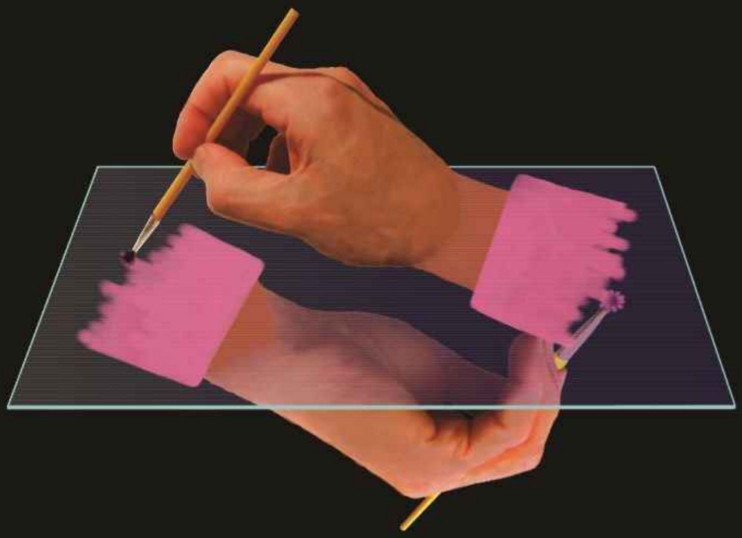
TOPOLOGY

illustrated



Peter Saveliev

Peter Saveliev

Topology Illustrated

With 1000 Illustrations

Peter Saveliev
Department of Mathematics
Marshall University
Huntington, WV 25755
USA

Topology Illustrated by Peter Saveliev
657 pages, includes index
ISBN 978-1-4951-8875-6

Mathematics Subject Classification (2010): 55-01, 57-01, 54-01, 58-01, 39A12

Dedicated to the memory of my parents

Preface

A first course in topology is usually a semester in point-set topology. Sometimes a chapter on the fundamental group is included at the end, with very little time left. For the student, algebraic topology often never comes.

The main part of the present text grew from the course *Topology I and II* that I have taught at Marshall University in recent years. This material follows a two-semester first course in topology with emphasis on algebraic topology. Some common topics are missing, though: the fundamental group, classification of surfaces, and knots. Point-set topology is presented only to a degree that seems necessary in order to develop algebraic topology; the rest is likely to appear in a, typically required, real analysis course. The focus is on *homology.*

Such tools of algebraic topology as chains and cochains form a foundation of *discrete calculus.* A through introduction is provided.

The presentation is often more detailed than one normally sees in a textbook on this subject, which makes the text useful for self-study or as a companion source.

There are over 1000 exercises. They appear just as new material is being developed. Some of them are quite straight-forward; their purpose is to slow down your reading.

There are over 1000 pictures. They are used – but only as *metaphors* – to illustrate topological ideas and constructions. When a picture is used to illustrate a proof, the proof still remains complete without it.

Applications are present throughout the book. However, they are neither especially realistic nor (with the exception of a few spreadsheets to be found on the author's website) computational in nature; they are mere *illustrations* of the topological ideas. Some of the topics are: the shape of the universe, configuration spaces, digital image analysis, data analysis, social choice, exchange economy, and, of course, anything related to calculus. As the core content is independent of the applications, the instructor can pick and choose what to include.

The way the ideas are developed may be called "historical", but not in the sense of what actually happened – it's been too messy – but rather what *ought to* have happened.

All of this makes the book a lot longer than a typical book with a comparable coverage. Don't be discouraged!

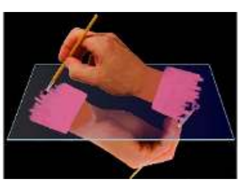
A rigorous course in linear algebra is an absolute necessity. In the early chapters, one may be able to avoid the need for a modern algebra course but not the maturity it requires.

- **Chapter I: Cycles** contains an informal introduction to homology as well as a sample: homology of graphs.
- **Chapter II: Topologies** is the starting point of point-set topology, developed as much as is needed for the next chapter.
- **Chapter III: Complexes** introduces first cubical complexes, cubical chains, unoriented and then oriented, and cubical homology. Then a thorough introduction to simplicial homology is provided.
- **Chapter IV: Spaces** continues to develop the necessary concepts of point-set topology and uses them to advance the ideas of algebraic topology, such as homotopy and cell complexes.
- **Chapter V: Maps** presents homology theory of polyhedra and their maps.
- **Chapter VI: Forms** introduces discrete differential forms as cochains, calculus of forms, cohomology, and metric tensors on cell complexes.
- **Chapter VII: Flows** presents applications of calculus of forms to ODEs, PDEs, and social choice.

By the end of the first semester, one is expected to reach the section on simplicial homology in Chapter III, but maybe not the section on the homology maps yet. For a single-semester first course, one might try this sequence: Chapter II, Sections III.4 - III.6, Chapter IV (except IV.3), Section V.1. For a one-semester course that follows point-set topology (and modern algebra), one can take an accelerated route: Chapters III - V (skipping the applications). For discrete calculus, follow: Sections III.1 - III.3, Chapters VI and VII.

The book is mostly undergraduate; it takes the student to the point where the tough proofs are about to start to become unavoidable. Where the book leaves off, one usually proceeds to such topics as: the axioms of homology, singular homology, products, homotopy groups, or homological algebra. *Geometry and Topology* by Bredon is a good choice.

Peter Saveliev

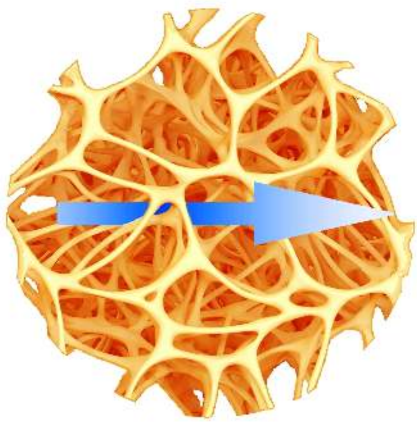


About the cover. Once upon a time I took a better look at the poster of *Drawing Hands* by Escher hanging in my office and realized that what it depicts looks symmetric but isn't! I decided to fix the problem and the result is called *Painting Hands*. This juxtaposition illustrates how the antipodal map (aka the central symmetry) reverses the orientation in the odd dimensions and preserves it in the even dimensions. That's why to be symmetric the original would have to have two right hands!

Contents

Chapter I

Cycles



1 Topology around us

1.1 Topology – Algebra – Geometry

In an attempt to capture the essence of topology in a single sentence, one usually says that *topology is the science of spatial properties that are preserved under continuous transformations*: you can bend, stretch, and shrink but not tear or glue. Thus, topology studies spatial properties of a special kind...

In order to illustrate this idea in the context of mathematics as a whole, let's take a look at these delicious donuts:

TOPOLOGY:  Count: *One, two, three, four... five!*

ALGEBRA:  Compute: *Two times three... six!*

GEOMETRY:  Measure: *Seven... eighths!*

In order to see how topology is necessary for counting, consider the fact that the first step is to recognize that these are separate objects – disconnected from each other! In fact, counting, computing, and measuring are all preceded by our ability to perceive the topology in these pictures. Furthermore, if we want to count the *holes*, we will need to recognize them as a different kind of topological feature.

Now, why do we need to know topology? The answer is, because we *already* know a lot and this knowledge deserves a precise, mathematical treatment. We will start with a few elementary examples. They come from four seemingly unrelated areas: vision and computer vision, cosmology, data analysis, and social choice theory.

1.2 The integrity of everyday objects

As we would like to be able to delegate some of the decision making to computers, we start with what we intuitively understand and try to describe it in absolutely unambiguous terms.

In an industrial context, one might need to consider the integrity of objects being manufactured or inspected.

The first question may be: this bolt is supposed to hold two things together; is it still capable of that, or *is there a crack in it*?

In other words: would the bolt hold a *hair* or would the hair slip through?

The second question may be: this porous material is supposed to stop a flow of liquid; is it still water-tight or *is there leakage*?

In other words: would the sheet hold *water* or might some permeate through?

The third question may be: to be strong, this alloy is supposed to be fully solid; is it still solid, or *are there air bubbles*?


In other words: would the alloy hold no *air*, or did some get in?

It is important to understand now that those are three *different* kinds of integrity loss – as there may be a crack but no hole or vice versa, etc.:



We can describe the three situations informally as follows. The objects have:
- *cuts*, or
- *tunnels*, or
- *voids*.

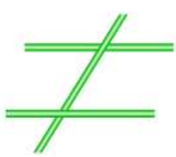
Exercise 1.1. Sketch examples for all possible combinations of cuts, tunnels, and voids with one or none of each and indicate corresponding real-life objects.

Of course, the presence of one of these features doesn't have to mean that the object is defective, as in the above examples: a rope, a bucket, and paint. The examples of objects that are *intended* to have such a topological feature are respectively: sliced bread, a strainer, and a balloon.



Next, we will classify these three types of "defects", according to their *dimensions*.

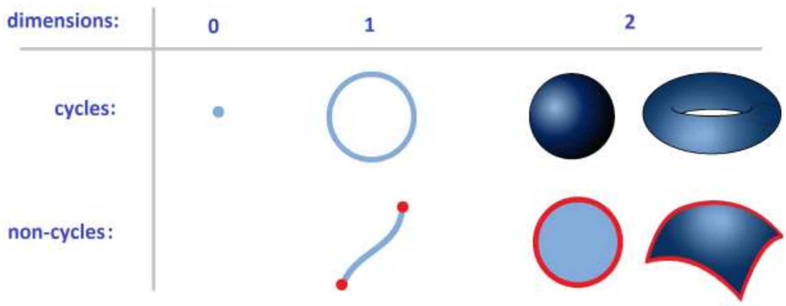
To understand why and how, let's recall from the elementary geometry (or linear algebra) this list 0-dimensional, 1-dimensional and 2-dimensional spaces:
- 0. points,
- 1. lines,
- 2. planes.


If these spaces are allowed to be deformed, the list becomes:
- 0. points,
- 1. curves,
- 2. surfaces.

In every dimension, some of those spaces are special – in the sense that they have no endpoints or edges, known as *boundaries*, such as these:

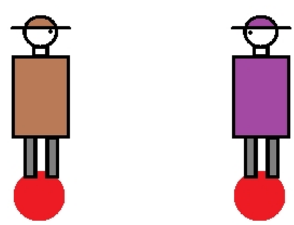
- 0. points,
- 1. circles,
- 2. spheres.



These, as well as their deformed versions, may collectively be called *cycles*. What makes them special is certain *topological features*. Meanwhile, objects or figures that don't have such features are called *acyclic*.

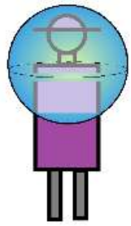
Finally, our conclusions.

• 0. Any two points form a 0-cycle and, since this is the simplest example of a *cut*, the latter is a 0-dimensional feature:



• 1. Any circle is a 1-cycle and, since this is the simplest example of a *tunnel*, the latter is a 1-dimensional feature:

• 2. Any sphere is a 2-cycle and, since this is the simplest example of a *void*, the latter is a 2-dimensional feature:

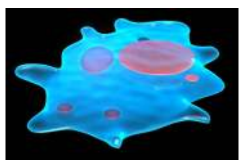


Exercise 1.2. Using this terminology, describe the topology of the following objects: a basketball, a football, a cannonball, a doughnut, an inner tire, a steering wheel, a bicycle wheel, a potter's wheel, a fingerprint, a tree, an envelope.

Exercise 1.3. Suggest your own examples of topological issues in everyday life and describe them using this terminology.

What if we *deform* these objects as if they are made of rubber? We can stretch and shrink them and, as long as we do not cut or glue, the number of topological features will remain the same. Indeed, the cuts, holes, and voids may narrow but won't disappear. Indentations may appear but they won't turn into topological features.

This property is exemplified by an amoeba – a single-cell organism able to freely change its form.



Exercise 1.4. Sketch a sequence of steps to show how this man – let us first appoint him amoeba-like abilities – can unlock his hands while his fingers remain together and continue to form the two loops.

What about more general *continuous transformations*?

Breaking a bolt is not continuous but welding it back together is. Digging a tunnel (all the way) through a wall is not continuous but filling it shut is. Piercing a bubble is not continuous but patching it is. Bread is cut, tires are punctured, paper is folded into an origami, fabric is sewed into a suit or an airbag, etc., etc.

As these examples show, *continuous transformations* can add and remove topological features.

Exercise 1.5. Describe what happens to the three topological features in the above examples.

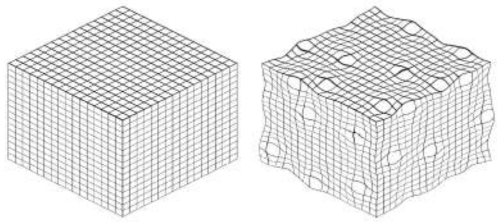
Exercise 1.6. Describe, in precise terms, what happens to the amoeba's topology as it feeds? Indicate which stages of development are continuous and which aren't.

1.3 The shape of the Universe

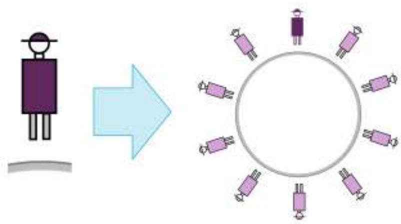
What is the shape of the Universe? What is its topology? Does the Universe have "cuts", "tunnels", or "voids"?

Looking around, we don't see any of these! But remember how, hundreds of years ago, sailors started to notice the curvature of the horizon? They later *proved* – by around-the-world travel and even later by orbiting – that the surface of the Earth curves all the way to meet itself on the other side: it encloses everything inside.

As for the Universe, what we *know* is that we are not living in the flat world of Euclid and Newton; we are living in the curved world of Riemann and Einstein:

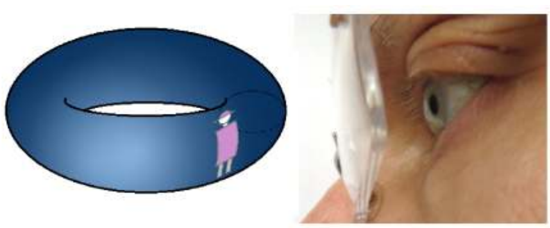


But the Universe that curves and bends *here* might curve and bend *everywhere!* Then, no matter how small the bend is, it might make the Universe *close on itself*:



It may be possible then to travel in a straight line and arrive at the starting point from the opposite direction. In fact, the light of the Sun may come to us from an unexpected direction and, as a result, appear to come from a distant star:

Events of this kind would provide us with clues about the topology of the Universe. We have to rely on such indirect observations because we can't step outside and observe the Universe from there:



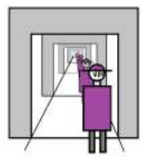
As we stay *inside* our Universe, we can't see its "cuts", "tunnels", or "voids". The only way we can discover its structure is by traveling or by observing others – such as light or radio waves – travel.

If multiple universes exist but there is no way of traveling there, we can't confirm their existence. So much of 0-cycles...

Traveling in various directions and coming (or not coming) back will produce information about loops in space. These loops, or 1-cycles, are used to detect tunnels in the Universe.

There might also be voids and, since the Universe is 3-dimensional, it might also have 3-dimensional topological features. Studying these features would require us to add to the existing list, currently containing the point, the line, and the plane, a new item: space, or, more accurately: a 3-dimensional space. How such a space forms a 3-*cycle* may be hard or impossible to visualize. Nonetheless, these cycles **are** detectable – once we arm ourselves with the methods presented in this book.

Example 1.7. What if the Universe is like a room with two doors and, as you exit through one door, you enter it through the other? You'll realize that, in fact, it's the same door! If you look through this doorway, this is what you see:



□

Exercise 1.8. What kind of cycle is this?

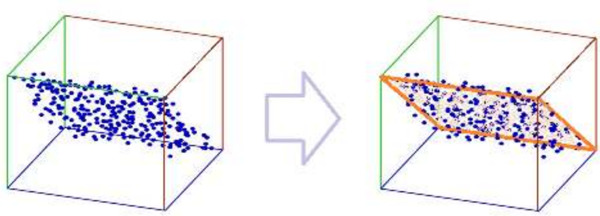
Exercise 1.9. What if, as you exit one door, you enter the other – but upside down? Sketch what you would see.

Elsewhere, we face even higher dimensions...

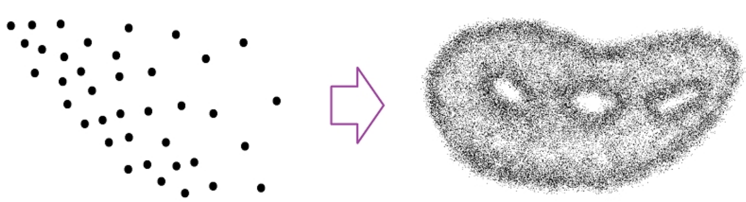
1.4 Patterns in data

Data resides outside of our tangible, physical, 3-dimensional world.

Imagine we conduct an *experiment* that consists of a set of 100 different measurements. We would, of course, like to make sense of the results. First, we put the experiment's results in the form of a string of 100 numbers. Next, thinking mathematically, we see this string as a point in the 100-dimensional space. Now, if we repeat the experiment 1000 times, the result is a "point cloud" of 1000 points in this space:



Now, as scientists, we are after *patterns* in data. So, is there a pattern in this point cloud and maybe a law of nature that has produced it? What shape is hinted by the data?



Unfortunately, our ability to see is limited to three dimensions!

Exercise 1.10. What is the highest dimension of data your spreadsheet software can visualize? Hint: don't forget the colors.

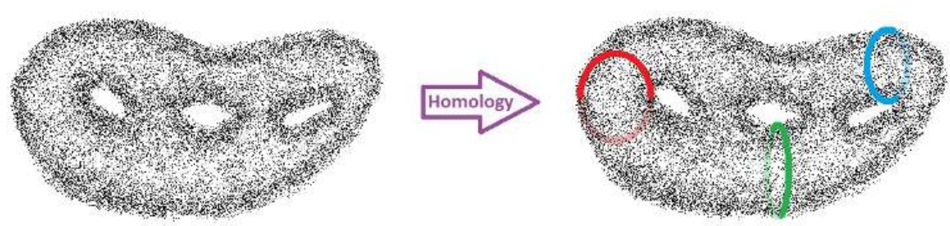
With this limited ability to visualize, we still need to answer the same questions about the shape outlined by the point cloud:

- Is it one piece or more?
- Is there a tunnel or a void?
- And what about possible 100-dimensional topological features?

Once again, we *can't* step outside and take a look at this object.

The first question is the most important as it is the question of *classification*. We assign to classes what we see: drugs, customers, movies, species, web sites, etc. The methods for solving this problem – often called "clustering" – are well-developed in data analysis:

The rest of the questions will require topological thinking presented in this book. We will record *adjacency relation* in the data and, through *purely algebraic* procedures, detect topological features in it:



These features are called *homology classes*.

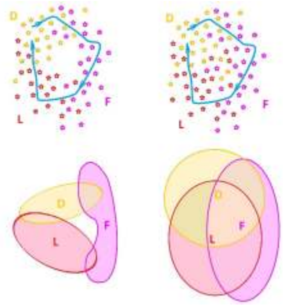
Example 1.11 (garden as data). Here is an example of data analysis one can do without a computer. Imagine you are walking through a field of flowers: daffodils, lilies, and forget-me-nots. You are blindfolded but you do detect the smells of the flowers as you walk, in this order:

- daffodils only, then
- daffodils and lilies, then
- lilies only, then
- lilies and forget-me-nots, then
- forget-me-nots only, then
- daffodils and forget-me-nots,
- daffodils only again.

Now, thinking topologically, you might arrive at an interesting conclusion: there is a spot in the field, for which it is either true that:

- none of these types of flowers grow there, or
- all three types of flowers grow there.

The reason why is illustrated below:



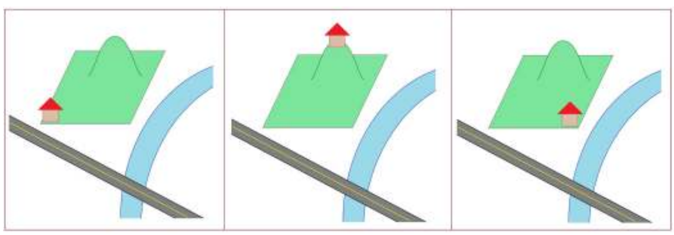
To be sure, the conclusion fails whenever any of the types grows in several separate, disconnected patches. □

Exercise 1.12. What if there is one patch for each flower but it might have a hole in it?

1.5 Social choice

Next, we present examples of how the presence of topological features in the space of choices may cause some undesirable outcomes.

Imagine the owner of a piece of land who needs to decide on the location for his new house based on a certain set of *preferences*. The land varies in terms of its qualities: grass/undergrowth/forest, hills/valleys, wet/dry, distance to the road/river/lake, etc. If the landowner only had a single criterion for his choice, it would be simple; for example, he would choose the highest location, or the location closest to the river, or closest to the road, etc.

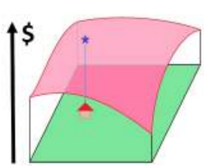


Now, what if, because of such a variety of possible locations and the complexity of the issues, the owner can only decide on the *relative* desirability of locations, and only for locations in *close proximity* to each other? Is it possible to make a decision, a decision that satisfies his preferences?

Let us first suppose that the piece of land consists of two separate parcels. Then no two sites located in different parcels can be considered "nearby"... Therefore, the comparison between them is impossible and the answer to our question is No. This observation alone suggests that the question may be topological in nature.

What if the land parcel is a single piece but there is a lake in the middle of it? We will prove later in this book that the answer is still No.

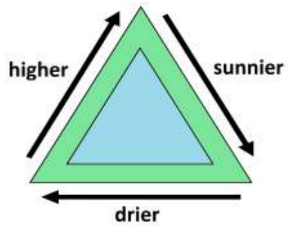
Ideally, the preferences are expressed by a single number assigned to each location: the larger its value, the better the choice. These numbers form what's called a "utility function":



This function may have a monetary meaning in the sense that building the house will increase the value of the real estate by an amount that depends on the chosen location.

We will prove two facts in this book. First, whenever the piece of land is known to be acyclic, there inevitably is a utility function. Second, a piece of land that always has a utility function must be acyclic.

So, what prevents us from constructing such a function for a non-acyclic parcel? Let's consider a piece of land that is a thin strip along a triangular lake. Suppose the locations within any of the three sides of the triangle are comparable but there is no comparison between any two edges:



The result could be a *cyclic preference*!

Such preferences are seen elsewhere. A small child may express, if asked in a certain way, a circular preference for a Christmas present:

$$\textit{bike} \quad > \quad \textit{videogame} \quad > \quad \textit{drum set} \quad > \quad \textit{bike} \quad > \quad ...$$

In another example, suppose we have three candidates, A,B, and C, running for office and there are three voters with the following preferences:

	First choice	Second choice	Third choice
Voter1 :	A	B	C
Voter2 :	B	C	A
Voter3 :	C	A	B

Who won? Is it candidate C? No, because one can argue that B should win instead as *two* voters (1 and 2) prefer B to C and only *one* voter (3) prefers C to B. Then the same argument proves that A is preferred to B, hence C is preferred to A. Our conclusion is: sometimes a seemingly reasonable interpretation of collective preferences can be cyclic, even if the preferences of individual voters are not.

And let's not forget about the familiar:

$$\textit{rock} \quad > \quad \textit{paper} \quad > \quad \textit{scissors} \quad > \quad \textit{rock} \quad > \quad ...$$

Exercise 1.13. Think of your own examples of cyclic preferences, in food, entertainment, or weather.

This is not to say that these preferences are unreasonable but only that they cannot be expressed by a utility function. In particular, we can't assign dollar values to the locations when the preferences are cyclic. We are also not saying that there is no way to make a choice when there is a lake in the middle of the parcel, but only that we can't guarantee that it will always be possible.

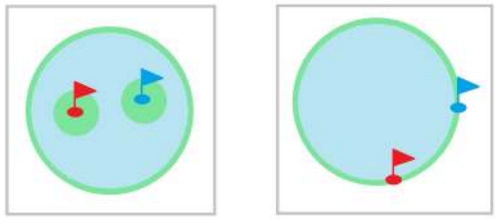
Next, let us assume that it *is* possible.

So, the landowner has chosen a location (based on his preferences or in some other way). What if, however, the landowner's wife has a different opinion? Can they find a *compromise*?

The goal is to have a procedure ready *before* their choices are revealed. The idea is that as the two of them place – simultaneously – two little flags on the map of the land, the decision is made automatically:

• to every pair of locations A and B, a third, compromise, location C is assigned ahead of time.

The obvious method is: chose the *midpoint*. Unfortunately, this method fails as before if the land consists of two separate parcels.



Exercise 1.14. Sometimes this midpoint method fails even if the parcel isn't disconnected, such a U-shape. Find an alternative solution.

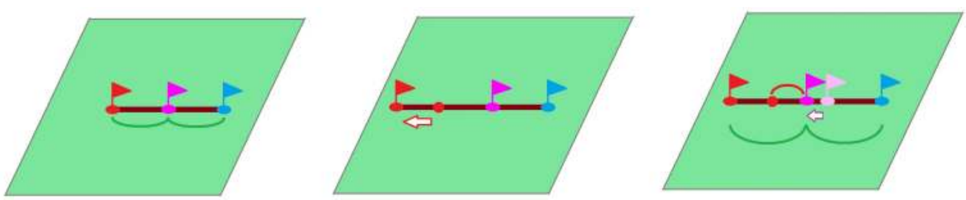
We already know what the next question ought to be: is it always possible to find a fair compromise when there is a lake in the middle? The method may be: choose the midpoint as measured along the shore. But what if they choose two diametrically opposite points? Then there are *two* midpoints!

An amended method may be: choose the midpoint and, in case of diametrically opposite choices, go clockwise from A to B. This method, however, treats the two participants unequally...

Exercise 1.15. Suggest your own alternative solutions.

Once again, we cannot guarantee that such a fair compromise-producing procedure is possible when the land parcel might not be acyclic. This subtle result will be proven in this book. The result implies that there may be a problem when a jointly owned satellite is to be placed on a (stationary) orbit. We will also see how these topological issues create obstacles to designing a reasonable electoral system.

This isn't the end of the story though... Imagine, the husband and wife have placed their little flags, A and B, on the map of their land and then applied the procedure designed in advance, to place the third flag, C, at the compromise location. Imagine now that as the wife leaves the room, the husband moves his flag in the direction away from the wife's. Then he also moves flag C as well to preserve the appearance of a fair compromise. As a result of this manipulation, flag C is now closer to the husband's ideal location!



However, as the husband leaves the room, the wife enters and makes a similar move! And the game is on...

Such a backward movement resembles a tug-of-war.

Now, as the game goes on, one of them reaches the edge of the parcel. Then he (or she) realizes that there is no further advantage to be gained. While the other person continues to improve her (or his) position, eventually she (or he) too discovers that she (or he) is not gaining anything by moving further back. It's a stalemate!

Exercise 1.16. Try this game yourself – on the square and other shapes – using the midpoint method for finding the compromise location. What if the "compromise" is chosen to be twice as close to your ideal location than to the other?

This stalemate is thought of as an *equilibrium* of the game: neither can improve one's position by a unilateral action.

Once again, we can guarantee that such an equilibrium exists but only if the land parcel is known to be acyclic.

Exercise 1.17. Try to play this game on the shore of a lake using the amended midpoint method described above. Show that there is no equilibrium.

---

2 Homology classes

2.1 Topological features of objects

At the very beginning of our study of topology we are interested in the following *visible* features of everyday objects:

- cuts/gaps,
- holes/tunnels, and
- voids/bubbles.

Let's try to describe these features mathematically.

In order to make this study independent of a lengthy development of point-set topology, we will initially limit ourselves to the study of

$$\textit{subsets of a Euclidean space},\ X \subset \mathbf{R}^N.$$

The continuity of functions

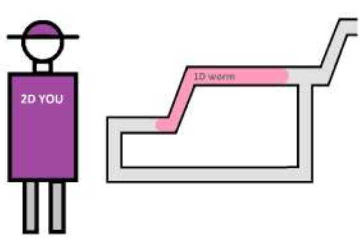
$$f : X \to Y \subset \mathbf{R}^M$$

between such sets is explained in calculus: it is understood via its coordinate functions

$$f_i : \mathbf{R}^N \to \mathbf{R},\ i = 1, 2, ..., M,$$

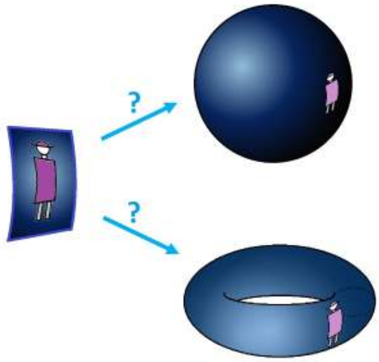
each of which is continuous with respect to each of its variables.

First, we need to recognize that even though we describe these properties as visible features, our interest is the *intrinsic properties.* The intrinsic properties are the ones that you can detect from *within the object.* As an illustration, in the 1-dimensional case one can think of a worm moving through a system of tubes. The topology the worm can feel and detect is limited to the forks of the tube. It might also discover that it can come back to the beginning point of its journey:



Meanwhile, we, the humans, can see the whole thing by being *outside*, in the 2-dimensional space.

On the other hand, we, the humans, couldn't tell whether the surface of the Earth is a sphere or a torus if it were always covered by fog or darkness:

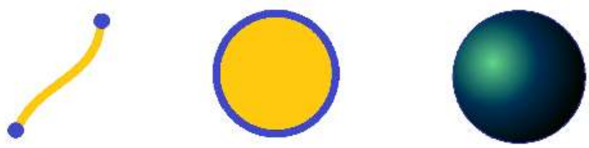


Finally, intrinsic properties are the only properties that we can hope to capture when we study the topology of our 3-dimensional universe. An example of such a property is a room you are walking out of and simultaneously entering through another door (maybe upside down!).

Let's examine, informally, a mathematical meaning of topological features of subsets of the Euclidean space. We will first consider *featureless objects*.

Recall that we represent topological features of space X by means of *cycles* of dimensions 0, 1, and 2. Of course, we can draw them anywhere in X; such as these:

- 0. two points,
- 1. circle, and other closed curves,
- 2. sphere, and other closed surfaces.

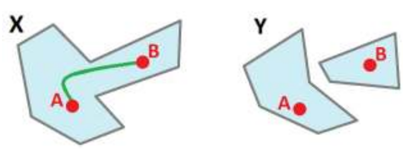


For X to be a featureless object or figure, all of its cycles have to be *trivial* in a sense. The triviality is understood as our ability to "fill" the opening of the cycle, or its ability to contain or "bound" something in X, as follows.

Cuts/gaps:

- the gap between every two points in X can be "filled" with a curve in X.

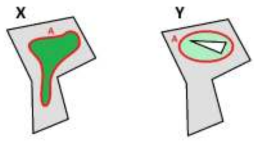
There is a simple and precise way to describe this: $X \subset \mathbf{R}^N$ is called *path-connected* if for any two points A, B in X there is a continuous function $q : [0, 1] \to X$ such that $q(0) = A$, $q(1) = B$; i.e., it is a path connecting these two points:



Exercise 2.1. Prove that any convex set is path-connected.

Exercise 2.2. Prove that the circle is path-connected.

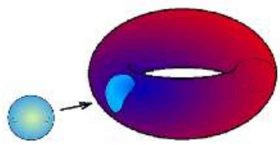
Holes/tunnels:
- the hole in every closed curve in X can be "filled" with a surface in X.



Exercise 2.3. Does the sphere have a non-trivial 1-cycle? Illustrate your answer with a picture. What about the torus?

Voids/bubbles:
- the void in every closed surface in X can be "filled" with a solid in X.

Even though we are after the bubbles here, we don't limit ourselves to just spheres. The reason is that otherwise we would miss the void of the torus. As you can see, any sphere on the torus has to be flattened:



Exercise 2.4. Give an example of a trivial 3-cycle.

"Filling" gaps is usually understood as the ability to connect the two points to each other – by a curve. "Filling" holes and voids can also be understood this way:

featureless	dim	a feature	is trivialized	by a filling	dim
no gaps	0	any two points	are connected by	a curve	1
no tunnels	1	any closed curve	is connected to itself by	a surface	2
no voids	2	any closed surface	is connected to itself by	a solid	3

The relation between the dimensions becomes clear: a k-dimensional *gap* is filled with a $(k+1)$-dimensional object.

2.2 How to define and count 0-dimensional features

Let's review. We want our mathematical interpretations of the topological features to meet the following criteria:
- 1. The description should include no measuring: then the features are preserved under stretching, bending etc.
- 2. The description should be verifiable from the *inside* of the object: then the features are independent of how X might fit into some Y.
- 3. The description should be quantitative: then, without measuring, what's left is *counting*.

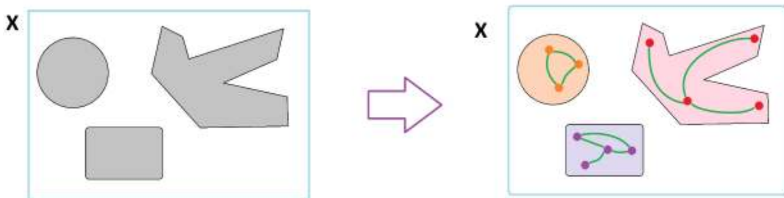
Let's test the above definitions against these criteria. The first two are easily fulfilled. But how quantitative is it? So far, we are only able to say Yes or No: connected or not, trivial or non-trivial, etc. Without counting of the features there is no *evaluation* of the topology of X!

We will have to go deeper.

We start with dimension 0 with the goal of learning how to *count* the topological features – based somehow on the definition of path-connectedness. Furthermore, we would like to count not the number of cuts but the number of pieces of the object instead. The reason is that the former concept is ambiguous: one can cut a cake into 4 pieces with 4 cuts or with 2 cuts...

As there may be infinitely many points in each piece, it wouldn't make sense to count those. We need a way to resolve this ambiguity. We observe that in the definition the two points are connected to each other by a curve. Hence the idea:

- instead of counting the pieces directly, *we count points*, but,
- we count two of them only once if they are connected to each other by a path.

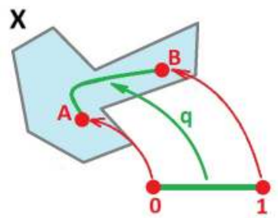


This means that we don't count points, we count *classes of points*.

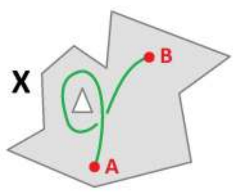
This analysis suggests that we are dealing with an *equivalence relation*. This equivalence relation is called *homology* and the equivalence classes are called *homology classes*.

Definition 2.5. We say that two points A and B in X are *homologous* if there is a continuous function

$$q : [0, 1] \to X \text{ such that } q(0) = A, q(1) = B.$$



These paths are nothing but parametric curves. Naturally, self-intersections are allowed:



Theorem 2.6. Homology is an equivalence relation.

Proof. We recall and prove the three axioms.

Reflexivity: $A \sim A$. A point is connected to itself; just pick the constant path, $q(t) = A, \ \forall t$.

Symmetry: $A \sim B \implies B \sim A$. If path q connects A to B then p connects B to A; just pick $p(t) = q(1 - t), \ \forall t$.

Transitivity: $A \sim B$, $B \sim C \implies A \sim C$. If path q connects A to B and path p connects B to C then there is a path r that connects A to C; just pick:

$$r(t) = \begin{cases} q(2t) & \text{for } t \in [0, 1/2], \\ p(2t - 1) & \text{for } t \in [1/2, 1]. \end{cases}$$

■

Exercise 2.7. Sketch illustrations for the three paths above.

Exercise 2.8. Provide the missing details in the above proof. Hint: use the definition of continuity from calculus.

Exercise 2.9. (a) Provide an alternative definition of homology based on a path given by $q : [a,b] \to X$. (b) Prove that this new definition gives an equivalence relation. (c) Prove that the two definitions are equivalent.

A couple of facts about equivalence relations:

- The equivalence classes are disjoint.
- Their union is the whole set.

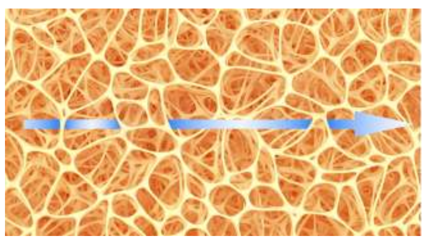
In other words, we have a *partition* of X. The homology class of point A in X is denoted by $[A]$. We have partitioned X into 0-dimensional homology classes, called *path-components*.

Exercise 2.10. Suppose $f : X \to Y$ is continuous. Prove that $A \sim B \Longrightarrow f(A) \sim f(B)$. Hint: the composition of two continuous functions is continuous.

2.3 How to define and count 1-dimensional features

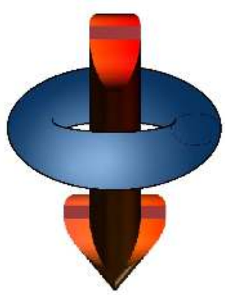
Next, how do we capture *holes and tunnels*?

Giving these simple words an unambiguous meaning may be a challenge, as this picture demonstrates:



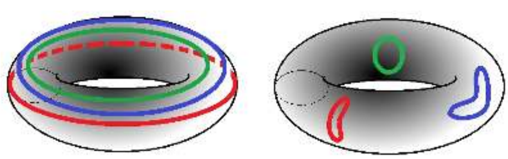
Where is the beginning and where is the end of a tunnel? If two tunnels merge, is this one tunnel or two? What if the tunnel splits? And then merges with another one? And so on.

To start with the simplest case, what is the hole of a doughnut?



Just pointing out that we are able to put a finger through the hole doesn't help. We want to learn to detect the hole *from inside* the object.

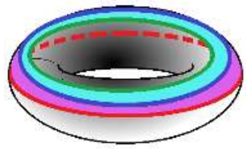
Curves in X could be useful. Indeed, the loops, i.e., 1-cycles, that go around the hole do capture that hole:



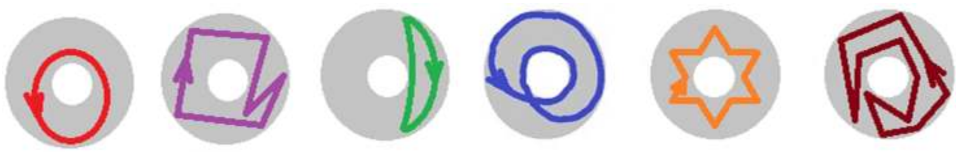
Meanwhile, loops that don't go around the hole capture nothing; they are "trivial".

This is a good start but, just as with the points, there are infinitely many of these loops and it would be pointless to try to count them. Fortunately, from the discussion of dimension 0, we know that a good way to handle this issue might be to establish an equivalence relation over the loops in X.

Recall that if 1-cycle L does not catch any tunnels; i.e., it's trivial, when it bounds a surface S in X. Now, what if 1-cycle L consists of *two loops*, A and B? And what if these two loops – together – bound a surface S? If such an S exists, we think of A and B as equivalent! This way we avoid double counting:

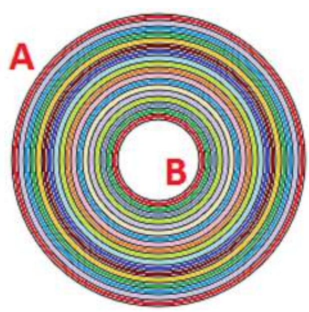


There is now exactly one, it appears, equivalence class of loops that goes around the hole. Or is there? What about the loops that go around it twice or thrice?

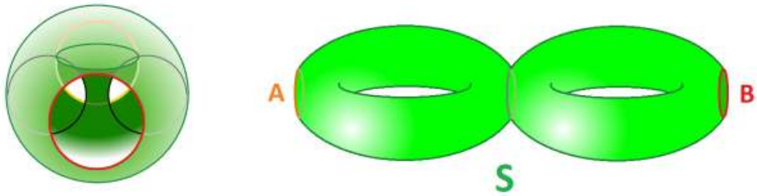


Are they all homologous to each other? How to define homology for these, more general loops is discussed in the next subsection. We will see that the answer is No: they are not all homologous to each other. Only the loops with the same number of turns around the hole (in the same direction!) are homologous. But then we, once again, have infinitely many homology classes for a single hole! We resolve this issue by taking an *algebraic* point of view; by counting the number of turns we have established a correspondence between the set of all these classes and the set of integers $\mathbf{Z}$. But the integers are all multiples of a single number, $1 \in \mathbf{Z}$. In that sense, the homology class that corresponds to 1 "generates" the rest of them. That's why it is the only one that counts.

The simplest example of a surface connecting two loops is a cylinder or a ring. One can then justify calling these loops A and B equivalent by observing that we can *gradually* transform one into the other using the surface that connects them:



These loops are also called *homotopic* and spaces with all loops homotopic to each other are called *simply connected*. Things aren't always this simple. In general, S might look like this:



Exercise 2.11. Give formulas of continuous functions: (a) of the line that creates a loop, (b) of the plane that creates a loop, (c) of the plane that creates a void.

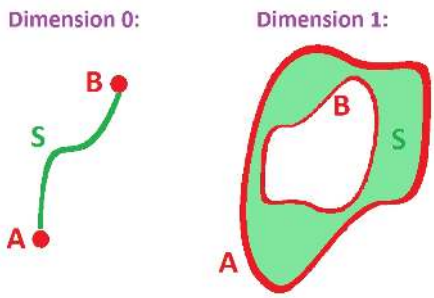
2.4 Homology as an equivalence relation

Thus, homology is an equivalence relation – of points. Points A and B are homologous if there is a path between them. What if A and B were curves?

Then we need to fill in the blank space below:

$$A \sim B \text{ if there is a } \_\_\_\_\_\_\_\_\_\_\_\_ \text{ between them.}$$

The analogy is visible:



Let's compare:

- dimension 0: *points* $A \sim B$ if there is a *curve* between them.
- dimension 1: *curves* $A \sim B$ if there is a *?????* between them.

We see that the dimensions go up by one:

- from 0 to 1: "point(s)" is replaced with "curve(s)", and
- from 1 to 2: "curve(s)" is replaced with its 2-dimensional counterpart: "surface(s)"!

This is what we have: two loops are equivalent, $A \sim B$, if there is a surface S "between them". The key insight here is the relation between S and $A \cup B$: the latter is the *boundary* of the former.

The definition then becomes:

- Two closed curves (loops) A and B are *homologous* if there is a surface S such that A and B form its boundary.

Exercise 2.12. Sketch illustrations for the axioms of equivalence relations for the 2-dimensional homology.

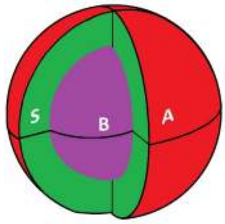
Our new understanding of the meaning of "between them" allows us to go back to the 0-dimensional case and discover that the endpoints A, B of the curve should also be understood as its boundary!

We start on the path toward higher dimensions with this list:

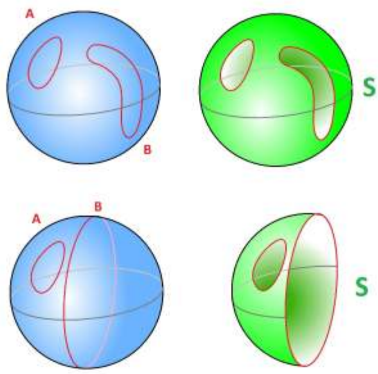
- dimension 0: *points* $A \sim B$ if there is a *curve* S between them.
- dimension 1: *curves* $A \sim B$ if there is a *surface* S between them.

• dimension 2: *surfaces* $A \sim B$ if there is a *solid* S between them.

The relation described in the last item is illustrated below by the inner and outer surfaces of this spherical shell (a "thick sphere"):



Example 2.13. Let's consider these two loops A, B on a sphere:



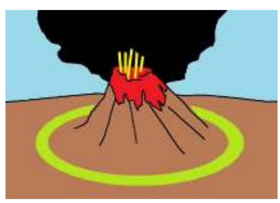
This is what we observe:

• They are homologous.

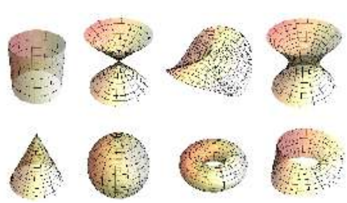
• The surface S is the whole sphere minus the "inside" of A, B.

We conclude that there seems to be no tunnels on this sphere. □

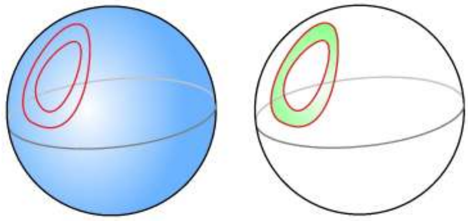
Or volcanoes: is this loop homologous to a point?



Exercise 2.14. Discuss the homology of surfaces by sketching two pairs of examples for each of the surfaces below: (a) two closed curves, preferably not too close to each other, that are homologous; (b) two closed curves that are not homologous, if possible.

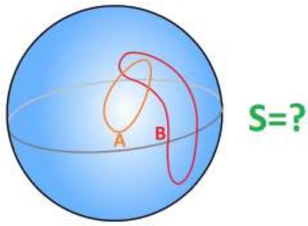


In general, the explanation that "S is the whole sphere minus the inside of A, B" in the last examples incorrect as this illustration shows:

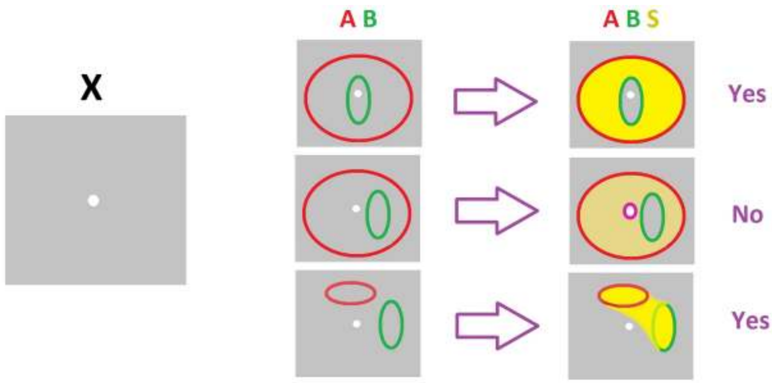


The reason is that it is impossible to be "inside" or "outside" of a curve that lies a *sphere*!

The next illustration suggests that whatever S is, it's not really a surface:



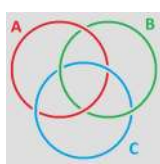
The problem is easier to recognize in these three examples of pairs of homologous and non-homologous loops in the *punctured plane*:



In the last example, the surface that connects the two loops, which is simply a cylinder, is flattened by being "crushed" into the plane.

The way to straighten this out is to realize that just as we allow the curves in the 0-dimensional case to self-intersect, we allow the same for these surfaces; we are dealing with parametric curves and parametric surfaces.

Exercise 2.15. What is the homology class of A in $\mathbf{R}^3 \setminus (B \cup C)$?



This approach taken to its logical end leads to a special kind of homology called "bordism". Instead, in the rest of the book we replace "curve", "surface", and "solid" with "something made of elementary n-dimensional pieces", $n = 1, 2, 3$, respectively. Then, the crucial part of our study will be about

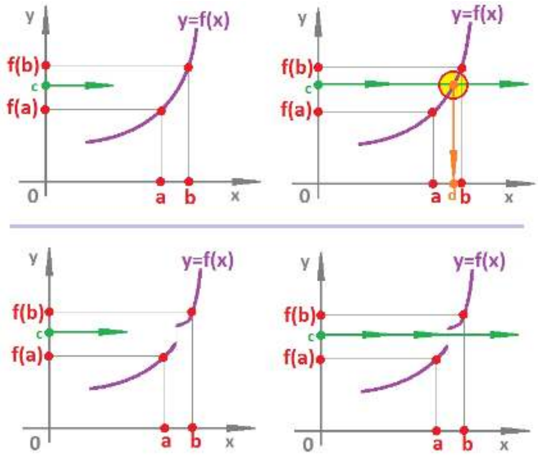
- how the n-dimensional pieces are attached to each other along their $(n-1)$-dimensional

boundaries.
We will start in the next section with *graphs* ($n = 1$).

2.5 Homology in calculus

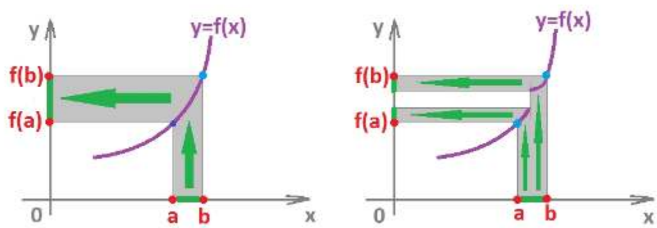
We will consider three well-known theorems from elementary calculus.

Theorem 2.16 (Intermediate Value Theorem). Suppose $f : [a, b] \to \mathbf{R}$ is continuous. Then for any c between $f(a)$ and $f(b)$, there is a $d \in [a, b]$ such that $f(d) = c$.



Exercise 2.17. Is continuity a necessary condition?

What does this have to do with homology? The theorem tells us that the image of a path-connected interval is also a path-connected interval, under a continuous function:



This new understanding of the theorem justifies the word "continuous"...

We can also use this theorem to *evaluate* the topology: a continuous function doesn't increase the number of path-components.

Next, let's test how well you know calculus.

Calculus 1 question: "What is the antiderivative of $1/x$?"

Calculus 1 answer: "It's simply $F(x) = \ln |x| + C$!"

Wrong!

Why wrong? Hint: $1/x$ is undefined at 0.

Exercise 2.18. The question is wrong, too! Explain why.

First, what is our formula intended to mean? The $+C$ in the formula indicates that

- we have found infinitely many – one for each real number C – antiderivatives, and

- we have found *all* of them.

And we would have – if it were any other, continuous, function f – the following result:

$$\text{the set of all antiderivatives of } f \text{ is } \{F_C(x) = F_0(x) + C : C \in \mathbf{R}\},$$

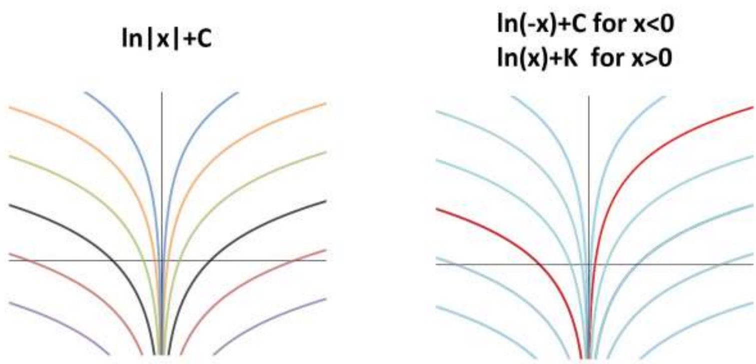
where F_0 is any given antiderivative of f. But the domain of $1/x$ consists of two rays $(-\infty, 0)$ and $(0, +\infty)$, two separate path-components!

As a result, we can, correctly, solve this problem *separately* on each, as follows. All antiderivatives are represented by: for every $C \in bfR$, we define

- $f_-(x) := \ln(-x) + C$ on $(-\infty, 0)$, and
- $f_+(x) := \ln(x) + C$ on $(0, +\infty)$.

But now we want to combine each of these pairs of functions into one, f, defined on $(-\infty, 0) \cup (0, +\infty)$. Then we realize that, every time, the constant might be different: after all, they have nothing to do with each other!

We visualize the wrong answer on the left and the right on the right:



The formula and the image on the left suggest that the antiderivatives are even functions. The image on the right shows only one antiderivative but its two branches don't have to match!

Algebraically, the correct answer is given by

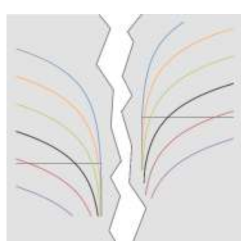
$$F(x) = F_{C,K}(x) := \begin{cases} \ln(-x) + C & \text{for } x \in (-\infty, 0), \\ \ln(x) + K & \text{for } x \in (0, +\infty) \end{cases}$$

This family of functions isn't parametrized by the reals $\mathbf{R}$ anymore, but, with two parameters, by $\mathbf{R}^2$. The exponent reflects *the number of components* of the domain!

Question: But isn't F supposed to be continuous? Because it doesn't look continuous...

Exercise 2.19. Verify that it is. Hint: check point-wise.

This is how we see the disconnected reality resulting in two disconnected calculi:



Next, one might recall something else from calculus:

"Theorem": If $F' = 0$ then F is constant.

But this isn't the whole story, as we just saw.

Let's state it properly.

Theorem 2.20. Given a function $F : X \to \mathbf{R}$, where X is an open subset of $\mathbf{R}$, and
- F is differentiable on X,
- $F'(x) = 0$ for all $x \in X$.

Then F is constant, provided X is path-connected.

Exercise 2.21. Give an example of a disconnected X for which the conclusion of the theorem fails.

Exercise 2.22. Restate the theorem for a disconnected X.

Less precise, for now, are the ideas related to the role of *holes* in calculus. Recall that a vector field is a function $f : \mathbf{R}^2 \to \mathbf{R}^2$. It is called *conservative* on a region D if it is the gradient of a scalar function: $F = \operatorname{grad} f$.

Theorem 2.23. Suppose $F = (P, Q)$ and $P_y = Q_x$ on region D. Then F is conservative provided D is simply connected.

Exercise 2.24. In light of this analysis, discuss Escher's *Waterfall*.

We have demonstrated the importance of the interaction between calculus and topology. Above are examples of something very common:
- $\longrightarrow$ Given X, if we know its topology, we know a lot about calculus on X.
- $\longleftarrow$ Given X, if we know calculus on it, we know a lot about the topology of X.

---

3 Topology of graphs

3.1 Graphs and their realizations

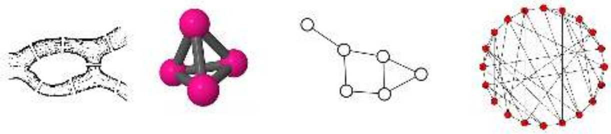
Here we take our first step toward *algebraic topology*.

We will concentrate initially on discrete structures, such as graphs. The reason is clear from our attempts to answer some obvious topological questions about specific spaces. Given a subset of the Euclidean space:
- verifying that it is path-connected requires testing *infinitely many pairs of points*; and
- verifying that it is simply connected requires testing *infinitely many pairs of loops*.

Even though there are a few theorems that we can use as shortcuts, we have no algebraic means for solving these problems computationally, nor for computing the number of path-components or the number of holes.

In the rest of the chapter we will:
- define discrete structures and find discrete analogs of these topological ideas,
- develop algebraic methods for solving the two topological problems above for discrete structures, and
- develop methods for studying transformations of these structures.

The very first data structure that needs to be analyzed topologically is *graphs*:



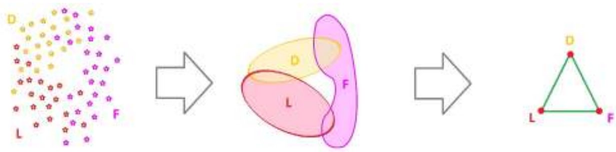
The examples of graphs shown come as *networks* of bridges, atoms in a molecule, related individuals, or computers.

Definition 3.1. A *graph* $G = (N, E)$ consists of two finite sets:

- the set of *nodes* $N = N_G$, and
- the set of *edges* $E = E_G$, which is a set of pairs of nodes.

We don't allow distinct edges for the same pair of nodes, nor doubles, such as AA, and we assume that $AB = BA$.

Example 3.2. In the example of a field of flowers, each type is represented by a node of a graph and the overlapping areas by its edges:

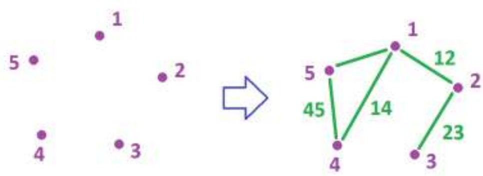


□

Example 3.3. Let's define a graph G by:

- $N := \{1, 2, 3, 4, 5\}$, and
- $E := \{12, 23, 45, 14\}$.

Let us be clear from the beginning that the dataset given above *is* the whole graph and what's below is just an illustration:



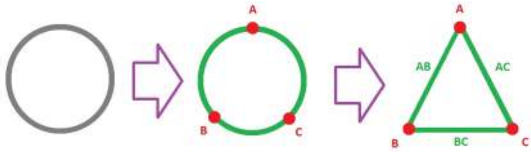
To draw an illustration, we first put the nodes in a more or less circular arrangement and then add the corresponding edges.

A common representation of a graph is via its *incidence matrix*, i.e., the matrix with a 1 in the (i, j)-entry if the graph contains edge ij and 0s elsewhere. For G, we have:

N	1	2	3	4	5
1	0	1	0	1	0
2	1	0	1	0	0
3	0	1	0	0	0
4	1	0	0	0	1
5	0	0	0	1	0

□

One can also think of a graph as just a collection of points in space, also called "vertices", or "nodes", connected by paths, called "edges". Thus, a graph is made of finitely many pieces that are known to be path-connected! One can then study the topology of such sets by means of graphs represented as discrete structures:



Definition 3.4. A *realization* $|G|$ of graph G is a subset of the Euclidean space that is the union of the following two subsets of the space:

- a collection of points $|N|$, one for each node N in G, and
- a collection of paths $|E|$, one for each edge E in G, with no intersections other than the points in $|N|$.

In other words, $a, b \in |N|$ are connected by a path $p \in |E|$ if and only if there is an edge in $AB \in E$, where A, B are the nodes corresponding to points a, b as follows:

- $a = |A|,\ b = |B|$, and
- $p = |AB|$.

The definition is justified by the following.

Theorem 3.5. Every finite graph can be realized in $\mathbf{R}^3$.

Certainly, there may be many different realizations of the same graph.

Exercise 3.6. Provide an explicit realization of a graph with 4 nodes in $\mathbf{R}^3$. Hint: start with two nodes.

Now we provide discrete analogs of our topological definitions.

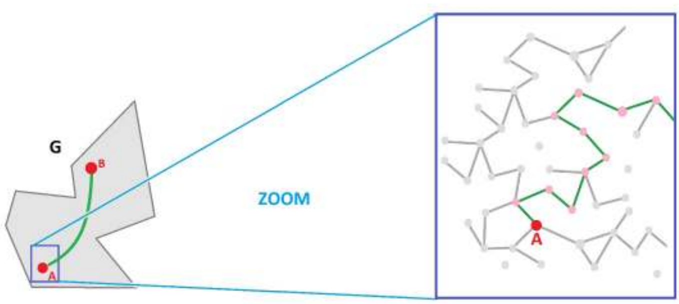
3.2 Connectedness and components

We start with the discrete analog of the most basic topological property.

Definition 3.7. A *edge-path*, or simply path, from node $A \in N_G$ to node $B \in N_G$ in graph G is a sequence of edges

$$A_0A_1,\ A_1A_2, \ldots , A_{n-1}A_n \in E_G$$

with $A_0 = A,\ A_n = B$. A graph G is called *edge-connected*, or connected, if for any pair of nodes $A, B \in N$, there is a path from A to B.



We now link this new concept to its origin.

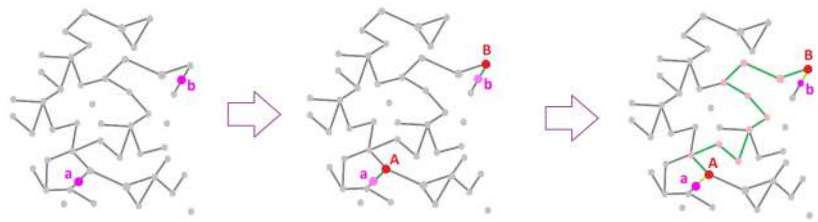
Theorem 3.8. A graph G is edge-connected if and only if its realization $|G|$ is path-connected.

Proof. [$\Longrightarrow$] Suppose G is an edge-connected graph and $a, b \in |G|$ are two points in its realization. Then a, b belong to the realizations of some edges of the graph: $a \in |AA'|, b \in |BB'|$. Since G is edge-connected, there is an edge-path

$$A_0A_1, A_1A_2, ..., A_{n-1}A_n$$

in the graph from A to B. Then a path from a to b is constructed from these three pieces:

- 1. the piece of $|AA'|$ that connects a to $|A|$,
- 2. the realizations $|A_iA_{i+1}|$ of the edges A_iA_{i+1}, and
- 3. the piece of $|BB'|$ that connects $|B|$ to b.



[$\Leftarrow$] Suppose we have a realization $|G|$ of G and it is path-connected and suppose $A, B \in N_G$ are two nodes of the graph. We know that there is a path $p : [0, 1] \to |G|$ from $|A|$ to $|B|$ in $|G|$. Now, starting from $t = 0$ every time $p(t)$ is a realization of a node Q, add Q to the list L of nodes. Now, this list may be infinite. Replace each repetition QQ in L with Q and L becomes finite:

$$L = \{Q_0 = A, Q_1, ..., Q_n = B\}.$$

Finally, our edge-path in G from A to B is

$$P := Q_0Q_1, Q_1Q_2, ..., Q_{n-1}Q_n$$

■

Exercise 3.9. Provide a formula for the path in the first part of the proof and prove that it is continuous.

Exercise 3.10. Show that, in the second part of the proof, the list L may indeed be infinite. Also prove that, after removing repetitions, L becomes finite.

Just as before, in order to *evaluate* the topology of the graph, we make a transition to homology.

Definition 3.11. Two nodes $A, B \in N_G$ of graph G are called *homologous* if there is an edge-path from A to B.

To avoid confusion, let's emphasize here that homology is a *relation between two nodes* of the graph. It certainly is not a relation between the points of a realization, nor is it a relation between the edges of the graph! To illustrate, the nodes in a graph normally represent some agents while the edges may represent some pairwise relations between them, such as: people and their family relations. We can't allow these beings to mix.

Theorem 3.12. Homology is an equivalence relation on the set of nodes of the graph.

Exercise 3.13. Prove the theorem.

The equivalence classes are called *edge-components*, or simply components, of G.

Theorem 3.14. For any graph G, we have

$$\# \text{ of edge-components of } G = \# \text{ of path-components of } |G|.$$

Exercise 3.15. Prove the theorem.

The problem is solved: we have expressed the topological property entirely in terms of data and now the computer can do this job.

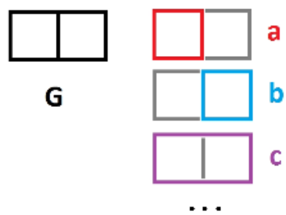
That's homology of dimension 0 (nodes) considered for objects of dimension 1 (graphs). Higher dimensions will pose a more significant challenge.

3.3 Holes vs. cycles

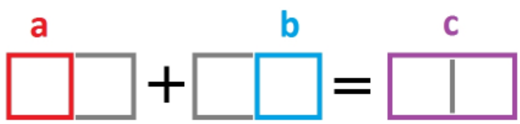
Definition 3.16. An edge-path in a graph G from node A to node A is called a *cycle* (i.e., a 1-cycle) in G.

In the last subsection, we studied holes and tunnels in objects by capturing them with 1-cycles. Then, in order to deal with the unavoidable redundancy, we declared 1-cycles homologous and, therefore, indistinguishable, if they form the boundary of some surface. In contrast, there is no meaningful way to fit a surface into a graph; so, homology as an equivalence relation won't be useful or necessary.

Is there still redundancy? Yes, but of a different kind. The realization of graph G below is the figure eight and it appears to have 2 holes. However, the graph itself has (at least) 3 cycles, a, b, c!



How do we eliminate the redundancy? We observe that combining a and b yields c:



Algebraically, it is:

$$a + b = c.$$

In other words, the cycles are *linearly dependent*. That's why c doesn't count.

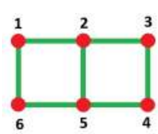
This brings us to the general idea:

$$\#\text{ of holes} = \#\text{ of linearly independent cycles.}$$

It is time now to start to recognize the *need for algebra* in topology.

In order to turn nodes and edges into algebraic entities, we could try something familiar first, such as the union. Unfortunately, the algebra of unions is inadequate as there is no appropriate meaning for subtraction: sometimes $(A \cup B) \setminus B \neq A$. The algebra that does work is familiar.

Example 3.17. Let's consider our graph again:



Now consider our two cycles:

$$\begin{aligned}
a &= 12 + 25 + 56 + 61 \\
b &= 23 + 34 + 45 + 52 \\
a + b &= 12 + 25 + 56 + 61 + 23 + 34 + 45 + 52 \\
&= 12 + (25 + 25) + 56 + 61 + 23 + 34 + 45 \\
&= 12 + 56 + 61 + 23 + 34 + 45 \\
&= c
\end{aligned}$$

Here we cancel the edge that appears twice. □

Such cancelling is assured if we use the *binary arithmetic*:

$$x + x = 0, \ \forall x.$$

It is the arithmetic of integers modulo 2, i.e., the algebra of $\mathbf{Z}_2$.

This idea takes us deep into algebra. One unexpected conclusion is that

0 is also a cycle!

Now, we list the four 1-cycles that have no repeated edges:

$$Z_1 := \{0, a, b, a + b\}.$$

It is a *group*. Because a and b generate the rest of the cycles ($Z_1 =< a, b >$), these two are the only ones that matter. Therefore,

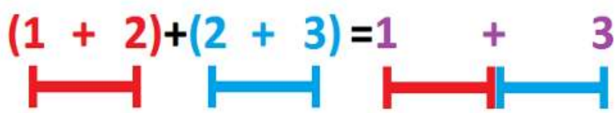
$$\# \text{ of holes} = 2.$$

Note: If we choose to stay within the realm of linear algebra, we have two options: either think of Z_1 as a vector space with coefficients over $\mathbf{Z}_2$ or enlarge Z_1 by including *all* of its real-valued linear combinations: $ra + sb, \ \forall r, s \in \mathbf{R}$. Then $\{a, b\}$ is a basis of this vector space and the idea becomes:

$$\# \text{ of holes } = \dim Z_1.$$

However, in order to assure cancellation, we would have to deal with the complexity of *directed* edges, with $AB \neq BA$. We will postpone following this idea until later.

Above, we have justified the use of binary arithmetic for edges but it is just as appropriate for the algebra of nodes. It suffices to consider the boundaries of edge-paths. Clearly, the boundary of an edge-path is the sum of its end-nodes. For instance, the path $12 + 23$ in the above graph has boundary $1 + 3$.



On the other hand, its boundary is the sum of the boundaries of 12 and 23. Those are $1 + 2$ and $2 + 3$. We end up with

$$1 + 3 = (1 + 2) + (2 + 3).$$

Hence, $2 + 2 = 0$. That's binary arithmetic again.

Let's review:
- for dimension 0, we captured components by means of nodes, and
- for dimension 1, we captured holes by means of cycles of edges.

In either case, there is redundancy: many nodes per component and many cycles per hole. To resolve the redundancy, we did the following:
- for dimension 0, we counted only the equivalence classes of nodes, and
- for dimension 1, we counted only the linearly independent cycles.

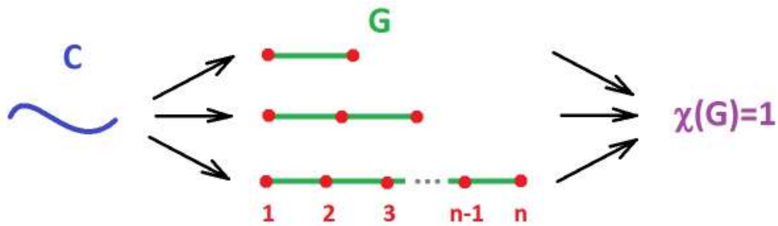
These are two very different solutions to two similar problems. In our study of dimension 2 (e.g., surfaces) and above, we'll need to use a combination of these two techniques.

3.4 The Euler characteristic

Definition 3.18. The *Euler characteristic* of graph G is

$$\chi(G) := \# \text{ of vertices } - \# \text{ of edges}.$$

The topological significance of this number is revealed if we consider a simple, i.e., without self-intersections, curve C. Let's suppose G is a sequence of n consecutive edges:



Then

$$\chi(G) = n - (n-1) = 1.$$

So, this number is independent of n! Therefore, $\chi(G)$ could be a topological attribute, i.e., a "characteristic", of the curve itself that distinguishes it from curves that do have self-intersections. To justify this conclusion, however, we need the following.

Theorem 3.19. If a simple path is a realization of a graph, this graph is a sequence of $n \geq 1$ consecutive edges.

Exercise 3.20. Prove the theorem.

Exercise 3.21. Show that if C consists of k path-components each of which is a simple curve and $C = |G|$, then $\chi(G) = k$.

Then, could $\chi(G)$ be the number of path-components of $|G|$? The example of a triangle T shows that this is not correct as $\chi(T) = 0$. Below is best we can do.

