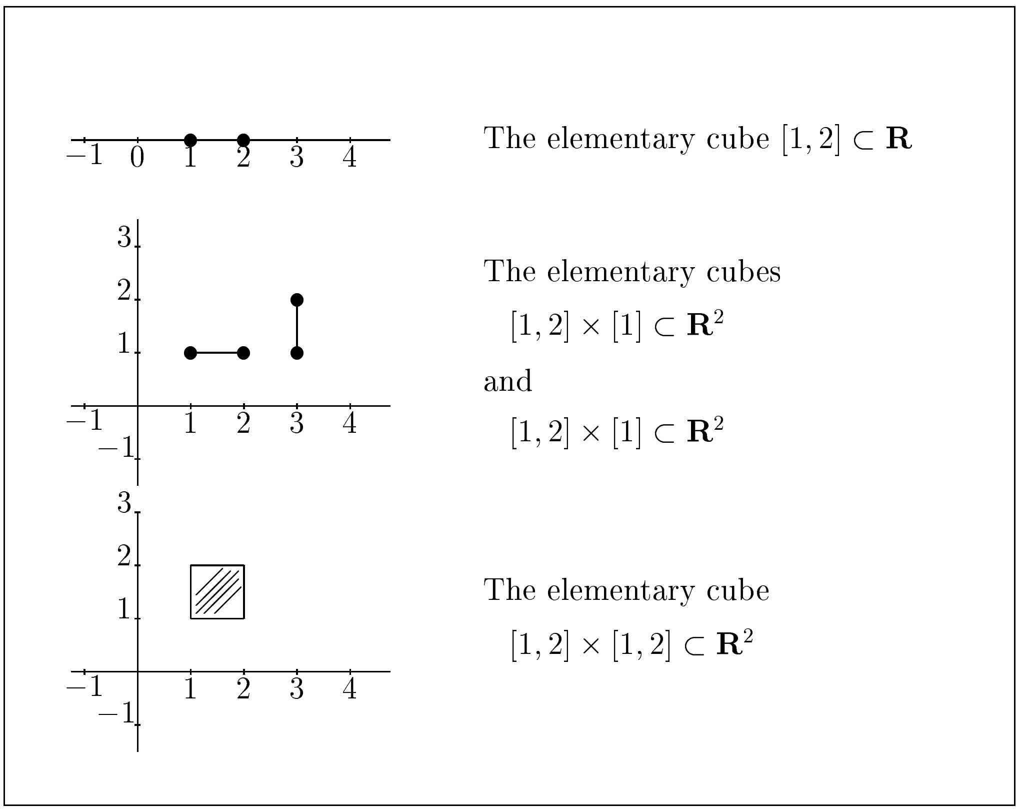


Figure 4.3: Elementary cubes in $\mathbf{R}$ and $\mathbf{R}^2$.

Of course there are many other elementary cubes, e.g.

$$Q_1 := [1,2] \times [0,1] \times [-2,-1] \subset \mathbf{R}^3$$

$$\begin{aligned}
Q_2 &:= [1] \times [1,2] \times [0,1] = \{1\} \times [1,2] \times [0,1] \subset \mathbf{R}^3 \\
Q_3 &:= [1,2] \times [0] \times [-1] = [1,2] \times \{0\} \times \{-1\} \subset \mathbf{R}^3 \\
Q_4 &:= [0] \times [0] \times [0] = (0,0,0) \in \mathbf{R}^3 \\
Q_5 &:= [-1,0] \times [3,4] \times [6] \times [1,2] = [-1,0] \times [3,4] \times \{6\} \times [1,2] \subset \mathbf{R}^4
\end{aligned}$$

which we shall not attempt to draw.

Definition 4.4 Let $Q = I_1 \times I_2 \times \cdots \times I_n \subset \mathbf{R}^n$ be an elementary cube. The *embedding number* of Q is denoted by $\operatorname{emb} Q$ and is defined to be n since $Q \subset \mathbf{R}^n$. The *dimension* of Q is denoted by $\dim Q$ and is defined to be the number of nondegenerate intervals I_i which are used to define Q. Using this notation we can write

$$\mathcal{K}^n := \{Q \in \mathcal{K} \mid \operatorname{emb} Q = n\}.$$

Similarly, we will let

$$\mathcal{K}_d := \{Q \in \mathcal{K} \mid \dim Q = d\}$$

and

$$\mathcal{K}_d^n := \mathcal{K}_d \cap \mathcal{K}^n.$$

Example 4.5 Refering to the elementary cubes defined above we have that

$$\begin{aligned}
\operatorname{emb} Q_1 = 3 \quad &\text{and} \quad \dim Q_1 = 3 \\
\operatorname{emb} Q_2 = 3 \quad &\text{and} \quad \dim Q_2 = 2 \\
\operatorname{emb} Q_3 = 3 \quad &\text{and} \quad \dim Q_3 = 1 \\
\operatorname{emb} Q_4 = 3 \quad &\text{and} \quad \dim Q_4 = 0 \\
\operatorname{emb} Q_5 = 4 \quad &\text{and} \quad \dim Q_5 = 3
\end{aligned}$$

In particular, the reader should observe that the only general relation between the embedding number and the dimension of an elementary cube Q is that

$$0 \le \dim Q \le \operatorname{emb} Q. \tag{4.1}$$

Proposition 4.6 *Let $Q \in \mathcal{K}_d^n$ and $P \in \mathcal{K}_k^m$, then*

$$Q \times P \in \mathcal{K}_{d+k}^{n+m}.$$

Proof: Since $Q \in \mathcal{K}^n$ it can be written as the product of n elementary intervals, i.e.

$$Q = I_1 \times I_2 \times \ldots \times I_n.$$

Similarly, we can write

$$P = J_1 \times J_2 \times \ldots \times J_m$$

where each J_i is an elementary interval. Hence,

$$Q \times P = I_1 \times I_2 \times \ldots \times I_n \times J_1 \times J_2 \times \ldots \times J_m$$

which is a product of elementary intervals.

It is left to the reader to check that $\dim(Q \times P) = \dim Q + \dim P$. ■

It should clear from the proof of Proposition 4.6 that though they lie in the same space $Q \times P \neq P \times Q$.

Exercises

4.1 Prove that any elementary cube is closed.

4.1.2 Representable Sets

Elementary cubes will be the building blocks for the homology theory that we will develop, however for technical reasons it will useful to have additional sets to work with. For this reason we introduce the notion of open cubes.

Definition 4.7 Let I be an elementary interval. The associated *open elementary interval* is

$$\mathring{I} := \begin{cases} (l, l+1) & \text{if } I = [l, l+1], \\ [l] & \text{if } I = [l, l]. \end{cases}$$

We extend this definition to a general elementary cube $Q = I_1 \times I_2 \times \ldots \times I_n \subset \mathbf{R}^n$ by defining the associated *open elementary cube* as

$$\mathring{Q} := \mathring{I}_1 \times \mathring{I}_2 \times \ldots \times \mathring{I}_n.$$

Example 4.8 An important word of warning: *An open cube need not be an open set.* Consider for example $[1] \in \mathcal{K}_0^1$. This is a single point and hence a closed set (see Exercise 1.4). Of course, as was shown in Exercise 1.4 any interval of the form $(l, l+1)$ is an open subset of $\mathbf{R}$. Thus, if $I \in \mathcal{K}_1^1$, then the open elementary interval $\overset{\circ}{I} \subset \mathbf{R}$ is an open set.

Consider now the elementary cube $Q = [1,2] \times [3] \in \mathcal{K}_d^2$. The associated open elementary cube is $\overset{\circ}{Q} = (1,2) \times [3] \subset \mathbf{R}^2$ which is clearly not an open set.

We can generalize this example to the following Proposition.

Proposition 4.9 *Let $Q \in \mathcal{K}$. The associated open elementary cube $\overset{\circ}{Q}$ is an open set if and only if $Q \in \mathcal{K}_n^n$ for $n \geq 1$.*

Proof: Since Q is an elementary cube it is the product of elementary intervals $Q = I_1 \times I_2 \times \cdots \times I_n \subset \mathbf{R}^n$. Let $I_i = [a_i, b_i]$ where $a_i \in \mathbf{Z}$ and $b_i = a_i$ or $b_i = a_i + 1$. Let $x_i = \frac{a_i + b_i}{2}$. Observe that $x = (x_1, x_2, \ldots, x_n) \in \overset{\circ}{Q}$.

Assume that $Q \in \mathcal{K}_d^n$ where $d < n$. Then, there exists i_0 such that $I_{i_0} = [a_{i_0}]$ is a degenerate interval. Observe that for any $\epsilon > 0$, $B(x, \epsilon) \not\subset \overset{\circ}{Q}$. Therefore, $\overset{\circ}{Q}$ is not open.

On the other hand, if $Q \in \mathcal{K}_n^n$, then by Exercise 1.4 $\overset{\circ}{Q}$ is an open set. ■

Proposition 4.10 *We have the following properties*

(i) $\mathbf{R}^n = \bigcup \{ \overset{\circ}{Q} \mid Q \in \mathcal{K}^n \}$,

(ii) $A \subset \mathbf{R}^n$ *bounded implies that* $\operatorname{card} \{ Q \in \mathcal{K}^n \mid \overset{\circ}{Q} \cap A \neq \emptyset \} < \infty$,

(iii) If $P, Q \in \mathcal{K}^n$, then $\overset{\circ}{P} \cap \overset{\circ}{Q} = \emptyset$ or $P = Q$,

(iv) For every $Q \in \mathcal{K}$, $\operatorname{cl} \overset{\circ}{Q} = Q$,

(v) $Q \in \mathcal{K}^n$ implies that $Q = \bigcup \{ \overset{\circ}{P} \mid P \in \mathcal{K}^n \; \overset{\circ}{P} \subset Q \}$.

Proof: (i) Obviously $\bigcup\{\overset{\circ}{Q} \mid Q \in \mathcal{K}^n\} \subset \mathbf{R}^n$. To prove the opposite inclusion take an $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ and put

$$I_i := \begin{cases} [x_i, x_i] & \text{if } x_i \in \mathbf{Z}, \\ [\text{floor}\,(x_i), \text{floor}\,(x_i) + 1] & \text{otherwise.} \end{cases}$$

Then $\overset{\circ}{Q} := \overset{\circ}{I}_1 \times \overset{\circ}{I}_2 \times \ldots \times \overset{\circ}{I}_n$ is an open cube and $x \in \overset{\circ}{Q}$. This proves (i).

(ii) The proof is straightforward.

(iii) For elementary cubes of dimesion one the result is obvious. Also, it extends immediately to elementary cubes of dimension greater than one, because the intersection of Cartesian products of intervals is the Cartesian product of the intersections of the corresponding intervals.

(iv) Observe that $\overset{\circ}{Q} \subset Q$, therefore $\text{cl}\,\overset{\circ}{Q} \subset Q$. To prove the opposite inclusion take an $x = (x_1, x_2 \ldots x_n) \in Q$. Let $Q = [k_1, l_1] \times [k_2, l_2] \times \ldots \times [k_n, l_n]$ and put

$$\begin{aligned} A &:= \{i = 1, \ldots, n \mid x_i = k_i\}, \\ B &:= \{i = 1, \ldots, n \mid x_i = l_i\}. \end{aligned}$$

Define $y^j := (y^j_1, y^j_2, \ldots y^j_d) \in \mathbf{R}^n$ by

$$y^j_i := \begin{cases} x_i + \frac{1}{2n} & i \in A \setminus B, \\ x_i & i \in A \cap B \text{ or } i \notin A \cup B, \\ x_i - \frac{1}{2n} & i \in B \setminus A. \end{cases}$$

Then $y^j \in \overset{\circ}{Q}$ and $\lim_{j \to \infty} y^j = x$. It follows that $x \in \text{cl}\,\overset{\circ}{Q}$.

(v) Consider $Q = I_1 \times I_2 \times \ldots \times I_n$ and let $x = (x_1, x_2, \ldots, x_n) \in Q$. Define

$$J_i := \begin{cases} [x_i, x_i] & \text{if } x_i \text{ is an endpoint of } I_i \\ I_i & \text{otherwise.} \end{cases}$$

and put $P := J_1 \times J_2 \times \ldots \times J_n$ Then obviously $x \in \overset{\circ}{P}$ and $\overset{\circ}{P} \subset Q$. Hence x belongs to the right-hand-side of (v). ∎

Using open cubes we can define a class of topological spaces.

Definition 4.11 A set $Y \subset \mathbf{R}^n$ is *representable* if it is a finite union of open elementary cubes. The family of representable sets in $\mathbf{R}^n$ is denoted by $\mathcal{R}^n$.

As an immediate consequence of Proposition 4.10(v) we get

Proposition 4.12 *Every elementary cube is representable.*

Definition 4.13 The *open hull* of a set $A \subset \mathbf{R}^n$ is

$$\operatorname{oh}(A) := \bigcup \{ \overset{\circ}{Q} \mid Q \in \mathcal{K}, Q \cap A \neq \emptyset \}, \tag{4.2}$$

and the *closed hull* of A is

$$\operatorname{ch}(A) := \bigcup \{ Q \mid Q \in \mathcal{K}, \overset{\circ}{Q} \cap A \neq \emptyset \}. \tag{4.3}$$

Example 4.14 Consider the vertex $P = [0] \times [0] \in \mathbf{R}^2$. Then,

$$\operatorname{oh}(P) = \{ (x_1, x_2) \in \mathbf{R}^2 \mid -1 < x_i < 1 \}.$$

Generalizing this example leads to the following result.

Proposition 4.15 *Let $P = [a_1] \times \cdots \times [a_n] \in \mathbf{R}^n$ be an elementary vertex. Then,*

$$\operatorname{oh}(P) = (a_1 - 1, a_1 + 1) \times \cdots \times (a_n - 1, a_n + 1).$$

The names chosen for $\operatorname{oh}(A)$ and $\operatorname{ch}(A)$ are justified by the following proposition.

Proposition 4.16 *Assume $A \subset \mathbf{R}^n$. Then*

(i) $A \subset \operatorname{oh}(A)$ *and* $A \subset \operatorname{ch}(A)$.

(ii) The set $\operatorname{oh}(A)$ *is open and representable.*

(iii) The set $\operatorname{ch}(A)$ *is closed and representable.*

(iv) $\operatorname{oh}(A) = \bigcap \{ U \in \mathcal{R}^n \mid U \text{ is open and } A \subset U \}$

(v) $\operatorname{ch}(A) = \bigcap \{ B \in \mathcal{R}^n \mid B \text{ is closed and } A \subset B \}$. *In particular, if K is a cubical set such that $A \subset K$, then* $\operatorname{ch}(A) \subset K$.

(vi) $\operatorname{oh}(\operatorname{oh}(A)) = \operatorname{oh}(A)$ *and* $\operatorname{ch}(\operatorname{ch}(A)) = \operatorname{ch}(A)$.

(vii) If $y \in \operatorname{oh}(x)$, then $\operatorname{ch}(x) \subset \operatorname{ch}(y)$.

(viii) $Q \in \mathcal{K}^n$ *and* $x \in \overset{\circ}{Q}$ *implies that* $\operatorname{ch}(x) = Q$.

(ix) Let $Q \in \mathcal{K}^n$ and let $x, y \in \overset{\circ}{Q}$. Then, $\operatorname{oh}(x) = \operatorname{oh}(y)$ *and* $\operatorname{ch}(x) = \operatorname{ch}(y)$.

Proof: (i) That $A \subset \operatorname{ch}(A)$ follows directly from the definition and $A \subset \operatorname{oh}(A)$ follows from Proposition 4.10(v).

(ii) By Proposition 4.10(ii) the union in (4.2) is finite. Therefore the set $\operatorname{oh}(A)$ is representable. To prove that $\operatorname{oh}(A)$ is open we will show that it satisfies **??**. Let $P \in \mathcal{K}^d$ be such that $\overset{\circ}{P} \cap \operatorname{oh}(A) = \emptyset$. Assume that $P \cap \operatorname{oh}(A) \neq \emptyset$. Then there exists a $Q \in \mathcal{K}$ such that $Q \cap A \neq \emptyset$ and $P \cap \overset{\circ}{Q} \neq \emptyset$. Since P is representable, it follows from Proposition **??** that $\overset{\circ}{Q} \subset P$. Therefore $Q = \operatorname{cl} \overset{\circ}{Q} \subset P$, i.e. $P \cap A \neq \emptyset$. This means that $\overset{\circ}{P} \subset \operatorname{oh}(A)$, a contradiction. It follows that $\operatorname{oh}(A)$ is open.

(iii) The set $\operatorname{ch}(A)$ is closed since it is the finite union of closed sets. By Proposition 4.12 $\operatorname{ch}(A)$ is representable.

(iv) Observe that since $\operatorname{oh}(A)$ is open, representable and contains A,

$$\bigcap \{U \in \mathcal{R}^n \mid U \text{ is open and } A \subset U\} \subset \operatorname{oh}(A).$$

To show the opposite inclusion take an open set $U \in \mathcal{R}^n$ such that $A \subset U$. Let $x \in \operatorname{oh}(A)$. Then there exists a $Q \in \mathcal{K}$ such that $A \cap Q \neq \emptyset$ and $x \in \overset{\circ}{Q}$. It follows that $\emptyset \neq Q \cap U = \operatorname{cl} \overset{\circ}{Q} \cap U$, i.e. $\overset{\circ}{Q} \cap U \neq \emptyset$. By Proposition **??** $\overset{\circ}{Q} \subset U$, hence $x \in U$. This shows that $\operatorname{oh}(A) \subset U$ and since U is arbitrary,

$$\operatorname{oh}(A) \subset \bigcap \{U \in \mathcal{R}^n \mid U \text{ is open and } A \subset U\}.$$

(v) Since $\operatorname{ch}(A)$ is closed, representable and contains A,

$$\bigcap \{B \in \mathcal{R}^n \mid B \text{is closed and } A \subset B\} \subset \operatorname{ch}(A).$$

Let $K \in \mathcal{R}^n$ be a closed set which contains A. We will show that $\operatorname{ch}(A) \subset K$. For this end take an $x \in \operatorname{ch}(A)$. Then there exists a $Q \in \mathcal{K}$ such that $\overset{\circ}{Q} \cap A \neq \emptyset$ and $x \in Q$. It follows that $\overset{\circ}{Q} \cap K \neq \emptyset$ and consequently $\overset{\circ}{Q} \subset K$. Hence $Q \subset K$ and $x \in K$. This shows that $\operatorname{ch}(A) \subset K$ and since K is arbitrary,

$$\operatorname{ch}(A) \subset \bigcap \{B \in \mathcal{R}^n \mid B \text{is closed and } A \subset B\}.$$

(vi) This follows immediately from (iv) and (v).

(vii) Observe that since $y \in \operatorname{oh}(x)$, there exists a $P \in \mathcal{K}$ such that $y \in \overset{\circ}{P}$ and $x \in P$. Take a $z \in \operatorname{ch}(x)$. Then there exists a $Q \in \mathcal{K}$ such that $z \in Q$

and $x \in \overset{\circ}{Q}$. It follows that $\overset{\circ}{Q} \subset P$, hence also $Q \subset P$ and consequently $z \in P$, which proves (vii).

(viii) This is straightforward.

(ix) Let $z \in \operatorname{oh}(x)$. Then there exists a $Q \in \mathcal{K}$ such that $z \in \overset{\circ}{Q}$ and $x \in Q$. It follows that $\overset{\circ}{P} \subset Q$, i.e. $y \in Q$. Consequently $z \in \operatorname{oh}(y)$ and $\operatorname{oh}(x) \subset \operatorname{oh}(y)$. The same way one proves that $\operatorname{oh}(y) \subset \operatorname{oh}(x)$. The equality $\operatorname{ch}(x) = \operatorname{ch}(y)$ follows from (viii).

■

4.1.3 Cubical Sets

As was mentioned before elementary cubes will make up the basic building blocks for our homology theory. This leads to the following definition.

Definition 4.17 A set $X \subset \mathbf{R}^n$ is *cubical* if X can be written as a finite union of elementary cubes.

If $X \subset \mathbf{R}^n$ is a cubical set, then we shall adopt the following notation.

$$\mathcal{K}(X) := \{Q \in \mathcal{K} \mid Q \subset X\}$$

and

$$\mathcal{K}_k(X) := \{Q \in \mathcal{K}(X) \mid \dim Q = k\}.$$

Observe that if $Q \subset X$ and $Q \in \mathcal{K}$ then $\operatorname{emb} Q = n$, since $X \subset \mathbf{R}^n$. This in turn implies that $Q \in \mathcal{K}^n$ so to use the notation $\mathcal{K}^n(X)$ is somewhat redundant, but it serves to reminds us that $X \subset \mathbf{R}^n$. Therefore, when it is convenient we will write $\mathcal{K}^n_k(X)$. In analogy with graphs, the elements of $\mathcal{K}_0(X)$ are the *vertices* of X and the elements of $\mathcal{K}_1(X)$ are the *edges* of X. More generally, the elements of $\mathcal{K}_k(X)$ are the *k-cubes* of X.

Example 4.18 Consider the set $X = [0,1] \times [0,1] \times [0,1] \subset \mathbf{R}^3$. This is an elementary cube, and hence, is a cubical set. It is easy to check that

$$\begin{aligned}
\mathcal{K}_3(X) &= [0,1] \times [0,1] \times [0,1] \\
\mathcal{K}_2(X) &= \{[0] \times [0,1] \times [0,1], [1] \times [0,1] \times [0,1], \\
&\quad [0,1] \times [0] \times [0,1], [0,1] \times [1] \times [0,1], \\
&\quad [0,1] \times [0,1] \times [0], [0,1] \times [0,1] \times [1]\}
\end{aligned}$$

$$\begin{aligned}
\mathcal{K}_1(X) &= \{[0]\times[0]\times[0,1], [0]\times[1]\times[0,1],\\
&\quad [0]\times[0,1]\times[0], [0]\times[0,1]\times[1],\\
&\quad [1]\times[0]\times[0,1], [1]\times[1]\times[0,1],\\
&\quad [1]\times[0,1]\times[0], [1]\times[0,1]\times[1],\\
&\quad [0,1]\times[0]\times[0], [0,1]\times[0]\times[1],\\
&\quad [0,1]\times[1]\times[0], [0,1]\times[1]\times[1]\}\\
\mathcal{K}_0(X) &= \{[0]\times[0]\times[0], [0]\times[0]\times[1],\\
&\quad [0]\times[1]\times[0], [0]\times[1]\times[1],\\
&\quad [1]\times[0]\times[0], [1]\times[0]\times[1],\\
&\quad [1]\times[1]\times[0], [1]\times[1]\times[1]\}.
\end{aligned}$$

Example 4.19 It should be noted that the definition of a cubical set is extremely restrictive. For example, the unit circle $x^2 + y^2 = 1$ is not a cubical set. In fact, even a simple set such as a point may or may not be a cubical set. In particular consider the point $P = (x, y, z) \in \mathbf{R}^3$. P is a cubical set if and only if x, y, and z are all integers.

Proposition 4.20 *If $X \subset \mathbf{R}^n$ is cubical, then X is closed and bounded.*

Proof: By definition a cubical set is the finite union of elementary cubes. By Exercise 4.1 an elementary cube is closed and by Theorem 1.15 the finite union of closed sets is closed.

To show that X is bounded, let $Q \in \mathcal{K}(X)$ then $Q = I_1 \times I_2 \times \cdots \times I_n$ where $I_i = [l_i]$ or $I_i = [l_i, l_i + 1]$. Let

$$\rho(Q) = \max_{i=1,\dots n} \{|l_i| + 1\}$$

Now set $R = \max_{Q\in\mathcal{K}(X)} \rho(Q)$. Then $X \subset B(0, R)$. ■

Definition 4.21 Any $Q \in \mathcal{K}(X)$ is called a *face* of X and is denoted by $Q \preceq X$. Q is a *proper face* in X, denoted by $Q \prec X$, if there exists $P \in \mathcal{K}(X)$ such that $P \neq Q$ and $Q \preceq \mathcal{K}(P)$. If Q is not a proper face, then it is a *maximal face*. $\mathcal{K}_{\max}(X)$ is the set of maximal faces in X. A face which is a proper face of exactly one elementary cube is a *free face*.

Example 4.22 Let $X = [0,1]\times[0,1]\times[0,1]$. Then, $\mathcal{K}_0(X)\cup\mathcal{K}_1(X)\cup\mathcal{K}_2(X)$ is the set of proper faces. The set of free faces is given by $\mathcal{K}_2(X)$.

Example 4.23 Refering to the cubical set $X \subset \mathbf{R}^2$ shown in Figure 4.4. The following elementary cubes are free faces

$$[-1] \times [2]$$
$$[0,1] \times [0], \quad [0,1] \times [1], \quad [0] \times [0,1], \quad [1] \times [0,1]$$

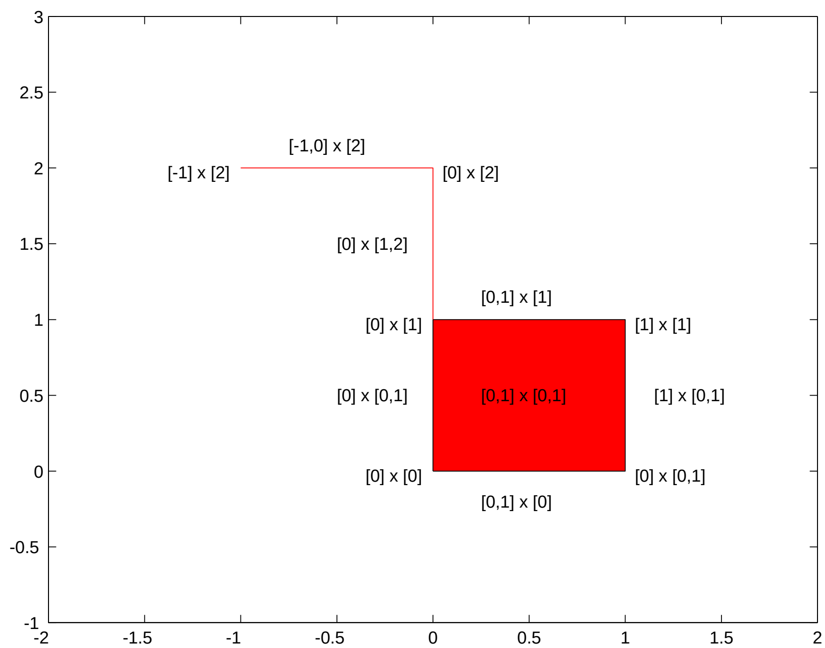


Figure 4.4: Elementary cubes of $X \subset \mathbf{R}^2$.

Exercises

4.2 In Example 4.19 it was noted that a given point need not be a cubical set. However, the set consisting of a point can be represented by a cubical set as follows. Let $X \subset \mathbf{R}^n$ consist of a single point, i.e. $X = \{x_0\}$. Let $f : X \to 0 \in \mathbf{R}^n$. Then, f is a homeomorphism and $f(X) = 0$ is a cubical set.

Prove that any abstract graph which is a tree can be represented as a cubical set.

4.3 Observe that any cubical set which consists of elementary cubes of dimension 0 or 1 is a graph and hence gives rise to an abstract graph. Give an example of an abstract graph which which does not arise as a cubical set.

4.2 The Algebra of Cubical Sets

In this section we finally present the formal definitions that we use to transition between the topology of a cubical set and the algebra of homology theory.

4.2.1 Cubical Chains

We begin by defining the algebraic objects of interest.

Definition 4.24 The group C_k of *k-dimensional chains* (k-chains for short) of X is the free abelian group generated by elements of $\mathcal{K}_k$, i.e.

$$C_k := \mathbf{Z}^{\mathcal{K}_k}.$$

If $c \in C_k$ *then* $\dim c := k$.

Observe that C_k is an infinitely generated free abelian group. In practice we will be interested in the chains generated by cubical sets.

Definition 4.25 Let $X \subset \mathbf{R}^n$ be a cubical set. $C_k(X)$ is the finitely generated free abelian group generated by the elements of $\mathcal{K}_k(X)$ and is refered to as the set of *k-chains of* X. Observe that $C_k(X)$ is a subgroup of C_k.

Recall from definition given in Chapter 3 that this implies that the basis for $C_k(X)$ is the set of functions $\widehat{Q} : \mathcal{K}_k(X) \to \mathbf{Z}$ defined by

$$\widehat{Q}(P) = \begin{cases} 1 & \text{if } P = Q \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Since $\mathcal{K}_k(X) = \emptyset$ for $k < 0$ and $k > n$, the corresponding group of k-chains is $C_k(X) = 0$.

Given an elementary cube Q we will refer to $\widehat{Q}$ as its dual elementary chain, and similarly, given an elementary chain $\widehat{Q}$ we will refer to Q as its dual elementary cube.

Let $\widehat{\mathcal{K}}_k(X) := \{\widehat{Q} \mid Q \in \mathcal{K}_k(X)\}$. Since X is a cubical set and $\widehat{\mathcal{K}}_k^n(X)$ is a basis for $C_k(X)$, $C_k(X)$ is finite dimensional. Furthermore, given any $c \in C_k(X)$ there are integers a_i such that

$$c = \sum_{\widehat{Q}_i \in \widehat{\mathcal{K}}_k(X)} a_i \widehat{Q}_i.$$

Definition 4.26 Let $c \in C_k(X)$ and let $c = \sum_{i=1}^m a_i\widehat{Q}_i$ where $a_i \neq 0$ for $i = 1, \ldots, m$. The *support* of the chain c is the cubical set

$$|c| := \bigcup_{i=1}^{m} Q_i \subset \mathbf{R}^n.$$

Proposition 4.27 *Support has the following properties:*

(i) $|0| = \emptyset$.

(ii) Let $a \in \mathbf{Z}$, *then*

$$|ac| = \begin{cases} \emptyset & \text{if } a = 0, \\ |c| & \text{if } a \neq 0. \end{cases}$$

(iii) If $Q \in \mathcal{K}$, *then* $|\widehat{Q}| = Q$.

(iv) $|c_1 + c_2| \subset |c_1| \cup |c_2|$.