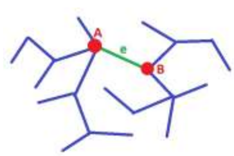
Theorem 3.22. If T is a *tree*, i.e., an edge-connected graph with no cycles, then

$$\chi(T) = 1.$$

Proof. *Idea:* Remove an edge and use induction.

So, we use induction on the number of edges on the graph.

First, tree T with a single edge has $\chi(T) = 1$. Now, we assume that any tree with fewer than n edges satisfies this identity. Suppose that T is a tree with n edges.



Remove *any* edge, $e = AB$, from T. The result is a new graph G and we have:

$$N_G = G_T, E_G = E_T - \{e\}.$$

What kind of graph is it?

It is disconnected. Indeed, let $H := [A]$ and $K := [B]$ be the edge-components of A, B respectively. As we know, either graph is connected. Secondly, removing an edge can't create cycles, so both are trees. Thirdly, there is no path from A to B, because if there was one, say, P, then the combination of P and e would be a path from A to A, a cycle in the tree T. Therefore, H and K are disjoint.

So, G splits into two trees H and K and each has fewer than n edges. Therefore, by assumption, we have

$$\chi(H) = \chi(K) = 1.$$

Let's compute now:

$$\begin{array}{lll}
\chi(T) & = & \#\text{ of vertices of } T - \#\text{ of edges of } T \\
& = & \#\text{ of vertices of } G - (\#\text{ of edges of } G + 1) \\
& = & \#\text{ of vertices of } G - \#\text{ of edges of } G - 1 \\
& = & \begin{pmatrix} \#\text{ of vertices of } H \\ +\#\text{ of vertices of } K \end{pmatrix} - \begin{pmatrix} \#\text{ of edges of } H \\ +\#\text{ of edges of } K \end{pmatrix} - 1 \\
& = & \Big(\#\text{ of vertices of } H - \#\text{ of edges of } H\Big) \\
& + & \Big(\#\text{ of vertices of } K - \#\text{ of edges of } K\Big) - 1 \\
& = & \chi(H) \\
& + & \chi(K) - 1 \\
& = & 1 + 1 - 1 \\
& = & 1.
\end{array}$$

■

Exercise 3.23. Find and fix two gaps in the proof. Hint: first prove that removing an edge can't create cycles.

Exercise 3.24. Repeat the proof of the theorem for e an "end-edge".

Exercise 3.25. What about the converse?

Corollary 3.26. If G consists of n disjoint trees, then $\chi(G) = n$.

Exercise 3.27. Prove the corollary.

Exercise 3.28. Prove the following generalization of the above theorem.

Theorem 3.29. If G is a graph, then

$$\chi(T) \leq 1.$$

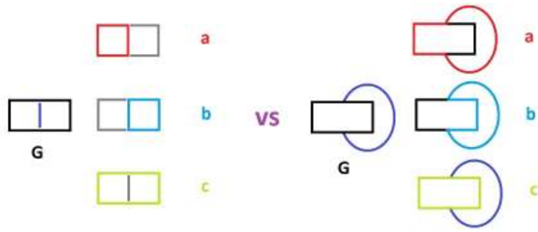
Moreover, $\chi(G) = 1$ if and only if G is a tree.

3.5 Holes of planar graphs

What is a hole in the graph? The question isn't as simple as it appears. Let's investigate.

First, a tree has no holes! At the same time, it has no cycles. This looks like a match. Maybe holes are cycles?

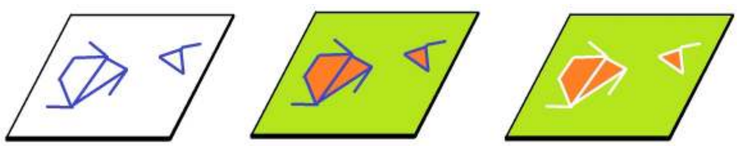
In general, we can't say what a hole *is*. Even though we now know how to *count* holes – as the number of linearly independent cycles – we can't point at one of the cycles and say "That's a hole and that one isn't". The example below shows that which of the cycles "naturally" look like holes depends on the realization of the graph:



Compare the following:
• 1. for the first realization, the holes "are" cycles a (red) and b (blue) but not c (green); while
• 2. for the second realization, it's b and c.

This ambiguity is caused by our *topological* point of view; we are allowed to bend, stretch, etc., and the results have to hold. But this ambiguity is, in fact, a good news because it is fully matched by the ambiguity of the *algebra*. Indeed, considering $Z_1 = \{0, a, b, c = a + b\}$, we have
• 1. a and b are generators, but
• 2. so are b and c!
(In the language of linear algebra, these are two bases of this vector space.)

We still would like to confirm our intuition about what holes in a graph are. For that, we will limit our attention to a simpler kind – *planar graphs*. Those are the graphs that can be realized in the plane:



The idea is, once a realization $|G|$ of planar graph G is chosen, the holes become visible as path-components of the *complement* of $|G|$.

We rely on the following two theorems (for the proof of the first see Munkres, *Topology. A First Course*, p. 374).

Theorem 3.30 (Jordan Curve Theorem). The complement of a simple closed curve in the plane has two path-components (one bounded and one unbounded).

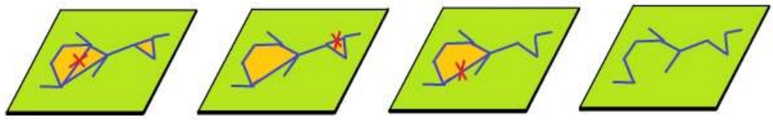
Theorem 3.31. If a simple closed curve is a realization of a graph, then this graph is a cycle of $n \geq 1$ consecutive edges.

Exercise 3.32. Prove the last theorem. Hint: try to answer these questions about G: how many components? cycles? holes? forks? endpoints?

So, we have one-to-one correspondences connecting these:
• the path-components of $\mathbf{R}^2 \setminus |G|$,
• the loops in G the realization of which bound them, and
• a certain basis of the space of cycles of G.

Let's count holes in this environment.

Suppose we are given an edge-connected planar graph G with a specific realization $|G|$ in the plane. The idea is to remove some edges one by one until we have a tree.



The process is inductive. We start with our graph $G_0 := G$. Then at each step k, we pick an edge in graph $G_k,\ k = 0, 1, ...$, that is a part of a cycle that bounds a hole in G_k and remove it from the graph. This act creates a new graph G_{k+1}. Clearly, removing an edge in a cycle that surrounds a hole "kills" the hole by merging it with another or with the outside. Then,

$$\#\text{ of holes in } |G_{k+1}| = \#\text{ of holes in } |G_k| - 1,$$

and

$$\#\text{ of edges in } G_{k+1} = \#\text{ of edges in } G_k - 1.$$

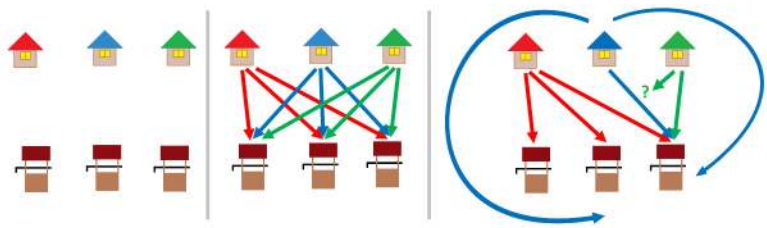
Exercise 3.33. Prove that after every step the new graph is connected.

Theorem 3.34. If a plane realization $|G|$ of an edge-connected graph G has n holes then

$$\chi(G) = 1 - n.$$

Proof. Above, we have shown that the graph with n holes will need n edges removed to turn it into a tree. Then the last theorem applies. ■

Example 3.35 (houses and wells). Suppose we have three houses and three wells arranged as shown below. All three wells are public but, since the house owners don't get along, they want to make paths from their houses to the wells that don't cross. The first attempt fails – there is no way to find a path from the third house to the second well:



Let's prove that this is indeed impossible. The graph has 6 nodes and 9 edges, therefore, by the formula above, there must be 4 holes in this planar graph. Next, considering the way the paths are to be arranged, each hole, and the outside area too, has to be bounded by at least 4 edges. This makes it necessary to have 20 edges, with some repetitions. How many? Each edge is shared by exactly two holes, counting the outside area. Therefore, there must be at least 10 edges, a contradiction. □

Exercise 3.36. Can we rearrange the houses and wells so that the paths can be found?

Exercise 3.37. Can this problem be solved on a sphere, a cylinder, a torus? Suggest others.

An important way to rephrase the last theorem is the following.

Corollary 3.38. For any edge-connected planar graph G, we have

$$\#\text{ of holes of } |G| = 1 - \chi(G).$$

Since the number on the right is independent of a choice of realization, so is the number on the left. This conclusion confirms the results of our algebraic analysis of cycles presented above.

Exercise 3.39. Suppose graph G is the disjoint union of m trees, find its Euler characteristic.

Exercise 3.40. Suppose graph G has m components and n holes, find its Euler characteristic.

3.6 The Euler Formula

We have demonstrated that the Euler characteristic of a graph G,

$$\chi(G) = \# \text{ of nodes} - \# \text{ of edges},$$

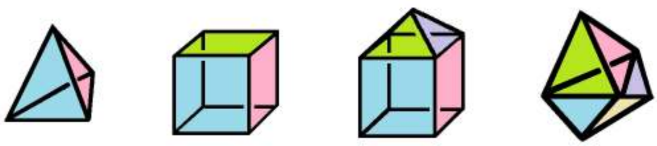
captures enough topology of G so that we can tell trees from non-trees. We have also noticed that it is the minus sign in the formula that ensures that this number is preserved when we add or remove "topologically irrelevant" edges.

Following these ideas, we move from dimension 1 to dimension 2 and define the *Euler characteristic of a polyhedron* P as the following number:

$$\chi(P) := \# \text{ of vertices } - \# \text{ of edges } + \# \text{ of faces},$$

in hope of capturing some or all of its topology.

Centuries ago, a pattern was noticed by observing a lot of various polyhedra:



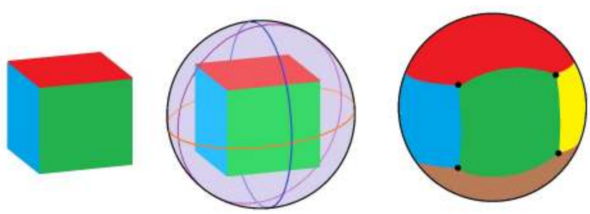
Theorem 3.41 (Euler Formula). For any convex polyhedron P, we have:

$$\chi(P) = 2.$$

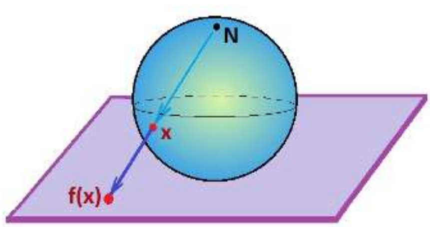
The formula has been proven to be useful, even though convexity isn't a topological property.

We will consider two proofs. The first one will use a result from the last subsection about planar graphs.

We start by observing that the collection of vertices (nodes) and edges of polyhedron P is always a realization of some graph G_P, but is it planar? It is demonstrated in tow steps. First, we see the graph realized on the sphere $\mathbf{S}^2$ – via the radial projection – illustrated for this cube:



Second, this spherical realization becomes planar under an appropriate choice of stereographic projection:



Exercise 3.42. Provide details of this construction.

Next, the corollary from the last subsection is used in the following computation:

$$\begin{aligned}\chi(P) &= \#\text{ of vertices of } P - \#\text{ of edges of } P + \#\text{ of faces of } P\\ &= \#\text{ of vertices of } G_P - \#\text{ of edges of } G_P + (\#\text{ of holes of } G_P + 1)\\ &= \chi(G_P) + \#\text{ of holes of } G_P + 1\\ &= \chi(G_P) + (1-\chi(G_P)) + 1\\ &= 1+1\\ &= 2.\end{aligned}$$

Unfortunately, this proof relies on the convexity of the polyhedron. This constraint is unnecessary: it might take the same number of vertices, edges, and faces to either

- add a sloped roof to a cube, or
- make an indentation in it.

Therefore, both constructions preserve the Euler characteristic:

While the former object is convex, the latter isn't! This fact suggests that the Euler characteristic has nothing to do with convexity and may depend only on the topology of the polyhedron.

Exercise 3.43. Try the torus.

An alternative approach to the **proof** of the formula is outlined below.

We assume that P is a polyhedron that satisfies these *topological* properties:

- 1. P is path-connected;
- 2. every loop cuts P into two disjoint pieces.

The idea of the proof is to build two trees, T and G, on P with the following properties:

- T contains all vertices of P and some edges;
- G contains all faces of P and the rest of the edges.

From what we know about trees, we conclude:

$$\begin{aligned}V_T - E_T &= 1,\\ V_G - E_G &= 1.\end{aligned}$$

We add the equations and rearrange the terms:

$$V_T - (E_T + E_G) + V_G = 2.$$

An examination reveals that these three terms are:

$$\#\text{ of vertices - }\#\text{ of edges + }\#\text{ of faces}$$

of P. And the theorem follows.

We just need to build T and G satisfying these conditions.

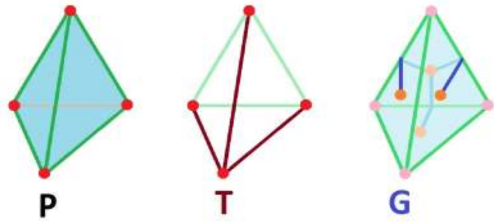
Exercise 3.44. Prove that every graph has a subgraph that

- contains all of the original vertices, and
- is a tree.

We apply this result to the graph of P. Then we have its subgraph T which is a tree and

$$\text{vertices of } T = \text{vertices of } P.$$

Next, we construct G from T:
- G's vertices are the middle points of the faces of T, and
- G's edges connect G's vertices by crossing the edges of P, but not the edges of T.



Then G is called the "dual graph" of T.

Exercise 3.45. It is always possible to get from one face to any adjacent face without crossing T. Why?

Exercise 3.46. A *regular polyhedron* is a polyhedron the faces of which have the same number of edges and the vertices of which have the same number of faces adjacent. Prove that there are exactly five of them (they called the *Platonic solids*):

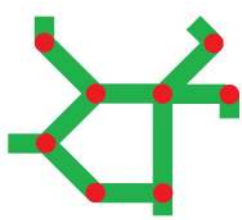
While useful, this computational but non-algebraic approach is a poor-man's substitute for homology.

4 Homology groups of graphs

4.1 The algebra of plumbing

We have been looking at the *subsets* of the sets of nodes and edges to study the topology of the graph. Next, we will pursue this analysis via a certain kind of *algebra*. We introduce this algebra with the following metaphor:

- One can think of a graph as a plumbing system that consists of a network of pipes and joints. There is no water yet.

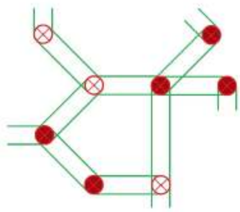


Let's assume that each joint is equipped with an on-off switch.

Suppose the plumbers are, when necessary, told to "flip the switch". Then two consecutive flips cause the switch to return to its original position. Furthermore, a sequence of such commands will result in another simple command: either "flip the switch" again or "do nothing". Naturally, these two commands are represented as 1 and 0 and combining these commands is represented by the binary arithmetic of $\mathbf{Z}_2$:

$$0+0=0,\ 1+0=1,\ 0+1=1,\ 1+1=0.$$

Now, this activity is happening at every joint. The plumber – call him the joints guy – receives complaints from the tenants about a possible leak in a given joint and a request to flip its switch.



Then a combined request is to flip a selection of switches, such as:

• flip switches at joints $A, B, ...$

The joints listed appear in the requests several times or never. Requests may be issued at any time and combined with the outstanding ones. The plumber may, using binary arithmetic, reduce his workload one joint at a time. Or, he can *combine* two or more work requests into one as follows. The joints are ordered and these requests are recorded in a vector format, coordinate-wise. For instance, $(1,0,1,0,...)$ means: flip the first and third switches, etc. Then,

$$(1,0,1,0,...)+(1,1,1,0,...)=(0,1,1,0,...).$$

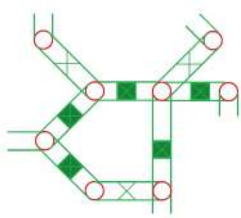
Here, 0 is the "do nothing" request. In other words, the algebra is that of this group:

$$\Big(\mathbf{Z}_2\Big)^n = \underbrace{\mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus ... \oplus \mathbf{Z}_2}_{\text{n times}},$$

where n is the number of the joints in the network.

Let' also suppose that each pipe has, too, an on-off switch. There is also another plumber – call him the pipes guy – who flips these switches by request from the tenants. A combined request is to take care of a selection of pipes, such as:

• flip switches of pipes $a, b, ...$



Once again, the pipes listed may appear several times and the requests may be issued at any time and combined with the outstanding one, using binary arithmetic. The set of all requests form a group: $\big(\mathbf{Z}_2\big)^m$, where m is the number of the pipes in the network.

Now, there is no overlap between the activities of the two plumbers so far, and there is no relation between any two joints, or any two pipes, or a joint and a pipe. However, what if the pipes guy doesn't show up for work? Then one may try to amend his request, such as

• flip the switch of pipe a,

with a request for the joints guy. The choice of a *substitution* is obvious, if imperfect:

• flip the switches at ends of pipe a.

The result is a function, from the pipe requests to the joint requests:

$$a = AB \ \mapsto \ A + B.$$

This is the key – an algebraic way to link the network together.

Exercise 4.1. How can the pipes guy try to help the joints guy?

No-one knows the complete topology of the network: each tenant only knows the plumbing in his apartment and the plumbers only flip switches. Nonetheless, if one has access to all plumbing requests and, especially, the substitutions, he can use the algebra described above to understand the topology of the whole network. We proceed to develop this theory.

4.2 Chains of nodes and chains of edges

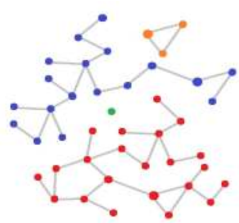
We will now approach topology of graphs in a fully algebraic way as the second step in our study of *homology theory*.

Suppose we are given a graph G:

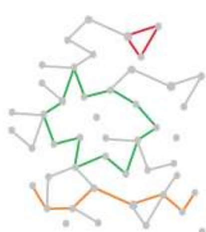
• the set of nodes N, and

• the set of edges $E \subset N \times N$, without repetitions.

We have seen the importance of some special *subsets* of these sets. First, combinations of nodes may form components:



Second, combinations of edges may form paths (and loops):



In order to deal with these sets, we follow the idea from the last section and impose algebra on vertices and edges in the following very deliberate manner.

Definition 4.2. We define *chains* as formal sums of nodes or edges

$$a_1 + a_2 + ... + a_s$$

that follow this **cancellation rule**:

$$x + x = 0.$$

It's as if revisiting a node or an edge removes it from consideration...

A chain of edges
- cannot include doubles (such as AA), and
- cannot include edges with opposite directions (we assume $AB = BA$), but
- can include repetitions, and
- can include 0.

Notation: The following notation will be used throughout the book. The set of chains of nodes, or 0-chains, is:

$$C_0 = C_0(G) := \Big\{ \sum_{A\in Q} A : A \subset N \Big\} \cup \{0\}.$$

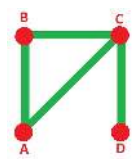
The set of chains of edges, or 1-chains, is

$$C_1 = C_1(G) := \Big\{ \sum_{AB\in P} AB : P \subset E \Big\} \cup \{0\}.$$

These sets include 0 understood as the chain with all coefficients equal to 0.

Example 4.3. Consider this graph G:
- $N = \{A, B, C, D\}$,
- $E = \{AB, BC, CA, CD\}$.

Its realization is shown below:



Then,

$$\begin{array}{ll} C_0 = & \{0, \\ & A, B, C, D, \\ & A + B, B + C, C + D, D + A, A + C, B + D, \\ & A + B + C, B + C + D, C + D + A, D + A + B, \\ & A + B + C + D\}, \\ C_1 = & \{0, \\ & AB, BC, CA, CD, \\ & AB + BC, BC + CA, ...\}. \end{array}$$

□

We recognize these as *abelian groups*. Recall that an abelian group is a set X with a binary operation called *addition*

$$x, y \in X \Longrightarrow x + y \in X,$$

such that
- the addition is associative and commutative;
- it contains the *identity element* $0 \in X$ that satisfies: $x + 0 = x, \forall x \in X$; and
- it contains the *inverse element* $-x$ of each $x \in X$ that satisfies $x + (-x) = 0$.

Exercise 4.4. Prove this conclusion.

In fact, we can rewrite the above groups via their generators:

$$\begin{array}{ll} C_0 = & < A, B, C, D >, \\ C_1 = & < AB, BC, CA, CD > . \end{array}$$

Recall that the *rank* of an abelian group is the minimal number of its generators. Then,

$$\operatorname{rank} C_0 = 4,\ \operatorname{rank} C_1 = 4.$$

Definition 4.5. The *group of chains of nodes*(or 0-chains) of the graph G is given by

$$C_0(G) =< N_G >,$$

and the *group of chains of edges* (or 1-chains) by

$$C_1(G) =< E_G > .$$

If we fix the *ordered* sets of nodes N and edges E as the bases of these groups, the chains can be written coordinate-wise as *column-vectors*:

$$A = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ ...,\ A+B = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, ...$$

We can rewrite these chains of nodes in a more compact way using transposition:

$$A = [1,0,0,0]^T,\ B = [0,1,0,0]^T,\ ..., A+B = [1,1,0,0]^T,\ ...;$$

as well as the chains of edges:

$$AB = [1,0,0,0]^T,\ BC = [0,1,0,0]^T,\ ..., AB+BC = [1,1,0,0]^T,\ ...$$

4.3 The boundary operator

Now, the relation between the edges E and the nodes N is simple:

the two endpoints of an edge are nodes of the graph.

Unfortunately, as there are two nodes A, B for each edge e, this relation is *not* a function $E \to N$. However, these two nodes do form a chain, **denoted** by

$$\partial e = A + B.$$

This is the key step:

the boundary of an edge, AB, is a *chain* of two nodes, $A+B$.

To define a function that will give us all boundaries of all edges of the graph G, we set its value for each edge, $AB \in E_G$, first:

$$\partial(AB) = A + B.$$

The next step is to extend it to the rest of the chains by "additivity" (or "linearity"):

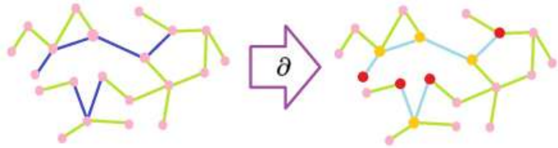
$$\partial(x+y) = \partial(x) + \partial(y).$$

Recall that a function that satisfies this identity "preserves the operations" of the group; i.e., it is a *homomorphism.* The extension step is simple:

$$\partial\Big(\sum_i x_i\Big) = \sum_i \partial(x_i),$$

where $\{x_i\}$ are the edges.

Example 4.6. Here is an example of a chain of edges (blue) and its boundary, a chain of nodes (red):



Some nodes appear twice (orange) and are cancelled, also seen in this algebra:

$$\begin{aligned}\partial(AB+BC+CD) &= \partial(AB)+\partial(BC)+\partial(CD)\\ &= (A+B)+(B+C)+(C+D)\\ &= A+(B+B)+(C+C)+D\\ &= A+0+0+D\\ &= A+D.\end{aligned}$$

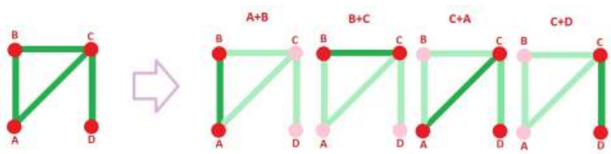
□

As a crucial step, for any graph G we have a homomorphism

$$\partial = \partial_G : C_1(G) \to C_0(G),$$

called the *boundary operator* of G. We will demonstrate that this operator contains – in the algebraic form – all information about the topology of the graph!

Example 4.7. Let's present the boundary operator of the graph from the last section:



We have, of course,

- 1. $\partial(AB) = A+B$,
- 2. $\partial(BC) = B+C$,
- 3. $\partial(CA) = C+A$,
- 4. $\partial(CD) = C+D$.

Rewrite these coordinate-wise:

- 1. $\partial\Big([1,0,0,0]^T\Big) = [1,1,0,0]^T$,
- 2. $\partial\Big([0,1,0,0]^T\Big) = [0,1,1,0]^T$,
- 3. $\partial\Big([0,0,1,0]^T\Big) = [1,0,1,0]^T$,
- 4. $\partial\Big([0,0,0,1]^T\Big) = [0,0,1,1]^T$.

The matrix of ∂ uses those as columns:

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

At this point, we (or the computer!) can easily find the boundary of any chain via simple matrix multiplication, as shown here:

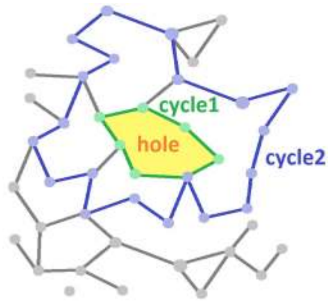
$$\partial(AB{+}BC) = \begin{bmatrix} 1 & 0 & 1 & 0\\ 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\1\\0\\0\end{bmatrix}.$$

□

Exercise 4.8. Compute all boundaries of all chains of edges for this graph.

4.4 Holes vs. cycles

We already know that the issue is challenging because multiple cycles may go around a single hole:



We won't examine the cycles all one by one but, instead, use the boundary operator to establish a relation between them.

In fact, the topology of (any realization of) the graph is hidden in the *algebra*, as we shall see. And this algebra is revealed by the behavior of graph's boundary operator. Indeed, just compare:

- topology: the boundary of a loop is empty,
- algebra: the boundary of a cycle is zero.

Example 4.9. Let's test the latter:

$$\begin{aligned}\partial(AB+BC+CA) &= \partial(AB)+\partial(BC)+\partial(CA)\\ &= (A+B)+(B+C)+(C+A)\\ &= A+A+B+B+C+C\\ &= 0.\end{aligned}$$

□

Exercise 4.10. State and prove this statement in full generality.

Definition 4.11. A 1-chain is called a 1-*cycle* if its boundary is zero. The *group of* 1-*cycles* of graph G is defined to be

$$Z_1 = Z_1(G) := \{x \in C_1(G) : \partial_G x = 0\}.$$

Of course, this is simply the *kernel* of the boundary operator:

$$Z_1(G) := \ker \partial_G.$$

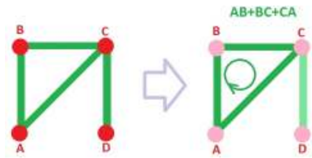
Note: "C" in C_0, C_1 stands for "chain", while "Z" in Z_1 stands for "Zyklus", "cycle" in German.

We know that kernels are subgroups. Indeed, a quick proof shows that the kernel is closed under the operations of the group:

$$\begin{aligned}x, y \in \ker\partial &\implies \partial(x) = \partial(y) = 0\\ &\implies \partial(x+y) = \partial(x)+\partial(y) = 0+0 = 0\\ &\implies x+y \in \ker\partial.\end{aligned}$$

This makes Z_1 a subgroup of C_1. The "size" of a group is measured by its rank (or, in the language of linear algebra, its dimension).

Example 4.12. Some simple algebra is sufficient to find the group of cycles for our example graph:



We use the result above, $\partial(AB + BC + CA) = 0$, and a direct examination of C_1 to conclude that

$$\partial(x) = 0 \Longrightarrow x = AB + BC + CA \text{ or } x = 0.$$

Hence,

$$\begin{array}{ll} Z_1 & =< AB + BC + CA >= \{0, AB + BC + CA\}, \\ \text{rank}\, Z_1 & = 1. \end{array}$$

That's how we know that there is only one hole! □

So, our algebraic-topological **conclusion** is that for any graph G,

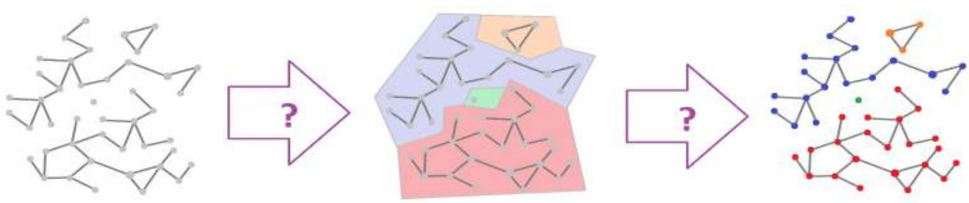
$$\# \text{ of holes of } |G| = \text{rank} \ker \partial_G.$$

In linear algebra, the number $\dim \ker \partial$ is known as the "nullity" of the linear operator.

Exercise 4.13. Compute $Z_1(G)$ for G the graph with 7 edges realized as the figure eight.

4.5 Components vs. boundaries

What about the components? The problem we are to solve is often called *component labeling*:



Let's review what we know. We defined, in parallel, two equivalence relations in two different realms and then counted their equivalence classes:

• Topology: Two *points* $x, y \in X$ in set X are homologous, $x \sim y$, if there is a path between them. Then the number of path-components of X is the number of homology classes of points in X.

• Combinatorics: Two *nodes* $A, B \in N_G$ in graph G are homologous, $A \sim B$, if there is a path of edges between them. Then the number of edge-components of G is the number of homology classes of nodes in G.

Now, let's try to understand what we need to change in the latter in order to make our analysis fully algebraic: we need to progress from nodes and edges to *chains* of nodes and edges. We approach the problem in a fashion similar to that in the last subsection:

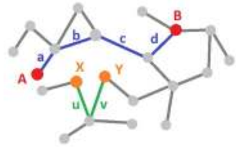
• Algebra: Two *chains of nodes* $P, Q \in C_0(G)$ in graph G are homologous, $P \sim Q$, if there is a path (or paths?) of edges between them. These equivalence classes should form some group and the number of components of G will be the rank of this group.

First we realize that

an edge-path from A to B = a *chain* of edges the boundary of which is $A + B$.

To confirm, consider the next example.

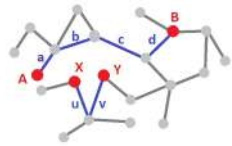
Example 4.14. Suppose we are given a graph:



Then,

$$\begin{array}{lll}\partial(a+b+c+d) & = A+B & \iff & A \sim B,\\ \partial(u+v) & = X+Y & \iff & X \sim Y.\end{array}$$

But such a chain of edges doesn't have to be a single edge-path. Why not *two paths*? We can join these two into one *chain*:



Does this make sense algebraically? Yes:

$$\partial(a+b+c+d+u+v) = A+B+X+Y \iff A+X \sim B+Y.$$

After all, ∂ is additive!

A more precise way to understand the result is given by this equivalence:

$$\partial(a+b+c+d+u+v) = A+B+X+Y \iff A+X+B+Y \sim 0. \qquad \square$$

Then, our definition is very simple.

Definition 4.15. Two chains of nodes $P, Q \in C_0(G)$ in graph G are *homologous*, $P \sim Q$, if there is a chain of edges $w \in C_1(G)$ with $\partial(w) = P + Q$.

In particular, a chain of nodes $P \in C_0(G)$ in graph G is homologous to zero 0 if there is a chain of edges w with $\partial(w) = P$. In other words,

$$P \sim 0 \iff P = \partial(w) \text{ for some 1-chain } w.$$

Exercise 4.16. Prove that $P \sim Q \iff P + Q \sim 0$.

Definition 4.17. A 0-chain is called a 0-*boundary* if it is the boundary of some 1-chain. The *group of* 0-*boundaries* of graph G is defined as

$$B_0 = B_0(G) := \{P \in C_0(G) : \ P = \partial_G(w) \text{ for some } w \in C_1(G)\}.$$

Of course, this definition can be reworded as follows:

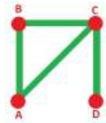
$$B_0(G) := \operatorname{Im} \partial_G.$$

We know that the images of groups under homomorphisms are subgroups. Indeed, a quick proof shows that this set is closed under the operations:

$$\begin{array}{rcl}x, y \in \operatorname{Im} \partial & \Longrightarrow & x = \partial(X), y = \partial(Y)\\ & \Longrightarrow & x + y = \partial(X) + \partial(Y) = \partial(X+Y)\\ & \Longrightarrow & x + y \in \operatorname{Im} \partial.\end{array}$$

This makes B_0 a subgroup of C_0.

Example 4.18. Let's find the group of boundaries for our example graph:



The group is generated by the values of the boundary operator on the edges, which we have already found:

$$\partial(AB) = A + B,\ \partial(BC) = B + C,\ \partial(CA) = C + A,\ \partial(CD) = C + D.$$

We compute:

$$\begin{array}{lll} B_0 & = & < A+B, B+C, C+A, C+D > \\ & = & < A+B, B+C, C+D > \\ & = & \{0, \\ & & A+B, B+C, C+A, C+D, B+D, A+D, \\ & & A+B+C+D\}, \\ \operatorname{rank} B_0 = 3. & & \end{array}$$

It follows that $\operatorname{rank} C_0 - \operatorname{rank} B_0 = 1$. That's how we know that there is only one component! □

Exercise 4.19. Modify the above analysis when edge AD is added.

The boundary group isn't the goal of our quest however. This group represents the redundancy that needs to be removed from the group of chains! How?

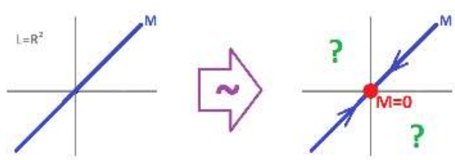
4.6 Quotients in algebra

Let's review the quotient construction for abelian groups and vector spaces as it will reappear many times.

We are given an abelian group L and a subgroup M (or a vector space L and a subspace M). How do we "remove" M from L? The simple answer of $L \setminus M$ won't work because it's not a group!

Instead, we "collapse" M to 0:

$$x \sim 0 \text{ if } x \in M.$$



The question now is, of course, what about the rest of L?

For the end result to be a group we need to make sure that the equivalence relation we are constructing respects the algebraic operations of L. We then extend the above equivalence $\sim$ to the whole L by invoking its algebra:

$$x \sim y \text{ if } x - y \sim 0,$$

or

$$x \sim y \text{ if } x - y \in M.$$

Then the equivalence class of $v \in L$ is

$$[v] := \{x : \ x - v \in M\} = \{x = v + m : \ m \in M\} = v + M.$$

We have proved the following.

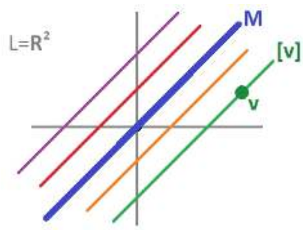
Theorem 4.20. The equivalence class of $v \in L$ in L/M is a *coset* of the group (or the affine subspace of the vector space) produced when M is "shifted" by v:

$$[v] := v + M.$$

Example 4.21. Suppose M is the *diagonal* in $L = \mathbf{R}^2$:

$$M = \{(r,r) : \ r \in \mathbf{R}\}.$$

Then each equivalence class is a line:

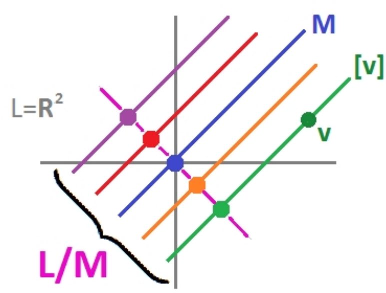


We can clearly see the partition induced by this equivalence relation.

Further, we have:

$$[v] := \{v + (r,r) : \ r \in \mathbf{R}\} = v + M.$$

In other words, the elements of L/M are the lines parallel to M. We notice, of course, that these lines appear to be in a one-to-one correspondence with the points of the other diagonal line $y = -x$:



A visualization of the case of free abelian groups, such as $L = \mathbf{Z}^2$, would be similar but those lines would be dotted. □

Theorem 4.22. The quotient set L/M is a group with the operation of addition:

- $[u] + [v] = [u + v], \ u, v \in L,$

and, in the case of vector spaces, also scalar multiplication:

- $q[u] = [qu], \ u \in L, q \in \mathbf{R}.$

The new group is **denoted** by L/M as we've chosen to "divide" rather than "subtract". Just consider:

$$\mathbf{R}^n / \mathbf{R}^m = \mathbf{R}^{n-m}, \ n > m.$$

Then it's easy to see that the operations on L/M are well-defined:

$$[v] + [u] := (v + M) + (u + M) = v + u + (M + M) = v + u + M =: [v + u],$$

and

$$q[v] := q(v+M) = qv + qM = qv + M =: [qv].$$

Example 4.23. Finite groups, such as $\mathbf{Z}_p$, are harder to visualize in comparison to vector spaces. The advantage is that such a group is nothing but a *list*. Let's compute, from the definition,

$$\Big(\mathbf{Z}_2 \oplus \mathbf{Z}_2\Big) / < (1,0) > .$$

Because $\mathbf{Z}_2 = \{0,1\}$, we have the numerator:

$$L = \Big\{(0,0),(1,0),(0,1),(1,1)\Big\},$$

and the denominator

$$M = \{0,1\}.$$

Then we compute the cosets:

$$\begin{array}{llll}
[(0,0)] & := (0,0)+M = (0,0)+\Big\{(0,0),(1,0)\Big\} & & \\
& = \Big\{(0,0)+(0,0),(0,0)+(1,0)\Big\} & = \Big\{(0,0),(1,0)\Big\} & = M =: 0; \\
[(1,0)] & := (1,0)+M = (1,0)+\Big\{(0,0),(1,0)\Big\} & & \\
& = \Big\{(1,0)+(0,0),(1,0)+(1,0)\Big\} & = \Big\{(1,0),(0,0)\Big\} & = M =: 0; \\
[(0,1)] & := (0,1)+M & = (0,1)+\Big\{(0,0),(1,0)\Big\} & \\
& = \Big\{(0,1)+(0,0),(0,1)+(1,0)\Big\} & = \Big\{(0,1),(1,1)\Big\}; & \\
[(1,1)] & := (1,1)+M & = (1,1)+\Big\{(0,0),(1,0)\Big\} & \\
& = \Big\{(1,1)+(0,0),(1,1)+(1,0)\Big\} & = \Big\{(1,1),(0,1)\Big\}. &
\end{array}$$

So, we've found the quotient with the generator explicitly presented:

$$L/M =< \Big\{(0,1),(1,1)\Big\} > .$$

□

Exercise 4.24. Compute, from the definition,

$$\Big(\mathbf{Z}_2 \oplus \mathbf{Z}_2\Big) / < (1,1) > .$$

Exercise 4.25. Compute, from the definition,

$$\mathbf{Z}/ < 2 > .$$

Theorem 4.26. For a finitely-generated abelian group L and its subgroup M, we have

$$\operatorname{rank} L/M = \operatorname{rank} L - \operatorname{rank} M,$$

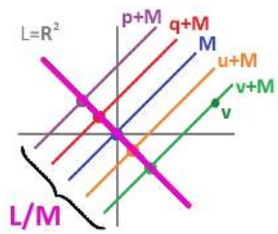
and for a finitely-dimensional vector space L and its subspace M,

$$\dim L/M = \dim L - \dim M.$$

Note that the dimension behaves like the logarithm!

Example 4.27. In the above example, we can collapse the equivalence classes in a particular way so that they "form" a line. The line is $y = -x$ and it is perpendicular to M. Then one may, informally, suggest an isomorphism of vector spaces:

$$L \cong M \oplus L/M.$$



□

Exercise 4.28. Show that a choice of any other line through 0, other than M itself, would be just as valid an illustration of this idea.

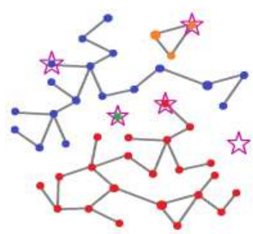
Example 4.29. To illustrate $L/M = \mathbf{R}^3/\mathbf{R}^2$, these are the equivalence classes, given as cosets and as subsets:

□

Exercise 4.30. Sketch an illustration for $\mathbf{R}^3/\mathbf{R}^1$.

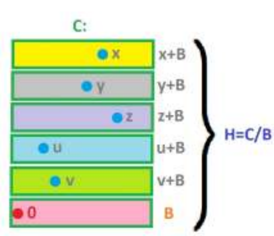
4.7 Homology as a quotient group

Below we visualize what would like to represent algebraically:



Here, each color corresponds to an element of the group we are after including zero (white). Meanwhile, we mark one – with a star – in each as a representative.

We are after the *components* of the graph and they are exactly what's left once the boundaries are removed from consideration. As discussed above, we remove a subgroup B from group C not by deletion but by taking its quotient $H = C/B$ modulo B:



Here the denominator B becomes the *zero* of the new group.

This is how the construction applies to the chains of nodes:

$$x \sim y \iff x - y \in B_0.$$

Since we are using the binary addition, this is equivalent to:

$$x \sim y \iff x + y \in B_0,$$

so $x+y$ is indeed a boundary. This quotient group is the *group of components* of the graph:

$$C_0(G)/_{\sim} := C_0(G)/B_0(G).$$

Its elements are the equivalence classes of our equivalence relation as well as the *cosets* of the group of boundaries:

$$[0] = B_0,\ [x] = x + B_0,\ [y] = y + B_0, ...$$

Example 4.31. For our example graph, $\operatorname{rank} B_0 = \operatorname{rank} C_0 - 1$ and, therefore, the group of components is 1-dimensional. Once again, a single component! □

Then our algebraic-topological *conclusion* is

$$\#\text{ of components of } G = \operatorname{rank} C_0 - \operatorname{rank} \operatorname{Im} \partial_G.$$

More commonly, the group we have constructed is called the 0th *homology group*:

$$H_0 = H_0(G) := C_0(G)/B_0(G).$$

Meanwhile, the cycle group Z_1 of a graph is also called the 1st *homology group*:

$$H_1 = H_1(G) := \ker \partial_G.$$

We will see later that, for an arbitrary dimension n and an arbitrary space X, the homology group $H_n(X)$ of X captures the interaction between:

- the n-cycles in X, that come from looking at dimension $n-1$, and
- the n-boundaries in X, that come from looking at dimension $n+1$.

It is defined as

$$H_n(G) := Z_n(G)/B_n(G).$$

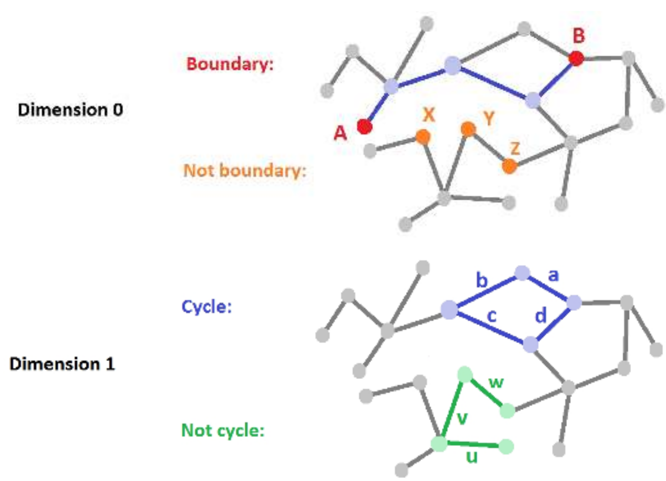
As we have seen, this analysis applied to graphs is much simpler than that because there are only two classes of chains. Then:

- in dimension $n=0$, "everybody is a cycle", $Z_0 = C_0$; while
- in dimension $n=1$, "nobody is a boundary", $B_1 = 0$.

Exercise 4.32. Explain why $Z_0 = C_0,\ B_1 = 0$.

Exercise 4.33. What conditions does a pair of abelian groups H_0, H_1 have to satisfy so that there is a graph G with $H_i = H_i(G),\ i = 0, 1$?

Here is the summary of homology theory for graphs:



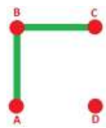
4.8 An example of homological analysis

Let's now present a complete write-up of analysis of the topology of a graph.

Example 4.34. Consider this simple graph G:

$$N = \{A, B, C, D\} \text{ and } E = \{AB, BC\}.$$

Its realization may be this:



The groups of chains are fully listed:

$$\begin{array}{ll} C_0 = & \{0, \\ & A, B, C, D, \\ & A+B, B+C, C+D, D+A, A+C, B+D, \\ & A+B+C, B+C+D, C+D+A, D+A+B, \\ & A+B+C+D\}, \\ \operatorname{rank} C_0 = 4; & \end{array}$$

$$\begin{array}{ll} C_1 = & \{0, \\ & AB, BC, \\ & AB+BC\}, \\ \operatorname{rank} C_1 = 2. & \end{array}$$

Based on the ranks of these groups, the boundary operator $\partial : C_1 \to C_0$ is given by a 4×2 matrix. It satisfies:

- 1. $\partial(AB) = A + B$,
- 2. $\partial(BC) = B + C$;

or, written coordinate-wise:

- 1. $\partial\big([1, 0]^T\big) = [1, 1, 0, 0]^T$,
- 2. $\partial\big([0, 1]^T\big) = [0, 1, 1, 0]^T$.

Therefore, the matrix of ∂ is

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

The boundaries of all the chains of edges are then computed:

$$\begin{array}{lll} \partial(AB) & = A+B & \neq 0, \\ \partial(BC) & = B+C & \neq 0, \\ \partial(AB+BC) & = A+B+B+C = A+C & \neq 0. \end{array}$$

Therefore, the group of cycles is

$$\begin{array}{ll} Z_1 & = \ker \partial = \{x \in C_1 : \partial x = 0\} = \{0\}, \\ \operatorname{rank} Z_1 & = 0. \end{array}$$

Conclusion: *There are no holes.*

The group of boundaries is computed next:

$$\begin{array}{ll} B_0 & = \operatorname{Im} \partial =< A+B, B+C >= \{0, A+B, B+C, C+A\}, \\ \operatorname{rank} B_0 & = 2. \end{array}$$

The homology group is then:

$$H_0 = C_0/B_0,$$

with the zero element

$$[0] = B_0 = \{0, A+B, B+C, C+A\}.$$

The rest of the cosets are computed by adding element by element:

$$\begin{array}{lll}
[A] & = A + B_0 & = \{A, B, A+B+C, C\}, \\
[D] & = D + B_0 & = \{D, D+A+B, D+B+C, D+C+A\}, \\
[A]+[D]=[A+D] & = A+D+B_0 & = \{A+D, D+B, A+D+B+C, D+C\}.
\end{array}$$

This gives us the list of elements of the homology group:

$$\begin{array}{ll}
H_0 & =< [A],[D] >= \{[0],[A],[D],[A]+[D]\}, \\
\operatorname{rank} H_0 & = 2.
\end{array}$$

We can also see this result in the following:

$$\operatorname{rank} H_0 = \operatorname{rank}(C_0/B_0) = \operatorname{rank} C_0 - \operatorname{rank} B_0 = 4 - 2 = 2.$$

Conclusion: *The number of components is* 2. □

Exercise 4.35. Modify the above analysis for the case when (a) a new node is added with no new edges, and (b) a new edge is added with no new nodes.

Exercise 4.36. Provide a similar analysis for the graph:

$$N = \{1, 2, 3, 4\},$$
$$E = \{23, 34, 42\}.$$

Exercise 4.37. In a similar fashion, compute the homology groups of the graph of n edges arranged in (a) a string, (b) a circle, (c) a star.

Exercise 4.38. In a similar fashion, compute the homology groups of the graph of edges arranged in an $n \times m$ grid.

Exercise 4.39. Compute the homology groups of an arbitrary tree.

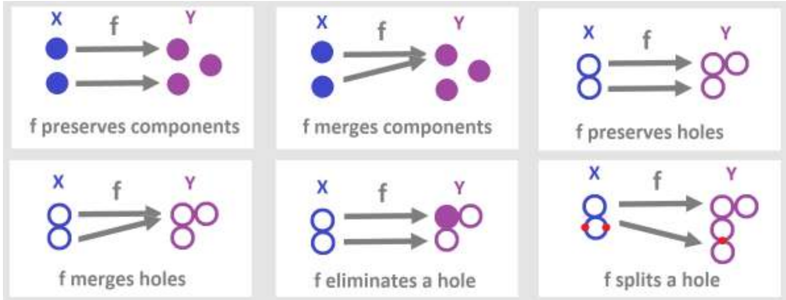
---

5 Maps of graphs

5.1 What is the discrete counterpart of continuity?

Any study of topology would be incomplete without considering continuous transformations of objects, also known as maps.

Below, we sketch a few simple examples of maps. One should notice that simply counting the topological features, i.e., components and holes, in the domain and the target spaces doesn't reveal the complete picture. Instead, we want to *track each of these features* and record what map does to it.



Exercise 5.1. We "split" the hole here, why don't we split a component?

In order to be able to employ the algebraic machinery we've started to develop, we would like to match these transformations with their discrete counterparts. These counterparts are functions of graphs:

$$f : G \to J,$$

and just have to make sure that these functions match the continuity of the originals.

To sort this out, we start with just the nodes.

Example 5.2. Let's illustrate the behavior of those continuous functions above that preserve and merge components. We consider two edge-less graphs G and J:

$$N_G = \{A, B\},\ E_G = \emptyset,\ N_J = \{X, Y, Z\},\ E_J = \emptyset,$$

and define a "function of nodes":

$$f_N : N_G \to N_J$$

by either

- $f_N(A) = X,\ f_N(B) = Y$, or
- $f_N(A) = X,\ f_N(B) = X$.

Then, of course, these functions generate point-by-point functions on the *realizations* of these graphs and those functions demonstrate the topological behavior depicted above. □

Next, we need to understand what may happen to an edge under such a "discrete" function.

It is natural to initially assume that the edges are to be taken to edges, by means of some function:

$$f_E : E_G \to E_J.$$

Such an assumption doesn't seem to cause any problems; in fact, it matches our basic interpretation of graphs as combinations of "agents" and the relations between them. We simply assign an agent to an agent and a relation to a relation.

However, what about a constant function? What is the discrete counterpart of the function

$$f : X \to Y,\ f(x) = y_0 \in Y,\ \forall x \in X?$$

An easy answer is to allow an edge to be taken to a node, the event that we will call a "collapse":

$$f(AB) = X \in N_J,\ \forall AB \in E_G.$$

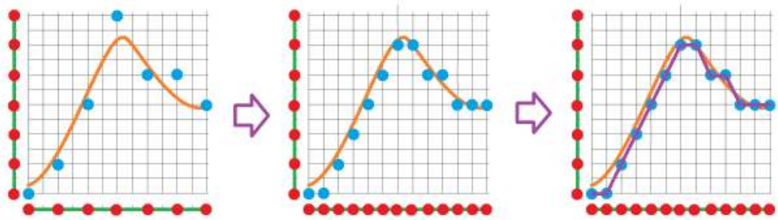
We just learned that a function $f : G \to J$ between graphs should be a combination of two, the node function:

$$f_N : N_G \to N_J.$$

and the edge function:

$$f_E : E_G \to E_J \cup N_J.$$

Finding a specific discrete representation, or a counterpart, of a given continuous transformation may require approximation and even refining the graph:



This issue is discussed at a later time and for now we will be content with an *analogy*.

5.2 Graph maps

We can show, however, that not every function of graphs as defined above is a valid discrete counterpart of a continuous function.

Example 5.3. Consider two single-edge graphs:

$$N_G = \{A, B\},\ E_G = \{AB\},\ N_J = \{X, Y\},\ E_J = \{XY\}.$$

What functions can we define between them?

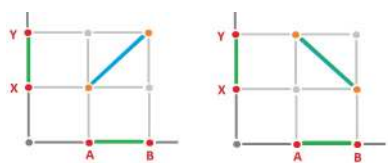
Let's we assume that

$$f_E(AB) = XY.$$

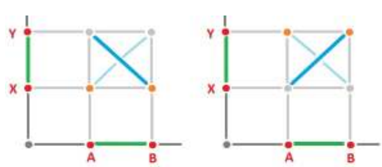
Then we can have several possible values for A, B:

- 1. $f_N(A) = X,\ f_N(B) = Y$,
- 2. $f_N(A) = Y,\ f_N(B) = X$,
- 3. $f_N(A) = X,\ f_N(B) = X$,
- 4. $f_N(A) = Y,\ f_N(B) = Y$.

The first two options make sense because they preserve the relation between the agents. This is not the case for options 3 and 4! By looking at the *graphs* of these functions (on graphs), we realize that what we accept here is *continuity*:



and what we reject is the *discontinuity*:



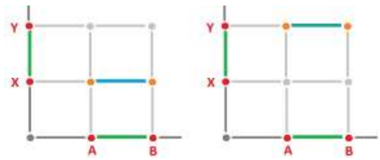
□

Example 5.4. Let's start over with the assumption:

- $f_N(A) = X, f_N(B) = X$.

The only way to make sense of these is to define the value of the edge to be a *node*:

- $f_E(AB) = X$.



And for

- $f_N(A) = Y, f_N(B) = Y$,

we set

- $f_E(AB) = Y$.

Then we simply have two constant functions here. □

What we have learned is that the values of the function on the nodes dictate the values on the edges, and vice versa.

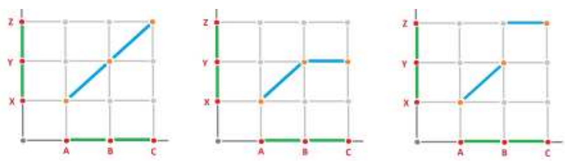
Keeping in mind that we are after a discrete analog of continuous functions, let's consider another example.

Example 5.5. Given two two-edge graphs:

$$N_G = \{A, B, C\},\ E_G = \{AB, BC\},\ N_J = \{X, Y, Z\},\ E_J = \{XY, YZ\}.$$

What functions can we define between them?

Consider just these three possibilities:



They are given by these functions:

- (1) $f_N(A) = X,\ f_N(B) = Y,\ f_N(C) = Z,\ f_E(AB) = XY,\ f_E(BC) = YZ$,
- (2) $f_N(A) = X,\ f_N(B) = Y,\ f_N(C) = Y,\ f_E(AB) = XY,\ f_E(BC) = Y$,
- (3) $f_N(A) = X,\ f_N(B) = Y,\ f_N(C) = Z,\ f_E(AB) = XY,\ f_E(BC) = Z$. □

Even a casual look at the graphs reveals the difference: the presence or absence of continuity. In fact, these three graphs look like they could have come from a calculus textbook as illustrations of the issue of continuity vs. discontinuity:

$$\lim_{x \to B-} f(x) = f(B) = \lim_{x \to B+} f(x).$$

We need to ensure this kind of continuity. We plot the nodes first and then attach the edges to them. If we discover that this is impossible, no realization of the function can be continuous and it should be discarded.

Therefore, we require from the edge function f_E the following:

- for each edge e, f_E takes its endpoints to the endpoints of $f_E(e)$.

Or, in a more compact form,

$$f_N(A) = X,\ f_N(B) = Y \iff f_E(AB) = XY.$$

We say that, in this case, the edge is *cloned*.

But what about the *collapsed* edges? It's easy; we just require:

- for each edge e, if e is taken to node X, then so are its endpoints, and vice versa.

Or, in a more compact form,

$$f_N(A) = f_N(B) = X \iff f_E(AB) = X.$$

Definition 5.6. A function of graphs $f : G \to J$,

$$f = \Big\{ f_N : \ N_G \to N_J, \ f_E : E_G \to E_J \cup N_J \Big\},$$

is called a *graph map* if

$$f_E(AB) = \begin{cases} f_N(A)f_N(B) & \text{if } f_N(A) \neq f_N(B), \\ f_N(A) & \text{if } f_N(A) = f_N(B). \end{cases}$$

In the most abbreviated form, this *discrete continuity condition* is:

$$f(AB) = f(A)f(B),$$

with the understanding that $XX = X$.

With this property assumed, we only need to provide f_N, and f_E will be derived.

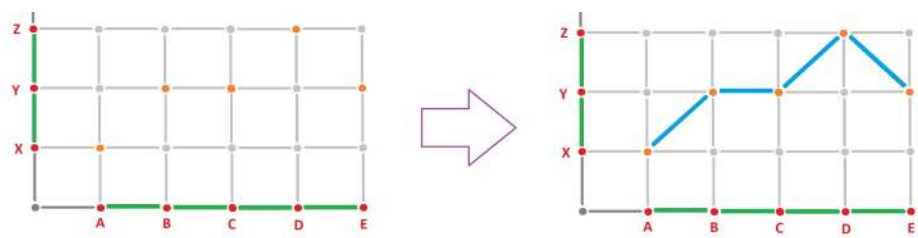
This idea also allows us to realize graph maps as continuous functions.

Example 5.7. Suppose two graphs G, J are realized as *intervals*:

$$|G| = [A, E] \subset \mathbf{R}, \ |J| = [X, Z] \subset \mathbf{R}.$$

Suppose $f : G \to J$ is a graph map, then we are looking for a continuous function

$$|f| : [A, E] \to [X, Z].$$



We will need the node function f_N only. Suppose

$$f_N(A) = X, \ f_N(B) = Y, \ f_N(C) = Y, \ f_N(D) = Z, \ f_N(E) = Y.$$

Then, naturally, we set

$$|f|(A) = X, \ |f|(B) = Y, \ |f|(C) = Y, \ |f|(D) = Z, \ |f|(E) = Y.$$

Now we can see how easy it is to reconstruct the rest of the function. We just need to connect these points and, since all we need is continuity, we connect them by *straight segments*. The formulas are familiar:

$$|f|(tA + (1-t)B) = tX + (1-t)Y, \ t \in [0, 1].$$

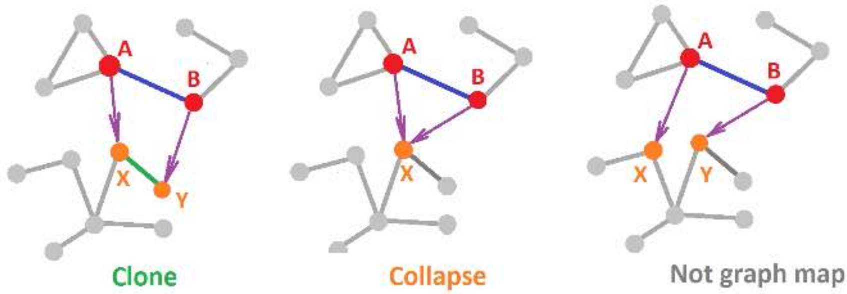
It's a bit trickier for the next edge:

$$|f|(tB + (1-t)C) = tY + (1-t)Y = Y, \ t \in [0, 1],$$

but it still works! We can handle collapses too. □

Exercise 5.8. Define a realization of a graph map and prove the continuity of $|f|$.

Below we illustrate these ideas for general graphs. Here are two *continuous* possibilities for a discrete function and one *discontinuous*:



Exercise 5.9. Prove that the composition of two graph maps is a graph map.

Exercise 5.10. Under what circumstances is there the inverse of a graph map which is also a graph map?

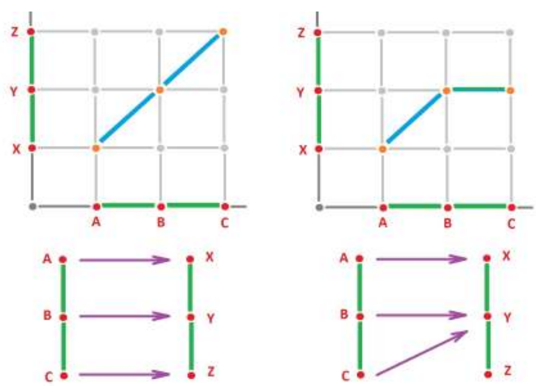
5.3 Chain maps

We can now move on to algebra.

Example 5.11. We consider the maps of the two two-edge graphs given above:

$$N_G = \{A, B, C\}, \quad E_G = \{AB, BC\},$$
$$N_J = \{X, Y, Z\}, \quad E_J = \{XY, YZ\};$$

and the two graph maps.



The key idea is to think of graph maps as *functions on the generators of the chain groups*:

$$C_0(G) =< A, B, C >, \quad C_1(G) =< AB, BC >,$$
$$C_0(J) =< X, Y, Z >, \quad C_1(J) =< XY, YZ > .$$

Let's express the values of the generators in G under f in terms of the generators in J.

The first function is given by

$$f_N(A) = X, \qquad f_N(B) = Y, \qquad f_N(C) = Z,$$
$$f_E(AB) = XY, \quad f_E(BC) = YZ.$$

It is then written coordinate-wise as follows:

$$f_0\Big([1,0,0]^T\Big) = [1,0,0]^T, \quad f_0\Big([0,1,0]^T\Big) = [0,1,0]^T, \quad f_0\Big([0,0,1]^T\Big) = [0,0,1]^T,$$
$$f_1\Big([1,0]^T\Big) = [1,0]^T, \qquad f_1\Big([0,1]^T\Big) = [0,1]^T.$$

The second is given by

$$\begin{array}{lll} f_N(A) = X, & f_N(B) = Y, & f_N(C) = Y, \\ f_E(AB) = XY, & f_E(BC) = Y. & \end{array}$$

It is written coordinate-wise as follows:

$$\begin{array}{lll} f_0\Big([1,0,0]^T\Big) = [1,0,0]^T, & f_0\Big([0,1,0]^T\Big) = [0,1,0]^T, & f_0\Big([0,0,1]^T\Big) = [0,0,1]^T, \\ f_1\Big([1,0]^T\Big) = [1,0]^T, & f_1\Big([0,1]^T\Big) = 0. & \end{array}$$

The very last item requires special attention: the collapsing of an edge in G does not produce a corresponding edge in J. This is why we make an *algebraic decision* to assign it the zero value.

As always, it suffices to know the values of a function on the generators to recover the whole homomorphism. These two functions f_N and f_E *generate* two homomorphisms:

$$f_0 : C_0(G) \to C_0(J),\ f_1 : C_1(G) \to C_1(J).$$

It follows that the first homomorphism is the identity and the second can be thought of as a projection. □

Exercise 5.12. Prove the last statement.

Exercise 5.13. Find the matrices of f_0, f_1 in the last example.

Exercise 5.14. Find the matrices of f_0, f_1 for $f : G \to J$ given by $f_E(AB) = XY,\ f_E(BC) = XY$.

Let's take another look at the "discrete continuity conditions" from the last subsection. No matter how compact these conditions are, one is forced to explain that the collapsed case appears to be an exception to the general case. To get around this inconvenience, let's find out what happens under f to the *boundary operator* of the graph.

Example 5.15. We have two here:

$$\begin{array}{ll} \partial_G : & C_1(G) \to C_0(G) \\ \partial_J : & C_1(J) \to C_0(J), \end{array}$$

one for each graph.

Now, we just use the fact that

$$\partial(AB) = A + B,$$

and the continuity condition takes this form:

$$\partial(f_1(AB)) = f_0(\partial(AB)).$$

It applies, without change, to the case of a collapsed edge. Indeed, if AB collapses to X, both sides are 0:

$$\begin{array}{lll} \partial(f_1(AB)) & = \partial(0) & = 0; \\ f_0(\partial(AB)) & = f_0(A+B) = f_0(A) + f_0(B) = X + X & = 0. \end{array}$$ □

We follow this idea and introduce a new concept.

Definition 5.16. The *chain map* generated by a graph map $f : G \to J$ is a pair of homomorphisms $f_\Delta := \{f_0, f_1\}$:

$$\begin{array}{l} f_0 : C_0(G) \to C_0(J), \\ f_1 : C_1(G) \to C_1(J), \end{array}$$

generated by f_N and f_E respectively.

By design, these two homomorphisms satisfy the following.

Theorem 5.17 (Algebraic Continuity Condition). For any graph map $f : G \to J$, its chain map satisfies:

$$\partial_J f_1(e) = f_0(\partial_G e),$$

for any edge e in G.

Whenever the boundary operator can be properly defined, we can use this condition in any dimension and not just for graphs.

Exercise 5.18. Suppose a graph map collapses all edges. Find its chain map.

Exercise 5.19. Find the matrix of the chain map of a graph map that shifts by one edge a graph of n edges arranged in (a) a string, (b) a circle.

Exercise 5.20. Find the matrix of the chain map of a graph map that *folds in half* a graph of $2n$ edges arranged in (a) a string, (b) a circle, (c) a figure eight; also (d) for a figure eight with $4n$ edges – the other fold.

Exercise 5.21. Find the matrix of the chain map of a graph map that *flips* a graph of $2n$ edges arranged in (a) a string, (b) a circle, (c) a figure eight; also (d) for a figure eight with $4n$ edges – the other flip.

Below are the main theorems of this theory.

Theorem 5.22. Transitioning to chain maps preserves the identity:

$$(\mathrm{Id}_G)_0 = \mathrm{Id}_{C_0(G)},\ (\mathrm{Id}_G)_1 = \mathrm{Id}_{C_1(G)}.$$

Theorem 5.23. Transitioning to chain maps preserves the compositions:

$$(fg)_0 = f_0 g_0,\ (fg)_1 = f_1 g_1.$$

Theorem 5.24. Transitioning to chain maps preserves the inverses:

$$(f^{-1})_0 = f_0^{-1},\ (f^{-1})_1 = f_1^{-1}.$$

Exercise 5.25. Prove these three theorems.

A wonderfully compact form of the above condition is below:

$$\boxed{\partial f = f \partial}$$

We will see this identity a lot. An appropriate way to describe what happens here is to say that the chain maps and the boundary operators *commute*. The idea is visualized with so-called "commutative diagrams".

5.4 Commutative diagrams

This very fruitful approach will be used throughout.

Compositions of functions can be visualized as flowcharts:

In general, we represent a function $f : X \to Y$ diagrammatically as a black box that takes an input and releases an output (same y for same x):

$$\begin{array}{ccccc} \text{input} & & \text{function} & & \text{output} \\ x & \longrightarrow & \boxed{\quad f \quad} & \longrightarrow & y \end{array}$$

or, simply,

$$X \xrightarrow{f} Y.$$

Suppose now that we have another function $g : Y \to Z$; how do we represent their composition $fg = g \circ f$?

To compute it, we "wire" their diagrams together *consecutively*:

$$x \longrightarrow \boxed{\quad f \quad} \longrightarrow y \longrightarrow \boxed{\quad g \quad} \longrightarrow z$$

The standard notation is the following:

$$X \xrightarrow{f} Y \xrightarrow{g} Z.$$

Or, alternatively, we may want to emphasize the resulting composition:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {}_{gf} \searrow & \downarrow g \\ & & Z \end{array}$$

We say that the new function "completes the diagram".

The point illustrated by the diagram is that, starting with $x \in X$, you can

- go right then down, or
- go diagonally;

and either way, you get the *same result*:

$$g(f(x)) = (gf)(x).$$

In the diagram, this is how the values of the functions are related:

$$\begin{array}{ccccc} x \in & X & \xrightarrow{f} & Y & \ni f(x) \\ & & {}_{gf} \searrow & \downarrow g & \\ & (gf)(x) \in & & Z & \ni g(f(x)) \end{array}$$

Example 5.26. As an example, we can use this idea to represent the *inverse* function f^{-1} of f. It is the function that completes the diagram with the identity function on X:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {}_{\mathrm{Id}_X} \searrow & \downarrow f^{-1} \\ & & X \end{array}$$

□

Exercise 5.27. Plot the other diagram for this example.

Example 5.28. The *restriction* of $f : X \to Y$ to subset $A \subset X$ completes this diagram with the inclusion of A into X:

$$\begin{array}{ccc} A & \hookrightarrow & X \\ & {\scriptstyle f|_A} \searrow & \downarrow {\scriptstyle f} \\ & & Y \end{array}$$

□

Diagrams like these are used to represent compositions of all kinds of functions: continuous functions, graph functions, homomorphisms, linear operators, and many more.

Exercise 5.29. Complete the diagram:

$$\begin{array}{ccc} \ker f & \hookrightarrow & X \\ & {\scriptstyle ?} \searrow & \downarrow {\scriptstyle f} \\ & & Y \end{array}$$

A commutative diagram may be of any shape. For example, consider this square:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow {\scriptstyle g'} & \searrow & \downarrow {\scriptstyle g} \\ X' & \xrightarrow{f'} & Y' \end{array}$$

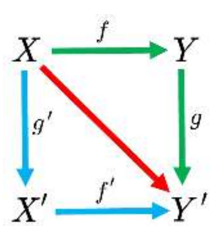
As before, go right then down, or go down then right, with the same result:

$$gf = f'g'.$$

Both give you the function of the diagonal arrow!

This identity is the reason why it makes sense to call such a diagram "commutative". To put it differently,

vertical then horizontal is same as horizontal then vertical.



The illustration above explains how the blue and green threads are tied together in the beginning – as we start with the same x in the left upper corner – and at the end – where the output of these compositions in the right bottom corner turns out to be the same. It is as if the commutativity turns this combination of loose threads into a *piece of fabric*!

The algebraic continuity condition in the last subsection

$$\partial_J f_1(e) = f_0(\partial_G e)$$

is also represented as a commutative diagram:

$$\begin{array}{ccc} C_1(G) & \xrightarrow{f_1} & C_1(J) \\ \Big\downarrow \partial_G & \searrow & \Big\downarrow \partial_J \\ C_0(G) & \xrightarrow{f_0} & C_0(J). \end{array}$$

Exercise 5.30. Draw a commutative diagram for the composition of two chain maps.

Exercise 5.31. Restate in terms of commutative diagrams the last two theorems in the last subsection.

5.5 Cycles and boundaries under chain maps

Ultimately, our interest is to track the topological changes as one graph is mapped to another. Given a graph map $f : G \to J$, the main two questions are about the two topological features we have been studying:

• 1. What is the counterpart in J of each *component* of G under f?

• 2. What is the counterpart in J of each *hole* of G under f?

Unfortunately, these questions are *imprecise*.

As we know, the topological features are fully captured by the homology groups of these graphs so that we can rephrase these questions as follows:

• 1. How does f transform the 0th homology group $H_0(G)$ (components) of G into the 0th homology group $H_0(J)$ of J?

• 2. How does f transform the 1st homology group $H_1(G)$ (cycles) of G into the 1st homology group $H_1(J)$ of J?

To answer these questions we need to conduct a further algebraic analysis of f.

Let's review. First, f is a pair of functions:

• 1. $f_N : N_G \to N_J$, and

• 2. $f_E : E_G \to E_J \cup N_J$.

These two generate the chain map f_Δ of f, which is a pair of homomorphisms:

• 1. $f_0 : C_0(G) \to C_0(J)$, and

• 2. $f_1 : C_1(G) \to C_1(J)$.

These two "commute" with the boundary operator:

$$f_0\partial = \partial f_1.$$

These homomorphisms reveal what happens to the cycles. First, we have no interest in the 1-chains that aren't 1-cycles. That's why we consider the *restriction* of f_1 to the group of cycles $Z_1(G)$:

$$f_1\Big|_{Z_1(G)} : Z_1(G) \to C_1(J);$$

it has the same values as the original. Now a crucial step.

Theorem 5.32. f_1 takes cycles to cycles.

Proof. Suppose $\partial_G(x) = 0$, then by the commutativity property above, we have So, we've found the quotient with the generator explicitly presented:

$$\partial_J(f_1(x)) = f_0(\partial_G(x)) = f_0(0) = 0. \qquad \blacksquare$$

Remember, zero is also a cycle...

So, not only the restriction makes sense but also the values of this new function lie in $Z_1(J)$. Reusing "f_1" for this new function, we state this fact as follows.

Corollary 5.33. The map of cycles $f_1 : Z_1(G) \to Z_1(J)$ is well-defined.

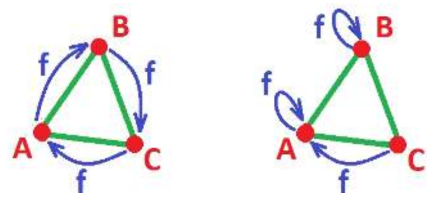
Example 5.34. Consider the graph in the form of a *triangle*:

$$N_G = N_J = \{A, B, C\},\ E_G = E_J = \{AB, BC, CA\};$$

and suppose f is this *rotation*:

$$f_N(A) = B,\ f_N(B) = C,\ f_N(C) = A.$$

Its realization is on the left:



A quick computation shows that

$$Z_1(G) = Z_1(J) =< AB + BC + CA > .$$

Now,

$$\begin{aligned} f_1(AB + BC + CA) &= f_1(AB) + f_1(BC) + f_1(CA) \\ &= BC + CA + AB = AB + BC + CA. \end{aligned}$$

So, $f_1 : Z_1(G) \to Z_1(J)$ is the identity (even though f isn't). Conclusion: *the hole is preserved.*

Now, the other realization is that of the *collapse of the triangle* onto one of its edges given by

$$f_N(A) = A,\ f_N(B) = B,\ f_N(C) = A.$$

Then

$$\begin{aligned} f_1(AB + BC + CA) &= f_1(AB) + f_1(BC) + f_1(CA) \\ &= AB + BA + 0 = 0. \end{aligned}$$

So, the homomorphism is zero. Conclusion: *the hole collapses.* □

Exercise 5.35. Carry out the "quick computation".

Exercise 5.36. Modify the computations for another rotation and another collapse.

Next, the components. This issue is more complicated because the 0th homology groups of G and J are both quotients. We already have a homomorphism

$$f_0 : C_0(G) \to C_0(J),$$

with the commutativity property, and we need to define somehow another homomorphism

$$? : H_0(G) = \frac{C_0(G)}{B_0(G)} \to H_0(J) = \frac{C_0(J)}{B_0(J)}.$$

The challenge is that a map on homology groups would have to handle *classes* of chains. Fortunately, the problem is fully resolved in group theory. We review quotient maps next.

5.6 Quotient maps in algebra

Suppose we have two groups A, B, and pairs of their subgroups

$$A'' \subset A' \subset A, \; B'' \subset B' \subset B.$$

Suppose also that there is a homomorphism between these two groups:

$$F : A \to B.$$

Then the only reasonable way to define the *quotient map*

$$[F] : A'/A'' \to B'/B''$$

of F is by setting its value for each equivalence class $[x]$ of x to be the equivalence class of $F(x)$; i.e.,

$$[F]([x]) := [F(x)], \; \forall x \in A'.$$

But is this function well-defined?

What can possibly go wrong here?

First, the new function might miss its target: $F(x) \notin B'$. We have to require that such a restriction of F to A' makes sense; i.e.,

- (1) $F(A') \subset B'$.

Second, the formula might produce a different result if we choose another representative of the equivalence class $[x]$ of x. We want to ensure that this doesn't happen:

$$[y] = [x] \Longrightarrow [F]([y]) = [F]([x]),$$

or

$$y \sim x \Longrightarrow F(y) \sim F(x).$$

Recall the definition of the quotient group modulo X:

$$a \sim b \mod (X) \Longleftrightarrow a - b \in X.$$

So, we want to ensure that

$$x - y \in A'' \Longrightarrow F(x) - F(y) \in B'',$$

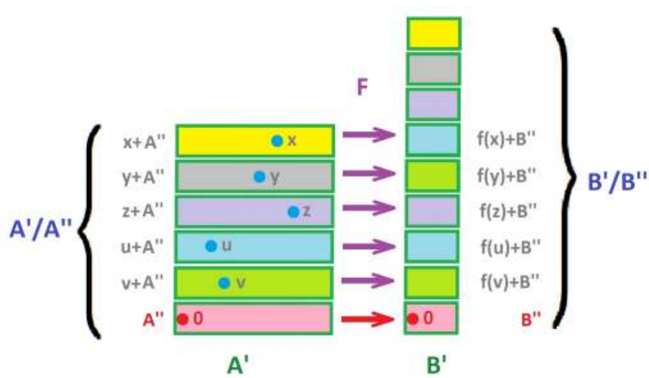
or

$$x - y \in A'' \Longrightarrow F(x - y) \in B''.$$

Thus, it suffices to ask, once again, that the restriction of F to A'' makes sense; i.e.,

- (2) $F(A'') \subset B''$.

This constraint isn't surprising. After all, the new homomorphism is supposed to take the zero of the first quotient group, which is $A'' = 0 \in A'/A''$, to the zero of the second quotient, which is $B'' = 0 \in B'/B''$. The idea is illustrated below:



Example 5.37. To illustrate condition (2), consider linear operators $f : \mathbf{R} \to \mathbf{R}^3$ and the quotient $\mathbf{R}^3/\mathbf{R}^2$. These are the equivalence classes:

Which operators have well-defined quotient operators with respect to this quotient? The reflections do, but the rotations do not, unless the axis of rotation is the z-axis. □

Exercise 5.38. What about the stretches?

To summarize, the quotient map is well-defined provided the conditions (1) and (2) are satisfied. These two conditions can be conveniently stated as, F is *a function of pairs*; i.e.,

$$F : (A', A'') \to (B', B'').$$

Exercise 5.39. Find all homomorphisms $f : \mathbf{Z}_2 \oplus \mathbf{Z}_2 \to \mathbf{Z}_2 \oplus \mathbf{Z}_2$ such that their quotient maps,

$$[f] : \Big(\mathbf{Z}_2 \oplus \mathbf{Z}_2\Big)/ < (1,1) > \to \Big(\mathbf{Z}_2 \oplus \mathbf{Z}_2\Big)/ < (1,0) >,$$

are well-defined.

Exercise 5.40. Do the same for homomorphisms f of integers and their quotients:

$$[f] : \mathbf{Z}/ < n > \to \mathbf{Z}/ < m > .$$

5.7 Homology maps

As we examine the two requirements in the last subsection in our particular situation, we realize that for maps of components we only need the latter:

$$f_0(B_0(G)) \subset B_0(J).$$

Theorem 5.41. The 0th chain map f_0 of a graph map f takes boundaries to boundaries.

Proof. Suppose $x = \partial_G(y)$, then by the commutativity formula above, we have

$$f_0(x) = f_0(\partial_G(y)) = \partial_J(f_1(y)).$$ ■

Recall that $[A]$ stands for the homology class of vertex (or a 0-cycle) A that consists of all 0-cycles homologous to A.

Corollary 5.42. The *map of components* $[f_0] : H_0(G) \to H_0(J)$ given by

$$[f_0]([A]) = [f_0(A)]$$

is well-defined.

Proof. If $x \in B_0(G)$ then $f_0(x) \in B_0(J)$. Hence, $f_0(B_0(G)) \subset B_0(J)$. Further,

$$[f_0]([A]) = [f_0](A + B_0(G)) = [f_0](A) + B_0(J) = [f_0(A)].$$ ■

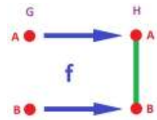
Example 5.43. Consider this *two-node graph* and this *single-edge graph*:

$$N_G = N_J = \{A, B\},\ E_G = \emptyset,\ E_J = \{AB\},$$

Merging of components is easy to illustrate with the following map between them:

$$f_N(A) = A, \ f_N(B) = B.$$

This is its realization:



It follows that

$$\begin{array}{l} C_0(G) = H_0(G) =< [A], [B] >, \\ C_0(J) =< A, B >, \ H_0(J) = \{[A] = [B]\}, \\ f_0([A]) = f_0([B]) = [A]. \end{array}$$

□

We now combine the maps of components with the map of cycles from the last subsection.

Definition 5.44. The pair $f_* := \{[f_0], f_1\}$ is called the *homology map* of the graph map f.

Exercise 5.45. Find the matrices of the homology map of a graph map that folds in half a graph of $2n$ edges arranged in (a) a string, (b) a circle, (c) a figure eight.

Exercise 5.46. Devise a graph map that "splits a hole" and then conduct a homological analysis to confirm this.

The following three theorems give us the main features of *homology theory*.

Theorem 5.47. Transitioning to homology maps preserves the identity:

$$\left[(\mathrm{Id}_G)_0\right] = \mathrm{Id}_{H_0(G)}, \ \left(\mathrm{Id}_G\right)_1 = \mathrm{Id}_{H_1(G)}.$$

Theorem 5.48. Transitioning to homology maps preserves the compositions:

$$\left[(fg)_0\right] = [f_0][g_0], \ \left(fg\right)_1 = f_1 g_1.$$

Theorem 5.49. Transitioning to homology maps preserves the inverses:

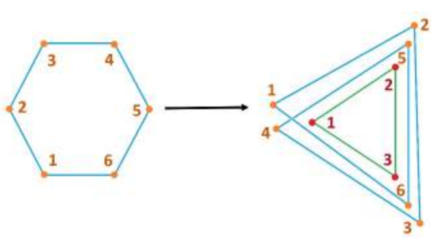
$$\left[(f^{-1})_0\right] = [f_0]^{-1}, \ \left(f^{-1}\right)_1 = f_1^{-1}.$$

Exercise 5.50. Prove these three theorems.

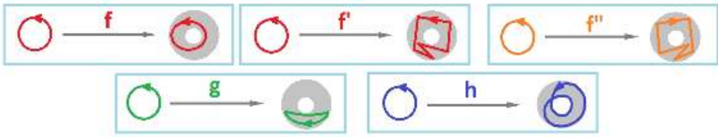
Example 5.51 (self-maps of circle). Let's try to classify the continuous functions of circle to circle. The best we can do, for now, is to think of the circles as realizations of two circular graphs G, J made of n, m nodes and n, m edges respectively:

$$\begin{array}{ll} N_G = \{A_0, A_1, ..., A_n = A_0\}, & N_J = \{B_0, B_1, ..., B_m = B_0\}, \\ E_G = \{A_0A_1, ..., A_{n-1}A_n\}, & E_J = \{B_0B_1, ..., B_{m-1}B_m\}, \end{array}$$

and then classify the graph maps $f : G \to J$. We have to allow all possible values of $n \geq m$ in order to be able to reproduce multiple wraps; one needs $n = 2m$ for a double wrap.



The outcome we might hope for is the one that reveals the difference of a loop that goes around 1 time from one that goes around 2 times from one that goes around 3 times, etc., and from those that go in the opposite direction:



For these functions, the numbers of turns are: $1, 1, -1, 0, 2$. Meanwhile, homology theory as it has been so far developed produces results that fall short of our goal. The homology groups of the graphs are very simple:

$$H_0(G) =< [A_0] >, \qquad H_0(J) =< [B_0] >,$$
$$H_1(G) =< A_0A_1 + ... + A_{n-1}A_n >, \quad H_1(J) =< B_0B_1 + ... + B_{m-1}B_m > .$$

Both of these 1st homology groups are isomorphic to $\mathbf{Z}_2$; therefore, there can be only two possible homology maps: the zero and the isomorphism. In the former case, the graph map doesn't make a full turn or makes an even number of turns and in the latter case, it makes an odd number of turns, clockwise or counterclockwise. In binary arithmetic, the numbers of turns for the above functions become: $1, 1, 1, 0, 0$.

To be able to count the turns, we'd have to use *directed edges* and the *integer arithmetic*. This development is presented in the later chapters. □

Exercise 5.52. Provide the missing details of the example: compute the homology groups of the graphs, represent the depicted functions as realizations of graph maps, compute their homology maps, etc.

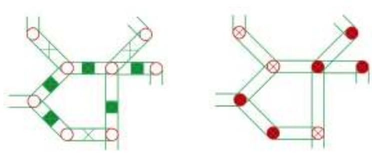
---

6 Binary calculus on graphs

6.1 The algebra of plumbing, continued

We introduce more algebra with the familiar metaphor.

We think of a graph as a plumbing system that consists of a network of pipes and joints, and each joint and each pipe is equipped with an on-off switch.



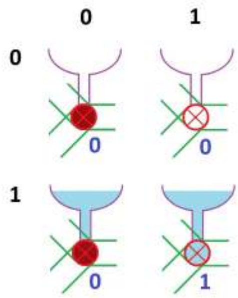
The two plumbers "flip the switch" on a joint or a pipe which is recorded with binary numbers and the repeated flips are cancelled via *binary addition*. Since this activity is happening at every joint and every pipe, we have the two groups: the chains of nodes and the chains of edges:

$$C_0 \cong (\mathbf{Z}_2)^n, \; C_1 \cong (\mathbf{Z}_2)^m.$$

Let's concentrate on a single switch. We assume that it is closed in the beginning of the day. Then the combined request will tell us the state of each switch in the end of the day. We use 0 for closed and 1 for open.

We add water now.

Behind the switch is a reservoir. We look at:
- the state of the reservoir (full or empty),
- the request for the switch, and
- the resulting flow of water.



All of these are binary numbers, 0 or 1. How this works is obvious:
- pump on (1) and switch flipped (1), then water flows (1);
- pump off (0) and switch flipped (1), then water doesn't flow (0); etc.

We discover that this isn't binary addition but *binary multiplication*:

$$1 \times 1 = 1,\ 0 \times 1 = 0,\ 1 \times 0 = 0,\ 0 \times 0 = 0.$$

It is important that the two numbers in the left-hand side refer to two very different entities and it would be misleading to suggest that these are group operations. Instead, let's rename the states of the reservoir as 0^* and 1^*. Then we have:

$$1^* \times 1 = 1,\ 0^* \times 1 = 0,\ 1^* \times 0 = 0,\ 0^* \times 0 = 0.$$

In fact, the best way to think of these two numbers is as *functions*:

$$1^*(1) = 1,\ 0^*(1) = 0,\ 1^*(0) = 0,\ 0^*(0) = 0.$$

Furthermore, the distributive property proves that these are homomorphisms

$$x^* : \mathbf{Z}_2 \to \mathbf{Z}_2.$$

These simple functions, 0^* and 1^*, are defined for every switch and, therefore, for all chains; these functions are called *cochains*.

6.2 The dual of a group

What follows is the algebra we need in order to understand these binary cochains.

Given a group L, its *dual* L^* (over $\mathbf{Z}_2$) is the set of all homomorphisms on L,

$$s : L \to \mathbf{Z}_2.$$

Let's start with this simplest group:

$$L = \{0, A\} \cong \mathbf{Z}_2,$$

with $A + A = 0$, etc. There are only two such homomorphisms: the zero and the identity. The former may be thought of as the multiplication by 0 and the latter by 1. We can denote them by 0^* and 1^*, just as in the last subsection. Then,

$$L^* := \{0^*, 1^*\}.$$

In other words, we see the following again:

$$0^*(0) = 0,\ 0^*(1) = 0,\ 1^*(0) = 0,\ 1^*(0) = 1.$$

A more general case is:

$$L := \underbrace{\mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus ... \oplus \mathbf{Z}_2}_{\text{n times}} = \Big(\mathbf{Z}_2\Big)^n.$$

Every element of L is written as

$$x = (x_1, ..., x_n),\ x_i \in \mathbf{Z}_2.$$

Then every element of L^* is written as

$$x = (x_1^*, ..., x_n^*),\ x_i \in \mathbf{Z}_2.$$

Therefore,

$$L^* \cong \Big(\mathbf{Z}_2\Big)^n.$$

Since the sum of two homomorphisms is a homomorphism, we make the following conclusion.

Proposition 6.1. The dual of $\mathbf{Z}_2$ is a group isomorphic to $\mathbf{Z}_2$.

So, $L \cong L^*$! However, this doesn't mean that those two are the same thing.

The first crucial difference is that we can *multiply* the elements of the latter, because the product of two homomorphisms to a given group is also a homomorphism. We make the following conclusion.

Proposition 6.2. The dual L^* is a ring.

Exercise 6.3. Prove the proposition.

Exercise 6.4. Define and find the dual over $\mathbf{Z}_2$ of the group $\mathbf{Z}_3$.

Further, what happens to the dual L^* under homomorphisms of L?

As it turns out, this is a wrong question to ask. Suppose there is another group K and suppose $h : L \to K$ is a homomorphism. Then, for any $s \in L^*$, what happens is seen in the following diagrams:

$$\begin{array}{ccccccc}
L & \xrightarrow{s} & \mathbf{Z}_2 & \qquad & A & \mapsto & 0 \text{ or } 1 \\
\downarrow{h} & & || & & \downarrow{h} & & ||\ ? \\
K & \xrightarrow{t} & \mathbf{Z}_2 & & h(A) & \mapsto & 0 \text{ or } 1
\end{array}$$

It would be natural to try to express the "new" function t in terms of the "old" s, but it's impossible to combine s with h. In fact, it is the opposite that we need to consider – th makes sense and ht doesn't. Then we simply require the first diagram to be commutative or, which is the same thing, that the outputs of the two homomorphisms coincide on the right in the second diagram. As a result, s is defined in terms of t:

$$s(A) := th(A).$$

Then $t \in K^*$ is the *input* and $s \in L^*$ is the *output* of the new homomorphism.

To summarize, the *dual homomorphism* of a homomorphism

$$h : L \to K$$

is the homomorphism (note the direction of the arrow)

$$h^* : K^* \to L^*,$$

defined by the identity:

$$h(y^*)(x) := y^* h(x),$$

for any $x \in L$ and any $y \in K$.

Exercise 6.5. For L and K chosen to be either $\mathbf{Z}_2$ or $\mathbf{Z}_2 \oplus \mathbf{Z}_2$, provide a formula for h^* for every possible h.

Exercise 6.6. How many homomorphisms $(\mathbf{Z}_2)^n \to \mathbf{Z}_2$ are there?

6.3 Cochains of nodes and cochains of edges

We apply the algebra developed in the last subsection to cochains.

Suppose we are given a graph G with the set of nodes N and the set of edges E. Then the group of chains of nodes, or 0-chains, and the group of chains of edges, or 1-chains, are respectively:

$$\begin{aligned} C_0 = C_0(G) &:= \Big\{ \sum_{A \in Q} A : A \subset N \Big\} \cup \{0\} &&=< N >, \\ C_1 = C_1(G) &:= \Big\{ \sum_{AB \in P} AB : P \subset E \Big\} \cup \{0\} &&=< E > . \end{aligned}$$

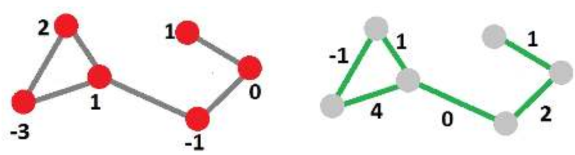
Definition 6.7. A *k-cochain* on graph G is a homomorphism $s : C_k(G) \to \mathbf{Z}_2$.

Since the sum of two homomorphisms is a homomorphism, we make the following conclusion.

Proposition 6.8. The 0- and 1-cochains on graph G form groups, **denoted** by $C^0(G)$ and $C^1(G)$ respectively (not to be confused with the set of differentiable functions).

If we consider one node (or edge) at a time, we are in the setting of the last subsection. As we know, there are only two such homomorphisms: the zero 0^* and the identity 1^*. Therefore,
- a 0-cochain assigns a number, 0 or 1, to each node, and
- a 1-cochain assigns a number, 0 or 1, to each edge.

Just like chains!



There is a special way to present the cochains in terms of the "basic chains", i.e., the nodes and the edges. For every such chain $x \in C_k,\ k = 0, 1$, define a "basic cochain" $x^* : C_k \to \mathbf{Z}_2$, by

$$x^*(t) := \begin{cases} 1 & \text{if } t = x, \\ 0 & \text{if } t \neq x. \end{cases}$$

Then every cochain is simply the sum of a set of the basic cochains.

Proposition 6.9.

$$\begin{aligned} C^0 &:= \Big\{ \sum_{A^* \in Q} A : A \subset N \Big\} \cup \{0\} &\cong \mathbf{Z}_2, \\ C^1 &:= \Big\{ \sum_{AB^* \in P} AB : P \subset E \Big\} \cup \{0\} &\cong \mathbf{Z}_2. \end{aligned}$$

Proposition 6.10. The function $x \mapsto x^*$ generates an isomorphism between the groups of chains and cochains:

$$C_k \cong C^k.$$

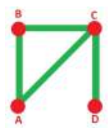
Since everything is expressed in terms of these generators, this isomorphism is given explicitly by

$$\Big(\sum_i A_i\Big)^* := \sum_i A_i^*.$$

Exercise 6.11. Prove the propositions.

Next, with ordered sets of nodes N and edges E fixed as bases of the groups of chains, both chains and cochains can be written coordinate-wise, as *column-vectors* and *row-vectors*.

Example 6.12. Let's once again consider this graph G:
- $N = \{A, B, C, D\}$,
- $E = \{AB, BC, CA, CD\}$.



Then its chains of nodes are

$$A = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ B = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ ..., \ A + B = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \ ...$$

Then the cochains of nodes are:

$$A^* = [1, 0, 0, 0], \ B^* = [0, 1, 0, 0], \ ..., \ (A + B)^* = [1, 1, 0, 0], \ ...$$

Similarly, the cochains of edges are:

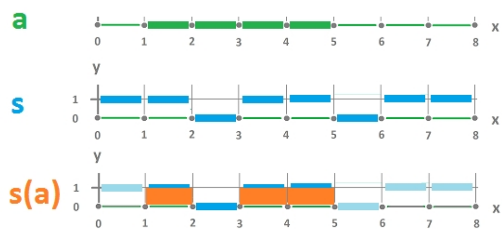
$$AB^* = [1, 0, 0, 0], \ BC^* = [0, 1, 0, 0], \ ..., \ (AB + BC)^* = [1, 1, 0, 0], \ ...$$ □

Proposition 6.13. The value of a cochain s at a given chain a is the dot-product of the two:

$$s(a) = \langle s, a \rangle.$$

We always put the cochain first in this dot product. This way, the output is simply the matrix product sa of a $1 \times n$ matrix s and an $n \times 1$ matrix a.

Example 6.14. To illustrate this formula, let's consider a very simple example:
- G, a graph with 8 edges numbered from 0 to 8,
- a, a chain equal to the sum of the edges from 1 to 5, and
- s, a cochain with 1s and 0's shown.



Then,

$$a=[0,1,1,1,1,1,0,0,0]^T,\ s=[1,1,0,1,1,0,1,1].$$

We compute:

$$\begin{array}{ll} s(a) & =\langle s,a\rangle=\sum_i s_ia_i\\ & =1\cdot 0+1\cdot 1+0\cdot 1+1\cdot 1+1\cdot 1+0\cdot 0+1\cdot 0+1\cdot 0\\ & =1\cdot 1+0\cdot 1+1\cdot 1+1\cdot 1=1.\end{array}$$

This is the (binary) area under the graph of s! (One can also imagine the plumber keeping a binary count of how many times he has flipped the switch.) □

The notation so justified is:

$$s(a)=\int_a s,\qquad \text{or } s(AB)=\int_A^B s.$$

Meanwhile, in this context, cochains are called *discrete differential forms*, to be discussed later.

Exercise 6.15. Verify the counterparts of the Linearity, Additivity, and other properties of the Riemann integral.

Exercise 6.16. Provide a similar, integral interpretation of 0-cochains.

6.4 Maps of cochains

Now, what happens to chains under graph maps is very simple. If $f:G\to J$ is such a map, its chain maps

$$f_k:C_k(G)\to C_k(J)$$

are defined as follows. If A is a node in G, then we have the following:

$$f_0(A)=f(A).$$

What about the cochains? We already know that the arrows will be reversed, in a "dual" way.

Suppose s is a k-cochain in G and t in J. Then the relation between them is seen in the following *commutative* diagram:

$$\begin{array}{ccc} C_k(G) & \xrightarrow{\ s\ } & \mathbf{Z}_2\\ \Big\downarrow{\scriptstyle f_k} & & ||\\ C_k(J) & \xrightarrow{\ t\ } & \mathbf{Z}_2\end{array}$$

Definition 6.17. The *k-cochain map*, $k=0,1,...$, of a graph map $f:G\to J$ are the homomorphisms (note the direction of the arrow):

$$f^k:C^k(J)\to C^k(G),$$

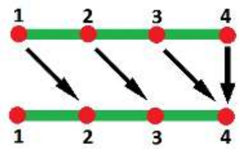
defined by

$$f^k(t)(a) := tf_k(a),$$

for any k-chain a in G and any k-cochain t in J.

Example 6.18. Let's consider the formula for a specific $f : G \to G$, the shift (with one collapse) of the three-edge graph G:

- $G := \{0, 1, 2, 3;\ 01, 12, 23\}$,
- $f(0) = 1,\ f(1) = 2,\ f(2) = 3,\ f(3) = 3.$



First, the nodes. We have

- $f_0(0) = 1,\ f_0(1) = 2,\ f_0(2) = 3,\ f_0(3) = 3.$

It suffices to compute the cochain map f^0 for

- all basic 0-chains of G, $a = p$, and
- all basic 0-cochains of G, $t = q^*$;

with $p, q = 0, ..., 3$. We use the formula:

$$f^0(q^*)(p) := q^* f_0(p),$$

with the latter equal to 0 (the number, not the node) unless $q = f_0(p)$. Then, $f^0(q^*)(p) = 0$ unless $q = p + 1$ or $q = p = 3$. The answer is:

- $f^0(0^*) = 0$ (i.e., it's trivial);
- $f^0(q^*) = (q - 1)^*,\ q = 1, 2;$
- $f^0(3^*) = 3^*.$

We next compare the matrices:

$$f_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad f^0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The second matrix is the transposes of the first!

Second, the edges. We have

- $f_1(01) = 12,\ f_1(12) = 23,\ f_1(23) = 0.$

It suffices to compute the cochain map f^1 for

- all basic 1-chains of G, $a = p$, and
- all basic 1-cochains of G, $t = q^*$;

with $p, q = 01, 12, 23$. We use the above formula:

$$f^1(q^*)(p) := q^* f_1(p),$$

which is equal to 0 unless $q = f_1(p) \neq 0$. The answer is:

- $f^1(01^*) = 0;$
- $f^1\big([m, m + 1]^*\big) = [m - 1, m]^*,\ m = 2, 3.$

Compare the matrices:

$$f_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad f^1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Once again, they are the transposes of each other. □

Exercise 6.19. Compute the cochain maps of other specific choices of f – shift, flip, fold – for the graph in the last example.

To make sense of the formula in the definition, let's use the integral notation above ($k = 1$):

$$\int_a f^1(t) = \int_{f_1(a)} t.$$

Then, if we set $a = AB$, we have

$$\int_A^B f^1(t) = \int_{f(A)}^{f(B)} t.$$

But this is just the formula of *integration by substitution*...

6.5 The coboundary operator

Recall that the boundary operator

$$\partial : C_1(G) \to C_0(G)$$

is set by its value for each edge:

$$\partial(AB) = A + B.$$

What happens to the cochains? Just as in the last subsection, we expect that the arrows will be reversed, in a "dual" way.

Suppose s is a 1-cochain in G and Q a 0-cochain. Then the relation between them is seen in the following commutative diagram, similar to the one in the last subsection:

$$\begin{array}{ccc}
C_1(G) & \xrightarrow{s} & \mathbf{Z}_2 \\
\downarrow \partial & & || \\
C_0(G) & \xrightarrow{Q} & \mathbf{Z}_2
\end{array}$$

Definition 6.20. The *coboundary operator* of a graph G is the homomorphism:

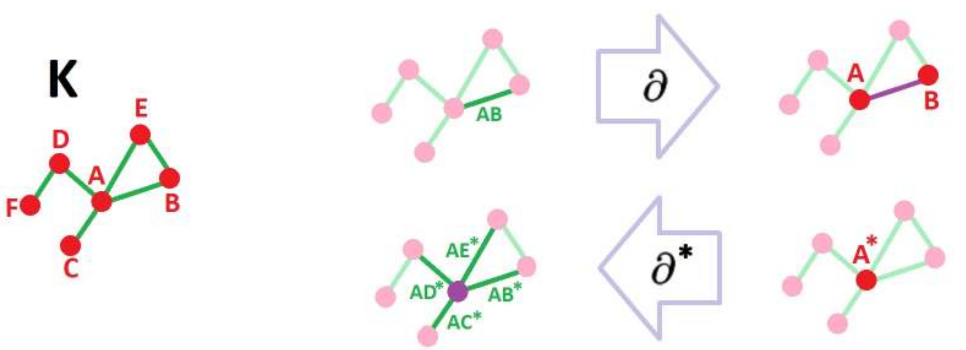
$$d = \partial^* : C^0(G) \to C^1(G),$$

defined by

$$d(Q)(a) := Q\partial(a),$$

for any 1-chain a and any 0-cochain Q in G.

Example 6.21. Here is an example:



Our claim is that

$$dA^* = AB^* + AC^* + AD^* + AE^*.$$

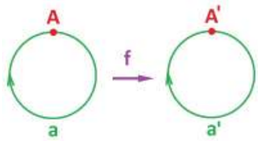
We prove it by evaluating dA^* for each of the four basic 1-chains:

$$dA^*(AB) = (AB^* + AC^* + AD^* + AE^*)(AB) = 1 + 0 + 0 + 0 = 1,$$

and, similarly:

$$dA^*(AC) = 1, \ dA^*(AD) = 1, \ dA^*(AE) = 1.$$

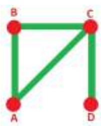
We can also use the definition with $Q = A^*$:



□

Exercise 6.22. Generalize this example.

Example 6.23. Consider the boundary operator for this graph one more time:



Then:

$$\begin{aligned}
\partial\Big([1,0,0,0]^T\Big) &= [1,1,0,0]^T, & d\Big([1,0,0,0]\Big) &= [1,0,1,0];\\
\partial\Big([0,1,0,0]^T\Big) &= [0,1,1,0]^T, & d\Big([0,1,0,0]\Big) &= [1,1,0,0];\\
\partial\Big([0,0,1,0]^T\Big) &= [1,0,1,0]^T, & d\Big([0,0,1,0]\Big) &= [0,1,1,1];\\
\partial\Big([0,0,0,1]^T\Big) &= [0,0,1,1]^T, & d\Big([0,0,0,1]\Big) &= [0,0,0,1];
\end{aligned}$$

Therefore, the matrices are

$$\partial = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad d = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

They are the transposes of each other. □

Exercise 6.24. Provide the details of the computations above.

To make sense of the formula in the definition we use the integral interpretation again, with $a = AB$:

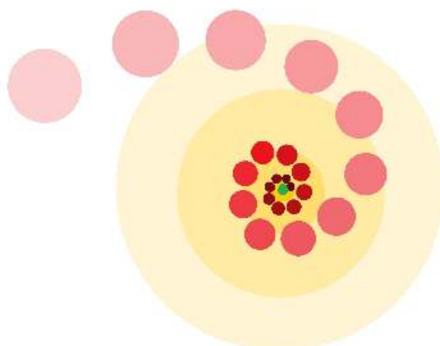
$$\int_{AB} dQ = \int_{\partial(AB)} Q = \int_{A+B} Q = Q(A) + Q(B).$$

But that's the (binary) *Fundamental Theorem of Calculus*! That's why, when we think of cochains as differential forms, d is called the *exterior derivative* (or the differential). In this sense, Q is an antiderivative of dQ.

We are in a position to conjecture that this isn't a coincidence...

Chapter II

Topologies



1 A new look at continuity

1.1 From accuracy to continuity

The idea of continuity can be introduced and justified by comparing the accuracy of direct and indirect measurements.

Let's imagine that we have a collection of square tiles of various sizes and we need to find the area A of each of them in order to know how many we need to cover the whole floor.

The answer is, of course, to measure the side, x, of each tile and then compute

$$A = x^2.$$

In particular, we may have (in inches):

$$x = 10 \Longrightarrow A = 100.$$

But what if the measurement isn't exactly accurate? What if there is *always* some error? It's never $x = 10$ but, say,

$$x = 10 \pm .3.$$

It follows that the computed value of the area of the tile – what we care about – will also have some error! Indeed, the area won't be just $A = 100$ but

$$A = (10 \pm .3)^2.$$

Hence,

$$\begin{aligned} A &= 10^2 \pm 2 \cdot 10 \cdot .3 + .3^2 \\ &= 100.09 \pm 6. \end{aligned}$$

This means that the actual area must be somewhere within the interval $(94.09, 106.09)$.

Suppose next that we are in a position to improve *the accuracy of the measurement* of the side of the tile x – as much as we like. The question is, can we also improve *the accuracy of the computed value* of A – to our, or somebody else's, satisfaction?

The standard of accuracy is subject to change... Suppose $x = 10$. The above computation shows that if the desired accuracy of A is ± 5, we haven't achieved it with the given accuracy of measurement x, which is of $\pm .3$. We can easily show, however, that $\pm .2$ would solve the problem:

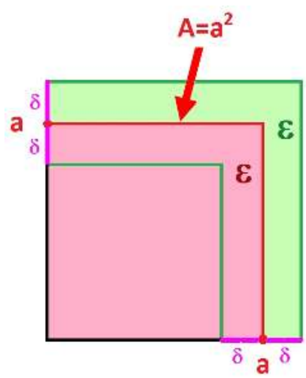
$$\begin{aligned} A &= (10 \pm .2)^2 = 10^2 \pm 2 \cdot 10 \cdot .2 + .2^2 \\ &= 100.04 \pm 4. \end{aligned}$$

It follows that the *actual* area must be within 4.04 from 100.

Let's rephrase this problem in order to solve it for all possible values of the desired accuracy of A.

Let's assume that the measurement of the side is a and, therefore, the assumed area A is a^2. Now suppose we want the accuracy of A to be some small value $\varepsilon > 0$ or better. What accuracy δ of x do we need to be able to guarantee that?

Suppose the actual length is x and, therefore, the actual area is $A = x^2$. Then we want to ensure that A is within ε from a^2 by making sure that x is within δ from a. What should δ be?



To rephrase algebraically, we want to find δ such that

$$|x - a| < \delta \Longrightarrow |x^2 - a^2| < \varepsilon.$$

The definition suggested by the above discussion is familiar from calculus.

Definition 1.1. A real-valued function f is called *continuous* at $x = a$ if for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|x - a| < \delta \Longrightarrow |f(x) - f(a)| < \varepsilon.$$

So, the answer to our question is:

• yes, we *can* always improve the accuracy of the computed value of $A = x^2$ – to anybody's satisfaction – by improving the accuracy of the measurement of x.

The reason to be quoted is that $f(x) = x^2$ is continuous at $x = 10$.

Note: The word "continuous" itself is justified on other grounds.

Exercise 1.2. Prove that $f(x) = x^2$ is continuous at $x = 0, x = 1, x = a$.

Exercise 1.3. Carry out this kind of analysis for: a thermometer put in a cup of coffee to find its temperature. Assume that the thermometer gives perfect readings. Hint: it'll take time for it to warm up.

To further illustrate this idea, consider a different situation. Suppose we don't care about the area anymore; we just want to fit these tiles into a strip 10 inches wide. We take a tile and if it fits, it is used; otherwise it is discarded.

So, we still get a measurement a of the side of the tile but our real interest is whether a is less or more than 10.

Just as in the previous example, we don't know the actual length x exactly; it's always within some limits: 5.0 ± 0.5 or $a \pm \delta$. Here δ is the accuracy of measurement of x. The algebra is much simpler than before. For example, if the length is measured as 11, we need the accuracy $\delta = 1$ or better to make the determination. It's the same for the length 9.

But what if the measurement is exactly 10? Even if we can improve the accuracy, i.e., δ, as long as $\delta > 0$, we can't know whether x is larger or smaller than 10.

Let's define a function f:

$$f(x) = \begin{cases} 1 \ \ \text{(pass)} & \text{if } x \leq 10, \\ 0 \ \ \text{(fail)} & \text{if } x > 10. \end{cases}$$

Suppose we need the accuracy of $y = f(x)$ to be $\varepsilon = 0.5$. Can we achieve this by decreasing δ? In other words, can we find δ such that

$$|x - 10| < \delta \Longrightarrow |f(x) - 1| < \varepsilon?$$

Of course not:

$$x > 10 \Longrightarrow |f(x) - 1| = |0 - 1| = 1.$$

So, the answer to our question is:

• No, we *cannot* always improve the accuracy of the computed value of $f(x)$ – to anybody's satisfaction – by improving the accuracy of the measurement of x.

The reason to be quoted is that f is discontinuous at $x = 10$.

Exercise 1.4. Carry out this kind of analysis for: the total test score vs. the corresponding letter grade. What if we introduce A-, B+, etc.?

Thus, the idea of continuity of the dependence of y on x is:

• We can ensure the desired accuracy of y by increasing the accuracy of x.

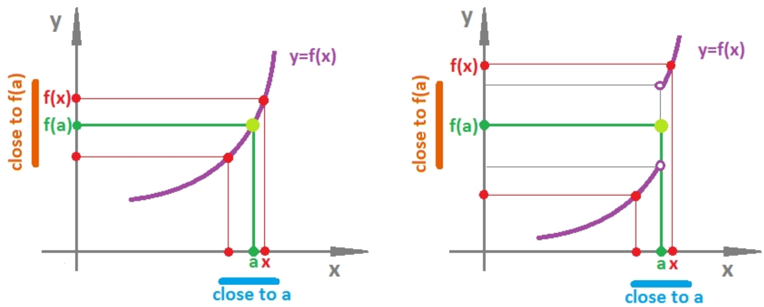
Exercise 1.5. In addition to being continuous, $f(x) = x^2$ is also differentiable. What additional information can we derive about the accuracy? Hint: there is a simple dependence between ε and δ.

1.2 Continuity in a new light

The discussion in the last subsection reveals that continuity of a function $f : \mathbf{R} \to \mathbf{R}$ at point $x = a$ can be introduced in terms of closeness (proximity) of its values, informally:

$$\textit{if } x \textit{ is close to } a \textit{ then } f(x) \textit{ is close to } f(a).$$

This description applies to the function shown on the left:



On the right, even though x is close to a, $f(x)$ does not have to be close to $f(a)$: the *discontinuity of the function does not allow us to finely control its output by controlling the input.*

Let's rephrase the second part of the continuity definition:
- for any $\varepsilon > 0$ there is a $\delta > 0$ such that
- $|x - a| < \delta \Longrightarrow |f(x) - f(a)| < \varepsilon$.

It is advantageous to restate the latter part as follows:

$$\text{every } x \in (a - \delta, a + \delta) \text{ is taken by } f \text{ to } y = f(x) \in \Big(f(a) - \varepsilon, f(a) + \varepsilon\Big).$$

Next, there is a concept that will help to make things more compact.

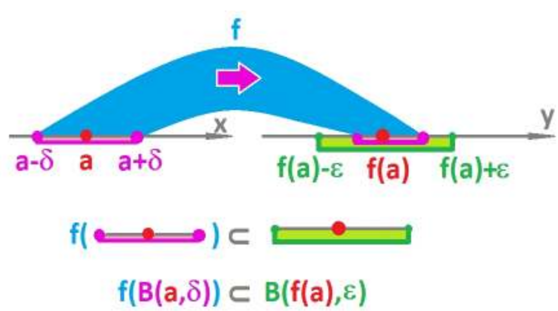
Definition 1.6. The *image of subset $A \subset X$ under function $f : X \to Y$*, is defined as

$$f(A) := \{f(x) : \ x \in A\}.$$

Now the last formulation of continuity means simply that f takes a certain subset of the x-axis to a subset of the y-axis:

$$f\Big((a - \delta, a + \delta)\Big) \subset \Big(f(a) - \varepsilon, f(a) + \varepsilon\Big).$$

Both sets are open intervals:



We look at these intervals as 1-dimensional "balls", while in general, we define an open ball in $\mathbf{R}^n$ to be

$$B(p, d) = \{u : ||u - p|| < d\}, \ d > 0.$$

We simply used the norm,

$$||x|| = ||(x_1, , ..., x_n)|| := \sqrt{\sum_{i=1}^{n} |x_i|^2},$$

to replace the absolute value.

In mathematics, we see a lot of examples of functions between higher-dimensional Euclidean spaces $f : \mathbf{R}^n \to \mathbf{R}^m$: parametric curves, parametric surfaces, vector fields, etc. We can accommodate the need to understand continuity of these functions by using the same idea of proximity:

$$||x - a|| < \delta \Longrightarrow ||f(x) - f(a)|| < \varepsilon.$$

Then our definition becomes even more compact, as well as applicable to all dimensions:

Definition 1.7. A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is continuous at point $x = a$ if

- for any $\varepsilon > 0$ there is a $\delta > 0$ such that
- $f\Big(B(a, \delta)\Big) \subset B\Big(f(a), \varepsilon\Big)$.

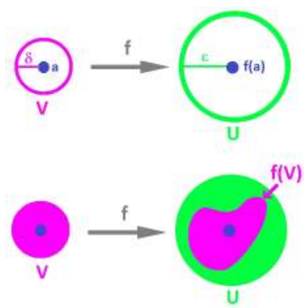
The final simplification of the latter part is:

$$f(V) \subset U,$$

where

$$V = B(a, \delta),\ U = B\Big(f(a), \varepsilon\Big).$$

In dimension 2, the relation between these sets is illustrated as follows:



Exercise 1.8. Using the definition, prove that constant functions $c : \mathbf{R}^n \to \mathbf{R}^m$ and the identity function are continuous.

Exercise 1.9. Prove that the norm is continuous.

Exercise 1.10. Prove that the composition of continuous functions is continuous.

Exercise 1.11. Provide an analog of this definition for "continuity from the left and right" of a function the domain of which is a closed interval. For a function of two variables, what if the domain is a closed square, a closed disk?

1.3 Continuity restricted to subsets

Our interest is mainly algebraic topology. Consequently, apart from pointing out some dangers to avoid, we will keep our attention limited to

- "nice" subsets of Euclidean spaces, such as *cells*, and
- those subsets "nicely" glued together.

We have already seen an example: realizations of graphs. When we come back to algebra, it will explain how these pieces fit together *globally* to form components, holes, voids, etc. Right now, in a way, we are interested in *local* topological properties, such as continuity.

Now, an example of an especially "nice" set is

$$\bar{B}(a,\delta)=\{u\in\mathbf{R}^n : ||u-a||\le\delta\},$$

the *closed ball* centered at $a\in\mathbf{R}^n$ of radius $\delta>0$. In dimension 1, this is simply a closed interval, yet functions defined on such a set won't fit into the last definition of continuity.

As a simple example, the continuity at the endpoints has to be treated as an exception, i.e., as a separate definition. Suppose we are considering continuity of a function f at $x=a$ when its domain is $[a,b]$. Then a half of the interval $(a-\delta,a+\delta)$ lies outside the domain! A way to simplify this is to verify the proximity condition only for the points inside $[a,b]$ and ignore the rest. This way, it doesn't matter how the domain fits into some bigger space.

To accommodate this idea, we retrace the analysis in the last subsection, for a function

$$f:X\to Y$$

with

$$X\subset\mathbf{R}^n,\ Y\subset\mathbf{R}^m.$$

Recall that this notation simply means that

$$x\in X\Longrightarrow f(x)\in Y.$$

We will need to adjust the second part of the definition a tiny bit more:
- $||x-a||<\delta,$ and $x\in X\Longrightarrow ||f(x)-f(a)||<\varepsilon.$

With this little addition, "$x\in X$", we have sidestepped the issue of continuity from the left/right and, in higher dimensions, the case when a point x is close to a but outside the domain of the function.

As before, the inequalities indicate simply that f takes x that lies in a certain subset of the domain of f to $y=f(x)$ in another subset of its range:
- if $x\in B(a,\delta)$ and $x\in X$ then $f(x)\in B(f(a),\varepsilon)$.

Of course, it was assumed from the beginning that $f(x)\in Y$. The condition can be further rewritten:
- $f\Big(B(a,\delta)\cap X\Big)\subset B\Big(f(a),\varepsilon\Big)\cap Y.$

Adding "$\cap Y$" is not a matter of convenience; we want to be clear that we don't care about anything outside X and Y.

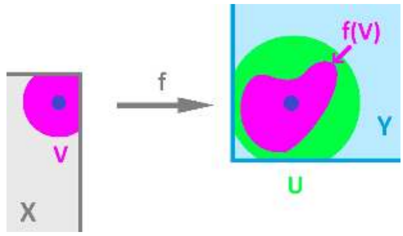
Let's state this most, for now, compact definition.

Definition 1.12. A function $f:X\to Y$ with $X\subset\mathbf{R}^n, Y\subset\mathbf{R}^m$ is called *continuous at* $x=a$ if
- for any $\varepsilon>0$, there is a δ such that
- $f(V)\subset U$,

where

$$V=B(a,\delta)\cap X,\quad U=B\Big(f(a),\varepsilon\Big)\cap Y.$$

In dimension 2, the relation between these sets is illustrated as follows:



Exercise 1.13. Show how this new definition incorporates the old definition of one-sided continuity.

The following is familiar.

Definition 1.14. A function f is called *continuous* if it is continuous at every point of its domain.

The question, "What do you know about continuity of $1/x$?" might be answered very differently by different students:

• a calculus 1 student: "$1/x$ is discontinuous... well, it's discontinuous at $x = 0$ and continuous everywhere else."

• a topology student: "The function $f : (-\infty, 0) \cup (0, \infty) \to \mathbf{R}$ given by $f(x) = 1/x$ is continuous."

...because we don't even consider continuity at a point where the function is undefined.

Exercise 1.15. Show that continuity of a function of two variables defined on the x-axis (its domain) as a function of (x, y) is equivalent to its continuity as a function of x only.

There is still an even more profound generalization to be made. We already test for continuity of f *from the inside* of its domain X, without mention of any $x \notin X$. Now we would like to take this idea one step further and avoid referring to open balls that come from the ambient Euclidean spaces.

1.4 The intrinsic definition of continuity

Let's start with the case of $X = \mathbf{R}^n$, $Y = \mathbf{R}^m$. Suppose that

• γ_X is the set of all open balls in $X = \mathbf{R}^n$, and

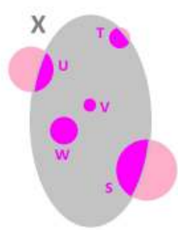
• γ_Y is the set of all open balls in $Y = \mathbf{R}^m$.

We can restate the definition in a more abstract form:

$$f : X \to Y \text{ is continuous if for any } U \in \gamma_Y \text{ there is a } V \in \gamma_X \text{ such that } f(V) \subset U.$$

Exercise 1.16. Prove that this definition is equivalent to the original.

Now, once again, what if the domain is a *subset* X of the Euclidean space?



Then, we just choose for γ_X, instead of the open balls, the *intersections of the open balls with the subset* $X \subset \mathbf{R}^n$:

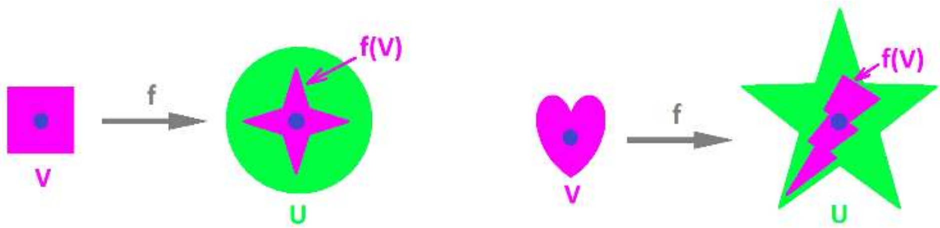
$$\gamma_X := \{B(x, \delta) \cap X : \; x \in X, \delta > 0\}.$$

The definition above remains the same.

This approach fits into our strategy of studying the *intrinsic properties* of objects. It's as if nothing outside X even exists!

Moreover, if our interest is to study X, we need to consider all and only functions with domain X. For the new definition, we will make no reference to the Euclidean space that contains X. For that, we just need to determine what collections γ_X are appropriate.

First we observe that these balls can be easily replaced with sets of any other shape:



The importance of these intervals, disks, and balls for continuity is the reason why we start studying topology by studying these collections of sets. We will call them *neighborhoods*. For example, $B(a,\delta)$ is a *ball neighborhood* of a.

Second, we notice that continuity is a local property and, therefore, we don't need a single yardstick to measure closeness (i.e., distances between points) throughout X (or Y). We just need the ability for each $a \in X$, to answer the question: "How close is x to a?" And the new answer doesn't look that different:

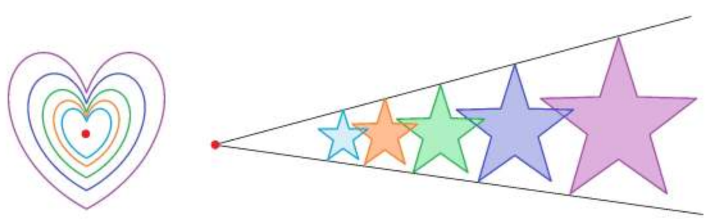
- Calculus 1: "It's within δ";
- Topology: "It's within V".

Thus, all references to numbers are gone!

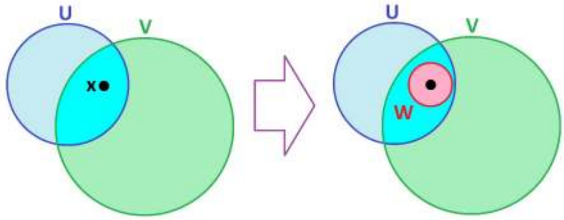
In order to have a meaningful concept of continuity, we need these collections of neighborhoods to satisfy certain properties...

First, no doubt, we want to be able to discuss continuity of $f : X \to Y$ at *every point* $a \in X$. So, these neighborhoods have to cover the whole set:

Second, we want to be able to *zoom in* on every point to ensure any desired accuracy (or to detect breaks in the graphs of functions). So, we must be able to "refine" the collection:



This is how we understand the refining process, illustrated below for the disks:



In summary, we will require:
- (B1) $\cup\gamma = X$;
- (B2) for any two neighborhoods $U, V \in \gamma$ and any point $c \in U \cap V$ in their intersection, there is a neighborhood $W \in \gamma$ of c such that $W \subset U \cap V$.

We will demonstrate that this is *all* the extra structure we need to have on sets X and Y to be able to define and study continuity!

Definition 1.17. Suppose X is any set. Any collection γ satisfying properties (B1) and (B2) is called a *basis of neighborhoods* in X.

And the new definition of continuity is wonderfully compact:

Definition 1.18. Given two sets X, Y with bases of neighborhoods γ_X, γ_Y, function $f : X \to Y$ is *continuous* if for any $U \in \gamma_Y$ there is a $V \in \gamma_X$ such that $f(V) \subset U$.

Exercise 1.19. Prove that the definition of continuity in the last subsection satisfies this definition.

---

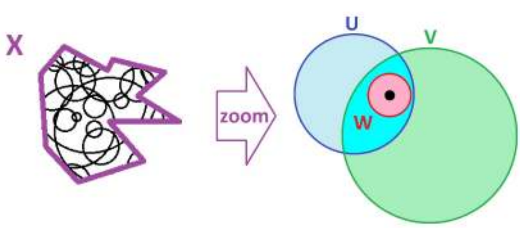
2 Neighborhoods and topologies

2.1 Bases of neighborhoods

Let's recall our main definition.

Definition 2.1. Suppose X is any set. A collection γ of subsets of X satisfying the properties of:
- (B1) *Covering*: $\cup\gamma = X$; and
- (B2) *Refining*: for any neighborhoods $U, V \in \gamma$ and any point $x \in U \cap V$ in their intersection, there is a neighborhood $W \in \gamma$ of x such that $W \subset U \cap V$;

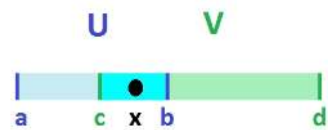
is called a *basis of neighborhoods* (or simply "basis") on X. The elements of γ are called *neighborhoods*, and if $x \in W \in \gamma$, we say that W is a *neighborhood of* x.



The purpose of this definition is to enable us to discuss continuity of functions in the most general setting. Indeed, X doesn't have to have either algebraic (such a vector space) or geometric (such as a metric space) structure.

In the case of the 1-dimensional Euclidean space $\mathbf{R}$, our neighborhoods γ have been simply open intervals, which makes the refining condition of γ especially short:

$$(a,b),(c,d) \in \gamma, a < c < x < b < d \Longrightarrow x \in (a,b) \cap (c,d) = (c,b) \in \gamma.$$



Now, we can modify this collection γ by choosing open intervals with the endpoints being:
- only rational, or
- only irrational, or
- left rational and right irrational, etc.

The proof of the refining condition above won't change!

However, closed intervals (of non-zero length) won't work:

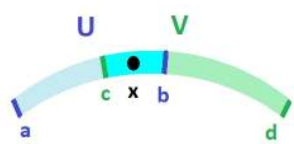
$$[a,b],[b,c] \in \gamma, a < b < c \Longrightarrow [a,b] \cap [b,c] = \{b\} \notin \gamma.$$

Another basis in $\mathbf{R}$ is the rays $R(s) = \{u \colon u > s\}$ (only left or only right). Indeed, we have:

$$(a,\infty),(c,\infty) \in \gamma, a < c \Longrightarrow (a,\infty) \cap (c,\infty) = (c,\infty) \in \gamma.$$

Exercise 2.2. Find more examples of γ for the line that satisfy (B1) but not (B2). Hint: try to narrow the choice of possible centers.

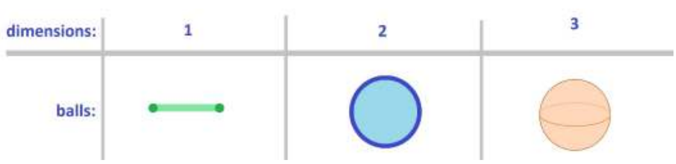
Exercise 2.3. Prove that the collection of all open *arcs* is a basis of the circle.



The collection of all *open balls* in $X = \mathbf{R}^n$,

$$\gamma_b = \{B(a,\delta) : a \in \mathbf{R}^n, \delta > 0\},$$

is a basis.



We will call it the *standard Euclidean basis.*

What are other bases in $\mathbf{R}^n$?

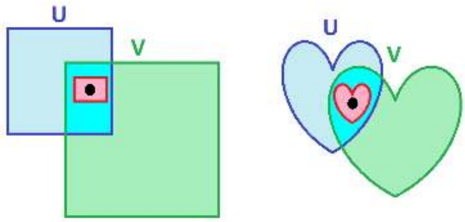
An important example is the squares. The collection of all squares (with sides parallel to the axes) in $\mathbf{R}^2$ is

$$\gamma_s = \{S\big((a,b),d\big) : \ (a,b) \in \mathbf{R}^2, d > 0\},$$

with

$$S\big((a,b),d\big) := \{(x,y):\ |x-a|<d, |y-b|<d\},$$

is a basis:



Exercise 2.4. Prove this statement. Hint: you will have to be specific.

Exercise 2.5. Is the collection of all rectangles (with sides parallel to the axes) in $\mathbf{R}^2$ a basis? What is its analog in $\mathbf{R}^n$? Hint: think "boxes".

The collection of *closed balls*,

$$\bar{B}(p,d) = \{u\colon ||u-p|| \leq d\},\ d>0,$$

however is not a basis. It is easy to see if we observe that two closed balls may intersect by a single point, which is not a closed ball. If we wanted to turn this collection into a basis, we'd have to add all *singletons*, i.e., the one-point sets, to the collection. Since those are simply balls of diameter 0, we end up with

$$\gamma_p = \{B(a,\delta):\ a\in\mathbf{R}^n, \delta\geq 0\}.$$

Exercise 2.6. Prove that this is a basis.

The sets of all:
- rectangles,
- ellipses,
- stars, etc.,

are also bases.

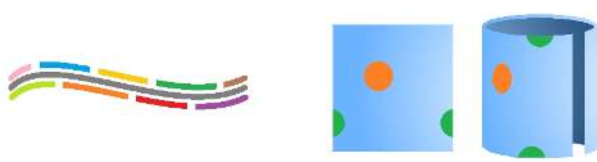
Exercise 2.7. Prove that disks with rational centers and/or radii form a basis.

Example 2.8. Suppose we have a parametric curve $p:\mathbf{R}\to\mathbf{R}^2$ in the plane without self-intersections. Then the image of the curve

$$\operatorname{Im} p := \{p(t):\ t\in\mathbf{R}\}$$

will acquire a basis from the standard basis of $\mathbf{R}$, as follows. Let

$$\gamma := \Big\{p\big((a,b)\big),\ a<b\Big\}.$$



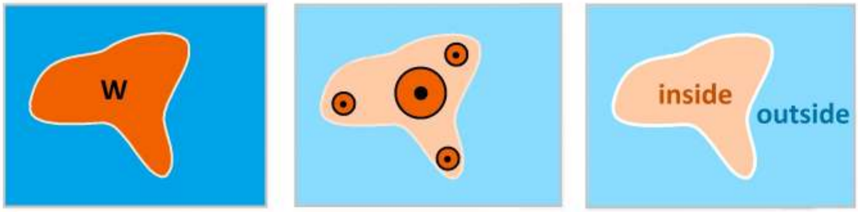
□

Exercise 2.9. Provide a similar construction for parametric surfaces. What happens to these bases if there are self-intersections? What difference does continuity of the parameters make?

2.2 Open sets

Below, we assume that a basis of neoghborhoods γ is fixed.

Now we define open sets as the ones where every point has its own neighborhood:



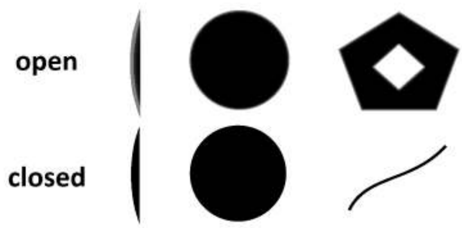
Definition 2.10. A subset W of X is called *open* (with respect to a given basis of neighborhoods γ) if for any $x \in W$ there is a neighborhood $U \in \gamma$ of x that lies entirely within W:

$$x \in U \subset W.$$

For example, all "open" balls and "open" squares are open.

Exercise 2.11. What sets are open with respect to the set of all right rays?

When drawing open sets, we use various visual clues to suggest that the points on the boundary of the sets don't belong to the set:



Theorem 2.12. The intersection of two open sets is open.

To approach the proof, lets start with some analysis. We will need to write the definition three times for each of the three sets involved.

Suppose U, V are open with respect to γ and let $W := U \cap V$. Suppose $x \in W$ is given.

Now, due to $W = U \cap V$, we have

$$x \in U \text{ and } x \in V.$$

Because both U and V are open, we conclude from the definition:

$$x \in N \subset U \text{ and } x \in N \subset V.$$

Can we conclude that

$$x \in N \subset U \cap V = W?$$

No, we can't! The flaw in this argument is that, applied separately, the definition will produce two *different* Ns.

Fortunately, the refining condition (B2) will give us the desired – smaller – neighborhood of x.

We are now prepared to write the proof.

Proof. Given $x \in W = U \cap V$. Since both U, V are open by assumption, we have:
- there is a $P \in \gamma$ such that $x \in P \subset U$;

• there is a $Q \in \gamma$ such that $x \in Q \subset V$.
In particular, we have $x \in P \cap Q$, and we can apply (B2) to this situation: there is $B \in \gamma$ with

$$B \subset P \cap Q.$$

To continue,

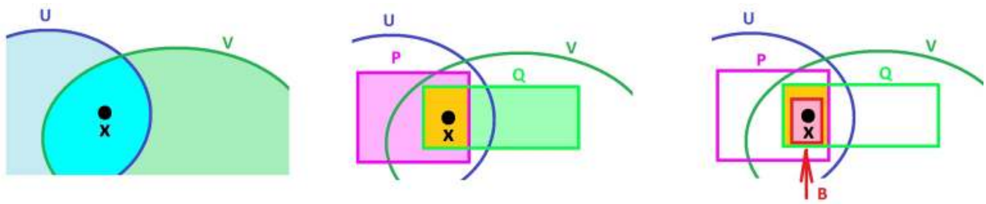
$$B \subset P \cap Q \subset U \cap V = W.$$

So, we have shown that there is a $B \in \gamma$ such that

$$x \in B \subset W.$$

By definition, W is open. ■

The construction is illustrated below:



The illustration is inserted after the proof to emphasize that the latter can't rely on – or even refer to – pictures.

What about the union? Let's try to recycle the proof for intersection. After all, we will need to use the definition of openness three times, once again.

Theorem 2.13. The union of two open sets is open.

Proof. We copy the proof from before and replace each $\cap$ with $\cup$. This is what we get:

"Suppose U, V are open with respect to γ and let $W := U \cup V$. Suppose $x \in W$ is given.

Given $x \in W = U \cup V$. Since both U, V are open by assumption, we have:
• there is a $P \in \gamma$ such that $x \in P \subset U$;
• there is a $Q \in \gamma$ such that $x \in Q \subset V$.
[In particular, we have $x \in P \cup Q$ and we can apply (B2) to this situation: there is a $B \in \gamma$ with

$$B \subset P \cup Q.$$

To continue,]

$$B \subset P \cup Q \subset U \cup V = W.$$

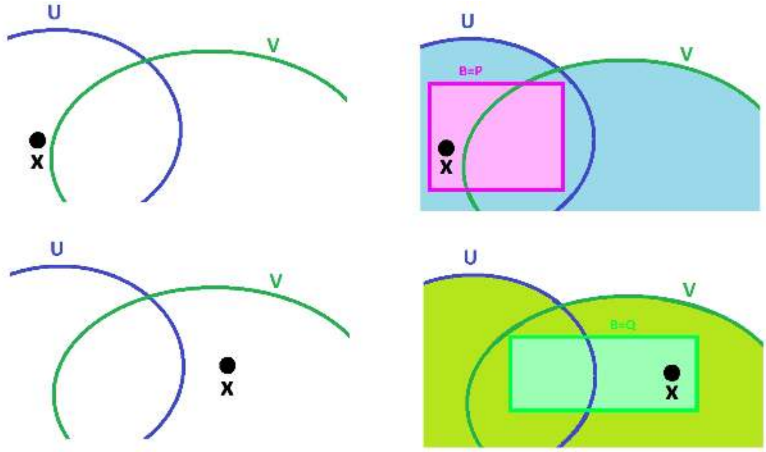
So, we have shown that there is a $B \in \gamma$ such that

$$x \in B \subset W.$$

By definition, W is open."

Turns out, the part in brackets is unnecessary. Instead, we just choose either $B = P$ or $B = Q$ depending on whether P or Q contains x. ■

Why the proof is easy is illustrated below:



Theorem 2.14. The union of *any* collection of open sets is open.

Exercise 2.15. Prove this theorem (a) by following the proof of the last theorem, and (b) as a corollary.

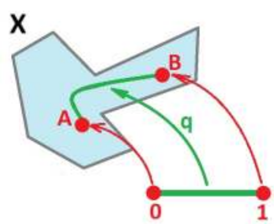
Naturally, a subset $A \subset X$ will be called *closed* if its complement $X \setminus A$ in X is open.

2.3 Path-connectedness

Since we defined continuity in terms of open sets, the idea of what functions are continuous will entirely depend on the choice of topology for our spaces. We now investigate path-connectedness in this environment.

We will need these three items here:

• A topological space X is path-connected, if for any two points $A, B \in X$ there is a continuous function $q : [0,1] \to X$ such that $q(0) = A$, $q(1) = B$.



• function $f : X \to Y$ is continuous if for any $U \in \gamma_Y$ there is a $V \in \gamma_X$ such that $f(V) \subset U$.

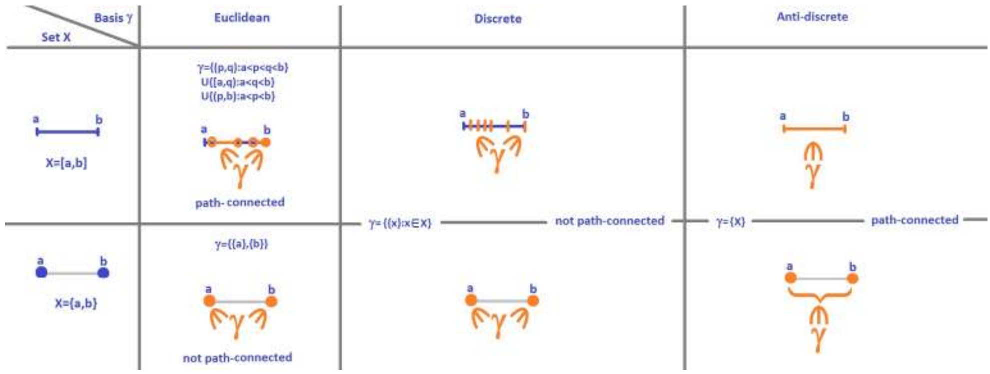
• the topology of $[0,1]$ is Euclidean generated by the set of all open intervals γ.

You are to demonstrate now how misleading the above illustration is.

Exercise 2.16. Suppose X, Y are two topological spaces, one of which is (a) discrete or (b) anti-discrete. Determine under what circumstances a function $f : X \to Y$ is continuous. Specifically, answer these questions:

- What topology for the domain or target space would guarantee that f is continuous?
- What function f would be continuous regardless of the topology these spaces?

The diagram below illustrates the idea that simple topological concepts, such as path-connectedness, are meaningless until the topology is specified. Two sets, the interval $[0,1]$ and two points $\{0,1\}$, are equipped with various topologies and then path-connectedness is determined for either of them.



It appears that the two topologies at the two extremes of our hierarchy are indeed *extreme*.

Exercise 2.17. Prove that
- with respect to discrete topology, only the singletons are a path-connected;
- with respect to anti-discrete topology, all sets are path-connected.

Exercise 2.18. For both discrete topology and anti-discrete topology, what are the path-components.

2.4 From bases to topologies

Let's summarize the results of this section.

Theorem 2.19. Given a basis of neighborhoods γ, let τ be the set of all open sets with respect to γ. Then τ satisfies the following conditions:
- (T1) $\emptyset, X \in \tau$;
- (T2) if $\alpha \subset \tau$ then $\cup\alpha \in \tau$;
- (T3) if $U, V \in \tau$ then $U \cap V \in \tau$.

Exercise 2.20. Prove (T1).

Definition 2.21. Given a basis of neighborhoods γ, the collection τ of all sets open in X with respect to γ is called *the topology of X generated by γ.*

The smallest possible basis on set X is

$$\gamma_a := \{X\}.$$

The topology it generates is called the *anti-discrete topology*:

$$\tau_a = \{\emptyset, X\}.$$

The largest possible basis is the set of all subsets of X:

$$\gamma_d := \{A \subset X\} = 2^X.$$

The topology it generates is called the *discrete topology*; it coincides with the basis:

$$\tau_a = \{A \subset X\} = 2^X.$$

Note that this topology is more economically generated by the basis of all *singletons*:

$$\gamma_s = \big\{\{x\} : \ x \in X\big\}.$$

Theorem 2.22. All neighborhoods are open; i.e., $\gamma \subset \tau$.