Proof: (i) By definition the 0 chain is the element of the free abelian group which is not generated by any cube.

(ii) This follows directly from the definition of support and (i).

(iii) This too follows directly from the definition of chains and support.

(iv) Let $c_1 = \sum_{i=1}^m a_i\widehat{Q}_i$ and let $c_2 = \sum_{j=1}^l b_j\widehat{P}_i$ where $a_i, b_j \neq 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, l$. Then

$$c_1 + c_2 = \sum_{i=1}^{m} a_i\widehat{Q}_i + \sum_{j=1}^{l} b_j\widehat{P}_i.$$

Thus, $x \in |c_1 + c_2|$ implies $x \in |c_1|$ or $x \in |c_2|$. ∎

Example 4.28 It is not true in general that $|c_1 + c_2| = |c_1| \cup |c_2|$. Consider any chain c such that $|c| \neq \emptyset$. Observe that

$$\emptyset = |c - c| \neq |c| \cup |c| = |c| \neq \emptyset.$$

Notice that while a chain c is an algebraic object, its support $|c|$ is a set. Thus, we have just defined a way to go from a cubical set to a finite dimensional free group, and from an element of the free group back to a cubical set.

Proposition 4.29 *The map $\phi : \mathcal{K}_k \to \widehat{\mathcal{K}}_k$ given by $\phi(Q) = \widehat{Q}$ is a bijection.*

Proof: Since $\widehat{\mathcal{K}}_k$ is defined to be the image of ϕ it is obvious that ϕ is surjective. To prove injectivity assume that $P, Q \in \mathcal{K}_k$ and $\widehat{P} = \widehat{Q}$. This implies that

$$1 = \widehat{P}(P) = \widehat{Q}(P)$$

and hence that $P = Q$. ∎

Remark 4.30 While the notation we are using for chains is consistent with that of earlier chapters some care must be taken when discussing 0-chains that are generated by elementary cubes in **R**. Let $X \subset \mathbf{R}$ be a cubical set. Consider $\widehat{[1]} \in C_0(X)$. By definition it is the function

$$\widehat{[1]}(Q) = \begin{cases} 1 & \text{if } Q = [1] \\ 0 & \text{otherwise.} \end{cases}$$

while

$$2\widehat{[1]}(Q) = \begin{cases} 2 & \text{if } Q = [1] \\ 0 & \text{otherwise.} \end{cases}$$

This is different from $\widehat{[2]} \in C_0(X)$, since

$$\widehat{[2]}(Q) = \begin{cases} 1 & \text{if } Q = [2] \\ 0 & \text{otherwise.} \end{cases}$$

In particular $|\widehat{[1]}| = |2\widehat{[1]}| = 1 \in \mathbf{R}$ while $|\widehat{[2]}| = 2 \in \mathbf{R}$.

Finally, $0 \in C_k(X)$ is the identity element of the group and hence $|0| = \emptyset$, while $\widehat{[0]}$ is the dual of the vertex located at the origin, i.e. $|\widehat{0}| = 0 \in \mathbf{R}$.

Example 4.31 Let $c = \widehat{A}_2 - \widehat{A}_1 + \widehat{B}_1 - \widehat{B}_2$, where

$$A_1 = [0] \times [0,1],\ A_2 = [1] \times [0,1],\ B_1 = [0,1] \times [0],\ B_2 = [0,1] \times [1]$$

Then $|c|$ is the contour of the square $[0,1]^2$ shown on Figure 4.5. In addition we have chosen to give a geometric interpretion of the signs appearing in the expression for c. In particular, in Figure 4.5 we included an orientation to the edges indicated by the arrows. Thus, positive or negative elementary chains represent the direction in which an edge is traversed. For example, we think of $\widehat{A}_1$ as indicating moving along the edge from $(0,0)$ to $(0,1)$ while $-\widehat{A}_1$ suggests covering the edge in the opposite direction. With this in mind, c represents a counter-clockwise closed path around the square.

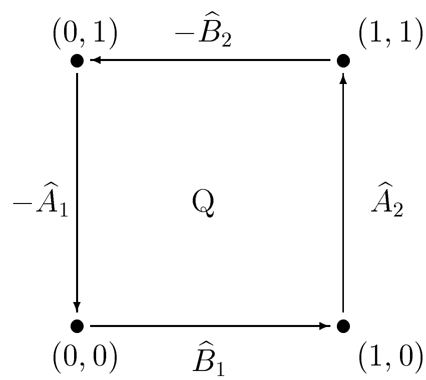


Figure 4.5: Boundary of the unit square.

Example 4.32 With the notation of the previous example, consider the chain $2c$. It is clear that $|2c| = |c|$ so both chains represent the same geometric object. The chain $2c$ can be interpreted as a path winding twice around the square in the counter-clockwise direction. Similarly, the chain

$$\widehat{A}_1 + \widehat{A}_2 + \widehat{B}_1 + \widehat{B}_2 = \widehat{A}_1 + \widehat{B}_2 + \widehat{A}_2 + \widehat{B}_1$$

could be interpreted as a "sum" of two different paths along the boundary of the square connecting $(0,0)$ to $(1,1)$.

Proposition 4.33 *If $K, L \subset \mathbf{R}^n$ are cubical sets, then*

$$C_k(K \cup L) = C_k(K) + C_k(L).$$

Proof: Let $\widehat{Q} \in \widehat{\mathcal{K}}_k(K)$. Then $Q \in \mathcal{K}_k(K)$ and hence $Q \in \mathcal{K}_k(K \cup L)$. The same argument applies to $\widehat{Q} \in \widehat{\mathcal{K}}_k(L)$ and so $C_k(K) + C_k(L) \subset C_k(K \cup L)$. To prove the opposite inclusion let $c \in C_k(K \cup L)$. In terms of the basis elements this can be written as

$$c = \sum_{i=1}^{m} a_i \widehat{Q}_i, \quad a_i \neq 0.$$

Let $A := \{i \mid Q_i \subset K\}$ and $B := \{1, 2, \ldots, m\} \setminus A$. Put $c_1 := \sum_{i \in A} a_i \widehat{Q}_i$, $c_2 := \sum_{i \in B} a_i \widehat{Q}_i$. Obviously $|c_1| \subset K$. Let $i \in B$. Then $Q_i \subset K \cup L$ and $Q_i \not\subset K$. In particular $\overset{\circ}{Q}_i \cap K = \emptyset$. Consequently $\overset{\circ}{Q}_i \subset L$ and since L is closed also $Q_i \subset L$. Hence $|c_2| \subset L$. It follows that $c = c_1 + c_2 \in C_k(K) + C_k(L)$. ■

From Proposition 4.6 we know that the product of two elementary cubes is again an elementary cube. This motivates the following definition.

Definition 4.34 Given two elementary cubes $P \in \mathcal{K}_k$ and $Q \in \mathcal{K}_{k'}$ set

$$\widehat{P} \diamond \widehat{Q} := \widehat{P \times Q}.$$

We can extend this product to $\diamond : C_k \times C_{k'} \to C_{k+k'}$ as follows. Let $c_1 \in \widehat{\mathcal{K}}_k$ and let $c_2 \in \widehat{\mathcal{K}}_{k'}$. By definition we can write

$$c_1 = \sum a_i \widehat{P_i} \quad \text{and} \quad c_2 = \sum b_j \widehat{Q_j}$$

where $\{P_i\} = \mathcal{K}_k$ and $\{Q_j\} = \mathcal{K}_{k'}$. Define

$$c_1 \diamond c_2 := \sum_{i,j} a_i b_j \widehat{P_i \times Q_j}.$$

The element $c \diamond c_2 \in C_{k+k'}$ is called the *cubical product* of c_1 and c_2.

Example 4.35 Let

$$P_1 = [0] \times [0,1],\ P_2 = [1] \times [0,1],\ P_3 = [0,1] \times [0],\ P_4 = [0,1] \times [1]$$

then $\widehat{P_i} \in \widehat{\mathcal{K}}_1$. Let $Q_1 = [-1,0]$ and $Q_2 = [0,1]$, then $\widehat{Q}_i \in \widehat{\mathcal{K}}_1$. This gives rise to chains $c_1 = P_1 + P_2 + P_3 + P_4$ and $c_2 = Q_1 + Q_2$. By definition we have

$$\begin{aligned} c_1 \diamond c_2 &= \widehat{P_1 \times Q_1} + \widehat{P_2 \times Q_1} + \widehat{P_3 \times Q_1} + \widehat{P_4 \times Q_1} + \\ &\quad \widehat{P_1 \times Q_2} + \widehat{P_2 \times Q_2} + \widehat{P_3 \times Q_2} + \widehat{P_4 \times Q_2} \end{aligned}$$

while

$$\begin{aligned} c_2 \diamond c_1 &= \widehat{Q_1 \times P_1} + \widehat{Q_1 \times P_2} + \widehat{Q_1 \times P_3} + \widehat{Q_1 \times P_4} + \\ &\quad \widehat{Q_2 \times P_1} + \widehat{Q_2 \times P_2} + \widehat{Q_2 \times P_3} + \widehat{Q_2 \times P_4} \end{aligned}$$

Figure 4.6 indicates the support of the chains c_1, c_2, $c_1 \diamond c_2$ and $c_2 \diamond c_1$.

The cubical product has the following properties.

Proposition 4.36 *Let c_1, c_2, c_3 be chains. Then*

(i) $c_1 \diamond 0 = 0 \diamond c_1 = 0$

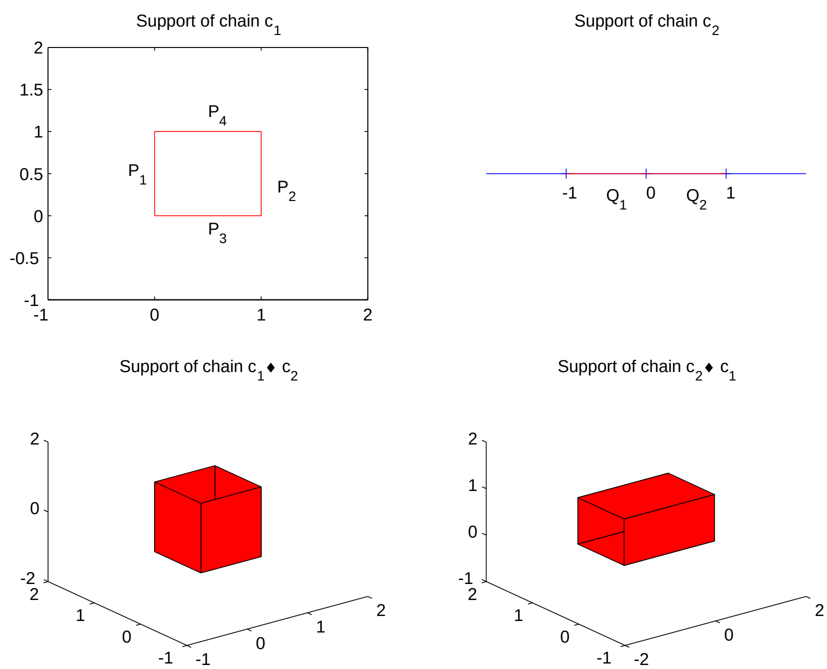


Figure 4.6: The support of the chains c, c', $c \diamond c'$ and $c' \diamond c$.

(ii) $c_1 \diamond (c_2 + c_3) = c_1 \diamond c_2 + c_1 \diamond c_3$

(iii) $(c_1 \diamond c_2) \diamond c_3 = c_1 \diamond (c_2 \diamond c_3)$

(iv) *if* $c_1 \diamond c_2 = 0$, *then* $c_1 = 0$ *or* $c_2 = 0$.

Proof: (i) and (ii) follow immediately from the definition.
(iii) The proof is straightforward.
(iv) Assume that $c_1 = \sum_{i=1}^{k} a_i \widehat{P}_i$ and $c_2 = \sum_{j=1}^{l} b_j \widehat{Q}_j$. Then

$$\sum_{i=1}^{k} \sum_{j=1}^{l} a_i b_j \widehat{P}_i \diamond \widehat{Q}_j = 0,$$

i.e. $a_i b_j = 0$ for any $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, l$. It follows that

$$(\sum_{i=1}^{k} a_i^2)(\sum_{j=1}^{l} b_j^2) = \sum_{i=1}^{k} \sum_{j=1}^{l} (a_i b_j)^2 = 0,$$

hence $\sum_{i=1}^k a_i^2 = 0$ or $\sum_{j=1}^l b_j^2 = 0$. Consequently $c_1 = 0$ or $c_2 = 0$. ■

Proposition 4.37 *Let $\widehat{Q}$ be an elementary cubical chain such that* $\operatorname{emb} Q > 1$. *Then, there exist unique elementary cubical chains $\widehat{I}$ and $\widehat{P}$ with* $\operatorname{emb} I = 1$ *and* $\operatorname{emb} P = d - 1$ *such that*

$$\widehat{Q} = \widehat{I} \diamond \widehat{P}.$$

Proof: Since $\widehat{Q}$ is an elementary cubical chain, Q is an elementary cube, i.e.

$$Q = I_1 \times I_2 \times \cdots \times I_n.$$

Set $I = I_1$ and $P := I_2 \times I_3 \times \cdots \times I_n$, then $\widehat{Q} = \widehat{I} \diamond \widehat{P}$.

We still need to prove that this is the unique decomposition. If $\widehat{Q} = \widehat{J} \diamond \widehat{P'}$ for some $J \in \mathcal{K}^1$ and $P' \in \mathcal{K}^{n-1}$ then $\widehat{I_1 \times P} = \widehat{J \times P'}$ and from Proposition 4.29 we obtain $I_1 \times P = J \times P'$. Since $I_1, J \subset \mathbf{R}$, it follows that $I_1 = J$ and $P = P'$. ■

4.2.2 The Boundary Operator

Given a cubical set $X \subset \mathbf{R}^n$, the chains $C_k(X)$ are the free groups which will be used to define the homology groups. To obtain a free chain complex we need to define boundary operators, i.e. linear maps $\partial_k : C_k(X) \to C_{k-1}^n(X)$ with the property that $\partial_k \circ \partial_{k+1} = 0$. Since ∂_k is supposed to be linear and $C_k(X)$ is a free group it is sufficient to give the definition in terms of the basis elements of $C_k(X)$.

At times the notation ∂_k is too cumbersome, so we will typically simplify it to ∂.

Definition 4.38 The *cubical boundary operator*

$$\partial_k : C_k \to C_{k-1}$$

is defined by induction on the embedding number. Notice that if ∂ is a linear map then it must be the case that

$$\partial\, 0 := 0.$$

Let $\widehat{Q} \in \widehat{\mathcal{K}}_k^1$, then Q is an elementary interval and hence $Q = [l] \in \mathcal{K}_0^1$ or $Q = [l, l+1] \in \mathcal{K}_1^1$ for some $l \in \mathbf{Z}$. Define

$$\partial \widehat{Q} := \begin{cases} 0 & \text{if } Q = [l], \\ \widehat{[l+1]} - \widehat{[l]} & \text{if } Q = [l, l+1]. \end{cases}$$

Now assume that $\widehat{Q} \in \widehat{\mathcal{K}}_k^n$ where $n > 1$. By Proposition 4.37 there exist unique elementary cubical chains $\widehat{I}, \widehat{P}$ with $\operatorname{emb} I = 1$ and $\operatorname{emb} P = n - 1$ such that

$$\widehat{Q} = \widehat{I} \diamond \widehat{P}.$$

Define

$$\partial \widehat{Q} := \partial \widehat{I} \diamond \widehat{P} + (-1)^{\dim I} \widehat{I} \diamond \partial \widehat{P}.$$

Finally, we extend the definition to all chains by linearity, i.e. if $c = a_1 \widehat{Q}_1 + a_2 \widehat{Q}_2 + \cdots + a_m \widehat{Q}_m$ then

$$\partial c := a_1 \partial \widehat{Q}_1 + a_2 \partial \widehat{Q}_2 + \cdots + a_m \partial \widehat{Q}_m.$$

Example 4.39 Let $Q = [l] \times [l']$. Then,

$$\begin{aligned} \partial \widehat{Q} &= \partial \widehat{[l]} \diamond \widehat{[l']} + (-1)^{\dim \widehat{[l]}} \widehat{[l]} \diamond \partial \widehat{[l']} \\ &= 0 \diamond \widehat{[l']} + \widehat{[l]} \diamond 0 \\ &= 0 + 0. \end{aligned}$$

Thus, the boundary of the dual to a vertex is trivial. This matches our intuitive notions developed for graphs.

Example 4.40 Let $Q = [l, l+1] \times [l', l'+1]$. Then,

$$\begin{aligned} \partial \widehat{Q} &= \partial \widehat{[l, l+1]} \diamond \widehat{[l', l'+1]} + (-1)^{\dim \widehat{[l,l+1]}} \widehat{[l, l+1]} \diamond \partial \widehat{[l', l'+1]} \\ &= (\widehat{[l+1]} - \widehat{[l]}) \diamond \widehat{[l', l'+1]} - \widehat{[l, l+1]} \diamond (\widehat{[l'+1]} - \widehat{[l']}) \\ &= \widehat{[l+1]} \diamond \widehat{[l', l'+1]} - \widehat{[l]} \diamond \widehat{[l', l'+1]} - \widehat{[l, l+1]} \diamond \widehat{[l'+1]} + \widehat{[l, l+1]} \diamond \widehat{[l']} \\ &= \widehat{[l+1] \times [l', l'+1]} - \widehat{[l] \times [l', l'+1]} + \widehat{[l, l+1] \times [l']} - \widehat{[l, l+1] \times [l'+1]}. \end{aligned}$$

Proposition 4.41 *Let c and c' be cubical chains, then*

$$\partial (c \diamond c') = \partial c \diamond c' + (-1)^{\dim c} c \diamond \partial c'.$$

Proof: Since ∂ is a linear operator it is sufficient to prove the proposition for elementary cubical chains, i.e. to show that

$$\partial(\widehat{Q} \diamond \widehat{Q'}) = \partial\widehat{Q} \diamond \widehat{Q'} + (-1)^{\dim Q}\widehat{Q} \diamond \partial\widehat{Q'}.$$

The proof will be done by induction on the embedding dimension of the corresponding cubes.

If $n = 1$, then the result follows from calculations similar to those of Example 4.40.

If $n > 1$, then we can decompose Q or Q' as in Proposition 4.37. Assume that it is Q that can be decomposed, i.e. $Q = I \times P$ where $\operatorname{emb} I = 1$ and $\operatorname{emb} P = n - 1$. Then,

$$\begin{aligned}
\partial(\widehat{Q} \diamond \widehat{Q'}) &= \partial(\widehat{I} \diamond \widehat{P} \diamond \widehat{Q'}) \\
&= \partial\widehat{I} \diamond \widehat{P} \diamond \widehat{Q'} + (-1)^{\dim I}\widehat{I} \diamond \partial(\widehat{P} \diamond \widehat{Q'}) \\
&= \partial\widehat{I} \diamond \widehat{P} \diamond \widehat{Q'} + (-1)^{\dim I}\widehat{I} \diamond \left(\partial\widehat{P} \diamond \widehat{Q'} + (-1)^{\dim P}\widehat{P} \diamond \partial\widehat{Q'}\right) \\
&= \partial\widehat{I} \diamond \widehat{P} \diamond \widehat{Q'} + (-1)^{\dim I}\widehat{I} \diamond \partial\widehat{P} \diamond \widehat{Q'} + (-1)^{\dim I + \dim P}\widehat{I} \diamond \widehat{P} \diamond \partial\widehat{Q'} \\
&= \left(\partial\widehat{I} \diamond \widehat{P} + (-1)^{\dim I}\widehat{I} \diamond \partial\widehat{P}\right) \diamond \widehat{Q'} + (-1)^{\dim Q}\widehat{Q} \diamond \partial\widehat{Q'} \\
&= \partial\widehat{Q} \diamond \widehat{Q'} + (-1)^{\dim Q}\widehat{Q} \diamond \partial\widehat{Q'}
\end{aligned}$$

■

Corollary 4.42 *If $\widehat{Q}_1, \widehat{Q}_2, \ldots, \widehat{Q}_m$ are elementary cubical chains, then*

$$\partial(\widehat{Q}_1 \diamond \widehat{Q}_2 \diamond \cdots \diamond \widehat{Q}_m) = \sum_{j=1}^{m} (-1)^{\sum_{i=1}^{j-1} \dim Q_i} \widehat{Q}_1 \diamond \cdots \diamond \widehat{Q}_{j-1} \diamond \partial\widehat{Q}_j \diamond \widehat{Q}_{j+1} \diamond \cdots \diamond \widehat{Q}_m.$$

As was indicated earlier we are really interested in $\partial_k : C_k(X) \to C_{k-1}(X)$ where X is a cubical set.

Definition 4.43 The boundary operator for the cubical set X is defined to be

$$\partial_k : C_k(X) \to C_{k-1}(X)$$

obtained by restricting $\partial : C_k \to C_{k-1}$ to $C_k(X)$.

Before we can employ this definition we need to be sure that $\partial_k\left(C_k(X)\right) \subset C_{k-1}(X)$. Observe that since ∂ is a linear operator the following proposition suffices.

Proposition 4.44 *Let $Q \subset \mathbf{R}^n$ be an elementary cube, then*

$$\partial_k : C_k(Q) \to C_{k-1}(Q).$$

Proof: Let $Q = I_1 \times I_2 \times \cdots \times I_n$. By Corollary 4.42

$$\partial(\widehat{Q}) = \sum_{j=1}^{m} (-1)^{\sum_{i=1}^{j-1} \dim I_i} \widehat{I}_1 \diamond \cdots \diamond \widehat{I}_{j-1} \diamond \partial \widehat{I}_j \diamond \widehat{I}_{j+1} \diamond \cdots \diamond \widehat{I}_m.$$

Consider each term of this sum separately. If I_j is a degenerate interval, then

$$\widehat{I}_1 \diamond \cdots \diamond \widehat{I}_{j-1} \diamond \partial \widehat{I}_j \diamond \widehat{I}_{j+1} \diamond \cdots \diamond \widehat{I}_m = 0 \in C_{k-1}(Q).$$

On the other hand if I_j is nondegenerate, then $I_j = [l_j, l_j + 1]$. This implies that

$$\begin{aligned}
\widehat{I}_1 \diamond \cdots \diamond \widehat{I}_{j-1} \diamond \partial \widehat{I}_j \diamond \widehat{I}_{j+1} \diamond \cdots \diamond \widehat{I}_m &= \widehat{I}_1 \diamond \cdots \diamond (\widehat{[l_j + 1]} - \widehat{[l_j]}) \diamond \cdots \diamond \widehat{I}_m \\
&= \widehat{I}_1 \diamond \cdots \diamond \widehat{[l_j + 1]} \diamond \cdots \diamond \widehat{I}_m + \\
&\quad \widehat{I}_1 \diamond \cdots \diamond \widehat{[l_j]} \diamond \cdots \diamond \widehat{I}_m
\end{aligned}$$

Both terms on the right side are in $C_{k-1}(Q)$ since

$$I_1 \times \cdots \times I_{j-1} \times [l_j] \times I_{j+1} \cdots \times I_n \subset Q$$

and

$$I_1 \times \cdots \times I_{j-1} \times [l_j + 1] \times I_{j+1} \cdots \times I_n \subset Q.$$

■

The following proposition shows that ∂ is a boundary operator.

Proposition 4.45

$$\partial \circ \partial = 0$$

Proof: Because ∂ is a linear operator it is enough to verify this property for elementary cubical chains. Again, the proof is by induction on the embedding number.

Let Q be an elementary interval. If $Q = [l]$, then by definition $\partial\widehat{Q} = 0$ so $\partial(\partial\widehat{Q}) = 0$. If $Q = [l, l+1]$, then

$$\begin{aligned}
\partial(\partial\widehat{Q}) &= \partial(\partial\widehat{[l, l+1]}) \\
&= \partial(\widehat{[l+1]} - \widehat{[l]}) \\
&= \partial\widehat{[l+1]} - \partial\widehat{[l]} \\
&= 0 - 0 \\
&= 0.
\end{aligned}$$

Now assume that $Q \in \mathcal{K}^n$ for $n > 1$. Then by Proposition 4.37 we can write $Q = I \times P$ where $\operatorname{emb} I = 1$ and $\operatorname{emb} P = n-1$. So

$$\begin{aligned}
\partial(\partial\widehat{Q}) &= \partial(\partial(\widehat{I \times P})) \\
&= \partial(\partial(\widehat{I} \diamond \widehat{P})) \\
&= \partial\left(\partial\widehat{I} \diamond \widehat{P} + (-1)^{\dim \widehat{I}}\widehat{I} \diamond \partial\widehat{P}\right) \\
&= \partial\left(\partial\widehat{I} \diamond \widehat{P}\right) + (-1)^{\dim \widehat{I}}\partial\left(\widehat{I} \diamond \partial\widehat{P}\right) \\
&= \partial\partial\widehat{I} \diamond \widehat{P} + (-1)^{\dim \partial\widehat{I}}\partial\widehat{I} \diamond \partial\widehat{P} + (-1)^{\dim \widehat{I}}\left(\partial\widehat{I} \diamond \partial\widehat{P} + \widehat{I} \diamond \partial\partial\widehat{P}\right) \\
&= (-1)^{\dim \partial\widehat{I}}\partial\widehat{I} \diamond \partial\widehat{P} + (-1)^{\dim \widehat{I}}\partial\widehat{I} \diamond \partial\widehat{P}.
\end{aligned}$$

The last step uses the induction hypothesis that the proposition is true if the embedding number is less than n.

Observe that if $\dim \widehat{I} = 0$, then $\partial\widehat{I} = 0$ in which case we have that each term in the sum is 0 and hence $\partial\partial\widehat{Q} = 0$. On the other hand, if $\dim \widehat{I} = 1$, then $\dim \partial\widehat{I} = 0$ and hence the two terms cancel each other giving the desired result. ∎

4.2.3 Homology of Cubical Sets

Let $X \subset \mathbf{R}^n$ be a cubical set. Then $\mathcal{K}(X)$ generates the cubical k-chains $C_k(X)$ and $\partial_k : C_k(X) \to C_{k-1}(X)$ is a boundary operator. Thus we can make the following definition.

Definition 4.46 The *cubical chain complex* for the cubical set $X \subset \mathbf{R}^n$ is

$$\mathcal{C}(X) := \{C_k(X), \partial_k\}$$

where $C_k(X)$ are the cubical k-chains generated by $\mathcal{K}(X)$ and ∂_k is the cubical boundary operator.

This allows us to immediately define the homology of X.

Definition 4.47 Let $X \subset \mathbf{R}^n$ be a cubical set. The *cubical k-cycles* of X are the elements of the subgroup

$$Z_k(X) := \ker \partial_k \subset C_k(X).$$

The *cubical k-boundaries* of X are the elements of the subgroup

$$B_k(X) := \operatorname{image} \partial_{k+1} \subset C_k(X).$$

The *cubical homology groups* of X are the quotient groups

$$H_k(X) := Z_k(X)/B_k(X).$$

We finish this section with the computation of the homology of two extremely simple cubical spaces.

Example 4.48 Let $X = \emptyset$. Then $C_k(X) = 0$ for all k and hence

$$H_k(X) = 0 \quad k = 0, 1, 2, \ldots$$

Example 4.49 Let $X = \{x_0\} \subset \mathbf{R}^n$ be a cubical set consisting of a single point. Then $x_0 = [l_1] \times [l_2] \times \cdots \times [l_n]$. Thus,

$$C_k(X) = \begin{cases} \mathbf{Z} & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, $Z_0(X) = C_0(X) = \mathbf{Z}$. Since $C_1 = 0$, $B_0 = 0$ and therefore, $H_0(X) \cong \mathbf{Z}$. Since, $C_k(X) = 0$ for all $k \geq 1$, $H_k(X) = 0$ for all $k \geq 1$. Therefore,

$$H_k(x_0) \cong \begin{cases} \mathbf{Z} & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Example 4.50 Recall the cubical set

$$\Gamma^1 = [0] \times [0, 1] \cup [1] \times [0, 1] \cup [0, 1] \times [0] \cup [0, 1] \times [1]$$

The set of elementary cubes is

$$\begin{aligned}
\mathcal{K}_0(\Gamma^1) &= \{[0]\times[0],[1]\times[0],[1]\times[0],[1]\times[1]\} \\
\mathcal{K}_1(\Gamma^1) &= \{[0]\times[0,1],[1]\times[0,1],[0,1]\times[0],[0,1]\times[1]\}
\end{aligned}$$

Thus, the bases for the sets of chains

$$\begin{aligned}
\widehat{\mathcal{K}}_0(\Gamma^1) &= \{\widehat{[0]\times[0]},\widehat{[0]\times[1]},\widehat{[1]\times[0]},\widehat{[1]\times[1]}\} \\
&= \{[0]\diamond[0],[0]\diamond[1],[1]\diamond[0],[1]\diamond[1]\} \\
\widehat{\mathcal{K}}_1(\Gamma^1) &= \{\widehat{[0]\times[0,1]},\widehat{[1]\times[0,1]},\widehat{[0,1]\times[0]},\widehat{[0,1]\times[1]}\} \\
&= \{[0]\diamond[0,1],[1]\diamond[0,1],[0,1]\diamond[0],[0,1]\diamond[1]\}
\end{aligned}$$

To compute the boundary operator we need to compute the boundary of the basis elements.

$$\begin{aligned}
\partial([0]\diamond[0,1]) &= -[0]\diamond[0]+[0]\diamond[1] \\
\partial([1]\diamond[0,1]) &= -[1]\diamond[0]+[1]\diamond[1] \\
\partial([0,1]\diamond[0]) &= -[0]\diamond[0]+[1]\diamond[0] \\
\partial([0,1]\diamond[1]) &= -[0]\diamond[1]+[1]\diamond[1]
\end{aligned}$$

We can put this into the form of a matrix

$$\partial_1 = \begin{bmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

To understand $Z_1(\Gamma^1)$ we need to know $\ker \partial_1$, i. e. we need to solve the equation

$$\begin{bmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This in turn means solving

$$\begin{bmatrix} -\alpha_1-\alpha_3 \\ \alpha_1-\alpha_4 \\ -\alpha_2+\alpha_3 \\ \alpha_2+\alpha_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The only non-trivial solution to this is

$$\alpha_1 = -\alpha_2 = -\alpha_3 = \alpha_4.$$

Thus, we have that $\dim Z_1(\Gamma^1) = 1$ and is generated by

$$[0] \diamond [0,1] - [1] \diamond [0,1] - [0,1] \diamond [0] + [0,1] \diamond [1].$$

Since, $C_2(\Gamma^1) = 0$, $B_1(\Gamma^1) = 0$ and hence

$$H_1(\Gamma^1) = Z_1(\Gamma^1) \cong \mathbf{Z}.$$

As we learned in Chapter 3, solving for the quotient space $\mathbf{Z}_0(\Gamma^1)/B_0(\Gamma^1)$ is a little more difficult. While we could compute the Smith normal form we shall take a slightly different tack here and concentrate on equivalence classes. We begin with the observation that there is no solution to the equation

$$\begin{bmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This implies that $[0] \diamond [0] \notin B_0(\Gamma^1)$. On the other hand

$$\{[0] \diamond [0] + [0] \diamond [1], [0] \diamond [0] + [1] \diamond [0], [0] \diamond [0] + [1] \diamond [1]\} \subset B_0(\Gamma^1).$$

From this, given any element $u \in C_0(\Gamma^1)$ such that $u \neq \alpha[0] \diamond [0]$ for some $\alpha \in \mathbf{Z}$ one can show that $u + [0] \diamond [0] \in B_0(\Gamma^1)$. In particular, $H_0(\Gamma^1) = Z_0(\Gamma^1)/B_0(\Gamma^1)$ is generated by $[0] \diamond [0]$ and thus $\dim H_0(\Gamma^1) = 1$. In particular, we have proven that

$$H_k(\Gamma^1) \cong \begin{cases} \mathbf{Z} & \text{if } k = 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

We could continue in this fashion for a long time computing homology groups, but as the reader hopefully has already seen this is a rather time consuming process. Furthermore, even if one takes a simple set such as

$$X = [0,1] \times [0,1] \times [0,1] \times [0,1]$$

the number of elementary cubes is quite large and the direct computation of its homology is quite tedious. Thus, we need to develop more efficient methods.

Exercises

4.4 Let $\Gamma^2 = \operatorname{bd}[0,1]^3$ be the boundary of the unit cube. Determine the cubical complex $\mathcal{C}(\Gamma^2)$ and compute $H_*(\Gamma^2)$.

4.5 Let X be a cubical set obtained by removing the center cube $(1,2) \times (1,2) \times [0,1]$ from the solid rectangle $[0,3] \times [0,3] \times [0,1]$. Let $T = \operatorname{bd} X$ be its boundary. (compare this set with a torus discussed in Example 4.81.

(a) Prepare the data file for computing the chain complex $\mathcal{C}(X)$ of X by the program cubchain. Run the program to find $\mathcal{C}(X)$ and $H_*(X, \mathbf{Z}_p)$ for several values of p. Make a guess about $H_*(X)$.

(b) Determine $\mathcal{C}(T)$ and compute $H_*(P)$.

4.6 The figure L in the file labirynth.bmp is composed of a large but finite number of pixels so it is a cubical set. Run the Pilarczyk programs to find the homology of it. Open two gates (i.e. remove two pixels) in opposite walls of the labyrinth and again run the program to find the homology of what is left. Make a guess about the solvability of the labyrinth. i.e. a possibility of passing inside from one gate to another without crossing a wall.

4.3 $H_0(X)$

This Chapter began with a discussion of cubical sets. These are a very special class of topological spaces. We then moved on to the combinatorics and algebra associated with these spaces and defined the homology of a cubical set. However, we have not said anything about the relationships between homology groups of a cubical set and topological properties of the set. The following theorem is a first step in this direction. It says that the zero dimensional homology group measures the number of connected components of the cubical set.

Theorem 4.51 *Let X be a cubical set. Then $H_0(X)$ is a free abelian group. Furthermore, if $\{P_i \mid i = 1, \ldots, d\}$ is a collection of vertices in X consisting of one vertex from each connected component of X, then*

$$\{[\widehat{P}_i] \in H_0(X) \mid i = 1, \ldots, d\}$$

forms a basis for $H_0(X)$.

Proof: The proof consists of two steps: (1) identifing elementary cubes with the connected components, and (2) using this to prove the theorem.

Step 1. Let P and P' be vertices in X. Define the equivalence class $P \sim P'$ if there is a sequence of vertices $R_0, \ldots, R_m$ of X such that $P = R_0$, $P' = R_m$, and there exist elementary edges Q_k with vertices R_{k-1} and R_k. For each vertex P in X, let

$$C_P := \bigcup_{Q \sim P} \mathrm{oh}\,(Q) \cap X.$$

Observe that $P \sim Q$ implies that $C_P = C_Q$. Also, by Proposition 4.16(ii) C_P is open.

We will now show that if $P \not\sim Q$, then $C_P \cap C_Q = \emptyset$. The proof is by contradiction, so let $x \in C_P \cap C_Q$. In particular, $x \in X$. Since X is a cubical set there exists an elementary cube $S \subset X$ such that $x \in S$. We also know that $x \in \mathrm{oh}\,(P') \cap X$ and $x \in \mathrm{oh}\,(Q') \cap X$ where $P \sim P'$ and $Q \sim Q'$. This implies that $opS \subset \mathrm{oh}\,(P') \cap \mathrm{oh}\,(Q') \cap X$. Thus, $P', Q' \in S$. Since S is convex there exists a path from P' to Q' made up of edges of S. Therefore, $P' \sim Q'$, a contradiction.

Finally, we need to show that C_P is a connected component. We do this by showing that it is path connected. Let $x, y \in C_P$. Then there exist vertices

P and Q such that $P \sim Q$, $x \in \operatorname{oh}(P) \cap X$ and $y \in \operatorname{oh}(Q) \cap X$. Since X is cubical, there exists an elementary cube $S \subset X$ such that $x \in S \cap \operatorname{oh}(P)$. Observe that this implies that $P \in S$. However, S is convex so there is a line segment from x to P. Similarly, there exists a path from y to Q. Since $P \sim Q$ there exist a sequence of vertices $R_0, \ldots, R_m$ and edges Q_k as above. The union of the line segments and edges forms a path from x to y. Therefore, C_P is path connected.

We can now conclude that the sets C_{P_i}, $i = 1, \ldots, d$ are connected, open, and disjoint. Therefore, they represent all the connected components of X.

Step 2. First recall that $Z_0(X) = C_0(X)$. Therefore, $\widehat{P}_i$ is a cycle for each $i = 1, \ldots d$.

Let P be a vertex in X. Then, there exists j such that $P \in C_{P_j}$. By construction, this implies that $P \sim P_j$ and hence there exist edges Q_k which form a path from P to P_j. Consider the chain

$$c = \sum_{k=1}^{m} \widehat{Q}_k.$$

Then, $\partial c = \widehat{P}_j - \widehat{P}$ and hence

$$[\widehat{P}_j] = [\widehat{P}] \in H_0(X).$$

The final step is to show that each $\widehat{P}_i$ is a distinct basis element. To do this we need to show that

$$c = \sum_{j=1}^{d} \alpha_i \widehat{P}_i$$

is a boundary element if and only if each $\alpha_i = 0$. Obviously, if $c = 0$, then $c \in B_0(X)$. So assume that at least one scalar $\alpha_i \neq 0$ and assume that $c = \partial b$ for some $b \in C_1(X)$. We can write b as a sum of chains as follows

$$b = \sum_{i=1}^{d} b_i$$

where $|b_i| \subset C_{P_i}$. Observe that $|\partial b_i| \subset C_{P_i}$ and therefore, since

$$\partial b = \sum_{i=1}^{d} \partial b_i$$

it must be that $\partial b_i = \alpha_i \widehat{P}_i$.

We need to show that the only way this can happen is for $\alpha_i = 0$. To do this, let $\epsilon : C_0(X) \to \mathbf{Z}$ be the group homomorphism defined by $\epsilon(\widehat{P}) = 1$ for every vertex $P \in X$. Let Q be an elementary edge. Then, $\partial\widehat{Q} = \widehat{R}_1 - \widehat{R}_0$ where R_0 and R_1 are vertices. Observe that

$$\begin{aligned}
\epsilon(\partial\widehat{Q}) &= \epsilon(\widehat{R}_1 - \widehat{R}_0) \\
&= \epsilon(\widehat{R}_1) - \epsilon(\widehat{R}_0) \\
&= 1 - 1 \\
&= 0.
\end{aligned}$$

This implies that $\epsilon(\partial b_i) = 0$ and hence

$$0 = \epsilon(\partial b_i) = \epsilon(\alpha_i \widehat{P}_i) = \alpha_i \epsilon(\widehat{P}_i) = \alpha_i.$$

Thus, $\widehat{P}_i$ generates nontrivial homology and $[\widehat{P}_i] \neq [\widehat{P}_j]$ if $i \neq j$. ■

4.4 Elementary Collapses

As the reader might have realized by now, even very “simple” cubical sets contain a large number of elementary cubes. We shall now discuss a method that allows us to reduce the number of elementary cubes needed to compute the homology of the set.

Lemma 4.52 *Let X be a cubical set. Let $Q \in \mathcal{K}(X)$ be a free face and assume $Q \prec P$. Then, P is not the proper face of any other cube in $\mathcal{K}(X)$ and* $\dim Q = \dim P - 1$.

Proof: Assume $P \prec R$. Then $Q \prec R$ contradicting the uniqueness of P.

Assume $\dim Q < \dim P - 1$. Then there exists $R \in \mathcal{K}(X)$ different from Q and P such that $Q \prec R \prec P$. ■

Definition 4.53 Let Q be a free face in $\mathcal{K}(X)$ and let Q be a proper face of P. Let $\mathcal{K}'(X) := \mathcal{K}(X) \setminus \{Q, P\}$. Define

$$X' := \bigcup_{R \in \mathcal{K}'(X)} R.$$

Then X' is a cubical space obtained from $\mathcal{K}(X)$ via an *elementary collapse of Q through P.*

Example 4.54 Let $X = [0,1] \times [0,1] \subset \mathbf{R}^2$ (see Figure 4.7). Then

$$\begin{aligned}
\mathcal{K}_2(X) &= \{[0,1] \times [0,1]\} \\
\mathcal{K}_1(X) &= \{[0] \times [0,1], [1] \times [0,1], [0,1] \times [0], [0,1] \times [1]\} \\
\mathcal{K}_0(X) &= \{[0] \times [0], [0] \times [1], [1] \times [0], [1] \times [1]\}
\end{aligned}$$

There are four free faces, the elements of $\mathcal{K}_1(X)$. Let $Q = [0,1] \times [1]$, then $Q \prec P = [0,1] \times [0,1]$. If we let X' be the cubical space obtained from $\mathcal{K}(X)$ via the elementary collapse of Q through P, then $X' = [0] \times [0,1] \cup [1] \times [0,1] \cup [0,1] \times [0]$ and

$$\begin{aligned}
\mathcal{K}_1(X') &= \{[0] \times [0,1], [1] \times [0,1], [0,1] \times [0]\} \\
\mathcal{K}_0(X') &= \{[0] \times [0], [0] \times [1], [1] \times [0], [1] \times [1]\}
\end{aligned}$$

Observe that the free faces of $\mathcal{K}(X')$ are different from those of $\mathcal{K}(X')$. In particular, $[0]\times[1]$ and $[1]\times[1]$ are free faces with $[0]\times[1] \prec [0]\times[0,1]$. Let X'' be the space obtained by collapsing $[0]\times[1]$ through $[0]\times[0,1]$. Then,

$$\begin{aligned}\mathcal{K}_1(X'') &= \{[1]\times[0,1],[0,1]\times[0]\}\\ \mathcal{K}_0(X'') &= \{[0]\times[0],[1]\times[0],[1]\times[1]\}\end{aligned}$$

On $\mathcal{K}(X'')$ we can now perform an elementary collapse of $[1]\times[1]$ through $[1]\times[0,1]$ to obtain X''' where

$$\begin{aligned}\mathcal{K}_1(X'') &= \{[0,1]\times[0]\}\\ \mathcal{K}_0(X'') &= \{[0]\times[0],[1]\times[0],\}\end{aligned}$$

A final elementary collapse of $[1]\times[0]$ through $[0,1]\times[0]$ results in the single point $X'''' = [0]\times[0]$. Thus, through this procedure we have reduce a 2-cube to a single point.

Theorem 4.55 *Let X' be obtained from X via an elementary collapse of Q_0 through P_0. Then*

$$H_*(X') \cong H_*(X).$$

Proof: Let ∂' and ∂ denote the boundary operators on $C_*(X')$ and $C_*(X)$, respectively. Assume $\dim P_0 = k$. By Lemma 4.52, $\dim Q_0 = k-1$.

Observe that

$$C_n(X') = C_n(X) \quad n \neq k, k-1.$$

Therefore, the domain and range of ∂_n and ∂'_n remain the same except for $n \neq k+1, k, k-1, k-2$. Thus,

$$H_n(X') = H_n(X) \quad n \neq k+1, k, k-1, k-2.$$

By Lemma 4.52, $\widehat{P}_0 \notin B_k(X)$, thus $B_k(X) = B_k(X')$. This means that the

$$\operatorname{image} \partial = \operatorname{image} \partial'.$$

Therefore, $Z_{k+1}(X') = Z_{k+1}(X)$ which implies that

$$H_{k+1}(X') = H_{k+1}(X).$$

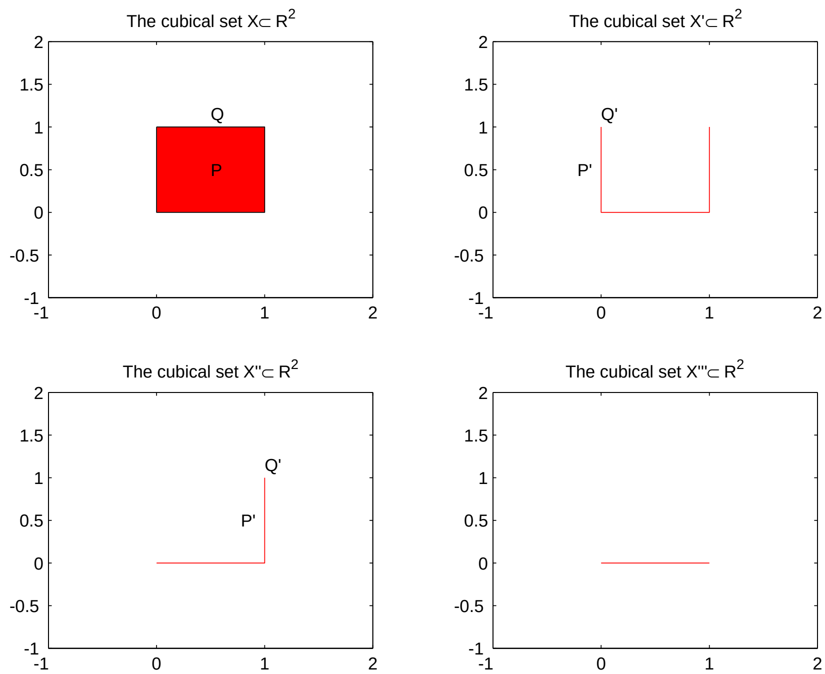


Figure 4.7: Sequence of Elementary Collapses of $[0,1] \times [0,1 \subset \mathbf{R}^2$.

Assume that

$$\partial_k \widehat{P}_0 = \widehat{Q}_0 + \sum_{i=1}^{k} a_i \widehat{R}_i, \tag{4.5}$$

where $R_i \neq Q_0$ and $a_i = \pm 1$ for all $i = 1, \ldots, k$. It should be noted that in writing this equation a choice has been made for orientations of $\widehat{P}_0$ and $\widehat{Q}_0$. The reader should check that the argument is, in fact, independent of this choice. Now

$$0 = \partial_{k-1} \circ \partial_k(\widehat{P}_0) = \partial_{k-1}(\widehat{Q}_0) + \partial_{k-1}\left(\sum_{i=1}^{k} a_i \widehat{R}_i\right).$$

This implies that

$$\partial_{k-1}(\widehat{Q}_0) = -\partial_{k-1}\left(\sum_{i=1}^{k} a_i \widehat{R}_i\right) \tag{4.6}$$

and hence, $B_{k-2}(X') = B_{k-2}(X)$. Therefore,

$$H_{k-2}(X') = H_{k-2}(X).$$

Now consider $c \in Z_k(X)$. We can write

$$c = \sum_{i=0}^{I} b_i \widehat{P}_i. \tag{4.7}$$

Then,

$$\begin{aligned} 0 &= \partial_k c \\ &= b_0 \partial_k \widehat{P}_0 + \sum_{i=1}^{I} b_i \partial_k (\widehat{P}_i) \\ &= b_0 \widehat{Q}_0 + b_0 \sum_{i=1}^{k} a_i \widehat{R}_i + \sum_{i=1}^{I} b_i \partial_k (\widehat{P}_i) \end{aligned}$$

where the last equality follows from (4.5). By Lemma 4.52, $\widehat{Q}_0$ does not appear in either of the summations. Thus, $b_0 = 0$. Observe that this means that

$$c = \sum_{i=1}^{I} b_i \widehat{P}_i$$

and hence that $c \in Z_k(X')$. This in turn implies that $Z_k(X') = Z_k(X)$. Since $B_k(X') = B_k(X)$.

$$H_k(X') = H_k(X).$$

The final step is to show that there exists a group isomorphism $f : H_{k-1}(X) \to H_{k-1}(X')$. We will do this as follows. Consider $\alpha \in H_{k-1}(X)$. Then $\alpha = [\theta]$ for some $\theta \in Z_{k-1}(X)$. We can write

$$\theta = b_0 \widehat{Q}_0 + \sum_{j=1}^{J} b_j \widehat{S}_j$$

where $S_j \neq Q_0$. Recall (4.5) and define

$$\theta' = -b_0 \sum_{i=1}^{k} a_i \widehat{R}_i + \sum_{j=1}^{J} b_j \widehat{S}_j.$$

Then, by (4.6)

$$[\theta] = \left[-\sum_{i=1}^{k} a_i \widehat{R}_i + \sum_{j=1}^{J} b_j \widehat{S}_j \right] = [\theta'] \in H_{k-1}(X). \tag{4.8}$$

But,

$$-\sum_{i=1}^{k} a_i \widehat{R}_i + \sum_{j=1}^{J} a_j \widehat{S}_j \in Z_{k-1}(X')$$

and thus we can view $[\theta'] \in H_{k-1}(X')$. So define

$$f([\theta]) = [\theta'].$$

It is straightforward to check that f is a group homomorphism, so all that remains is to show that it is an isomorphism. Since $Z_{k-1}(X') \subset Z_{k-1}(X)$ it is clear that f is surjective. To show it is a monomorphism assume that $\theta_1, \theta_2 \in Z_{k-1}(X)$ and that $f([\theta_1]) = f([\theta_2])$. The same argument that led to (4.8) shows that $[\theta_1] = [\theta_2] \in H_{k-1}(X)$. ■

Corollary 4.56 *Let $Y \subset X$ be cubical sets. Furthermore, assume that Y can be obtained from X via a series of elementary collapses, then*

$$H_*(Y) \cong H_*(X).$$

From Examples 4.54, 4.49 and Corollary 4.56 we can conclude that

$$H_k([0,1] \times [0,1]) \approx \begin{cases} \mathbf{Z} & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Up to this point the discussion of elementary collapses has been purely combinatorial and algebraic. We have not indicated how an elementary collapse is related to a topological operation. This is the purpose of following discussion.

Let $Q \subset \mathbf{R}^n$ be an elementary cube of the form

$$Q = I_1 \times \cdots \times I_n$$

where $I_i = [a_i, b_i]$ is an elementary interval. To simplify the formulas for the continuous maps that will be used we want to move Q to the origin. Thus we define the translation

$$T_Q(x_1, x_2, \ldots, x_n) = (x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n). \tag{4.9}$$

Let $P = T_Q(Q)$. Then, P has the form,

$$P = J_1 \times \cdots \times J_n$$

where $J_i = [0, d_i]$ and $d_i \in \{0,1\}$. If $\dim P > 0$, then there exists i_0 such that $d_{i_0} = 1$. Let

$$R = K_1 \times \cdot \times K_n$$

where

$$K_i = \begin{cases} [1] & \text{if } i = i_0, \\ J_i & \text{otherwise.} \end{cases}$$

The R is both a free and a proper face of P.

Lemma 4.57 *Let P' be obtained from P via the elementary collapse of R through P. Then, P' is a deformation retract of P.*

Proof: If $\dim P = 1$, then this is just restating the fact that a point is a deformation retract of an edge. So we can assume that $\dim P > 1$.

Let $\mathcal{I} = \{i \mid d_i = 1\} \setminus \{i_0\}$. Define $F : P \times [0,1] \to P$ by

$$F(x_1, \ldots, x_n, t) = \begin{cases} (x_1, \ldots, \left(2 \max_{i \in \mathcal{I}} \left\{x_i - \frac{1}{2}\right\}\right)^{\tan \frac{\pi}{2} t} x_{i_0}, \ldots, x_n) & \text{if } 0 \leq t < 1, \\ \lim_{t \to 1}(x_1, \ldots, \left(2 \max_{i \in \mathcal{I}} \left\{x_i - \frac{1}{2}\right\}\right)^{\tan \frac{\pi}{2} t} x_{i_0}, \ldots, x_n) & \text{if } t = 1. \end{cases} \tag{4.10}$$

Observe that $F(\cdot, 0) = \mathrm{id}_P$, $F|_{P' \times [0,1]} = \mathrm{id}_{P'}$, and $F(P, 1) \subset P'$. We leave it to the reader to check that F is continuous. ■

Proposition 4.58 *Let Q be an elementary cube. Let Q' be obtained from Q through an elementary collapse. Then Q' is a deformation retract of Q.*

Proof: Let $Q \subset \mathbf{R}^n$. Since Q is an elementary cube it has the form

$$Q = I_1 \times \cdots \times I_n$$

where $I_i = [a_i, b_i]$ is an elementary interval. Let S be the proper free face of Q such that Q' is obtain by the elementary collapse of S through Q. Then S has the form $S = J_1 \times \cdots \times J_n$ where

$$J_i = \begin{cases} [\alpha] & \text{if } i = i_0, \\ J_i & \text{otherwise.} \end{cases}$$

and $\alpha \in \{a_{i_0}, b_{i_0}\}$. We will present the proof in the case that $\alpha = b_{i_0}$. The case that $\alpha = a_{i_0}$ is left to the reader.

Define $G : Q \times [0,1] \to Q$ by

$$G(x,t) = T_Q^{-1}(F(T_Q(x),t)) \tag{4.11}$$

where F is given by (4.10) and T_Q is given by (4.9). That this is the desired deformation retraction follows from Lemma 4.57. ■

Proposition 4.59 *Let X be a cubical set. Let X' be obtained from X through an elementary collapse. Then X' is a deformation retract of X.*

Proof: Let X' be obtained by the elementary collapse of the proper free face S through the elementary cube Q. Define $H : X \times [0,1] \to X$ by

$$H(x,t) = \begin{cases} G(x,t) & \text{if } x \in Q \\ x & \text{otherwise,} \end{cases}$$

where G is given by (4.11). We leave it to the reader to check that H is continuous. ■

Exercises

4.7 Use the elementary collapses to show that the elementary cube $[0,1]^3$ is acyclic.

4.8 Let X be the solid cubical set discussed in Exercise **??**. Here is an alternative way of computing the homology of X: Use the elementary collapses of X onto the simple closed curve Γ defined as the union of four line segments $[1,2] \times [1] \times [0]$, $[2] \times [1,2] \times [0]$, $[1,2] \times [2] \times [0]$, $[1] \times [1,2] \times [0]$. Compute the homology of Γ and deduce what is the homology of X.

4.5 Acyclic Cubical Spaces

We finish this chapter with a class of important cubical sets; those which have trivial homology, i.e. the homology of a point

Definition 4.60 A cubical set X is *acyclic* if

$$H_k(X) \approx \begin{cases} \mathbf{Z} & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 4.61 *Elementary cubes are acyclic.*

Proof: Let $Q = I_1 \times I_2 \times \cdots \times I_n$ be an elementary cube. We can assume that $I_i = [0, b_i]$ where $b_i \in \{0, 1\}$. (If Q is not of this form, then use the translation T_Q to move it to the origin.)

The proof is by induction on the dimension of Q.

If $\dim Q = 0$, then the result follows from Example 4.49.

Now assume that the result is true for every elementary cube of dimension less than d and that $\dim Q = d$. Since, it is possible that $d < n$ not all elementary intervals need be nondegenerate. Let

$$\mathcal{J} := \{i \mid I_i = [0, 1]\}.$$

Let $m = \max\{i \in \mathcal{J}\}$.

Observe that

$$F := I_1 \times \cdots I_{m-1} \times [1] \times [0] \times \cdots \times [0]$$

is a free face. Let Q^1 be the cubical set obtained by collapsing F through Q. Q^1 can now be written as the union of $d-1$ dimensional elementary cubes. To be precise if $i \in \mathcal{J}$, set

$$G_i^0 := I_1 \times \cdots \times I_{i-1} \times [0] \times I_{i+1} \times \cdots \times I_{m-1} \times [0, 1] \times [0] \times \cdots \times [0]$$

and let

$$G_i^1 := I_1 \times \cdots \times I_{i-1} \times [1] \times I_{i+1} \times \cdots \times I_{m-1} \times [0, 1] \times [0] \times \cdots \times [0].$$

set

$$P = I_1 \times \cdots I_{m-1} \times [0] \times [0] \times \cdots \times [0].$$

Then

$$Q^1 = P \cup \bigcup_{i \in \mathcal{J}} G_i^\alpha$$

where $\alpha \in \{0,1\}$. Now observe that each G_i^α has a proper free face

$$F_i^1 := I_1 \times \cdots \times I_{i-1} \times [1] \times I_{i+1} \times \cdots \times I_{m-1} \times [1] \times [0] \times \cdots \times [0]$$

and

$$F_i^0 := I_1 \times \cdots \times I_{i-1} \times [0] \times I_{i+1} \times \cdots \times I_{m-1} \times [1] \times [0] \times \cdots \times [0].$$

Let Q^2 be the cubical set obtained by collapsing each F_i^α through G_i^α. Q^2 can be written as a union of P and $d-2$ dimensional elementary cubes. Again, to be precise, for each pair $i_1, i_2 \in \mathcal{J}$ with $i_1 < i_2$, let $\alpha = (\alpha_1, \alpha_2) \in \{0,1\}^2$ and set

$$\begin{aligned} G_{i_1,i_2}^\alpha \quad := \quad & I_1 \times \cdots \times I_{i_1-1} \times [\alpha_1] \times I_{i_1+1} \times \cdots I_{i_2-1} \times [\alpha_2] \times I_{i_2+1} \times \cdots \\ & \times I_{m-1} \times [0,1] \times [0] \times \cdots \times [0] \end{aligned}$$

Then,

$$Q^2 = P \cup \bigcup_{\substack{i_1, i_2 \in \mathcal{J} \\ i_1 < i_2 \\ \alpha \in \{0,1\}^2}} G_{i_1,i_2}^\alpha$$

Once again, each G_{i_1,i_2}^α has a free face

$$\begin{aligned} F_{i_1,i_2}^\alpha \quad := \quad & I_1 \times \cdots \times I_{i_1-1} \times [\alpha_1] \times I_{i_1+1} \times \cdots I_{i_2-1} \times [\alpha_2] \times I_{i_2+1} \times \cdots \\ & \times I_{m-1} \times [1] \times [0] \times \cdots \times [0] \end{aligned}$$

which allows for an elementary collapse. After k steps we have that

$$Q^k = P \cup \bigcup_{\substack{i_1, i_2, \ldots, i_k \in \mathcal{J} \\ i_1 < i_2 < \cdots < i_k \\ \alpha \in \{0,1\}^k}} G_{i_1,i_2,\ldots,i_k}^\alpha$$

where $G_{i_1,i_2,\ldots,i_k}^\alpha$ is the elementary cube of the form $J_1 \times \cdots \times J_n$ with

$$J_i = \begin{cases} [\alpha_j] & \text{if } i = i_j \in \{i_1, i_2, \ldots, i_k\} \\ [0] & \text{if } i \notin \mathcal{J} \\ [0,1] & \text{otherwise.} \end{cases}$$

Furthermore, $G^\alpha_{i_1,i_2,\ldots,i_k}$ has a proper free face $F^\alpha_{i_1,i_2,\ldots,i_k} = K_1 \times \cdots \times K_n$ of the form

$$K_i = \begin{cases} [\alpha_j] & \text{if } i = i_j \in \{i_1, i_2, \ldots, i_k\} \\ [0] & \text{if } i \notin \mathcal{J} \\ [1] & \text{if } i = m \\ [0,1] & \text{otherwise.} \end{cases}$$

After, d iterations we have that

$$Q^d = P$$

and by the induction step P is acyclic. ■

While the reduction process that we used in the previous proof is simple to implement, it is rather difficult to comprehend. Therefore, we would like to have conceptually easier way to conclude that a cubical set is acyclic. The following theorem provides us with such a method. As we shall see in Chapter 6 this is a simple version of a much more general and powerful theorem called the Meyer-Vietoris sequence.

Proposition 4.62 *Assume $X, Y \subset \mathbf{R}^n$ are cubical sets. If X, Y and $X \cap Y$ are acyclic, then $X \cup Y$ is acyclic.*

Proof: We will first prove that $H_0(X \cup Y) \approx \mathbf{Z}$. By Theorem 4.51 the assumption that X and Y are acyclic implies that X and Y are connected. $X \cap Y$ is acyclic implies that $X \cap Y \neq \emptyset$. Therefore, $X \cup Y$ is connected and hence by Theorem 4.51, $H_0(X \cup Y) \approx \mathbf{Z}$.