Proof.
- (T1) $\Longrightarrow$ (B1);
- (T3) $\Longrightarrow$ (B2). ■

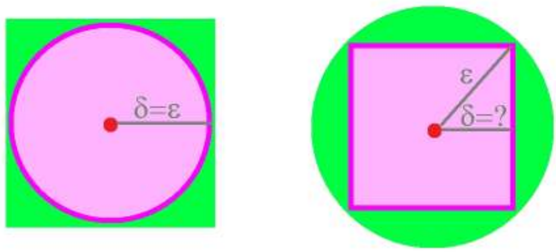
A basis can now be easily enlarged with open sets.

Proposition 2.23. If U is open with respect to basis γ, then $\gamma \cup \{U\}$ is also a basis.

Exercise 2.24. Prove the proposition.

We can *refine* the basis of disks $B(x,\varepsilon)$ with that of squares $S(x,\delta)$ and vice versa:

Lemma 2.25. For any $x \in \mathbf{R}^2$,
- for any $\varepsilon > 0$ there is a $\delta > 0$ such that $B(x,\delta) \subset S(x,\varepsilon)$; and
- for any $\varepsilon > 0$ there is a $\delta > 0$ such that $S(x,\delta) \subset B(x,\varepsilon)$.



Theorem 2.26. The topology τ_d generated by the disks γ_d is the same as the topology generated by the squares.

Exercise 2.27. Prove the lemma and the theorem.

Another basis in $\mathbf{R}^2$ is γ_q, the set of all quadrants:

$$Q(r,s) = \{(x,y) \in \mathbf{R}^2 : \ x > r, y > s\}.$$

Exercise 2.28. Prove that it is.

Disks refine quadrants but not vice versa! This fact suggests the following.

Theorem 2.29. The topology τ_d generated by the disks γ_d is *not* the same as the topology τ_q generated by the quadrants γ_q.

Exercise 2.30. Prove the theorem. Hint: you just need a single example of an open set.

When two bases generate the same topology they are called *equivalent*.

For a given set X, the "sizes" of topologies on X vary – between the anti-discrete and the discrete. This is the hierarchy for $X = \mathbf{R}^2$:

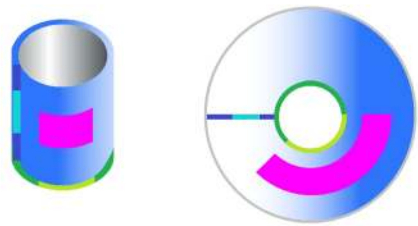
$$\tau_a \subset \tau_q \subset \tau_d \subset \tau_d.$$

Exercise 2.31. Determine which of the following collections are bases of neighborhoods; find their topologies when they are, and identify the equivalent ones:
- $\{[a,b) \subset \mathbf{R} : \ a \leq b\}$,
- $\{(n,n+2) \subset \mathbf{R} : \ n \in \mathbf{Z}\}$,
- $\{[a,b] \subset \mathbf{R} : \ a < b\}$,
- $\{[a,b] \subset \mathbf{R} : \ ab\}$,
- $\{(-x,x) \subset \mathbf{R} : \ x \in \mathbf{R}\}$,
- $\{(-\infty,q) \subset \mathbf{R} : \ q \in \mathbf{Q}\}$,
- $\big\{\{a\} \times (b,c) \subset \mathbf{R}^2 : \ a,b,c \in \mathbf{R}\big\}$.

Exercise 2.32. Prove that the set of all right half-intervals in **R** is a basis of neighborhoods:

$$\gamma = \{[a,b) : \ a < b\}.$$

Exercise 2.33. Provide formulas for these neighborhoods on the cylinder and the ring and prove that they form bases:



2.5 From topologies to bases

We may choose to enter topology, as we did, through the door that leads to the bases of neighborhoods and then to the topologies, or we might start with the topologies. We simply turn the theorem into a definition, as follows:

Definition 2.34. Given a set X, any collection τ of subsets of X is called *a topology* on X if it satisfies the following three conditions:

- (T1) $\emptyset, X \in \tau$;
- (T2) if $\alpha \subset \tau$ then $\cup\alpha \in \tau$;
- (T3) if $U, V \in \tau$ then $U \cap V \in \tau$.

The elements of τ are still called *open sets*. The set paired up with a topology, (X, τ), is called a *topological space*.

Just as before, the smallest possible topology on X is the *anti-discrete topology*:

$$\tau_a := \{\emptyset, X\},$$

and the largest possible is the *discrete topology*:

$$\tau_d := \{A \subset X\} = 2^X.$$

We restate the theorem from the last subsection as one about a relation between bases and topologies.

Theorem 2.35. Given a basis γ, the collection τ of all sets open with respect to γ is a topology.

In this case, we still say that basis γ *generates topology* τ.

The issue of "topological space vs. its basis" is similar to that of "vector space vs. its basis". Indeed, in either case there are many possible bases for the same space. The dissimilarity is in the fact that there is no requirement for a basis of neighborhoods to be as small as possible. In fact, τ is generated by τ.

Exercise 2.36. Suppose

- X is a set,
- γ is a basis of neighborhoods on X,
- τ is a topology on X (unrelated to γ),
- A is a subset of X.

Define:

- $\gamma' := \{N \cap A : \ N \in \gamma\}$,
- $\tau' := \{U \cap A : \ U \in \tau\}$.

Prove:

- (a) γ' satisfies condition (B2),
- (b) τ' satisfies conditions (T2) and (T3).

Theorem 2.37. The intersection of any *finite* collection of open sets is open.

Exercise 2.38. Prove the theorem.

---

3 Topological spaces

3.1 Open and closed sets

In order to define and study

- continuity,
- path-connectedness,

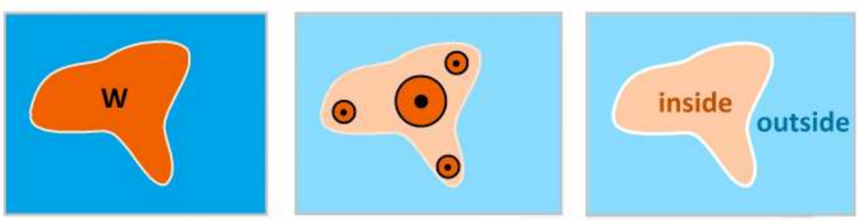
and other related issues, all we need is to equip each of the sets involved with an additional structure.

Definition 3.1. Given an arbitrary set X, a collection τ of subsets of X is called a *topology on* X if it satisfies the following conditions:

- (T1) $\emptyset, x \in \tau$;
- (T2) if $\alpha \subset \tau$ then $\cup\alpha \in \tau$;
- (T3) if $U, V \in \tau$ then $U \cap V \in \tau$.

The pair (X, τ) is called a *topological space*. The elements of τ are called *open sets*.

Of course, instead, we can, and did, start with the set of all neighborhoods γ in X and then define open sets as ones where every point has its own neighborhood: a subset W of X is called *open* (with respect to γ) if for any x in W there is a neighborhood U $(\in \gamma)$ of x that lies entirely within W:



Then, as we have proven, the set of all such sets, say τ_γ, satisfies (T1) - (T3).

We will use both approaches, whichever is more convenient at the time.

As an example, all "open", as we have known them, intervals, finite or infinite, are open in $\mathbf{R}$:

$$(0,1), ..., (0,\infty), (-\infty,0), ...$$

"Open" disks on the plane, and balls in the Euclidean space, are also open.

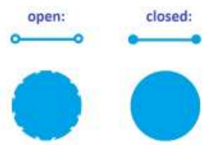
Definition 3.2. A subset is called *closed* if its complement is open.

In particular, all "closed" intervals, finite or infinite, are closed in **R**:

$$[0,1], ..., [0,\infty), (-\infty,0], ...$$

"Closed" disks on the plane, and "closed" balls in the Euclidean space are also closed. Points too.

There is a standard way to visualize open vs. closed sets. For the former, the borders are drawn "hollow" or dashed in order to suggest that they are not included (for dimensions 1 and 2):



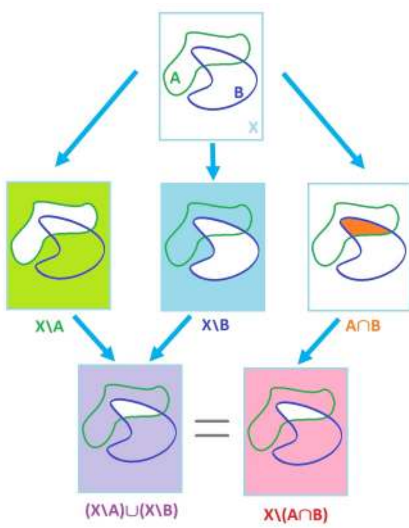
Some sets are *neither closed nor open* (unlike doors!). Examples in **R** are:
- s half-open interval: $[0,1)$, and
- a convergent sequence: $S = \left\{\frac{1}{n} :\ n = 1,2,...\right\}$ that doesn't include its limit.

To see that the latter is neither (with respect to the standard Euclidean topology), consider these two facts:
- $x = 1$ does not have an interval around it that lies entirely inside an open S, $\Longrightarrow$ not open;
- $x = 0$ is in the complement of S but it does not have an interval around it that lies inside in $\mathbf{R} \setminus S$, $\Longrightarrow$ not closed.

However, $S \cup \{0\}$ is closed.

Theorem 3.3. The set σ of all closed sets satisfies the following conditions:
- (T1′) $\emptyset, X \in \sigma$;
- (T2′) if $\alpha \subset \sigma$ then $\cap\alpha \in \sigma$;
- (T3′) if $P, Q \in \sigma$ then $P \cup Q \in \sigma$.

Proof. (T3′) follows from (T3), illustrated below:



We simply used the de Morgan's Laws. ■

Exercise 3.4. Given a basis of neighborhoods γ, prove directly, i.e., without using the de Morgan's Laws, that the set σ_γ of all closed sets with respect to γ satisfies (T1′) - (T3′). Hint: imitate the proof for τ_γ.

Plainly, condition (T3′) implies that the union of any *finite* collection of closed sets is closed.

To see why the (T3′) cannot be extended to infinite collections of sets, consider these examples:

$$\begin{aligned}\bigcap \left\{ \left(-\tfrac{1}{n}, \tfrac{1}{n} \right) :\ n = 1,2,... \right\} &= \{0\};\\ \bigcup \left\{ \left[-1+\tfrac{1}{n}, 1-\tfrac{1}{n} \right] :\ n = 1,2,... \right\} &= (-1,1).\end{aligned}$$

So, just as there is no (T3) for closed sets, there is no (T3′) for open sets.

A note on **terminology**. Once we assume that the topology τ is fixed, all references to it will be routinely omitted. Instead of a topological space (X,τ), we will speak of a topological space X, along with some assumed fixed τ, and

- "open sets in X" = the elements of τ,
- "closed sets in X" = the complements of the elements of τ,
- "neighborhoods of points in X" = the elements of some fixed basis γ that generates τ.

A profound (but not unexpected) benefit of limiting yourself to using only the language of open sets is that all properties you are *able* to discuss are intrinsic! They are called *topological invariants.*

3.2 Proximity of a point to a set

There are no measurements in topology. Does the distance between a point and a set make any sense?

Let's narrow down this problem and just try to decide whether the distance is 0 or not.

Suppose we are given a subset A of a topological space X. We can state, without hesitation, that, if $x \in A$, then the distance between them is 0!

Now, for $X := \mathbf{R}$, suppose A is the set of points of the sequence

$$A := \left\{ \tfrac{1}{n} :\ n = 1,2,... \right\}.$$

What is the distance between A and $0 \notin A$? Since the sequence *converges* to this point, the distance also appears to be 0... But the distance to -1 can't be 0.

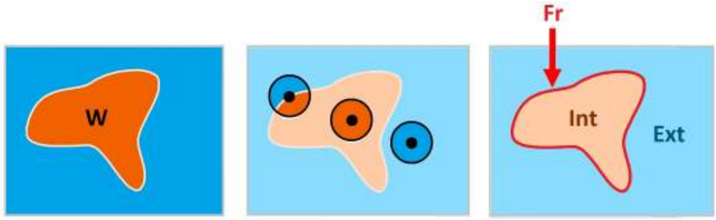
What makes a difference topologically? There is a neighborhood of -1, or any number $x < 0$, that separates it from the set A. That's what makes it "far from" A.

This idea of separation of a point from a set, or from its complement, by a neighborhood is developed below.

Definition 3.5. A point $x \in X$ is called an *interior point* of A if there is a neighborhood W of x that lies entirely in A. The set of interior points of a set is called its *interior* **denoted** by

$$\operatorname{Int}(U).$$

In other words, these are the points that are "far from" $X \setminus A$ as each is isolated from the complement of A by a neighborhood.



Then an open set U coincides with its interior. In fact, we have the following:

Theorem 3.6. $\text{Int}(A)$ is open for any A.

Exercise 3.7. Prove the theorem.

Example 3.8. The interior points of $A = [0,1]$ in $\mathbf{R}$ are $(0,1)$. Same for $(0,1], [0,1)$, and $(0,1)$.

In $\mathbf{R}^2$, let

- $A := \{(x,y) : x^2 + y^2 \leq 1\}$ (the closed disk). Then,
- $\text{Int}(A) = D = \{(x,y) : x^2 + y^2 < 1\}$ (the open disk); also
- $\text{Int}(D) = D$.

The interior of a circle $C = \{(x,y) : x^2 + y^2 = 1\}$ in the plane is empty. □

Definition 3.9. A point x is called an *exterior point* of A if there is a neighborhood W of x that lies entirely in $X \setminus A$. We will use notation

$$\text{Ext}(A)$$

for this set called the *exterior*.

These are the points that are "far from" A. Thus, exterior points are interior points of the complement.

Example 3.10. In $\mathbf{R}$, suppose P is the set of the points of the sequence

$$P := \left\{ \tfrac{1}{n} : \ n = 1, 2, ... \right\}.$$

Then the exterior points of P are all points in the complement of P except for 0. Why not 0? Because $\frac{1}{n} \to 0$; hence, for any $\varepsilon > 0$ there is a large enough n so that $\frac{1}{n} \in B(0,\varepsilon)$. Also, $\text{Int}(P)$ is empty. □

Definition 3.11. A point x is called a *limit point* of A if for any neighborhood W of x

$$W \cap A \neq \emptyset.$$

Certainly, 0 is a limit point for the set P above. But we also note that any point x in A is a limit point of A because $x \in W \cap A$. To exclude this possibility and to be able to study convergence of sequences in sets, we introduce one more concept.

Definition 3.12. A point x is an *accumulation point* of A if for any neighborhood W of x

$$W \cap (A \setminus \{x\}) \neq \emptyset.$$

A point x that is a limit point but not an accumulation point of A satisfies: there is a neighborhood W of x such that

$$W \cap A = \{x\}.$$

Definition 3.13. A point x in a set A is called *isolated* if there is a neighborhood W of x the intersection of which with the set A is $\{x\}$.

Example 3.14. Suppose Q is the set that consists of the points of the same sequence as above but combined with its limit, 0:

$$Q := \left\{\tfrac{1}{n}:\ n = 1, 2, ...\right\} \cup \{0\}.$$

Then all points of Q are isolated except for 0. □

Definition 3.15. A point x is called a *frontier point* of A if it is a limit point for both the set A and its complement. The set of all frontier points is called the *frontier* $\text{Fr}(A)$ of subset A.

The condition can be reworded: any neighborhood W of x intersects both A and $X \setminus A$.

Example 3.16. The frontier $\text{Fr}(D)$, of the open disk A is the circle C. □

The frontier is not to be confused with the "boundary" (to be discussed later), even though they are sometimes equal.

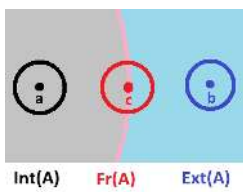
Given a subset A of a topological space X, all points of X can now be classified according to how close (0 or not 0) they are to A and to its complement.

3.3 Interior - frontier - exterior

Theorem 3.17. For any set A in X, the interior, the frontier, and the exterior of A form a partition of X:

$$X = \text{Int}(A) \sqcup \text{Fr}(A) \sqcup \text{Ext}(A).$$

Proof. We want to show that the complement of the union of the interior and exterior consists of all points that are limit points of both A and its complement. This is just a matter of manipulating the definitions in a predictable way:



Suppose γ is the basis of neighborhoods. Then
- (1) $a \in \text{Int}(A) \iff \exists N_a \in \gamma, N_a \subset A;$
- (2) $b \in \text{Ext}(A) \iff \exists N_a \in \gamma, N_b \subset X \setminus A.$

Now,
- $x \in \text{Int}(A) \cup \text{Ext}(A) \iff$
- (1) $\exists N_a \in \gamma, x \in N_a,\ N_a \subset A$; OR
- (2) $\exists N_b \in \gamma, x \in N_b,\ N_b \subset X \setminus A.$

Therefore, a point is in the complement if and only if it satisfies the negation of (1) and (2):
- $c \notin \text{Int}(A) \cup \text{Ext}(A) \iff c \in X \setminus (\text{Int}(A) \cup \text{Ext}(A)) = (X \setminus \text{Int}(A)) \cap (X \setminus \text{Ext}(A)) \iff$
  - ⋄ (1′) $\forall N_c \in \gamma, N_c \cap A \neq \emptyset$; AND
  - ⋄ (2′) $\forall N_c \in \gamma, N_c \cap X \setminus A \neq \emptyset$;

Putting these two together:
- $c \notin \text{Int}(A) \cup \text{Ext}(A) \iff$
  - ⋄ $\forall N_c \in \gamma, N_c \cap A \neq \emptyset$ and $N_c \cap (X \setminus A) \neq \emptyset$. ■

Definition 3.18. The *closure*, $\text{Cl}(A)$, of A is the set of all limit points of A.

The following two theorems are obvious:

Theorem 3.19.

$$\text{Int}(A) \subset A \subset \text{Cl}(A).$$

Theorem 3.20.

$$\mathrm{Cl}(A) = \mathrm{Int}(A) \cup \mathrm{Fr}(A).$$

Theorem 3.21.
- 1. $\mathrm{Int}(A)$ is open.
- 2. $\mathrm{Fr}(A)$ is closed.
- 3. $\mathrm{Cl}(A)$ is closed.

Proof. (1) is obvious. (2) and (3) are below. ■

To summarize these definitions, let's consider the gaps in this sentence, for a given $x \in A$, and the three ways they can be filled:

"If $\exists W_x$ so that	...	, x is an	...	point."
...	$W_x \subset A$	...	interior	...
...	$W_x \subset X \setminus A$	...	exterior	...
...	$W_x \cap A = \{x\}$	...	isolated	...

Now consider another sentence with gaps and the three ways they can be filled:

"If $\forall W_x$we have	...	, x is a	...	point."
...	$W_x \cap A \neq \emptyset$	...	limit	...
...	$W_x \cap (A \setminus \{x\}) \neq \emptyset$	...	accumulation	...
...	$W_x \cap A \neq \emptyset, W_x \cap (X \setminus A) \neq \emptyset$	...	frontier	...

Theorem 3.22.
- (1) The closure is the intersection of a certain collection of closed sets. $\Longrightarrow$
- (2) The closure is closed. $\Longrightarrow$
- (3) The frontier is closed.

The theorem follows from this result.

Theorem 3.23. The closure of a set is the *smallest closed set* containing the set:

$$\mathrm{Cl}(A) = \bigcap \{G : G \text{ closed in } X, A \subset G\}.$$

Proof. Let

$$\begin{aligned} \alpha :=& \quad \{G : G \text{ closed in } X, A \subset G\}, \\ \bar{A} :=& \quad \cap \alpha. \end{aligned}$$

Then we need to show
- $\bar{A} = \mathrm{Cl}(A)$.

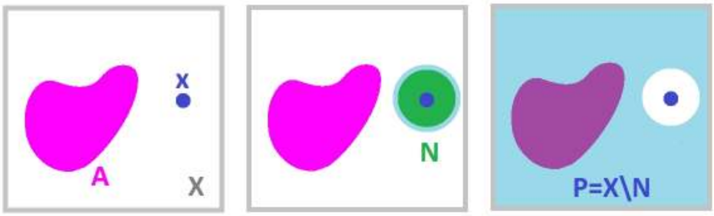
A proof of equality of sets often contains two parts:
- 1. $\bar{A} \subset \mathrm{Cl}(A)$, and
- 2. $\bar{A} \supset \mathrm{Cl}(A)$.

Part 1. We start on the right. Suppose $x \notin \mathrm{Cl}(A)$. Then, by definition, x is not a limit point of A, so
- there is an open neighborhood N of x such that
- $N \cap A = \emptyset$.

Now, we need to connect N to $\bar{A}$ and, hence, to α. But the latter consists of *closed* sets, while N is *open*! So, let's try

$$P := X \setminus N.$$

This clever idea is illustrated below:



Observe:
- P is closed because N is open, and
- P contains A because $N \cap A = \emptyset$.

Hence, $P \in \alpha$. It follows that
$$\bar{A} = \cap \alpha \subset P.$$
Hence, $x \notin \bar{A}$.

To summarize what has been proven,
$$x \notin \text{Cl}(A) \Longrightarrow x \notin \bar{A}.$$
Finally, we rephrase this as follows:
$$\text{Cl}(A) \supset \bar{A}.$$
Note: except for choosing to consider P, the proof can be seen as routine and based entirely on the definition. ■

Exercise 3.24. Prove:
$$\bar{A} \supset \text{Cl}(A).$$

Exercise 3.25. (a) Prove:
$$\text{Cl}(A \cap B) \subset \text{Cl}(A) \cap \text{Cl}(B).$$
(b) What if we face infinitely many sets?

Theorem 3.26. The interior of a set is the *largest open set* contained by the set:
$$\text{Int}(A) = \bigcup \{U : U \text{ open in } X, A \supset G\}.$$

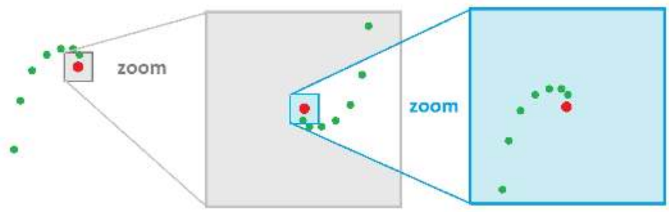
Exercise 3.27. Prove this theorem based entirely on the definition, just as the last theorem.

Exercise 3.28. Prove:
$$\overline{B(p,d)} = \bar{B}(p,d).$$

3.4 Convergence of sequences

Recall from calculus that a sequence of numbers $\{x_n : n = 1, 2, ...\}$ converges to number a if for any $\varepsilon > 0$ there is an $N > 0$ such that for any integer $n > N$ we have
$$|x_n - a| < \varepsilon.$$
In other words, the elements of the sequence will be *eventually within* ε from a. In the case of an m-dimensional Euclidean space, the last part will be
$$||x_n - a|| < \varepsilon,$$
with the norm $|| \cdot ||$ replacing the absolute value.

This idea is visualized through a repeated zooming in on the "tail" of the sequence:



Based on our experience with continuity, we know how to recast this definition in terms of bases of neighborhoods. We just observe that the last part is simply:

$$x_n \in B(a, \varepsilon).$$

This means that the elements of the sequence will be *eventually within* $U = B(a, \varepsilon)$, an element of the Euclidean basis, from a. Finally, we just replace the standard Euclidean neighborhood with one of arbitrary nature, as follows.

Definition 3.29. Suppose X is a set with a basis of neighborhoods γ. Suppose also $\{x_n : n = 1, 2, ...\}$ is a sequence of elements of X and $a \in X$. Then we say that the *sequence converges to* a,

$$\lim_{n \to \infty} x_n = a, \text{ or } x_n \to a \text{ as } n \to \infty,$$

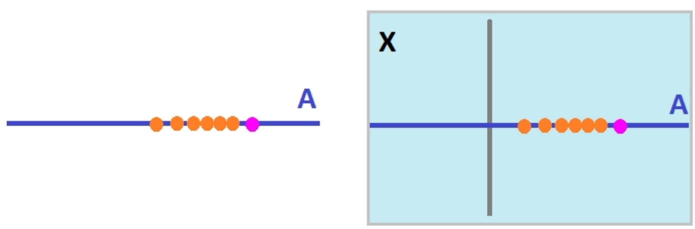
if for any $U \in \gamma$ there is an $N > 0$ such that

$$n > N \Longrightarrow x_n \in U.$$

Exercise 3.30. Prove that the definition is remains equivalent if we replace "basis γ" with "topology τ".

As squares refine disks, every convergent (in the usual sense) sequence in $\mathbf{R}^2$ will also converge with respect to the squares.

Exercise 3.31. Prove that if $x_n \to a$ then $(x_n, 0) \to (a, 0)$:



Exercise 3.32. Show that if a sequence converges with respect to a basis γ, it also converges with respect to any basis equivalent to γ.

Let's consider convergence in the two extreme cases: discrete and anti-discrete.

Example 3.33. When do we have $x_n \to a$ in *discrete* topology? So many choices for U... Let's try $U := \{a\}$! Then, $x_n \in U$ means simply that $x_n = a$. So, the only sequences that converge to a are those that *are equal* to a, eventually:

- $x_1, x_2, ..., x_k, a, ..., a, ...$ converges to a; however,
- $1, 1/2, 1/3, ..., 1/n, ...$ diverges. □

Example 3.34. When do we have $x_n \to a$ in *anti-discrete* topology? There is only one choice for us, $U := X$! But then $x_n \in U$ means that there is no constraint whatsoever. Consequently, *every* sequence converges ... to *every* point ... at the same time. Yes, even this one:

$$\lim_{n \to \infty} 1 = 0.$$ □

Just as the analysis of path-connectedness in the last section, these examples are intended to demonstrate how little we can rely on intuition and visualization at this level of abstraction. They also set the standard of proof in point-set topology.

Exercise 3.35. What condition on the topology of X would guarantee the uniqueness of limits?

Theorem 3.36. *Continuous functions preserve convergence*; i.e., given $f : X \to Y$ which is continuous, if $x_n \to a$ as $n \to \infty$ in X then $f(x_n) \to f(a)$ as $n \to \infty$ in Y.

Exercise 3.37. Prove the theorem.

Exercise 3.38. Given a sequence $\alpha = \{x_n : \ n = 1, 2, ...\}$ in a topological space X. View the sequence as a *subset* $A = \{x_n : \ n \in \mathbf{N}\}$ of X and compare the meanings of the limit points of A and the accumulation points of A. What about the limit of α?

Exercise 3.39. Suppose we have a set X. Suppose also a collection α of sequences in X is given and each sequence q in α is associated with a point $x_q \in X$. What is an appropriate topology for X? Construct a topology on X such that $p \to x_p$ for any $p \in \alpha$ and the rest diverge. Hint: what constraints on α would guarantee that this is possible?

3.5 Metric spaces

Notice that, in the Euclidean space, convergence of *points*

$$x_n \to a$$

simply means convergence of certain *numbers* (the distances to the limit point):

$$||x_n - a|| \to 0.$$

These numbers converges, in $\mathbf{R}$. Then, in a given topological space, can we express convergence of its sequences in terms of an appropriate *distance function*?

Definition 3.40. Given a set X, any function

$$d : X \times X \to \mathbf{R}$$

is called a *metric* (or a "distance function") if it satisfies the following conditions, for every $x, y, z \in X$,

- (M1) *positivity*: $d(x, y) \geq 0$, with equality if and only if $x = y$;
- (M2) *symmetry*: $d(x, y) = d(y, x)$;
- (M3) *triangle inequality*: $d(x, z) \leq d(x, y) + d(y, z)$.

In this case, the pair (X, d) is called a *metric space.*

For $a \in X$ and $\varepsilon > 0$, the *open ball around* a *of radius* ε is the set

$$B(a, \varepsilon) := \{x \in X : \ d(x, a) < \varepsilon\}.$$

These sets form a basis of neighborhoods γ_m for X. Indeed, (B2) follows from (M3). Then the topology on X generated by γ_m, or any other metric, is called a *metric topology.*

The Euclidean space $\mathbf{R}^n$ is a metric space with the metric defined by

$$d(x, y) = d_E(u, v) := ||u - v||.$$

Example 3.41 (taxicab metric). Let's compare the Euclidean metric on the plane,

$$d_E(u, v) := \sqrt{|u_1 - v_1|^2 + |u_2 - v_2|^2}$$

to the *taxicab metric* (aka "Manhattan metric"). To find the distance between two vectors, we go *along the grid*:

$$d_M(u,v) := |u_1 - v_1| + |u_2 - v_2|.$$

Taxicab's circle is a square... but, as we know, the squares generate the same topology as the disks: *the geometry is different but the topology is the same!* □

Exercise 3.42. The *post office metric* is based on the idea that postman's has to always go through the post office. Find the formula and prove that this is a metric.

Exercise 3.43. Prove that (a) the discrete topology is generated by this trivial metric:

$$d_D(x,y) := \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y; \end{cases}$$

while (b) the anti-discrete topology (with more than one point) cannot be generated by a metric.

Exercise 3.44. Suppose d is a metric on X. Prove that the function $d_1 : X \times X \to \mathbf{R}$ given by

$$d_1(x,y) := \min\{d(x,y), 1\}$$

is also a metric. Are they equivalent?

3.6 Spaces of functions

Notation: Let $C(X)$ stand for the set of all real-valued continuous functions on X.

Consider $C[a,b]$,the set of all continuous functions $f : [a,b] \to \mathbf{R},\ b > a$.

One may equip this set with a variety of topologies, starting with discrete and anti-discrete. However, the main "meaningful" examples come from the two primary ways sequences of functions may converge.

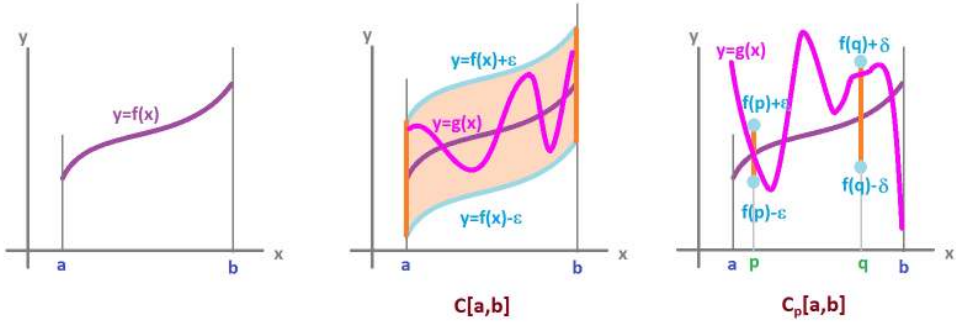
The first is the uniform convergence: a sequence $f_n,\ n = 1,2,...,$ *uniformly converges* to f if

$$\max_{x\in[a,b]} |f_n(x) - f(x)| \to 0.$$

This convergence is produced by the *sup-metric* (or sup-norm) on $C[a,b]$:

$$d_S(f,g) := \max_{x\in[a,b]} |f(x) - g(x)|.$$

The picture below (middle) shows an ε-neighborhood of f: a function g can behave arbitrarily as long as it stays within the "tunnel" with the two "walls" determined by: $y = f(x) - \varepsilon$ and $y = f(x) + \varepsilon$.



Exercise 3.45. Why is the sup-norm well-defined?

The following result, which we accept without proof, underscores the importance of this topology and the reason why it is the default choice.

Theorem 3.46. The uniform limit of continuous functions is continuous.

The second way is the point-wise convergence: a sequence f_n, $n = 1, 2, ...$, *point-wise converges* to f if

$$|f_n(x) - f(x)| \to 0, \ \forall x \in [a, b].$$

The picture above (right) shows a neighborhood of f: a function g can behave arbitrarily as long as it passes through, not a "tunnel", but a series of "gates", with these "posts": $f(p) - \varepsilon$ and $f(p) + \varepsilon$, $f(q) - \delta$ and $f(q) + \delta$, etc.

Exercise 3.47. Explain why the depicted set is a neighborhood.

There can be multiple values of x for which the "gates" are built, with different widths, but not infinitely many! That's why they can't hold the functions as tight as a "tunnel". There is no metric for this topology.

Example 3.48. The point-wise limit of continuous functions doesn't have to be continuous. Just choose:

$$a := 0, \ b := 1, \ f_n(x) := x^n,$$

and

$$f(x) := \begin{cases} 0 & \text{if } x \in [0, 1), \\ 1 & \text{if } x = 1. \end{cases}$$

The limit function is discontinuous! □

Exercise 3.49. (a) Prove that uniform convergence implies point-wise convergence. (b) Give an example of a sequence $f_n \in C[a, b]$ with

- $f_n \to f \in C[a, b]$ point-wise, but
- $f_n \not\to f \in C[a, b]$ uniformly.

Exercise 3.50. (a) Define a metric on $C[a, b]$ based on the definite integral. (b) Compare it to the uniform and point-wise convergence.

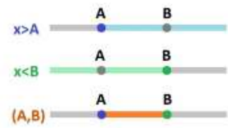
3.7 The order topology

The standard basis of the topology of the reals consists of all open intervals:

$$\{(a, b) : \ a, b \in \mathbf{R}, a < b\}.$$

But what is an interval? It is a set of all elements in $\mathbf{R}$ that lie strictly *between* two given elements of $\mathbf{R}$. That's not a coincidence: this is how a basis is formed for all partially ordered sets.

The three types of intervals in a linearly ordered set, such as $\mathbf{R}$ or $\mathbf{Z}$, are illustrated below:



What about partially ordered sets? Recall that a *partial order* is a binary relation "$\leq$" between pairs of elements of a set X that satisfies these three axioms:

- 1. *reflexivity*: $x \leq x$;
- 2. *antisymmetry*: if $x \leq y$ and $y \leq x$ then $x = y$;
- 3. *transitivity*: if $x \leq y$ and $y \leq z$ then $x \leq z$.

Naturally, we write $x < y$ when $x \leq y$ and $x \neq y$. Then X is called a *partially ordered set* or poset.

Given an order on a set X, the basis β of topology of X is chosen to be the collection of all intervals of all three types:

- $(A, \infty) = \{x \in X : x > A\}$,
- $(-\infty, B) = \{x \in X : x < B\}$,
- $(A, B) = \{x \in X : x > A, x < B\}$.

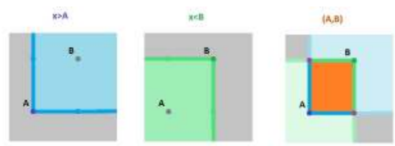
It is called the *order topology generated by the order.*

Exercise 3.51. Prove that β is a basis.

As an example, $\mathbf{R}^2$ and $\mathbf{Z}^2$ are ordered by:

$$(a, b) \leq (a', b') \text{ if and only if } a \leq a' \text{ and } b \leq b'.$$

Consider the three types of intervals in $\mathbf{R}^2$:



Observe that, except for the points A, B themselves, the (Euclidean) boundaries of these regions are included in the open intervals!

Exercise 3.52. Show how "open rays" define the same topology.

Exercise 3.53. Define a partial order on $C[a, b]$ and compare the order topology with the sup-metric topology.

Above we used for $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ and $\mathbf{Z}^2 = \mathbf{Z} \times \mathbf{Z}$ the *product order* on the product set $X \times Y$ set of two ordered sets X, Y:

$$(a, b) \leq (a', b') \text{ if and only if } a \leq a' \text{ and } b \leq b'.$$

Alternatively, this is how the product $X \times Y$ set of two ordered sets X, Y is ordered via the so-called the *lexicographical order*:

$$(a, b) \leq (a', b') \text{ if and only if } a < a' \text{ or } (a = a' \text{ and } b \leq b').$$

Exercise 3.54. Present the intervals with respect to the lexicographical order on $\mathbf{R}^2$.

Example 3.55. A different kind of ordered set is any group of people ordered by their ancestor-descendant relations: $A > B$ if A is a direct descendant of B. Then some typical open intervals are:

- (parent, child) $= \emptyset$;

- (grandparent, grandchild) = {parent};
- (husband, wife) = $\emptyset$;
- (cousin, cousin) = $\emptyset$; etc. □

Exercise 3.56. Given an arbitrary set X, let P be any collection of subsets of X. Prove that P is partially ordered by inclusion and describe the intervals of its order topology.

---

4 Continuous functions

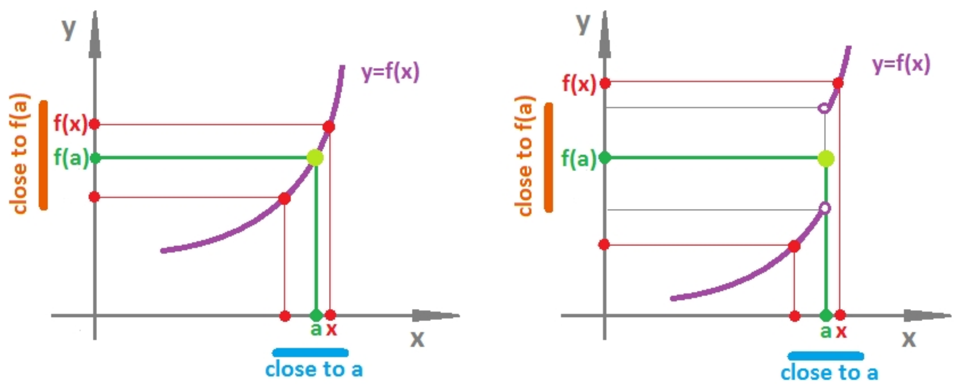
4.1 Continuity as preservation of proximity

We are not done with continuity!

Let's review the stages our definition of continuity went through.

First we, informally, discussed continuity of a function as a transformation that does not tear things apart and interpreted this idea in terms of closeness (proximity) of the input values vs. that of the output values:

if x is close to a then $f(x)$ is close to $f(a)$:



Then, this informal idea brought us to the calculus definition of continuity of f at point $x = a$:

$$\text{for any } \varepsilon > 0 \text{ there is a } \delta > 0 \text{ such that } |x - a| < \delta \Longrightarrow |f(x) - f(a)| < \varepsilon,$$

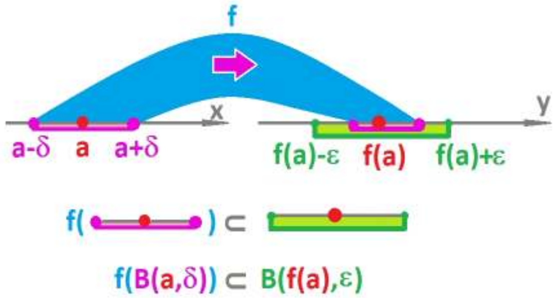
and, further, the absolute value replaced by the norm:

$$\text{for any } \varepsilon > 0 \text{ there is a } \delta > 0 \text{ such that } ||x - a|| < \delta \Longrightarrow ||f(x) - f(a)|| < \varepsilon.$$

Next we realized that these inequalities are simply inclusions:

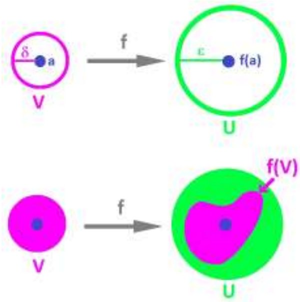
$$\text{for any } \varepsilon > 0 \text{ there is a } \delta > 0 \text{ such that } x \in B(a, \delta) \Longrightarrow f(x) \in B\big(f(a), \varepsilon\big),$$

where $B(p, d) = \{u : \ ||u - p|| < d\}$ is an open ball in $\mathbf{R}^n$:



Furthermore, using the simple concept of the image of subset, we rewrote the definition one more time:

$$\text{for any } \varepsilon > 0 \text{ there is a } \delta > 0 \text{ such that } f\big(B(a,\delta)\big) \subset B\big(f(a),\varepsilon\big).$$



As the last step, we interpreted these Euclidean balls as "neighborhoods" of points, i.e., elements of a basis that generates the topology. In the purely topological setting that we have developed, the definition becomes as follows.

Definition 4.1. Given two sets X, Y with bases of neighborhoods γ_X, γ_Y, a function $f : X \to Y$ is continuous if

$$\text{for any } U \in \gamma_Y \text{ there is a } V \in \gamma_X \text{ such that } f(V) \subset U.$$

We further explore this definition next.

We start with a simple observation that if we replace neighborhoods with open sets in the definition, it would still guarantee that the function is continuous. In fact, the converse is also true.

Theorem 4.2. Given two topological spaces X, Y, a function $f : X \to Y$ is continuous if and only if

$$\text{for any open } U \subset Y \text{ there is an open } V \subset X \text{ such that } f(V) \subset U.$$

Exercise 4.3. Prove the theorem.

4.2 Continuity and preimages

The concept of image has proven to be very convenient, but the following, related, concept is even more fruitful in this context.

Definition 4.4. For any function $f : X \to Y$ the *preimage of subset* $B \subset Y$ under f is defined to be

$$f^{-1}(B) := \{x : f(x) \in B\}.$$

This set is defined explicitly and is independent of the existence of the inverse function f^{-1}. However, then f^{-1} dies exist, $f^{-1}(B)$ is indeed the image of B under f^{-1}.

To appreciate the difference, consider this diagram of consecutive functions:

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} U \longrightarrow ...$$

Here, we can trace both points and sets in the forward direction:

$$\begin{array}{llll} x \mapsto y & := f(x) \mapsto z & := g(y) \mapsto u & := h(z) \mapsto ..., \\ A \mapsto B & := f(A) \mapsto C & := g(B) \mapsto D & := h(C) \mapsto ..., \end{array}$$

but, since the preimages of points don't have to be points, we can only follow *sets* in the reverse direction:

$$f^{-1}(T) \leftarrow T := g^{-1}(R) \leftarrow R := h^{-1}(S) \leftarrow S...$$

To be sure, preimages appear implicitly in many areas of mathematics. For example, the solution set of the equation $f(x) = b$ is $f^{-1}(\{b\})$. In particular, the kernel of a linear operator is the preimage of the zero:

$$\ker f = f^{-1}(0).$$

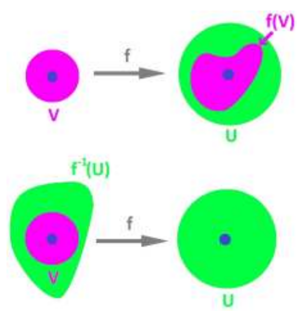
We will now consider *the preimage of an open set* under a continuous function.

This is actually quite simple: the inclusion

$$f(V) \subset U$$

of the definition can now be rewritten as

$$V \subset f^{-1}(U):$$



The fact that, according to the definition, such a V exists for any point in $f^{-1}(U)$ suggests that this set is *open*.

Theorem 4.5. Suppose $f : X \to Y$ is continuous. Then, if $U \subset Y$ is open then so is $f^{-1}(U)$.

Proof. Suppose U is open in Y. Let $a \in f^{-1}(U)$ be arbitrary. Let $b := f(a)$. Then, $b \in U$. Since U is open, b has a neighborhood E that lies entirely in U:

$$b \in E \subset U.$$

Now, the last version of the definition of continuity applies: there is a neighborhood D of a such that $f(D) \subset E$. The last inclusion can be rewritten,

$$D \subset f^{-1}(E),$$

and, due to $E \subset U$, we have

$$D \subset f^{-1}(U).$$

Hence, $f^{-1}(U)$ is open. ■

Warning: the image of an open set does not have to be open, as the example of a constant function shows.

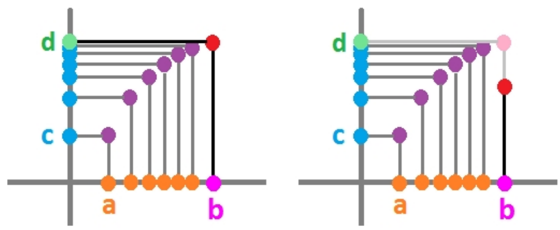
We've gone far from the original definition of continuity! Let's illustrate the new idea, "preimages of open sets are open", by trying to think of a function f with

- open range, such as (c,d), but
- closed domain, such as $[a,b]$.

Then one can imagine a scenario:

$$d_n \to d \quad \rightsquigarrow \quad d_n = f(b_n) \quad \rightsquigarrow \quad b_n \to b \quad \rightsquigarrow \quad f(b) \neq d.$$

So, the values of f approach d but never reach it, thus exhibiting a familiar discontinuous behavior.



The converse of the last theorem is also true:

Theorem 4.6. Suppose a function $f : X \to Y$ satisfies: if $U \subset Y$ is open then so is $f^{-1}(U)$. Then f is continuous.

Exercise 4.7. Prove the theorem.

Exercise 4.8. Prove that f is continuous if and only if the preimage of any closed set is closed.

Exercise 4.9. Is the image of a closed set under a continuous function closed?

Exercise 4.10. Prove that the image of a closed set under a continuous *periodic* function from $\mathbf{R}$ to $\mathbf{R}$ is closed (i.e., the function is "closed").

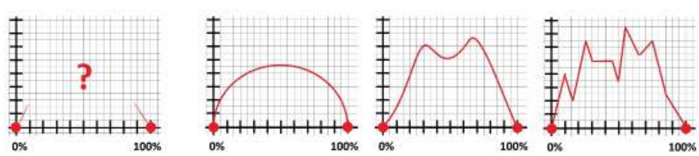
Exercise 4.11. Suppose $f : X \to Y$ is continuous. Prove

- for any $A \subset X$, $f(\text{Cl}(A)) \subset \text{Cl}(f(A))$, and
- for any $B \subset Y$, $\text{Cl}(f^{-1}(B)) \subset f^{-1}(\text{Cl}(B))$.

4.3 Continuous functions everywhere

A typical continuity argument is as follows.

Example 4.12 (Laffer curve). The exact dependence of the revenue on the tax rate is unknown. It is, however, natural to assume that both 0% and 100% rates will produce zero revenue.



If we also assume that the dependence is *continuous*, we conclude that increasing the rate may yield a decrease in revenue. □

Exercise 4.13. Show that the two hands of a clock will meet at least once a day. What if the clock is malfunctioning?

In our study, we will borrow results from calculus.

First, we have a long list of continuous functions.

Proposition 4.14. The following functions are continuous on their domains:
- polynomials;
- rational functions;
- $\sin x, \cos x, \tan x$, and other trigonometric functions;
- $e^x, \ln x$, and other exponential and logarithmic functions.

Second, we know that the four algebraic operation on functions preserve continuity – within the intersection of their domains – as follows.

Proposition 4.15. If two functions $f : A \to \mathbf{R}, g : B \to \mathbf{R},\ A, B \subset \mathbf{R}$, are continuous, then so are
- $f + g : A \cap B \to \mathbf{R}$;
- $f - g : A \cap B \to \mathbf{R}$;
- $f \cdot g : A \cap B \to \mathbf{R}$;
- $f/g : A \cap B \cap \{x : g(x) \neq 0\} \to \mathbf{R}$.

Exercise 4.16. Interpret the proposition in terms of $C[a, b]$ (assume $A := B := [a, b]$). These operations make $C[a, b]$ into?...

Third, the compositions of functions preserve continuity, as long as the domains and range match, as follows.

Proposition 4.17. If two functions $f : A \to \mathbf{R},\ g : B \to \mathbf{R},\ B \subset \mathbf{R}$, are continuous, then so is $gf : A \to \mathbf{R}$ provided $f(A) \subset B$.

Exercise 4.18. Relax the last condition.

Fourth, the minimum and the maximum preserve continuity.

Proposition 4.19. If two functions $f, g : X \to \mathbf{R}$, are continuous, then so are these:

$$M(x) := \max\{f(x), g(x)\}, \quad m(x) := \min\{f(x), g(x)\}.$$

Exercise 4.20. Prove the proposition.

Fifth, a function of several variables is continuous if and only if it is continuous with respect to each of its variables. And a vector-valued function is continuous if and only if every of its coordinate functions is continuous. And so on.

More advanced calculus results are the following.

Proposition 4.21. *Evaluation is continuous*; i.e., given a $c \in [a, b]$, the function $\Psi : C[a, b] \to \mathbf{R}$ defined by

$$\Psi(f) := f(c)$$

is continuous.

Proposition 4.22. *Definite integration is continuous*; i.e., the function $\Phi : C[a, b] \to \mathbf{R}$ given by

$$\Phi(f) := \int_a^b f(x)dx$$

is continuous.

Exercise 4.23. (a) Prove the two propositions – for the topology of uniform convergence. (b) Consider their analogs for the point-wise convergence. (c) What about anti-differentiation? (d) What about differentiation?

Definition 4.24. The *graph of a function* $f : X \to Y$ is the set

$$\operatorname{Graph} f := \{(x, y) \in X \times Y : \ y = f(x)\}.$$

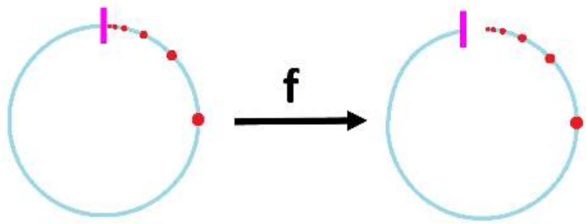
Proposition 4.25. The graph of a continuous function $f : \mathbf{R} \to \mathbf{R}$ is closed in $\mathbf{R}^2$.

Exercise 4.26. Prove the proposition.

Proposition 4.27. In case of metric spaces, a function $f : X \to Y$ is continuous if and only if *the function commutes with limits*; i.e.,

$$\lim_{n\to\infty} f(x_n) = f\Big(\lim_{n\to\infty} x_n\Big),$$

for any convergent sequence $\{x_n\}$ in X.



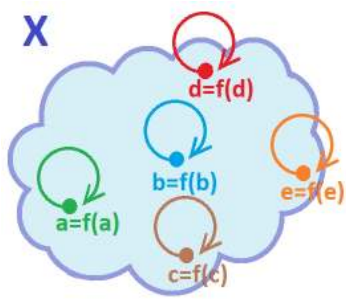
Exercise 4.28. Prove the proposition.

We know that continuity depends on the choice of topology. Nonetheless, there are functions – between topological spaces – that are *always* continuous.

Terminology: In the context of the above examples, the term commonly used is *continuous functions* but in topology it is *maps*. The latter will be used predominantly throughout.

Definition 4.29. Suppose we have a topological space X, then we can define the *identity function* $\operatorname{Id}_X : X \to X$ by

$$\operatorname{Id}_X(x) := x \ \text{ for all } x \in X.$$



Proposition 4.30. The identity function is continuous.

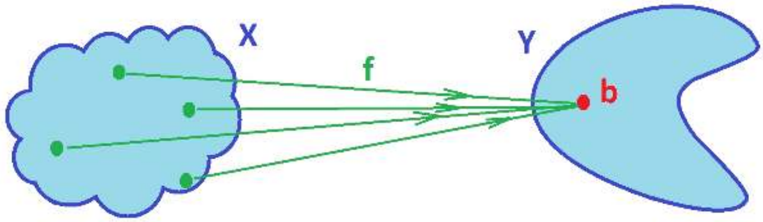
Proof. For every open U,

$$f^{-1}(U) = U$$

is open. ■

Definition 4.31. Suppose we have topological spaces X and Y, then for any given $b \in Y$, we can define a *constant function* $f : X \to Y$ by

$$f(x) = b \text{ for all } x \in X \text{ and some fixed } b \in Y.$$



Proposition 4.32. Any constant function is continuous.

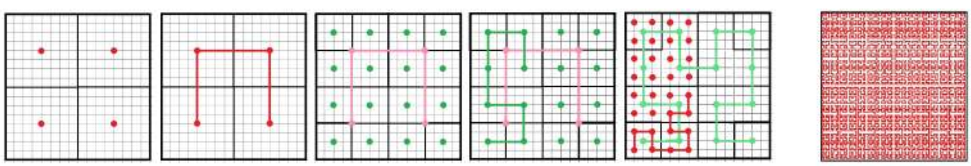
Proof. Suppose U is open in Y. Then we have two cases:
- Case 1: $b \in U \Longrightarrow f^{-1}(U) = X$, which is open.
- Case 2: $b \in X \setminus U \Longrightarrow f^{-1}(U) = \emptyset$, which is open too. ■

Exercise 4.33. Under what circumstances is the function $f : X \to X$ given by (a) $f(x) = x$, or (b) $f(x) = c$ for a fixed $c \in X$, continuous?

Exercise 4.34. Suppose X is a topological space.
- Suppose $f : X \to \mathbf{R}$ is continuous. Show that the solution set of any equation with respect to f, i.e., $\{x \in X : f(x) = r\}$ for some $r \in \mathbf{R}$, is closed.
- Show that this conclusion fails for a general continuous $f : X \to Y$, where Y is an arbitrary topological space.
- What condition on Y would guarantee that the conclusion holds?

Continuity isn't synonymous with good behavior.

Example 4.35 (space filling curve). These continuous curves are constructed iteratively and we will accept without proof that the limit is also continuous:



We will later demonstrate however that even such a monstrous function can be slightly "deformed" so that, while still continuous, it isn't onto anymore. □

Exercise 4.36. Given $g_a(x) = x^3 - x + a$, $a \in \mathbf{R}$, let $f(a)$ be the largest root of g_a. Is f continuous?

4.4 Compositions and path-connectedness

This is our new definition.

Definition 4.37. A function $f : X \to Y$ is *continuous* if for any open U, its preimage $f^{-1}(U)$ is also open.

It isn't only much more compact than the original but also allows us to make some theorems quite transparent.

Theorem 4.38. The composition of continuous functions is continuous.

Proof. Suppose we have two continuous functions $f : X \to Y$ and $g : Y \to Z$. Let $h : X \to Z$

be their composition: $h = gf$. Observe first that, in the commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {}_{h=gf}\searrow & \downarrow g \\ & & Z, \end{array}$$

we can trace back any set $U \subset Z$:

$$h^{-1}(U) = (gf)^{-1}(U) = f^{-1}(g^{-1}(U)).$$

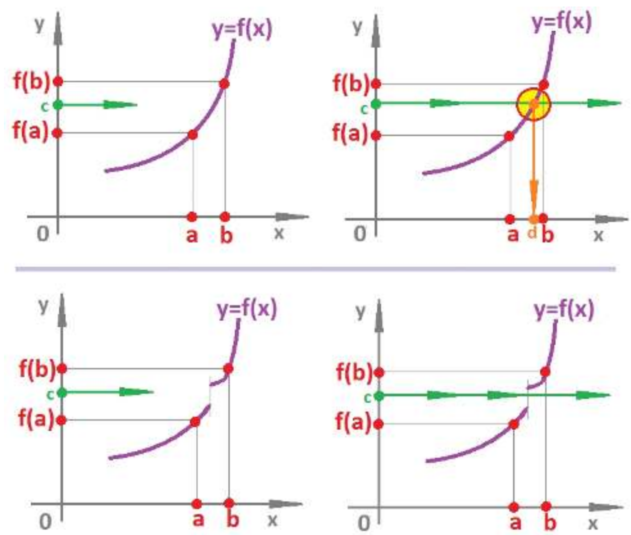
Suppose U is open in Z. Then $V := g^{-1}(U)$ is open in Y, since g is continuous. Then $W := f^{-1}(V)$ is open in X, since f is continuous. Therefore, the set

$$h^{-1}(U) = f^{-1}(g^{-1}(U)) = f^{-1}(V) = W$$

is open in X; h is continuous. ■

Our original understanding of a continuous function in calculus was as one with no gaps in its graph. The idea was confirmed (without proof) by the following.

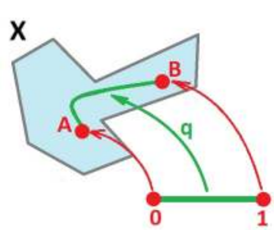
Theorem 4.39 (Intermediate Value Theorem). Suppose $f : [a, b] \to \mathbf{R}$ is continuous. Then, for any c between $f(a)$ and $f(b)$, there is a $d \in [a, b]$ such that $f(d) = c$.



We are ready to prove it now, in a more general context.

Recall the definition: a topological space Q is path-connected if for any $A, B \in Q$ there is a continuous function $p : [0, 1] \to Q$ such that

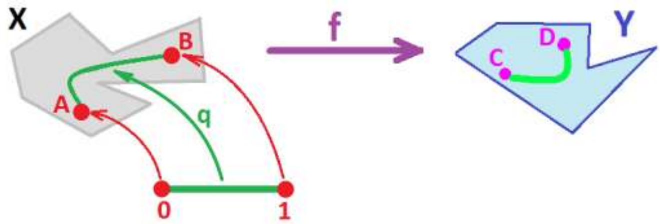
$$p(0) = A \text{ and } p(1) = B.$$



Exercise 4.40. Prove first that $[a, b]$ is path-connected, and so are $(a, b), [a, b)$.

Theorem 4.41. *Path-connectedness is preserved* by continuous functions; i.e., given $f : X \to Y$ which is continuous and onto, if X is path-connected then so is Y.

Proof. The idea is to restate the problem of finding the necessary path in Y as a problem of finding a certain path in X:



Then, the problem of finding a path from A to B can be solved in X. But we are facing the problem of finding a path from C to D in Y instead! Where would it come from? The answer is, from X via f.

Since f is onto, every point is Y comes from X, hence there are $A, B \in X$ such that

$$C = f(A),\ D = f(B).$$

We then use the path-connectedness of X and find a continuous function

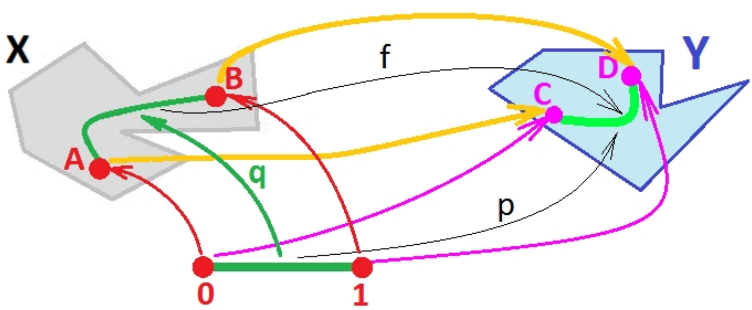
$$q : [0, 1] \to X$$

such that

$$q(0) = A,\ q(1) = B.$$

Now choose

$$p := fq.$$



To sum up,

- 1. p is continuous as the composition of two continuous functions (last theorem),
- 2. $p(0) = f(q(0)) = f(A) = C$,
- 3. $p(1) = f(q(1)) = f(B) = D$.

Thus, function p satisfies the above definition. ■

The proof is summarized in this commutative diagram:

$$\begin{array}{ccc} \{0,1\} & \xrightarrow{q} & X \\ & _{p=fq} \searrow & \downarrow f \\ & & Y \end{array} \quad \leadsto \quad \begin{array}{ccc} [0,1] & \xrightarrow{q} & X \\ & _{p=fq} \searrow & \downarrow f \\ & & Y \end{array}$$

Exercise 4.42. Derive the Intermediate Value Theorem from the last theorem.

Exercise 4.43. A topological space is called *connected* if it can't be represented as a disjoint union of two closed sets. Prove that every path-connected space is connected.

4.5 Categories

The results we have proven about topological spaces and continuous functions, especially the one about the composition, have a far-reaching generalization.

Suppose we have a collection $\mathscr{C}$ that consists of
- a collection $\operatorname{Obj}(\mathscr{C})$ of *objects*;
- a collection $\operatorname{Hom}(\mathscr{C})$ of *morphisms* between some pairs of objects.

Morphisms are also called "arrows" sometimes, which emphasizes the point that they are not functions just as the objects aren't sets.

Each morphism f has unique source object $X \in \operatorname{Obj}(\mathscr{C})$ and target object $Y \in \operatorname{Obj}(\mathscr{C})$. We write then $f : X \to Y$.

Then, the category as a whole can be illustrated by this diagram:

$$\mathscr{C} = \left\{ \begin{array}{cl} X & \in \operatorname{Obj}(\mathscr{C}) \\ \downarrow f & \in \operatorname{Hom}(\mathscr{C}) \\ Y & \in \operatorname{Obj}(\mathscr{C}) \end{array} \right\}$$

We write $\operatorname{Hom}_{\mathscr{C}}(X, Y)$ (or simply $\operatorname{Hom}(X, Y)$ when there is no confusion about to which category it refers) to denote the class of all morphisms from X to Y.

Suppose we have three objects $X, Y, Z \in \operatorname{Obj}(\mathscr{C})$. Then, for any two morphisms $f : X \to Y$ and $g : Y \to Z$, there is a third called their *composition* written as $gf : X \to Z$. The composition completes the following commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {}_{gf}\searrow & \downarrow g \\ & & Z \end{array}$$

Definition 4.44. The collection $\mathscr{C}$ is a *category* if the following axioms hold:
- *identity*: for every object $X \in \operatorname{Obj}(\mathscr{C})$, there exists a morphism $\operatorname{Id}_X \in \operatorname{Hom}_{\mathscr{C}}(X, X)$, called the *identity morphism* for X, such that for every morphism $f : X \to Y$, we have

$$\operatorname{Id}_Y f = f = f \operatorname{Id}_X .$$

- *associativity*: if $f : X \to Y, g : Y \to Z$ and $h : Z \to W$ then $h(gf) = (hg)f$.

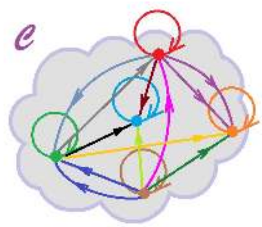
The axioms can be illustrated with diagrams. For the former, it's simply this:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & f \searrow & \downarrow \operatorname{Id}_Y \\ & & Y \end{array} \qquad \begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow \operatorname{Id}_X & \nearrow f & \\ X & & \end{array}$$

For the latter, it's a commutative diagram of commutative diagrams:

$$
\begin{array}{ccccc}
\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & & \downarrow g \\ W & \xleftarrow{h} & Z \end{array} & \Longrightarrow & \begin{array}{ccc} X & & Y \\ & \searrow^{gf} & \\ W & \xleftarrow{h} & Z \end{array} \\
\Downarrow & & \Downarrow \\
\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \swarrow hg & \\ W & & Z \end{array} & \Longrightarrow & \begin{array}{ccc} X & & Y \\ \downarrow hgf & & \\ W & & Z \end{array}
\end{array}
$$

The axioms justify the illustration of objects and morphisms below:



Exercise 4.45. Prove that these are categories with morphisms as indicated:
- sets with functions,
- topological spaces with continuous functions,
- groups with group homomorphisms,
- rings with ring homomorphisms,
- vector spaces with linear operators,
- graphs with graph maps.

Exercise 4.46. Prove that there is exactly one identity morphism for every object.

Exercise 4.47. Define isomorphisms on a category in such a way that it incorporates group isomorphisms when the category is groups.

The axioms appear similar to those of groups. Indeed, the composition is a binary operation:

$$\text{Hom}_{\mathscr{C}}(X, Y) \times \text{Hom}_{\mathscr{C}}(Y, Z) \to \text{Hom}_{\mathscr{C}}(X, Z).$$

The difference is that morphisms don't have to produce a binary operation on the whole $\text{Hom}(\mathscr{C})$ because not all compositions have to be available. The reverse connection is valid though as this example of a category $\mathscr{C}$ produced from a group $(G, *)$ shows:
- $\text{Obj}(\mathscr{C}) = \{*\}$,
- $\text{Hom}(\mathscr{C}) = G$.

Exercise 4.48. Prove that this is a category.

4.6 Vector fields

This is how we commonly describe motion.

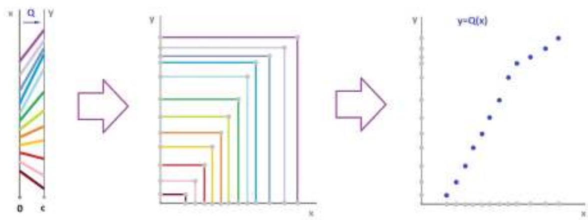
We take as a model a *fluid flow*. The *phase space* $S \subset \mathbf{R}^n$ of the system is the space of all possible locations. Then the position of a given particle is a function $x : I \to S$ of time t, where I is some interval. Meanwhile, the dynamics of the particle is governed by the velocity of the flow, at each location and each moment of time. Let $f : [0, \infty) \times S \to \mathbf{R}^n$ be a function and let the velocity at time t of a particle, if it happens to be at point x, be $f(t, x)$.

Let's start with dimension 1 and $S := \mathbf{R}$. Then this dynamics is given by an *ordinary differential equation* (ODE):

$$x'(t) = f(t, x(t)),$$

Suppose we have such a system in $\mathbf{R}$. Now, suppose we are given a current state a and a forward propagation time t. Then, by the Euler method, there is a map, called the *Euler map*, that approximates the solution of the ODE at a given point:

$$Q_t(a) = a + tf(a).$$



The Euler map is continuous if the right-hand side f of the ODE is continuous!

In dimension 2, we are talking about *vector fields.* A vector field is given when there is a vector attached to each point of the plane:

$$\text{point} \mapsto \text{vector}.$$

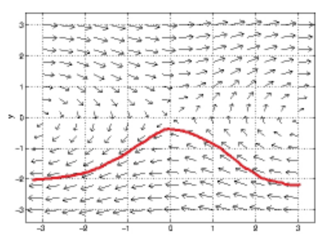
So, vector field is just a function

$$V : \mathbf{R}^2 \to \mathbf{R}^2.$$

The following vector field

$$V = (f_x, f_y)$$

could be the velocity field of a flow:



One can think of vector fields as systems of time-independent ODEs:

$$\begin{aligned} x'(t) &= f_x(x(t), y(t)), \\ y'(t) &= f_y(x(t), y(t)). \end{aligned}$$

This is still a model for a flow: $u = (x, y)$ is the position of a given particle as a function of time t and $(f_x(x, y), f_y(x, y))$ is the velocity of this particle (or any other) if it happens to be at point (x, y).

For any vector field in $\mathbf{R}^n$, the *Euler map*

$$Q_t : \mathbf{R}^n \to \mathbf{R}^n$$

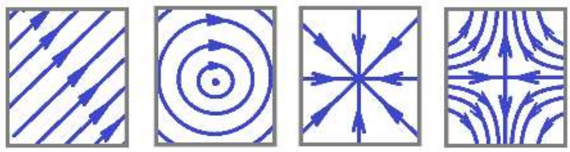
is defined by the same formula

$$Q_t(a) = a + tV(a).$$

Theorem 4.49. If V is continuous, the Euler map $Q_t : \mathbf{R}^n \to \mathbf{R}^n$ is continuous too, for each t.

Exercise 4.50. Prove this theorem.

This map is then studied to discover such properties of the system as the existence of stationary points:



Exercise 4.51. What can you say about the Euler map for the motion patterns above?

4.7 The dynamics of a market economy

Let's imagine that we have a "closed" market with n commodities freely traded. The possible prices,

$$p_1, ..., p_n,$$

of these commodities form a *price vector*,

$$p = (p_1, ..., p_n),$$

at each moment of time. Then, the *space of prices* is

$$P := \mathbf{R}^n_+ := \{(p_1, ..., p_n) : \ p_i \in \mathbf{R}_+\} \subset \mathbf{R}^n.$$

Suppose also that there are m "agents" in the market, each owning a certain amount of each commodity. We suppose that

- the ith agent owns c^i_j of the jth commodity.

These amounts form the ith *commodity vector*,

$$c^i \in \mathbf{R}^n_+.$$

Then, the ith agent's *wealth* is

$$w^i = \langle p, c^i \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the dot product. We suppose also that this agent wants to own a different combination of commodities. This vector, d^i, constitutes the demand of the agent. The demand, of course, depends on the current prices p.

One can argue now that the demand function $d^i = d^i(p)$ is *continuous*. The argument is based on the idea that the demand is elastic enough so that a small change in the price of a commodity will produce a small change in the person's demand for this commodity, as well as other commodities. Even though one-time purchases may require a certain threshold to be reached to become affordable, we argue that approaching this threshold is accompanied by an increase of the likelihood of the purchase. In other words, the person's desire for a certain commodity is unchanged but the affordability does change – continuously – with his wealth $w^i = \langle p, c^i \rangle$ as the price vector p changes.

It follows that the market-wide demand function $d = \sum_i d^i$ is continuous too. Now, the continuous change of the demand will have a continuous effect on the prices.

Next, let's examine the dynamics of prices.

First, we assume that *the tomorrow's prices depend only on the today's prices*. We need to justify this assumption. First, we think of the market as truly "closed" so that there is no outside influence on the prices by non-market events, such as wars, natural disasters, new laws and regulations etc. We can simply think that the time gap is so short that these events don't affect the prices. At the same time, the time gap has to be long enough for the information about price changes to disseminate to all participants. In addition, we assume that the prices don't depend on time – of the day, week, or year.

This dependence, then, is a *price function* $F : P \to P$. And the analysis above suggests that this *dependence is continuous*.

Exercise 4.52. What difference would price controls have?