Now consider the case of $H_1(X \cup Y)$. Let $z \in Z_1(X \cup Y)$ be a cycle. We need to show that $z \in B_1(X \cup Y)$. By Proposition 4.33, $z = z_X + z_Y$ for some $z_X \in C_1(X)$ and $z_Y \in C_1(Y)$. Since z is a cycle, $\partial z = 0$. Thus,

$$\begin{aligned} 0 &= \partial z \\ &= \partial(z_X + z_Y) \\ &= \partial z_X + \partial z_Y \\ -\partial z_Y &= \partial z_X. \end{aligned}$$

Observe that $-\partial z_Y, \partial z_X \in C_0(Y \cap X) = Z_0(Y \cap X)$. From the assumption of acyclicity, $H_0(Y \cap X) \approx \mathbf{Z}$. Therefore, as an element of $H_0(Y \cap X)$, $[\partial z_X] = n \in \mathbf{Z}$.

We will now show that $n = 0$. $\partial z_X \in C_0(X \cap Y)$ implies that $\partial z_X = \sum a_i \widehat{P}_i$ where $P_i \in \mathcal{K}_0(X \cap Y)$. By Theorem 4.51, $[\partial z_X] = n \in \mathbf{Z}$ implies that

$\sum a_i = n$. Define the group homomorphism $\epsilon : C_0(X) \to \mathbf{Z}$ by $\epsilon(\widehat{P}) = 1$ for each $P \in \mathcal{K}_0(X \cup Y)$. Then for any $Q \in \mathcal{K}_0(X \cup Y)$ $\epsilon(\partial\widehat{Q}) = 0$. Therefore, $\epsilon(\partial z_X) = 0$, but

$$\epsilon(\partial z_X) = \sum a_i = n.$$

Therefore, $n = 0$.

Since $[\partial z_X] = 0 \in H_0(X \cup Y)$, there exists $b \in C_1(X \cap Y)$ such that $\partial b = \partial z_X$. Now observe that

$$\partial(-b + z_X) = -\partial b + \partial z_X = 0.$$

Therefore, $-b + z_X \in \mathbf{Z}_1(X)$. But, $H_1(X) = 0$ which implies that there exists $b_X \in C_2(X)$ such that $\partial b_X = -b + z_X$. The same argument shows that there exits $b_Y \in C_2(X)$ such that $\partial b_Y = b + z_Y$. Finally, observe that $b_X + b_Y \in C_2(X \cup Y$ and

$$\begin{aligned}
\partial(b_X + b_Y) &= \partial b_X + \partial b_Y \\
&= b + z_Y + -b + z_Y \\
&= z_Y + z_X \\
&= c.
\end{aligned}$$

Therefore, $c \in B_1(X \cup Y)$ which implies that $[z] = 0 \in H_1(X \cup Y)$. Therefore, $H_1(X \cup Y) = 0$.

We now show that $H_n(X \cup Y) \approx 0$ for all $n > 1$. Let $z \in Z_n(X \cup Y)$ be a cycle. Then by Proposition 4.33.2, $z = z_X + z_Y$ for some $z_X \in C_n(X)$ and $z_Y \in C_n(Y)$. Since z is a cycle, $\partial z = 0$. Thus,

$$\begin{aligned}
0 &= \partial z \\
&= \partial(z_X + z_Y) \\
&= \partial z_X + \partial z_Y \\
-\partial z_Y &= \partial z_X.
\end{aligned}$$

Of course, this does not imply that $\partial z_X = 0$. However, since $z_Y \in C_n(Y)$ and $z_X \in C_n(X)$ we can conclude that $-\partial z_Y, \partial z_X \in C_{n-1}(Y \cap X)$. Let $c = \partial z_X$. Since

$$\partial c = \partial \circ \partial z_X = 0$$

$c \in Z_{n-1}(Y \cap X)$.

Since $X \cap Y$ is acyclic, $H_{n-1}(X \cap Y) \approx 0$. Therefore, $c \in B_{n-1}(X \cap Y)$. i.e. there exists a $c' \in C_n(X \cap Y)$ such that $c = \partial c'$. It follows that $z_X - c' \in Z_n(X)$

and $z_Y + c' \in Z_n(Y)$. By the acyclicity of X and Y there exist $c'_X \in C_{n+1}(X)$ and $c'_Y \in C_{n+1}(Y)$ such that $z_X - c' = \partial c'_X$ and $z_Y - c' = \partial c'_Y$. Therefore

$$z = z_X + z_Y = \partial(c'_X + c'_Y) \in B_n(X \cup Y).$$

■

Proposition 4.63 *If $X \subset \mathbf{R}^n$ is a convex cubical set, then X is acyclic.*

Proof: Since X is a convex cubical set, it can be written as the product of intervals, i.e.

$$X = [a_1, b_1] \times \cdots \times [a_n, b_n]$$

where $a_i, b_i \in \mathbf{Z}$. (Note: we are not assuming that these are elementary intervals.) Let the dimension of X be d, the number of intervals such that $b_i > a_i$. The proof will be by induction, both on the dimension of the convex set and the number of d-dimensional elementary cubes in X.

Observe that if X is 1 dimensional, then X is a line segment in R^n, which is easily checked to be acyclic.

If X consists of a single d dimensional elementary cube, then by Proposition 4.61 X is acyclic.

So assume that there are q elementary d dimensional cubes in X and that the proposition is true for every convex cubical set with less than q elementary d dimesional cubes and every convex set of dimension less than d.

Observe that for some i_0, $b_{i_0} - a_{i_0} \geq 2$. If not, then X is an elementary cube. Let

$$X_1 := [a_1, b_1] \times \cdots \times [a_{i_0}, a_{i_0} + 1] \times \cdots \times [a_n, b_n]$$

and

$$X_1 := [a_1, b_1] \times \cdots \times [a_{i_0} + 1, b_{i_0}] \times \cdots \times [a_n, b_n].$$

Then, X_1, X_2, and $X_1 \cap X_2$ are convex cubical sets. Furthermore, since the number of d dimesional elementary cubes in X_1 and X_2 are less than q, X_1 and X_2 are acyclic. The dimesion of $X_1 \cap X_2$ is less than d, and hence by induction is also acyclic. The result follows from Proposition 4.62. ■

Since convex cubical sets are always the products of intervals they represent a small class of cubical sets. A slightly larger collection that is topologically simple is as follows.

Definition 4.64 A cubical set $X \subset \mathbf{R}^n$ is *starshaped* with respect to a point $x \in \mathbf{Z}^n$ if X is the union of a finite number of convex cubical sets each of which contains the point x.

Proposition 4.65 *Let X_i, $i = 1, \ldots, n$ be a collection of starshaped sets with respect to the same point x. Then,*

$$\bigcup_{i=1}^{n} X_i \quad \text{and} \quad \bigcap_{i=1}^{n} X_i$$

are starshaped.

Proof: Since X_i is starshaped be can write $X_i = \cup R_{i,j}$ where $R_{i,j}$ is convex and $x \in R_{i,j}$. Thus, if $X = \cup_i X_i$, then $X = \cup_{i,j} R_{i,j}$ and hence is starshaped.

So assume that $X = \cap_i X_i$. Then

$$\begin{aligned} X &= \bigcap_i X_i \\ &= \bigcap_i \left(\bigcup_j R_{i,j} \right) \\ &= \bigcup_j \left(\bigcap_i R_{i,j} \right). \end{aligned}$$

But, since $x \in R_{i,j}$ for each i, j, for each j, $\bigcap_i R_{i,j}$ is a convex and contains x. Again, this means that X is starshaped. ■

Proposition 4.66 *Every starshaped set is acyclic.*

Proof: Let X be a starshaped cubical set. Then, $X = \bigcup_{i=1}^{k} R_i$ where each R_i is a cubical convex set, there exists $x \in X$ such that $x \in R_i$ for all $i = 1, \ldots k$, and k is the minimal number of convex sets needed to obtain X. The proof is by induction on k.

If $k = 1$ then X is convex and hence by Propostion 4.63 is acyclic.

So assume that every starshaped cubical set which can be written as the union of $k - 1$ convex sets containing the same point is acyclic. Let $Y = \bigcup_{i=1}^{k-1} R_i$. Then by the induction hypothesis, Y is acyclic. R_k is convex and hence by Propostion 4.63 is acyclic. Furthermore, $R_i \cap R_k$ is convex for each $i = 1, \ldots, k-1$ and

$$Y \cap R_k = \bigcap_{i=1}^{k-1} (R_i \cap R_k).$$

Therefore, $Y \cap R_k$ is a starshaped. region which can be written in terms of $k-1$ convex sets. By the induction hypothesis it too is acyclic. Therefore, by Proposition 4.62, X is acyclic. ■

Proposition 4.67 *Assume that $\mathcal{C}$ is a family of rectangles in $\mathbf{R}^n$ such that the intersection of any two of them is non-empty. Then $\bigcap \mathcal{C}$ is non-empty.*

Proof: First consider the case when $d = 1$. Then rectangles become intervals. Let a denote the supremum of the set of left endpoints of the intervals and let b denote the infimum of the set of right endpoints. We cannot have $b < a$, because then one can find two disjoint intervals in the family. Therefore $\emptyset \neq [a, b] \subset \bigcap \mathcal{C}$.

If $d > 1$ then each rectangle is a Cartesian product of intervals, the intersection of all rectangles is the Cartesian product of the intersections of the corresponding intervals, and the conclusion follows from the previous case. ■

Proposition 4.68 *Let $X \subset \mathbf{R}^n$ be a cubical set. Let $A \subset X$ such that* $\operatorname{diam} A < 1$. *Then,* $\operatorname{ch}(A) \cap X$ *is acyclic.*

Proof: Let

$$\mathcal{C} := \{Q \in \mathcal{K}(X) \mid \overset{\circ}{Q} \cap A \neq \emptyset\}.$$

Since X is cubical

$$\operatorname{ch}(A) \cap X = \bigcup_{Q \in \mathcal{C}} Q.$$

Observe that for any two elementary cubes $P, Q \in \mathcal{C}$ the intersection $P \cap Q$ is non-empty, because otherwise $\operatorname{diam} A \geq 1$. Therefore by Proposition 4.67 also $\bigcap \mathcal{C}$ is non-empty. It follows that $\operatorname{ch}(A)$ is star-shaped and consequently acyclic by Proposition 4.66. ■

Exercises

4.9 Give an example where X and Y are acyclic cubical sets, but $X \cup Y$ is not acyclic.

4.10 Consider the capital letter **H** as a 3-dimensional cubical complex. Compute its homology.

4.6 Reduced Homology

In the proofs of Theorem 4.51 and Proposition prop:acyclicM-V we used a specific group homomorphism to deal with the fact that the 0-th homology group was isomorphic to **Z**. In mathematics seeing a particular trick being employed to overcome a technicality in different contexts suggests that the possibility of a general procedure to take care of the problem. As was mentioned the difficulty arose because $H_0 \cong \mathbf{Z}$ rather than being trivial. We can therefore, as the following question: Is there a different homology theory such that in the previous two examples we would have trivial 0th level homology?

Hopefully, this question does not seem too strange. We spent most of Chapter 2 motivating the homology theory that we are using and as we did so we had to make choices of how to define our algebraic structures. From a purely algebraic point of view, given $\mathcal{K}(X)$ all we need inorder to define homology groups is a chain complex $\{C_k(X), \partial_k\}_{k\in\mathbf{Z}}$. This means that if we change our chain complex, then we will have a new homology theory. The trick we employed involved the group homomorphism $\epsilon : C_0(X) \to \mathbf{Z}$ defined by sending each elementary cubical chain to 1. Furthermore, we showed in each case that $\epsilon \circ \partial_1 = 0$, which means that

$$\operatorname{image} \partial_1 \subset \ker \epsilon.$$

It is with this in mind that we introduce the following definition.

Definition 4.69 Let X be a cubical set. The *reduced cubical chain complex* of X is given by $\{\tilde{C}_k(X), \tilde{\partial}_k\}_{k\in\mathbf{Z}}$ where

$$\tilde{C}_k(X) = \begin{cases} \mathbf{Z} & \text{if } k = -1, \\ C_k(X) & \text{otherwise,} \end{cases}$$

and

$$\tilde{\partial}_k := \begin{cases} \epsilon & \text{if } k = 0, \\ \partial_k & \text{otherwise.} \end{cases}$$

The corresponding homology groups form the *reduced homology* of X and are denoted by

$$\tilde{H}_k(X).$$

The following theorem indicates the relationship between the two homology groups we now have at our disposal.

Theorem 4.70 *Let X be a cubical set. $\tilde{H}_0(X)$ is a free abelian group and*

$$H_k(X) \approx \begin{cases} \tilde{H}_0(X) \oplus \mathbf{Z} & \text{for } k = 0 \\ \tilde{H}_k(X) & \text{otherwise.} \end{cases}$$

Furthermore, if $\{P_i \mid i = 0, \ldots, d\}$ is a collection of vertices in X consisting of one vertex from each connected component of X, then

$$\{[P_i - P_0] \in \tilde{H}_0(X) \mid i = 1, \ldots, d\}$$

forms a basis for $\tilde{H}_0(X)$.

Proof: Let $c \in C_0(X)$. Then, by Theorem 4.51 there exists

$$c' = \sum_{i=0}^{d} \alpha_i \widehat{P}_i$$

such that $[c] = [c'] \in H_0(X)$. In other words, there exists $b \in C_1(X)$ such that $c = c' + \partial_1 b$. Furthermore, $[c'] = 0$ if and only if $\alpha_i = 0$ for all $i = 0, \ldots, d$.

Since $\tilde{C}_0(X) = C_0(X)$, $c \in \tilde{C}_0(X)$. However, $c \in \tilde{Z}_0(X)$ only if $\epsilon(c) = 0$. But,

$$\begin{aligned} \epsilon(c) &= \epsilon(c' + \partial_1 b) \\ &= \epsilon(c') + \epsilon\partial_1 b \\ &= \epsilon(\sum_{i=0}^{d} \alpha_i \widehat{P}_i \\ &= \sum_{i=0}^{d} \alpha_i. \end{aligned}$$

Now assume that X has exactly one connected component. Then, $c \in \tilde{Z}_0(X)$ if and only if $c' = 0$. Therefore, in this case $H_0(X) = 0$.

So assume that $d \geq 1$. $c \in \tilde{Z}_0(X)$ implies that $\sum_{i=0}^{d} \alpha_i = 0$. Thus, $0 = -\sum_{i=0}^{d} \alpha_i \widehat{P_0}$. Thus, we can write

$$\begin{aligned} c' &= \sum_{i=0}^{d} \alpha_i \widehat{P}_i - \sum_{i=0}^{d} \alpha_i \widehat{P_0} \\ &= \sum_{i=0}^{d} \alpha_i (\widehat{P}_i - \widehat{P_0}). \end{aligned}$$

■

This theorem allows us to give an alternative characterization of acyclic spaces.

Corollary 4.71 *Let X be a nonempty acyclic cubical set, then*

$$\tilde{H}_*(X) = 0.$$

4.7 Comparison with Simplicial Homology

4.7.1 Simplexes and triangulations

We present here basic definitions and results of Simplicial Homology Theory. The proofs of the presented results and more examples may be found in most of standard textbooks in Algebraic Topology, e.g. [Munkres,Keesee,Rotman].

A subset C of $\mathbf{R}^n$ is called *convex* if, given any two points $x, y \in C$, the line segment

$$[x, y] := \{tx + (1-t)y \mid 0 \leq t \leq 1\}$$

is contained in C.

Definition 4.72 The *convex hull* $\mathrm{co}A$ of a subset A of $\mathbf{R}^n$ is the intersection of all closed and convex sets containing A.

There is at least one closed convex set containing C, the whole space $\mathbf{R}^n$, hence $\mathrm{co}A \neq \emptyset$. It is easy to see that an intersection of any family of convex sets is convex and we already know that the same is true about intersections of closed sets. Thus $\mathrm{co}A$ is the smallest closed convex set containing A. It is intuitively clear that the convex hull of two points is a line segment joining those points, a convex hull of three non-colinear points is a triangle, and a convex hull of four non-coplanar points is a tetrahedron. We shall generalize those geometric figures to an arbitrary dimension under the name simplex.

Theorem 4.73 *Let $\mathcal{V} = \{v_0, v_1, \ldots, v_n\} \in \mathbf{R}^n$ be a finite set. Then co$\mathcal{V}$ is the set of those $x \in \mathbf{R}^n$ which can be written as*

$$x = \sum_{i=0}^{n} \lambda_i v_i,\ 0 \leq \lambda_i \leq 1, ; \sum_{i=0}^{n} \lambda_i = 1\ . \tag{4.12}$$

In general, the coefficients λ_i are not unique. If, for example a, b, c, d are four vertices of the unit square on Figure 2.2 then

$$(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}a + 0b + \frac{1}{2}c + 0d = 0a + \frac{1}{2}b + 0c + \frac{1}{2}d\ .$$

Definition 4.74 A finite set $\mathcal{V} = \{v_0, v_1, \ldots, v_n\}$ in $\mathbf{R}^n$ is *geometrically independent* if, for any $x \in \mathrm{co}\mathcal{V}$, the coefficients λ_i in Equation 4.12 are unique. If this is the case, λ_i are called *barycentric coordinates* of x.

Theorem 4.75 *Let* $\mathcal{V} = \{v_0, v_1, \ldots, v_n\} \in \mathbf{R}^n$. *Then* $\mathcal{V}$ *is geometrically independent if and only if the set of vectors* $\{v_1 - v_0, v_2 - v_0, \ldots, v_n - v_0\}$ *is linearly independent. When this is the case, the barycentric coordinates of* $x \in \mathcal{V}$ *are continuous functions of* x.

Definition 4.76 Let $\mathcal{V} = \{v_0, v_1, \ldots, v_n\}$ be geometrically independent. The set $s = \mathrm{co}\mathcal{V}$ is called *simplex* or, more specificaly, *n-simplex spanned by vertices* $v_0, v_1, \ldots, v_n$. The number n is called *the dimension of* $\mathcal{V}$. If $\mathcal{V}'$ is a subset of $\mathcal{V}$ of $k \leq n$ vertices, the set $\mathrm{co}\mathcal{V}'$ is called *k-face* of $\mathrm{co}\mathcal{V}$.

The union $\mathrm{bd}\,(\sigma)$ of all $(k-1)$-faces of a k-simplex s is called *geometric boundary* of s. It is easy to verify that a point $x \in s$ is in $\mathrm{bd}\, s$ if and only if at least one of its barycentric coordinates is equal to zero.

From Theorem 4.75 we get the following

Corollary 4.77 *Any two n-simplexes are homeomorphic.*

Proof: . Let $s = \mathrm{co}\{v_0, v_1, \ldots, v_n\}$ and $t = \mathrm{co}\{w_0, w_1, \ldots, w_n\}$ be two n-simplexes. Let $\lambda_i(x)$ be barycentric coordinates of $x \in s$ and $\lambda_i(y)$ barycentric coordinates of $y \in t$. By the definition of geometric independence and by Theorem 4.75 the formula

$$f(x) := \sum_{i=0}^{n} \lambda_i(x) w_i$$

defines a linear continuous map $f : s \to t$ with the contiuous inverse

$$f^{-1}(y) := \sum_{i=0}^{n} \lambda_i(y) v_i \ .$$

∎

we will later make use of the following

Definition 4.78 Given any $n \geq 0$ the *standard n-simplex* Δ_n is given by $\Delta_n := \mathrm{co}\{e_1, e_2, \ldots e_{n+1}\}$ where $\{e_1, e_2, \ldots e_{n+1}\}$ is the canonical basis for $\mathbf{R}^{n+1}$. It is easy to see that any linearly independet set is also geometrically independent so Δ_n is an n-simplex indeed. Its special property is that the barycentric coordinates of any point x in Δ_n coincide with the cartesian coordinates $x_1, x_2, \ldots x_{n+1}$.

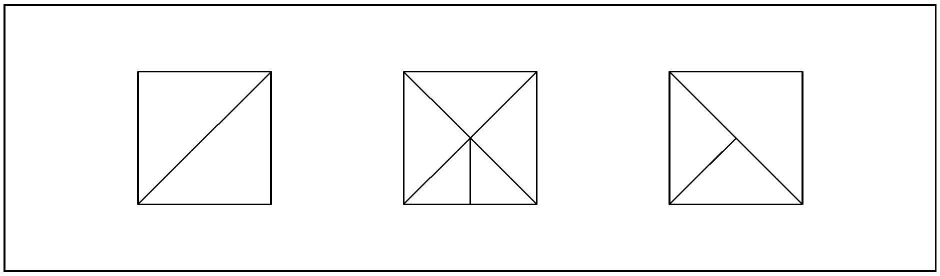


Figure 4.8: Subdivisions of a square to triangles: the first two are triangulations, the last one is not.

Definition 4.79 A *simplicial complex* $\mathcal{S}$ is a finite collection of simplexes such that

1. Every face of a simplex in in $\mathcal{S}$ is in $\mathcal{S}$;
2. The intersection of any two simplexes in $\mathcal{S}$ is a face of each of them.

The subset of $\mathbf{R}^n$ being the union of all simplexes of $\mathcal{S}$ is called the space of $\mathcal{S}$ and is denoted by $|\mathcal{S}|$.

Definition 4.80 A subset $P \in \mathbf{R}^n$ is called *polytope* or *polyhedron* if $P = |\mathcal{S}|$ for some simplicial complex $\mathcal{S}$. In this case $\mathcal{S}$ is called a *triangulation* of P.

Obviously, a polytope may have different triangulations. The Figure 4.8 shows examples of subdivisions of a square to triangles. The first two are triangulations but the last one is not since the intersection of a triangle in the lower-left corner with the triangle in the upper-right corner is not an edge of the latter one but a part of it.

One may expect that any cubical set can be triangulated. We leave the construction as an exercice.

Example 4.81 By a *torus* we mean any space homeomorphic to the product $S^1 \times S^1$ of two circles. Since $S^1 \times S^1 \in \mathbf{R}^4$, it is hard to visualise it. However one can show, by means of polar coordinates, that this space is homeomorphic to the surface in $\mathbf{R}^3$ obtained by rotation of the circle $(x-2)^2 + z^2 = 1,\ y = 0$ about the Y-axis. This set can be described as the surface of a donat. Neither of the above surfaces is a polytope but we shall construct one which is. Let G be the boundary of any triangle in $\mathbf{R}^2$. Then G is a simple closed curve

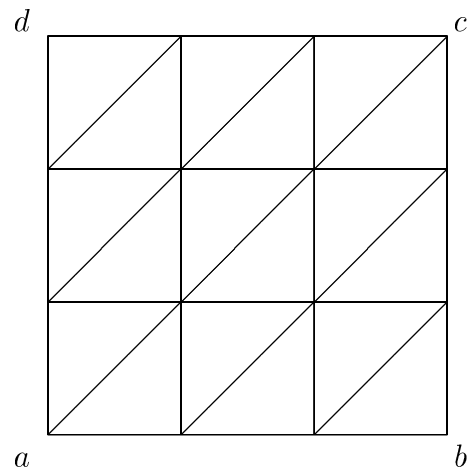


Figure 4.9: Triangulation of a torus

hence it is homeomorphic to the unit cicle. Thus $T = G \times G \in \mathbf{R}^4$ is a torus. In order to construct a triangulation of T we may visualise T as a square on Figure 4.9 with pairs of parallel sides glued together. More precisely, consider the square $[0,3]^2 = \mathrm{co}\{a,b,c,d\}$ where $a = (0,0), b = (0,3), c = (3,3), d = (0,3)$. Bend the square along the lines $x = 1$ and $x = 2$ and glue the directed edge $[a,d]$ with $[b,c]$ so that the vertex a is identified with b and d with c. We obtain a cylinder in $\mathbf{R}^3$ with a boundary of a unilateral triangle in the plane $y = 0$ as the base. We bend the cylinder along the lines $y = 1$ and $y = 2$ (this cannot be done in $\mathbf{R}^3$ without stretching but we may add another axis) and glue the edge $[a,b]$ with $[d,c]$. Note that the four vertices a,b,c,d of the square became one. The bend lines divide the square to nine unitary squares. Each of them can be divided to two triangles as shown on Figure 4.9. Let $\mathcal{S}$ be the collection of all vertices, edges, and triangles of T obtained in this way. Although some vertices and edges are identified by gluing, the reader may verify that $\mathcal{S}$ satisfies the definition of simplicial complex.

4.7.2 Simplicial Homology

The term simplicial complex suggests that there should be some natural structure of chain complex associated with it. That is not so easy to define due to problems with orientation which do not appear when we study cubical sets. We shall therefore proceed as we did with graphs in Chapter 2, that is, we shall start from chain complexes with coefficients in $\mathbf{Z}_2$. This will make definitions much more simple and, historically, this is the way homology was first introduced.

Let $C^k(\mathcal{S}, \mathbf{Z}_2)$ be the vector space generated by the set $\mathcal{S}^k$ of k-dimensional simplexes of $\mathcal{S}$ as the canonical basis. We put $C^k(\mathcal{S}, \mathbf{Z}_2) := 0$ if $\mathcal{S}$ has no simplexes od dimension k. The boundary map $\partial_k : C^k(\mathcal{S}, \mathbf{Z}_2) \to C^{k-1}(\mathcal{S}, \mathbf{Z}_2)$ is defined on any basic element $s = \mathrm{co}\{v_0, v_1, \ldots, v_n\}$ by the formula

$$\partial_k(s) = \sum_{i=0}^{n} \mathrm{co}(\mathcal{V} \setminus \{v_i\}) \ .$$

Thus, in modulo 2 case, the algebraic boundary of a simplex corresponds precisely to its geometric boundary. We have the following

Proposition 4.82 $\partial_{k-1}\partial_k = 0$ *for all* k.

Proof: For any basic element $s = \mathrm{co}\{v_0, v_1, \ldots, v_n\}$,

$$\partial_{k-1}\partial_k(\sigma) = \sum_{j \neq i} \sum_{i=0}^{n} \mathrm{co}(\mathcal{V} \setminus \{v_i, v_j\}) \ .$$

Each $(k-1)$-face of s appears in the above sum twice, therefore the sum modulo 2 is equal to zero. ■

Thus $\mathcal{C}(\mathcal{S}, \mathbf{Z}_2) := \{C^k(\mathcal{S}, \mathbf{Z}_2), \partial_k\}_{k \in \mathbf{Z}}$ has the structure of a chain complex with coefficients in $\mathbf{Z}_2$. The homology of that chain comlex is the sequence of vector spaces

$$H_*(\mathcal{S}, \mathbf{Z}_2) = \{H_n(\mathcal{S}, Z_2)\} = \{\ker \partial_n / \mathrm{im}\, \partial_{n+1}\}$$

The modulo 2 homology of graphs discussed in Section 2,2,2 is a special case of what we did above. The real goal however is to construct a chain complex corresponding to $\mathcal{S}$ with coefficients in $\mathbf{Z}$ as defined in Section 3.7. As we did it with graphs, we want to impose an orientation of vertices $v_0, v_1, \ldots, v_n$ spanning a simplex. In case of graphs that was easy since each edge joining vertices a, b could be written in two ways, as $[a, b]$ or $[b, a]$ and it was sufficient to tell which vertex do we want to write as the first and which as the last. In case of simplexes of dimension higher than one, there are many different ways of ordering the set of vertices.

Definition 4.83 Two orderings $(v_0, v_1, \ldots, v_n)$ and $(v_{p_0}, v_{p_1}, \ldots, v_{p_n})$ of vertices of an n-simplex s are said to have the same *orientation* if one can get one from another by an even number of permutations of neighboring terms

$$(v_{i-1}, v_i) \to (v_i, v_{i-1}) \ .$$

This defines an equivalence relation on the set of all orderings of vertices od s. An *oriented simplex* $\sigma = [v_0, v_1, \ldots, v_n]$ is is an equivalence class of the ordering $(v_0, v_1, \ldots, v_n)$ of vertices of a simplex $s = \mathrm{co}\{v_0, v_1, \ldots, v_n\}$.

It is easy to see that for $n > 0$ the above equivalance relation divides the set of all orderings to two equivalence classes. Hence we may say that the orderings which are not in the same equivalence class have the *oposite orientation.* We shall denote the pairs or oposite oriented simplexes by σ, σ' or τ, τ'. An oriented simplicial complex in a simplicial complex $\mathcal{S}$ with one of the two equivalence clsses chosen for each simplex of $\mathcal{S}$. The orientations of a simplex and its faces may be done arbitrarily, they do not need to be related.

Example 4.84 Let s be a triangle in $\mathbf{R}^2$ spanned by vertices a, b, c. Then the orientation equivalence class $\sigma = [a, b, c]$ contains the orderings (a, b, c), (b, c, a), (c, a, b) and the oposite orientation σ' contains (a, c, b), (b, a, c), (c, b, a). One may graphically distinguish the two orientations by tracing a closed path around the boundary of the triangle s following the order of vertices. The first equivalence class gives the counter-clockwise direction and the second one the clockwise direction. However, the meaning of clockwise or counterclockwise orientation is lost when we consider a triangle in a space of higher dimension. Let $\mathcal{S}$ be the complex consisting of s and all of its edges and vertices. Here are some among possible choices of orientations and their graphical representations on Figure 4.10:

1. $[a, b, c], [a, b], [b, c], [c, a]$
2. $[a, b, c], [a, b], [b, c], [a, c]$
3. $[a, c, b], [a, b], [b, c], [a, c]$

On the first sight second and third orientation seem wrong since the arrows on the edges of the triangle do not close a cycle but do not worry: when we get to algebra, the "wrong" direction of the arrows will be corrected by the minus sign in the formula for the boundary operator.

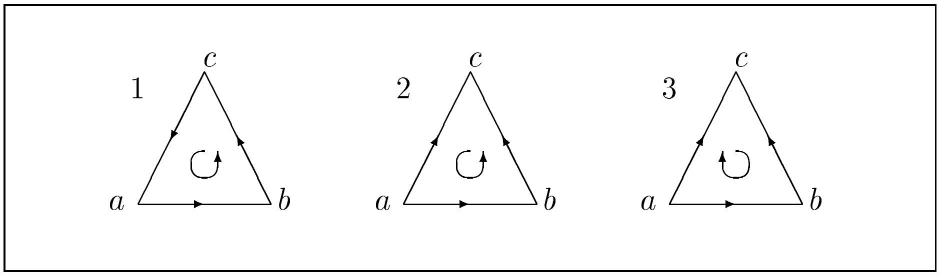


Figure 4.10: Some orientations of simplexes in a triangle

Let now S^n be the set of all oriented n-simplexes of $\mathcal{S}$. Recall from Section 3.2 that a free abelian group $\mathbf{Z}^{S^n}$ generated by S^n is the set of all functions $c : S^n \to \mathbf{Z}$, generated by basic elements $\hat{\sigma}$ which can be identified with $\sigma \in S^n$. We would like to call this the group of n-chains but there is a complication: If $n \geq 0$, each n-simplex of $\mathcal{S}$ corresponds to two elements of of S^n. We therefore adapt the following definition.

Definition 4.85 The group of *n-chains* denoted by $\mathcal{C}^n(\mathcal{S})$ is the subgroup of $\mathbf{Z}^{S^n}$ consisting of those functions c which satisfy the identity

$$c(\sigma) = -c(\sigma')$$

if σ and σ' are oposite orientations of the same n-simplex s.

Proposition 4.86 *The group $\mathcal{C}^n(\mathcal{S})$ is a free abelian group generated by fuctions $\tilde{\sigma} = \hat{\sigma} - \hat{\sigma'}$ given by the formula*

$$\tilde{\sigma}(\tau) := \begin{cases} 1 & \text{if } \tau = \sigma, \\ -1 & \text{if } \tau = \sigma', \\ 0 & \text{otherwise,} \end{cases}$$

where $\sigma, \sigma', \tau \in S^n$ and σ, σ' are oposite orientations of the same simplex. This set of generators is not a basis since $\tilde{\sigma'} = -\tilde{\sigma}$ for any pair σ, σ'. A basis is obtained by chosing either one.

The choice of a basis in Proposition 4.86 is related to the choice of an orientation in $\mathcal{S}$. Upon identification of the basic elements $\tilde{\sigma}$ with σ we get the identification of σ' with $-\sigma$. We put $\mathcal{C}^n(\mathcal{S}) := 0$ if $\mathcal{S}$ contains no n-simplexes. The boundary map $\partial_k : C^k(\mathcal{S}) \to C^{k-1}(\mathcal{S})$ is defined on any basic element $[v_0, v_1, \ldots, v_n]$ by the formula

$$\partial_k[v_0, v_1, \ldots, v_n] = \sum_{i=0}^{n}(-1)^i[v_0, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n] \; .$$

There is a bit of work involved in showing that this formula actually defines a boundary map: First, one needs to show that the formula is correct i.e. that it does not depend on the choice of a representative of the equivalence class $[v_0, v_1, \ldots, v_n]$. Secondly, one needs to show that $\partial_{k-1}\partial_k = 0$. The reader may consult [Munkers] for the proofs.

Thus $\mathcal{C}(\mathcal{S}) := \{C^k(\mathcal{S}), \partial_k\}_{k \in \mathbf{Z}}$ has the structure of a chain complex as defined in Section 3.7. The homology of that chain complex is the sequence of abelian groups.

$$H_*(\mathcal{S}) = \{H_n(\mathcal{S}\} = \{\ker \partial_n / \operatorname{im} \partial_{n+1}\} \; .$$

An important and difficult problem is to show that different triangulations of the same polytope have isomorphic homology comology complexes. That is proved by means of so called barycentric subdivisions and is too involved for the scope of this overview. The concept of barycentric subdivision will appear as a by-product of the proof of Theorem 4.87 in the next section.

4.7.3 Comparison of Cubical and Simplicial Homology

Cubical complexes have several nice properties which are not shared by simplicial complexes:

1. As we already mentioned in the introduction to this chapter, numerical computations and computer graphics naturally lead to cubical sets. Since they already have a sufficient combinatorial structure to define homology, further subdivision to triangulations becomes artificial and increases the complexity of data.

2. A product of elementary cubes is an elementary cube but a product of simplexes is not a simplex. For example, a product of a triangle by an interval is a cylinder and it has to be triangulated in order to compute the simplicial homology. That feature of elementary cubes makes many proves easier and lists of data shorter.

3. We shall talk later about cubical subdivisions. That will be done very naturally by changing the scale on each coordinate so that the grid $\mathbf{Z}^n$ of integer coordinates is replaced by the grid $(\frac{1}{2}\mathbf{Z})^n$. Each elementary cube is then subdivided to 2^n smaller cubes by cutting the length of each side by half. The notion of barycentric sudivision in the simplicial theory is much more complicated both numerically and concepionally.

4. As we have seen in the previous section, the notion of orientation in simplicial complexes is not an easy concept to learn. Why does this problem not appear in the study of cubical complexes? The answer is in the fact that the definition of a cubical set is dependent on a particular choice of coordinates in the space. First, already in $\mathbf{R}$, we have unknowingly chosen a particular orientation by having written an elementary interval as $[l, l+1]$ and not $[l+1, l]$. In other words, a linear order of real numbers imposes a choice of an orientation on each coordinate axis in $\mathbf{R}^n$. Secondly, by having written a product of inervals $I_1 \times I_2 \times \cdots \times I_n$ we have implicitely chosen the ordering of the canonical basis for $\mathbf{R}^n$.

There is one important weak point of cubical complexes: Every polytope can be triangulated but not every polytope can be expressed as a cubical set. In aprticular, a triangle is not a cubical set.

We have however the following result which will help us to define homology of a polytope via cubical homology when we later introduce homology of a map:

Theorem 4.87 *Every polytope P is homeomorphic to a cubical set. Moreover, given any triangulation $\mathcal{S}$ of P, there exists a homeomorphism $h : P \to X$, where X is a cubical set, such that the restriction of h to any simplex of $\mathcal{S}$ is a homeomorphism of that simplex onto a cubical subset of X.*

Proof: In order to keep the idea transparent we skip several technical verifications. The construction of h is done in two steps.

Step 1. We construct a homeomorphic embedding of P into a standard simplex in a space of a sufficiently high dimension.

Indeed, let $\mathcal{S}$ be a triangulation of P and let $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ be the set of all vertices of $\mathcal{S}$. Let Δ_N be the standard N-simplex in $\mathbf{R}^{N+1}$ described in Definition 4.78. Consider the bijection f_0 of $\mathcal{V}$ onto the canonical basis of

$\mathbf{R}^{N+1}$ given by $f_0(v_i) = e_i$. Given any n-simplex $s = (co)\{v_{p_0}, v_{p_1} \dots v_{p_n}\}$ of $\mathcal{S}$, f_0 extends to a map $f_s : s \to \mathbf{R}^{N+1}$ by the formula

$$f_s(\sum \lambda_i v_{p_i}) = \sum \lambda_i e_{p_i}$$

where λ_i are barycentric coordinates of a point in s. It follows that $f_s(s)$ is a n-simplex and f_s is a homeomorphism of s onto $f_s(s)$. If s and t are any two simplexes of $\mathcal{S}$, $s \cap t$ is empty or is their common face so if $x \in s \cap t$ then

$$f_s(x) = f_{s\cap t}(x) = f_t(x) \; .$$

Thus the maps f_s match on intersections of simplexes. Since simplexes are closed and there are finitely many of them, the maps f_s extend to a map $f : P \to \tilde{P} := f(P)$. By the linear independence of $\{e_1, e_2, \dots, e_N\}$, one shows that $\tilde{P}$ is a polytope triangulated by $\{f(s)\}$ and f is a homeomorphism. Moreover, by its construction, f maps simplexes to simplexes.

Step 2. We construct a homeomorphism g of Δ_N onto the cubical set $Y \subset \operatorname{bd}[0,1]^{N+1}$ consisting of those faces of $[0,1]^{N+1}$ which have the degenerate interval $[1]$ on one of the components and such that any face of Δ_N is mapped to a cubical face of Y. Once we do that, it will be sufficient to take $X := g(\tilde{P})$ and define the homeomorphism h as the composition of f and g.

Consider the diagonal line L parametrised by $t \to (t, t, \dots, t) \in \mathbf{R}^{N+1}$, $t \in \mathbf{R}$. The idea is to project a point $x \in \Delta_N$ to a face of Y along the line L in the direction away from the origin. Recall that the barycentric coordinates of $x \in \Delta_N$ coincide with its cartesian coordinates, thus $\sum x_i = 1$ and $0 \leq x_i \leq 1$ for all i. The image $y = g(x)$ should have coordinates $y_i = x_i + t$ for all i and some $t \geq 0$. This point is in Y if $0 \leq x_i + t \leq 1$ for all i and $x_j + t = 1$ for some j. Note that the supremum norm of x is $||x|| = max\{x_1, x_2, \dots x_{N+1}\}$. Thus the number $t := 1 - ||x||$ has the desired property and the coordinates of $y = g(x)$ are given by

$$y_i = 1 + x_i - ||x|| \; .$$

It is clear that g is continuous. The injectivity of g is proved by noticing that any line parallel to L intersects Δ_N at a unique point. The surjectivity of g is a by-product of the construction of its inverse g^{-1}. Let $y \in Y$. In order to define $x = g^{-1}(y)$ we must find a number $t \in [0,1]$ such that the point x whose coordinates are given by $x_i = y_i - t$ is in Δ_N. For this, we must have $0 \leq y_i - t \leq 1$ for all i and $\sum_{j=1}^{N+1}(y_j - t) = 1$. Thus

$$t = \frac{1}{N+1}(\sum_{j=1}^{N+1} y_j - 1) \; .$$

Since $0 \leq y_i \leq 1$ for all i and $y_j = 1$ for some j, t has the desired properties.
■

We finish this section by discussing an ineresting by-product of the above proof. A reader unfamiliar with the simplicial theory may skip it or just try to grasp the main idea. Note that the inverse image of the vertex $(1, 1, \ldots, 1)$ of Y in g is the point

$$\bar{x} := \frac{1}{N+1}(1, 1, \ldots, 1)$$

called *barycenter* of Δ_N. By continuing along these lines one can show that, for each face Q of Y, $g^{-1}(Q)$ is a so called star of a vertex of Δ_N with respect to the barycentric subdivision Δ'_N of Δ_N. The first homeomorphism f^{-1} preservers the barycentric coordinates of points in each simplex so it preserves barycenters and barycentric sudivisions. These observations permit to define a homomorphism of chain complexes $\mathcal{C}(X) \to \mathcal{C}(\mathcal{S}')$ which induces an isomorphism $H_*(\mathcal{S}') \cong H_*(X)$ in homology. If we take for granted the result of the simplicial theory saying that that the simplicial homology $H_*(P)$ of a polytope is independent on the choice of a triangulation, we get

$$H_*(P) \cong H_*(X) \ .$$

In the last chapter we shall be able to arrive at this conlusion without the necessity of applying the simplicial theory.

Exercises

4.11 Define the chain complex $\mathcal{C}(T, \mathbf{Z}_2)$ for the triangulation discussed in Example 4.81 and use the homchain program to compute $H_*(T, \mathbf{Z}_2)$.

4.12 * Prove that any cubical set can be triangulated.

4.13 Label vertices, edges, and triangles of the triangulation of the torus in Example˜refex:torus2 displayed on Figure 4.9. Define the chain complex $\mathcal{C}(T)$. Use the homology program to compute $H_*(T)$.

4.14 Let K be a polytope constructed as T in Example 4.81 but with one pair of sides twisted before gluing so that the directed edge $[a, d]$ is identified with $[c, b]$. The remaining pair of edges is glued as before, $[b, c]$ with $[a, d]$. Compute $H_*(K)$. What happens if we try to use the homchain program for computing $H_*(K, \mathbf{Z}_2)$?

This K is called *Klein bottle.* Note that K cannot be visualised in $\mathbf{R}^3$, we need an extra dimension in order to glue two circles limiting a cylinder with twisting and without cutting the side surface of the cylinder.

4.15 Let P be a polytope constructed as T in Example 4.81 but with sides twisted before gluing so that the directed edge $[a, d]$ is identified with $[c, b]$ and $[b, c]$ with $[d, a]$. Compute $H_*(P)$. What happens if we try to use the homchain program for computing $H_*(P, \mathbf{Z}_2)$?
This P is called *projective plane.* Note that P cannot be visualised in $\mathbf{R}^3$.

Chapter 5

Homology of Maps

If homology is a natural invariant of a topological space, then at the very least given homeomorphic spaces X and Y it should be true that $H_*(X)$ and $H_*(Y)$ are isomorphic as groups. To prove this we need to be able to pass from continuous maps $h : X \to Y$ to group homomorphisms $h_* : H_*(X) \to H_*(Y)$. Of course, for the time being the set of topological spaces that we can consider is restricted to cubical sets.

As we have indicated many times by now an element of a homology group is a cycle, i.e. a chain which lies in the kernel of the boundary map. Thus, it is reasonable to expect that to define a map on homology one first needs to define a map on the chains. We shall do this by first constructing a multivalued map on cubes, and then providing an algorithm for obtaining a linear map on cubical chains. We begin, however, with a purely algebraic discussion of the latter.

5.1 Chain Maps

Let X and Y be cubical sets with associated cubical chain complexes $\mathcal{C}(X) = \{C_k(X), \partial_k^X\}$ and $\mathcal{C}(Y) = \{C_k(Y), \partial_k^Y\}$. We need to define a special class of group homomorphisms between the chain complexes that will induce maps on the homology groups. While we will use the notation $F : C(X) \to C(Y)$ to represent such a map, it must be kept in mind that F really consists of a collection of group homomorphisms

$$F_k : C_k(X) \to C_k(Y).$$

Since F is supposed to induce a map on homology, it must be the case that F maps cycles to cycles and boundaries to boundaries. As was discussed in Section 5.1 this leads to the notion of a chain map.

Definition 5.1 Let X and Y be cubical sets with associated cubical chain complexes $\mathcal{C}(X) = \{C_k(X), \partial_k^X\}$ and $\mathcal{C}(Y) = \{C_k(Y), \partial_k^Y\}$. $F : C(X) \to C(Y)$ is a *cubical chain map* if for every $k \in \mathbf{Z}$

$$\partial_k^Y \circ F_k = F_{k-1} \circ \partial_k^X. \tag{5.1}$$

Another way to describe an equality such as (5.1) is through the language of *commutative diagrams*. More precisely to say that the diagram

$$\begin{array}{ccc} C_k(X) & \xrightarrow{F_k} & C_k(Y) \\ \downarrow \partial_k^X & & \downarrow \partial_k^Y \\ C_{k-1}(X) & \xrightarrow{F_{k-1}} & C_{k-1}(Y) \end{array}$$

commutes is equivalent to saying that $\partial_k^Y \circ F_k = F_{k-1} \circ \partial_k^X$.

Proposition 5.2 *If $F : C(X) \to C(Y)$ is a chain map, then*

$$F_k : Z_k(X) \to Z_k(Y)$$

and

$$F_k : B_k(X) \to B_k(Y).$$

Proof: If $c \in Z_k(X)$, then $\partial_k^X c = 0$. Thus

$$0 = \partial_k^X c = F_k \partial_k^X c = \partial_k^Y F_k c$$

which implies that $F_k c \in Z_k(Y)$.

Let $c \in B_k(X)$. Then, there exists $b \in C_{k+1}(X)$ such that $\partial_{k+1}^X b = c$. Thus,

$$F_k c = F_k \partial_{k+1}^X b = \partial_k^Y F_k b$$

which implies that $F_k(c) \in B_k(Y)$. ∎

Definition 5.3 Let $F : C(X) \to C(Y)$ be a chain map. Define $F_* : H_*(X) \to H_*(Y)$ by

$$F_*([\gamma]) = [F(\gamma)].$$

That this map is well defined follows essentially from Proposition 5.2. More precisely, if $[\gamma] \in H_k(X)$, then $\gamma \in Z_k(X)$. By Proposition 5.2 $[F_k(\gamma)] \in H_k(Y)$. Now assume that $[\gamma] = [\alpha]$. Then, $\alpha = \gamma + b$ where $b \in B_k(X)$. But,

$$F_*[\gamma] = [F_k\gamma] = [F_k\gamma + F_k b] = [F_k(\gamma + b)] = [F_k\alpha] = F_*[\alpha].$$

We now know that cubical chain maps $F, G : C(X) \to C(Y)$ generate homology maps $F_*, G_* : H_*(X) \to H_*(Y)$. It is natural to ask under what conditions does $F_* = G_*$?

Motivate the following defintion

Definition 5.4 Let $F, G : C(X) \to C(Y)$ be chain maps. A collection of functions

$$D_k : C_k(X) \to C_{k+1}(Y)$$

is a *chain homotopy* between F and G if for all k

$$\partial^Y_{k+1} D_k + D_{k-1}\partial^X_k = G - F.$$

Restating this definition in terms of a diagram gives

$$\begin{array}{ccccc}
 & & & & C_{k+1}(Y) \\
 & & \nearrow D_k & & \downarrow \partial^Y_{k+1} \\
C_k(X) & & \xrightarrow{G_k - F_k} & & C_k(Y) \\
\downarrow \partial^X_k & & \nearrow D_{k-1} & & \\
C_{k-1}(X) & & & &
\end{array}$$

Theorem 5.5 *If there exists a chain homotopy between F and G, then $F_* = G_*$.*

Proof: Let $[\gamma] \in H_k(X)$. Then

$$\begin{aligned} G(\gamma) - F(\gamma) &= \partial_{k+1}^Y D_k(\gamma) + D_{k-1}\partial_k^X(\gamma) \\ &= \partial_{k+1}^Y D_k(\gamma) \in B_k(Y) \end{aligned}$$

Therefore, $[G(\gamma)] = [F(\gamma)]$. ■

Example 5.6 Let $X \subset \mathbf{R}^2$ be the boundary of the unit square $[0,1]\times[0,1]$. Then

$$\mathcal{K}_1(X) = \{[0,1]\times[0], [0]\times[0,1], [1]\times[0,1], [0,1]\times[1]\}.$$

Let $\mathrm{id} : C(X) \to C(X)$ be the identity map and let $F : C(X) \to C(X)$ be the chain map which one can think of as being generated by rotating X by 90 degrees in a clockwise direction. More precisely,

$$\begin{aligned} F_0 : C_0(X) &\to C_0(X) \\ \widehat{[0]\times[0]} &\mapsto \widehat{[0]\times[1]} \\ \widehat{[0]\times[1]} &\mapsto \widehat{[1]\times[1]} \\ \widehat{[1]\times[1]} &\mapsto \widehat{[1]\times[0]} \\ \widehat{[1]\times[0]} &\mapsto \widehat{[0]\times[0]} \end{aligned}$$

$$\begin{aligned} F_1 : C_1(X) &\to C_1(X) \\ \widehat{[0,1]\times[0]} &\mapsto -\widehat{[0]\times[0,1]} \\ \widehat{[0]\times[0,1]} &\mapsto \widehat{[0,1]\times[1]} \\ \widehat{[0,1]\times[1]} &\mapsto -\widehat{[1]\times[0,1]} \\ \widehat{[1]\times[0,1]} &\mapsto \widehat{[0,1]\times[0]} \end{aligned}$$

We will show that $\mathrm{id}_* = F_*$ by producing a chain homotopy $D_k : C_k(X) \to C_{k+1}(X)$ from F to id. Observe that $\mathcal{K}_2(X) = \emptyset$, therefore $D_k = 0$ for $n \geq 1$. This means that only D_0 needs to be defined. By definition it must satisfy

$$D_0\partial_1 = F - \mathrm{id}.$$

Let

$$D_0 : C_0(X) \to C_1(X)$$

$$
\begin{aligned}
[0]\widehat{\times}[0] &\mapsto [0]\widehat{\times}[0,1]\\
[0]\widehat{\times}[1] &\mapsto [0,1]\widehat{\times}[1]\\
[1]\widehat{\times}[1] &\mapsto -[1]\widehat{\times}[0,1]\\
[1]\widehat{\times}[0] &\mapsto -[0,1]\widehat{\times}[0]
\end{aligned}
$$

Observe that

$$
\begin{aligned}
D_0\partial_1[0]\widehat{\times}[0,1] &= D_0([0]\widehat{\times}[1]-[0]\widehat{\times}[0])\\
&= [0,1]\widehat{\times}[1]-[0]\widehat{\times}[0,1]\\
&= (F-\mathrm{id}\,)([0]\widehat{\times}[0,1]).
\end{aligned}
$$

The remaining cases are left to the reader to check. ■

Proposition 5.7 *Assume X, Y, Z are cubical sets and $F : C(X) \to C(Y)$ and $\psi : C(Y) \to C(Z)$ are chain maps. Then $\psi \circ F : C(X) \to C(Z)$ is a chain map and*

$$(\psi \circ F)_* = \psi_* \circ F_*.$$

The proof is left to Exercise 5.1.

Definition 5.8 A chain map $F : C(X) \to C(Y)$ is called a *chain equivalence* if there exists a chain map $G : C(Y) \to C(X)$ such that $G \circ F$ is chain homotopic to $\mathrm{id}_{C(X)}$ and $F \circ G$ is chain homotopic to $\mathrm{id}_{C(Y)}$

Exercises

5.1 Prove Proposition 5.7

5.2 If $F : C(X) \to C(Y)$ is a chain equivalence then $f_* : H_*(X) \to H_*(Y)$ is an isomorphism.

5.3 * Let X be a cubical set and X' obtained from X via an elementary collapse of a free face $Q \in \mathcal{K}_{n-1}(X)$ through $P \in \mathcal{K}_n(X)$ as in Theorem 4.55. Let $j : \mathcal{C}_{n-1}(X) \to \mathcal{C}_n(X)$ be the inclusion homomorphism and $p : \mathcal{C}_n(X) \to \mathcal{C}_{n-1}(X)$ the projection homomorphism given on generators by $p(\widehat{P}) := 0$ and $p(\widehat{S}) := \widehat{S}$ if $S \neq P$. Show that $p \circ j = \mathrm{id}_{\mathcal{C}_{n-1}(X)}$ and that $j \circ p$ is chain homotopic to $\mathrm{id}_{\mathcal{C}_n(X)}$. Conclude from Exercise 5.2 that $H_*(X') \cong H_*(X)$. This gives an alternative shorter proof of Theorem 4.55.

5.2 Cubical Multivalued maps.

In the last section we discussed chain maps and the maps they induce on homology. We did not however, discuss how one goes from a continuous map to a chain map. The are a variety of possibilities, each with it advantages and disadvantages. The approach we will adopt involves using multivalued maps to approximate the continuous map. The motivation for this was given in Chapter 2. We now want to formalize these ideas.

Let X and Y be cubical sets. A *multivalued map* $\mathcal{F} : X \overrightarrow{\to} Y$ from X to Y is a function from X to subsets of Y, i.e. for every $x \in X$, $\mathcal{F}(x) \subset Y$. However, this notion is far too general to be of use in defining a homology theory. In particular, we want to make sure that points, which have simple topology, get mapped to sets that have simple topology. In the previous chapter we introduced the notion of acyclic sets, i.e. sets with the same homology as that of a point. With this in mind we restrict ourselves to the following types of multivalued maps.

Definition 5.9 Let X and Y be cubical sets. A multivalued map $\mathcal{F} : X \overrightarrow{\to} Y$ is *cubical* if:

1. For every $x \in X$, $\mathcal{F}(x)$ is an acyclic cubical set.

2. For every $Q \in \mathcal{K}(X)$, $\mathcal{F} \mid_{\overset{\circ}{Q}}$ is constant, i.e. if $x, x' \in \overset{\circ}{Q}$, then $\mathcal{F}(x) = \mathcal{F}(x')$.

Observe that since $\mathcal{F}(x)$ is cubical, $\mathcal{F}(x)$ is closed. If $A \subset X$, then

$$\mathcal{F}(A) := \bigcup_{x \in A} \mathcal{F}(x).$$

Example 5.10 Let $X = [0, 2]$ and let $Y = [-5, 5]$. Define $\mathcal{F} : X \overrightarrow{\to} Y$ by

$$\begin{array}{rcl}
[0] & \mapsto & [-5] \\
[1] & \mapsto & [0] \\
[2] & \mapsto & [5] \\
(0,1) & \mapsto & [-4,-1] \\
(1,2) & \mapsto & [1,4]
\end{array}$$

The graph of this function is given in Figure 5.1. Observe that $\mathcal{F}$ is a cubical map. However, from several points of view this is not satisfactory for our needs. The first is that intuitively it should be clear that a map of this type cannot be thought of as being continuous. The second is that we are interested in multivalued maps because we use them as outer approximations of continuous maps. But it is obvious that there is no continuous map $f : X \to Y$ such that $f(x) \in \mathcal{F}(x)$ for all $x \in X$.

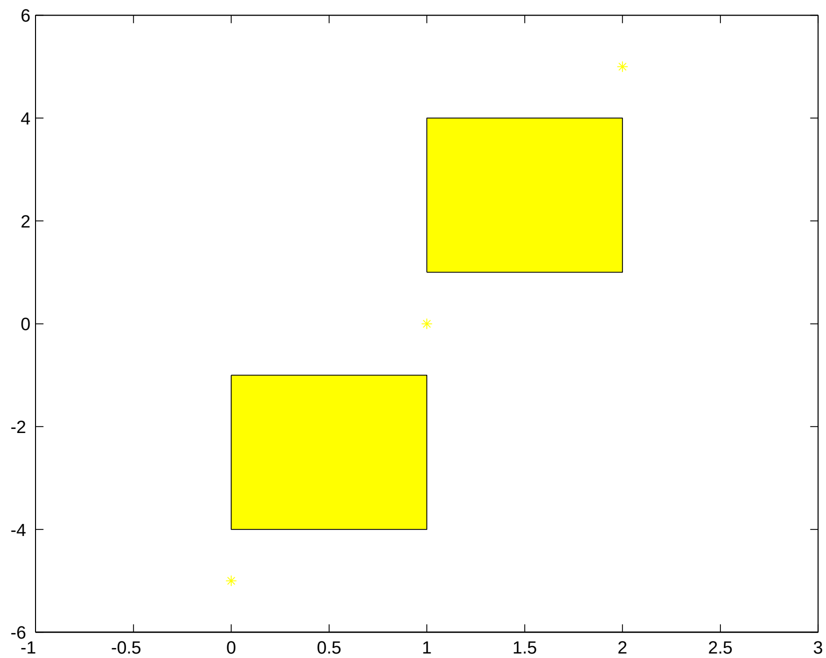


Figure 5.1: The graph of the cubical map $f : [0, 2] \to [-5, 5]$.

To overcome these problems we need to introduce a notion of continuity for multivalued maps. Recall that for single valued functions, continuity is defined in terms of the pre-images of open sets. We want to do something similar for multivalued maps. However, the first problem is that there are at least two reasonable ways to define a pre-image.

Let $\mathcal{F} : X \rightrightarrows Y$ and let $B \subset Y$. The *weak pre-image* of B under $\mathcal{F}$ is

$$\mathcal{F}^{*-1}(B) := \{x \in X \mid \mathcal{F}(x) \cap B \neq \emptyset\},$$

while the *pre-image* of B is

$$\mathcal{F}^{-1}(B) := \{x \in X \mid \mathcal{F}(x) \subset B\}.$$

Definition 5.11 A multivalued map $\mathcal{F}$ is *upper semicontinuous* if $\mathcal{F}^{-1}(U)$ is open for any open $U \subset Y$ and it is *lower semicontinuous* if the set $\mathcal{F}^{*-1}(U)$ is open for any open $U \subset Y$.

Example 5.12 With our goal of using multivalued maps to enclose the image of a continuous function there is in some a canonical choice of constructing upper or lower semicontinuous cubical maps. To make this clear let us return to the discussion in Section **??** where the use of multivalued maps was first presented. We considered the function $f(x) = (x - \sqrt{2})(x+1)$ as a map from $X = [-2, 2] \subset \mathbf{R}$ to $Y = [-2, 4] \subset \mathbf{R}$.

Using a Taylor approximation we derived bounds on f that applied to the elementary intervals (see Table 2.7 and Figure 2.7). These bounds were used to define $\mathcal{F}(Q)$ for each $Q \in \mathcal{K}_1(X)$. There are two simple ways to define $\mathcal{F}$ acting on vertices. Let $P \in \mathcal{K}_0(X)$ and let $Q_P^{\pm} \in \mathcal{K}_1(X)$ be the two edges for which $P \in Q^{\pm}$ (if $P = [0]$ or $P = [2]$, then set $Q^- = Q^+$). Define

$$\mathcal{F}^u(P) := \mathcal{F}(Q_P^+) \cup \mathcal{F}(Q_P^-)$$

and

$$\mathcal{F}^l(P) := \mathcal{F}(Q_P^+) \cap \mathcal{F}(Q_P^-).$$

Then, $\mathcal{F}^u$ is upper semicontinuous and $\mathcal{F}^l$ is lower semicontinuous.

Proposition 5.13 *Assume* $\mathcal{F} : X \overrightarrow{\to} Y$ *is a cubical lower semicontinuous map. If* $P, Q \in \mathcal{K}(X)$ *and* P *is a face of* Q*, then* $\mathcal{F}(\overset{\circ}{P}) \subset \mathcal{F}(\overset{\circ}{Q})$.

Proof: Since $\mathcal{F}(\overset{\circ}{Q}) = \mathcal{F}(x)$ for $x \in \overset{\circ}{Q}$, the set $\mathcal{F}(\overset{\circ}{Q})$ is cubical and consequently closed. Thus the set $U := Y \setminus \mathcal{F}(\overset{\circ}{Q})$ is open. By the lower semicontinuity of $\mathcal{F}$,

$$V := \mathcal{F}^{*-1}(U) = \{z \in X \mid \mathcal{F}(z) \cap U \neq \emptyset\}$$

is open.