What about the commodities? The total amount is the vector

$$\bar{c} := \sum_i c^i \in \mathbf{R}^n_+.$$

It may be fixed or, assuming continuous production and continuous consumption, it varies continuously. If we also assume that the trading is also continuous, we conclude that the commodity vectors $c^i,\ i = 1, 2, ..., m$, vary continuously as well. Then the dynamics is represented by a continuous function

$$G : C \to C,$$

where

$$C := \{(c^1, ..., c^m) :\ c^i \in \mathbf{R}^n_+\} = \left(\mathbf{R}^n_+\right)^m,$$

is the *space of distributions of commodities*.

Exercise 4.53. Incorporate into the model the money as a new commodity.

The *state* of the system is then a combination of the prices and a distribution of the commodities. Then the *state space* is

$$S := P \times C,$$

and the state dynamics is given by a continuous function $D : S \to S$. In other words, tomorrow's prices and distributions are determined by today's prices and distributions.

One can easily incorporate other quantities into this extremely oversimplified model. As an illustration, suppose the interest rates are to be included as they influence and are being influenced by the prices. Then the state space becomes $P \times C \times [0, \infty)$. Any new quantity can be added to this model by simply extending P in this fashion.

Exercise 4.54. What happens if we add currency exchange rates to this model? Suggest your own parameters.

A more abstract way to model motion and dynamics is a *discrete dynamical system*, which is simply any continuous function $F : P \to P$ on any topological space P, applied repeatedly:

$$x_{n+1} = F(x_n).$$

Dynamical systems are known for exhibiting a complex behavior even with a simple choice of F. Modelling is uncomplicated for dimension 1; one can simply use a calculator. Just enter a function and then repeatedly apply it. A spreadsheet will allow you visualize the dynamics for dimensions 1 and 2.

4.8 Maps

We know that path-connectedness is preserved under continuous transformations. This means that we can't *tear* things!

That's the reason why continuous functions are also called *maps* or mappings. After all, mapping is a process of transferring the terrain – point by point – to a piece of paper, a map.

This process should better be continuous!

What if it's not? Then we won't be able to answer the most basic question maps are intended to answer:

Can we get from point A to point B?

The simplest case is when the map itself is torn:

Such a "discontinuous" map might give the answer "No" while, in fact, the terrain it came from answers "Yes". A more subtle case is when the actual process of mapping was erroneous, such as when a peninsula is turned into an island:

Thus, the continuity of the mapping prevents it from tearing the map and guarantees that a "No" is correct.

Furthermore, even though we can't *tear* the map, we might still end up *gluing* its pieces. A "glued" map might give the answer "Yes" to our question while, in fact, the terrain it is supposed to depict answers "No". One might glue – by mistake – islands together, or even continents:

That's a different topologically – kind of mapping error!

Example 4.55. Consider these two examples of gluing. First, merging two points:

$$f : \{A, B\} \to \{C\},\ f(A) = f(B) = C,$$

and, second, parametrization of the circle:

$$f : [0, 2\pi] \to \mathbf{S}^1 \subset \mathbf{R}^2,\ f(t) = (\cos t, \sin t).$$ □

Exercise 4.56. Prove that these are continuous functions.

Thus, while a continuous map can't bring the number of components up, it can bring it down.

As we just saw, sometimes gluing is bad – even though it is a continuous operation. The reason is that *the inverse of gluing is tearing.* Then the match between the map and the terrain is inadequate.

That's why we will define a class of maps so that both the function

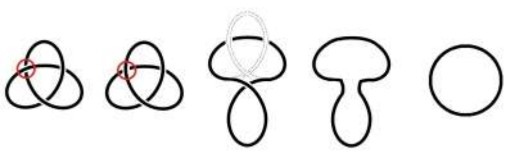
$$f : X \to Y$$

and its inverse

$$f^{-1} : Y \to X$$

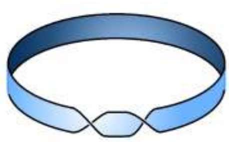
are continuous.

Note that even though tearing isn't allowed, we can cut – if you glue it back together exactly as before. For example, this is how you can un-knot this knot:



One can achieve this without cutting by manipulating the knot in the 4-dimensional space.

Exercise 4.57. What is this, topologically?



4.9 Homeomorphisms

Let's make this idea more precise. Recall a couple of definitions.

A function $f : X \to Y$ is called *one-to-one*, or injective, if

$$f(x) = f(y) \Longrightarrow x = y.$$

Or, the preimage of a point, if non-empty, is a point.

A function $f : X \to Y$ is called *onto*, or surjective, if

$$\text{for every } y \in Y \text{ there is an } x \in X \text{ such that } y = f(x).$$

Or, the image of the domain space is the whole target space, $f(X) = Y$.

A function that is one-to-one and onto is also called *bijective*.

Theorem 4.58. A bijective function $f : X \to Y$ has the *inverse*, i.e., a unique function $f^{-1} : Y \to X$ such that $f^{-1}f = \mathrm{Id}_X$ and $ff^{-1} = \mathrm{Id}_Y$.

Definition 4.59. Suppose X and Y are topological spaces and $f : X \to Y$ is a function. Then f is called a *homeomorphism* if
- f is bijective,
- f is continuous
- f^{-1} is continuous.

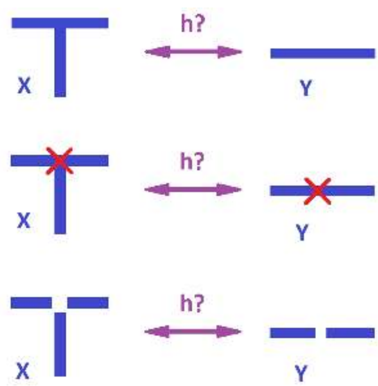
Then X and Y are said to be *homeomorphic*, as well as *topologically equivalent*.

The word "homeomorphism" comes from the Greek words "homoios" for "similar" and "morphō" for "shape" or "form". One may understand the meaning of this word as follows: the shapes of two figures are *similar* by being topologically *identical*.

Let's examine informally how the definition may be applied to demonstrate that spaces are *not* homeomorphic.

Example 4.60. Consider X a segment and Y two segments. We already know that the number of path-components is supposed to be preserved. Therefore, these two aren't homeomorphic. □

Example 4.61. Consider now X letter "T" and Y a segment. Both are path-connected so this information doesn't help. The question remains, are they homeomorphic? One trick is to pick a point a in X, remove it, and see if $X \setminus \{a\}$ is homeomorphic to $Y \setminus \{b\}$ for *any* $b \in Y$. Let's choose a to be in the cross of "T":



Removing this point creates a mismatch of components regardless what b is.

Therefore, the method is to examine how many components can be produced by removing various points from the space. □

Exercise 4.62. Using just this approach classify the symbols of the standard computer keyboard (think of them as if they are made of wire):
- ‘ 1 2 3 4 5 6 7 8 9 0 - =

- q w e r t y u i o p [ ]
- a s d f g h j k l ; ’
- z x c v b n m , . /

and

- ~ ! @ # $ % ^& * ( ) _ +
- Q W E R T Y U I O P {} —
- A S D F G H J K L : “
- Z X C V B N M < > ?

More general is this almost trivial fact.

Proposition 4.63. If spaces X and Y are homeomorphic, then for any $a \in X$ there is a $b \in Y$ such that $X \setminus \{a\}$ is homeomorphic to $Y \setminus \{b\}$.

Exercise 4.64. Prove the proposition. Suggest an example when counting components, as we did above, is no longer sufficient, but the proposition still helps.

Proposition 4.65. An *isometry*, i.e., a bijection between two metric spaces that preserves the distance, is a homeomorphism.

Exercise 4.66. (a) Prove the proposition. (b) Show that self-isometries form a group.

Example 4.67 (non-example). It might be hard to quickly come up with an example of *a continuous bijection that isn't a homeomorphism.* The reason is that our intuition always takes us to the Euclidean topology. Instead, consider the identity function for any set X,

$$f = \mathrm{Id} : (X, \tau) \to (X, \kappa),$$

where κ is the anti-discrete topology and τ isn't. Then, on the one hand,

- $f^{-1}(X) = X$ is open in (X, τ),

so that f is continuous. On the other hand, there is a proper subset A of X open in τ, but

- $f(A) = A$ isn't open in (X, κ)

Therefore, $f^{-1} : (X, \kappa) \to (X, \tau)$ isn't continuous. A specific example is:

- $X = \{a, b\}$,
- $\tau = \big\{\emptyset, \{a\}, \{b\}, X\big\}$,
- $\kappa = \big\{\emptyset, \{a\}, X\big\}$.

This example suggests a conclusion: the identity function will be a desired example when κ is a proper subset of τ. Consider, for example, $X = \mathbf{R}$, while τ is the Euclidean topology and κ any topology with fewer open sets (it's “coarser” or “sparser”), such as the topology of rays. □

Exercise 4.68. Provide another non-example: a continuous bijection $f : X \to Y$ that isn't a homeomorphism; this time both X and Y are subspaces of $\mathbf{R}^2$.

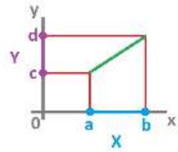
Exercise 4.69. We are given a continuous function $f : \mathbf{R} \to \mathbf{R}$. Define a function $g : \mathbf{R} \to \mathbf{R}^2$ by $g(x) = (x, f(x))$. Prove that g is continuous and that its image, the graph of f, is homeomorphic to $\mathbf{R}$.

Exercise 4.70. To what space is the Möbius band with the center line cut out homeomorphic?

4.10 Examples of homeomorphic spaces

Theorem 4.71. All closed intervals of a non-zero, finite length are homeomorphic.

Proof. Let $X := [a,b], Y := [c,d],\ b > a, d > c$. We will find a function $f : [a,b] \to [c,d]$ with $f(a) = c$ and $f(b) = d$. The simplest function of this kind is linear:



To find the formula, use the point-slope formula from calculus. The line passes through the points (a,c) and (b,d), so its slope is $m = \frac{d-c}{b-a} \neq 0$. Hence, the line is given by

$$f(x) = c + m(x - a).$$

We can also give the inverse explicitly:

$$f^{-1}(y) = a + \frac{1}{m} \cdot (y - b).$$

Finally we recall that linear functions are continuous. ∎

Theorem 4.72. Open intervals of finite length are homeomorphic.

Exercise 4.73. Prove the theorem.

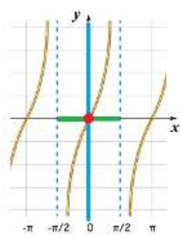
Theorem 4.74. An open interval is not homeomorphic to a closed interval (nor half-open).

Exercise 4.75. Prove the theorem. Hint: focus on the endpoint(s).

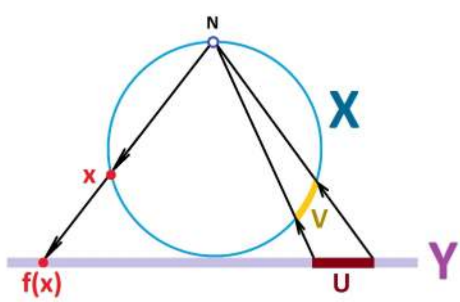
No interval of non-zero length can be homeomorphic to a point because there can be no bijective function between them.

Theorem 4.76. All open intervals, even infinite ones, are homeomorphic.

Proof. The tangent gives you a homeomorphism between $(-\pi/2, \pi/2)$ and $(-\infty, \infty)$. ∎



Example 4.77. Another way to justify this conclusion is given by the following construction:



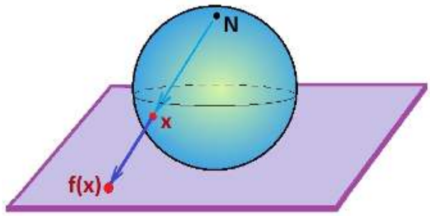
Here the "north pole" N is taken out from a circle to form X. Then X is homeomorphic to a finite open interval, and to an infinite interval, Y. The function $f : X \to Y$ is defined as follows:

- given $x \in X$, draw a line through x and N, find its intersection y with Y, then let $y := f(x)$.

□

Exercise 4.78. Prove that this f is a homeomorphism.

Example 4.79 (stereographic projection). The above construction is a 2-dimensional version of what is, literally, a map:



This construction is used to prove that the sphere with a pinched point is homeomorphic to the plane.

In the n-dimensional case, we place $\mathbf{S}^n$ in $\mathbf{R}^{n+1}$ as the unit "sphere":

$$\mathbf{S}^n = \{x \in \mathbf{R}^{n+1} : \ ||x|| = 1\},$$

and then remove the north pole $N = (0, 0, ..., 0, 1)$. The stereographic projection

$$P : \mathbf{S}^n \setminus N \to \mathbf{R}^n$$

is defined by the formula:

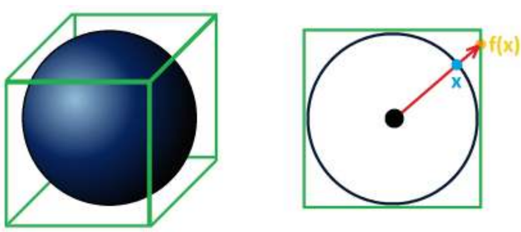
$$P(x) := \left(\frac{x_1}{1 - x_{n+1}}, \frac{x_2}{1 - x_{n+1}}, ..., \frac{x_n}{1 - x_{n+1}} \right), \ \forall x = (x_1, ...x_{n+1}).$$

Its inverse is

$$P^{-1}(y) = \frac{1}{1 + ||y||} \big(2y_1, 2y_2, ..., 2y_n, ||y||^2 - 1\big), \ \forall y = (y_1, ..., y_n). \qquad \square$$

Exercise 4.80. Prove that this is a homeomorphism.

Exercise 4.81. Show that the sphere and the hollow cube are homeomorphic. Hint: imagine inserting a balloon inside a box and then inflating it so that it fills the box from the inside:



Or we can concentrate on the inverse of f. Let's illustrate the idea in dimension 2, i.e., square Y and circle X. One can give an explicit representation of the function:

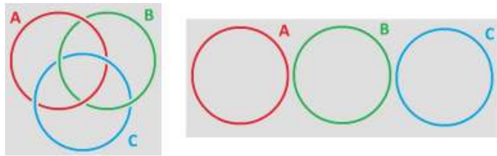
$$f^{-1}(u) = \frac{u}{||u||},$$

where $||u||$ is the norm of u. This function is a *radial projection.*

Exercise 4.82. Prove that f^{-1} is continuous.

Exercise 4.83. Let a, b be two distinct points on the sphere $X = \mathbf{S}^2$. Find a homeomorphism of the sphere to itself that takes a to b. What if X is the plane, or the torus?

Exercise 4.84. Are these two spaces, $A \sqcup B \sqcup C$, homeomorphic?



4.11 Topological equivalence

Theorem 4.85. Homeomorphisms establishes an equivalence relation on the set of all topological spaces.

Exercise 4.86. Prove the theorem.

This is the reason why it makes sense to call two spaces *topologically equivalent* when they are homeomorphic. We will use the following **notation** for that:

$$X \approx Y.$$

There is more to this equivalence... Suppose

$$f : (X, \tau_X) \to (Y, \tau_Y)$$

is a homeomorphism. From the definition,
- if U is open in Y then $f^{-1}(U)$ is open in X, and
- if V is open in X then $f(V)$ is open in Y.

In other words,

$$V \in \tau_X \iff f(V) \in \tau_Y \quad \text{ and } f^{-1}(U) \in \tau_X \iff U \in \tau_Y.$$

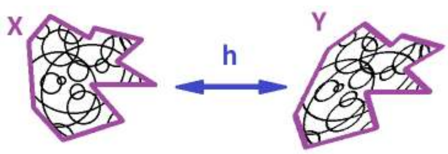
Therefore, we can define a function between the two topologies, as sets of open sets:

$$f_\tau : \tau_X \to \tau_Y,$$

by setting

$$f_\tau(V) := f(V).$$

It is a *bijection*!



The result is that whatever is happening, topologically, in X has an exact counterpart in Y: open and closed sets, their unions and intersections, the exterior, interior, closure of sets, and convergent and divergent sequences, continuous and discontinuous functions, etc. To put it informally,

homeomorphic spaces are topologically indistinguishable.

The situation is similar to that in algebra when isomorphic groups (or vector spaces) have the exact matching of every algebraic operation and, in that sense, they are algebraically indistinguishable.

Exercise 4.87. Prove that the classes of homeomorphic spaces form a category.

This idea justifies the common language about topological spaces that is often appropriate:

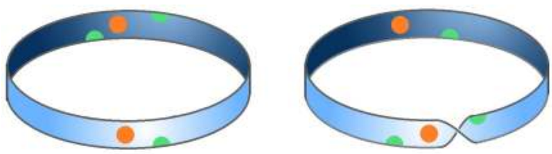
"X is homeomorphic to Y" means "X *is* Y".

This is the reason why we speak of some classes of homeomorphic spaces as if they are represented by specific topological spaces:

- *the* circle $\mathbf{S}^1$,
- *the* disk $\mathbf{B}^2$,
- *the* sphere $\mathbf{S}^2$,
- *the* torus$\mathbf{T}^2$, etc.

The attitude is similar to that of algebra: a group *is* $\mathbf{Z}, \mathbf{Z}_p, \mathcal{S}_n$, etc. (Furthermore, "the lion" refers to all the lions as a species, etc.)

It would be dangerous, however, to apply this idea to *bases* instead of topologies. Consider the fact that the bases of this cylinder and the Möbius band may seem indistinguishable but the spaces aren't homeomorphic:



Properties "preserved" under homeomorphisms are called *topological invariants*.

Exercise 4.88. Prove that the "number of points (including infinity)" is a topological invariant.

In anticipation of our future studies, we mention that *the homology groups of homeomorphic spaces are isomorphic*. In other words, homology is a topological invariant. However, the converse isn't true. Indeed, even though a point and a segment aren't homeomorphic, their homology groups coincide.

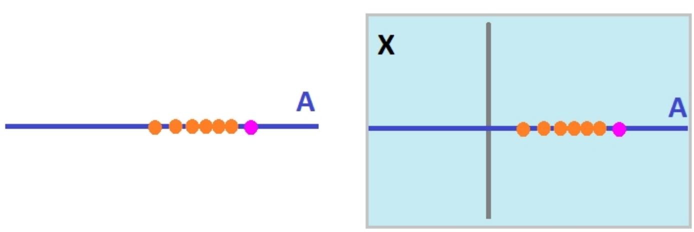
Exercise 4.89. Prove that for a given topological space, the set of all of its self-homeomorphisms form a group with respect to composition.

5 Subspaces

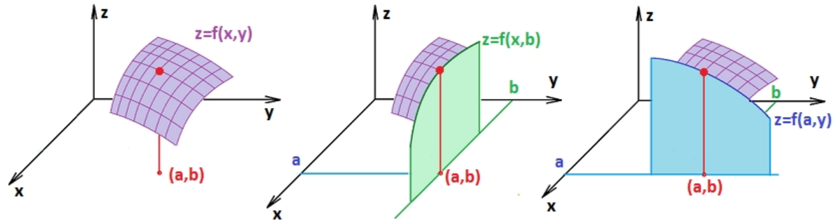
5.1 How a subset inherits its topology

Is there a relation between the topology of the xy-plane and the topology of the x-axis?

The question may be that of convergence of sequences: if $x_n \to a$ then $(x_n, 0) \to (a, 0)$:



Or that of continuity of functions: if $f(\cdot,\cdot)$ is continuous, then $f(\cdot, 0)$ is continuous too:



Or it may be about the interior, exterior, closure, etc.

There must be a relation!

Example 5.1. Let's start with a simple example. Suppose X is the xy-plane and A is the x-axis. We know that the closeness (proximity) of points in X is measured simply by the distance formula, the Euclidean metric. The distance between (x,y) and (a,b) is

$$d\Big((x,y),(a,b)\Big) := \sqrt{(x-a)^2+(y-b)^2}.$$

But if these two points happen to lie on the x-axis ($y=b=0$), the formula becomes:

$$d_A(x,a) = |x-a|,$$

which is the Euclidean metric of the x-axis. Not by coincidence, the proximity of points in A is measured in a way that matches that for X. □

The case of metric spaces is thus clear: we take the metric d of X and apply it to the elements of A. The result is a metric for A.

Exercise 5.2. Prove that this construction does produce a metric.

Next, we consider the respective topologies. Suppose we are given a point a in A.

• If we treat a as a point in X, then a neighborhood W of a in X consists of points in X within some $d>0$ from a.

• If we treat a as a point in A, then a neighborhood W_A of a in A consists of points in A within some $d>0$ from a.

Of course, this is the same d. Then,

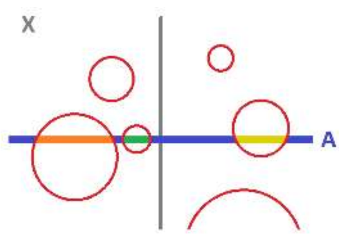
$$W_A = W \cap A!$$

Example 5.3. In our example, the construction is easy to visualize:

• the Euclidean basis of $X := \mathbf{R}^2$ is the open disks;

• the Euclidean basis of $A := \mathbf{R}$ is the open intervals.

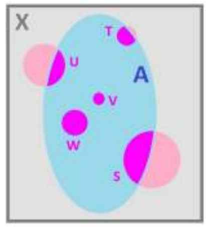
Now, intersections of disks with the x-axis produce intervals and, conversely, every interval in the x-axis is its intersection with a disk:



To be sure, there are many different disks for each interval and some disks have no matching intervals. □

This idea applies to all topological spaces.

Following our original approach to topology via *neighborhoods*, a subset A of a topological space X with basis γ will acquire its own collection γ_A as the set of all of its intersections with the elements of γ:



In other words,

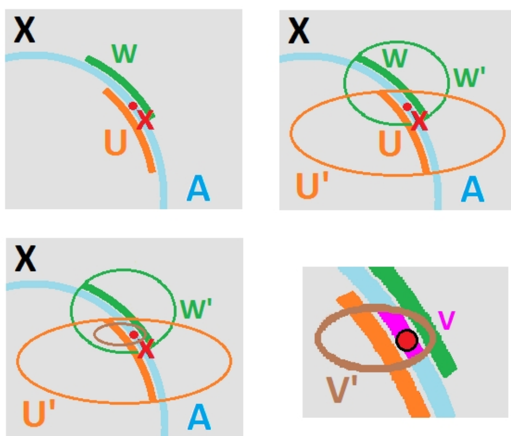
$$\gamma_A := \{W \cap A : \ W \in \gamma_X\}.$$

Alternatively, we acquire the new *topology* τ_A for A from the topology τ of X, in the exact same manner.

Theorem 5.4. The collection γ_A so defined is a basis of neighborhoods in A.

Proof. If $x \in A \subset X$, then also $x \in X$. Since γ_X is a basis on X, it satisfies (B1): there is some $W \in \gamma_X$ such that $x \in W$. Thus, we have $x \in A$ and $x \in W$, hence $x \in A \cap W \in \gamma_A$. Hence, γ_A satisfies (B1).

We need to prove (B2). Suppose we are given $U, W \in \gamma_A$ and a point $x \in U \cap W$. We need to prove that there is some $V \in \gamma_A$ such that $x \in V \subset U \cap W$.



By definition of γ_A, we have

$$U = U' \cap A, \ W = W' \cap A \text{ for some } U', W' \in \gamma.$$

From (B2) for γ_X we conclude:

$$\text{there is some } V' \in \gamma \text{ such that } x \in V' \subset U' \cap W'.$$

Now let

$$V := V' \cap A.$$

Then, $x \in A$ and $x \in V$, hence

$$x \in A \cap V = V' \in \gamma_A.$$

Also, $V' \subset U' \cap W'$ implies that

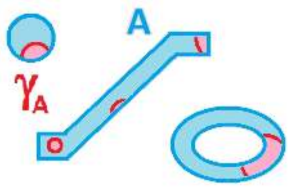
$$V = V' \cap A' \subset U' \cap W' \cap A = (U' \cap A) \cap (W' \cap A) = U \cap W.$$

Hence, γ_A satisfies (B2). ■

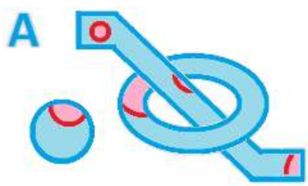
Definition 5.5. The topology generated by γ_A is called the *relative topology* on A generated by the basis γ of X.

Let's review what happens to the neighborhoods in X as we pass them over to A. Suppose X is Euclidean with nice round neighborhoods:

The neighborhoods in A are now explicit while X is gone:



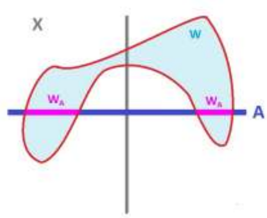
In fact, since it doesn't matter how A used to fit into X anymore, we can even rearrange its pieces:



5.2 The topology of a subspace

What if, instead of a basis, the topology is given directly, i.e., by a collection τ of open sets? Then how do we set up a topology for a subset?

In our example, all open sets in the x-axis A can be seen as intersections of open sets in the xy-plane X:



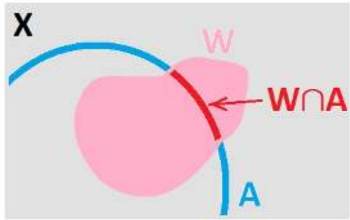
Whether this is true in general is less obvious and will need a proof.

Generally, suppose we have a set X and suppose A is a subset of X. As we know, there are many ways to define a topology on A – if it is just another set. But suppose X already has a topology, a collection of open sets τ. Then,

what ought to be *the* topology for A?

All we need to do is to point out the sets we deem open within the subset. We choose, just as before,

$$W_A := W \cap A \text{ for all } W \text{ open in } X.$$



Suppose τ is the topology of X, then we have a collection of subsets of A:

$$\tau_A := \{W \cap A : W \in \tau\}.$$

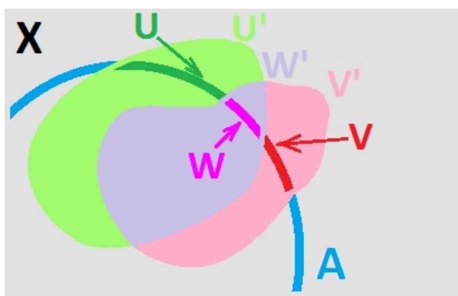
Theorem 5.6. τ_A is a topology on A.

Proof. Needless to say, the axioms of topology (T1) - (T3) for τ_A will follow from (T1) - (T3) for τ.

For (T1):
- $\emptyset \in \tau \Longrightarrow \emptyset = A \cap \emptyset \in \tau_A$,
- $X \in \tau \Longrightarrow A = A \cap X \in \tau_A$.

For (T3):
- $U, V \in \tau_A \Longrightarrow$
- $U = U' \cap A, V = V' \cap A$ for some $U', V' \in \tau \Longrightarrow$
- $W = U \cap V = (U' \cap A) \cap (V' \cap A) = (U' \cap V') \cap A \in \tau_A$,

since
- $W' = U' \cap V' \in \tau$. ■



Exercise 5.7. Prove (T2).

Definition 5.8. Given a topological space (X, τ) and a subset $A \subset X$, the topological space (A, τ_A) is called a *subspace* of X and we say that its topology is the *relative topology induced by* X.

Compare this idea to that of a subspace of a vector space. In either case, the structure (topological or algebraic) is inherited by a subset. The main difference is that for vector spaces (or groups) the subset needs to be closed under the operations while for topological spaces a subset can *always* be seen as a subspace.

Note: In the illustrations above, we attempt to choose a subset A to be depicted as something "thin" in comparison to a "thick" X. The reason is that a thick A could implicitly suggest that A is an open subset of X and that might make you rely on this subconscious assumption in your proofs.

Exercise 5.9. What is so special about the relation between τ and τ_A when A is an *open* subset of X?

Once the topology on A is established, all the topological concepts become available. We can speak of open and closed sets "in A" (as opposed to "in X"), the interior, exterior, closure, frontier, continuity, etc.

The following is obvious.

Theorem 5.10. The sets closed in A are the intersections of the closed, in X, sets with A.

Exercise 5.11. If A is closed in X then the sets closed in A are...

Theorem 5.12. If $x_n \to a$ in $A \subset X$ then $x_n \to a$ in X.

Exercise 5.13. Prove the theorem.

The converse isn't true: while $x_n = 1/n$ converges to 0 in $X := \mathbf{R}$, it diverges in $A := (0, 1)$. This affects the limit points of subsets and, therefore, the closure.

Example 5.14. Here's a simple example that illustrates the difference between the closure in X and the closure in $A \subset X$. Suppose

- $X := \mathbf{R},\ A := (0, 1),\ B := (0, 1)$, then
- in X, $\text{Cl}(B) = [0, 1]$,
- in A, $\text{Cl}(B) = (0, 1)$. □

Exercise 5.15. Find a non-Euclidean topology (or a basis) on $X := \mathbf{R}^2$ that generates the Euclidean topology on A the x-axis.

Exercise 5.16. Describe the relative topology of the following subsets of $X := \mathbf{R}$:

- $A := \{1, 2, ..., n\}$,
- $A := \mathbf{Z}$,
- $A := \{1, 1/2, 1/3, ...\}$,
- $A := \{1, 1/2, 1/3, ...\} \cup \{0\}$,
- $A := \mathbf{Q}$.

Exercise 5.17. Show that if B is a subspace of A and A is a subspace of X, then B is a subspace of X.

Exercise 5.18. Suppose A, B are disjoint subsets of a topological space X. Compare these three topologies:

- A relative to X,
- B relative to X, and
- $A \cup B$ relative to X.

It is a **default assumption** that any subset $A \subset X$ is a topological space with the relative topology acquired from X.

5.3 Relative neighborhoods vs. relative topology

An important, but easy to overlook, question remains, do γ_A and τ_A match?

What do we mean by "match"? Starting with base γ of X, we have acquired the topology τ_A of A in two different ways:

- 1. taking all sets open with respect to γ creates τ and then taking the latter's intersections with A creates τ_A; or
- 2. taking intersections with A of γ creates γ_A and then taking all sets open with respect to the latter creates τ_A.

These two methods should produce the same result, as informally illustrated by the commutative diagram below:

$$\begin{array}{ccc} \gamma & \xrightarrow{\ open\ } & \tau \\ \downarrow{\scriptstyle \cap A} & & \downarrow{\scriptstyle \cap A} \\ \gamma_A & \xrightarrow{\ open\ } & \tau_A \end{array}$$

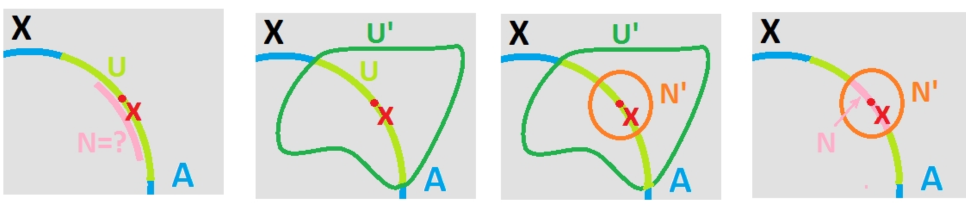
We can now reword the question we started with, as follows: why are all the open sets in the x-axis the intersections of all the open sets in the xy-plane with the x-axis?

Theorem 5.19.

- τ_A
- $=$ the set of all intersections of the elements of τ with A
- $=$ the set of all sets open with respect to γ_A.

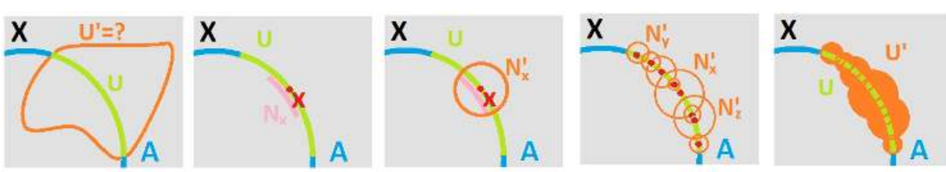
Proof. To prove the equality we prove the mutual inclusions for these two sets.

Part 1: $[\subset]$ Given $U \in \tau_A$, show that U is open with respect to γ_A. Suppose $x \in U$, we need to find $N \in \gamma_A$ such that $x \in N \subset U$. Naturally, we construct N as the intersection with a neighborhood of x in X. Since U is open in A, there is an open U' in X such that $U = U' \cap A$. Since U' is open, there is a neighborhood $N \in \gamma$ such that $x \in N' \subset U'$. Now we take $N := N' \cap A$.



Part 2: $[\supset]$ It is less straight-forward and will require a subtle step. Suppose $U \in \tau_A$; i.e., it's open with respect to γ_A. We need to find U' open in X such that $U = U' \cap A$. Since U is open in A, it follows that for every $x \in U$ there is an $N_x \in \gamma_A$ such that $x \in N_x \in U$. Further, $N_x = N'_x \cap A$ for some $N'_x \in \gamma_X$. But, even though this set is open in X, it can't possibly give us the desired U'. The ingenious idea is to take all of these neighborhoods together:

$$U' := \bigcup_{x \in U} N'_x.$$



It is open by (T2). ■

Relative topology is the simplest way to create new topological spaces from old. The second simplest is the *disjoint union.*

Suppose two (unrelated to each other) topological spaces X and Y are given. One can form the disjoint union $Z := X \sqcup Y$ of the underlying sets but what about the topology? What the definition of the topology of the disjoint union of these spaces *ought to* be? In other words, given topologies τ_X and τ_Y, what ought to be *the* topology τ_Z on set Z? Out of the multitude of possibilities, we need to choose the one that makes most sense. What we want is for the topologies of X and Y to remain *intact* in Z. But just taking the union of $\tau_X \cup \tau_X$ would not produce a topology as (T1) fails...

In light of the above discussion, we want the relative topology of X with respect to Z to coincide with τ_X and same for Y. So,

$$W \in \tau_Z \Longleftrightarrow W \cap X \in \tau_X,\ W \cap Y \in \tau_Y.$$

The choice becomes clear:

$$\tau_Z = \{U \cup V :\ U \in \tau_X, V \in \tau_Y\}.$$

Exercise 5.20. Prove that this is a topology. Also, what happens to the bases?

5.4 New maps

In light of this new concept of relative topology, we can restate a recent theorem, as follows.

Theorem 5.21. Suppose $f : X \to Y$ is continuous. If X is path-connected the so is $f(X)$.

Proposition 5.22. If $f : X \to Y$ is continuous then so is $f' : X \to f(X)$ given by $f'(x) = f(x),\ \forall x \in X$.

To follow the idea of piecewise defined functions, two continuous functions can be "glued together" to create a new continuous function, as follows.

Theorem 5.23 (Pasting Lemma). Let A, B be two closed subsets of a topological space X such that $X = A \cup B$, and let Y also be a topological space. Suppose $f_A : A \to Y,\ f_B : B \to Y$ are continuous functions, and

$$f_A(x) = f_B(x),\ \forall x \in A \cap B.$$

Then the function $f : X \to Y$ defined by

$$f(x) := \begin{cases} f_A(x) & \text{if } x \in A, \\ f_B(x) & \text{if } x \in B, \end{cases}$$

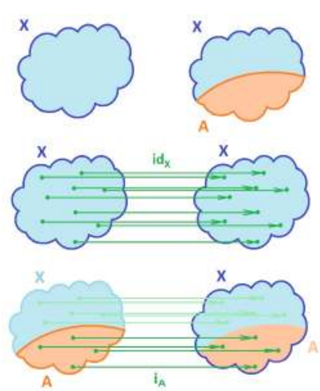
is continuous.

Exercise 5.24. (a) Prove the lemma. Hint: you'll have to use relative topology. (b) State and prove the analog of the theorem with "closed" replaced by "open".

Exercise 5.25. Show that the theorem fails if A, B aren't closed. What if just one of them isn't?

Definition 5.26. Suppose we have a topological space X and a subset A of X. Then the *inclusion* $i = i_A : A \to X$ of f is given by

$$i_A(x) := x,\ \forall x \in A.$$



For any open set U in X, we have

$$i_A^{-1}(U) = U \cap A.$$

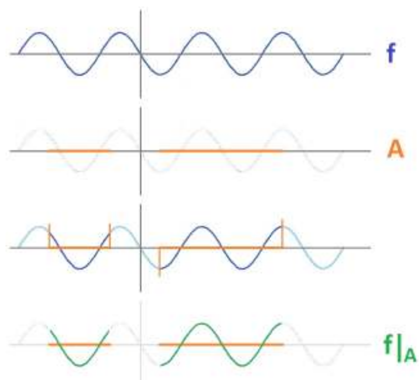
Then the result below follows from the definition of relative topology.

Theorem 5.27. The inclusion function is continuous.

Notation: The notation often used for the inclusion is

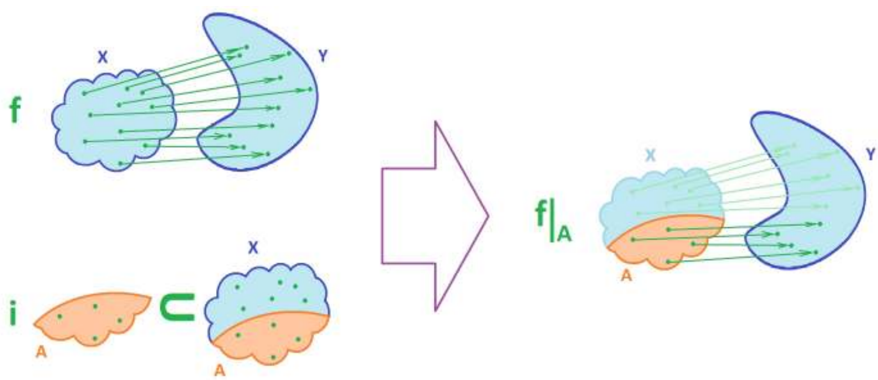
$$i = i_A : A \hookrightarrow X.$$

Next, notice how the picture below seems to imply that limiting the domain of a function will preserve its continuity:



Definition 5.28. Suppose we have topological spaces X and Y, a map $f : X \to Y$, and a subset A of X. Then the *restriction* $f\Big|_A : A \to Y$ *of* f *to* A is a function defined by

$$f\Big|_A(x) := f(x),\ \forall x \in A.$$



In particular, we can understand inclusions as restrictions of the identity functions.

Theorem 5.29. Any restriction of a continuous function is continuous.

Proof. Suppose we have a continuous function $f : X \to Y$ and a subset A of X. Suppose U is open in Y. Then,

$$\left(f\Big|_A\right)^{-1}(U) = f^{-1}(U) \cap A.$$

As the intersection of an open, in X, set with A, this set is open in A. Therefore, $f\Big|_A$ is continuous.
□

A less direct but more elegant proof is to observe that the restriction of a map to a subset is its composition with the inclusion of the subset:

$$f\Big|_A = fi_A,$$

so that the diagram commutes:

$$\begin{array}{ccc} A & \xrightarrow{i_A} & X \\ & f|_A \searrow & \downarrow f \\ & & Y \end{array}$$

Both are continuous and so is their composition.

Exercise 5.30. Suppose $i_A : A \hookrightarrow X$ is the inclusion. Suppose the set A is given such a topology that for every topological space Y and every function $f : Y \to A$,

$$f : Y \to A \text{ is continuous} \iff \text{the composition } i_A f : Y \to X \text{ is continuous.}$$

Prove that this topology coincides with the relative topology of A in X.

More general is the following concept.

Definition 5.31. A one-to-one map $f : X \to Y$ is called an *embedding* if it yields a homeomorphism $f' : X \to f(X)$, where $f'(x) = f(x)$ for all x.

Let X be a topological space and A a subspace of X. A map $r : X \to A$ such that $r(x) = x$ for all $y \in A$ is called a *retraction*. We also say that A is a *retract* of X. Another way of looking at this condition is:

$$ri_A = \mathrm{Id}_A,$$

where i_A is the inclusion of A into X and Id_A is the identity map of A. The identity is represented by this commutative diagram:

$$\begin{array}{ccc} A & \xrightarrow{i_A} & X \\ & \mathrm{Id}_A \searrow & \downarrow r \\ & & A \end{array}$$

Exercise 5.32. Prove the following.
- Any point in a convex set is its retract.
- The $(n-1)$-sphere $\mathbf{S}^{n-1}$ is a retract of $\mathbf{R}^n \setminus \{0\}$.
- The circle $\mathbf{S}^1$ is a retract of the Möbius band $\mathbf{M}^2$.
- The union of the equator and a meridian of the torus $\mathbf{T}^2$ is a retract of the torus with a point removed.

5.5 The extension problem

Just as we can restrict maps, we can try to "extend" them, continuously.

Definition 5.33. For $A \subset X$ and a given function $f : A \to Y$, a function $F : X \to Y$ is called an *extension* of f if $F\Big|_A = f$.

The relation between the function and its extension can be visualized with a diagram:

$$\begin{array}{ccc} A & \xrightarrow{i_A} & X \\ & f \searrow & \downarrow F=? \\ & & Y \end{array}$$

Extension Problem: Is there a continuous F to complete the diagram so that it is commutative?

Unlike the restriction problem above, the extension problem can't be solved by simply forming the composition. In fact, such an extension might simply not exist.

Of course, it is simple to extend a continuous function $f : [0,1] \to \mathbf{R}$ to $F : [0,2] \to \mathbf{R}$. But what about $f : (0,1) \to \mathbf{R}$ and $F : [0,1] \to \mathbf{R}$? For $f(x) = x^2$, yes; for $f(x) = 1/x$ or $f(x) = \sin(1/x)$, no.

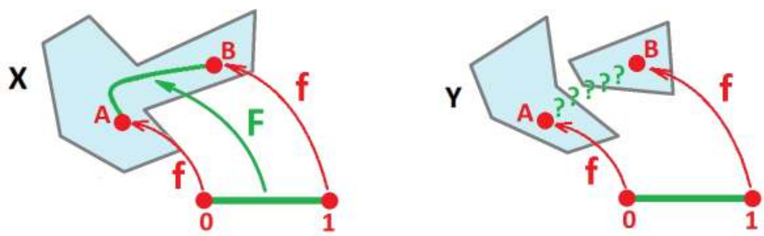
Exercise 5.34. State and prove the sufficient and necessary conditions for existence of an extension of $f : (a,b) \to \mathbf{R}$ to $F : (a,b] \to \mathbf{R}$.

A more interesting example comes from our recent study. We can recast the definition of path-connectedness of a topological space X as the solvability of a certain extension problem: can every map defined on the endpoints of the interval

$$f : \{0,1\} \to X$$

be extended to a map on the whole interval

$$F : [0,1] \to X?$$



Example 5.35 (path-connectedness of graphs). Let's sneak a peek into how this approach will be used in algebraic topology. Suppose X is a graph. The first two diagrams below are:
- the extension diagram – with topological spaces and maps, and
- its *algebraic* counterpart – with groups and homomorphisms:

$$\begin{array}{ccc} \{0,1\} & \xrightarrow{i_A} & [0,1] \\ & {}_f\searrow & \downarrow{\scriptstyle F=?} \\ & & X \end{array} \qquad \begin{array}{ccc} H_0(\{0,1\}) & \longrightarrow & H_0([0,1]) \\ & \searrow & \downarrow{\scriptstyle ?} \\ & & H_0(X) \end{array} \qquad \begin{array}{ccc} <\{0,1\}> & \longrightarrow & <0> \\ & \searrow & \downarrow{\scriptstyle ?} \\ & & H_0(X) \end{array}$$

These groups are evaluated in the last diagram. But the way we construct the homology groups and homology maps tells us that we can complete the first diagram only if we can complete the last one. But to do that – for any given f – $H_0(X)$ would have to be cyclic. Therefore, X is path-connected. □

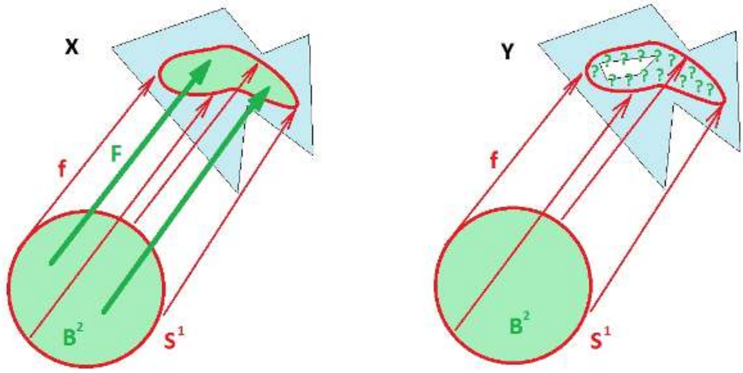
The value of this approach becomes clearer as we start climbing dimensions.

Example 5.36. A worthwhile example is recasting simple connectedness as an extension problem: can every map of the circle

$$f : \mathbf{S}^1 \to X$$

be extended to a map of the disk

$$F : \mathbf{B}^2 \to X?$$



And so on for all dimensions, even those impossible to visualize... □

Exercise 5.37. State and prove the sufficient and necessary condition for existence of a *linear* extension of a linear operator $f : A \to W$ to $F : V \to W$, where V, W are vector spaces and A is a subset of V.

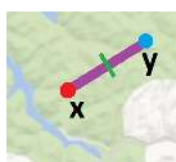
5.6 Social choice: looking for compromise

A topic we will consider – in several stages – is *social choice*. By that we understand various schemes that are meant to combine the choices several individuals into one.

Let's imagine two hikers who are about to go into the woods and they need to decide where to set up their camp. Their preferred places don't match and we need to develop a procedure for them to find a fair compromise:

We are not speaking of a one-time compromise but a compromise-making procedure. The reason is that we want to solve the problem once and for all – in case they decide to take a trip again.

The solution may be simple: *take the middle point between the two locations, if possible*:



Because of the last part, the answer depends on the *geometry* (and the *topology*!) of the set of possible choices, the forest.

So, this is what we have so far:
- the forest: a closed subset W of $\mathbf{R}^2$,
- the location chosen by the first hiker: $x \in W$, and
- the location chosen by the second hiker: $y \in W$.

We need to find
- a compromise location $C \in W$,

and not just once, for these particular x, y, but for all possible pairs $x, y \in W$. It is then a function, which we will call the *choice function*,

$$f : W \times W \to W,$$

that we are after.

Now, whenever W allows, we choose the mid-point function:

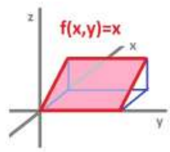
$$f(x,y) := \frac{1}{2}x + \frac{1}{2}y.$$

Such a choice is, of course, always possible when W is convex. When it's not, the answer is unclear. We then want to understand *when* this problem does have a solution, a choice function f. The answer will depend on the set W.

But first, let's make sure that f that we find will make sense as the answer in this particular situation.

First, the interests of both hikers should be taken into account equally. Certainly, a *dictatorship* of either one of the hikers can't be allowed:

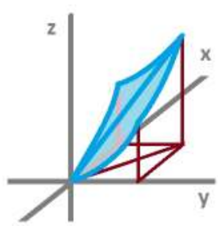
$$f(x,y) = x, \ \forall x, y.$$



We accomplish this by requiring that, if the two flip their preferences, the result will remain unchanged. Traditionally, this condition is called "anonymity".

Anonymity Axiom (Symmetry): The choice function is symmetric:

$$f(x,y) = f(y,x), \ \forall x \in W.$$



Exercise 5.38. Restate the axiom for the case of m hikers.

Unfortunately, a constant function

$$f(x,y) := P, \ \forall x, y,$$

for some $P \in W$, demonstrates a kind of dictatorship (by a third party?) that passes the test of the Anonymity Axiom. This is why we also need to ensure that these two persons are the ones and the only ones who decide. We require that, if the two choices happen to coincide, that's the choice that will be made. Traditionally, this condition is called "unanimity".

Unanimity Axiom (Diagonality): The choice function restricted to the diagonal is the identity:

$$f(x,x) = x, \ \forall x \in W.$$

In other words, the *diagonal* of $W \times W$,

$$\Delta(W) := \{(x,y) \in W \times W : \ x = y\},$$

is copied to W by f.

Define the *diagonal map* $\delta : W \to W \times W$ by $\delta(x) = (x,x)$ (its image is the diagonal $\Delta(W)$ of W). Then, we can also interpret the axiom as the commutativity of this diagram:

$$\begin{array}{ccc} W & \xrightarrow{\ \delta\ } & W \times W \\ & {}_{\text{Id}}\searrow & \downarrow f \\ & & W \end{array}$$

Exercise 5.39. Prove that the diagonal map is an embedding.

Exercise 5.40. Restate the axiom for the case of m hikers.

Naturally, we will require f to be continuous. The justification is the same as always: a small change in the input – the choice of one of the two hikers – should produce only a small change in the output – the compromise choice. As another way to look at this, the error (caused by limited accuracy) of the evaluation of the preferences won't dramatically change the result. Traditionally, this condition is called "stability".

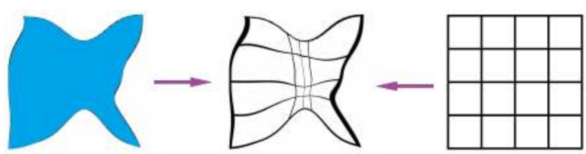
Stability Axiom (Continuity): The choice function $f : W \times W \to W$, where $W \times W \subset \mathbf{R}^4$, is continuous.

Exercise 5.41. Prove that the mid-point choice function, when it exists, satisfies the axioms.

Finally, the *social choice problem* is:

Is there a function $f : W \times W \to W$ that satisfies these three axioms?

A positive answer won't tell us a lot. Consider how, when W is homeomorphic to a convex set Q, we solve the problem for Q and then bring the solution back to W:



Specifically, if $h : W \to Q$ is such a homeomorphism, our solution is

$$f(x,y) = h^{-1}\left(\tfrac{1}{2}h(x) + \tfrac{1}{2}h(y)\right).$$

The solution is fair as it doesn't favor one hiker over the other, but would they be happy with it? As all the geometry of Q is lost in W, such a "compromise" might be far away from the choices of either of the hikers.

Exercise 5.42. Prove that the nature of the problem is topological: if the problem has a solution for W, then it has a solution for every space homeomorphic to W.

We try to keep the constraints on f as broad as possible and try to investigate what constraints on the topology of W will make it possible that f does exist. The example of the mid-point solution above already suggests that W might have to be *acyclic*.

Example 5.43. Starting with 0-homology, does W have to be path-connected? Suppose not:

$$W = W_1 \sqcup W_2,$$

where W_1, W_2 are convex closed subsets of the plane. Suppose we pick a point $P \in W_1$ and define

$$f(x,y) = \begin{cases} \frac{1}{2}x + \frac{1}{2}y & \text{if } x, y \in W_1 \text{ or } x, y \in W_2, \\ P & \text{otherwise.} \end{cases}$$

It is easy to see that the axioms are satisfied. □

Exercise 5.44. Find a fairer choice function for the above situation.

The existence of the solution above indicates that, in a way, the axioms are *too weak*. In real life, the two persons may choose two different forests and there can be no fair compromise. The unfairness is seen from the fact that, even though the two hikers are treated equally, the two forests aren't (as if a third person chose a candidate ahead of time).

Exercise 5.45. Strengthen the Unanimity Axiom in such a way that we avoid this kind of dictatorship.

Next, does W have to be simply connected?

In practical terms, is it always possible to find a fair compromise between the two choices when there is a *lake* in the middle of the forest? To keep things simple, let's reduce the problem to the following:

- the hikers need to decide on a camping location on the shore of a circular lake:

When the two choices are close to each other, the solution can still be chosen to be the mid-point (along the arch). But what if they choose two diametrically opposite points, $x = -y$? Then there are *two* mid-points. Which one do we choose for $f(x, -x)$? A possible solution is to choose the one, say, clock-wise from first to second but that would violate the symmetry requirement.

Exercise 5.46. Show that it also violates the continuity requirement.

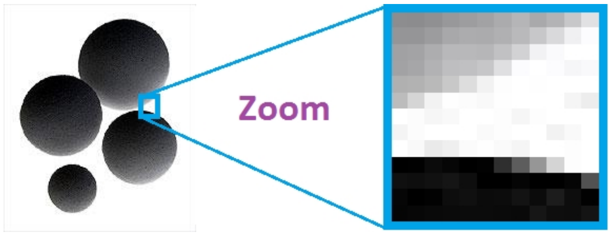
Exercise 5.47. What if, in the case of diametrically opposite choices x, y, we always select $f(x, y)$ to be the mid-point that lies in, say, the upper half of the circle?

We will demonstrate later that, even with these loose constraints, there is no fair decision aggregation rule. Homology theory will be used as the tool to answer – in the negative – this and related social choice questions.

5.7 Discrete decompositions of space

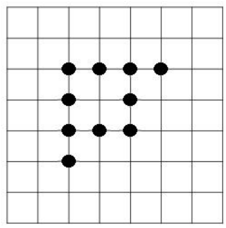
The need to discretize the Euclidean space comes from our desire to study the world computationally. In order to understand what is the right way to discretize the space, we take, as an example, *image processing*.

An image is a grid of "pixels" marked in black and white (binary) or shades of gray (gray-scale):

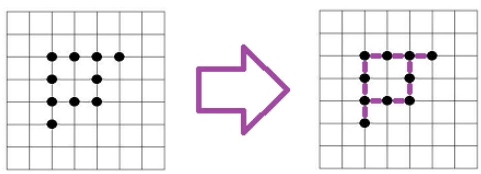


Now, how do you represent objects depicted in the images? The idea is to break our continuous universe, the plane, into small, discrete (but still continuous!) elementary pieces and then use them to build what we want. For that to work, each of these pieces must be some kind of object or figure itself and not just a location. What are those elementary pieces?

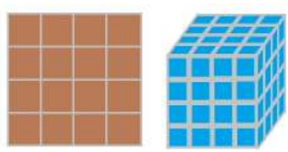
We start our decomposition by placing dots, or *vertices*, at the nodes of the grid:



To be able to form graphs, we also add the *edges*:



To form higher dimensional objects, we think of pixels as tiles and we think of voxels as bricks:



We want these building blocks, called *cells*, to be:

- 1. Euclidean,
- 2. acyclic, and
- 3. "small".

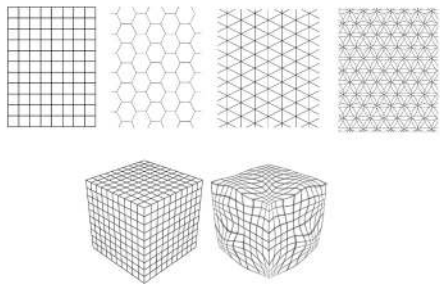
Note that, even though the pieces are cut from the Euclidean space, we might end up with subsets with "exotic" topology.

Exercise 5.48. Show that the closure of the graph of $y = \sin(1/x)$ has three path-components but two of them aren't closed subsets.

To avoid these abnormalities, it makes sense to start, in the next subsection, with *open* subsets of $\mathbf{R}^n$. The advantage is that every point has a neighborhood homeomorphic to $\mathbf{R}^n$. These pieces are indeed *Euclidean*.

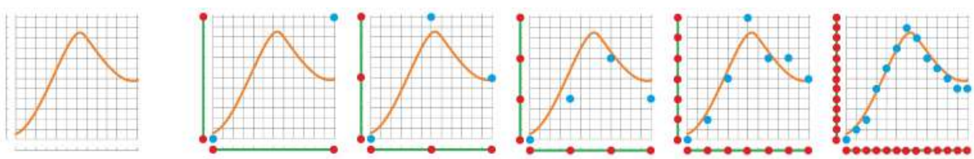
In the above example, vertices, edges, squares, and cubes are the cells. Because these pieces are discrete, we can model and study the universe with computers. On the other hand, we can put them all back together to reconstruct the space in full. And if we use only some of them, they will form objects with similar, Euclidean, properties.

In general, the shapes of these pieces may be arbitrary. They may be squares, triangles, hexagons, etc., the shapes may vary within the same decomposition, and they may even be curved:



We also want these pieces to be *acyclic*: path-connected, no tunnels, no voids, etc. The idea is that the topological features of a space formed by the cells should come from the way they are put together not from the topology of each cell.

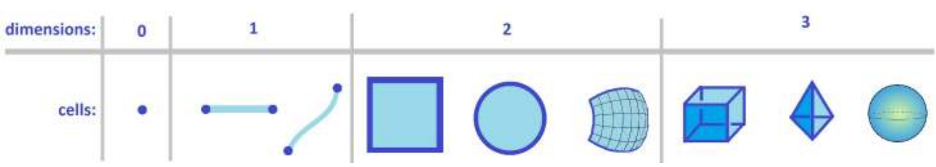
What makes cell *small* is the possibility of refinement. Refining these decompositions into ones with smaller and smaller cells allows us to *approximate* the space – with discrete structures – better and better. Moreover, discrete functions defined on these discrete structures approximate continuous functions defined on the space:



5.8 Cells

Next, we formalize these ideas.

Definition 5.49.
• An *open n-cell* is a topological space homeomorphic to the Euclidean n-space $\mathbf{R}^n$.
• A *closed n-cell*, or simply an n-cell, is a topological space homeomorphic to the closed n-ball $\mathbf{B}^n$.

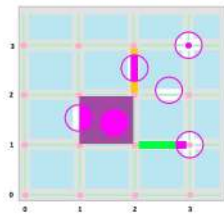


Initially, we look at cells as subsets of this Euclidean space $\mathbf{R}^n$. The simplest closed n-cell is then the closed n-ball located and the simplest open n-cell is its interior, the open n-ball. As subsets, we can study their openness, closedness, interior, frontier, and closure. What they are for these simplest cells is quite obvious. However, if the ambient n-dimensional Euclidean space is fixed but the dimensions of the cells aren't, a cell of a lower dimension will have very different attributes.

Example 5.50. In dimension $n = 2$, we have for open cells:

0-cell, vertex P	closed, not open
1-cell, edge a	neither open nor closed
2-cell, face σ	open, not closed

To see why, it suffices to look at how the Euclidean (disk) neighborhoods intersect these cells:



Also,

$$\begin{aligned}&\text{Int}(P) = \emptyset, \text{Fr}(P) = P, \text{Cl}(P) = P;\\ &\text{Int}(a) = \emptyset, \text{Fr}(a) = a, \text{Cl}(a) = a \cup \{\text{the two endpoints}\};\\ &\text{Int}(\sigma) = \sigma;\\ &\text{Fr}(\sigma) = \cup\ 4 \text{ edges } \cup\ 4 \text{ vertices of the square};\\ &\text{Cl}(\sigma) = \text{the closed square.}\end{aligned}$$

□

Next, we are to build new things with the cells by attaching them to each other along their "boundaries". It is important to distinguish between:

- the *frontier* $\text{Fr}(a)$ of a k-cell a if it is treated as a subset of $\mathbf{R}^n$, $n \geq k$, and
- the *boundary* ∂a of a k-cell a if it is treated as a topological space of a special kind.

The former depends on n while the latter doesn't, but it has to match the former under homeomorphisms when $k = n$. Then what is the boundary?

Definition 5.51. A point x in a n-cell a is an *interior point* of a if it has a neighborhood homeomorphic to $\mathbf{R}^n$; the rest are *boundary points*. The boundary ∂a is the set of all boundary points of a.

This definition of the boundary is *intrinsic*!

Exercise 5.52. Show that each boundary point of an n-cell has a neighborhood homeomorphic to the half-space $\mathbf{R}^n_+$.

Exercise 5.53. Represent the boundary of an n-ball in $\mathbf{R}^n$ as the union of (a) closed $(n-1)$-cells, (b) open $(n-1)$-cells.

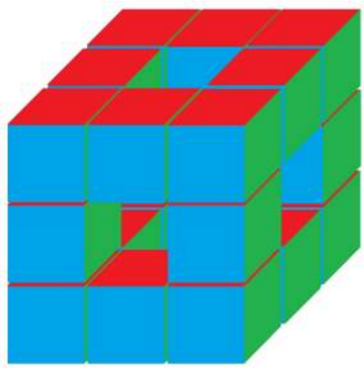
Exercise 5.54. Prove that the graph of $y = \sin \frac{1}{x}$, $0 < x < 1$, can't be a part of the boundary of a 2-cell in the plane.

Theorem 5.55. A homeomorphism of a n-cell a to the n-ball B in $\mathbf{R}^n$ maps the boundary points of a to the points of the frontier of B. In other words, *boundary goes to boundary.*

Exercise 5.56. Prove the theorem. Hint: it's an embedding.

Chapter III

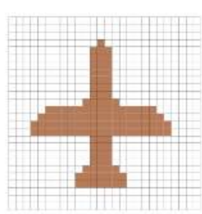
Complexes



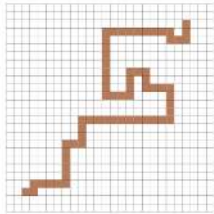
1 The algebra of cells

1.1 Cells as building blocks

We know that we can decompose the N-dimensional Euclidean space into blocks, the N-cells. For instance, this is how an object can be represented with tiles, or pixels ($N = 2$):

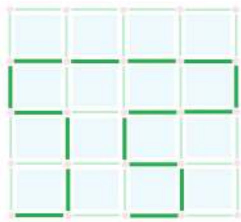


Some questions remain, such as: what about curves? Without making sense of curves we can't study such topological concepts as path-connectedness. *Are curves also made of tiles?* Such a curve would look like this:

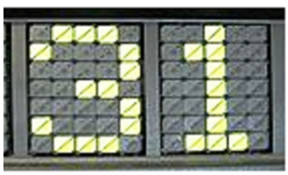


If we look closer, however, this "curve" isn't a curve in the usual topological sense! After all, a curve is supposed to be a continuous image of the 1-dimensional interval $[a, b]$.

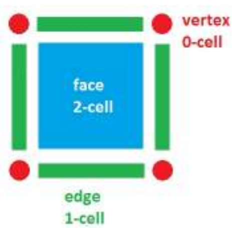
The answer comes from our study of graphs: *curves should be made of edges*. Then a "discrete" – but still Euclidean – curve will look like this:



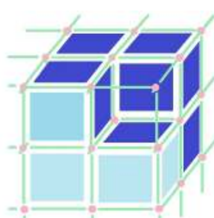
This idea isn't entirely abstract; just contrast these two images:



The conclusion is that we need to include the *lower dimensional* cells as additional building blocks. The complete cell decomposition of the pixel is shown below; the edges and vertices are shared with adjacent pixels:

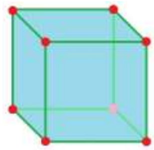


Meanwhile, a mismatch of the dimensions also appears in the 3-dimensional case if you try to use bricks, or voxels, to make surfaces. Surfaces should be made of faces of our bricks, i.e., the tiles. This is what a "discrete" surface looks like:



It's 2-dimensional!

The cell decomposition of the voxel follows and here, once again, the faces, edges, and vertices are shared:



One can interpret cells as, literally, the building blocks for everything:

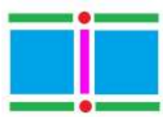
- dimension 1: sticks and wires;
- dimension 2: tiles and plywood boards;
- dimension 3: bricks and logs.

Exercise 1.1. Sketch this kind of discrete, "cellular", representation for: a knotted circle, two interlinked circles, the sphere, the torus, the Möbius band.

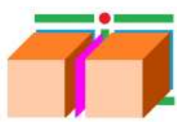
Now, we are on a solid ground to describe adjacency of cells.

Two adjacent edges, 1-cells, share a vertex, i.e., a 0-cell:

Two adjacent pixels, 2-cells, share an edge, i.e., a 1-cell:



Two adjacent voxels, 3-cells, share a face, i.e., a 2-cell:



The hierarchy becomes clear.

Thus, our approach to decomposition of space, in any dimension, boils down to the following:

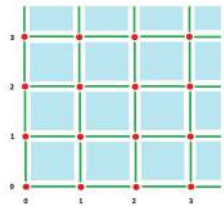
The N-dimensional space is composed of cells in such a way that k-cells are attached to each other along $(k-1)$-cells, $k = 1, 2, ..., N$.

Exercise 1.2. What are the open sets of this space?

Exercise 1.3. Explain – in terms of open, closed sets, etc. – the topological meaning of adjacency.

1.2 Cubical cells

From many possible decompositions of the plane, $\mathbf{R}^2$, into "cells" we choose the squares:



Here and elsewhere we use the following **coloring convention**:

- red vertices,
- green edges,
- blue faces,
- orange cubes.

Example 1.4 (dimension 1). We should start with dimension $N = 1$ though:

Here the cells are:

- a vertex, or a 0-cell, is $\{n\}$ with $n = ... - 2, -1, 0, 1, 2, 3, ...$;
- an edge, or a 1-cell, is $[n, n + 1]$ with $n = ... - 2, -1, 0, 1, 2, 3, ...$.

And,

- 1-cells are attached to each other along 0-cells. □

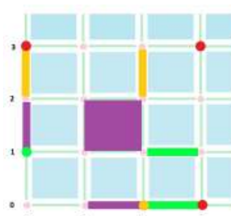
Example 1.5 (dimension 2). Meanwhile, for the dimension $N = 2$ grid, we define (closed) *cubical cells* for all integers n, m as products:

- a vertex, or a 0-cell, is $\{n\} \times \{m\}$;
- an edge, or a 1-cell, is $\{n\} \times [m, m + 1]$ or $[n, n + 1] \times \{m\}$;
- a square, or a 2-cell, is $[n, n + 1] \times [m, m + 1]$.

And,

- 2-cells are attached to each other along 1-cells and, still,
- 1-cells are attached to each other along 0-cells. □

Example 1.6.



Cells shown above are:

- 0-cell $\{3\} \times \{3\}$;
- 1-cells $[2, 3] \times \{1\}$ and $\{2\} \times [2, 3]$;
- 2-cell $[1, 2] \times [1, 2]$. □

We can define *open cells* as products too:

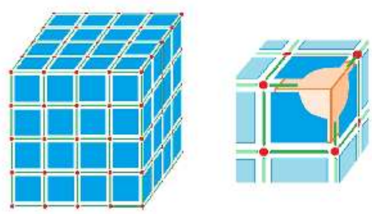
- a vertex is $\{n\} \times \{m\}$;
- an open edge is $\{n\} \times (m, m + 1)$ or $(n, n + 1) \times \{m\}$;
- the inside of a square is $(n, n + 1) \times (m, m + 1)$.

The difference is that under the former approach, the closed cells overlap and, therefore, can be glued together to form objects, while under the latter, the open cells are disjoint and produce a partition of the space. The way we will combine these cells, *it won't matter which decomposition is used.*

Exercise 1.7. Show that intersections of the elements of the basis of the Euclidean space $\mathbf{R}^N$ with an open cell form its basis. What about the closed cells?

Example 1.8 (dimension 3). For all integers n, m, k, we have:

- a vertex, or a 0-cell, is $\{n\} \times \{m\} \times \{k\}$;
- an edge, or a 1-cell, is $\{n\} \times [m, m+1] \times \{k\}$ etc.;
- a square, or a 2-cell, is $[n, n+1] \times [m, m+1] \times \{k\}$ etc.;
- a cube, or a 3-cell, is $[n, n+1] \times [m, m+1] \times [k, k+1]$.



□

What we see here is the N-dimensional Euclidean space decomposed into 0-, 1-, ..., N-cells in such a way that

- N-cells are attached to each other along $(N-1)$-cells,
- $(N-1)$-cells are attached to each other along $(N-2)$-cells,
- ...
- 1-cells are attached to each other along 0-cells.

Definition 1.9. In the N-dimensional unit grid, $\mathbf{Z}^N$, a *cube* is the subset of $\mathbf{R}^N$ given by the product with N components:

$$P := I_1 \times ... \times I_N,$$

such that its kth component is either

- *an edge* $I_k = [m_k, m_k + 1]$, or
- *a vertex* $I_k = \{m_k\}$,

in the kth coordinate axis of the grid. The cube is n-dimensional, $\dim P = n$, and is also called a (cubical) *cell* of dimension n, or an n-cell, when there are n edges and $N - n$ vertices on this list. A *face* of P is also a cube given by the same product as P but with some of the edges replaced with vertices.

Notation: The set of all unit cubes (of all dimensions) in $\mathbf{R}^N$ will be denoted by $\mathbb{R}^N$.

Exercise 1.10. Show that any $(k-1)$-dimensional face of a $(k+1)$-dimensional cube P is a common face of exactly two k-dimensional faces of P.

Exercise 1.11. For $k \leq 6$, determine the number of vertices and the number of edges of a k-cube.

Exercise 1.12. Count the faces of a 4-cube.

1.3 Boundaries of cubical cells

Before we start with the topological analysis of subsets of the Euclidean space made of cubes, let's review what we have learned studying the *topology of graphs*.

The topology of a graph is determined by what the edges do. The list of what can happen is short:

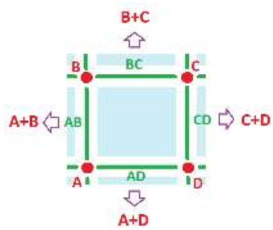
• some edges connect components of the graph and some don't, and

• some edges form loops and some don't.