Now consider $x \in \overset{\circ}{P}$. Since $\mathcal{F}$ is cubical, $\mathcal{F}(x) = \mathcal{F}(\overset{\circ}{P})$. Therefore, it is sufficient to prove that $\mathcal{F}(x) \subset \mathcal{F}(\overset{\circ}{Q})$. This is equivalent to showing that $x \notin V$, which will be proved by contradiction.

So, assume that $x \in V$. Since $x \in \overset{\circ}{P}$ and $P \prec Q$, it follows that $x \in Q = \operatorname{cl}(\overset{\circ}{Q})$. Thus, $V \cap \operatorname{cl}(\overset{\circ}{Q}) \neq \emptyset$. But V is open, hence $V \cap \overset{\circ}{Q} \neq \emptyset$. Let $z \in V \cap \overset{\circ}{Q}$. Then, because $\mathcal{F}$ is cubical, $\mathcal{F}(z) = \mathcal{F}(\overset{\circ}{Q})$, and hence, $\mathcal{F}(z) \cap U = \emptyset$. Thus, $z \notin V$, a contradiction. ∎

Exercises

5.4 Let $X = [-1,1]^2 \subset \mathbf{R}^2$. Let $Y = [-2,2]^2 \subset \mathbf{R}^2$. Consider the map $A : X \to Y$ given by

$$A = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}.$$

Find a lower semicontinuous multivalued map $\mathcal{F} : X \rightrightarrows Y$ with the property that $Ax \in \mathcal{F}(x)$ for every $x \in X$.

5.5 Let $X, Y \subset \mathbf{R}$ be cubical sets. Let $f : X \to Y$ be a continuous function. Assume that for each $Q \in \mathcal{K}_1(X)$, $\mathcal{F}(Q)$ is defined and is an acyclic cubical set. Let $P \in \mathcal{K}_0(X)$ and let $Q_P^{\pm} \in \mathcal{K}_1(X)$ be the two edges for which $P \in Q^{\pm}$ (if $P = [0]$ or $P = [2]$, then set $Q^- = Q^+$).

(a) For $P \in \mathcal{K}_0(X)$, define

$$\mathcal{F}(P) := \mathcal{F}(Q_P^+) \cap \mathcal{F}(Q_P^-)$$

and assume that $f(x) \in \mathcal{F}(x)$ for all $x \in X$. Prove that $\mathcal{F}$ is lower semicontinuous.

(b) For $P \in \mathcal{K}_0(X)$, define

$$\mathcal{F}(P) := \mathcal{F}(Q_P^+) \cup \mathcal{F}(Q_P^-)$$

and assume that $f(x) \in \mathcal{F}(x)$ for all $x \in X$. Prove that $\mathcal{F}$ is upper semicontinuous.

(c) Show that the assumption $f(x) \in \mathcal{F}(x)$ is necessary.

5.3 Chain Selectors.

As has been indicated since Chapter 2, our purpose for introducing multi-valued maps is to obtain an outer approximation for continuous functions. Of course, we still need to indicate how this outer approximation can be used to generate homology. By Section 5.1 it is sufficient to indicate how a multivalued map induces a chain map.

Theorem 5.14 *Assume $\mathcal{F} : X \overrightarrow{\to} Y$ is a lower semicontinuous cubical map. Then, there exists a chain map $F : C(X) \to C(Y)$ with the property*

$$|F(\widehat{Q})| \subset \mathcal{F}(\overset{\circ}{Q}) \tag{5.2}$$

for all $Q \in \mathcal{K}(X)$.

Proof: We will construct the homomorphisms $F_k : C_k(X) \to C_k(Y)$ by induction in k.

For $k < 0$, $C_k(X) = 0$, therefore there is no choice but to define $F_k := 0$.

Consider $k = 0$. For each $Q \in \mathcal{K}_0$, choose $P \in \mathcal{K}_0(\mathcal{F}(Q))$ and set

$$F_0(\widehat{Q}) := \widehat{P}. \tag{5.3}$$

Clearly, $|F_0\widehat{Q}| = P \in \mathcal{F}(Q)$. Since, $Q \in \mathcal{K}_0$, $\overset{\circ}{Q}= Q$ and hence $\mathcal{F}(Q) = \mathcal{F}(\overset{\circ}{Q})$. Therefore,

$$|F_0\widehat{Q}| \subset \mathcal{F}(\overset{\circ}{Q}).$$

Furthermore,

$$F_{-1}\partial_0 = 0 = \partial_0 F_0.$$

To continue the induction, suppose now that the homomorphisms $F_i : C_i(X) \to C_i(Y)$, $i = 0, 1, 2, \ldots, k-1$, are constructed in such a way that

$$|F_i\widehat{Q}| \subset \mathcal{F}(\overset{\circ}{Q}) \text{ for all } Q \in \mathcal{K}_i(K),$$

and

$$F_{i-1}\partial_i = \partial_i F_i.$$

Let $\widehat{Q} \in \mathcal{K}_k(X)$. Then $\partial\widehat{Q} = \sum_{j=1}^{m} a_j\widehat{Q}_j$ for some $a_j \in \mathbf{Z}$ and $\widehat{Q}_j \in \mathcal{K}_{k-1}(X)$. Since $\mathcal{F}$ is lower semicontinuous, we have by Proposition 5.13

$$|F_{k-1}\widehat{Q}_j| \subset F(\overset{\circ}{Q}_j) \subset F(\overset{\circ}{Q})$$

for all $j = 1, \ldots, m$. Thus

$$|F_{k-1}\partial\widehat{Q}| \subset F(\overset{\circ}{Q}).$$

$\mathcal{F}(\overset{\circ}{Q}) = \mathcal{F}(x)$ for any $x \in \overset{\circ}{Q}$, hence the set $\mathcal{F}(\overset{\circ}{Q})$ is acyclic. By the induction assumption F_{k-1} is a chain map, so $F_{k-1}\partial\widehat{Q}$ is a cycle. However, by acyclicity, there exists a chain $c \in C_k(F(\overset{\circ}{Q}))$ such that $\partial c = F_{k-1}\partial\widehat{Q}$. Define

$$F_k\widehat{Q} := c.$$

By definition, the homomorphism F_k satisfies the property

$$\partial_k F_k = F_{k-1}\partial_k.$$

Also, if $Q \in \mathcal{K}_k(X)$, then $F_k\widehat{Q} \in C_k(F(\overset{\circ}{Q}))$, hence

$$|F_k\widehat{Q}| \subset F(\overset{\circ}{Q}).$$

Therefore the chain map $F = \{F_k\}_{k\in\mathbf{Z}} : C(X) \to C(Y)$ satisfying (5.2) is well defined. ∎

Observe that in the first nontrivial step (5.3) of the inductive construction of F we were allowed to choose any $P \in \mathcal{K}_0(\mathcal{F}(Q))$. Thus, this procedure allows us to produce many chain maps of the type described in Theorem 5.14. This leads to the following definition.

Definition 5.15 A chain map $F : C(X) \to C(Y)$ satisfying 5.2 is called a *chain selector* of $\mathcal{F}$.

Proposition 5.16 *Assume $\mathcal{F} : X \overrightarrow{\to} Y$ is a lower semicontinuous cubical map and F is a chain selector for $\mathcal{F}$. Then, for any $c \in C(X)$*

$$|F(c)| \subset \mathcal{F}(|c|).$$

Proof: Let $c = \sum_{i=1}^m a_i\widehat{Q}_i$, where $a_i \in \mathbf{Z}$, $a_i \neq 0$. Then

$$\begin{aligned} |F(c)| &= |\sum_{i=1}^m a_i F(\widehat{Q}_i)| \\ &\subset \bigcup_{i=1}^m |F(\widehat{Q}_i)| \end{aligned}$$

$$
\begin{aligned}
&\subset \bigcup_{i=1}^{m} \mathcal{F}(\overset{\circ}{Q}_i) \\
&\subset \bigcup_{i=1}^{m} \mathcal{F}(Q_i) \\
&= \mathcal{F}(\bigcup_{i=1}^{m} Q_i) = \mathcal{F}(|c|).
\end{aligned}
$$

■

The following theorem justifies the use of chain selectors.

Theorem 5.17 *Let $F, G : C(X) \to C(Y)$ be chain selectors for the lower semicontinous cubical map $\mathcal{F} : X \overrightarrow{\to} Y$. Then, F is chain homotopic to G, and hence, they induce the same homomorphism in homology.*

Proof: A chain homotopy $D = \{D_k : C_k(X) \to C_{k+1}(Y)\}_{k \in \mathbf{Z}}$ joining F to G can be constructed by induction.

For $k < 0$, there is no choice but to set $D_k := 0$.

Thus assume $k \geq 0$ and D_i is defined for $i < k$ in such a way that

$$\partial_{i+1} \circ D_i + D_{i-1} \circ \partial_i = G_i - F_i, \tag{5.4}$$

and for all $Q \in \mathcal{K}_i(X)$ and $c \in C_i(Q)$,

$$|D_i(c)| \subset \mathcal{F}(\overset{\circ}{Q}). \tag{5.5}$$

Let $\widehat{Q} \in C_k(X)$ be an elementary k-cube. Put

$$c := G_k(\widehat{Q}) - F_k(\widehat{Q}) - D_{k-1}\partial_k(\widehat{Q}).$$

It follows easily from the induction assumption that c is a cycle. Moreover, if $\partial\widehat{Q} = \sum_{i=1}^{m} a_i \widehat{P}_i$ for some $a_i \neq 0$ and $P_i \in \mathcal{K}_{k-1}(X)$, then P_i are faces of Q and by Proposition 5.13

$$|D_{k-1}\partial(\widehat{Q})| \subset \bigcup_{i=1}^{m} |D_{k-1}(\widehat{P}_i)| \subset \bigcup_{i=1}^{m} \mathcal{F}(\overset{\circ}{P}_i) \subset \mathcal{F}(\overset{\circ}{Q}).$$

Consequently,

$$|c| \subset |G_k(\widehat{Q})| \cup |F_k(\widehat{Q})| \cup |D_{k-1}\partial_k(\widehat{Q})| \subset \mathcal{F}(\overset{\circ}{Q}).$$

It follows that $c \in Z_k(\mathcal{F}(\mathring{Q}))$. Since $\mathcal{F}(\mathring{Q})$ is acyclic, we conclude that there exists a $c' \in C_{k+1}(\mathcal{F}(\mathring{Q}))$ such that $\partial c' = c$. We put $D_k(\widehat{Q}) := c'$. One easily verifies that the induction assumptions are satisfied, therefore the construction of the required homotopy is completed. ∎

The above theorem lets us make the following fundamental definition.

Definition 5.18 Let $\mathcal{F} : X \overrightarrow{\to} Y$ be a lower semicontinuous cubical maps. Let $F : C(X) \to C(Y)$ be a chain selector of $\mathcal{F}$. The *homology map* of $\mathcal{F}$, $\mathcal{F}_* : H_*(X) \to H_*(Y)$ is defined by

$$\mathcal{F}_* := F_*.$$

Keep in mind that the purpose of introducing multivalued maps is in order to be able to compute the homology of a continuous map by some systematic method of approximation. Obviously, and we saw this in Chapter 2, what procedure one uses or the amount of computation one is willing to do determines how sharp an approximation one obtains. An obvious question is how much does this matter.

Definition 5.19 Let X and Y be cubical spaces and let $\mathcal{F}, \mathcal{G} : X \overrightarrow{\to} Y$ be lower semicontinuous cubical maps. $\mathcal{F}$ is a *submap* of $\mathcal{G}$ if

$$\mathcal{F}(x) \subset \mathcal{G}(x)$$

for every $x \in X$. This is denoted by $\mathcal{F} \subset \mathcal{G}$.

Proposition 5.20 *If $\mathcal{F}, \mathcal{G} : K \overrightarrow{\to} L$ are two lower semicontinuous cubical maps and $\mathcal{F}$ is a submap of $\mathcal{G}$, then $\mathcal{F}_* = \mathcal{G}_*$.*

Proof: Let F be a chain selector of $\mathcal{F}$. Then, F is also a chain selector of $\mathcal{G}$. Hence, by definition

$$\mathcal{F}_* = F_* = \mathcal{G}_*.$$

∎

A fundamental property of maps is that they can be composed. In the case of multivalued maps $\mathcal{F} : X \overrightarrow{\to} Y$ and $\mathcal{G} : Y \overrightarrow{\to} Z$ we will construct the multivalued map $\mathcal{G} \circ \mathcal{F} : X \overrightarrow{\to} Z$, called the *superposition* of $\mathcal{F}$ and $\mathcal{G}$ by setting

$$\mathcal{G} \circ \mathcal{F}(x) := \mathcal{G}(\mathcal{F}(x))$$

for every $x \in X$.

Proposition 5.21 *If $\mathcal{F} : X \overrightarrow{\rightarrow} Y$ and $\mathcal{G} : Y \overrightarrow{\rightarrow} Z$ are lower semicontinuous cubical maps and $\mathcal{G} \circ \mathcal{F}$ is acyclic then $(\mathcal{G} \circ \mathcal{F})_* = \mathcal{G}_* \circ \mathcal{F}_*$.*

Proof: Let $F \in C(\mathcal{F})$ and $G \in C(\mathcal{G})$. Then by Proposition 5.16 for any $Q \in \mathcal{K}(X)$

$$|(G(F(\widehat{Q}))| \subset \mathcal{G}(|F(\widehat{Q})|) \subset \mathcal{G}(\mathcal{F}(\overset{\circ}{Q})).$$

Hence $G \circ F \in C(\mathcal{G} \circ \mathcal{F})$. But we can compose chain maps and hence

$$(GF)_* = G_* F_* = \mathcal{G}_* \mathcal{F}_*.$$

■

5.4 Homology of continuous maps.

We are finally in the position to discuss the homology of continuous maps. Recall the discussion of maps in Chapter 2. There we started with a continuous function and used Taylor's theorem to get bounds on images of the function. These bounds were then used to construct a multivalued map. We would like our discussion of the construction of the multivalued map to be independent of the particular approximation method that is employed. In particular, the simplest possibility would be to describe the approximation directly in terms of the image of the function and the cubes in the cubical spaces. This leads to the following definitions.

Definition 5.22 Let X and Y be cubical sets and let $f : X \to Y$ be a continuous function. A *cubical approximation* to f is a lower semicontinuous multivalued cubical map $\mathcal{F} : X \overrightarrow{\rightarrow} Y$ such that

$$f(x) \in \mathcal{F}(x) \tag{5.6}$$

for every $x \in X$.

We will define the homology of a continuous map in terms of cubical approximations.

Definition 5.23 Let X and Y be cubical sets and let $f : X \to Y$ be a continuous function. Let $\mathcal{F} : X \overrightarrow{\rightarrow} Y$ be a cubical approximation of f. Then, the induced *homology map*, $f_* : H_*(X) \to H_*(Y)$, is given by

$$f_* := \mathcal{F}_*.$$

As is often the case, it is easy to make a definition, but showing that it is well defined or even applicable is harder. There are at least two questions that need to be answered before we can be content with this approach to defining the homology of a continuous map. First, observe that given a continuous function, there may be many cubical approximations. Thus, we will need to show that all cubical approximations of a given function give rise to the same homomorphism on homology. This will be the content of Section 5.4.1. Second, given cubical sets and a continuous map between them it need not be the case that there exists a cubical approximation. We will deal with this problem in Section 5.4.2.

5.4.1 Cubical Approximations

From the point of view of computations one typically wants a cubical approximation whose images are as small as possible.

Definition 5.24 Let X and Y be cubical sets and let $f : X \to Y$ be a continuous function. The *minimal approximation*, $\mathcal{M}_f : X \overrightarrow{\to} Y$, of f is defined by

$$\mathcal{M}_f(x) := \operatorname{ch}(f(\operatorname{ch}(x))). \tag{5.7}$$

If $\mathcal{M}_f$ is a cubical map, then $\mathcal{M}_f$ is refered to as the *minimal cubical approximation.*

Example 5.25 Consider the continuous function $f : [0, 3] \to [0, 3]$ given by $f(x) = x/3$. Figure 5.2 indicates the graph of f and its minimal cubical approximation $\mathcal{M}_f$. To verify that $\mathcal{M}_f$ really is the minimal cubical approximation just involves checking the definition on all the elementary cubes in $[0, 3]$. To begin with consider $[0] \in \mathcal{K}_0$. $\operatorname{ch}[0] = [0]$ and $f(0) = 0$, therefore $\mathcal{M}_f(0) = 0$. On the other hand, while $\operatorname{ch}[1] = [1]$, $f(1) = 1/3 \in [0, 1]$ and hence $\operatorname{ch} f(1) = [0, 1]$. Therefore, $\mathcal{M}_f([1]) = [0, 1]$. The rest of the elementary cubes can be checked in a similar manner.

Observe that if any cube from the graph of $\mathcal{M}_f$ were removed, then the graph of f would no longer be contained in the graph of $\mathcal{M}_f$. In this sense $\mathcal{M}_f$ is minimal.

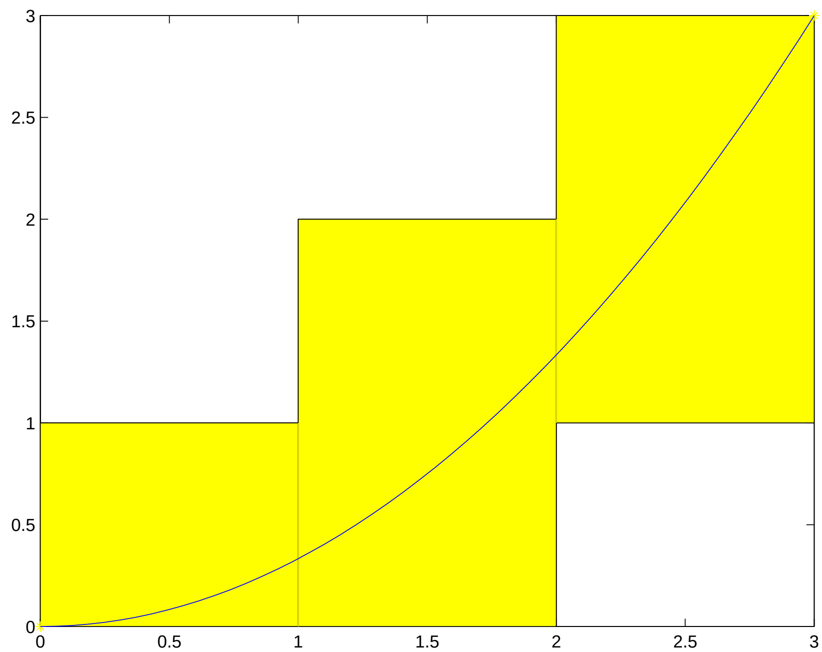


Figure 5.2: The graph of the continuous function $f : [0,3] \to [0,3]$ and the graph of $\mathcal{M}_f$.

Example 5.26 As is suggested by the previous definition, it is not true that a minimal approximation is necessarily a cubical map. Consider the cubical set X consisting of the union of the elementary cubes:

$$K_1 := [0] \times [0,1] \qquad K_2 := [0,1] \times [1]$$
$$K_3 := [1] \times [0,1] \qquad K_4 := [0,1] \times [0].$$

Define the map $\lambda : [0,1] \to X$ for $t \in [0,1]$ by

$$\lambda(t) := \begin{cases} (0, 4t) & \text{if } t \in [0, 1/4] \\ (4t-1, 1) & \text{if } t \in [1/4, 1/2] \\ (1, 3-4t) & \text{if } t \in [1/2, 3/4] \\ (4-4t, 0) & \text{if } t \in [3/4, 1] \end{cases}$$

and the map $f : X \to X$ for $(x_1, x_2) \in X$ by

$$f(x_1, x_2) := \begin{cases} \lambda(x_2) & \text{if } (x_1, x_2) \in K_1 \cup K_3 \\ \lambda(x_1) & \text{if } (x_1, x_2) \in K_2 \cup K_4. \end{cases}$$

Then f is continuous and for $(x_1, x_2) \in \overset{\circ}{K}_i$

$$\mathcal{M}_f(x_1, x_2) = \mathrm{ch}\,(f(\mathrm{ch}\,(x_1, x_2))) = \mathrm{ch}\,(f(K_i)) = \mathrm{ch}\,(X) = X.$$

Since X is not acyclic, it follows that $\mathcal{M}_f$ is not acyclic and consequently not a cubical map.

This example shows that given two cubical sets and a continuous map between them, the minimal approximation need not be a cubical approximation. One can ask if there is a different cubical approximation for the continuous map. As the following proposition indicates, the answer is no.

Proposition 5.27 *Let X and Y be cubical sets, let $f : X \to Y$ be a continuous function and let $\mathcal{F} : X \overrightarrow{\to} Y$ be a cubical approximation to f. Then, $\mathcal{M}_f$ is a submap of $\mathcal{F}$.*

Proof: Let $x \in X$. Then, there exists $Q \in \mathcal{K}(X)$ such that $x \in \overset{\circ}{Q}$. Since $\mathcal{F}$ is a cubical map, $\mathcal{F}(x) = \mathcal{F}(\overset{\circ}{Q})$ and in particular, $\mathcal{F}(x)$ is closed. Now, let $\{x_n\} \subset \overset{\circ}{Q}$ such that $x_n \to \bar{x}$. By continuity of f, $f(\bar{x}) \in \mathcal{F}(\overset{\circ}{Q})$ which in turn implies that $f(Q) \subset \mathcal{F}(\overset{\circ}{Q})$.

Since $x \in \overset{\circ}{Q}$, $\mathrm{ch}\,(x) = Q$. Thus,

$$\mathcal{M}_f(x) = \mathrm{ch}\,(f(\mathrm{ch}\,(x))) = \mathrm{ch}\,(f(Q)) \subset \mathcal{F}(\overset{\circ}{Q}) = \mathcal{F}(x).$$

■

One way to interpret this proposition is to realize that it implies that a cubical approximation for a continuous function f exists if and only if $\mathcal{M}_f$ is a cubical approximation. We have, of course, given a formula for $\mathcal{M}_f$, therefore what remains is to understand when $\mathcal{M}_f$ can fail to be a cubical approximation. The failure in Example 5.26 was due to fact that the map was not acyclic. The next proposition indicates that this is the only reason that $\mathcal{M}_f$ can fail to be cubical.

Proposition 5.28 *Let X and Y be cubical sets and let $f : X \to Y$ be a continuous function. If $\mathcal{M}_f(x)$ is acyclic for each $x \in X$, then $\mathcal{M}_f$ is a cubical approximation.*

Proof: Let $x \in X$. Obviously $f(x) \in f(\operatorname{ch}(x)) \subset \mathcal{M}_f(x)$. The fact that $\mathcal{M}_f$ restricted to $\overset{\circ}{Q}$ is constant follows from the fact that if $x, y \in \overset{\circ}{Q}$, then $\operatorname{ch}(x) = \operatorname{ch}(y)$. The lower semicontinuity of $\mathcal{M}_f$ follows from the fact that if P is a face of Q, then $\operatorname{ch}(P) \subset \operatorname{ch}(Q)$. ■

Corollary 5.29 *Let X and Y be cubical sets, let $f : X \to Y$ be a continuous function. If f has a cubical approximation, then $f_* : H_*(X) \to H_*(Y)$ is well defined.*

Proof: By 5.27, if there exists a cubical approximation $\mathcal{F} : X \overrightarrow{\to} Y$ to f, then $\mathcal{M}_f$ is a submap of $\mathcal{F}$. Since $\mathcal{F}$ is acyclic, $\mathcal{M}_f$ is acyclic and hence by 5.28 $\mathcal{M}_f$ is a cubical map. Thus, $\mathcal{M}_{f*}$ is defined. By Proposition 5.20, $\mathcal{F}_* = \mathcal{M}_{f*}$.

Now assume that $\mathcal{G} : X \overrightarrow{\to} Y$ is another cubical approximation to f, then $\mathcal{M}_f$ is a submap of $\mathcal{G}$ and so

$$\mathcal{F}_* = \mathcal{M}_{f*} = \mathcal{G}_*.$$

■

Proposition 5.30 *Let X be a cubical set. Conside the identity map $\operatorname{id}_X : X \to X$. Then, $\mathcal{M}_{\operatorname{id}_X}$ is a cubical approximation of id_X and*

$$(\operatorname{id}_X)_* = \operatorname{id}_{H_*(X)}$$

Proof: By Proposition 4.16

$$\mathcal{M}_{\operatorname{id}_X}(x) = \operatorname{ch}(x),$$

which, by Proposition 4.68 is acyclic. Therefore, $\mathcal{M}_{\operatorname{id}_X}$ is a cubical approximation of id_X and

$$(\operatorname{id}_X)_* = \mathcal{M}_{\operatorname{id}_X *}.$$

Let $Q \in \mathcal{K}(X)$. Then

$$|\operatorname{id}_{C(X)}(\widehat{Q})| = Q = \operatorname{ch}(\overset{\circ}{Q}) = \mathcal{M}_{\operatorname{id}_X}(\overset{\circ}{Q}).$$

Therefore, $\operatorname{id}_{C(X)}$ is a chain selector for $\mathcal{M}_{\operatorname{id}_X}$. Finally, it is easy to check that $\operatorname{id}_{C(X)}$ induces the identity map $\operatorname{id}_{H_*(X)}$ on homology. ■

Proposition 5.31 *Let $f : X \to X$ be a continous map on a connected cubical set. If $\mathcal{M}_f$ is a cubical approximation of f, then $f_* : H_0(X) \to H_0(X)$ is the identity map.*

Proof: The homology map $f_* : H_0(X) \to H_0(X)$ is determined by an appropriate chain map $F_0 : C_0(X) \to C_0(X)$. Which in turn can be determined by $\mathcal{M}_f$ acting on $\mathcal{K}_0(X)$. So let $Q \in \mathcal{K}_0(X)$. By definition $\mathcal{M}_f(Q) = \mathrm{ch}\,(f(Q))$ which is an elementary cube. Let $P \in \mathcal{K}_0(\mathrm{ch}\,(f(Q)))$. Then, we can define $F_0(\widehat{Q}) = \widehat{P}$, in which case $f_*([\widehat{Q}]) = [\widehat{P}]$. By Theorem 4.51, $[\widehat{Q}] = [\widehat{P}] = 1 \in H_0(X)$, and hence we have the identity map on $H_0(X)$. ∎

Proposition 5.32 *Let X, Y, and Z be cubical sets. Assume $f : X \to Y$ and $g : Y \to Z$ are continuous maps such that $\mathcal{M}_f$, $\mathcal{M}_g$ and $\mathcal{M}_{g\circ f}$ are cubical approximations. Then,*

$$(g \circ f)_* = g_* \circ f_*$$

Proof: Observe that

$$\mathcal{M}_{g\circ f}(x) = \mathrm{ch}\,(g(f(\mathrm{ch}\,(x)))) \subset \mathrm{ch}\,(g(\mathrm{ch}\,(f\mathrm{ch}\,(x)))) = \mathcal{M}_g(\mathcal{M}_f(x)),$$

i.e. $\mathcal{M}_{g\circ f} \subset \mathcal{M}_g \circ \mathcal{M}_f$. Therefore, from Propositions 5.20 and 5.21

$$(g \circ f)_* = (\mathcal{M}_{g\circ f})_* = (\mathcal{M}_g \circ \mathcal{M}_f)_* = (\mathcal{M}_g)_* \circ (\mathcal{M}_f)_* = g_* \circ f_*.$$

∎

5.4.2 Rescaling

So far we are able to define the homology map of a continuous function when a cubical approximation exists. Unfortunately, as was indicated in Example 5.26 not every map admits a cubical approximation. We encountered this problem before in Section 2. There we adopted the procedure of subdividing the intervals of our graph. We could do the same thing here, i.e. we could try to make the images of all elementary cubes acyclic by subdividing the domain of the map into smaller cubes. However, that would require developing the homology theory for cubical sets defined on fractional grids. Obviously, this could be done, but it is not necessary. Instead we take an equivalent path based on rescaling the domain of the function to a large size. Observe that if we make the domain large, then as a fraction of the size of the domain the elementary cubes become small. This leads to the following notation.

Definition 5.33 A *scaling vector* is a vector of positive integers

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbf{Z}^n$$

and gives rise to the *scaling* $\Lambda^\alpha : \mathbf{R}^n \to \mathbf{R}^n$ defined by

$$\Lambda^\alpha(x) := (\alpha_1 x_1, \alpha_2 x_2, \ldots \alpha_n x_n).$$

If $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ is another scaling vector, then set

$$\alpha\beta := (\alpha_1\beta_1, \alpha_2\beta_2, \ldots, \alpha_d\beta_n).$$

The following properties of scalings are straighforward and left as an exercise.

Proposition 5.34 *Let α and β be a scaling vector. Then, Λ^α maps cubical sets onto cubical sets and $\Lambda^\beta \circ \Lambda^\alpha = \Lambda^{\alpha\beta}$.*

Definition 5.35 Let $X \subset \mathbf{R}^n$ be a cubical set and let $\alpha \in \mathbf{Z}^n$ be a scaling vector. Define $\Lambda^\alpha_X := \Lambda^\alpha \mid_\Lambda$. The *scaling of X* by α is

$$X^\alpha := \Lambda^\alpha_X(X) = \Lambda^\alpha(X).$$

Example 5.36 Recall that Example 5.26 described a function f for which $\mathcal{M}_f$ was not a cubical approximation. The first step in dealing with this problem involves rescaling the space X. Figure 5.3 shows the effect of scaling X using the scaling vector $\alpha = (4, 4)$.

We begin by establishing that scalings are nice continuous maps in the sense that they have cubical approximations.

Proposition 5.37 *Let X, Y, and Z be cubical sets and let α and β be scaling vectors. If $\Lambda^\alpha(X) \subset Y$, then $\mathcal{M}_{\Lambda^\alpha_X}$ is a cubical approximation. Moreover, if $\Lambda^\beta(Y) \subset Z$, then $\mathcal{M}_{\Lambda^\beta_Y \circ \Lambda^\alpha_X}$ is a cubical approximation.*

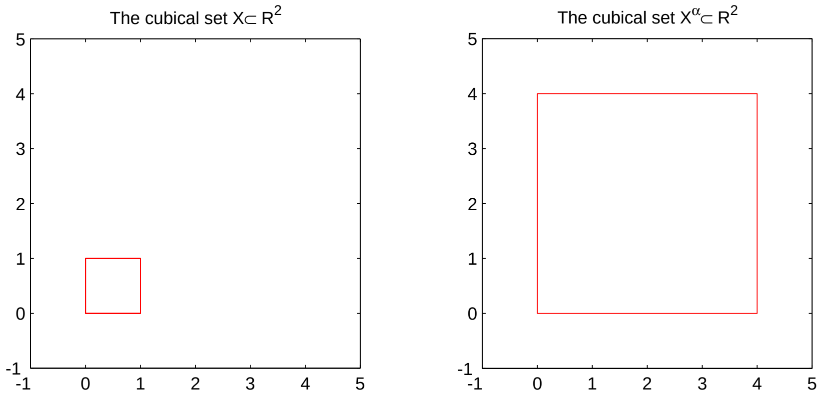


Figure 5.3: The space X and X^α where $\alpha = (4, 4)$.

Proof: By definition, for any $x \in X$

$$\mathcal{M}_{\Lambda_X^\alpha}(x) = (\mathrm{ch}\,(\Lambda_X^\alpha(\mathrm{ch}\,(x))).$$

Since $\mathrm{ch}\,(x)$ is a cube, it follows that $\Lambda_X^\alpha(\mathrm{ch}\,(x))$ is a convex cubical set. Therefore, by Corollary **??** the set $\mathcal{M}_{\Lambda_X^\alpha}(x)$ is convex and consequently acyclic by Proposition 4.63.

Since $\Lambda_Y^\beta \circ \Lambda_X^\alpha = \Lambda_X^{\alpha\beta}$, the map $\Lambda_Y^\beta \circ \Lambda_X^\alpha$ is also simple. To show that $\mathcal{M}_{\Lambda_Y^\beta} \circ \mathcal{M}_{\Lambda_X^\alpha}$ is acyclic, observe that

$$\mathcal{M}_{\Lambda_Y^\beta} \circ \mathcal{M}_{\Lambda_X^\alpha}(x) = (\mathrm{ch}\,(\Lambda_Y^\beta(\mathrm{ch}\,(\Lambda_X^\alpha(\mathrm{ch}\,(x)))))).$$

Therefore the acyclicity of $\mathcal{M}_{\Lambda_Y^\beta} \circ \mathcal{M}_{\Lambda_X^\alpha}$ follows by the same argument as in the previous paragraph. ■

Since scalings have cubical approximations they induce maps on homology. Furthermore, since scalings just change the size of the space one would expect that they induce isomorphisms on homology. The simplest way to check this is to show that their homology maps have inverses. Therefore, given a cubical set X and a scaling vector α let $\Omega_X^\alpha : X^\alpha \to X$ be defined by

$$\Omega_X^\alpha(x) := (\alpha_1^{-1}x_1, \alpha_2^{-1}x_2, \ldots, \alpha_n^{-1}x_n).$$

Obviously, $\Omega_X^\alpha = (\Lambda_X^\alpha)^{-1}$. However, we need to know that it induces a map on homology.

Lemma 5.38 $\mathcal{M}_{\Omega_X^\alpha} : X^\alpha \overrightarrow{\rightarrow} X$ *is a cubical approximation of* Ω_X^α

Proof: Let $x \in X^\alpha$. Then $x \in \overset{\circ}{P} \in \mathcal{K}_k(X^\alpha)$. Since $X^\alpha = \Lambda^\alpha(X)$, there exists $Q \in \mathcal{K}_m(X)$, $m \geq k$, such that $P \in \Lambda^\alpha(Q)$. Now

$$\begin{aligned}\mathcal{M}_{\Omega_X^\alpha}(x) &= \operatorname{ch}(\Omega_X^\alpha(\operatorname{ch}(x))) \\ &= \operatorname{ch}(\Omega_X^\alpha(P)) \\ &= Q\end{aligned}$$

which is acyclic. ∎

Proposition 5.39 *If X is a cubical set and α is a scaling vector, then*

$$(\Lambda_X^\alpha)_* : H_*(X) \to H_*(X^\alpha) \quad \text{and} \quad (\Omega_X^\alpha)_* : H_*(X^\alpha) \to H_*(X)$$

are isomorphisms. Furthermore,

$$(\Lambda_X^\alpha)_*^{-1} = (\Omega_X^\alpha)_*.$$

Proof: It follows from Proposition 5.37 and Lemma 5.38 that $\mathcal{M}_{\Lambda_X^\alpha}$ and $\mathcal{M}_{\Omega_X^\alpha}$ are cubical approximations. Thus, by Proposition 5.37, $\mathcal{M}_{\Lambda_X^\alpha \circ \Omega_X^\alpha}$ and $\mathcal{M}_{\Omega_X^\alpha \circ \Lambda_X^\alpha}$ are cubical approximations. Hence, by Propositions 5.32, and 5.30,

$$(\Lambda_X^\alpha)_* \circ (\Omega_X^\alpha)_* = (\Lambda_X^\alpha \circ \Omega_X^\alpha)_* = \operatorname{id}_{X^\alpha *} = \operatorname{id}_{H_*(X^\alpha)}$$

and

$$(\Omega_X^\alpha)_* \circ (\Lambda_X^\alpha)_* = (\Omega_X^\alpha \circ \Lambda_X^\alpha)_* = \operatorname{id}_{X*} = \operatorname{id}_{H_*(X)}.$$

∎

As was indicated in the introduction, the purpose of scaling is to allow us to define the homology of an arbitrary continuous map between cubical sets. Thus, given a continuous map $f : X \to Y$ and a scaling vector α define

$$f^\alpha := f \circ \Omega_X^\alpha$$

Observe that $f^\alpha : X^\alpha \to Y$.

Example 5.40 To indicate the relationship between f and f^α we return to Example 5.26. Consider $\alpha = (4,4)$. As was already mentioned Figure 5.3 shows X^α and $f^\alpha : X^\alpha \to X$. Now consider $\mathcal{M}_{f^\alpha}$. Consider $Q = [0,1] \times [4]$. Let $(x_1, x_2) \in \overset{\circ}{Q}$. Then

$$\begin{aligned}
\mathcal{M}_{f^\alpha}(x_1, x_2) &= \operatorname{ch}(f^\alpha(\operatorname{ch}(x_1, x_2))) \\
&= \operatorname{ch}(f^\alpha(Q)) \\
&= \operatorname{ch}(\lambda([0, 1/4])) \\
&= \operatorname{ch}([0] \times [0,1]) \\
&= [0] \times [0,1]
\end{aligned}$$

which is acyclic. Similar checks at all the points on X^α shows that $\mathcal{M}_{f^\alpha}$ is acyclic and hence $\mathcal{M}_{f^\alpha}$ is a cubical approximation.

Proposition 5.41 *Let X and Y be cubical sets and $f : X \to Y$ be continuous. Then there exists a scaling vector α such that $\mathcal{M}_{f^\alpha}$ is a cubical approximation of f^α. Moreover, if β is another scaling vector such that $\mathcal{M}_{f^\beta}$ is a cubical approximation of f^β, then*

$$f^\alpha_*(\Lambda^\alpha_X)_* = f^\beta_*(\Lambda^\beta_X)_*$$

Proof: Choose $\delta > 0$ such that for $x, y \in K$

$$\operatorname{dist}(x,y) \leq \delta \quad \Rightarrow \quad \operatorname{dist}(f(x), f(y)) \leq \frac{1}{2} \tag{5.8}$$

and let α be a scaling vector such that $\min\{\alpha_i \mid i = 1, \ldots, n\} \geq 1/\delta$. Since $\operatorname{diam} \operatorname{ch}(x) \leq 1$, we get from (5.8) that

$$\operatorname{diam} f^\alpha(\operatorname{ch}(x)) \leq \frac{1}{2}.$$

Therefore it follows from Proposition 4.68 that $\mathcal{M}_{f^\alpha}$ is acyclic, i.e. $\mathcal{M}_{f^\alpha}$ is a cubical approximation of f^α.

Now assume that the scaling vector β is such that $\mathcal{M}_{f^\beta}$ is also a cubical approximation. Also, assume for the moment that for each $i = 1, \ldots, n$, $\alpha_i | \beta_i$. Let $\gamma_i := \frac{\beta_i}{\alpha_i}$. Then $\gamma = (\gamma_1, \ldots, \gamma_n)$ is a scaling vector. Clearly, $\Lambda^\beta_X = \Lambda^\gamma_{K^\alpha} \circ \Lambda^\alpha_X$. Therefore it follows from Proposition 5.32 that

$$(\Lambda^\beta_X)_* = (\Lambda^\gamma_{X^\alpha})_* \circ (\Lambda^\alpha_X)_*.$$

On the other hand we also have

$$f^\alpha = f^\beta \circ \Lambda^\gamma_{X^\alpha},$$

hence

$$\mathcal{M}_{f^\beta}(\mathcal{M}_{\Lambda^\gamma_{X^\alpha}}(x)) = \mathrm{ch}\,(f^\beta(\mathrm{ch}\,(\Lambda^\gamma_{X^\alpha}(\mathrm{ch}\,(x))))) = \mathrm{ch}\,(f^\alpha(\mathrm{ch}\,(x))) = \mathcal{M}_{f^\alpha}(x).$$

Therefore we get from Proposition 5.32 that

$$f^\alpha_* = f^\beta_* \circ (\Lambda^\gamma_{X^\alpha})_*$$

and consequently

$$f^\alpha_* \circ (\Lambda^\alpha_X)_* = f^\beta_* \circ (\Lambda^\gamma_{X^\alpha})_* \circ (\Lambda^\alpha_X)_* = f^\beta_* \circ (\Lambda^\beta_X)_*.$$

Finally, if it is not true that $\alpha_i \mid \beta_i$ for each $i = 1, \ldots, n$, then let $\theta = \alpha\beta$. By what we have just proven

$$f^\alpha_* \circ (\Lambda^\alpha_X)_* = f^\theta_* \circ (\Lambda^\theta_X)_* = f^\beta_* \circ (\Lambda^\beta_X)_*$$

which settles the general case. ∎

We can now give the general definition for the homology map of a continuous function.

Definition 5.42 Let X and Y be cubical sets and let $f : X \to Y$ be a continuous function. Let α be a scaling vector such that $\mathcal{M}_{f^\alpha}$ is a cubical approximation to f^α. Then, $f_* : H_*(X) \to H_*(Y)$ is defined by

$$f_* = f^\alpha_* \circ \Lambda^\alpha_X.$$

By Proposition 5.41, this definition is independent of the particular scaling vector used. However, we need to reconcile this definition of the homology map with that of Definition 5.23. So assume that $\mathcal{M}_f$ is a cubical approximation of f. Let α be the scaling vector where each $\alpha_i = 1$. Then $f^\alpha = f$ and hence the two definitions of f_* agree.

The final issue we need to deal with involves the composition of continuous functions. We will need the following technical lemma.

Lemma 5.43 *Let X and Y be cubical sets and let $f : X \to Y$ be continuous. Let α be a scaling vector. If $\mathcal{M}_f$ and $\mathcal{M}_{\Lambda^\alpha_Y \circ f}$ are cubical approximations, then $\mathcal{M}_{\Lambda^\alpha_Y} \circ \mathcal{M}_f$ is a cubical map.*

Proof: We only need to verify that $\mathcal{M}_{\Lambda^\alpha_Y} \circ \mathcal{M}_f$ is acyclic. Observe that

$$\mathcal{M}_{\Lambda^\alpha_Y} \circ \mathcal{M}_f(x) = \operatorname{ch}(\Lambda^\alpha_Y(\mathcal{M}_f(x))) = \Lambda^\alpha_Y(\mathcal{M}_f(x)).$$

Since $\mathcal{M}_f(x)$ is acyclic, it follows from Proposition 5.39 that $\Lambda^\alpha_Y(\mathcal{M}_f(x))$ is also acyclic. ■

Proposition 5.44 *Assume $f : X \to Y$ and $g : Y \to Z$ are continuous maps between cubical sets. Then*

$$(g \circ f)_* = g_* \circ f_*$$

Proof: Select a scaling vector β such that $\mathcal{M}_{g^\beta}$ is a cubical approximation and for any $x, y \in Y^\beta$

$$\operatorname{dist}(x, y) \leq 2 \quad \Rightarrow \quad \operatorname{dist}(g^\beta(x), g^\beta(y)) \leq \frac{1}{2}. \tag{5.9}$$

Similarly, select a scaling vector α such that $\mathcal{M}_{f^\alpha}$ and $\mathcal{M}_{h^\alpha}$ are cubical approximations, and for any x, y in X^α

$$\operatorname{dist}(x, y) \leq 2 \quad \Rightarrow \quad \operatorname{dist}(\Lambda^\beta \circ f^\alpha(x), \Lambda^\beta \circ f^\alpha(y)) \leq \frac{1}{2}. \tag{5.10}$$

Then the maps $\Lambda^\beta \circ f^\alpha$ and $g^\beta \circ (\Lambda^\beta \circ f^\alpha) = h^\alpha$ have cubical approximations. Moreover, by 5.9 and 5.10, for any $x \in K^\alpha$

$$\operatorname{diam}(g^\beta \circ \operatorname{ch} \circ \Lambda^\beta \circ f^\alpha \circ \operatorname{ch}(x)) < 1.$$

Therefore by Proposition 4.68

$$\mathcal{M}_{g^\beta} \circ \mathcal{M}_{\Lambda^\beta \circ f^\alpha}(x) = \operatorname{ch} \circ g^\beta \circ \operatorname{ch} \circ \Lambda^\beta \circ f^\alpha \circ \operatorname{ch}(x)$$

is acyclic, i.e. the composition $\mathcal{M}_{g^\beta} \circ \mathcal{M}_{\Lambda^\beta \circ f^\alpha}$ is acyclic. Hence

$$(g^\beta \circ f^\alpha)_* = g^\beta_* \circ f^\alpha_*.$$

By Lemma 5.43 we also have that

$$(\Lambda^\beta_{K^\alpha} \circ f^\alpha)_* = \Lambda^\beta_{K^\alpha *} \circ f^\alpha_*.$$

Let $h := g \circ f$. It follows from Proposition 4.68 that

$$h^\alpha_* = (g^\beta \circ \Lambda^\beta_{K^\alpha} \circ f^\alpha)_* = g^\beta_* \circ (\Lambda^\beta_{K^\alpha} \circ f^\alpha)_*.$$

Hence,

$$\begin{aligned}
(g \circ f)_* &= h_* \\
&= h_*^\alpha \circ (\Lambda_K^\alpha)_* \\
&= g_*^\beta \circ (\Lambda_{K^\alpha}^\beta \circ f^\alpha)_* \circ (\Lambda_K^\alpha)_* \\
&= g_*^\beta \circ (\Lambda_{K^\alpha}^\beta)_* \circ f_*^\alpha \circ (\Lambda_K^\alpha)_* \\
&= g_* \circ f_*.
\end{aligned}$$

■

Exercises

5.6 Prove Proposition 5.34

5.7 Let X, λ, and f be as in Example 5.26.
a) Verify that the scaling by $\alpha := (2, 2)$ is sufficient for $\mathcal{M}_{f^\alpha}$ to be a cubical approximation of f.
b) Find a chain selector of $\mathcal{M}_{f^\alpha}$.
c) Compute the homology map of f. You may either compute it by hand or use the homology program for that.

5.8 Do the same as in Exercise 5.7 for the map given by

$$f(x_1, x_2) := \begin{cases} \lambda(x_2) & \text{if } (x_1, x_2) \in K_1 \\ \lambda(x_1) & \text{if } (x_1, x_2) \in K_2 \\ \lambda(1 - x_2) & \text{if } (x_1, x_2) \in K_3 \\ \lambda(1 - x_1) & \text{if } (x_1, x_2) \in K_4 \end{cases}$$

5.4.3 Homotopy Invariance of Maps

We now have a homology theory at our disposal. Given a cubical set X we can compute its homology groups $H_*(X)$ and given a continuous map f between cubical sets we can compute the induced map on homology f_*. What is missing is how these algebraic objects relate back to topology. Section 4.3 was a partial answer in that we showed that $H_0(X)$ counts the number of connected components of X. In this section we shall pursue the question of when do two continuous maps induce the same homomorphism on homology. In particular, we shall prove the following theorem.

Theorem 5.45 *Let X and Y be cubical sets and let $f, g : X \to Y$ be homotopic maps. Then*

$$f_* = g_*.$$

We shall break up the proof into two cases. The first is trivial, but contains the essential observation. The second case is merely an elaboration of the first needed to overcome some technical difficulties.

By definition $f \sim g$ implies that there exists a continuous function $\Phi : X \times [0,1] \to Y$ such that $\Phi(x, 0) = f(x)$ and $\Phi(x, 1) = g(x)$. Observe that $X \times [0,1]$ is a cubical set. Assume for the moment that there exists a cubical approximation $\mathcal{H} : X \times [0,1] \rightrightarrows Y$ to Φ. Define $\mathcal{H}' : X \rightrightarrows Y$ by

$$\mathcal{H}'(Q) := \mathcal{H}(Q \times [0,1])$$

for every $Q \in \mathcal{K}(X)$.

Lemma 5.46 *$\mathcal{H}' : X \rightrightarrows Y$ is a cubical approximation to both f and g.*

Proof: Clearly, Q is a face of $Q \times [0,1]$. Therefore, by Proposition 5.13, $\mathcal{H}(\mathring{Q}) \subset \mathcal{H}(\overset{\circ}{Q \times [0,1]})$. However, $f(x) \in \mathcal{H}(x, 0)$ for all $x \in X$. Therefore, $f(\mathring{Q} \times [0]) \subset \mathcal{H}(\mathring{Q})$ and in particular, for any $x \in X$, $f(x) \in \mathcal{H}'(x)$. A similar argument holds for g. ∎

Corollary 5.47 *If the homotopy from f to g has a cubical approximation then*

$$f_* = g_*.$$

Proof: By definition

$$f_* = \mathcal{H}'_* = g_*.$$

∎

This was the easy case. What makes this simple is that an approximation for the homotopy provides an approximation for both f and g. Of course, Φ need not admit a cubical approximation. However, as was made clear in the previous section, we can obtain a cubical approximation for an appropriate scaling of Φ.

With this in mind, choose a scaling vector α such that $\mathcal{M}_{\Phi^\alpha}$ is a cubical approximation. Observe that $\Phi^\alpha \circ \Lambda^\alpha$ is a homotopy between $f^\alpha \circ \Lambda^\alpha$ and

$g^\alpha \circ \Lambda^\alpha$. If $X \subset \mathbf{R}^n$ then $\alpha \in \mathbf{Z}^{n+1}$. Let $\beta = (\alpha_1, \ldots, \alpha_n)$ and let $\alpha_{n+1} = k$. Set $Z = \Lambda^\beta(X)$. Then

$$\Lambda^\alpha(X) = Z \times [0, k].$$

For $i = 0, \ldots, k$, let $f_i : X \times \{i\}$ be defined by $f_i(x) = f^\alpha(x, i)$ and let $\Psi_i : X \times [i, i+1] \to Y$ be given by $\Psi_i(x, s) = \Phi^\alpha(x, s)$. Then, Ψ_i is a homotopy from f_i to f_{i+1}. However, by assumption Ψ_i has a cubical approximation. Therefore, $f_{i*} = f_{i+1*}$ and hence, $f_{0*} = f_{k*}$. Now observe that $f_* = f_{0*}$ and $g_* = f_{k*}$, therefore

$$f_* = g * .$$

This proves Theorem 5.45 in the general case.

Exercises

5.9 In one of the two previous exercises you should get the trivial homology map. Prove this in a different way, by showing that your f is homotopic to a constant map.

5.5 Lefschetz Fixed Point Theorem

We are now in the position to prove one of the most important results in algebraic topology, the Lefschetz fixed point theorem. Let $f : X \to X$ be a continuous map. $x \in X$ is a *fixed point* of f if $f(x) = x$. The Lefschetz theorem gives conditions on f_* that imply that f has a fixed point. We need a few algebraic preliminaries before we can state and prove the theorem.

Let $A = [a_{ij}]$ be an $n \times n$ matrix. The *trace* of A is defined to be the sum of the diagonal entries, i.e.

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

It is easy to check that if A and B are $n \times n$ matrices, then

$$\operatorname{tr} AB = \sum_{i,j} a_{ij} b_{ji} = \operatorname{tr} BA.$$

Let G be a finitely generated free abelian group and let $\phi : G \to G$ be a group homomorphism. Since G is free abelian, it has a basis and for a particular choice of basis ϕ can written as a matrix A. So in this case define

$$\operatorname{tr} \phi = \operatorname{tr} A.$$

To check that this is a well defined concept, let $\{b_, \ldots, b_n\}$ be a different basis for G. Let $B : G \to G$ be the isomorphism corresponding to the change of basis. In this second basis the matrix representation of ϕ is given by $B^{-1}AB$. Thus,

$$\operatorname{tr}(B^{-1}AB) = \operatorname{tr}(B^{-1}(AB)) = \operatorname{tr}((AB)B^{-1}) = \operatorname{tr} A.$$

We will need to apply these ideas in the context of homology groups. Consider a free chain complex $\{C_k(X), \partial_k\}$ and a chain map $F : C(X) \to C(X)$. Let $H_k(X)$ be the induced homology groups and $f_* : H_*(X) \to H_*(X)$ the induced homology map.

Since $C_k(X)$ is a free abelian group, $\operatorname{tr} F_k$ is well defined for each k. However, the homology groups $H_k(X)$, while abelian need not be free. Let $T_k(X)$ denote the torsion subgroup of $H_k(X)$. Then, $H_k(X)/T_k(X)$ is free abelian. Furthermore, $f_* : H_*(X) \to H_*(X)$ induces a homomorphism

$$\phi_k : H_k(X)/T_k(X) \to H_k(X)/T_k(X).$$

Thus, ϕ_k can be represented as a matrix, and hence $\operatorname{tr} \phi_k$ is well defined.

Definition 5.48 Let X be a cubical set and let $f : X \to X$ be a continuous map. The *Lefschetz number* of f is

$$L(f) := \sum_k (-1)^k \operatorname{tr} \phi_k.$$

Theorem 5.49 Lefschetz Fixed Point Theorem *Let X be a cubical set and let $f : X \to X$ be a continuous map. If $L(f) \neq 0$, then f has a fixed point.*

This theorem is an amazing example of how closely the algebra is tied to the topology. To prove it we need to understand how to relate the topology in the form of the map on the chain complexes to the algebra in the form of the induced homology maps on the free part of the homology groups.

We begin with a technical lemma.

Lemma 5.50 *Let G be a free abelian group, let H be a subgroup and assume that G/H is free abelian. Let $\phi : G \to G$ be a group homomorphism such that $\phi(H) \subset H$. Then, ϕ induces a map $\phi' : G/H \to G/H$ and*

$$\operatorname{tr} \phi = \operatorname{tr} \phi' + \operatorname{tr} \phi \mid_H .$$

Proof: The first step is to understand ϕ'. Since G is free abelian, and H is a subgroup, H is also free abelian. Let $\{\alpha_1, \dots, \alpha_k\}$ be a basis for H and let $\{\beta_1 + H, \dots, \beta_n + H\}$ be a basis for G/H. Then, ϕ' is defined by

$$\phi'(\beta_i + H) = \phi(\beta_i).$$

It is left to the reader to check that ϕ' is a well defined group homomorphism. Given our choice of basis we can represent ϕ' as a matrix $B = [b_{ij}]$. In particular,

$$\phi'(\beta_j + H) = \sum_{i=1}^{n} b_{ij}(\beta_i + H).$$

Similarly, $\phi \mid_H : H \to H$ has the form

$$\phi \mid_H (\alpha_i) = \sum_{i=1}^{k} a_{ij}\alpha_i$$

and so we can write $\phi \mid_H = A = [a_{ij}]$.

Since $G = G/H \oplus H$, $\{\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_n\}$ is a basis for G. Thus,

$$\begin{aligned} \phi(\alpha_j) &= \sum_{i=1}^{k} a_{ij}\alpha_i \\ \phi(\beta_j) &= \sum_{i=1}^{n} b_{ij}\beta_i = h \end{aligned}$$

where $h \in H$. This means that as matrix ϕ has the form

$$\phi = \begin{bmatrix} A & * \\ 0 & B \end{bmatrix}.$$

Clearly, $tr\phi = \operatorname{tr}\phi' + \operatorname{tr}\phi\mid_H$. ■

Theorem 5.51 (Hopf trace theorem) *Let $\{C_k(X), \partial_k\}$ be a free chain complex and $F : C(X) \to C(X)$ a chain map. Let $H_k(X)$ denote the corresponding homology groups with torsion subgroups $T_k(X)$. Let $\phi_k : H_k(X)/T_k(X) \to H_k(X)/T_k(X)$ be the induced homomorphism. Then*

$$\sum_k (-1)^k \operatorname{tr} F_k = \sum_k (-1)^k \operatorname{tr}\phi_k.$$

Proof: We will use the notation from Section 3.7 where $W_k(X)$ denotes the weak boundaries. Recall that

$$B_k(X) \subset W_k(X) \subset Z_k(X) \subset C_k(X).$$

Furthermore, since F is a chain map, each of these subgroups is invariant under F_k, i.e. $F_k(B_k(X)) \subset B_k(X)$, $F_k(W_k(X)) \subset W_k(X)$, etc. From Lemma 5.50 F_k induces maps

$$\begin{aligned} F_k\mid_{W_k(X)} &: W_k(X) \to W_k(X), \\ F'_k &: Z_k(X)/W_k(X) \to Z_k(X)/W_k(X) \\ F''_k &: C_k(X)/Z_k(X) \to C_k(X)/Z_k(X). \end{aligned}$$

From Lemma 3.63 and the following comments we have that for each k, $Z_k(X)/W_k(X)$ and $C_k(X)/Z_k(X)$ are free abelian groups. Therefore, applying Lemma 5.50 twice gives

$$\operatorname{tr} F_k = \operatorname{tr} F''_k + \operatorname{tr} F'_k + \operatorname{tr} F_k\mid_{W_k(X)}. \tag{5.11}$$

However, from Lemma 3.63 $Z_k(X)/W_k(X) \cong H_k(X)/T_k(X)$ and furthermore under this isomorphism, F_k becomes ϕ_k. Therefore, (5.11) becomes

$$\operatorname{tr} F_k = \operatorname{tr} F_k'' + \operatorname{tr} \phi_k + \operatorname{tr} F_k \mid_{W_k(X)} . \tag{5.12}$$

Similarly, $C_k(X)/Z_k(X)$ is isomorphic to $B_{k-1}(X)$ and under this isomorphism F_k'' becomes $F_{k-1} \mid_{B_{k-1}(X)}$. An therefore, (5.12) can be written as

$$\operatorname{tr} F_k = \operatorname{tr} F_{k-1} \mid_{B_{k-1}(X)} + \operatorname{tr} \phi_k + \operatorname{tr} F_k \mid_{W_k(X)} . \tag{5.13}$$

We will now show that $\operatorname{tr} F_k \mid_{W_k(X)} = \operatorname{tr} F_k \mid_{B_k(X)}$. As was done explicitly in Section 3.7 there exists a basis $\{\alpha_1, \ldots, \alpha_l\}$ for $W_k(X)$ and integers $m_1, \ldots, m_l$, such that $\{m_1\alpha_1, \ldots, m_l\alpha_l\}$ is a basis for $B_k(X)$.

Observe that

$$F_k \mid_{W_k(X)} (\alpha_j) = \sum_{i=1}^{l} a_{ij}\alpha_i \tag{5.14}$$

and

$$F_k \mid_{B_k(X)} (m_j\alpha_j) = \sum_{i=1}^{l} b_{ij} m_i \alpha_i \tag{5.15}$$

for appropriate constants a_{ij} and b_{ij}. Both these maps are just restrictions of F_k to the appropriate subspaces. So multiplying (5.14) by m_j must give rise to (5.15) and hence $m_j a_{ij} = b_{ij} m_i$ and in particular $m_i a_{ii} = b_{ii} m_i$. Therefore, $\operatorname{tr} F_k \mid_{W_k(X)} = \operatorname{tr} F_k \mid_{B_k(X)}$. Applying this to (5.13) give

$$\operatorname{tr} F_k = \operatorname{tr} F_{k-1} \mid_{B_{k-1}(X)} + \operatorname{tr} \phi_k + \operatorname{tr} F_k \mid_{B_k(X)} . \tag{5.16}$$

The proof is finished by multiplying (5.16) by $(-1)^k$ and summing. ■

The Hopf trace formula is the key step in the proof of the Lefschetz fixed point theorem. However, before beginning the proof let us discuss the basic argument that will be used. Observe that an equivalent statement to the Lefschetz fixed point theorem is the following: *if f has no fixed points, then* $L(f) = 0$. This is what we will prove. The Hopf trace formula provides us with a means of relating a chain map $F : C(X) \to C(X)$ for f with $L(f)$. In particular, if we could show that $\operatorname{tr} F = 0$, then it would be clear that $L(f) = 0$. Of course, the easiest way to check that $\operatorname{tr} F = 0$ is for all the diagonal entries of F to be zero. However, the diagonal entries of F indicate how the duals of elementary cubes are mapped to themselves. If f has no fixed points then the image of a small cube will not intersect itself and so the diagonal entries are zero. With this argument in mind we turn to the proof, which as is often the case in mathematics, is presented in the reverse order.

Proof of Lefschetz Fixed Point Theorem. Assume f has no fixed points. We want to show that $L(f) = 0$.

The first step is to establish some constants that will used in the proof. Let

$$\epsilon := \min_{x \in X} ||x - f(x)||.$$

Since we are assuming that f has no fixed points and X is cubical, $\epsilon > 0$. Similarly, since X is cubical there exists $\delta > 0$ such that

$$||x - y|| < \delta \quad \Rightarrow \quad ||f(x) - f(y)|| < \epsilon/3.$$

Set $\mu := \min\{\delta, \epsilon/6\}$. Let α be a scaling vector with the property that $\alpha_i > \mu^{-1}$ for each i.