To detect these events, we looked at *how each edge is attached to the nodes* and vice versa. This information is found by detecting which nodes are the endpoints of which edges. It is simple:

• the endpoints of an edge $e = AB$, a 1-cell, are the two nodes A, B, 0-cells, that form its boundary ∂e.

The idea is fully applicable to our new setting. Starting with the 0- and 1-cells of our grid, we are in the exact same situation: the boundary of an edge is still the sum of its endpoints:



The algebraic tool we used was the *boundary operator*:

$$\partial(AB) = A + B.$$

The main lesson of our study of the topology of graphs is:

boundaries are chains of cells,

i.e., formal sums of cells. More precisely,

the boundary of a k-cell is a chain of $(k-1)$-cells.

Exercise 1.13. How do the homology groups of a graph change if we add an edge? In other words, compare $H_0(G), H_1(G)$ with $H_0(K), H_1(K)$, where H, K are two graphs satisfying $N_K = N_G,\ E_K = E_G \cup \{e\}$. Hint: consider two cases.

Example 1.14. Let's show what this theory looks like in our new, cubical notation. It is simple in $\mathbb{R}^1$:

$$\partial\Big([n, n+1]\Big) = \{n\} + \{n+1\}.$$

And it looks similar for $\mathbb{R}^2$:

$$\begin{aligned}\partial\Big([n, n+1] \times \{k\}\Big) &= \Big(\{n\} + \{n+1\}\Big) \times \{k\} \\ &= \{n\} \times \{k\} + \{n+1\} \times \{k\}.\end{aligned}$$

□

Further, our approach is uniform throughout all dimensions:

• The boundary of an edge, 1-cell, consists of its two endpoints, 0-cells;

• The boundary a square, 2-cell, consists of its four edges, 1-cells;

• The boundary of a cube, 3-cell, consists of its six faces, 2-cells;

• etc.

The *boundary*, denoted by $\partial\sigma$, can be thought of in three different ways depending on the context:

• if the cell σ, or the union of cells $\sigma := \cup_i \sigma_i$, is thought of as a *subset* of the Euclidean space $\mathbf{R}^N$, then its boundary $\partial\sigma$ is also a subset of $\mathbf{R}^N$;

• if the cell σ (dimension m), or cells $\{\sigma_i\}$, is thought of as a *collection of cells*, then its boundary $\partial\sigma$ is also a collection of cells (dimension $(m-1)$);

• if the cell σ, or a sum of cells $\sigma := \sum_i \sigma_i$, is thought of as a *chain*, then its boundary

$\partial\sigma = \sum_i \partial\sigma_i$ is also a chain.
The last interpretation is the one we will use.

Example 1.15. Let's compute the boundaries of the cells from last subsection ($\mathbb{R}^2$):

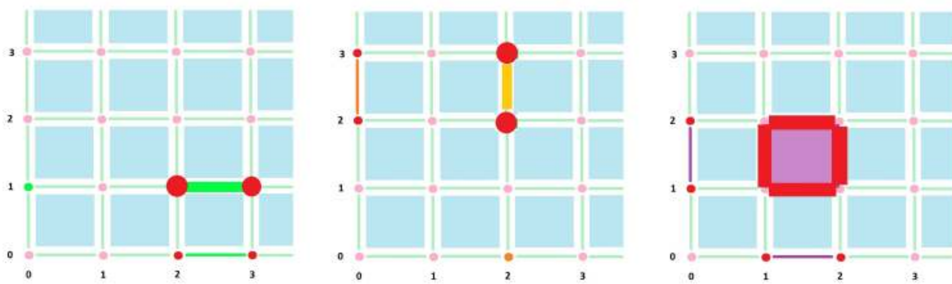
- 0-cell $\{3\} \times \{3\}$;
- 1-cells $[2,3] \times \{1\}$ and $\{2\} \times [2,3]$;
- 2-cell $[1,2] \times [1,2]$.

The boundary of a vertex is empty. Algebraically,

$$\partial\{(3,3)\} = 0.$$

The boundary of an edge is a combination of its endpoints, or algebraically,

$$\begin{aligned} \partial\Big([2,3] \times \{1\}\Big) &= \{(2,1)\} + \{(3,1)\}, \\ \partial\Big(\{2\} \times [2,3]\Big) &= \{(2,2)\} + \{(2,3)\}. \end{aligned}$$



Finally, the square:

$$\partial\Big([1,2] \times [1,2]\Big) = [1,2] \times \{1\} + \{1\} \times [1,2] + [1,2] \times \{1\} + \{2\} \times [1,2].$$

To confirm that this makes sense, let's factor and rearrange the terms:

$$\begin{aligned} &= [1,2] \times \Big(\{1\} + \{2\}\Big) &&+ \Big(\{1\} + \{2\}\Big) \times [1,2] \\ &= [1,2] \times \partial[1,2] &&+ \partial[1,2] \times [1,2]. \end{aligned}$$

We have linked the boundary of the product to the boundaries of its factors! □

The last result may be stated as a formula:

$$\partial(a \times b) = a \times \partial b + \partial a \times b.$$

It looks very much like the *Product Rule* for derivatives, except this time the order of factors matters. We know how to compute the derivative of the product of, say, three functions. We simply apply the product rule twice, as follows:

$$(fgh)' = ((fg)h)' = (fg)'h + fgh' = (f'g + fg')h + fgh' = f'gh + fg'h + fgh'.$$

As you can see, there are as many terms as the functions involved and each contains all three, except one of them is replaced with its derivative. Applying the rule for boundaries above will have the exact same effect. This motivates the following crucial definition.

Definition 1.16. The *boundary of a cubical cell* P in $\mathbb{R}^N$ given as the product:

$$P = \quad I_1 \times ... \times I_{k-1} \times I_k \times I_{k+1} ... \times I_N,$$

is given by:

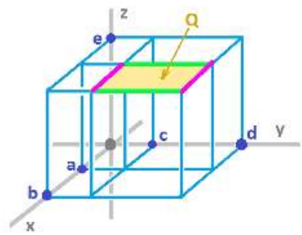
$$\partial P := \sum_{k=1}^{N} I_1 \times ... \times I_{k-1} \times \partial I_k \times I_{k+1} ... \times I_N.$$

Here, each term in the sum of ∂P has the same components as P except for one that is replaced with its boundary. Each component I_k is either

- a vertex A, then $\partial I_k = \partial(A) = 0$, or
- an edge AB, then $\partial I_k = \partial(AB) = A + B$.

Example 1.17. Let's illustrate the formula with an example of a 2-cell in $\mathbb{R}^3$, below. Suppose P is this horizontal square:

$$P := [a,b] \times [c,d] \times \{e\}.$$

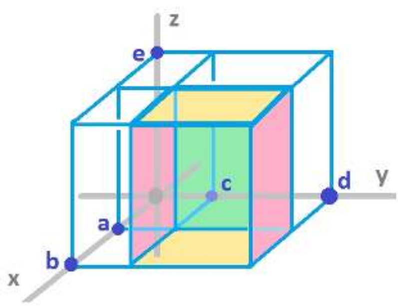


Below, the formula is used, followed by simplification:

$$\begin{array}{lcl}
\partial P & := & \partial([a,b] \times [c,d] \times \{e\}) \\
& = & \partial[a,b] \times [c,d] \times \{e\} \\
& + & [a,b] \times \partial[c,d] \times \{e\} \\
& + & [a,b] \times [c,d] \times \partial\{e\} \\
& = & \big(\{a\} + \{b\}\big) \times [c,d] \times \{e\} \\
& + & [a,b] \times \big(\{c\} + \{d\}\big) \times \{e\} \\
& + & [a,b] \times [c,d] \times 0 \\
& = & \{a\} \times [c,d] \times \{e\} + \{b\} \times [c,d] \times \{e\} \\
& + & [a,b] \times \{c\} \times \{e\} + [a,b] \times \{d\} \times \{e\}.
\end{array}$$

The last two lines give the two pairs of opposite edges that surround square P. □

Exercise 1.18. Compute the boundary of the cube below indicating in your answer its opposite faces:



Exercise 1.19. Show that if P is an m-cell, then every non-zero term in ∂P is the sum of two $(m-1)$-cells that are a pair of opposite faces of P.

1.4 Binary chains and their boundaries

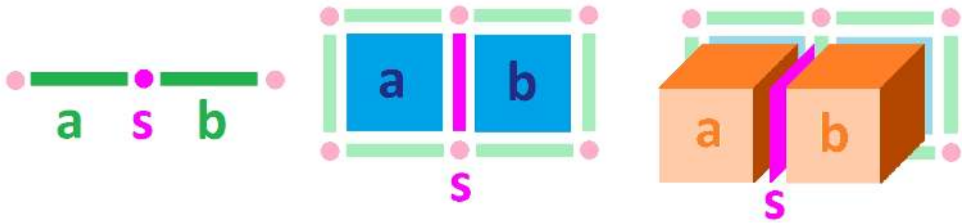
Now, what about boundaries of more complex objects?

They are still made of cells and we simply think of them as formal sums of cells. Just in the case of graphs, we will call them *chains*. Since boundaries of chains of *all* dimensions can be empty, we adopt the **convention** that

$$0 \textit{ is a } k\textit{-chain, for any } k.$$

Further, we need to combine their boundaries somehow to form the boundary of the object.

Take a look at these examples of chains that consist of two adjacent k-cells, a and b, for $k = 1, 2, 3$:



The two k-cells share a common face, a $(k-1)$-cell, s (in the middle). This cell is special. In the combined boundary of the chain,

- it appears twice, but
- it shouldn't appear at all!

Our conclusion is that, for computing the boundary, this **algebraic rule of cancellation** should apply:

$$2x = 0.$$

In other words, if a cell appears twice, it is canceled. The computation is carried out as if we do algebra with *binary arithmetic*. That's why we can think of our chains both as finite "combinations" and finite binary sums of cells. The latter idea is preferable:

$$a = \sum_i s_i \sigma_i, \ \ s_i \in \mathbf{Z}_2.$$

Then, as the boundary operator is defined on each of the cells, we can now extend this definition to all k-chains:

$$\partial_k(a) := \sum_i s_i \partial_k(\sigma_i),$$

subject to the cancellation rule.

Let's test this idea with examples.

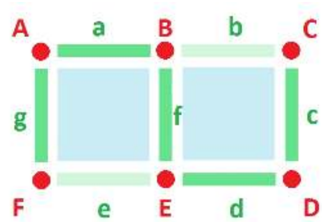
Example 1.20. A formal computation of the boundary of the chain representing two adjacent edges is below:

$$\begin{aligned} \partial\Big(\{n\} \times [m, m+1] + [n, n+1] \times \{m\}\Big) &= \partial\Big(\{n\} \times [m, m+1]\Big) + \partial\Big([n, n+1] \times \{m\}\Big) \\ &= \{(n, m)\} + \{(n, m+1)\} + \{(n, m)\} + \{(n+1, m)\} \\ &= \{(n, m+1)\} + \{(n+1, m)\}. \end{aligned}$$

□

We might prefer a less complex notation just to give an idea of what is going on.

Example 1.21. Let's compute the boundary of a curve formed by edges:

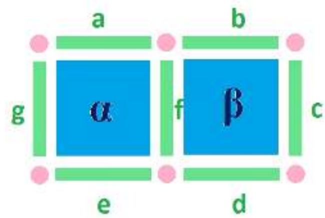


$$\begin{aligned} \partial(a + b + f + d + c) &= F + A + A + B + B + E + E + D + D + C \\ &\text{...cancellation...} \\ &= F + C. \end{aligned}$$

Only the *endpoints* of the curve remain, as expected. □

Exercise 1.22. State this conclusion as a theorem and prove it.

Example 1.23. Compute the boundary of the union of two adjacent squares:

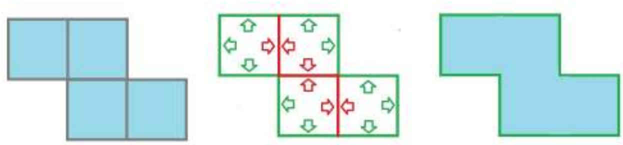


$$\begin{array}{lcccc}\partial(\alpha+\beta) & = & \partial\alpha & + & \partial\beta \\ & = & (a+f+e+g) & & +(b+c+d+f) \\ & = & a+e+g & +(\mathbf{f}+\mathbf{f}) & +b+c+d \\ & = & a+e+g & +\mathbf{0} & +b+c+d \\ & = & a+b & +c+d & +e+g.\end{array}$$

Only the *outside edges* appear, as expected. □

In conclusion, the boundary of a region formed by a collection of squares, i.e., a 2-chain with $N=2$, is made of its "external" edges. Now, each edge is shared by two squares, but
- an "external" edge is shared by one square in the collection and one not in the collection, while
- an "internal" edge is shared by two squares in the collection.

After application of the boundary operator,
- each "external" edge appears once and stays, while
- each "internal" edge appears twice and cancels:



Exercise 1.24. Provide a similar analysis (with sketches) for a collection of edges and a collection of cubes.

Beyond dimension 3, however, sketches and other visualization becomes impossible and we will have to rely exclusively on algebra.

1.5 The chain groups and the chain complex

Suppose the ambient dimension N is fixed as well as the integer grid $\mathbf{Z}^N$ of the Euclidean space $\mathbf{R}^N$.

Notation:
- $\mathbb{R}^N$ is the set of all cells (of all dimensions) in the grid,
- $C_k(L)$ is the set of all k-chains in a given collection $L \subset \mathbb{R}^N$ of cells, and, in particular,
- $C_k := C_k(\mathbb{R}^N)$ is the set of all k-chains in $\mathbb{R}^N$.

We will call $C_k(L)$ the kth *chain group of* L and we call, within this section, C_k the *total kth chain group*.

Definition 1.25. Given a k-chain $s \in C_k$,

$$s = a_1 + a_2 + ... + a_n,$$

where $a_1, a_2, ..., a_n$ are k-cells, the *boundary* ∂s of s is given by

$$\partial s := \partial a_1 + \partial a_2 + ... + \partial a_n.$$

So, the boundaries of k-cells are $(k-1)$-chains, and, moreover, the boundaries of k-chains are $(k-1)$-chains as well. Then, boundaries are found by a well-defined function

$$\partial = \partial_k : C_k \to C_{k-1}.$$

It is called the kth *boundary operator.*

Just as we did before, we notice that this function is simply an extension of a function defined on the generators of a group to the rest of the group – by additivity. Therefore, we have the following.

Theorem 1.26. The boundary operator is a homomorphism.

We now present the big picture of *the algebra of chains and their boundaries.*

The homomorphisms $\partial_k : C_k \to C_{k-1},\ k = 1, 2, ...,$ if written consecutively, form a sequence:

$$C_N \xrightarrow{\partial_N} C_{N-1} \xrightarrow{\partial_{N-1}} ... \xrightarrow{\partial_{k+2}} C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} ... \xrightarrow{\partial_1} C_0.$$

As always, consecutive arrows are understood as compositions. We would like to treat these groups uniformly and we append this sequence with two more items, one in the beginning and one at the end. Both are zero groups:

$$0 \xrightarrow{\partial_{N+1}} ... \quad ... \quad ... \quad ... \xrightarrow{\partial_0} 0.$$

The two groups are attached to the rest of the sequence with homomorphisms that are, of course, zero. The result is this sequence of groups and homomorphisms called the *chain complex* of $\mathbb{R}^N$, or the total chain complex:

$$0 \xrightarrow{\partial_{N+1}=0} C_N \xrightarrow{\partial_N} ... \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} ... \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0=0} 0.$$

As we know, the chain groups $C_k(L),\ k = 0, 1, 2, ...,$ can be defined for any set $L \subset \mathbb{R}^N$ of k-cells. They are subgroups of $C_k = C_k(\mathbb{R}^N)$. Moreover, the chain complex comprised of these groups can be constructed in the identical way as the one for the total chain groups – if we can make sense of how the boundary operators is defined:

$$\partial_k^L : C_k(L) \to C_{k-1}(L),\ \forall k = 1, 2, 3, ...$$

Of course, these are the same cells and they have the same boundaries, so we have:

$$\partial_k^L := \partial_k \Big|_{C_k(L)}.$$

To make sure that these are well-defined, we need the boundaries of the chains in $C_k(L)$ to be chains in $C_{k-1}(L)$:

$$\partial_k(C_k(L)) \subset C_{k-1}(L).$$

This algebraic condition is ensured by a simple completeness requirement on the set L itself, below.

Definition 1.27. Suppose a set of cells $L \subset \mathbb{R}^N$ satisfies:
- if L contains a cell a, it also contains all of a's faces.

Then L is called a *cubical complex*.

These groups and homomorphisms of the chain complex of L contain *all* topological information encoded in the cubical complex.

Exercise 1.28. For $L = \mathbb{R}^N$, find out when ∂_k is surjective or injective.

1.6 Cycles and boundaries

Following the ideas that have arisen from our study of topology of graphs, we introduce the following definitions for a given cubical complex K and each $k = 0, 1, ...$

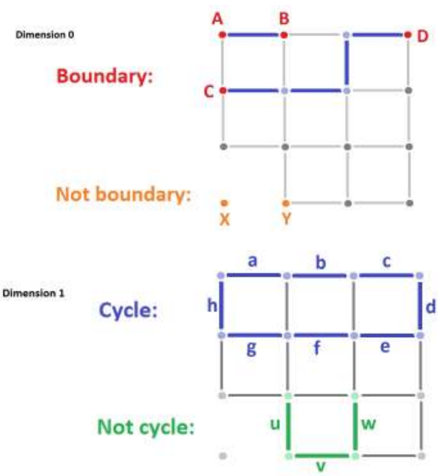
Definition 1.29.
- A k-chain is called a k-*cycle* if its boundary is zero, and the kth *cycle group* is a subgroup of C_k given by

$$Z_k = Z_k(K) := \ker \partial_k.$$

- A k-chain is called a k-*boundary* if it's the boundary of a $(k+1)$-chain, and the kth *boundary group* is a subgroup of C_k given by

$$B_k = B_k(K) := \operatorname{Im} \partial_{k+1}.$$

Example 1.30. Let's see what form the chain complex takes if we limit ourselves to *graphs*. To do that, we simply choose the cubical complex K to consist of only 0- and 1-cells. This is how we visualize boundaries and cycles, in the only two relevant dimensions:



The chain complex is short:

$$0 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0.$$

Of course, the diagram immediately leads to these two "triviality" conditions:
- $\partial_0 = 0$,
- $\partial_2 = 0$.

This is the reason why, for graphs, we only had to deal with a single boundary operator, ∂_1. Another consequence of those two identities is these two familiar observations:
- every 0-chain is a cycle, and
- the only 1-boundary is 0.

In other words,
- $Z_0 = C_0$, and
- $B_1 = 0$.

The result would significantly simplify our analysis. □

Exercise 1.31. What happens to the chain complex and these identities if we add to K a single 2-cell?

In the general case, the last two identities do reappear:
- $Z_0 = C_0$, and
- $B_N = 0$,

as a result of the triviality of the boundary operators at the ends of the chain complex:

$$0 \xrightarrow{\partial_{N+1}=0} C_N \xrightarrow{\partial_N=?} \dots \xrightarrow{\partial_1=?} C_0 \xrightarrow{\partial_0=0} 0.$$

The simplification this brings, however, is limited to these extreme dimensions.

Further, both the cycle groups Z_k as the kernels and the boundary groups B_k as the images of the boundary operators can be connected to each other, as shown in this diagram:

$$\begin{array}{ccccccc}
\to & 0 \in Z_{k+1} = \ker \partial_{k+1} & \xrightarrow{\partial_{k+1}} & 0 \in Z_k = \ker \partial_k & \xrightarrow{\partial_k} & 0 \in Z_{k-1} = \ker \partial_{k-1} & \to \\
 & \cap & & \cap & & \cap & \\
\to & \operatorname{Im} \partial_{k+2} = B_{k+1} \subset C_{k+1} & \xrightarrow{\partial_{k+1}} & \operatorname{Im} \partial_{k+1} = B_k \subset C_k & \xrightarrow{\partial_k} & \operatorname{Im} \partial_k = B_{k-1} \subset C_{k-1} & \to
\end{array}$$

But these pairs of groups have an even more intimate relation...

1.7 Cycles = boundaries?

What is the relation between cycles and boundaries?

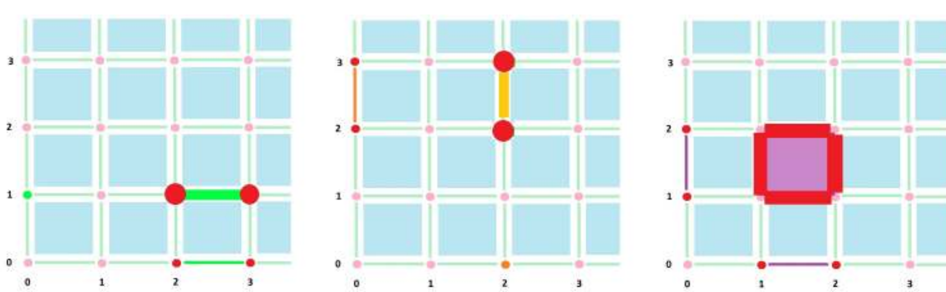
To approach this question, let's try to answer another first:

What is the boundary of a boundary?

Dimension 0. The boundary of a vertex is zero, so the answer is *zero.*

Dimension 1. The boundary of an edge is the sum of its two endpoints and the boundaries of those are zero. So the answer is *zero* again.

Dimension 2. The boundary of a square is the sum of its four edges. The boundary of each of those is the sum of its two endpoints, or the vertices of the square. Each vertex appears twice and... the answer is *zero* again.



Example 1.32. To confirm this algebraically, we compute the boundary of the boundary of a 2-cell:

$$\begin{aligned}&\partial_1\partial_2\Big([n,n+1]\times[m,m+1]\Big)\\&=\partial_1\Big([n,n+1]\times\{m\}+\{n\}\times[m,m+1]+[n,n+1]\times\{m+1\}+\{n+1\}\times[m,m+1]\Big)\\&\dots\\&=0.\end{aligned}$$

□

We can go on.

Dimension 3. The boundary of a cube is the sum of its six square faces. The boundary of each of those is the sum of its four edges. Since these are the edges of the cube and each appears twice, ... the answer is *zero* once again:

Exercise 1.33. Prove that for the n-dimensional case.

So, at least for the cells, the boundary of the boundary is zero. So,

$$\partial_k\partial_{k+1}(x)=0$$

for any $(k+1)$-cell x. Colloquially,

the boundary of a boundary is zero.

Since the $(k+1)$-cells are the generators of the total chain group C_{k+1} and the values of the homomorphism $\partial_k\partial_{k+1}$ are zeroes, it follows that the whole homomorphism is 0:

$$\partial_k\partial_{k+1}\Big(\sum_i s_i\Big)=\sum_i\partial_k\partial_{k+1}(s_i)=\sum_i 0=0.$$

The composition is trivial!

Theorem 1.34 (Double Boundary Identity). The composition of two boundary operators

$$\partial_k\partial_{k+1}:C_{k+1}\to C_{k-1}$$

is zero.

Definition 1.35. Any sequence of groups and homomorphisms:

$$G_N\xrightarrow{f_N}G_{N-1}\xrightarrow{f_{N-1}}\dots\xrightarrow{f_2}G_1\xrightarrow{f_1}G_0,$$

that satisfies $f_kf_{k+1}=0,\ \forall k=1,...,N-1$, is called a *chain complex*.

Exercise 1.36. Give an example of a chain complex with $f_k\neq 0,\ \forall k=2,...,N-1$.

The compact form of the above identity is:

$$\boxed{\partial\partial=0}$$

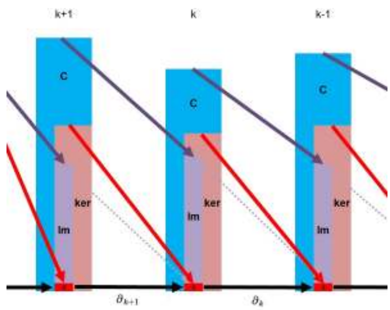
Then the answer to the question posed in the beginning is *every boundary is a cycle.*

In other words, we have the following:

Corollary 1.37.

$$B_k \subset Z_k, \ \forall k = 0, 1, 2, ...$$

The condition $\operatorname{Im} \partial_{k+1} \subset \ker \partial_k$ is commonly illustrated with this informal diagram:



This property is what make defining homology possible: two k-cycles are equivalent, *homologous*, if they form the boundary of a $(k+1)$-chain.

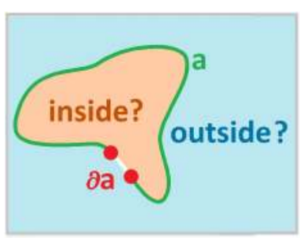
Exercise 1.38. Demonstrate that this definition generalizes the one for graphs.

The chain complex together with the kernels and the images of the boundary operators is given below:

$$\begin{array}{ccccccc}
\to & C_{k+1} & \xrightarrow{\partial_{k+1}} & C_k & \xrightarrow{\partial_k} & C_{k-1} & \to \\
 & \cup & & \cup & & \cup & \\
\to & 0 \in Z_{k+1} = \ker \partial_{k+1} & \xrightarrow{\partial_{k+1}} & 0 \in Z_k = \ker \partial_k & \xrightarrow{\partial_k} & 0 \in Z_{k-1} = \ker \partial_{k-1} & \to \\
 & \cup & & \cup & & \cup & \\
\to & \operatorname{Im} \partial_{k+2} = B_{k+1} \subset C_{k+1} & \xrightarrow{\partial_{k+1}} & \operatorname{Im} \partial_{k+1} = B_k \subset C_k & \xrightarrow{\partial_k} & \operatorname{Im} \partial_k = B_{k-1} \subset C_{k-1} & \to
\end{array}$$

1.8 When is every cycle a boundary?

We have demonstrated that the boundary of every cycle is zero. Indeed, it is plausible that a chain can't possibly enclose something inside if it has a boundary. Informally, this is because at the boundary is where the inside would be connected to the outside and, therefore, there is no inside and outside!



So,

$$\text{not cycle} \Longrightarrow \text{not boundary},$$

or

$$\text{boundary} \Longrightarrow \text{cycle},$$

or

$$B_k \subset Z_k.$$

Now, the other way around:

is every cycle a boundary?

Not if there is some kind of a hole in the space! Then let's limit ourselves to the whole grid $\mathbb{R}^N$, which has no topological features: components, tunnels, voids, etc.; so, it may be true.

We start with 1-cycles. The question we are asking is illustrated below:

The picture suggests that the answer is Yes. Indeed, a rubber band standing on its edge would enclose a certain area:

To specify what the band bounds, we fill it with water to make it look like an above-the-ground pool:

But what if the rubber band is twisted?

In $\mathbf{R}^3$, it is better to think of this problem as a soap film spanned on a possibly curved frame:

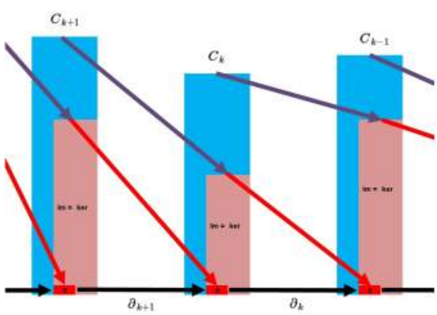
The answer still seems to be Yes, but try to stretch a film on a frame that isn't just curved but knotted too:

We will see that the answer is still "Yes, in a Euclidean space of any dimension cycles are boundaries". The conclusion applies equally to chains of all dimensions. For example, an air-balloon when closed (cycle!) will keep the air inside (boundary!). Or one can think instead of a wrapper containing something solid:

Our conclusion affects the chain complex of the entire grid $\mathbb{R}^N$ as follows:

$$\operatorname{Im} \partial_{k+1} = \ker \partial_k, \ k > 0.$$

In other words, the image isn't just a subset of, but is equal to the kernel, *exactly*. The diagram is much simpler than in the general case:



Definition 1.39. A sequence of groups and homomorphisms $f_k : G_k \to G_{k-1}$ is called *exact* if $\operatorname{Im} f_{k+1} = \ker f_k$.

All exact sequences are chain complexes but not vice versa.

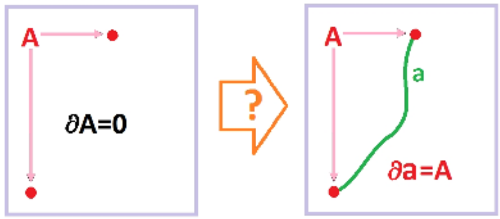
Exercise 1.40. Complete these as exact sequences:

$$\begin{array}{ccccccccc} 0 & \xrightarrow{\ 0\ } & ? & \xrightarrow{\ ?\ } & \mathbf{R}^n & \xrightarrow{\ projection\ } & \mathbf{R}^m & \xrightarrow{\ 0\ } & 0, \\ 0 & \xrightarrow{\ 0\ } & 2\mathbf{Z} & \hookrightarrow & \mathbf{Z} & \xrightarrow{\ ?\ } & ? & \xrightarrow{\ 0\ } & 0, \end{array}$$

where $2\mathbf{Z}$ is the group of even integers.

In general, the chain complex of a proper subset of the grid is *inexact* as the presence of topological features causes some cycles not to be boundaries. The degree of this inexactness will be measured via homology.

But what about the simplest case – chains of dimension $k = 0$?



This case is much easier to visualize:

Mathematically, it's also simple. On the one hand, any one of the vertices on the grid is a cycle. On the other, every two vertices can be connected by a 1-chain (just go vertical then horizontal).

Proposition 1.41. If 0-chain a has an even number of vertices present, then it is the boundary of some $s \in C_1$.

Exercise 1.42. Prove the proposition.

But what about the chains with an odd number of vertices, such as a single-vertex 0-chain? It's not a boundary! The reason is that the boundary operator of any 1-chain will always produce an even number of vertices, including the 0 chain.

So, the answer to our question for 0-cycles is "No, some of them aren't boundaries". As it turns out, the 0 dimension is an exception.

To make the 0 dimension unexceptional, one can modify the definition of 0-cycle to reduce them to those that contain an even number of vertices:

$$\tilde{Z}_0 := \{z \in Z_0 : \ z = r_1c_1 + ... + r_nc_n, r_1 + ... + r_n = 0\}.$$

Exercise 1.43. Prove that that all non-boundaries are homologous to each other.

---

2 Cubical complexes

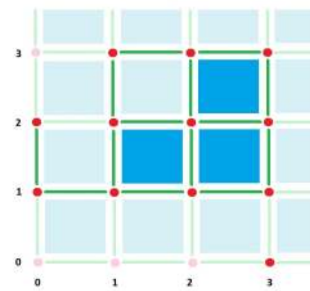
2.1 The definition

For objects located in a Euclidean space, we would like to devise a data structure that we can use to first represent and then topologically analyze them.

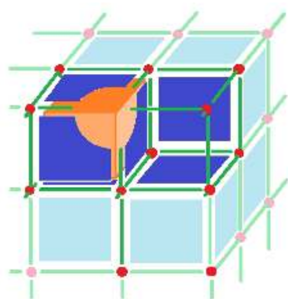
Suppose the Euclidean space $\mathbf{R}^N$ is given and so is its cubical grid $\mathbf{Z}^N$. Suppose also that we have its decomposition $\mathbb{R}^N$, a list of all (closed) cubical cells in our grid.

Definition 2.1. A cubical complex is a collection of cubical cells $K \subset \mathbb{R}^N$ for which the boundary operator is well-defined; i.e., K includes all faces of the cells it contains.

$N = 2:$



$N = 3:$



Example 2.2. The cubical complex K of the pixel at the origin is given by a list of cells of all dimensions:

- 0. $\{0\} \times \{0\},\ \{1\} \times \{0\},\ \{0\} \times \{1\}, \{1\} \times \{1\}$;
- 1. $\{0\} \times [0,1],\ [0,1] \times \{0\},\ [0,1] \times \{1\},\ \{1\} \times [0,1]$;
- 2. $[0,1] \times [0,1]$.

Now, their boundaries are defined *within* the complex:

- 0. $\partial\Big(\{(0,0)\}\Big) = 0$, etc.,
- 1. $\partial\Big(\{0\} \times [0,1]\Big) = \{(0,0)\} + \{(0,1)\}$, etc.,
- 2. $\partial\Big([0,1] \times [0,1]\Big) = [0,1] \times \{0\} + \{0\} \times [0,1] + [0,1] \times \{1\} + \{1\} \times [0,1]$. □

Note: Cell decomposition of digital images produces cubical complexes. These complexes however are special in the sense that their 1-cells can only appear as boundaries of its 2-cells. In general, this is not required.

For $N = 3$, we have:

- 0-cell $\{n\} \times \{m\} \times \{k\}$;
- 1-cell $\{n\} \times \{m\} \times [k, k+1]$, or $\{n\} \times [m, m+1] \times \{k\}$, or $\{n\} \times \{m\} \times [k, k+1]$;
- 2-cell $[n, n+1] \times [m, m+1] \times \{k\}$, or $\{n\} \times [m, m+1] \times [k, k+1]$, or $[n, n+1] \times \{m\} \times [k, k+1]$;
- 3-cell $[n, n+1] \times [m, m+1] \times [k, k+1]$;
- etc.

Exercise 2.3. Provide a list representation of the complex of the unit cube.

Recall that, more generally, a cubical cell in the N-dimensional space is the product of vertices and edges:

$$P := A_1 \times A_2 \times A_3 \times ... \times A_N,$$

where $A_i = [n_i, n_i + 1]$ or $A_i = \{n_i\}$. If the former case happens m times, this cell is an *m-cube*.

The boundary cells, also known as *$(m-1)$-faces*, of this m-cell are certain $(m-1)$-cells that can be computed from the above representation of the cube P. For each edge $A_i = [n_i, n_i + 1]$ in the product, we define a pair of *opposite faces* of P:

$$\begin{array}{ll} P_i^- & := B_1 \times B_2 \times B_3 \times ... \times B_N, \\ P_i^+ & := C_1 \times C_2 \times C_3 \times ... \times C_N, \end{array}$$

where
- $B_k = C_k = A_k$ for $k \neq i$,
- $B_i = \{n_i\}$, and
- $C_i = \{n_i + 1\}$.

Then the list of faces of P is

$$\{P_i^-, P_i^+ : \ A_i = [n_i, n_i + 1], \ i = 1, 2, ..., N\}.$$

It follows that the total number of $(m-1)$-faces is $2m$.

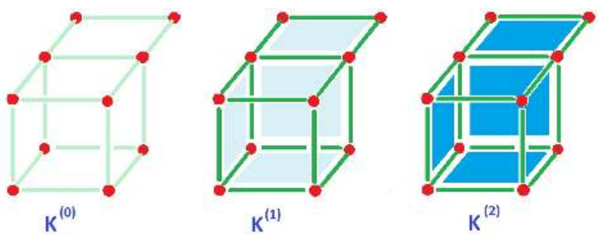
One can define k-faces, $k < m$, of P inductively, as faces of faces.

Exercise 2.4. Give a direct construction of (a) $(m-1)$-faces of m-cube P, (b) all k-faces for $k < m$ of P. (c) How many of each dimension?

Exercise 2.5. Define *cubical maps* in such a way that together with cubical complexes they form a category.

Definition 2.6. The *dimension* $\dim K$ *of a cubical complex* K is the highest among the dimensions of its cells.

Definition 2.7. For a given n, the collection of all k-cells with $k \leq n$ of complex K is called the *n-skeleton* of K **denoted** by $K^{(n)}$.

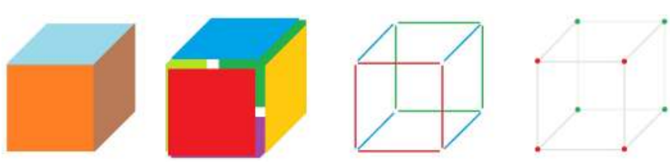


The sequence of skeleta can be understood as a blueprint of the complex or even a step-by-step algorithm for building it, from the ground up:

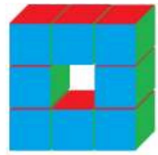
$$K^{(0)} \subset K^{(1)} \subset \ ... \ \subset K^{(N-1)} \subset K^{(N)} = K.$$

Exercise 2.8. Show that skeleta are also cubical complexes and $\dim K^{(n)} \leq n$.

Once the skeleta of the cube are found, we can use them to build new things:



For example, the skeleta of this solid torus are built from those of the cube:



Just keep in mind that the shared vertices, edges, faces, etc. appear only once.

Exercise 2.9. Find the cubical complex representation of the regular, hollow, torus.

2.2 Realizations

Recall that a realization of a graph G is a topological space $|G|$ with
- a point for each node of G, and
- a path for each edge of G between those points that doesn't intersect the other paths except at the endpoints.

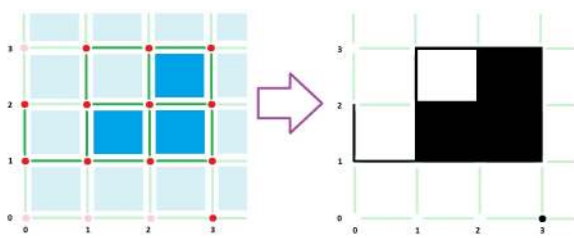
This construction looks somewhat similar to that of the 1-skeleton.

The idea is the same for cubical complexes. A realization of a cubical complex K is a topological space $|K|$ with
- a point for each vertex of K, and
- a path for each edge of K,
- and so on for all cells of K,

put together according to the data encoded in K. What makes things so much simpler is the fact that the cells come directly from $\mathbb{R}^N$ and there is only one way to put them together.

Definition 2.10. The union of the cells of a given cubical complex K is called its *realization*:

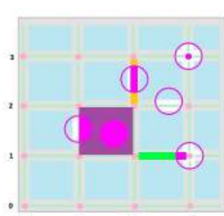
$$|K| := \cup K.$$



A cubical complex is called *finite* when it has a finite number of cells. Meanwhile, $\mathbb{R}^N$ is an infinite cubical complex, and

$$|\mathbb{R}^N| = \mathbf{R}^N.$$

Recall that an open k-cell is homeomorphic to $\mathbf{R}^k$. Both open and closed cells are subsets of the Euclidean space $\mathbf{R}^N$ and that's where their topology comes from.



Exercise 2.11. Explain the difference between the *boundary* of a cubical cell and its frontier.

These are the "open" cells:

- 0. $\{0\}\times\{0\},\{1\}\times\{0\},\{0\}\times\{1\},\{1\}\times\{1\}$,
- 1. $\{0\}\times(0,1),(0,1)\times\{0\},(0,1)\times\{1\},\{1\}\times(0,1)$,
- 2. $(0,1)\times(0,1)$.

Exercise 2.12. Show that we can replace "closed" with "open" in the definition of cubical complex and its realization won't change.

It follows then that, even though cells may be *open*, the realization produced is *closed* (which is a result of the complex being closed under the boundary operator). The difference is, as you can see, that this latter approach results in a partition of the realization.

Theorem 2.13. The realization of a cubical complex is a closed subset of $\mathbf{R}^N$.

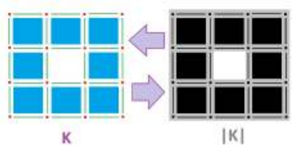
Exercise 2.14. The conclusion is obvious for a finite cubical set as the finite unions of its closed cells. What about infinite? Hint: unlike the union of $[-1/n,1/n]$, $n>0$, the union of cells doesn't produce *new* limit points. This kind of collection is called "locally finite" (why?).

Proposition 2.15. The realization of a finite cubical complex is bounded.

How do we use cubical complexes? If we want to study the topology of a subset of the Euclidean space, we do that by representing it as a realization of a cubical complex, if possible:

- Given $X\subset\mathbf{R}^n$, find a cubical complex K such that $|K|=X$, evaluate its homology, then set $H(X):=H(K)$.

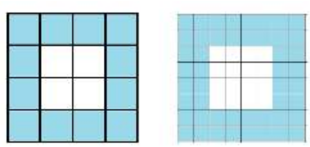
In a sense, a complex and its realization are *the same*:



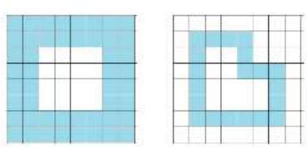
Indeed, a cubical complex yields a subset of $\mathbf{R}^N$ via realization but that union of cubical cells can be decomposed in one and only one way producing the same cubical complex. The distinction should still be maintained:

- K is a collection of subsets of the Euclidean space: $K\subset 2^{\mathbf{R}^N}$, while
- $|K|$ is a collection of points in the Euclidean space: $|K|\subset\mathbf{R}^N$.

The deceptively simple idea of realization conceals some difficult issues. To begin with, it is possible that $X=|K|=|L|$ is the realization of either of two *different cubical complexes* K,L created with two *different grids* of $\mathbf{R}^N$:



For our analysis to make sense, we'll have to show that the resulting homology is the same: $H(K)\cong H(L)$. Moreover, the homology groups should be the same for topologically equivalent spaces:



The following statement will be our beacon:

Theorem 2.16 (Invariance of homology). Suppose two homeomorphic subsets X, Y of $\mathbf{R}^N$ are realizations

$$X = |K|, \ = |L|$$

of two cubical complexes K, L, then they have isomorphic homology groups:

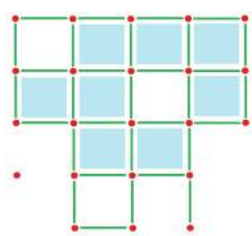
$$|K| \approx |L| \Longrightarrow H(K) \cong H(L).$$

Exercise 2.17. Prove that any tree or any plane graph can be represented as a one-dimensional cubical complex. Give an example of a graph that cannot be represented by a cubical complex.

Exercise 2.18. Each open cell has Euclidean topology. But is the topology formed on a cubical set as the disjoint union of these cells Euclidean?

2.3 The boundary operator

A cubical complex represents a figure in a finite form. It is a list of "cells" combined with the information on how they are attached to each other.

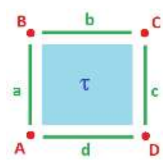


Now, a k-chain is a combination (a formal sum) of finitely many k-cells, such as $a + b + c$.

The *boundary operator* ∂ represents the relation between chains of consecutive dimensions that captures the topology of the cubical complex. We often specify the dimension of the cell involved: $\partial_k(x)$ stands for the boundary of a k-chain x.

Let's review.

Example 2.19.



The boundary of a vertex is empty, so the boundary operator of a 0-chain is 0:

$$\partial_0(A) = 0.$$

The boundary of a 1-cell consists of its two endpoints, so we have:

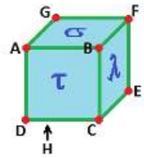
$$\partial_1(a) = \partial(AB) = A + B.$$

The boundary of a 2-cell consists of its four edges, so we have:

$$\partial_2(\tau) = \partial(ABCD) = a + b + c + d.$$

□

Example 2.20. Consider the cube:



In dimension 3, there is only one

- 3-cell: $\alpha = ABCDEFG$.

Next,

- 2-cells: $\tau = ABCD,\ \sigma = BCFG,\ \lambda = CFED$.

To compute the boundary of α, we need to express its faces in terms of these 2-cells:

$$\begin{aligned}\partial\alpha &= ABCD + BCFG + CDEF + 3 \text{ more}\\ &= \tau + \sigma + \lambda + (3 \text{ more}).\end{aligned}$$

□

Suppose the ambient dimension N is fixed, as well as the grid of the Euclidean space $\mathbf{R}^N$ and the "standard cubical complex" $\mathbb{R}^N$ of $\mathbf{R}^N$. Suppose $K \subset \mathbb{R}^N$ is a cubical complex.

Notation: We denote $C_k = C_k(K)$ the set of all k-chains in K.

We will call $C_k(K)$, as before, the kth *chain group of cubical complex* K. Its relation to the previously discussed "total" complex is simple.

Theorem 2.21. The chain group $C_k(K)$ of a cubical complex K is the subgroup of $C_k(\mathbb{R}^N)$ generated by the k-cells in K.

Given a k-chain $s \in C_k(K)$,

$$s = a_1 + a_2 + ... + a_n,$$

where $a_1, a_2, ..., a_n$ are k-cells in K, the *boundary* ∂s of s is given by

$$\partial s := \partial a_1 + \partial a_2 + ... + \partial a_n.$$

What we know is that the boundaries of k-cells are $(k-1)$-chains, and, moreover, the boundaries of k-chains are $(k-1)$-chains as well. Now, as a cubical complex, if K contains a cell s, it also contains all of s's faces. Hence,

$$s \in K \Longrightarrow \partial s \in C_k(K),$$

and

$$s \in C_k(K) \Longrightarrow \partial s \in C_k(K).$$

So, the *restrictions* of the boundary operators that we defined for the whole space are now the boundary operators of K. They are

$$\partial_k^K := \partial_k\Big|_{C_k(K)} : C_k(K) \to C_{k-1}(K),\ k = 0, 1, 2, ...$$

We use the same notation, ∂, for these functions, whenever possible.

Since this is a restriction of a homomorphism to a subgroup, we have the following.

Theorem 2.22. The boundary operator is a homomorphism.

Moreover, it follows that the main property,

every boundary is a cycle,

of the boundary operator remains true for this restriction.

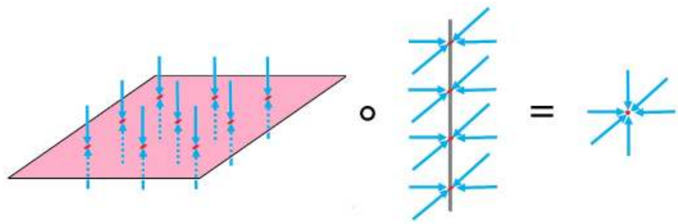
Theorem 2.23 (Double Boundary Identity). The composition of two boundary operators

$$\partial_k \partial_{k+1} : C_{k+1}(K) \to C_{k-1}(K),\ k = 1, 2, 3, ...$$

is zero. Or, simply put:

$$\partial\partial = 0.$$

Example 2.24. As an illustration, consider this example of two operators below. Neither operator is 0, but their composition is:



Here,

- A is the projection on the xy-plane, which isn't 0;
- B is the projection on the z-axis, which isn't 0;
- BA is 0. □

2.4 The chain complex

The rest of the definitions from the last section also apply.

Definition 2.25. For a given $k = 0, 1, 2, 3, ...,$

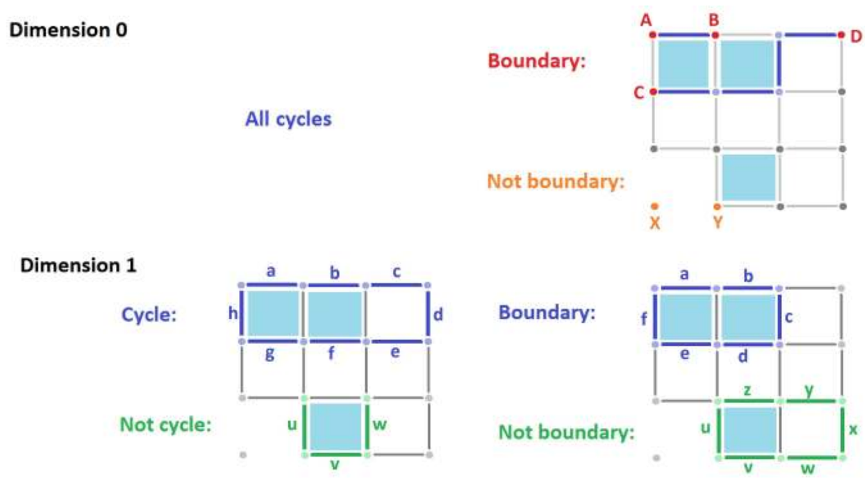
- the kth *cycle group* of K is the subgroup of $C_k = C_k(K)$ defined to be

$$Z_k = Z_k(K) := \ker \partial_k;$$

- the kth *boundary group* of K is the subgroup of $C_k(K)$ defined to be

$$B_k = B_k(K) := \operatorname{Im} \partial_{k+1}.$$

This is how cycles and boundaries are visualized:



Exercise 2.26. For the complex shown above, sketch a few examples of cycles and boundaries.

As before, the big picture of the algebra of chains and their boundaries is given by the *chain complex of the cubical complex* K (an unfortunate reuse of the word "complex"):

$$... \xrightarrow{\partial_{k+2}} C_{k+1}(K) \xrightarrow{\partial_{k+1}} C_k(K) \xrightarrow{\partial_k} C_{k-1}(K) \xrightarrow{\partial_{k-1}} ... \xrightarrow{\partial_1} C_1(K) \xrightarrow{\partial_0} 0.$$

Remember, it is the property $\partial\partial = 0$ what makes this sequence a chain complex.

Corollary 2.27. For any cubical complex K,

$$B_k(K) \subset Z_k(K), \ \forall k = 0, 1, 2, ...$$

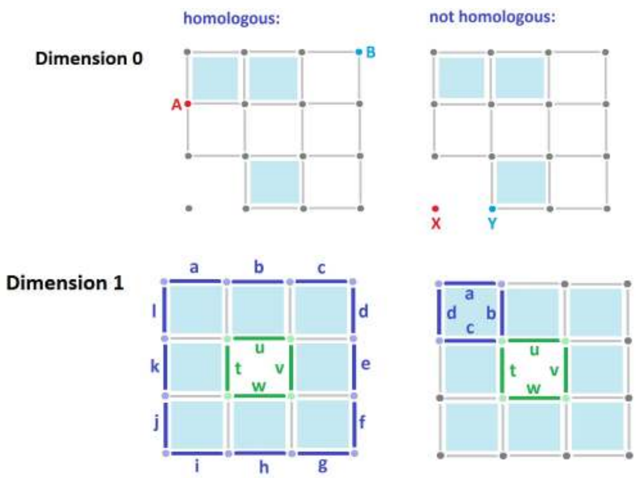
That's what makes defining homology groups possible.

Definition 2.28. The kth *homology group*, $k = 0, 1, 2, 3, ...$, of a cubical complex K is defined to be

$$H_k = H_k(K) := Z_k(K)/B_k(K).$$

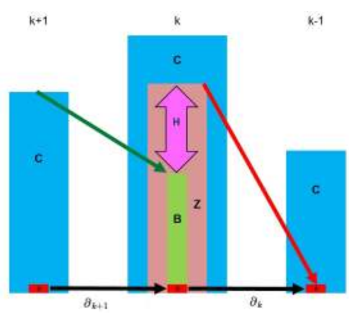
Exercise 2.29. Demonstrate that this definition generalizes the one for graphs.

According to the definition, two k-cycles are equivalent, *homologous*, if they form the boundary of a $(k+1)$-chain. This is how the idea is visualized:



Exercise 2.30. For the complex shown above, sketch a few examples of homologous and non-homologous cycles.

For a given k, the part of the chain complex that affects Z_k, B_k, H_k is often illustrated with this diagram:



Then H_k captures the *difference* between Z_k and B_k.

Exercise 2.31. (a) Prove that

$$H_m(K) = H_m(K^{(m+1)}).$$

(b) Give an example that shows that replacing $K^{(m+1)}$ with $K^{(m)}$ fails.

We will need the following classification theorem as a reference.

Theorem 2.32 (Fundamental Theorem of Finitely Generated Abelian Groups). Every finitely generated abelian group G is isomorphic to a direct sum of primary cyclic groups and infinite cyclic groups:

$$\mathbf{Z}^n \oplus \mathbf{Z}_{q_1} \oplus ... \oplus \mathbf{Z}_{q_s},$$

where $n \geq 0$ is the *rank* of G and the numbers $q_1, ..., q_s$ are powers of (not necessarily distinct) prime numbers. Here $\mathbf{Z}^n$ is the *free part* and the rest is the *torsion part* of G.

Proposition 2.33. For a finite cubical complex K, the groups

$$C_k(K), Z_k(K), B_k(K), H_k(K)$$

are direct sums of finitely many copies of $\mathbf{Z}_2$:

$$\mathbf{Z}_2 \oplus \mathbf{Z}_2 \oplus ... \oplus \mathbf{Z}_2.$$

Exercise 2.34. Prove the proposition. Hint: what is the order of the elements?

The number of the factors in such a group G is also the number of linearly independent generators. That's why, for the rest of this section, we will refer to this number as the *dimension* $\dim G$ of the group (after all it is a vector space over $\mathbf{Z}_2$).

Exercise 2.35. Find the formula for $\dim G$ in terms the number of elements $\#G$ of G.

Definition 2.36. The dimension of $H_k(K)$ is the number of topological features in K of dimension k, the kth *Betti number*:

$$\beta_k := \dim H_k(K).$$

Recall that to make the 0 dimension unexceptional, we modified the definition of 0-cycle to reduce them to those that contain an even number of vertices:

$$\tilde{Z}_0 := \{z \in Z_0 : \ z = r_1c_1 + ... + r_nc_n, r_1 + ... + r_n = 0\}.$$

The result is the *reduced homology* that is sometimes more convenient:

$$\tilde{H}_0 := \tilde{Z}_0 / B_0.$$

In particular, the homology of the whole space and of a single point is trivial, in all dimensions:

$$\tilde{H}_0(\{p\}) = \tilde{H}_0(\mathbf{R}^N) = 0.$$

In other words, the reduced homology groups are defined via the identical formulas but applied to the *augmented chain complex*:

$$... \xrightarrow{\partial_k} C_{k-1}(K) \xrightarrow{\partial_{k-1}} ... \xrightarrow{\partial_1} C_1(K) \xrightarrow{\varepsilon} \mathbf{Z}_2 \xrightarrow{0} 0,$$

with ε given by

$$\varepsilon(r_1c_1 + ... + r_nc_n) := r_1 + ... + r_n, \ r_i \in \mathbf{Z}_2.$$

Then, in particular,

$$\tilde{H}_k(K) = H_k(K), \ k = 1, 2, ...,$$

but

$$\dim \tilde{H}_0(K) = \dim H_k(K) - 1.$$

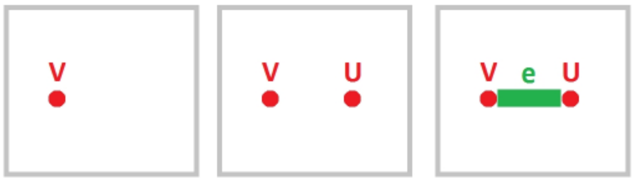
2.5 Examples

All these groups can be represented as *lists*!

Indeed, $C_k(K)$ is generated by the finite set of k-cells in K and $Z_k(K), B_k(K)$ are its subgroups. Therefore, they all have only finite number of elements. Meanwhile, their quotient $H_k(K)$ lists the cosets, so *homology is a list of lists.*

These lists can be accompanied by illustrations, for small enough cubical complexes, with each cell on those lists shown. Yet, below we will intentionally ignore the pictures – as soon as the chain complex is found. The goal is to demonstrate that the second step – computing $Z_k(K), B_k(K), H_k(K)$ – is purely algebraic.

We will start with these three, progressing from the simplest to more complex, in order to reuse our computations:



Example 2.37 (single vertex). Let $K := \{V\}$. Then

$$\begin{array}{lll} C_0 & =< V > & = \{0, V\}, \\ C_i & & = 0, \qquad \forall i > 0. \end{array}$$

Then we have the whole chain complex here:

$$0 \xrightarrow{\partial_2=0} C_1 = 0 \xrightarrow{\partial_1=0} C_0 =< V > \xrightarrow{\partial_0=0} 0.$$

From this complex, working algebraically, we deduce:

$$\begin{array}{llll} Z_0 := & \ker \partial_0 & =< V > & = \{0, V\}, \\ B_0 := & \operatorname{Im} \partial_1 & & = 0. \end{array}$$

Hence,

$$H_0 := Z_0/B_0 =< V > /0 =< [V] >= \{[0], [V]\} = \{\{0\}, \{V\}\},$$

and $\dim H_0 = 1$; i.e., K has a single path-component. Also,

$$H_i = 0, \ \forall i > 0,$$

so no holes. □

Example 2.38 (two vertices). Let $K := \{V, U\}$. We copy the last example and make small corrections:

$$\begin{array}{lll} C_0 & =< V, U > & = \{0, V, U, V + U\}, \\ C_i & & = 0, \qquad \forall i > 0. \end{array}$$

Then we have the whole chain complex here:

$$0 \xrightarrow{\partial_2=0} C_1 = 0 \xrightarrow{\partial_1=0} C_0 =< V, U > \xrightarrow{\partial_0=0} 0.$$

Now using only algebra, we deduce:

$$\begin{array}{llll} Z_0 := & \ker \partial_0 & =< V, U > & = \{0, V, U, V + U\}, \\ B_0 := & \operatorname{Im} \partial_1 & & = 0. \end{array}$$

Hence,

$$\begin{array}{llll} H_0 & := Z_0/B_0 & =< V,U > /0 & =< [V],[U] > \\ & = \{[0],[V],[U],[V+U]\} & = \{\{0\},\{V\},\{U\},\{V+U\}\}. & \end{array}$$

So, as the rank of this group is 2, K has two path-components. Also,

$$H_i = 0,\ \forall i > 0,$$

so no holes. □

Example 2.39 (two vertices and an edge). Let $K := \{V, U, e\}$. We copy the last example and make some corrections:

$$\begin{array}{lll} C_0 & =< V,U > & = \{0, V, U, V+U\}, \\ C_1 & =< e > & = \{0, e\}, \\ C_i & = 0, & \forall i > 1. \end{array}$$

Then we have the whole chain complex shown:

$$0 \xrightarrow{\partial_2=0} C_1 =< e > \xrightarrow{\partial_1=?} C_0 =< V,U > \xrightarrow{\partial_0=0} 0.$$

First we compute the boundary operator:

$$\partial_1(e) = V + U.$$

Hence,

$$\partial_1 = [1,1]^T.$$

Now the groups.

Dimension 0 (no change in the first line):

$$\begin{array}{llll} Z_0 := & \ker \partial_0 & =< V,U > & = \{0, V, U, V+U\}, \\ B_0 := & \operatorname{Im} \partial_1 & =< V+U > & = \{0, V+U\}. \end{array}$$

Notice the inexactness of our chain complex: $Z_0 \neq B_0$ (not every cycle is a boundary!). It follows,

$$H_0 := Z_0/B_0 =< V,U > / < V+U >=< [V] >= \big\{\{0, V+U\}, \{V,U\}\big\}.$$

So, with the dimension of this group equal to 1, complex K has one component.

Dimension 1:

$$\begin{array}{lll} Z_1 := & \ker \partial_1 & = 0, \\ B_1 := & \operatorname{Im} \partial_2 & = 0, \end{array}$$

therefore

$$H_1 = 0/0 = 0.$$

Still no holes. □

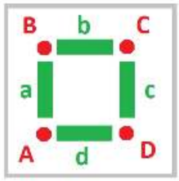
The transition from the second example to the third illustrates how a 0-cycle can be "killed" by adding a new edge. Let's consider this idea in more generality.

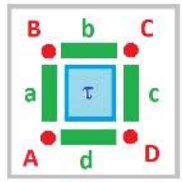
Theorem 2.40. Suppose K is a cubical complex and suppose $L = K \cup \{e\}$, where $e \notin K$ is an edge, is another. Suppose $e = UV$, where U, V are two vertices of K. Then there are two cases:

- Case 1: $U \sim V$ in K, then
- ⋄ $\dim H_0(K) = \dim H_0(L)$ (no new 0-boundaries),
- ⋄ $\dim H_0(K) + 1 = \dim H_0(L)$ (a new 1-cycle);
- Case 2: $U \not\sim V$ in K, then
- ⋄ $\dim H_0(K) - 1 = \dim H_0(L)$ (a new 0-boundary),
- ⋄ $\dim H_0(K) = \dim H_0(L)$ (no new 1-cycles).

Exercise 2.41. Prove the theorem.

Two, slightly more complex, examples:





Example 2.42 (hollow square). Let $K := \{A, B, C, D, a, b, c, d\}$. Then we have (too many cells to list all elements):

$$\begin{array}{ll} C_0 & =< A, B, C, D >, \\ C_1 & =< a, b, c, d >, \\ C_i & = 0, \quad \forall i > 1. \end{array}$$

Note that we can think of these two lists of generators as *ordered* bases.

We have the chain complex below:

$$0 \xrightarrow{\partial_2 = 0} C_1 =< a, b, c, d > \xrightarrow{\partial_1 = ?} C_0 =< A, B, C, D > \xrightarrow{\partial_0 = 0} 0.$$

First we compute the boundary operator:

$$\begin{array}{ll} \partial_1(a) & = A + B, \\ \partial_1(b) & = B + C, \\ \partial_1(c) & = C + D, \\ \partial_1(d) & = D + A. \end{array}$$

Hence,

$$\partial_1 = \left[\begin{array}{cccc} 1, & 0, & 0, & 1 \\ 1, & 1, & 0, & 0 \\ 0, & 1, & 1, & 0 \\ 0, & 0, & 1, & 1 \end{array} \right].$$

The chain complex has been found, now the groups.

Dimension 0:

$$\begin{array}{llll} Z_0 := & C_0 & & =< A, B, C, D >, \\ B_0 := & \operatorname{Im} \partial_1 & =< A + B, B + C, C + D, D + A > & =< A + B, B + C, C + D > \end{array}$$

(because $D + A$ is the sum of the rest of them). It follows,

$$\begin{array}{lll} H_0 := & Z_0/B_0 & =< A, B, C, D > / < A + B, B + C, C + D > \\ & =< [A] > & = \{B_0, \{A, B, C, D\}\}. \end{array}$$

So, with the dimension of this group equal to 1, complex K has one path-component.

Dimension 1:

$$\begin{array}{lll} Z_1 := & \ker \partial_1 & =?, \\ B_1 := & \operatorname{Im} \partial_1 & = 0. \end{array}$$

To find the kernel, we need to find all $e = (x, y, z, u) \in C_1$ such that $\partial_1 e = 0$. That's a (homogeneous) system of linear equations:

$$\left\{ \begin{array}{lllll} x & & & +u & = 0, \\ x & +y & & & = 0, \\ & y & +z & & = 0, \\ & & z & +u & = 0. \end{array} \right.$$

Solving from bottom to the top:

$$z = -u \Longrightarrow y = u \Longrightarrow x = -u,$$

so $e = (-u, u, u, -u)^T,\ u \in \mathbf{R}$. Then, as signs don't matter,

$$Z_1 =< e >=< (1,1,1,1) >=< a+b+c+d > .$$

Hence,

$$H_1 = Z_1/0 =< [a+b+c+d] > .$$

There is a hole! □

Example 2.43 (solid square). Let $K := \{A, B, C, D, a, b, c, d, \tau\}$. We copy the last example and make some corrections. We have:

$$\begin{array}{ll} C_0 & =< A, B, C, D >, \\ C_1 & =< a, b, c, d >, \\ C_2 & =< \tau >, \\ C_i & = 0, \quad \forall i > 2. \end{array}$$

We have the chain complex:

$$0 \xrightarrow{\partial_3=0} C_2 =< \tau > \xrightarrow{\partial_2=?} C_1 =< a, b, c, d > \xrightarrow{\partial_1} C_0 =< A, B, C, D > \xrightarrow{\partial_0=0} 0.$$

First we compute the boundary operator:

$$\partial_2(\tau) = a + b + c + d,$$

therefore,

$$\partial_2 = [1, 1, 1, 1]^T.$$

As ∂_1 is already known, the chain complex has been found. Now the groups:

Dimension 0. Since the changes in the chain complex don't affect these groups, we have answers ready:

$$\begin{array}{lll} Z_0 := & C_0 & =< A, B, C, D >, \\ B_0 := & \operatorname{Im} \partial_1 & =< A+B, B+C, C+D >, \\ H_0 = & < [A] > & = \big\{B_0, \{A, B, C, D\}\big\}. \end{array}$$

So, again, K has one component.

Dimension 1 (no change in the first line):

$$\begin{array}{llll} Z_1 := & \ker \partial_1 & & =< a+b+c+d >, \\ B_1 := & \operatorname{Im} \partial_1 & =< \partial_2(\tau) > & =< a+b+c+d > . \end{array}$$

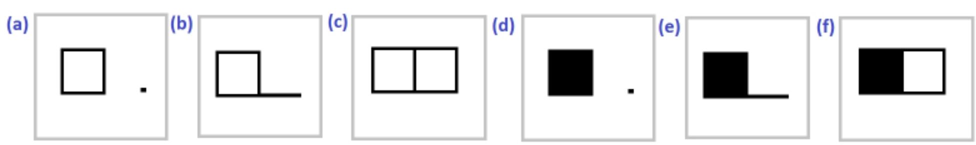
Hence,

$$H_1 = 0.$$

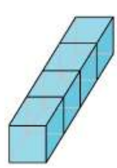
There is no hole! □

Exercise 2.44. Represent the sets below as realizations of cubical complexes. In order to demonstrate that you understand the algebra, for each of them:

- (a) find the chain groups and find the boundary operator as a matrix;
- (b) using only part (a) and algebra, find Z_k, B_k, H_k for all k, including the generators.

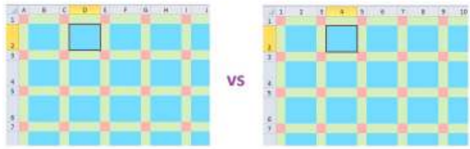


Exercise 2.45. Compute the homology of a "train" with n cars:



2.6 Computing boundaries with a spreadsheet

Cells on the plane can be presented in a spreadsheet such as Excel. Below, some of the rows and columns are narrowed in order to emphasize the 1-dimensional nature of the edges and the 0-dimensional nature of the vertices:



The cells are represented:
- left: in the standard style, and
- right: in the `R1C1` reference style.

The highlighted cell is given by
- `D2`, and
- `R2C4`,

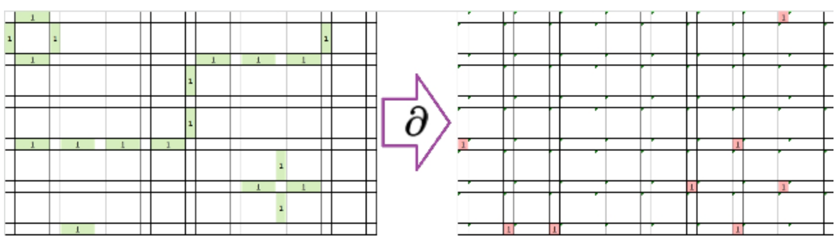
respectively. The latter is clearly more mathematical and that's the one we will use. Notice, however, that *the coordinate axes go down and right* as in a matrix, instead of right and up as we are used to in the Cartesian system.

Also, the coordinates differ from those in the above discussion by a factor of two: the highlighted cell is located at $(1,2)$. Then
- 0-cells are (odd, odd),
- 1-cells are $(odd, even)$ and $(even, odd)$,
- 2-cells are $(even, even)$.

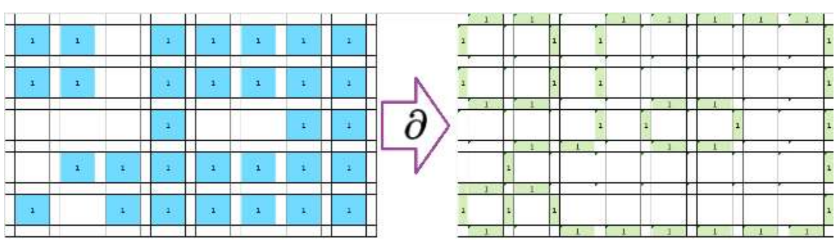
The chains are encoded by putting 0s and 1s in all cells. In the former case, the cell is white and the number is invisible while in the latter case, the cell is colored according to its dimension.

Next, we implement the boundary operator.

These are the results of computing boundaries, for a 1-chain (green):



and a 2-chain (blue):



The approach to the algorithm for computing the boundary operator is very different from the algebra we have done. Compare the following.

- Normally, to compute the boundary of a k-chain, we evaluate the boundary of each of its k-cells as the sum of this cell's boundary cells (dimension $k-1$), add them, and then cancel the repetitions.

• Now, we look at each $(k-1)$-cell present in the spreadsheet, add the values of the k-cells of the chain adjacent to it, and then find out whether the result makes this cell a part of the boundary of the chain.

The results are of course the same.

The code is *very simple.*

Dimension 0: For each vertex, we compute the sum, modulo 2, of the values at the four edges adjacent to it:

```
= MOD( RC[-21] + R[-1]C[-20] + RC[-19] + R[1]C[-20], 2)
```

Then the result is placed on the right. It follows that,

- if the curve isn't there, we have $0+0+0+0=0$;
- if the curve is passing by once, we have $1+0+1+0=0$;
- if the curve is passing by twice, we have $1+1+1+1=0$;
- if the curve ends here, we have $1+0+0+0=1$.

Dimension 1: For each edge, we, similarly, compute the sum of the values at the two adjacent faces; horizontal:

```
= MOD( R[-1]C[-20] + R[1]C[-20], 2)
```

and vertical:

```
= MOD( RC[-21] + RC[-19], 2)
```

Then,

- if the edge is away from the region, we have $0+0=0$;
- if the edge is inside the region, we have $1+1=0$;
- if the edge is on the boundary of the region, we have $1+0=1$.

Note: See the files online.