With these constants in mind, consider

$$g := \Lambda_X^\alpha \circ f \circ \Omega_X^\alpha : X^\alpha \to X^\alpha.$$

We will now show that for any $x \in X^\alpha$,

$$\mathcal{M}_g(x) \cap \operatorname{ch}(x) = \emptyset.$$

Let $y \in \operatorname{ch}(x)$. This implies that $||y - x|| \leq 1$ and hence

$$||\Omega_X^\alpha(y) - \Omega_X^\alpha(x)|| < \mu.$$

Therefore,

$$||f \circ \Omega_X^\alpha(y) - f \circ \Omega_X^\alpha(x)|| \leq \epsilon/3.$$

By the definition of ϵ followed by the triangle inequality we have

$$\begin{aligned} \epsilon &< ||\Omega_X^\alpha(x) - f \circ \Omega_X^\alpha(x)|| \\ &< ||\Omega_X^\alpha(x) - f \circ \Omega_X^\alpha(y)|| + ||f \circ \Omega_X^\alpha(y) - f \circ \Omega_X^\alpha(x)|| \end{aligned}$$

This implies that

$$||\Omega_X^\alpha(x) - f \circ \Omega_X^\alpha(y)|| > 2\epsilon/3$$

and therefore,

$$||\Lambda_X^\alpha \circ \Omega_X^\alpha(x) - \Lambda_X^\alpha \circ f \circ \Omega_X^\alpha(y)|| = ||x - \Lambda_X^\alpha \circ f \circ \Omega_X^\alpha(y)|| > \mu^{-1}\frac{2\epsilon}{3} = 4.$$

This inequality holds for all $y \in \operatorname{ch}(x)$, and therefore $\mathcal{M}_g(x) \cap \operatorname{ch}(x) = \emptyset$.

Observe that the this argument is true for any α sufficiently large. Therefore, by Proposition 5.41 we can assume that α was chosen large enough that $\mathcal{M}_g$ is a cubical approximation of g. Let $G : C(X^\alpha) \to C(X^\alpha)$ be a corresponding chain map. The standard basis for $C(X^\alpha)$ is $\widehat{\mathcal{K}}(X^\alpha)$, but $|G(\widehat{Q})| \cap Q = \emptyset$ for all $\widehat{Q} \in \widehat{\mathcal{K}}(X^\alpha)$, and therefore the diagonal entries of G are zero. In particular, $\operatorname{tr} G = 0$ and therefore by the Hopf trace formula, $L(g_*) = 0$.

Finally, by Proposition 5.39

$$L(f_*) = L(g_*).$$

■

Theorem 5.52 *Let X be an acyclic cubical set. Let $f : X \to X$ be continuous. Then, f has a fixed point.*

Proof: Since X is acyclic, the only nonzero homology group is $H_0(X) \cong \mathbf{Z}$. But, by Proposition 5.31, $f_* : H_0(X) \to H_0(X)$ is the identity map. Therefore, $L(f) = 1$. ■

Chapter 6

Homological Algebra

We finished the previous chapter with the Lefschetz fixed point theorem that allowed us to prove the existence of fixed points of maps from the homology map. As we noted before this is a remarkable theorem, but as the following example indicates it has its limitations. Consider for the moment the following almost trivial example. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear map

$$f = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} : \mathbf{R}^2 \to \mathbf{R}^2$$

Obviously the origin is a fixed point. Unfortunately, there is no direct way to apply the Lefschetz theorem to detect this fixed point. To begin with $\mathbf{R}^2$ consists of an infinite number of cubes and hence, is not a cubical set. We could try to get around this problem by restricting the domain and range of the function. We know that the origin is the fixed point, so we could, for example, consider $X := \{x \in \mathbf{R}^2 \mid ||x|| \leq 4\}$. Unfortunately, $f(X) \not\subset X$. We leave it to the reader to check that it is impossible to find a cubical set X that contains a neighborhood of the origin such that $f(X) \subset X$. But to talk about a fixed point we need to have a map of the form $f : X \to X$.

However, the Lefschetz fixed point theorem is too nice a tool to give up trying to extend it to an example such as this. So lets study the problem is little further. In Figure 6.1 the set $X := \{x \in \mathbf{R}^2 \mid ||x|| \leq 4\}$ is indicated in red and its image under f in blue. Obviously, there is a problem in that

$$f([-4,-2) \times [-4,4] \cup (2,4] \times [-4,4]) \cap X = \emptyset.$$

Yellow shows the set

$$E := \mathrm{ch}\,([-4,-2) \times [-4,4] \cup (2,4] \times [-4,4]).$$

Of course, the fixed point that is of interest lies in $X \setminus E$. This suggests that we try to develop a homology theory that begins with the set X but "ignores" the set E. This leads to the notion of relative homology.

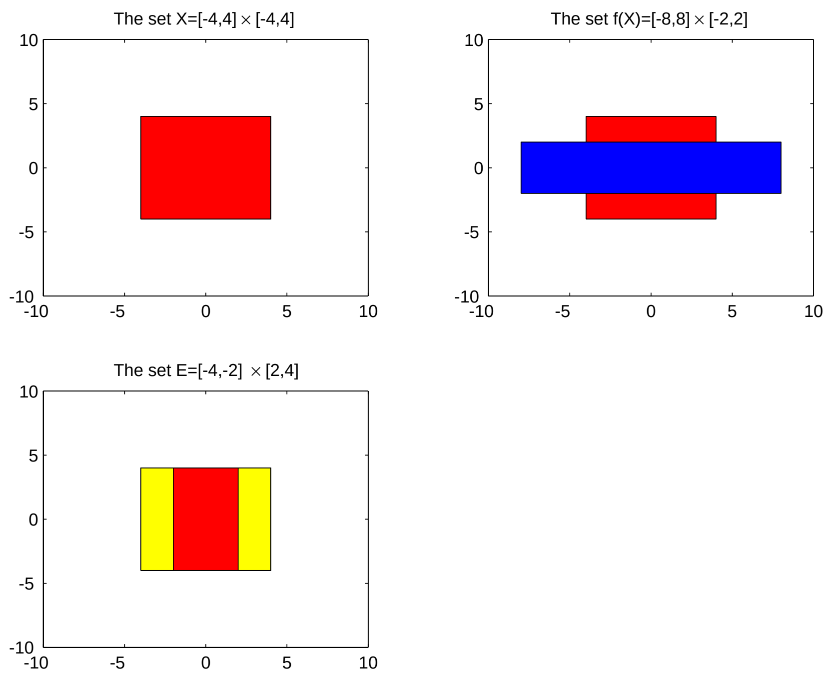


Figure 6.1: The image of the linear map f.

6.1 Relative Homology

Let X and $E \subset X$ be cubical sets. They generate the sets of elementary cubes $\mathcal{K}(X)$ and $\mathcal{K}(E)$ which in turn define the chains $C(X)$ and $C(E)$. Since $\widehat{\mathcal{K}}(X)$ is a basis for $C(X)$ and $\widehat{\mathcal{K}}(E) \subset \widehat{\mathcal{K}}(X)$, the quotient group $C(X)/C(E)$ is a free abelian group. Thus we can make the following definition.

Definition 6.1 Let X and $E \subset X$ be cubical sets. The *relative chains of* X *modulo* E are the free abelian groups

$$C_k(X, E) := C_k(X)/C_k(E).$$

The *relative chain complex of X modulo E* is given by

$$\{C_k(X,E), \partial_k\}$$

where $\partial_k : C_k(X,E) \to C_{k-1}(X,E)$ is the boundary map induced by the standard boundary map on $C_k(X)$.

The relative chain complex gives rise to the *relative k-cycles*,

$$Z_k(X,E) := \ker \partial_k : C_k(X,E) \to C_{k-1}(X,E),$$

the *relative k-boundaries*,

$$B_k(X,E) := \operatorname{image} \partial_{k+1} : C_{k+1}(X,E) \to C_k(X,E),$$

and finally the *relative homology groups*

$$H_k(X,E) := Z_k(X,E)/B_k(X,E).$$

Proposition 6.2 *Let X be a connected cubical set and let E be a non-empty cubical subset of X. Then,*

$$H_0(X,E) = 0.$$

Proof: To compute $H_0(X,E)$ we begin by examining the associated set of cycles $Z_0(X,E)$. Since $\partial_0 = 0$,

$$Z_0(X,E) = C_0(X,E) := C_0(X)/C_0(E).$$

From the proof of Theorem **??**, X connected implies that for any pair $\widehat{P}, \widehat{Q} \in C_0(X)$, there exists $c \in C_1(X)$ such that

$$\partial c = \widehat{P} - \widehat{Q}. \tag{6.1}$$

$E \neq \emptyset$, hence there exists $Q \in \mathcal{K}_0(E)$. By definition, $0 = \widehat{Q} \in Z_0(X,E)$. Therefore, by (6.1)

$$0 = [\widehat{Q}] = [\widehat{P}] \in H_0(X,E)$$

for any $P \in \mathcal{K}_0(X)$. Therefore, $H_0(X,E) = 0$. ∎

Example 6.3 Let $X = [-1,1] \subset \mathbf{R}$ and let $E = \{\pm 1\}$. $H_K(X,E) = 0$ for all $k \geq 2$, since $\mathcal{K}_k(X) = \emptyset$. By Proposition 6.2, $H_0(X,E) = 0$.

Therefore, all that remains to be computed is $H_1(X,E)$. Observe that $C_1(X) \approx \mathbf{Z}^2$ with a basis given by $\{\widehat{[-1,0]}, \widehat{[0,1]}\}$. Since $C_1(E) = 0$, $C_1(X,E) = C_1(X)$. The computation of $C_0(X,E)$ is a little more interesting. The standard basis for $C_0(X)$ is $\{\widehat{[\pm 1]}, \widehat{[0]}\}$, while the corresponding basis for $C_0(E)$ is $\{\widehat{[\pm 1]}\}$. Therefore,

$$C_0(X,E) \cong \mathbf{Z}$$

and generated by $\widehat{[0]}$. Using these bases the matrix representation for $\partial_1 : C_1(X,E) \to C_0(X,E)$ is

$$\partial_1 = [1 \;\; -1].$$

The chain $\widehat{[-1,0]} + \widehat{[0,1]} \in C_1(X,E)$ clearly generates the kernel of ∂_1. Therefore,

$$H_1(X,E) \approx \mathbf{Z}.$$

As will become clear as we progress, relative homology groups are a very powerful tool. So much so that we want a simple shorthand notation for discussing pairs of cubical sets. With this in mind the statement that (X,E) is a *cubical pair* or a *pair of cubical sets* means that X and E are cubical sets and $E \subset X$.

Example 6.4 Let $X = [-3,3]$ and $E = [-3,-1] \cup [1,3]$. The exact same arguments as in the previous example show that $H_k(X,E) = 0$ if $k \neq 1$.

Let us compute $H_1(X,E)$. $C_1(X) \cong \mathbf{Z}^6$ with a basis given by $\{\widehat{[i,i+1]} \mid i = -3,\ldots,2\}$. In contrast to Example 6.3, $C_1(E) \cong \mathbf{Z}^4$ with a basis $\{\widehat{[i,i+1]} \mid i = -3,-2,1,2\}$. This implies that a basis for $C_1(X,E)$ consists of the equivalence classes containing $\{\widehat{[-1,0]}, \widehat{[0,1]}\}$. Repeating this type of argument on the level of the 0-chains we see that $C_0(X,E) \cong \mathbf{Z}$ with a basis consisting of the equivalence class defined by $\widehat{[0]}$. Observe that on the level of relative chains we have the same chain complex as in Example 6.3. Therefore,

$$H_*([-3,3],[-3,-1] \cup [1,3]) \cong H_*([-1,1],\{\pm 1\}).$$

One can ask whether these two examples are merely a coincidence or represent a deeper fact. Since in the relative chains of the pair (X,E) one quotients out by those elementary chains which lie in the subspace, it seems reasonable to conjecture that if one adds the same cubes to both X and E, then the group of relative chains does not chain and hence the homology should not change. Theorem 6.5, presented shortly confirms this, though at first glance its statement may appear somewhat different.

Of course to compare the relative homology groups of different pairs we need to be able to talk about maps. So let (X, E) and (Y, B) be cubical pairs. Let $f : X \to Y$ be a continuous map. The most basic question is whether f induces a map from $H_*(X, E)$ to $H_*(Y, B)$. If this is to be the case then there must be an associated chain map $F : C(X, E) \to C(Y, B)$. However, this can only occur if $F(C(E)) \subset C(B)$. This leads to the following condition.

$$f : (X, E) \to (Y, B)$$

is a *continuous map between cubical pairs* if $f : X \to Y$ is continuous and $f(E) \subset B$.

To generate a map on the level of relative homology, i.e. $f_* : H_*(X, E) \to H_*(Y, B)$ we proceed as before. $f : X \to Y$ is continuous and so for an appropriate scaling vector α, $\mathcal{M}_{f^\alpha} : X \overrightarrow{\to} Y$ is a cubical approximation. Since $f(E) \subset B$ and B is cubical, $\mathcal{M}_{f^\alpha}(E) \subset B$. Now let $F : C(X) \to C(Y)$ be a chain selector for $\mathcal{M}_{f^\alpha}$. For any $Q \in \mathcal{K}(E)$, $|F(\widehat{Q})| \subset \mathcal{M}_{f^\alpha}(Q) \subset B$, and hence $F(C(E)) \subset C(B)$. Thus, F induces a chain map between the relative chain complexes, i.e. with a slight abuse of notation we can write $F : C(X, E) \to C(Y, B)$. Then we define $f_* : H_*(X, E) \to H_*(Y, B)$ by

$$f_*([\gamma]) := [F(\gamma)].$$

Theorem 6.5 (Excision Isomorphism Theorem) *Let (X, E) be a cubical set. Let $U \subset E$ be a representable set such that $E \setminus U$ is a cubical set. Then, the inclusion map $\iota : (X \setminus U, E \setminus U) \to (X, E)$ induces an isomorphism*

$$e_* : H_*(X \setminus U, E \setminus U) \to H_*(X, E).$$

Proof: Since $\iota : (X \setminus U, E \setminus U) \to (X, E)$ is the inclusion map, $\mathcal{M}_\iota(Q) = Q$ for every $Q \in \mathcal{K}(X \setminus U)$. Thus, the inclusion map $I : C(X \setminus U) \to C(X)$ is a chain selector for $\mathcal{M}_e$. Let $\pi : C(X) \to C(X, E)$ be the projection map. Then $\pi \circ I : C(X \setminus U) \to C(X, E)$ is surjective. To see this observe that a basis for $C(X, E)$ consists of all

$$\widehat{Q} \in \widehat{\mathcal{K}}(X) \setminus \widehat{\mathcal{K}}(E) \subset \widehat{\mathcal{K}}(X \setminus U).$$

Furthermore, the kernel of $\pi \circ I$ is exactly $\widehat{\mathcal{K}}(E \setminus U)$. Therefore, $\pi \circ I$ induces an isomorphism

$$e : C(X \setminus U)/C(E \setminus U) \to C(X)/C(E)$$

and hence

$$e_* : H_*(X \setminus U, E \setminus U) \to H_*(X, E)$$

is an isomorphism. ■

We began this chapter with a simple example of a linear map on the plane and asked the question whether it is possible to detect the fixed point using algebraic topological methods. Referring to Figure 6.1, we see that $X = [-4, 4] \times [4, 4]$ is the region we want to study. Unfortunately, $f(X) \not\subset X$. However, we identified $E = [-4, 2] \times [4, 4] \cup [2, 4] \times [4, 4]$ as the smallest cubical set with the property that if $x \in X$ and $f(x) \notin X$, then $x \in E$. Thus, E is a cubical representation of the exit set for X, i.e. those points which leave X under one iteration.

As was noted before, $f(X) = [-8, 8] \times [-2, 2]$. Clearly, $f(E) = [-8, 4] \times [-2, 2] \cup [4, 8] \times [-2, 2]$. Combining these two observations, we can write

$$f(X) \subset X \cup f(E).$$

So let $Y = X \cup f(E)$ and let $B = E \cup f(E)$. Then $f : (X, E) \to (Y, B)$ is a continuous map between cubical pairs. Now let $U = Y \setminus X$. This is a representable set and $B \setminus U = E$ which is a cubical set. Therefore by the Excision Isomorphism Theorem

$$e_* : H_*(Y \setminus U, B \setminus U) \to H_*(Y, B)$$

is an isomorphism. But $(Y \setminus U, B \setminus U) = (X, E)$, therefore, $e_*^{-1} : H_*(Y, B) \to H_*(X, E)$ is an isomorphism. Define

$$f_{(X,E)*} : H_*(X, E) \to H_*(X, E)$$

by $f_{(X,E)*} := e_*^{-1} \circ f_*$.

We now have a map, at least on the level of homology, that goes from a space to itself and we can hope to develop a Lefschetz fixed point theorem for this map that would tell us about the existence of fixed points for f restricted to $X \setminus E$.

Exercises

6.1 Let $Q \in \mathcal{K}_q$ be an elementary cube. Let $E = \{P \in \mathcal{K}_{q-1} \mid P$ is a proper face of $Q\}$. Prove that

$$H_k(Q,E) \cong \begin{cases} \mathbf{Z} & \text{if } k = q \\ 0 & \text{otherwise.} \end{cases}$$

6.2 Let $f : (X,E) \to (Y,B)$ be a continuous map between pairs. Choose a scaling vector α such that $\mathcal{M}_{f^\alpha} : X \rightrightarrows Y$ is a cubical approximation. Prove that $\mathcal{M}_{f^\alpha}(E) \subset B$.

6.3 Let $X = [-4,4] \times [4,4]$, $E = [-4,2] \times [4,4] \cup [2,4] \times [4,4]$ and

$$f = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} : \mathbf{R}^2 \to \mathbf{R}^2.$$

Compute $H_*(X,E)$ and $f_{(X,E)*}$.

6.2 Exact Sequences

We finished the last section with a suggestion that we were close to being able to develop a Lefschetz fixed point theorem for pairs of spaces. However, if the reader solved Exercise 6.3, then it is clear that our ability to compute relative homology groups, is rather limited. Thus, before continuing our quest for a fixed point theorem we will look for more efficient methods of computing relative homology groups. Given a pair of cubical sets (X,E), ideally, we would have a theorem that by which we could determine $H_*(X,E)$ in terms of $H_*(X)$ and $H_*(E)$. As we shall see in Section 6.3 such a theorem exists, but before we can state it we need to develop some more tools in homological algebra.

From the algebraic point of view, homology begins with a chain complex $\{C_k, \partial_k\}$ which can be thought of as an sequence of abelian groups and maps

$$\ldots \to C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \to \ldots$$

with the property that

$$\operatorname{image} \partial_{k+1} \subset \ker \partial_k.$$

A very special case of this is the following.

Definition 6.6 A sequence (finite or infinite) of groups and homomorphisms

$$\ldots \to G_3 \xrightarrow{g_3} G_2 \xrightarrow{g_2} G_1 \to \ldots$$

is *exact* at G_2 if

$$\text{image}\, g_3 = \ker g_2.$$

It is an *exact sequence* if it is exact at every group. If the sequence has a first or last element, then it is automatically exact at that group.

To develop our intuition concerning exact sequences we will prove a few simple lemmas.

Lemma 6.7 *$G1 \xrightarrow{g_1} G_0 \xrightarrow{\phi} 0$ is an exact sequence if and only if g_1 is an epimorphism.*

Proof: ($\Rightarrow$) Assume that $G1 \xrightarrow{g_1} G_0 \xrightarrow{\phi} 0$ is an exact sequence. Since $\phi : G_0 \to 0$, $\ker \phi = G_0$. By exactness, $\text{image}\, g_1 = \ker \phi = G_0$, i.e. g_1 is an epimorphism.

($\Leftarrow$) If g_1 is an epimorphism, then ■

Lemma 6.8 *$0 \longrightarrow G1 \xrightarrow{g_1} G_0$ is an exact sequence if and only if g_1 is an monomorphism.*

Proof: ■

Lemma 6.9 *Assume that*

$$G3 \xrightarrow{g_3} G2 \xrightarrow{g_2} G1 \xrightarrow{g_1} G_0$$

is an exact sequence. Then the following are equivalent:

1. *g_3 is an epimorphism,*

2. *g_1 is a monomorphism,*

3. *g_2 is the zero homomorphism.*

Proof: ■

Definition 6.10 A *short exact sequence* is an exact sequence of the form

$$0 \to G_3 \xrightarrow{g_3} G_2 \xrightarrow{g_2} G_1 \to 0.$$

Example 6.11 Stated as a definition, it may appear the a short exact sequence is a rather obscure notion. However, it appears naturally in many examples. Consider a cubical pair (X, E) and for each k the following sequence

$$0 \to C_k(E) \xrightarrow{I_k} C_k(X) \xrightarrow{\pi_k} C_k(X, E) \to 0 \tag{6.2}$$

where I_k is the inclusion map and π_k is the projection map. That this is a short exact sequence follows from simple applications of the previous lemmas. To begin with, I_k is a monomorphism since $\widehat{\mathcal{K}}(E) \subset \widehat{\mathcal{K}}(X)$. Therefore, by Lemma 6.8

$$0 \to C_k(E) \xrightarrow{I_k} C_k(X)$$

is exact. Similarly, by definition of relative chains π_k is an epimorphism. Hence, Lemma 6.7 implies that

$$C_k(X) \xrightarrow{\pi_k} C_k(X, E) \to 0$$

is exact. So all that remains is to show that the sequence is exact at $C_k(X)$.

By definition the kernel of π_k is $C_k(E)$. Similarly, since I_k is a monomorphism, $\operatorname{image} I_k = C_k(E)$, i.e. $\operatorname{image} I_k = \ker \pi_k$.

The short exact sequence (6.2) is called the *short exact sequence of a pair.*

Lemma 6.12 *Let*

$$0 \to G_3 \xrightarrow{g_3} G_2 \xrightarrow{g_2} G_1 \to 0$$

be a short exact sequence. Then, g_2 induces an isomorphism from $G_2/g_3(G_3)$ to G_1. Conversely, if $K := \ker g_2$, then the sequence

$$0 \to G_3 \xrightarrow{\iota} G_2 \xrightarrow{g_2} G_1 \to 0$$

is short exact where ι is the inclusion map.

Proof: ■

We now turn to the question of maps between exact sequences. Again, in search of the natural definitions we recall the case of maps between chain complexes. Let $\{C_k, \partial_k\}$ and $\{C'_k, \partial'_k\}$ be chain complexes. Recall that the maps of interest between chain complexes are chain maps $F : C \to C'$. To begin to view this in the context of exact sequences, observe that the two chain complexes and chain map form the following commutative diagram.

$$\begin{array}{ccccccccc} \dots & \to & C_{k+1} & \xrightarrow{\partial_{k+1}} & C_k & \xrightarrow{\partial_k} & C_{k-1} & \to & \dots \\ & & \downarrow{F_{k+1}} & & \downarrow{F_k} & & \downarrow{F_{k-1}} & & \\ \dots & \to & C'_{k+1} & \xrightarrow{\partial'_{k+1}} & C'_k & \xrightarrow{\partial'_k} & C'_{k-1} & \to & \dots \end{array} \tag{6.3}$$

This leads us to the following definition for the more restrictive case of exact sequences.

Definition 6.13 Let

$$\dots \to G_{k+1} \xrightarrow{g_{k+1}} G_k \xrightarrow{g_k} G_{k-1} \to \dots$$

and

$$\dots \to G'_{k+1} \xrightarrow{g'_{k+1}} G'_k \xrightarrow{g'_k} G'_{k-1} \to \dots$$

be exact sequences. A *homomorphism* F from the first sequence to the second is a collection of group homomorphisms $F_k : G_k \to G'_k$ such that the following diagram commutes

$$\begin{array}{ccccccccc} \dots & \to & G_{k+1} & \xrightarrow{g_{k+1}} & G_k & \xrightarrow{g_k} & G_{k-1} & \to & \dots \\ & & \downarrow{F_{k+1}} & & \downarrow{F_k} & & \downarrow{F_{k-1}} & & \\ \dots & \to & G'_{k+1} & \xrightarrow{g'_{k+1}} & G'_k & \xrightarrow{g'_k} & G'_{k-1} & \to & \dots \end{array} \tag{6.4}$$

F is an isomorphism, if F_k is an isomorphism for each k.

6.3 The Connecting Homomorphism

In the previous section we defined the notion of an exact sequence and proved some simple lemmas. In this section we shall prove a theorem that is fundamental to all of homology theory. As a corollary we will answer the motivating question of how relative homology groups are related to the homology groups of the each of spaces in the pair.

Definition 6.14 Let $\mathcal{A} = \{A_k, \partial_k^A\}$, $\mathcal{B} = \{B_k, \partial_k^B\}$, and $\mathcal{C} = \{C_k, \partial_k^C\}$ be chain complexes. Let 0 denote the trivial chain complex, i.e. the chain complex in which each group is the trivial group. Let $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{C}$ be chain maps. The sequence

$$0 \to \mathcal{A} \xrightarrow{F} \mathcal{B} \xrightarrow{G} \mathcal{C} \to 0$$

is a *short exact sequence of chain complexes* if for every k

$$0 \to A_k \xrightarrow{F_k} B_k \xrightarrow{G_k} C_k \to 0$$

is a short exact sequence.

Theorem 6.15 *Let*

$$0 \to \mathcal{A} \xrightarrow{F} \mathcal{B} \xrightarrow{G} \mathcal{C} \to 0$$

be a short exact sequence of chain complexes. Then, for each k there exist a map

$$\partial_* : H_k(\mathcal{C}) \to H_{k-1}(\mathcal{A})$$

such that

$$\ldots \to H_k(\mathcal{A}) \xrightarrow{F_*} H_k(\mathcal{B}) \xrightarrow{G_*} H_k(\mathcal{C}) \xrightarrow{\partial_*} H_{k-1}(\mathcal{A}) \to \ldots$$

is a long exact sequence.

Proof: ■

Corollary 6.16 (The exact homology sequence of a pair) *Let (X, E) be a cubical pair. Then there exists a long exact sequence*

$$\ldots \to H_k(E) \xrightarrow{I_*} H_k(X) \xrightarrow{\pi_*} H_k(X, E) \xrightarrow{\partial_*} H_{k-1}(E) \to \ldots$$

where $I : E \to X$ and $\pi : (X, \emptyset) \to (X, E)$ are inclusion maps.

Proof: ■

6.4 Relative Lefschetz Theorem

6.5 Mayer-Vietoris Sequence

Appendix A

Equivalence Relations

Let X and Y be sets. The *cartesian product* of X and Y consists of all ordered pairs (x, y) with $x \in X$ and $y \in Y$. It is denoted by

$$X \times Y := \{(x, y) \mid x \in X \text{ and } y \in Y\}.$$

Let X be any set. A *relation* on X is a subset $R \subset X \times X$.

Example A.1 1. Consider the set of integers $\mathbf{Z}$ and let $R = \{(n, m) \mid m = 2n\}$.

2. Consider the set of positive integers $\mathbf{Z}^+$ and let

$$R = \{(n, m) \mid n \text{ and } m \text{ share a prime factor}\}.$$

3. Consider a the set of integers $\mathbf{Z}$ and let $R = \{(n, m) \mid m - n \text{ is a multiple of } 2\}$.

Definition A.2 R is an *equivalence relation* on X if

1. R is *reflexive*, i.e. $(x, x) \in R$ for all $x \in X$.

2. R is *symmetric*, i.e. $(x, y) \in R$ implies that $(y, x) \in R$.

3. R is *transitive*, i.e. $(x, y) \in R$ and $(y, z) \in R$ implies that $(x, z) \in R$

When R is an equivalence relation, the standard convention is to write $x \sim y$ if and only if $(x, y) \in R$.

Example A.3 The relation R defined by Example A.1.3 is an equivalence relation. To see this we must check the three conditions. For every integer $n \in Z$, $n \sim n$ since $n - n = 0$ which is a multiple of 2. Observe that if $n \sim m$ then $m - n$ is divisible by 2. But this means that $n - m$ is divisible by two and so $m \sim n$. Finally, if $n \sim m$ and $m \sim k$ then there exist integers i and j such that $m - n = 2i$ and $k - m = 2j$. But this implies that

$$k - n = k - m + m - n = 2i + 2j = 2(i + j),$$

and hence $n \sim k$.

Given an equivalence relation $\sim$ on a set X there is a natural way to partition X into disjoint subsets. Namely, for every $x \in X$ define the *equivalence class* of x to be the subset

$$[x] := \{y \in X \mid x \sim y\}.$$

Because, an eqivalence relation is reflexive it is clear that $x \in [x]$. It is easy to check that the equivalence classes are disjoint. To be more precise. Let $[x]$ and $[y]$ be equivalence classes. Assume that there exists $z \in Z$ such that $z \in [x]$ and $z \in [y]$. Then $[x] = [y]$. By definition $z \in [x]$ means that $z \sim x$. Similarly, $z \in [y]$ means that $z \sim y$. By transitivity and symmetry, $x \sim y$ and hence $[x] = [y]$. Another way of saying this is that if $[x] \neq [y]$ then $x \not\sim y$.

A final important comment concerning equivalence relations has to do with the functions they induce. Let X be a set with an equivalence relation $\sim$. Let E denote the set of equivalence classes. Let $\rho : X \to E$ be given by $\rho(x) = [x]$. Since equivalence classes are disjoint, ρ is a function. Furthermore, ρ is surjective, since any element of E is an equivalence class which can be represented by $[x]$ and therefore, $\rho(x) = [x]$. Another standard notation for the set E is $X/\sim$.

Bibliography

[1] M. Herlihy and N. Shavit, The topological structure of asynchronous computability, preprint.

[2] J. R. Munkres, *Topology: a first course*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey 1975.

[3] J. R. Munkres, *Elements of Algebraic Topology* Addison-Wesley, 1984.