Exercise 2.46. Modify the code to compute the homology groups of graphs as 1-dimensional subcomplexes of $\mathbb{R}^2$.

Exercise 2.47. Devise and implement a similar algorithm for 3-chains. Hint: use the "sheets" as layers.

---

3 The algebra of oriented cells

3.1 Are chains just "combinations" of cells?

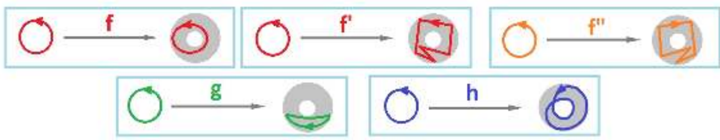
Consider the following *topological* problem:

How many times does a rubber band go around the finger (or a pony-tail)?

The nature of the problem is topological because the *rubber* band tends to stretch and shrink on its own and the answer remains the same.

But do we even understand the meaning of "go around"? After all, the band might go back and forth, cancelling some of the wraps. Intuitively, we can allow the band to fully contract and then count... We will sort this out later but, for now, we'll just try to see if we can make any progress with the tools already available.

Mathematically, what we are dealing with is *parametric curves*:



We are considering maps from a closed interval to the ring (annulus), $f : [0,1] \to A$, with the ends taken to the same point: $f(0) = f(1)$. Then we want to somehow count the number of turns.

At this point, the only available topological tool that seems to match the idea of a parametric curve is the 1-*cycles*. Let's try to classify the 1-cycles in the circle, or a ring, in terms of their homology and see if it solves our problem.

In the "hollow square" on the left, it seems clear that

- "the chain $q = a + b + c + d$ goes once around the hole".

Let's do this twice. The binary arithmetic then yields unexpected results:

- $p := q + q = (a + b + c + d) + (a + b + c + d) = 0$, even though it seems that "the chain p goes twice around the hole".

Exercise 3.1. Verify the following:

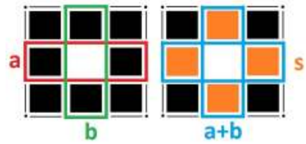
- if "the chain r goes thrice around the hole", then $r = p$;
- if "the chain s goes once in the opposite direction", then $s = p$; ...

Continue.

Thus, we have only one 1-cycle that's not homologous to 0.

Our conclusion is now clear: without a way to record the directions of the edges of a cycle, we can't speak of how many turns it makes!

Example 3.2. Let's make it clear that the problem isn't that we chose the cycles poorly and their edges got cancelled. In the "thick hollow square" below, we can choose these two 1-cycles with no overlaps:

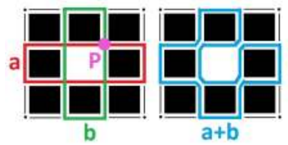


With the algebra we have learned, we can easily conclude the following about the homology of these cycles:

- $a \not\sim 0$ as a isn't a boundary, and
- $b \not\sim 0$ as b isn't a boundary, but
- $c = a + b \sim 0$ as $c = a + b$ is the boundary of the 2-chain s.

Again, homology theory, as we know it, is unable to classify the cycles according to the number of times they go around the hole.

To see this example from another angle, we can think of the curve $c = a+b$ as the curve a followed by the curve b. In that sense, the curve c circles the hole twice. But with directionless edges, we can also think of $c = a + b$ as a sequence of edges that form a curve that doesn't complete the full circle:



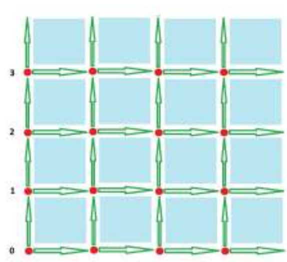
□

To address this problem, we need to learn to count *directed* edges – even at the expense of added complexity. We will rebuild the algebra of chains and boundaries for directed edges and, furthermore, "oriented" cells.

Note: Later we will solve the original problem by looking at the loops as maps and examining their homology maps. The number of turns is called the degree of a map.

3.2 The algebra of chains with coefficients

It appears that all we need to do in order to be able to count turns is to *assign a direction to each edge* of our cubical grid. Fortunately, they are already chosen for us; they are the directions of the axes:



Nonetheless, we should be able to choose any other combination of directions and our evaluation of the topology of a given cubical complex should remain the same.

No matter how it is done, this step – by design – has dramatic implications. We notice that

$$AB \neq BA!$$

In fact,

$$AB = -BA.$$

It follows that $AB + AB \neq 0$, which forces us to discard the cancellation rule we have replied on: $x + x = 0$. By doing so we are abandoning the binary arithmetic itself.

This time, chains are still "combinations of cells" but now these cells can, without cancelling, accumulate: $x + x = 2x$, etc. In fact, one can think informally of chain $2x$ as if cell x is visited twice. These are the interpretations of the new, directed chains:

- chain x visits cell x once going in the positive direction (that of the axis);
- chain $2x$ visits cell x twice;
- chain $3x$ visits cell x thrice;
- chain $-x$ visits cell x once going in the negative direction;
- chain $0 = 0x$ doesn't visit cell x (or any other);
- chain $x+2y+5z$ visits cell x once, cell y twice, and cell z five times – in no particular order.

Definition 3.3. A *k-chain* in a cubical complex K is a "formal linear combination of k-cells" in the complex:

$$S = r_1a_1 + r_2a_2 + ... + r_na_n,$$

where $a_1, a_2, ..., a_n$ are k-cells in K and the coefficients $r_1, r_2, ..., r_n$ are some scalars.

These "scalars" are chosen from some collection R. What kind of set is R? First, we need to be able to *add chains*,

$$\begin{aligned}(r_1a_1 + r_2a_2 + ... + r_na_n) &+ (s_1a_1 + s_2a_2 + ... + s_na_n)\\ &= (r_1 + s_1)a_1 + (r_2 + s_2)a_2 + ... + (r_n + s_n)a_n,\end{aligned}$$

and to *multiply chains by scalars*,

$$\begin{aligned}p(r_1a_1 + r_2a_2 + ... + r_na_n)&\\ &= (pr_1)a_1 + (pr_2)a_2 + ... + (pr_n)a_n.\end{aligned}$$

These requirements make us choose R to be a *ring*. Clearly, if we choose $R = \mathbf{Z}_2$, we are back to the binary theory, which remains a valid choice.

As a preview, the consequences of different choices of coefficients are outlined below.

Suppose K is a finite cubical complex.

Main choices for coefficients, $R =$	$\mathbf{Z}_2$ (binary)	$\mathbf{Z}$ (integral)	$\mathbf{R}$ (real)
Observations:	$\mathbf{Z}_2$ is a mere projection of $\mathbf{Z}$.	$\mathbf{Z}$ is the best for counting.	$\mathbf{R}$ contains $\mathbf{Z}$ but patches over finer details.
Algebraically, R is:	a finite field	an infinite cyclic group	an infinite field
Notation: add "; R"	$C_k(K;\mathbf{Z}_2), H_k(K;\mathbf{Z}_2)$	$C_k(K;\mathbf{Z}), H_k(K;\mathbf{Z})$	$C_k(K;\mathbf{R}), H_k(K;\mathbf{R})$
Algebraically, $C_k(K;R)$ is:	an abelian group with all elements of order 2, or a vector space over $\mathbf{Z}_2$ (also a list)	a free finitely-generated abelian group	a finite-dimensional vector space over $\mathbf{R}$
Decomposition of $C_k(K;R) =$	$\mathbf{Z}_2 \oplus ... \oplus \mathbf{Z}_2$	$\mathbf{Z} \oplus ... \oplus \mathbf{Z}$	$\mathbf{R} \oplus ... \oplus \mathbf{R}$
Studied in:	Modern algebra		Linear algebra

Similar choices for R:	other finite rings and fields: $\mathbf{Z}_p, p = 3, 4, ...$	*the* choice (the others can be derived from it)	other infinite fields: $\mathbf{C}, \mathbf{Q}$
Accessibility:	1. Chains are lists of cells. 2. All groups are lists of chains. 3. But the direction of a cell in a chain is meaningless.	1. The meaning of chains $2x, 3x, -x$ is clear. 2. The homology groups capture more features, such as twists, than the other two. 3. But the homology is harder to compute.	1. Vector spaces are more familiar than groups. 2. There is a connection to the familiar parts of calculus (next subsection). 3. But the meaning of $x/2$ is unclear.
Boundary operator (and homology maps) is:	same, mod 2, matrix as the other two	same integer-valued matrix	
	a homomorphism		a linear operator
Basis of the kth chain group is	the set of all k-cells of K		
Bases of cycle and boundary groups are	same, mod 2, chains as the other two	same integer-valued chains	
Algebraically, $H_k(K;R)$ is:	a finite abelian group with all elements of order 2 (a list)	a finitely-generated abelian group with a possible torsion part ($\mathbf{Z}_2$ for the Klein bottle)	a finite-dimensional vector space
Decomposition of $H_k(K;R) =$	$\mathbf{Z}_2 \oplus ... \oplus \mathbf{Z}_2$	$\mathbf{Z} \oplus ... \oplus \mathbf{Z} \oplus \mathbf{Z}_{p_1} \oplus ... \oplus \mathbf{Z}_{p_s}$	$\mathbf{R} \oplus ... \oplus \mathbf{R}$
Outcomes:	Same generators, same dimension/rank, same count of topological features: components, tunnels, voids, etc. (except for the torsion part)!		

Exercise 3.4. What are possible drawbacks of using $R = \mathbf{Z}_3$? $\mathbf{Z}_4$?

For simplicity and convenience, we choose to deal with chains and, therefore, homology with

real coefficients

for the rest of this chapter. We will omit "$;\mathbf{R}$" in our notation when there is no confusion: $H_k(K) = H_k(K;\mathbf{R})$.

3.3 The role of oriented chains in calculus

As we now understand, in every cubical complex there are two versions of the same edge present: positive and negative. It is positive when the direction of the edge coincides with the direction of the axis. So, for the same cell $[a, b] \subset \mathbf{R}$,

- if $a < b$, then the direction of $[a, b]$ is positive, and
- if $a > b$, then the direction of $[a, b]$ is negative.

In other words,

$$[b, a] = -[a, b].$$

In light of this approach, we can rewrite the *orientation of property of integral* from elementary calculus:

$$\int_a^b f(x)dx = -\int_b^a f(x)dx,$$

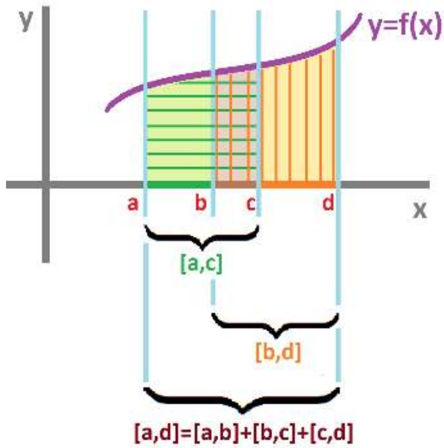
in terms of chains:

$$\int\limits_{-[a,b]} f(x)dx = -\int\limits_{[a,b]} f(x)dx,$$

We can also rewrite the *additivity property of integrals*:

$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx.$$

But first we observe that, if the intervals overlap, the property still makes sense, as the integral over the overlap is counted twice:



Then the property takes this form:

$$\int\limits_{[a,b]} f(x)dx + \int\limits_{[c,d]} f(x)dx = \int\limits_{[a,b]+[c,d]} f(x)dx.$$

Even the *scalar multiplication property* is given an additional meaning:

$$\int\limits_{r[a,b]} f(x)dx = r\int\limits_{[a,b]} f(x)dx.$$

The left-hand sides of these three identities can now be understood as a certain *function evaluated on these chains*, as follows. Suppose a cubical complex K represents interval $[A,B] \subset \mathbf{R}$. Then, for any given integrable function $f:[A,B]\to\mathbf{R}$, we can define a new function:

$$F : C_1(K) \to \mathbf{R}$$

by

$$F(s) := \int\limits_s f(x)dx, \quad \forall s \in C_1(K).$$

According to the above identities, this function is *linear*! Indeed, we've shown:

$$F(ru+tv) = rF(u) + tF(v), \quad r,t \in \mathbf{R}, u,v \in C_1(K).$$

Exercise 3.5. (a) Define a similar function for 0-chains. (b) Interpret the Fundamental Theorem of Calculus.

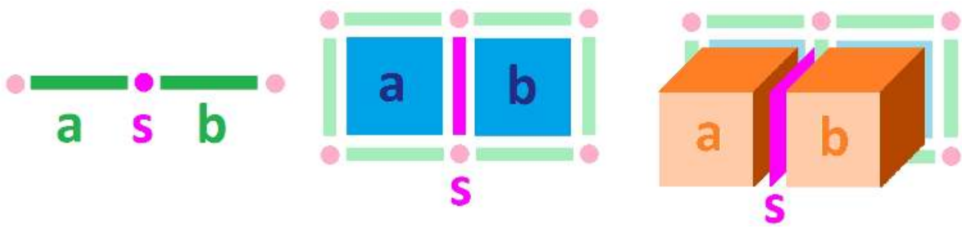
A linear function $F : C_1(K) \to \mathbf{R}$ is called 1-*cochain* on K. In the context of calculus, it is also called a *discrete differential form* of degree 1. This idea is the basis of the modern approach to calculus.

Conclusion: *oriented chains serve as domains of integration.*

3.4 Orientations and boundaries

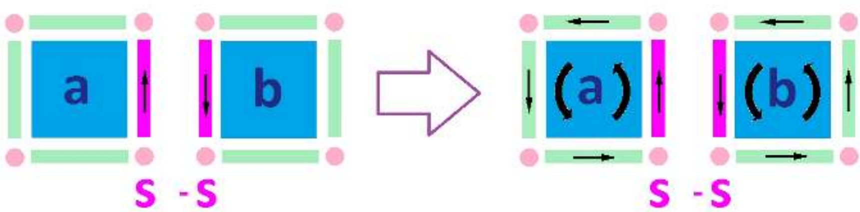
A crucial part of homology theory has been the cancellation rule of $R = \mathbf{Z}_2$. With this rule, all cells that appear twice in a given chain, such as the boundary of a cell, cancel. Without this rule, it will no longer automatically happen!

The idea is then to assume that – before any computations start – an "orientation" is assigned to each cell in K. Then each cell s may appear in computations with either positive, as s, or negative orientation, as $-s$. This should be done in such a way that, when computing the boundary of a chain $a + b$ with two adjacent cells a, b present, each "internal face" s (shared by a and b) will appear twice, but with the opposite orientations and, therefore, cancel: $s + (-s) = 0$.



But what *is* an orientation?

We already know the answer for 1-cells. The orientation means a choice of the direction of the edge and this is how they are supposed to cancel:



To have them cancel, let's choose the orientation for either of the 2-cells involved to mean the choice of the counterclockwise direction around it. Then clockwise direction will be the opposite orientation. Next, the orientations of the boundary 1-cells of this 2-cell are "read" by going around this cell. Then, indeed, there will be s coming from the boundary of the first cell and $-s$ from the boundary of the second.

What about 0-cells? The idea of an orientation of a *point* might seem strange. Instead, let's ask again, how do these cells cancel?

The boundary of a 1-cell AB is, as before, a linear combination of its two endpoints A and B, but what are the coefficients? What helps here is that we also expect the boundary of $AB+BC$ to be a linear combination of A and C, so B has to cancel:

$$\begin{array}{lll} ?A+?C & = \partial(AB+BC) & = \partial(AB)+\partial(BC) \\ & = (?A+?B)+(?B+?C) & =?A+(?B+?B)+?C. \end{array}$$

Therefore, B has to appear with the opposite signs! It might appear with "+" in $\partial(AB)$ and with "−" in $\partial(BC)$. But the only difference between them is the position of B with respect to the direction of these edges, the beginning or the end. We choose then the boundary of an edge to be equal to its final point minus the initial point:

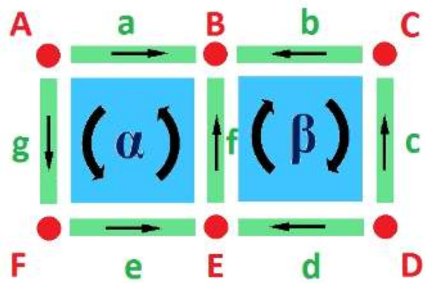
$$\partial(a)=\partial(AB):=B-A.$$

Then the orientation of a 0-cell works as expected even though it doesn't have a visible meaning. (It seems that the orientation means what the algebra says it means!)

No doubt, the boundaries of 0-chains are all, as before, equal to 0;

$$\partial(A):=0.$$

Example 3.6. Let's confirm that the definition works. We assigned random orientations to 1- and 2-cells:



Then,

$$\partial(a-b)=\partial a-\partial b=(A-B)-(B-C)=A-C,$$

and the interior vertex B is cancelled. And so it is for the chain $b-a$, but not for $a+b$.

Next,

$$\partial(\alpha-\beta)=\partial\alpha-\partial\beta=(-a+f+e+g)-(-b-c+d+f)=-a+b+c-d+e+g,$$

and the interior edge f is cancelled. And so it is for the chain $\beta-\alpha$, but not for $\alpha+\beta$. □

Exercise 3.7. Find three linearly independent 1-cycles in the last example and compute their boundaries.

We are now to reconstruct the homology theory for the new setting.

Proposition 3.8. For any cubical complex K (and any choice of orientations of its cells), the composition of boundary operators

$$\partial_1\partial_2 : C_2(K)\to C_0(K)$$

is zero.

Exercise 3.9. (a) Prove the proposition. (b) Show that the homology groups are well-defined.

Definition 3.10. The *homology groups* of a cubical complex K are the vector spaces defined by

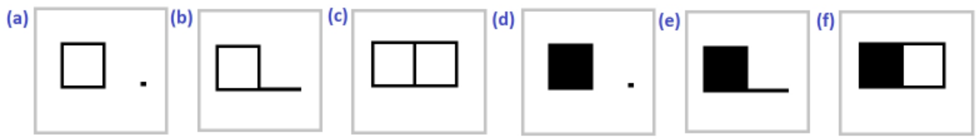
$$H_k(K) := \ker \partial_k / \operatorname{Im} \partial_{k+1}, \ k = 0, 1.$$

The *Betti numbers* are the dimensions of these vector spaces:

$$\beta_k(K) := \dim H_k(K), \ k = 0, 1.$$

Exercise 3.11. Represent the sets below as realizations of cubical complexes. For each of them:

- (a) find the chain groups and find the boundary operator as a matrix;
- (b) using only part (a) and linear algebra, find Z_k, B_k, H_k for all k, including the generators;
- (c) repeat the computations for two more choices of coefficients, $R = \mathbf{Z}, \mathbf{Z}_3$.



Exercise 3.12. Starting with the cubical complex K from the last exercise, add/remove cells so that the new complexes K' satisfy:

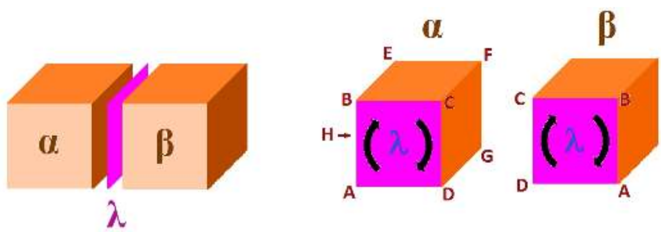
- $\beta_0(K') = 0, 1, 2$, or
- $\beta_1(K') = 1, 2, 3$.

Present the chain complexes and use linear algebra, such as

$$\dim \left(\mathbf{R}^n / \mathbf{R}^m \right) = n - m,$$

to prove the identities.

The 3-dimensional case is more complex. Even though we understand the meaning of the orientation of these 2-cells, what can we way about these two adjacent 3-cells α and β?



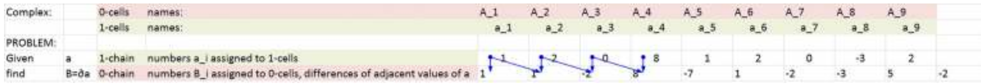
We need to define orientation for them in such a way that the 2-cell λ would appear with two opposite orientations – as a boundary face of α and β. How? Notice that understanding orientation as an *order of its vertices* works for 1-cells, say, $a = AB$, and for 2-cells, say, $\tau = ABCD$. So, maybe it's a good idea to try the same for the 3-cells, say, $\alpha = ABCDEFG$?

Exercise 3.13. Show that such a definition won't produce what we expect. Hint: does it really work for dimension 2?

Defining orientation for the case of higher dimensions will require using the product structure of the space.

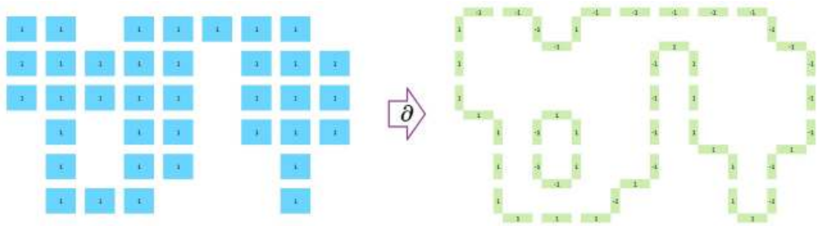
3.5 Computing boundaries with a spreadsheet, continued

This is how these simple formulas for dimension 1 are implemented with a spreadsheet program such as Excel:



Note: See the files online.

Example 3.14. As an illustration, this is a 2-chain (blue) and its boundary, a 1-chain (green), computed and presented in a spreadsheet:



The direction of an edge is given by 1 if it goes along the axis and -1 if it goes against it. Observe how the directions change as the boundary curve goes counterclockwise. $\square$

Exercise 3.15. What happens to the values of the boundary chain if we replace all 1's in the 2-chain on the left with 2's? What happens to the shape of the boundary if we make them random?

Just as in the binary case, the approach to the algorithm for computing the boundary operator is the opposite to that of the algebra we have seen. Instead of computing the boundary of each of the chain's k-cells as the sum of this cell's boundary cells and then simplifying, we just add the values (with appropriate signs) of the k-cells of the chain adjacent to each $(k-1)$-cell. As before, the code is *very simple*.

Dimension 0: For each vertex, we compute the sum of the values at the four adjacent edges with appropriate signs:

```
= -s!RC[-1] + s!RC[1] - s!R[-1]C + s!R[1]C
```

where s refers to the worksheet that contains the chain. Then,

- if the curve isn't there, we have $0+0+0+0=0$;
- if the curve is passing by once, we have $1+0-1+0=0$ (in some order);
- if the curve is passing by twice, we have $1-1+1-1=0$;
- if the curve begins or ends here, we have $1+0+0+0=1$ or -1.

Dimension 1: For each edge, we, similarly, compute the sum of the values at the two adjacent faces with appropriate signs, horizontal:

```
= s!R[-1]C - s!R[1]C
```

vertical:

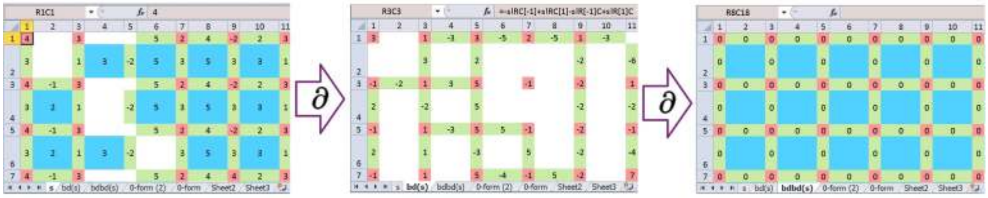
```
= -s!RC[-1] + s!RC[1]
```

Then,

- if the edge is away from the region, we have $0+0=0$;
- if the edge is inside the region, we have $1-1=0$;
- if the edge is on the boundary of the region, we have $1+0=1$.

Exercise 3.16. Modify the code to compute boundaries over $\mathbf{Z}_p$.

Example 3.17. We can compute boundaries of more complex chains:



We demonstrated here that the result of two boundary operators applied consecutively is, again, 0. □

Exercise 3.18. Define the group C of *all* chains (as opposed to the collection of all groups C_k of k-chains, $k = 0, 1, ...$). Define the boundary operator for C. Hint: consider the picture above.

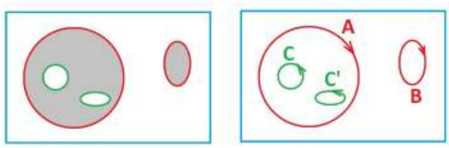
3.6 A homology algorithm for dimension 2

The image below has two path-components, the first one with two holes:

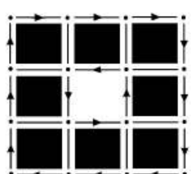
We will assume that *the binary images have black objects on white background.*

Exercise 3.19. How is the count affected if we take the opposite approach?

We will use 1-cycles, understood as circular sequences of edges, to partition the image. In particular, below, A and B are 0-cycles, C and C are 1-cycles:



The result is an unambiguous representation of the regions by the curves that enclose them. A 0-cycle is traversed clockwise and a 1-cycle is traversed counterclockwise.

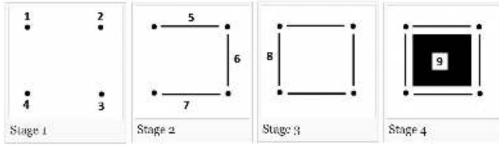


Observe that, in the either case, black is on the left.

The algorithm is incremental. The cycles are constructed as pixels, one by one, are added to the image. The result is an increasing sequence of complexes called a *filtration*:

$$K_1 \hookrightarrow K_2 \hookrightarrow ... \hookrightarrow K_m.$$

According to the way images are decomposed into cells, every pixel also contains edges and vertices; therefore, the process of adding a pixel starts with adding its vertices and then its edges, unless those are already present as parts of other pixels.



Exercise 3.20. Suggest another order of cells so that we still have a cubical complex at every stage.

Once all the edges have been added, there is always a 1-cycle inside the pixel. It is "removed" as the square closes the hole.

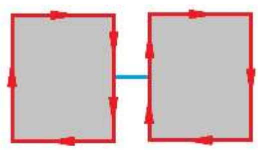
As the new pixels are being added, components merge, holes split, etc. In other words, the topology of the image changes as we watch cycles appear and disappear, merge and split.

Outline of the algorithm:

- All pixels in the image are ordered in such a way that all black pixels come before white ones.
- Following this order, each pixel is processed:
- • add its vertices, unless those are already present as parts of other pixels;
- • add its edges, unless those are already present as parts of other pixels;
- • add the face of the pixel.
- At every step, there are three possibilities:
- • adding a new vertex creates a new component;
- • adding a new edge may connect two components, or create, or split a hole;
- • adding the face to the hole eliminates the hole.

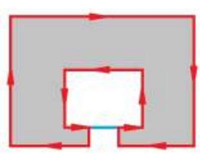
Only adding an edge presents any challenge. Thanks to the cell decomposition of the image, there are only 4 cases to consider.

Case (a): the new edge connects two different 0-cycles.



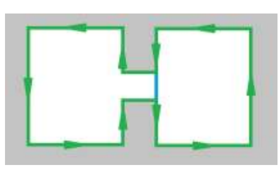
These two 0-cycles merge.

Case (b): the new edge connects a 0-cycle to itself.



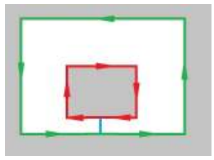
This 0-cycle gives birth to a new 1-cycle.

Case (c): the new edge connects a 1-cycle to itself.



This 1-cycle splits into two.

Case (d): the new edge connects a 1-cycle to a 0-cycle (inside).



This 0-cycle is absorbed into the 1-cycle.

Then, the new edge is associated with the new 0- of 1-cycle as described.

Exercise 3.21. Indicate how the Betti numbers change in these four cases.

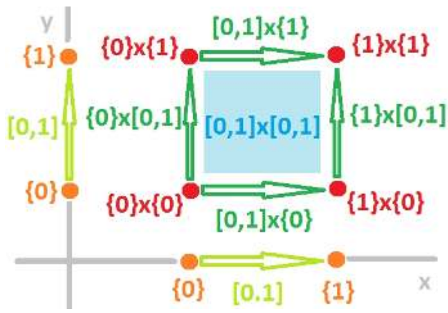
Exercise 3.22. Describe such an algorithm for the (a) triangular and (b) hexagonal grid.

3.7 The boundary of a cube in the N-dimensional space

In the presence of the product structure, the decision about orientations can be made for us and in a uniform fashion: all edges are directed along the coordinate axes. For 1-cells in the plane, the formulas are:

$$\begin{array}{ll}\partial\big(\{0\}\times[0,1]\big) & = -\{(0,0)\}+\{(0,1)\},\\ \partial\big([0,1]\times\{0\}\big) & = -\{(0,0)\}+\{(1,0)\}.\end{array}$$

For a 2-cell, its boundary should, as we know, be the four oriented edges appearing as we go counterclockwise around the square:



Then,

$$\begin{array}{l}\partial\big([0,1]\times[0,1]\big)\\ =[0,1]\times\{0\}+\{1\}\times[0,1]-[0,1]\times\{1\}-\{0\}\times[0,1].\end{array}$$

Here the edges are always oriented along the axes and going counterclockwise around the square produces these signs. To confirm that this makes sense, let's factor and rearrange the terms:

$$\begin{array}{lll} = & [0,1]\times\big(\{0\}-\{1\}\big) & +\big(\{1\}-\{0\}\big)\times[0,1]\\ = & -[0,1]\times\partial[0,1] & +\partial[0,1]\times[0,1].\end{array}$$

That's how products behave under the boundary operator.

Warning: A common mistake in elementary calculus is to assume that the derivative of the product is the product of the derivatives. Don't make a similar mistake and assume that the boundary of the product is the product of the boundaries! Consider:

$$\partial(a\times b)\neq\partial a\times\partial b.$$

Then, the left-hand side is the four *edges* of the square and the right-hand side is its four *vertices*. Even the dimensions/units don't match!

The correct formula of the *product rule for boundaries* is:

$$\partial(a \times b) = \partial a \times b + (-1)^{\dim a} a \times \partial b,$$

for cells and chains of all dimensions.

Exercise 3.23. Use to the formula to compute the boundary of a cube $\partial\big([0,1] \times [0,1] \times [0,1]\big)$ and confirm that the opposite faces appear with opposite orientations.

We will rely on the product structure to compute the boundary of the k-dimensional cube in the N-dimensional space.

Recall that a k-cube is the product of k edges and some vertices. In $\mathbb{R}^N$, a cubical k-cell can be represented as the product if the N-dimensional space is seen as the product of N copies of $\mathbf{R}$:

$$\begin{array}{ccccccccccccccc}
Q^k & =\prod_{i=1}^{N} A_i & = & A_1 & \times & ... & \times & A_{s-1} & \times & A_s & \times & A_{s+1} & \times & ... & \times & A_N \\
\cap & \cap & & \cap & & & & \cap & & \cap & & \cap & & & & \cap \\
\mathbf{R}^N & =\prod_{i=1}^{N} \mathbf{R}^1 & = & \mathbf{R} & \times & ... & \times & \mathbf{R} & \times & \mathbf{R} & \times & \mathbf{R} & \times & ... & \times & \mathbf{R}.
\end{array}$$

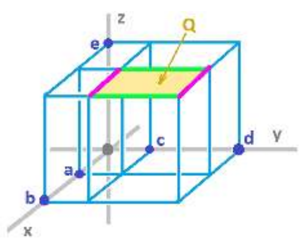
Here, each A_i is a subset of the ith axis, a copy of $\mathbf{R}$, which is either
- an edge, such as $[m, m+1]$, or
- a vertex, such as $\{m\}$.

There are exactly k edges on the list. Note that this representation is unique.

Example 3.24. Here's an example of a 2-cube, a square, in the 3-dimensional space:

$$Q^2 = [a,b] \times [c,d] \times \{e\},$$

illustrated below:



□

Even though the meaning of the boundary as shown above is clear, we proceed to give a general definition.

We already understand products of cubes. If
- A^p is an p-cube in $\mathbb{R}^n$,
- B^q is an q-cube in $\mathbb{R}^m$, then
- $A^q \times B^p$ is an $(p+q)$-cube in $\mathbb{R}^{n+m}$.

We also need to understand products of chains. If
- $a \in C_p(\mathbb{R}^n)$,
- $b \in C_q(\mathbb{R}^m)$, then
- $a \times b \in C_{p+q}(\mathbb{R}^{n+m})$.

Predictably, this product is defined via summation of the pairwise products of the cubes. If
- $a = \sum_i a_i P_i$, and
- $b = \sum_i b_i Q_i$, where a_i, b_i are the scalars and P_i, Q_i are the cells (finitely many!), then
- $a \times b := \sum_i a_i b_i (P_i \times Q_i)$.

Definition 3.25. The definition of the boundary is inductive.

As the base of induction, we define the boundary for 1-cells, the edges. Suppose

$$Q^1 := \prod_{i=1}^{N} A_i,$$

with
- $A_s := [m_s, m_s + 1]$ and
- $A_i := \{m_i\}$ for $i \neq s$,

for some particular choice of s. Its boundary is also a product:

$$\partial Q^1 = \prod_{i=1}^{N} B_i.$$

What are the values of these components? We define, as before, the boundary with a change in a single component:
- $B_s = \partial A_s = \{m_s + 1\} - \{m_s\}$ and
- $B_i = A_i = \{m_i\}$ for $i \neq s$.

We rephrase the formula:

$$\begin{aligned}&\partial\Big(\{m_1\} \times ... \times \{m_{s-1}\} \times [m_s, m_s + 1] \times \{m_{s+1}\} \times ... \times \{m_N\}\Big)\\ &= \{m_1\} \times ... \times \{m_{s-1}\} \times \big(\{m_s + 1\} - \{m_s\}\big) \times \{m_{s+1}\} \times ... \times \{m_N\}.\end{aligned}$$

This is, as expected, the endpoint

$$\{m_1\} \times ... \times \{m_{s-1}\} \times \{m_s + 1\} \times \{m_{s+1}\} \times ... \times \{m_N\}$$

minus the beginning point of the edge:

$$\{m_1\} \times ... \times \{m_{s-1}\} \times \{m_s\} \times \{m_{s+1}\} \times ... \times \{m_N\}.$$

Thus, we have defined the boundary for all edges in the N-dimensional space for all N. We also know the boundary of any vertex: $\partial V := 0$.

Next we assume that the boundary is defined for all k-cells – in the N-dimensional space for all N – and then extend the definition to $(k+1)$-cells, as follows. We simply use the fact that any $(k+1)$-cell Q^{k+1} is the product of a k-cell Q^k with an edge E:

$$Q^{k+1} = Q^k \times E.$$

More precisely, we deal with these cells in these Euclidean spaces:
- $Q^k \subset \mathbf{R}^n$ is a k-cube, and
- $A = E \subset \mathbf{R}^1$ is an edge, or
- $A = V \subset \mathbf{R}^1$ is a vertex, then
- $Q^k \times A \subset \mathbf{R}^n \times \mathbf{R}^1$ is a k- or a $(k+1)$-cube.

Finally, the *boundary of a cubical cell* $Q^{k+1} = Q^k \times A$ constructed this way is defined in terms of its components.
- Case 1: if A is an edge E, then

$$\partial(Q^k \times E) := \partial Q^k \times E + (-1)^k Q^k \times \partial E.$$

- Case 2: if A is a vertex V, then

$$\partial(Q^k \times V) := \partial Q^k \times V.$$

□

Plainly, both cases are just manifestations of the more general product rule (as $\partial V = 0$).

Using this definition, we can compute the boundary of any cubical cell:

$$A_1 \times ... \times A_{s-1} \times A_s \times A_{s+1} \times ...A_n$$

by adding one component A_s at a time and, depending on whether this is an edge or a vertex, and use one of the two formulas given above:

- A_1,
- $A_1 \times A_2$,
- $A_1 \times A_3 \times A_3$,
- ...
- $A_1 \times ... \times A_{s-1} \times A_s$,
- ...
- $A_1 \times ... \times A_{s-1} \times A_s \times A_{s+1} \times ... \times A_n$.

Example 3.26. Let's find the boundary of

$$Q^2 := [0,1] \times \{5\} \times [3,4].$$

Here

- $A_1 := [0,1]$,
- $A_2 := \{5\}$,
- $A_3 := [3,4]$.

Starting from the left. For ∂A_1:

$$\partial\big([0,1]\big) = \{1\} - \{0\}.$$

Next for $\partial(A_1 \times A_2)$. From the definition, as the second component is a vertex, it's case 2. We substitute the last equation:

$$\partial\big([0,1] \times \{5\}\big) = \partial[0,1] \times \{5\} = \big(\{1\} - \{0\}\big) \times \{5\}.$$

Now for $\partial(A_1 \times A_2 \times A_3)$. The last component is an edge, so it's case 1. We substitute the last equation:

$$\begin{array}{l}\partial\Big(([0,1] \times \{5\}) \times [3,4]\Big)\\ = \partial\big([0,1] \times \{5\}\big) \times [3,4] + (-1)^1\big([0,1] \times \{5\}\big) \times \partial[3,4]\\ = \big(\{1\} - \{0\}\big) \times \{5\} \times [3,4] - \big([0,1] \times \{5\}\big) \times \big(\{4\} - \{3\}\big).\end{array}$$

To check that the answer makes sense, verify these facts: it's in 3-space; its dimension is 1; there are 4 edges; all lie in plane $y = 5$, etc. □

Exercise 3.27. Illustrate this computation with a picture similar to the one in the beginning of the subsection. From that picture, compute the boundary of square Q and of the whole cube.

3.8 The boundary operator

As usual, we extend the boundary operator from the one defined for the cells only to one defined for all chains. We say that "we use linearity" and this is what we mean: as every k-chain is a formal linear combination of k-cells, the k-cells form the basis of the space of k-chains C_k. The idea is familiar from linear algebra:

- a linear operator is fully defined by its values on the basis elements; and
- any choice of the values on the basis determines a linear operator.

The latter is exactly what we have. Indeed, ∂_k is defined on the k-cells and now these values are combined to produce the values for all k-chains:

$$\partial_k\Big(\sum_i r_i p_i\Big) = \sum_i r_i \partial_k p_i,$$

where p_i are the k-cells (finitely many) and r_i are the real scalars.

With that, we have a sequence of vector spaces and linear operators:

$$... \quad \to \quad C_{k+1}(K) \xrightarrow{\partial_{k+1}} C_k(K) \xrightarrow{\partial_k} C_{k-1}(K) \xrightarrow{\partial_{k-1}} ... \xrightarrow{\partial_1} C_1(K) \xrightarrow{\partial_0} 0,$$

for any cubical complex K. Then, just as before, we can define and compute the *cycle groups and the boundary groups* of K.

But in order to continue to homology, we need to reestablish the fact that *all boundaries are cycles.*

From the definition of the boundary, we have the following:

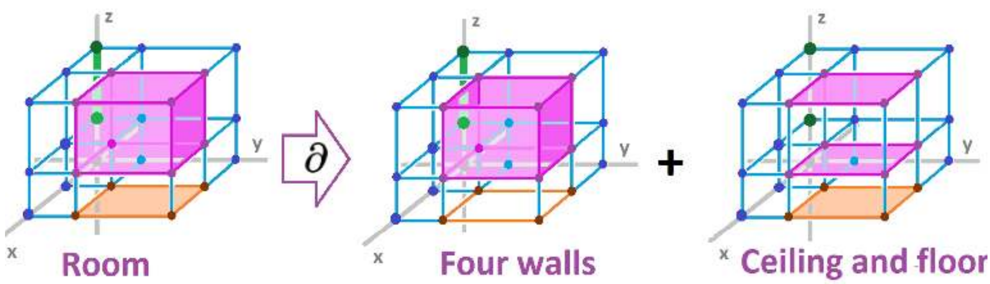
Proposition 3.28. Suppose:
- $Q^k \subset \mathbf{R}^n$ is a k-chain, and
- $A \subset \mathbf{R}^1$ is 0- or 1-chain, and
- $Q^k \times A \subset \mathbf{R}^n \times \mathbf{R}^1$ is a k- or a $(k+1)$-chain.

Then

$$\partial(Q^k \times A) = \partial Q^k \times A + (-1)^k Q^k \times \partial A.$$

Exercise 3.29. Prove the proposition.

The formula is illustrated below for $k = 2$. The cube as the product of a square and an edge is on the left and the two terms of the formula are shown on the right:



An even more general result is below:

Theorem 3.30 (Product Formula for Boundaries). Suppose
- $a \in C_i(\mathbb{R}^n)$,
- $b \in C_j(\mathbb{R}^m)$, and
- $a \times b \in C_{i+j}(\mathbb{R}^{n+m})$.

Then

$$\partial(a \times b) = \partial a \times b + (-1)^i a \times \partial b.$$

Theorem 3.31 (Double Boundary Identity). For oriented chains of cubical complexes,

$$\partial\partial = 0.$$

Proof. The proof is based on the same induction as the definition itself.

The *inductive assumption* is that in any N-dimensional space

$$\partial\partial Q^i = 0$$

for all cubical i-cells, for all N, for all $k \leq N$, and all $i = 0, 1, ..., k$.

Now we represent a $(k+1)$-cube as the product of a k-cube and an edge as the last component:

$$Q^{k+1} = Q^k \times E.$$

Now we just compute what we want. From the definition:

$$\partial Q^{k+1} = \partial(Q^k \times E) = \partial Q^k \times E + (-1)^k Q^k \times \partial E.$$

Next,

$$\begin{aligned}
\partial\partial Q^{k+1} &= \partial\Big(\partial Q^k \times E + (-1)^k Q^k \times \partial E\Big) && \text{and by linearity of } \partial... \\
&= \partial\Big(\partial Q^k \times E\Big) + (-1)^k \partial\Big(Q^k \times \partial E\Big) && \text{now use the last proposition, twice...} \\
&= \partial\partial Q^k \times E + (-1)^{k-1}\partial Q^k \times \partial E \\
&\quad + (-1)^k\Big(\partial Q^k \times \partial E + (-1)^k Q^k \times \partial\partial E\Big) && \text{now use the assumption...} \\
&= 0 \times E + (-1)^{k-1}\partial Q^k \times \partial E \\
&\quad + (-1)^k\Big(\partial Q^k \times \partial E + (-1)^k Q^k \times 0\Big) \\
&= (-1)^{k-1}\partial Q^k \times \partial E + (-1)^k \partial Q^k \times \partial E \\
&= 0.
\end{aligned}$$

■

Is it a coincidence that this computation feels like differentiation in elementary calculus?

Exercise 3.32. Supply the missing part of the above proof. Hint: the last component of the cube may be a vertex.

Exercise 3.33. Derive from the theorem the same formula for the binary arithmetic.

The meaning of the theorem is that the sequence above is a chain complex. Then, just as before, we can define and compute the *homology groups* of the cubical complex.

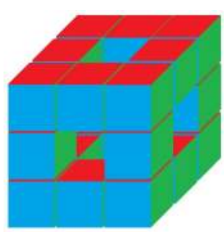
Definition 3.34. The *homology groups* of a cubical complex K are the vector spaces defined by

$$H_k(K) := \ker \partial_k / \operatorname{Im} \partial_{k+1}, \quad k = 0, 1, ...$$

The *Betti numbers* are the dimensions of the homology groups:

$$\beta_k(K) := \dim H_k(K), \ \ k = 0, 1, ...$$

Exercise 3.35. Compute the homology of the $3 \times 3 \times 3$ cube with 7 cubes removed:



Just as before, the *reduced homology groups* $\tilde{H}_k(K)$ of complex K are defined via the same formulas except for dimension 0:

$$\tilde{H}_k(K) := \begin{cases} \ker \partial_k / \operatorname{Im} \partial_{k+1}, & \text{for } k = 1, 2, ..., \\ \ker \varepsilon / \operatorname{Im} \partial_1 & \text{for } k = 0, \end{cases}$$

with ε given by

$$\varepsilon(r_1c_1 + ... + r_nc_n) := r_1 + ... + r_n, \ r_i \in \mathbf{Z}.$$

The reduced homology groups of the point are all trivial and we can even write them as one:

$$\tilde{H}(\{p\}) = 0.$$

Exercise 3.36. Write the chain complex for the reduced homology and express $\tilde{H}_k(K)$ in terms of $H_k(K)$.

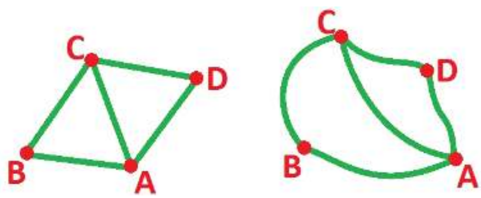
---

4 Simplicial complexes

4.1 From graphs to multi-graphs

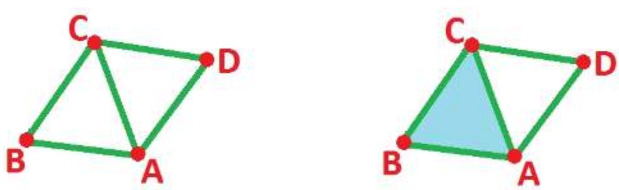
A graph is pure data. It consists of two sets:
- the nodes, say, $N = \{A, B, C, D\}$, representing some agents, and
- the edges, say, $E = \{AB, BC, CA, DA, CD\}$, representing some pairwise relation between them.

The topology is hidden in the data and, in order to see it, we often have to illustrate the data by a subset of the Euclidean space, as follows. Each node is plotted as a distinct point, but otherwise arbitrarily, and these points are connected by simple curves (paths) that can have only the nodes in common:



This set is called a *realization* of the graph, which is no more than an illustration of the data. Unfortunately, illustrations are only revealing when the data they represent is small in size...

Now, with a graph on the left, what could be the meaning of the *data* behind the picture on the right?



There must be some new kind of relation present! This three-way relation exists for A, B, C but not for A, C, D.

We now have a collection K of three sets:
- the nodes $N = \{A, B, C, D\}$ representing some agents,
- the edges $E = \{AB, BC, CA, DA, CD\}$ representing some pairwise relations, and

• the triangular face(s) $F = \{ABC\}$ representing some three-way relations.

This data set is called a *simplicial complex* (or sometimes even a "multi-graph"). Its elements are called 0-, 1-, and 2-simplices.

Example 4.1 (graph). The graph G above may come from a series of phone conversation between the four individuals:

	A	B	C	D
$A:$	$-$	$+$	$+$	$+$
$B:$	$+$	$-$	$+$	$-$
$C:$	$+$	$+$	$-$	$+$
$D:$	$+$	$-$	$+$	$-.$

Here, the fact that a conversation has occurred, it is marked with "+" and visualized with an edge. □

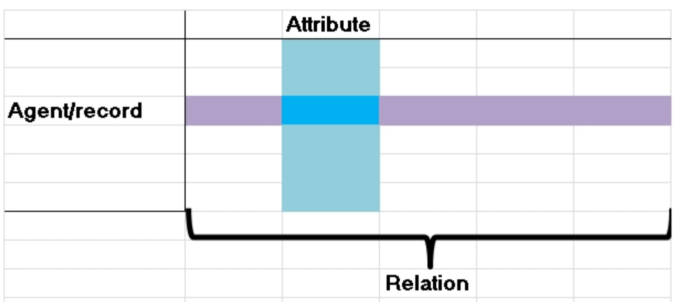
Example 4.2 (simplicial complex). Meanwhile, the simplicial complex K above may come from, say, the list of stores visited by these individuals:

	K	M	J	W
$A:$	$+$	$+$	$+$	$-$
$B:$	$+$	$-$	$-$	$-$
$C:$	$+$	$+$	$-$	$+$
$D:$	$-$	$-$	$+$	$+.$

Here, "K" may stand for Kroger (the triangle), "M" for Macy's, "J" for JCPenney, and "W" for Walmart (the edges). □

Exercise 4.3. Create a similar table for yourself and four of your best friends documenting your joint activities. Create a simplicial complex.

Example 4.4 (database). A similar construction is possible for any "relational database", which is simply a table:



For a single column table, every collection of n agents that happen to have an identical attribute form an element of our simplicial complex, an $(n-1)$-simplex. □

Based on these examples, it seems that a simplicial complex is nothing but *a collection of subsets of some finite set*. However, anticipating the need for defining the boundaries of simplices, we want all the boundary cells of each simplex to be present as well. These cells are called *faces* of the simplex. For example, the 1-faces of ABC are AB, BC, CA and the 0-faces are A, B, C. This is the notation we will use:

• for $\tau, \sigma \in K$, we write $\sigma < \tau$ if σ is a face of τ.

But this means simply that σ *is a subset of* τ:

$$\sigma < \tau \Longleftrightarrow \sigma \subset \tau.$$

An examination of this definition reveals the following properties:

• if $\tau \in K$ and $\sigma < \tau$ then $\sigma \in K$;

• if $\tau, \sigma \in K$ then $\tau \cap \sigma < \tau$.

These properties are reflected in a realization of this simplicial complex:

• the complex contains all faces of each simplex;
• two simplices can only share a single face.

Not every list of subsets of S would work then. Consider $\{AB, BC\}$ for example. This can't be a simplicial complex because the 0-faces of these 1-simplices are absent. But, once we add those to the list, the first condition of simplicial complex above is met and the second too, automatically. The list becomes:

$$\{AB, BC, A, B, C\}.$$

So, as it turns out, the first condition is the only one we need to verify.

Definition 4.5. A collection K of subsets of a finite set S is called an *abstract simplicial complex* if all subsets of any element of K are also elements of K; i.e.,

$$\text{if } \tau \in K \text{ and } \sigma < \tau \text{ then } \sigma \in K.$$

A subset with exactly $n+1$ elements, $n = 0, 1, ...$, is called an *n-simplex*.

Exercise 4.6. (a) Demonstrate that a relational database with a single attribute produces a simplicial complex as described above. (b) Give an example how simplices (and a simplicial complex) can be formed without requiring the attributes to be identical. (c) How is a complex formed if there are more than one attribute?

4.2 Simplices in the Euclidean space

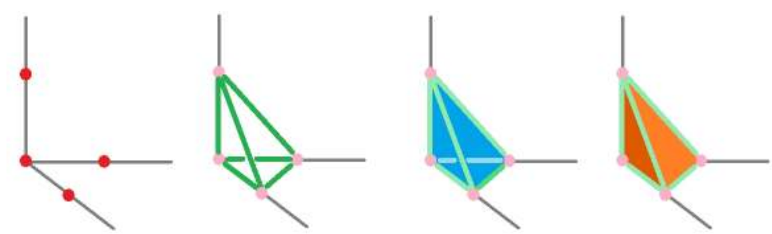
Next, we consider realizations of simplicial complexes as a way of giving them a topology. We start with simplices.

The most direct way to represent an n-simplex is to start in $\mathbf{R}^n$ with $n+1$ vertices which are located at the endpoints of the basis vectors of the n-dimensional Euclidean space and zero:

$$\begin{array}{c}(0,0,0,0,...,0,0,0),\\(1,0,0,0,...,0,0,0),\\(0,1,0,0,...,0,0,0),\\...\\(0,0,0,0,...,0,1,0),\\(0,0,0,0,...,0,0,1).\end{array}$$

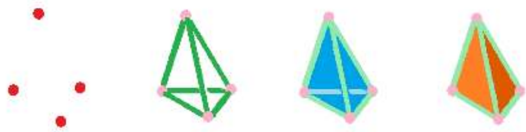
These vertices are then connected by edges, faces, etc.

A 3-simplex built this way is illustrated below:



Exercise 4.7. Represent this set as a set of inequalities.

Alternatively, we can use the first $n+1$ basis vectors of the N-dimensional Euclidean space with $N > n$. Suppose such a space is given. Now, we will build a similar figure starting with an arbitrary collection of points:



Given a set of $n+1$ points $\{A_0, A_1, ..., A_n\}$, a *convex combination* of these points is any point x given by

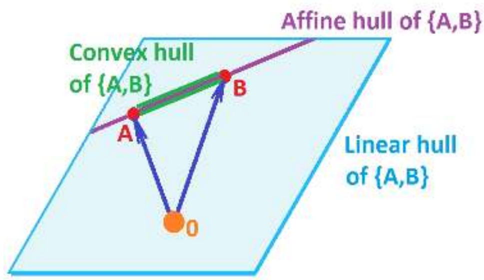
$$x = \sum_{i=0}^{n} \lambda_i A_i,$$

with the real coefficients that satisfy:
- 1. $0 \leq \lambda_i \leq 1, \ \forall i$, and
- 2. $\sum_i \lambda_i = 1$.

Then the *convex hull* is the set of all convex combinations.

These two conditions are important. To illustrate the idea, let's compare the convex hull of two points in $\mathbf{R}^3$ to
- the linear hull (aka the span) – no constraints on the coefficients, and
- the affine hull – only the second constraint present.

They are shown below:



The convex hull of these points is **denoted** by

$$\text{conv}\{A_0, A_1, ..., A_n\},$$

and the convex hull of any set Q is the set of all of its convex combinations; in fact, we have:

$$\text{conv}(A) := \bigcup \{\text{conv}(P) : \ P \subset Q, \ P \text{ finite}\}.$$

Recall that a set Q is *convex* if it contains all of its pairwise convex combinations:

$$\lambda x + (1-\lambda)y \in Q, \ \forall x, y \in Q, \ 0 \leq \lambda \leq 1.$$

It follows that for any convex set Q, we have

$$\text{conv}(Q) = Q.$$

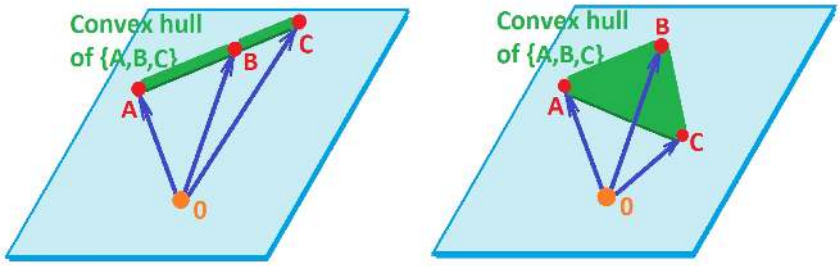
Convex sets are important in topology because they are "acyclic": they are path-connected, have no holes, voids, etc.

That's why it is a good idea to choose the building blocks to be convex, just as the ones of cubical complexes. But we also want to control the *dimensions* of these blocks!

This is the reason why we need a condition that prevents a "degenerate" situation when, for example, the convex hull of three points isn't a triangle, such as this:

$$\text{conv}\{(0,0), (1,0), (2,0)\}.$$

Below the issue is illustrated in $\mathbf{R}^3$:



We require these $n+1$ points $\{A_0, A_1, ..., A_n\}$ to be in *general position*, which means:

vectors $A_1 - A_0, ..., A_n - A_0$ are linearly independent.

One may say that this is a *generic* arrangement in the sense that if you throw three points on the plane, the probability that they line up is zero. They are also called "geometrically independent".

Proposition 4.8. The affine hull

$$\left\{ \sum_{i=0}^{n} r_i A_i : \sum_i r_i = 1 \right\}$$

of $n+1$ points in general position is an n-dimensional affine subspace (i.e., $M = v_0 + L$, where L is a linear subspace and v_0 is some vector).

Definition 4.9. A *geometric n-simplex* is defined as the convex hull of $n+1$ points $A_0, A_1, ..., A_n$ in general position:

$$s := A_0 A_1 ... A_n := \operatorname{conv}\{A_0, A_1, ..., A_n\}.$$

Exercise 4.10. Prove that the order of vertices doesn't matter.

Suppose s is an n-simplex:

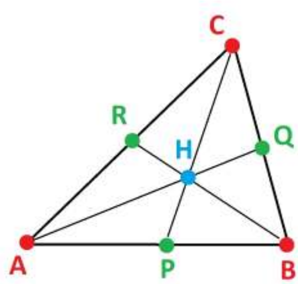
$$s = A_0 ... A_n.$$

An arbitrary point in s is a convex combination of the vertices:

$$x = \sum_i r_i A_i \text{ with } \sum_i r_i = 1, \ r_i \geq 0, \ \forall i.$$

These coefficients $r_0, ..., r_n$ are called the *barycentric coordinates* of x.

Example 4.11. An example of dimension 2 is below:



Here:

- $P = \frac{1}{2}A + \frac{1}{2}B, Q = \frac{1}{2}B + \frac{1}{2}C, R = \frac{1}{2}C + \frac{1}{2}A,$
- $H = \frac{1}{3}A + \frac{1}{3}B + \frac{1}{3}C.$ □

Exercise 4.12. Prove that the barycentric coordinates are unique for any given point.

Then it makes sense to write:

$$A = (1, 0, 0), \quad B = (0, 1, 0), \quad C = (0, 0, 1);$$

$$P = \left(\frac{1}{2}, \frac{1}{2}, 0\right), \quad Q = \left(0, \frac{1}{2}, \frac{1}{2}\right), \quad R = \left(\frac{1}{2}, 0, \frac{1}{2}\right);$$

$$H = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

Exercise 4.13. Show that the point with barycentric coordinates $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ is the center of mass of the triangle.

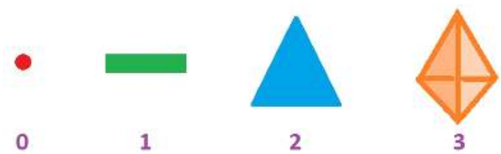
The point with equal barycentric coordinates is called the *barycenter* of the simplex.

Theorem 4.14. The n-simplex is homeomorphic to the n-ball $\mathbf{B}^n$.

Exercise 4.15. Prove the theorem.

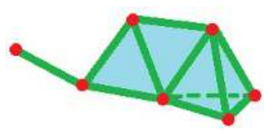
4.3 Realizations

The realizations of simplices up to dimension 3 are simple topological spaces:



Just as realizations of cubical complexes, a realization of a simplicial complex K is made of cells, but instead of

- edges, squares, cubes, ..., n-cubes; they are
- edges, triangles, tetrahedra, ..., n-simplices:



Definition 4.16. A *geometric simplicial complex* is a finite collection of points in space along with some of the geometric simplices defined by them. We will refer by the same name to the union of these simplices. Topological spaces homeomorphic to geometric simplicial complexes are called *polyhedra*.

A metric complex acquires its topology from the ambient Euclidean space. But how do we study its homology?

There is an additional structure here; a simplex has *faces*.

Example 4.17 (dimension 1). Suppose a is a 1-simplex,

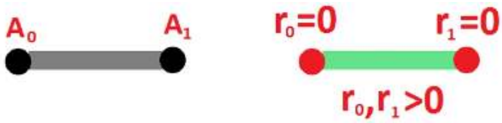
$$a = A_0 A_1.$$

Then its faces are the vertices A_0 and A_1. They can be easily described algebraically. An arbitrary point in a is a convex combination of A_0 and A_1:

$$x = r_0 A_0 + r_1 A_1 \text{ with } r_0 + r_1 = 1, \ r_0, r_1 \geq 0.$$

What about A_0 and A_1? They are convex combinations too but of a special kind:

$$\begin{aligned} A_0 &= r_0A_0 + r_1A_1 \quad \text{with } r_0 = 1, r_1 = 0, \\ A_1 &= r_0A_0 + r_1A_1 \quad \text{with } r_0 = 0, r_1 = 1. \end{aligned}$$



□

In other words,

every face has one of the barycentric coordinates equal to zero.

Example 4.18 (dimension 2). Suppose τ is a 2-simplex,

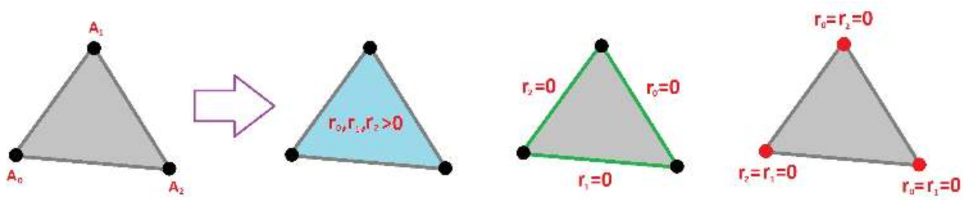
$$\tau = A_0A_1A_2.$$

An arbitrary point in τ is a convex combination of A_0, A_1, A_2:

$$x = r_0A_0 + r_1A_1 + r_2A_2 \text{ with } r_0 + r_1 + r_2 = 1, \; r_i \geq 0.$$

In order to find all 1-faces, set *one* of these coefficients equal to 0; the brackets indicate which:

$$\begin{aligned} f_2 &:= A_0A_1[A_2] &= \{r_0A_0 + r_1A_1 + r_2A_2 : r_0 + r_1 + r_2 = 1, r_2 = 0\}, \\ f_1 &:= A_0[A_1]A_2 &= \{r_0A_0 + r_1A_1 + r_2A_2 : r_0 + r_1 + r_2 = 1, r_1 = 0\}, \\ f_0 &:= [A_0]A_1A_2 &= \{r_0A_0 + r_1A_1 + r_2A_2 : r_0 + r_1 + r_2 = 1, r_0 = 0\}. \end{aligned}$$



So,

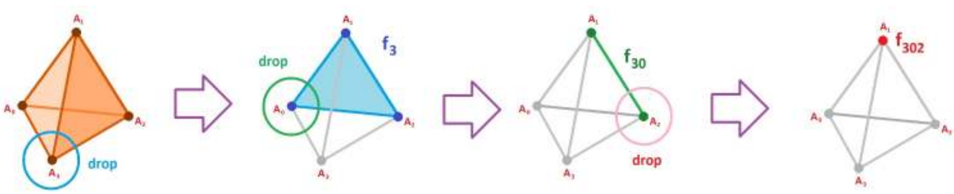
$$f_0, f_1, f_2 < \tau.$$

In order to find all 0-faces, set *two* of these coefficients equal to 0:

$$A_0 = f_{12} = A_0[A_1][A_2] = 1{\cdot}A_0 + 0{\cdot}A_1 + 0{\cdot}A_2, \text{ etc.}$$

□

Similar ideas apply to higher dimensions as we drop vertices one by one from our convex combinations.



Notation: For each $k = 0, 1, ..., n$, we denote by f_k the kth face of s which is an $(n-1)$-simplex acquired by setting the kth barycentric coordinate equal to 0:

$$f_k := A_0...[A_k]...A_n = \left\{ \sum_i r_iA_i : \sum_i r_i = 1, r_i \geq 0, r_k = 0 \right\}.$$

This way, the kth vertex is removed from all convex combinations. Further, for each pair $i, j = 0, 1, ..., n,\ i \neq j$, let $f_{ij} = f_{ji}$ be the $(n-2)$-simplex acquired by setting the ith and the jth barycentric coordinates equal to 0:

$$f_{ij} := A_0 ... [A_i] ... [A_j] ... A_n.$$

Here, the ith and the jth vertices are removed from all convex combinations.

Exercise 4.19. Show that f_{ij} is a face of f_i and a face of f_j.

We can continue this process and drop more and more vertices from consideration. The result is faces of lower and lower dimensions.

Definition 4.20. Given a geometric n-simplex τ, the convex hull f of any $m+1, m < n$, vertices of τ is called an *m-face* of τ.

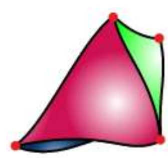
Exercise 4.21. How many m-faces does an n-simplex have?

Definition 4.22. The *boundary* of a geometric n-simplex is the union of all of its $(n-1)$-faces.

This union can be seen as either:
- the formal binary sum, or
- a linear combination with real coefficients.

In that latter case, it is a combination of all of the $(n-1)$-faces of the simplex in the original abstract simplicial complex. Then the boundary operator can be defined and the rest of the homology theory can be developed. Of course, we would rather carry out this construction with the abstract simplicial complex itself.

Meanwhile, the topological issues, as opposed to the geometrical issues, of realizations of simplicial complexes will be discussed later under *cell complexes*:



4.4 Refining simplicial complexes

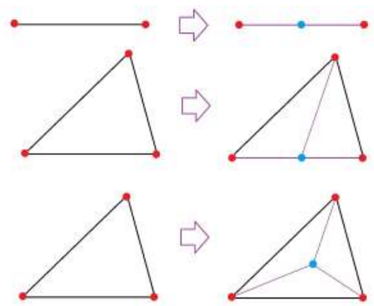
Definition 4.23. The *Euler characteristic* $\chi(K)$ of an n-dimensional simplicial complex K is defined as the alternating sum of the number of simplices in K of each dimension:

$$\chi(K) = \#0\text{-simplices } - \#1\text{-simplices } + \#2\text{-simplices } - ... \pm \#n\text{-simplices}.$$

We will prove later that the Euler characteristic is a topological invariant; i.e., if complexes K and M have homeomorphic realizations, $|K| \approx |M|$, then their Euler characteristics coincide, $\chi(K) = \chi(M)$. The converse of this theorem isn't true, which means that the Euler characteristic is not a "complete topological invariant".

Exercise 4.24. Find examples of non-homeomorphic complexes (of dimensions $1, 2, 3$) with the same Euler characteristic.

We will provide evidence in support of this theorem:



This is what "elementary subdivisions" of a 2-dimensional complex do to its Euler characteristic:

- Adding a vertex, $+1$, in the middle of an edge will split the edge and, therefore, increase the number of edges by one, -1.
- When this edge is in the boundary of a face, an extra edge has to be added, -1, but the face is also split in two, $+1$.
- Adding a vertex in the middle of a face, $+1$, will require 3 more edges, -3, and the face split into three, -2.

In all three cases the net effect on the Euler characteristic is nil.

Exercise 4.25. Provide a similar analysis for a 3-simplex.

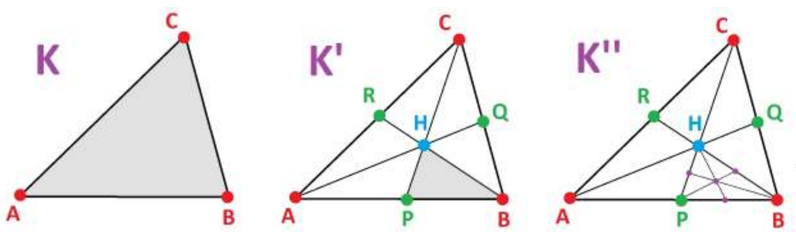
Exercise 4.26. What is the analog of barycentric subdivision for cubical complexes? Provide a similar analysis for 2-dimensional cubical complexes.

There is a standard method for refining geometric simplicial complexes of all dimensions. First, we cut every n-simplex into a collection of n-simplices, as follows. The process is inductive, for $k = 0, 1, 2, ...$. For each k-simplex $\tau \in K$,

- we remove τ and all of its boundary simplexes (except for the vertices) from K;
- we add a vertex V_τ to K, and then
- we create a new simplex from V_τ and each boundary simplex a of τ.

The result is a new collection K' of simplices.

Example 4.27 (dimension 2). This is what it looks like in dimension 2:



Here:

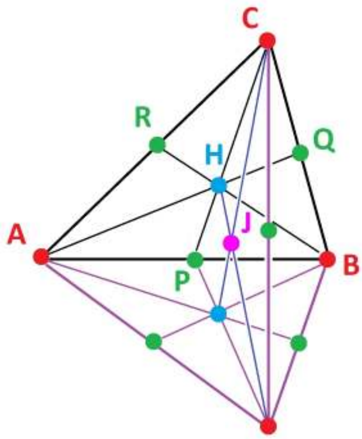
- for $\tau = ABC$ we choose $V_\tau = H$;
- for $\tau = AB$ we choose $V_\tau = P$;
- for $\tau = BC$ we choose $V_\tau = Q$;
- for $\tau = AC$ we choose $V_\tau = R$.

We also add the new edges: AP, PB, PH, etc., and new faces: AHP, PHB, etc. □

Exercise 4.28. Prove that the new collection is a simplicial complex.

A specific version of this operation is called the *barycentric subdivision*: the new vertex V_τ chosen for each cell τ is its barycenter – the point with equal barycentric coordinates – of the cell.

Example 4.29 (dimension 3). The structure of the subdivision of a 3-simplex is more complex:



For the 3-simplex above, we:

- keep the 4 original vertices,
- add 6 new vertices, one on each edge,
- add 4 new vertices, one on each face, and
- add 1 new vertex inside the simplex.

Then one adds *many* new edges and faces. □

Exercise 4.30. Given an abstract simplicial complex K, define a *discrete* version of barycentric subdivision. Hint: the vertices of the new complex are the simplices of K.

The following important result suggested by this analysis is to be proven later.

Theorem 4.31 (Invariance of Euler characteristic). The Euler characteristic is preserved under barycentric subdivision; i.e., if K' is the simplicial complex that results from the barycentric subdivision of simplicial complex K, then

$$\chi(K') = \chi(K).$$

Exercise 4.32. Prove the theorem (a) for a 3-dimensional simplicial complex and (b) similarly for a 3-dimensional cubical complex.

4.5 The simplicial complex of a partially ordered set

Suppose P is a finite partially ordered set (poset). We now would like to define a simplicial complex $\Delta(P)$ in such a way that the (Euclidean) topology of its realization would well correspond to the order topology on P. We will build complex $\Delta(P)$ *on* this set: its elements become the vertices of $\Delta(P)$ while some of the higher-dimensional cells are also added to mimic the topology of P.

The topology is meant to reflect the *proximity* of the elements of P and this proximity can't rely on any distance. Instead we define the topology of $\Delta(P)$ in terms of the order relation on P:

two element are "close" if and only if they are related.

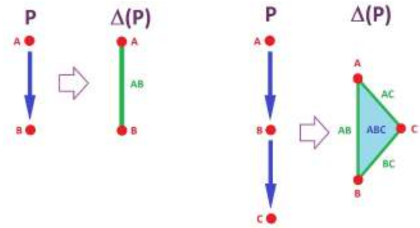
The later means $A < B$ or $A > B$.

Then, in order to encode this fact in $\Delta(P)$, we include an edge between these elements:

$$A < B \Longrightarrow AB \in \Delta(P).$$

Notice that it doesn't matter how many steps it takes to get from A to B.

Example 4.33. With just two, related, elements present, there is just one edge:



For the second example, we have $P = \{A, B, C : \ A < B < C\}$, and all the pairs AB, BC, AC are present in the complex.

But the resulting complex has a hole! Such a mismatch of the two topologies is what we want to avoid – and we add the 2-simplex ABC to $\Delta(P)$. □

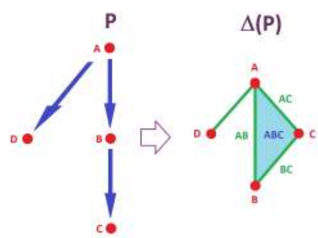
What does this last simplex have to do with the 1-simplices we have added so far? All of its vertices are related, pairwise. But this is only possible when they are linearly ordered.

Definition 4.34. We define the *order complex* $\Delta(P)$ of the ordered set P as the simplicial complex on P with the "monotone sequences" of P as its simplices:

$$\Delta(P) := \{A_1 A_2 ... A_n : \ A_i \in P, \ a_1 < a_2 < ... < a_n\},$$

Exercise 4.35. Prove that $\Delta(P)$ is a simplicial complex.

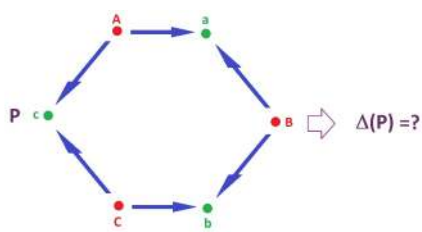
Example 4.36.



□

Is it possible to have $\Delta(P)$ with a non-trivial topology?

Exercise 4.37. Find the order complex for the following poset:



Exercise 4.38. Indicate what happens to the intervals of the posets in the above examples.

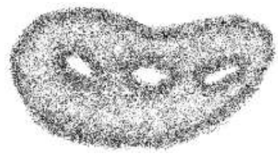
Conversely, the *face poset* $P(K)$ of a simplicial complex K is the poset of nonempty simplices of K ordered by inclusion.

Exercise 4.39. Find the $\Delta(P(S))$, where S is a 2-simplex.

Exercise 4.40. Describe the order topology of P in terms of the simplices of $\Delta(P)$.

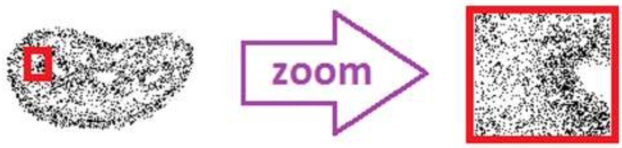
4.6 Data as a point cloud

Consider this object:



It appears to be a triple torus.

But what if we zoom in?



We realize that the "object" is just a loose collection of points suspended in space! It is called a *point cloud*.

Where do point clouds come from? They may come from scanning, as well as radar, sonar, etc. But they also come from *data*.

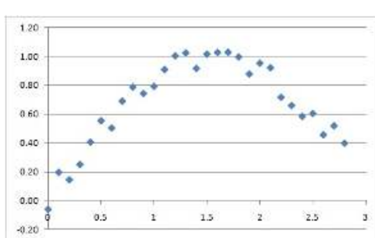
If a scientist has conducted 1000 experiments each of which consists of a sequence of 100 specific measurements, the results are combined into a collection of 1000 disparate points in $\mathbf{R}^{100}$. Then the scientist may be looking for a pattern in this data.

It is impossible to visualize higher dimensional data because any representation is limited to dimension 3 (by using colors one gets 6, time 7). In search of a pattern, we might still ask the same topological questions:

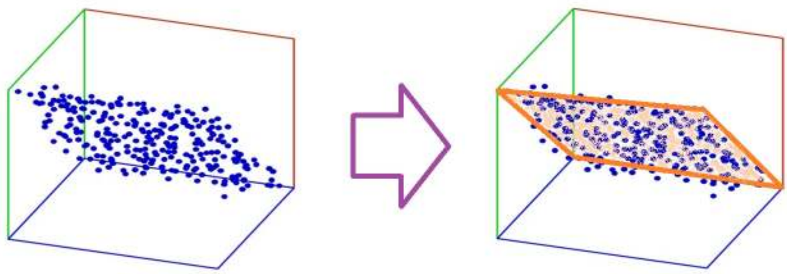
- Is it one piece or more?
- Is there a tunnel?
- Or a void?
- And what about possible 100-dimensional features?

"Clustering" answers the first question. The rest need the homology groups of the complex but where is it?

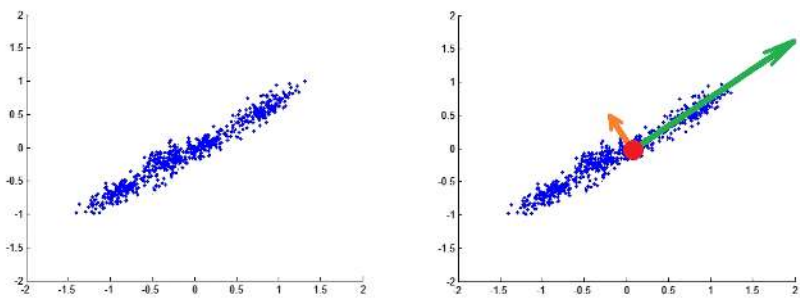
This data may hide a "manifold" behind it. It may be a *curve* (dimension 1):



or a *surface* (dimension 2):



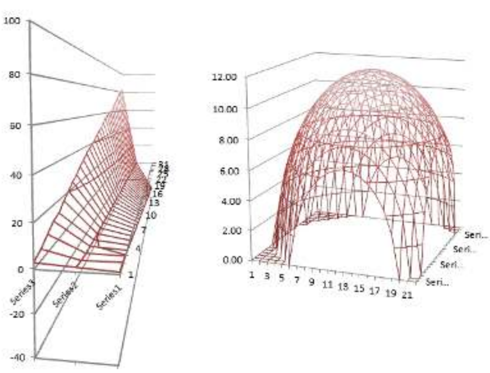
There are some well-developed approaches to this problem. One is a mathematical procedure (called the principal component analysis) for finding a linear transformation of the Euclidean space so that the features of the shape of the point cloud are aligned with the new coordinate axes. Such analysis will reveal the interdependence in data:



If the spread along an axis is less than some threshold, this axis may be ignored, which leads to "dimensionality reduction". Unfortunately, the linearity assumption of the method may cause it to fail when the manifold is highly curved.

The topological approach is to create a simplicial complex that "approximates" the topological space behind the point cloud.

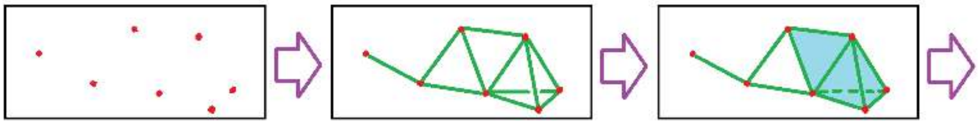
When the data isn't too noisy, a common spreadsheet software builds a complex by a simple interpolation:



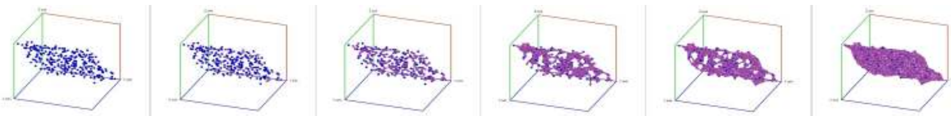
However, the software has to rely on the assumption that this is a surface!

What if the data is noisy? The method is as follows. We pick a threshold $r > 0$ so that any two points within r from each other are to be considered "close". Then we

- add an edge between any two points if they are within r from each other,
- add a face spanning three points if the "diameter" of the triangle is less than r, etc.



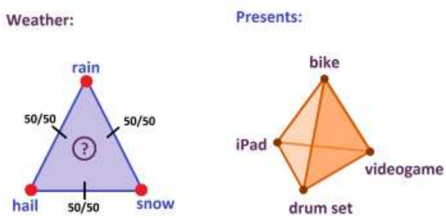
The result is a simplicial complex that might reveal the topology of whatever is behind the incomplete and noisy data. It is called the *Vietoris-Rips complex*.



Exercise 4.41. For a few values of r, construct the Vietoris-Rips complex for: four points on a line, or at the corners of a square, five points at the corners of a pentagon, five random points in the plane, six points at the corners of a cube. Hint: even though the points are in the plane, the complex might be of higher dimension.

4.7 Social choice: the lottery of life

Suppose we face $n+1$ possible events or outcomes, $A_0, ..., A_n$. They are conveniently located at the vertices of an n-simplex $\sigma^n = A_0...A_n$.



In addition to these "primary" events, we can also to consider their *combinations*. In what sense? These events are seen as mutually exclusive but *all* of them may come true. What happens is determined by the probabilities assigned to the primary events. These convex combinations of primary events are called *lotteries*, or, in the context of game theory, *mixed strategies*. The lotteries are conveniently represented by the points of the simplex σ^n and the barycentric coordinates of a point are the probabilities of the outcomes of the lottery. Then σ^n becomes the *probability simplex*.

Example 4.42. The midpoint between "rain" and "snow" represents the lottery when either is equally likely to appear while "hail" is impossible. The probabilities of the three events give the vector $(\frac{1}{2}, \frac{1}{2}, 0)$ of barycentric coordinates of this point. □

We next consider possible preferences of a person among these primary choices and, furthermore, among the lotteries.

The person's *set of preferences* is given by an order relation on the set of lotteries, i.e., the probability simplex σ^n. This means, as we know, that the following two axioms are satisfied:

- completeness: for any lotteries x, y, exactly one of the following holds:

$$x < y,\ y > x, \quad \text{or } y = x;$$

- transitivity: for any lotteries x, y, z, we have:

$$x \le y,\ y \le z \Longrightarrow x \le z.$$

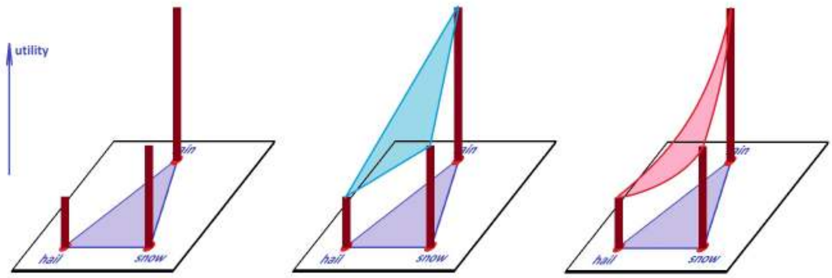
Sometimes these preferences are expressed by a single function.

Definition 4.43. A *utility* of a given set of preferences is a function

$$u : \sigma^n \to \mathbf{R},$$

that satisfies:

$$x < y \Longleftrightarrow u(x) < u(y).$$



Example 4.44. One may prefer:

$$rain \ > \ snow \ > \ hail,$$

and define the utility function u by assigning:

$$u(r) := 3, \quad u(s) := 2, \quad u(h) := 1.$$

Then, we may evaluate the lotteries by extending u to the whole 2-simplex, by linearity:

$$u(t_1 r + t_2 s + t_3 h) := t_1 \cdot 3 + t_2 \cdot 2 + t_3 \cdot 1.$$

Of course, a circular preference, such as

$$rain \ > \ snow \ > \ hail \ > \ rain,$$

doesn't have a corresponding utility function. □

Under certain special circumstances the preferences can be represented by a utility function of a very simple form – a linear combination of the utilities of the primary outcomes:

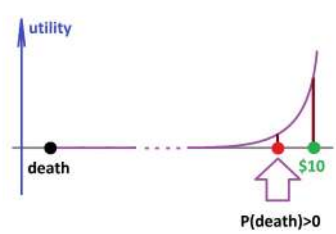
$$U(p_0, ..., p_n) = \sum_i p_i u_i,$$

where $u_0, ..., u_n$ are some numbers. This function is called the *expected utility.*

Of course, a person may have preferences that vary non-linearly but it is always assumed that *the utility function is continuous.*

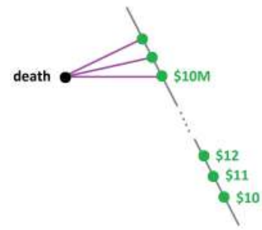
In the study of human behavior, this assumption may present a problem. Even though one would always choose \$10 over death, the continuity implies that, for a small enough probability α, he would see a positive value in the following extreme lottery:

- death: probability $\alpha > 0$; and
- \$10: probability $1 - \alpha$.



Exercise 4.45. Show that, moreover, such a lottery (with death as a possible outcome) will have a higher utility than that of \$1, for a small enough α.

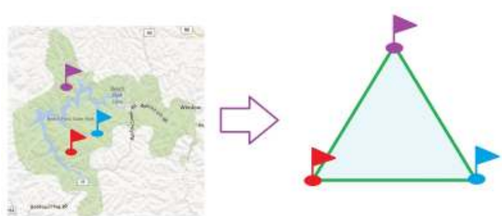
The reasons why we find this conclusion odd may be that, even though they are clearly comparable, death and \$10 seem *incompatible*! On the other hand, death and \$10 million may seem compatible to some people and these people would take part in such a lottery. To account for these observations, we should have an edge between "death" and \$10M but no edge between "death" and \$10:



Then, instead of a single simplex, *the space of outcomes is a simplicial complex*. The complex is meant to represent all possible or plausible lotteries. (Note that one can start with \$10 and continue gradually, in a compatible fashion, to worse and worse things until reaching death.)

Do we ever face a space of outcomes with a more complex topology, such as one with holes, voids, etc.?

Let's recall the problem of social choice for two hikers: we are to develop a procedure for finding a fair compromise on the location of the camp on the shore of a lake. In the original statement of the problem, the choices of sites are arbitrary locations. Still, a simplicial interpretation can be justified too, as follows. Suppose there are three camp sites already set up on the shore of the lake and let's assume that they are the only ones available.



Either hiker has a preference location on the shore, but, for the record, he may have to make this choice by assigning an appropriate probability, or weight, to each camp site. For example, if the person's choice is located between camps A and B and twice as close to A than to B, he may assign: $2/3$ to A, $1/3$ to B, and 0 to C. That sets up a lottery for him. By allowing no more than two non-zero weights, we limit ourselves to the simplicial complex of the hollow triangle. (For the original problem, once the aggregation of the choices is completed and the compromise location is chosen, its coordinates are used for a lottery on the whole 2-simplex.)

To sum up, we have disallowed some lotteries/mixed strategies in order to prevent the participants from doing some silly things: betting \$10 against death or placing a camp in the middle of a lake. This decision has produced a space of choices with a possibly *non-trivial homology*. As we shall see later, this will make impossible some compromises or other desirable social arrangements.

5 Simplicial homology

5.1 Simplicial complexes

Recall that a chain complex is a sequence of vector spaces and linear operators:

$$... \xrightarrow{\partial_{k+2}} C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} ... \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0.$$

that satisfies the *double boundary identity*:

$$\partial_{k+1}\partial_k = 0, \ \forall k.$$

This property allows us to study the *homology* of this chain complex:

$$H_k := \ker \partial_k / \operatorname{Im} \partial_{k+1}.$$

Where do chain complexes come from?

We started with chain complexes for *graphs*. They are very short:

$$0 \xrightarrow{\partial_2=0} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0=0} 0.$$

Here $C_0(G)$ and $C_1(G)$ are the groups of chains of nodes and of edges of the graph G.

We also constructed chain complexes for *cubical complexes*:

$$C_k = C_k(K),$$

where $C_k(K)$ is the group of k-chains of cubical complex K.

Graphs and cubical complexes are very different in nature. Cubical complexes are built via *discretization of the Euclidean space* in order to study its topology and the topology of its subsets. Meanwhile, we approach topology of graphs from the opposite direction, which may be called *Euclidization of data*, and yet we can still end up studying the topology of subsets of the Euclidean space – via realizations of graphs. We will follow this latter route with simplicial complexes.

Let's review the definitions.

Definition 5.1.

• A collection K of subsets of a finite set S is called an *abstract simplicial complex*, or just a simplicial complex, if all subsets of any element of K are also elements of K:

$$\tau \in K, \sigma \subset \tau \Longrightarrow \sigma \in K.$$

• The elements of these subsets are called *simplices*. If such a subset a has exactly $n+1$ elements, it is called an *n-simplex*, or simplex of dimension $n = \dim a$.

• The highest dimension of a simplex in K is called the *dimension* of K:

$$\dim K := \max_{a \in K} \dim a.$$

• For a given n, the collection of all k-simplices in K with $k \leq n$ is called the *n-skeleton* of K denoted by $K^{(n)}$:

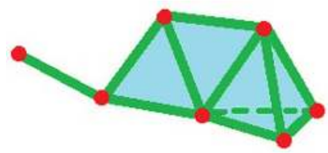
$$K^{(n)} := \{a \in K : \ \dim a \leq n\}.$$

• A *realization* of a simplicial complex K is a geometric simplicial complex $Q = |K|$ along with such a one-to-one correspondence f of the simplices (starting with the vertices) of K with the simplices of Q that the faces are preserved:

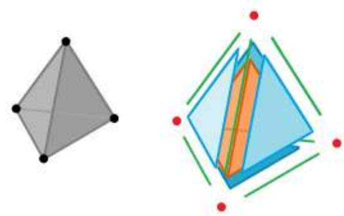
$$\sigma \subset \tau \Longrightarrow f(\sigma) \subset f(\tau).$$

• A representation of a topological space X as a homeomorphic image of a realization of a simplicial complex K is called its *triangulation* and these spaces are called *polyhedra*. We still say that X is a *realization* of K.

Here is an example of a realization of a 2-dimensional simplicial complex as a metric complex in $\mathbf{R}^3$:



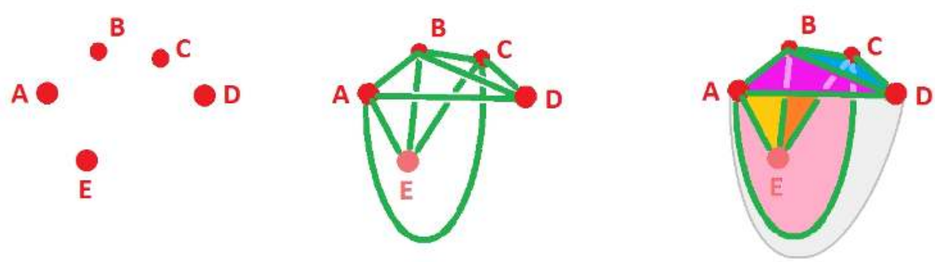
And here is a complex that may seem to contain only one 3-simplex, blown up:



Example 5.2. A realization of the complex

$$K = \{ABD, BCD, ACD, ABE, BCE, ACE\}$$

is found by gluing these 6 "triangles" to each other. Or we can start by listing the lower-dimensional cells:
• 0: {A,B,C,D,E},
• 1: {AB,BD, AD,BC,CD,AE,BE,AC,CE}.
Then we build a realizaiton of K skeleton by skeleton:



We throw the vertices (the singletons from K) around in $\mathbf{R}^3$ and then connect them by the paths (the pairs in K) trying to keep them unknotted for simplicity. Finally, we add the faces (the triples in K) as pieces of fabric stretched on these wire-frames.

The result is homeomorphic to the sphere. □

The topology of realizations of simplicial complexes is similar to those of cubical complexes.

Proposition 5.3. A polyhedron in a Euclidean space is closed and bounded.

Exercise 5.4. Prove the proposition. Hint: prove for a simplex first.

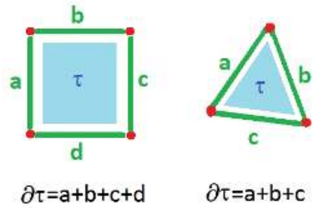
Exercise 5.5. Sketch realizations of these complexes:
• AB, BC, CD, DA, CA, BD; and
• ABC, ABD, ABE.

Exercise 5.6. Is there a 2-dimensional simplicial complex that can't be realized in $\mathbf{R}^3$?

Exercise 5.7. Prove that any abstract simplicial complex K has a realization. Hint: try $\mathbf{R}^N$, where N is the number of vertices in K.

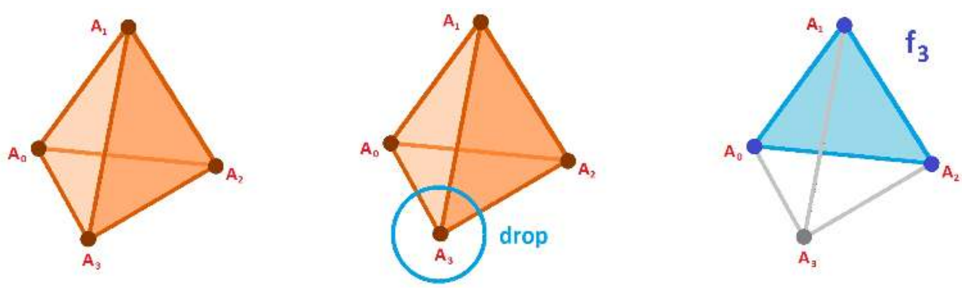
5.2 Boundaries of unoriented chains

Since they have fewer edges, simplices are indeed *simpler* than anything else, even cubes. Just consider the boundary:



Indeed,
- in dimension 2, a triangle has 3 sides, while a square has 4;
- in dimension 3, a tetrahedron has 4 faces, while a cube has 6;
- ...
- in dimension n, an n-simplex has $n+1$ $(n-1)$-faces, while an n-cube has $2n$.

The boundary of a n-simplex is made of its $(n-1)$-faces. Those are easy to construct by choosing one vertex at a time and considering the opposite face:



As a data structure, the n-simplex is just a list of $n+1$ items called vertices:

$$s = A_0A_1...A_n,$$

and any of its $(n-1)$-faces is the same list with one item dropped (indicated by the brackets):

$$\begin{array}{ll} f_0 & = [A_0]A_1...A_n, \\ f_1 & = A_0[A_1]...A_n, \\ & ... \\ f_n & = A_0A_1...[A_n]. \end{array}$$

Such a simple construction doesn't exist for cubes. Instead, we relied on their product structure to prove our theorems.

In the unoriented case, a chain is, as before, defined as a "combination" of cells, i.e., a formal, binary sum of cells. In the last example, there are only these 1-chains:

$$0,\ a,\ b,\ c,\ a+b,\ b+c,\ c+a,\ a+b+c.$$

It is clear now that the boundary of an n-simplex is the sum of its $(n-1)$-faces. It's an $(n-1)$-chain:

$$\partial_n s := \sum_{i=0}^{n} f_i = \sum_{i=0}^{n} A_0...A_{i-1}[A_i]A_{i+1}...A_n.$$

This happens for any simplex of a given simplicial complex K. As before, we extend the operator from cells to chains and the familiar groups. They have the same names with "simplicial" and "over $\mathbf{Z}_2$" attached whenever necessary. So, we have the *simplicial...*

- *chain groups* $C_k(K),\ \forall k$;
- *chain complex*:

$$... \quad \to \quad C_{k+1}(K) \xrightarrow{\partial_{k+1}} C_k(K) \xrightarrow{\partial_k} C_{k-1}(K) \xrightarrow{\partial_{k-1}} ... \quad \to \quad 0;$$

- *cycle groups* $Z_k(K) := \ker \partial_k,\ \forall k$;
- *boundary groups* $B_k(K) := \operatorname{Im} \partial_{k+1},\ \forall k$;
- *homology groups* $H_k(K) := Z_k(K)/B_k(K),\ \forall k$,

... *over* $\mathbf{Z}_2$.

The last item is made possible by the following result.

Theorem 5.8 (Double Boundary Identity). For unoriented chains and simplicial complexes, we have:

$$\partial_k \partial_{k+1} = 0.$$

Exercise 5.9. Prove the theorem algebraically for $k = 0, 1, 2$. Also provide sketches of the construction.

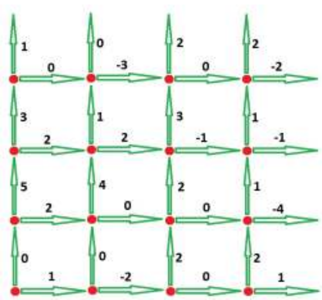
The homology theory of simplicial complexes over $\mathbf{Z}_2$ is now complete!

Still, our main interest remains homology over $\mathbf{R}$ and $\mathbf{Z}$.

5.3 How to orient a simplex

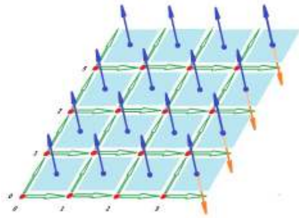
The first step in developing the homology over the reals is to add more structure to the complex – *orientation* of cells.

Recall that for cubical complexes we initially defined the orientation for 1-cells only as their directions. It can be arbitrary or it can be aligned with the axes:



These directions, however, don't suggest any particular way of orienting the squares. Instead, it is defined as clockwise or counterclockwise orderings of the vertices. This approach, however, proved itself too complex for higher dimensions and we instead relied on the representation of each cube as a product of edges and vertices.

Defining such orientation is also possible based the geometry of the Euclidean space; for example, a plane (and a square) in the 3-dimensional space is oriented by a choice of one of the two unit normal vectors.



Exercise 5.10. Show that this is the choice between two equivalence classes of normal vectors.

The idea extends to subspaces of dimension $n-1$ in $\mathbf{R}^n$. Regrettably, this "external" approach to orientation fails when the dimension of the subspace is less than $n-1$ because the normal vectors don't split into two equivalence classes anymore.

The most important thing we know is that

there should always be exactly *two* orientations.

Following our treatment of cubical complexes, we first define the *directions of all edges* of our simplicial complex and, this time, the direction of an edge isn't dictated by the ambient space (there is none) – the choice is entirely ours!

We illustrate this idea below. In this simplicial complex representation (triangulation) of the circle, all three edges have been oriented, at random:

Here, the 1-simplices aren't just edges a, b, c but

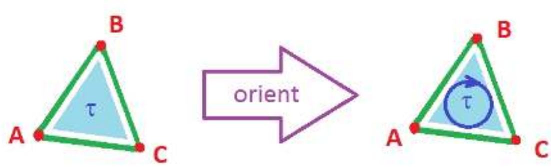
$$a = AB = -BA,\ b = CB = -BC,\ c = AC = -CA.$$

We could have chosen any other order of vertices: $a = BA, b = CB, c = CA$, etc.

Early on, let's make it clear that another choice of cells' orientations will produce different but isomorphic groups of chains. In fact, choosing an orientation of a complex is just a special case of choosing a basis of a vector space.

Next, after using directions to define orientation of edges, how do we handle orientations of higher-dimensional simplices?

As there is no such thing as "the direction of a triangle", we need to think of something else. Looking at the picture below, the answer seems obvious: we go around it either clockwise or counterclockwise:



Unfortunately, this idea has no analog for a 3-simplex. There is no such thing as a "clockwise path" in a cube. In fact, there is no circular path through all of its vertices that capture all possible orderings.

Exercise 5.11. Prove the last statement. Hint: just count.

We find the answer in the definition of simplicial complex: it relies only on the data structure and so should the definition of orientation.

First, let's observe that speaking of edge AB being an element of complex K is misleading because the elements of K are *subsets* of set S. Therefore, to be precise we wouldn't write:

$$AB \in K,$$

until the orientation has been already introduced, but instead:

$$\{A, B\} \in K.$$

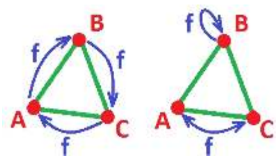
Considering the fact that these are equal, $\{A, B\} = \{B, A\}$, as sets, here's our conclusion: choosing AB over BA is equivalent to choosing an ordering of the set $\{A, B\}$.

So, the idea is that an orientation of a simplex is simply a specific choice of order of its vertices.

Let's consider the triangle. For a 2-simplex with vertices A, B, C, we have 6 possible orderings of the vertices:

- ABC, BCA, and CAB for clockwise, and
- ACB, CBA, and BAC for counter-clockwise.

But there should be only two orientations! It follows then that an orientation should be a choice of one of these *two classes* of orderings. What are they exactly? They are simply "reorderings" of ABC, i.e., self-bijections of $\{A, B, C\}$. We recognize them as *permutations*.



Recall, an even/odd permutation is the composition of an even/odd number of *transpositions*, elementary permutations that "flips" exactly two elements. For example,

- it takes two flips to get CAB from ABC, so it's even; but
- it takes one flip to get ACB from ABC, so it's odd.

We know that the even permutations form the subgroup $\mathcal{A}_n$ of the *symmetric group* $\mathcal{S}_n$ of all permutations of a set of n elements.

Exercise 5.12. Prove the above statement.

Exercise 5.13. A permutation $\sigma \in \mathcal{S}_n$ of the basis vectors of $\mathbf{R}^n$ defines a linear operator. What can you say about its determinant? What about $\sigma \in \mathcal{A}_n$?

So, the definition of orientation of simplices is based on the fact that there are *two equivalence classes of orderings* of any set and, in particular, on the set of vertices of a simplex:

two orderings are equivalent if they differ by an even permutation.

Of course, these two classes are the two cosets that form the following quotient group:

$$\mathcal{S}_n/\mathcal{A}_n \cong \mathbf{Z}_2.$$

Definition 5.14. Suppose an n-simplex τ is given as an ordered list of $n+1$ elements, $\tau = A_0A_1...A_n$. Then we say that τ is *an oriented simplex* if either the class of even permutations of the vertices $\{A_0, A_1, ..., A_n\}$ is chosen or the class of odd permutations.

Therefore, we will always have to deal with this ambiguity in all our constructions as each ordering that we use is, in fact, an *equivalence class* of orderings. For example,

- ABC means $[ABC] = \{ABC, BCA, CAB\}$; and
- ACB means $[ACB] = \{ACB, CBA, BAC\}$.

The good news is, it suffices to list a single *representative* of the class to orient the simplex.

Exercise 5.15. Demonstrate how this scheme fails for cubical complexes. Hint: try dimension 2.

Exercise 5.16. With the meaning of orientation of cells of dimensions 1 and 2 clear, what could be the meaning of orientation of a vertex?

5.4 The algebra of oriented chains

In this section, our ring of coefficients is either $\mathbf{R}$ or $\mathbf{Z}$.

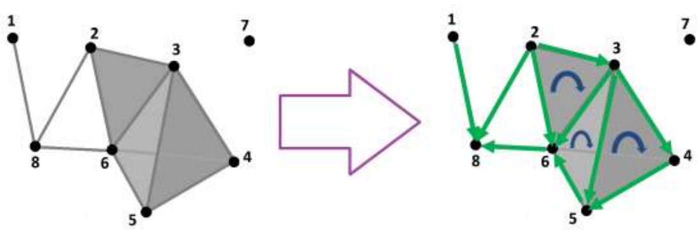
Definition 5.17. Suppose a simplicial complex K consists of simplices as subsets of set S. We say that K is an *oriented simplicial complex* if its every simplex is oriented. (We will omit "oriented" when there can be no confusion.)

Example 5.18. The simplicial complex K of the triangle is

unoriented:	$K = \{\{A\},$	$\{B\},$	$\{C\},$	$\{A,B\},$	$\{C,B\},$	$\{A,C\},$	$\{A,B,C\}\}$; or
oriented:	$K = \{A,$	$B,$	$C,$	$AB,$	$CB,$	$AC,$	$[ABC]\}$.

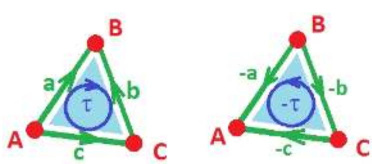
Note that the orientations of the edges of τ don't match the orientation of τ itself. The reason is, they don't have to... We shouldn't expect such a requirement because there may be another 2-simplex adjacent to one of these edges with the opposite orientation. □

To acquire an orientation for complex K, one can just order its vertices and then use that order to orient each simplex in K:



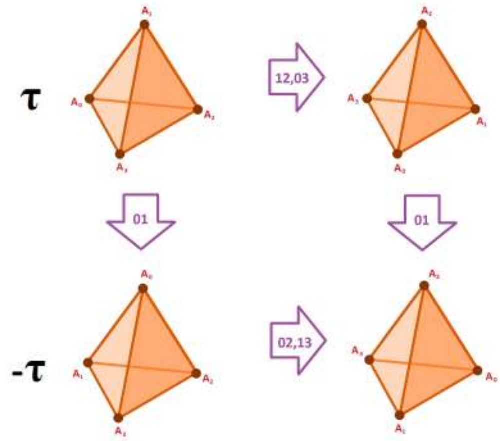
Exercise 5.19. In how many ways can this be done?

Up to this point, the development of the algebra of chains follows the same path as in the case of oriented *cubical* complexes. Let's point out some differences. First, *all* simplices have to be supplied with orientations not just 1-dimensional ones. Second, the meaning of $-\tau$ for a simplex τ is clear: it's τ with the opposite orientation. For dimensions 1 and 2, the explicit meaning is also clear:



For higher dimensions, it's a transposition of two vertices in the ordering:

$$A_1A_0A_2...A_n = -A_0A_1A_2...A_n.$$



It is worth repeating what oriented simplices are and aren't.

- Oriented simplices aren't sets of the vertices: not $\{A, B, C\}$.
- Oriented simplices aren't orderings of the vertices: not ABC.
- Oriented simplices are *classes* of orderings of the vertices, such as $[ABC] = \{ABC, BCA, CAB\}$.

This is how we make it precise.

Theorem 5.20. For any permutation $s \in \mathcal{S}_{n+1}$, we have:

$$A_0A_1...A_n = \pi(s)A_{s(0)}A_{s(1)}...A_{s(n)},$$

where $\pi(s)$ is its *parity*, i.e.,

$$\pi(s) := \begin{cases} 1 & \text{if } s \text{ is even,} \\ -1 & \text{if } s \text{ is odd.} \end{cases}$$

Recall that the point of using oriented chains is to be able to capture all possible ways to go around the circle: going once, twice, or thrice around it, or going in the opposite direction.

An oriented simplicial chain is a "formal" linear combination of finitely many *oriented* simplices, such as $3a + 5b - 17c$. The coefficients come from some ring R. From this point of view, the chains previously discussed are over $R = \mathbf{Z}_2$, i.e., binary. We will concentrate on *real* chains, i.e., ones with real coefficients. Nonetheless, the analysis will, with just a few exceptions, apply also to the integers $R = \mathbf{Z}$, rational numbers $R = \mathbf{Q}$, etc.

Next, for a given simplicial complex K and $k = 0, 1, 2, ...$, let $C_k(K)$ denote the set of all real k-chains:

$$C_k(K) := \left\{ \sum_i s_i\sigma_i : \ s_i \in \mathbf{R}, \sigma_i \text{ is a } k\text{-simplex in } K \right\}.$$

It is easy to define the *sum of two chains* by assigning appropriate coefficients of each simplex: if

$$A = \sum_i s_i\sigma_i, B = \sum_i t_i\sigma_i \Longrightarrow A + B := \sum_i (s_i + t_i)\sigma_i.$$

To see that the operation above is well-defined, one can assume that each sum lists *all* simplices present in both sums – some with zero coefficients.

Theorem 5.21. $C_k(K)$ is an abelian group with respect to chain addition.

Proof. We verify the axioms of group below.

- (1) Identity element:

$$0 = \sum_i 0 \cdot \sigma_i.$$

- (2) Inverse element:

$$A = \sum_i s_i \sigma_i \Longrightarrow -A = \sum_i (-s_i)\sigma_i.$$

- (3) Associativity:

$$\begin{aligned} A + (B + C) &= \sum_i (s_i + t_i + u_i)\sigma_i \\ &= (A + B) + C. \end{aligned}$$

■

It is also easy to define the *scalar multiplication of chains*: if

$$A = \sum_i s_i \sigma_i, \text{ and } r \in \mathbf{R},$$

then we let

$$rA := \sum_i (rs_i)\sigma_i.$$

Theorem 5.22. (1) $C_k(K; \mathbf{R})$ is a vector space with respect to chain addition and scalar multiplication with a basis that consists of all k-simplices in K. (2) $C_k(K; \mathbf{Z})$ is an abelian group with respect to chain addition with the set of generator that consists of all k-simplices in K.

Exercise 5.23. Prove the theorem.

Example 5.24. In particular, in the example of a triangle representing the circle, we have

- $C_0(K) = < A > \cong \mathbf{R}$,
- $C_1(K) = < a, b, c > \cong \mathbf{R} \oplus \mathbf{R} \oplus \mathbf{R}$,
- $C_2(K) = 0$, etc.

□

Exercise 5.25. Prove that choosing a different orientation of a simplicial complex produces a chain group isomorphic to the original. Present the matrix of the isomorphism.

Exercise 5.26. What does the algebra of oriented chains over $\mathbf{Z}_2$ look like?

Until some connection is established between chains of different dimensions, this algebra has nothing to say about the topology of the complex. This connection is given by the boundary operator.

5.5 The boundary operator

The boundary operator of a simplicial complex records the relations between cells/chains of consecutive dimensions.

Let's review what we already know.

Example 5.27. Suppose we have a simplicial complex K of the solid triangle seen before:

- 0-simplices: A, B, C;
- 1-simplices: $a = AB,\ b = CB,\ c = AC$;
- 2-simplex: $\tau = ABC$.

The boundary of a vertex empty, hence the boundary operator of a 0-chain is 0:

$$\partial(A) = 0, \ \forall A \in C_0(K).$$

The boundary of a 1-cell consists of its two endpoints, so, in the binary setting, this was simple:

$$\partial(a) = \partial(AB) = A + B.$$

In the oriented setting, the direction of the edge matters ($AB \neq BA$); so we define:

$$\partial(a) = \partial(AB) := B - A, \ \forall a = AB \in C_1(K).$$

In dimension 2, the boundary $\partial\tau$ is defined as a linear combination of its faces:

$$\partial\tau = AB + BC + CA = a - b - c,$$

or:

$$\begin{array}{lllll} \partial\tau & = & \partial\Big(\triangle\Big) & & \\ & = & \nearrow + & \searrow + & \leftarrow \\ & = & a + & (-b) + & (-c) \\ & = & a - b - c. & & \end{array}$$

Here, we first follow the chain clockwise as indicated by its orientation writing the 1-cells as they appear, and then interpret these 1-cells in terms of the 1-cells, a, b, c, that we listed in the complex K. □

Let's define the boundary of a simplex of an arbitrary dimension.

Recall first what we mean by a *face* of a simplex. As the n-simplex is just a list of $n+1$ vertices, any of its $(n-1)$-faces is a list of the same vertices with one dropped:

$$\begin{array}{ll} f_0 & = [A_0]A_1...A_n, \\ f_1 & = A_0[A_1]...A_n, \\ & ... \\ f_n & = A_0A_1...[A_n]. \end{array}$$

The question now is, what are the *oriented* faces of an *oriented* simplex? The answer is, of course, they are *oriented* simplices.

The idea is very simple: removing an item from an ordering gives an ordering of the smaller set, just as shown above. Yet, to see that this answer is justified isn't as simple as it seems. Indeed, we aren't dealing with orderings but with *classes* of orderings. Does this operation make sense?

Proposition 5.28. Removing an item from an even/odd ordering gives us an even/odd ordering of the smaller set; or, algebraically,

$$\mathcal{A}_n = \mathcal{S}_n \cap \mathcal{A}_{n+1}.$$

Exercise 5.29. Prove the proposition.

Now, the boundary of an n-simplex s is the sum of its faces, in the unoriented case. As we see in the last example, this time the faces appear with various signs. As it turns out, the boundary now is the *alternating* sum of its faces:

$$\partial s = f_0 - f_1 + f_2 - ... \pm f_n.$$

Definition 5.30. The *boundary of an oriented n-simplex* is an oriented $(n-1)$-chain defined to be

$$\partial(A_0A_1...A_i...A_n) := \sum_i (-1)^i A_0A_1...[A_i]...A_n.$$

Example 5.31. Let's apply the definition to the complex of the triangle, as in the last example. These are the oriented simplices based on the obvious ordering of the 3 vertices:

- 0-simplices: A_0, A_1, A_2;
- 1-simplices: A_0A_1, A_1A_2, A_0A_2;
- 2-simplex: $A_0A_1A_2$.

Then

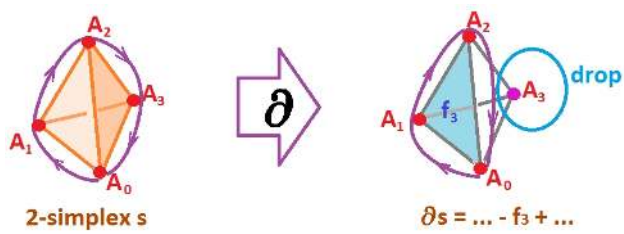
$$\begin{aligned}\partial(A_0A_1A_2) &= [A_0]A_1A_2 - A_0[A_1]A_2 + A_0A_1[A_2]\\ &= A_1A_2 - A_0A_2 + A_0A_1\\ &= A_0A_1 + A_1A_2 + A_2A_0.\end{aligned}$$

So, these edges go clockwise around the triangle, as expected. □

Notice that the input of the formula in the definition is an ordering while it is supposed to be an oriented simplex, i.e., a class of orderings. What happens if we use another representative from the class of orderings as the input of this formula?

Exercise 5.32. Confirm that the result is the same for $\partial(A_2A_0A_1)$.

A single term in the boundary formula for a 2-simplex is shown below:



To further understand what's going on, let's flip the first two vertices:

$$\begin{aligned}&\partial(A_1A_0A_2...A_n)\\ &= \textstyle\sum_i (-1)^i A_1A_0A_2...[A_i]...A_n\\ &= [A_1]A_0A_2...A_n - A_1[A_0]A_2...A_n \quad + \textstyle\sum_{i>1}(-1)^i A_1A_0A_2...[A_i]...A_n\\ &= A_0A_2...A_n - A_1A_2...A_n \quad + \textstyle\sum_{i>1}(-1)^i A_1A_0A_2...[A_i]...A_n\\ &= -A_1A_2...A_n + A_0A_2...A_n \quad + \textstyle\sum_{i>1}(-1)^i(-1)A_0A_1A_2...[A_i]...A_n\\ &= -[A_0]A_1A_2...A_n + A_0[A_1]A_2...A_n \quad - \textstyle\sum_{i>1}(-1)^i A_0A_1A_2...[A_i]...A_n\\ &= -\textstyle\sum_i (-1)^i A_1A_0A_2...[A_i]...A_n\\ &= -\partial(A_0A_1A_2...A_n).\end{aligned}$$

The sign of the boundary chain has reversed! Therefore, we have the following:

Theorem 5.33. The boundary chain of an ordered simplex is well-defined. In particular, for any oriented simplex s, we have

$$\partial(-s) = -\partial(s).$$

Exercise 5.34. Provide the rest of the proof.

With the operator defined on each of the simplices, one can extend this definition to the whole chain group by linearity, thus creating a linear operator:

$$\partial : C_n(K) \to C_{n-1}(K).$$

Indeed, since we know $\partial(\sigma_i)$ for each simplex $\sigma_i \in C_n(K)$, we set:

$$\partial\Big(\sum_i s_i\sigma_i\Big) := \sum_i s_i\partial(\sigma_i).$$

5.6 Homology

The key fact needed for homology theory is:

all boundaries are cycles.

Algebraically, it is given by the following theorem:

Theorem 5.35 (Double Boundary Identity). $\partial\partial = 0$ over $\mathbf{R}$ or $\mathbf{Z}$.

Proof. It suffices to prove that $\partial\partial(s) = 0$ for any simplex s in K. The idea is that in the expansion of $\partial\partial s$, each $(n-2)$-face appears twice but with opposite signs.

Suppose s is an n-simplex:

$$s = A_0 ... A_n.$$

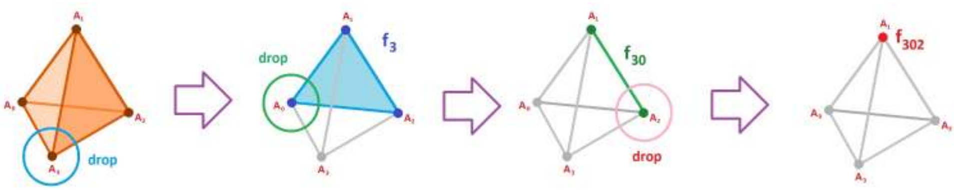
For each $i = 0, 1, ..., n$, f_i is the ith face of s:

$$f_i = A_0 ... [A_i] ... A_n.$$

It is an oriented $(n-1)$-simplex. For each pair $i, j = 0, 1, ..., n$, $i \neq j$, let $f_{ij} = f_{ji}$ be the jth face of f_i or, which is the same thing, the ith face of f_j:

$$f_{ij} = A_0 ... [A_i] ... [A_j] ... A_n.$$

It is an oriented $(n-2)$-simplex.



First, we use the definition and then the linearity of the boundary operator:

$$\begin{aligned}\partial\partial s &= \partial\Big(\textstyle\sum_i (-1)^i f_i\Big)\\ &= \textstyle\sum_i (-1)^i \partial(f_i).\end{aligned}$$

Consider the ith term, i.e., the summation over j, and watch how the signs alternate:

$$\begin{aligned}\partial f_i &= \partial(A_0A_1...[A_i]...A_n)\\ &= \sum_{j<i}(-1)^j A_0A_1...[A_j]...[A_i]...A_n \qquad [A_j] \text{ in the } j\text{th term of the sum...}\\ &+ \sum_{j>i}(-1)^{j-1} A_0A_1...[A_i]...[A_j]...A_n \qquad [A_j] \text{ in the } (j-1)\text{st term...}\\ &= \sum_{j<i}(-1)^j f_{ij} + \sum_{j<i}(-1)^{j-1} f_{ij}.\end{aligned}$$

So, the sign attached to the face becomes the opposite as j goes past i. That's what will ensure cancelling...

We substitute now and then deal with double summation, over i and j:

$$\begin{aligned}\partial\partial s &= \sum_i (-1)^i \Big(\sum_{j<i} (-1)^j f_{ij} + \sum_{j<i} (-1)^{j-1} f_{ij} \Big) \\ &= \sum_{i<j} (-1)^{i+j} f_{ij} - \sum_{i>j} (-1)^{i+j} f_{ij} \\ &= 0.\end{aligned}$$

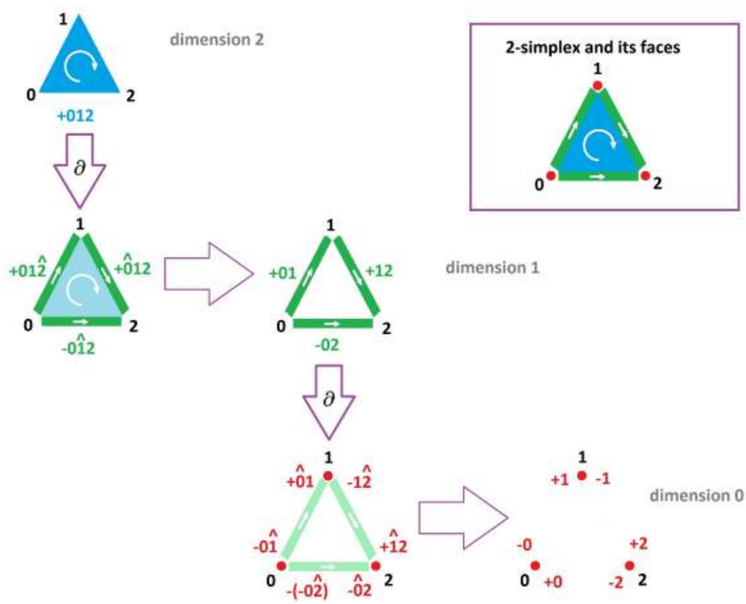
Let's summarize what has happened. The $(n-2)$-faces of our simplex s form a symmetric $(n+1)\times(n+1)$ matrix. Then $\partial\partial s$ is the sum of the elements of this matrix, with appropriate signs, after the diagonal is removed:

$$\{f_{ij}\} = \left[\begin{array}{c|ccc|c|c|c|c|c} & 0 & 1 & \dots & i & \dots & j & \dots & n \\ \hline 0 & 0 & -f_{01} & \dots & +f_{0i} & \dots & -f_{0j} & \dots & +f_{0n} \\ 1 & +f_{10} & 0 & \dots & -f_{1i} & \dots & +f_{1j} & \dots & -f_{1n} \\ & . & . & \dots & . & \dots & . & \dots & . \\ & . & . & \dots & . & \dots & . & \dots & . \\ \hline i & -f_{i0} & +f_{i1} & \dots & 0 & \dots & -f_{ij} & \dots & +f_{in} \\ \hline & . & . & \dots & . & \dots & . & \dots & . \\ & . & . & \dots & . & \dots & . & \dots & . \\ \hline j & +f_{j0} & -f_{j1} & \dots & +f_{ji} & \dots & 0 & \dots & -f_{jn} \\ \hline & . & . & \dots & . & \dots & . & \dots & . \\ & . & . & \dots & . & \dots & . & \dots & . \\ n & -f_{n0} & +f_{n1} & \dots & -f_{ni} & \dots & +f_{nj} & \dots & 0 \end{array}\right]$$

Finally, the symmetric entries have opposite signs and, therefore, cancel. ■

Exercise 5.36. Fill the blanks in the table to demonstrate how the signs alternate.

The diagram below illustrates the proof for $n = 2$:



Corollary 5.37. The theorem holds for chains over $\mathbf{Z}_2$.

Exercise 5.38. Prove the corollary (a) by rewriting the above proof, (b) directly from the theorem. Hint: think of an appropriate function $\mathbf{Z} \to \mathbf{Z}_2$.

So, the boundary of the boundary of a cell is zero. Since every chain is a linear combination of cells, it follows from the linearity of ∂ that *the boundary of the boundary of any chain is zero*; i.e.,

$$\begin{array}{lllll} C_k(K) & \xrightarrow{\partial} & C_{k-1}(K) & \xrightarrow{\partial} & C_{k-2}(K) \\ \tau & \mapsto & \partial(\tau) & \mapsto & \partial\partial(\tau) = 0. \end{array}$$

Thus, every boundary is a cycle:

$$B_k(K) \subset Z_k(K).$$

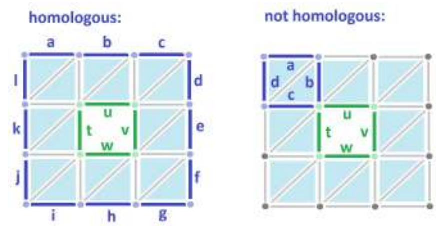
That's why the *homology groups* make sense:

$$H_k(K) := Z_k(K)/B_k(K).$$

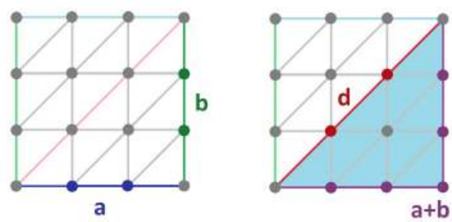
As before, two k-chains are *homologous* if they form the boundary of a $(k+1)$-chains; or, algebraically, if

$$a \sim b \Longleftrightarrow a - b \in B_k.$$

It is easy to make up examples of simplicial homology from those for cubical complexes:



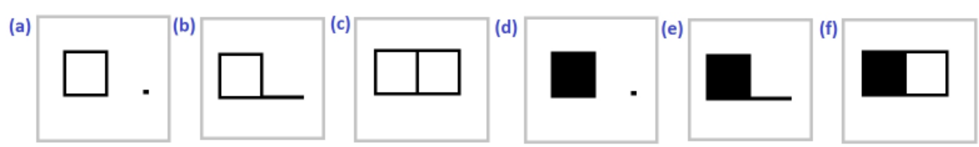
Example 5.39. Below, we visualize the idea why the sum of the two standard generators of the torus is homologous to the diagonal; i.e., $a + b \sim d$, and $(1,1) = (1,0) + (0,1)$ in the homology group:



□

Exercise 5.40. Represent the subsets of the plane given below as realizations of simplicial complexes. Then, for each of them and coefficients in $\mathbf{R}$:

- (1) Find the chain groups and find the boundary operator as a matrix;
- (2) Using only part (1) and linear algebra, find Z_k, B_k for all k.
- (3) Using only part (2) and linear algebra, find the Betti numbers of these complexes.
- (4) Confirm that, even though the chain groups are different, the results match those for cubical complexes. Point out what part of your computation makes it so.



Exercise 5.41. Represent the circle $\mathbf{S}^1$ as a hollow triangle and the sphere $\mathbf{S}^2$ as a hollow pyramid:

and confirm that the homology groups are:

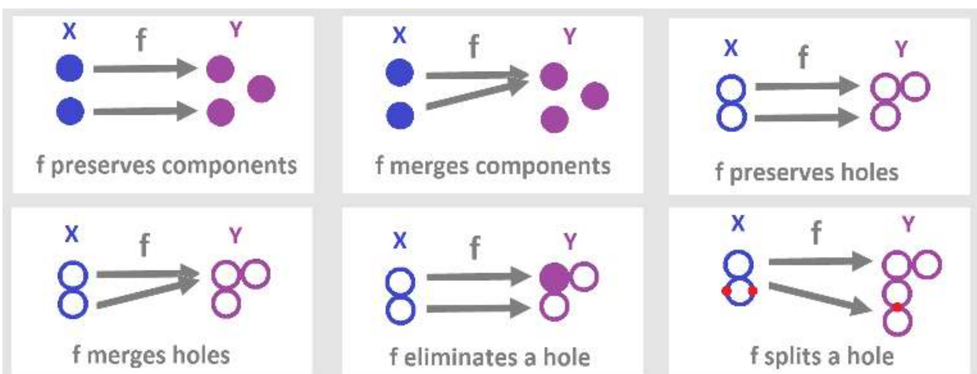
$$H_0(\mathbf{S}^1) = \mathbf{R}, \quad H_1(\mathbf{S}^1) = \mathbf{R}, \quad H_2(\mathbf{S}^1) = 0;$$
$$H_0(\mathbf{S}^2) = \mathbf{R}, \quad H_1(\mathbf{S}^2) = 0, \quad H_2(\mathbf{S}^2) = \mathbf{R}.$$

---

6 Simplicial maps

6.1 The definition

Suppose we have a map $h : X \to Y$ between two topological spaces. To fully understand this map, we want to track each of the topological features, i.e., the homology classes, in X as they are transformed into the ones in Y:



These are the "topological events" that we previously saw happen under graph maps. In the general setting, there will be much more happening. For example, can a void be transformed into a tunnel, or vice versa?

Now, suppose that X and Y are triangulated by two simplicial complexes: $X = |K|, Y = |L|$. We want to see map h as a "realization" $h = |f|$ of a "simplicial" map $f : K \to L$ between these complexes:

$$\begin{array}{ccc} K & \xrightarrow{realization} & |K| \\ \downarrow f & \xrightarrow{realization} & \downarrow |f| \\ L & \xrightarrow{realization} & |L| \end{array}$$

Let's first review what we know about graph maps.

Recall that those are simply functions of nodes of the graph that produce also functions of edges – allowing an edge to be taken to a node, by means of a "collapse". Since graphs are 1-dimensional simplicial complexes, we can rewrite those definitions using the language of simplices.

A graph map $f : K \to L$ is a function between graphs K, L that satisfies, for each edge e, either:

- 1. (cloning) $f(e)$ is an edge g and f takes the endpoints of e to the endpoints of g; or
- 2. (collapsing) $f(e)$ is a node P and f takes the endpoints of e to P.

These are the axioms written in a more compact form:

- 1. $f(AB) = f(A)f(B)$;
- 2. $f(AB) = P \Longrightarrow f(A) = f(B) = P$.

The second axiom is contained in the first if we understand that an edge given by two identical nodes is a node, that node:

$$PP := P.$$

Following this idea, we introduce the following **notation:** we allow *repetition of vertices* in the list that defines a simplex. A list of vertices

$$s = A_0 ... A_n$$

is an m-simplex if there are exactly $m + 1$ distinct vertices on the list.

Now, to understand what a simplicial map $f : K \to L$ between two (so far unoriented) simplicial complexes is, we first assume that f is defined on vertices only: if A is a vertex in K then $f(A)$ is a vertex in L. This function can be, so far, arbitrary but not every such function can be *extended* to the whole complex K!

Suppose now we have a 2-simplex:

$$s = A_0A_1A_2 \in K.$$

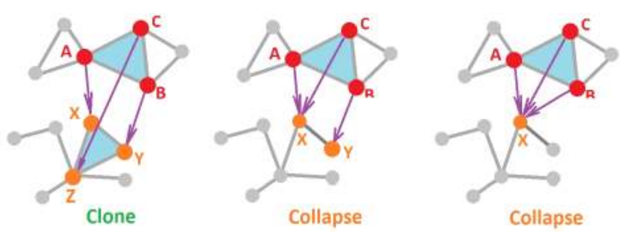
Let's define $f(s) \in L$. We already know what happens under f to the vertices and to the faces of s:

$$f(A_iA_j) = f(A_i)f(A_j).$$

This new simplex $f(s)$ will have to be determined by the values of f at these vertices – the values are already known as vertices in L – as follows:

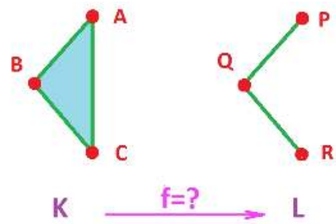
$$f(s) = f(A_0A_1A_2) := f(A_0)f(A_1)f(A_2).$$

The latter is an m-simplex ($m \leq 2$), which is required to belong to L. With the three possibilities for m, there are three possibilities for this new simplex. We see that, unlike a 1-simples, a 2-simplex can collapse in two ways:



Example 6.1. Let's consider possible maps between these two complexes:

- $K = \{A, B, C, AB, BC, CA, ABC\}$,
- $L = \{P, Q, R, PQ, QR\}$.



We start with:

$$f(A) = P, \ f(B) = Q, \ f(C) = R.$$

Then, by the above formula, we have

$$f(AB) = PQ \in L, \ f(BC) = QR \in L, \ f(CA) = RP \notin L.$$

Such a simplicial map is impossible!

Let's try another value for C:

$$f(A) = P, \ f(B) = Q, \ f(C) = Q.$$

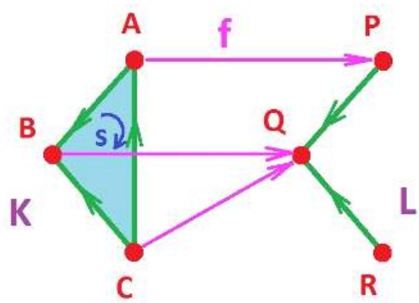
Then two edges clone and one collapses:

$$f(AB) = PQ \in L, \ f(BC) = Q \in L, \ f(CA) = PQ \in L;$$

and the triangle collapses too:

$$f(ABC) = PQ \in L.$$

That works! The result is below:



□

Exercise 6.2. Consider three more possibilities.

Definition 6.3. A function $f : K \to L$ between two (unoriented) simplicial complexes is called a *simplicial map* if it satisfies the following:

$$A_0...A_n \in K \Longrightarrow f(A_0)...f(A_n) \in L;$$

then, the value of f on a simplex of K is a simplex of L given by

$$f(A_0...A_n) := f(A_0)...f(A_n).$$

A bijective simplicial map is called a *simplicial isomorphism*.

The above condition is a kind of *discrete continuity.*

Theorem 6.4. A function $f : K \to L$ is a simplicial isomorphism if and only if it is bijective on vertices.

Theorem 6.5. A constant function $C : K \to L$ (i.e., a map defined by $f(s) = V$ for some fixed vertex $V \in L$) is a simplicial map.

Theorem 6.6. The composition $gf : K \to M$ of two simplicial maps $f : K \to L$ and $g : L \to M$ is a simplicial map.

Exercise 6.7. Prove these theorems.

Exercise 6.8. Provide a simplicial map that "represents" the following map $f : \mathbf{S}^1 \to \mathbf{T}^2$:

Exercise 6.9. Prove that simplicial complexes, together with simplicial maps, form a category.

Exercise 6.10. Give the graph of a simplicial map a structure of a simplicial complex.

Exercise 6.11. Define cubical maps.

As we shall see, simplicial maps are realized as *continuous* functions. But first we will consider the effect of these maps on the homology classes.

6.2 Chain maps of simplicial maps

Suppose a simplicial map

$$f : K \to L$$

between two *oriented* simplicial complexes is given. The map records where every simplex goes, but what about chains? We want now to define the corresponding linear operators:

$$f_k : C_k(K) \to C_k(L),\ k = 0, 1, ...$$

Since the k-simplices of K form a basis of the k-chain group $C_k(K)$, we only need to know the values of f_k for these simplices. And we do, from f.

The only problem is that, while s is a k-simplex in K, its image $f(s)$ might be a simplex in L of a *lower* dimension. Therefore, if s collapses, then $f(s) \notin C_k(L)$, so that $f(s)$ appears to be undefined... To resolve this conundrum, even though $f(s)$ is not a *k-simplex*, can we think of it as a *k-chain*? There is only one legitimate choice for such a chain: it's 0!

So, this is how we define the chain maps $f_0, ..., f_k, ...$ for f.

Definition 6.12. The nth *chain map* f_n of a simplicial maps f is defined by its values on each n-simplex in K,

$$s = A_0A_1...A_n,$$

where $A_0, A_1, ..., A_n$ are its vertices, as follows;

$$f_n(s) := \begin{cases} f(A_0)...f(A_n), & \text{if } f(A_i) \neq f(A_j)\ \forall i \neq j; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, a simplex is cloned or it is taken to 0.

Example 6.13. Let's consider these two complexes from last subsection but first orient them, arbitrarily:

- $K = \{A, B, C, AB, CB, CA, ACB\}$,
- $L = \{P, Q, R, PQ, RQ\}$.

This is the simplicial map we considered above:

$$f(A) = P,\ f(B) = Q,\ f(C) = Q.$$

These three identities immediately give us the three columns of the matrix of the linear operator $f_0 : C_0(K) \to C_0(L)$, or $f_0 : \mathbf{R}^3 \to \mathbf{R}^3$:

$$f_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Further, the simplicial identities

$$f(AB) = PQ,\ f(BC) = Q,\ f(CA) = PQ,$$

give us these chain identities:

$$f_1(AB) = PQ,\ f_1(BC) = 0,\ f_1(CA) = QP.$$

These three give us the three columns of the matrix of the linear map $f_1 : C_1(K) \to C_1(L)$, or $f_1 : \mathbf{R}^3 \to \mathbf{R}^2$:

$$f_1 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Finally, the simplicial identity

$$f(ABC) = PQ$$

gives us the chain identity:

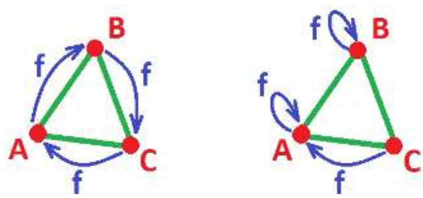
$$f_2(ACB) = 0.$$

Therefore,

$$f_2 = 0.$$

□

Exercise 6.14. Add a 2-simplex to either of these two simplicial complexes, orient the new complexes, and then find the matrices of the chain maps of these two simplicial maps:



The definition of f_n is simple enough, but we can make it even more compact if we observe that

- case 1: $f_n(s) = f(A_0)...f(A_n)$, if $f(A_i) \neq f(A_j)$, for all $i \neq j$;

implies

- case 2: $f_n(s) = 0$, if $f(A_i) = f(A_j)$, for some $i \neq j$.

The reason is that having two coinciding vertices in a simplex – understood as a chain – make it equal 0. Indeed, when we flip them, the sign has to flip too; but it's the same simplex!

$$\begin{array}{ll} B_i = B_j \Longrightarrow & \\ B_0...B_i...B_j...B_n & = -B_0...B_j...B_i...B_n \\ & = -B_0...B_i...B_j...B_n. \end{array}$$

It must be zero then.

We amend the convention from the last subsection.

Notation: We allow *repetition of vertices* in the list that defines a simplex. But, a list of vertices

$$s = A_0 ... A_n$$

is the 0 chain unless there are exactly $n+1$ distinct vertices on the list.

Definition 6.15. The *n-chain map $f_n : C_n(K) \to C_n(L)$ induced by a simplicial map $f : K \to L$* is defined by its values on the simplices, as follows:

$$f_n(A_0 ... A_n) := f(A_0) ... f(A_n).$$

Theorem 6.16. The chain map induced by a simplicial isomorphism is an isomorphism. Moreover, the chain map induced by the identity simplicial map is the identity.

Theorem 6.17. The chain map induced by a simplicial constant map $C : K \to L$ with K, L edge-connected is *acyclic*: $C_0 = \text{Id},\ C_k = 0,\ \forall k > 0$.

Theorem 6.18. The chain map induced by the composition $gf : K \to M$ of two simplicial maps $f : K \to L$ and $g : L \to M$ is the composition of the chain maps induced by these simplicial maps:

$$(gf)_k = g_k f_k.$$

Exercise 6.19. Prove these theorems.

Since we have two simplicial complexes, there are two *chain complexes*:

$$\begin{array}{ccccccccccc}
... & \to & C_{k+1}(K) & \xrightarrow{\partial_{k+1}} & C_k(K) & \xrightarrow{\partial_k} & C_{k-1}(K) & \to & ... & \to & 0, \\
... & \to & C_{k+1}(L) & \xrightarrow{\partial_{k+1}} & C_k(L) & \xrightarrow{\partial_k} & C_{k-1}(L) & \to & ... & \to & 0.
\end{array}$$

The chain maps connect these two chain complexes item by item, "vertically", like this:

$$\begin{array}{ccccccccccc}
... & \to & C_{k+1}(K) & \xrightarrow{\partial_{k+1}} & C_k(K) & \xrightarrow{\partial_k} & C_{k-1}(K) & \to & ... & \to & 0 \\
... & & \downarrow{f_{k+1}} & & \downarrow{f_k} & & \downarrow{f_{k-1}} & & ... & & \\
... & \to & C_{k+1}(L) & \xrightarrow{\partial_{k+1}} & C_k(L) & \xrightarrow{\partial_k} & C_{k-1}(L) & \to & ... & \to & 0.
\end{array}$$

The arrows here are maps and paths of arrows are their compositions. The question is then, are the results "path-independent"? In other words, is the diagram commutative?

Exercise 6.20. Produce such a diagram for the map in the last example.

Exercise 6.21. Define chain maps between chain complexes in such a way that together they form a category.

6.3 How chain maps interact with the boundary operators

We need to understand how the chain map of a simplicial map fits into the chain complexes.

There are two boundary operators to deal with for a given simplicial map

$$f : K \to L,$$

for either of the complexes:

$$\partial_k^K : C_k(K) \to C_{k-1}(K)$$

and

$$\partial_k^L : C_k(L) \to C_{k-1}(L).$$

(Both the subscripts and the superscripts may be omitted.) Along with the chain maps of f, the boundary operators can be combined in a single diagram:

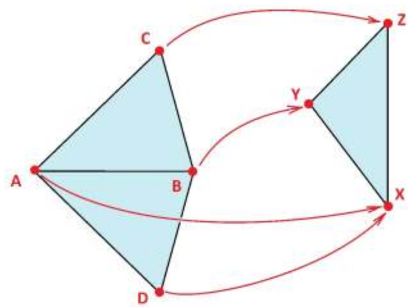
$$\begin{array}{ccc}
C_k(K) & \xrightarrow{f_k} & C_k(L) \\
\Big\downarrow \partial_k^K & \searrow & \Big\downarrow \partial_k^L \\
C_{k-1}(K) & \xrightarrow{f_{k-1}} & C_{k-1}(L)
\end{array}$$

As we shall see, the diagram is commutative. The familiar and wonderfully compact way to present this condition is below:

$$\boxed{\partial f = f\partial}$$

Definition 6.22. A *chain map* is any collection of homomorphisms between the elements of any two chain complexes that satisfies the above condition.

This behavior is expected from simplicial maps. After all, when a simplex is cloned, so is its boundary, face by face:



It is the case of collapse that needs special attention. We will prove the general case.

For the proof we will need the following two facts. Both boundary operators are defined by the same formula for $k = 0, 1, ...$:

$$\partial_k(A_0...A_k) := \sum_{i=0}^{k} (-1)^i A_0...[A_i]...A_k. \qquad (1)$$

We will also need the definition of the chain map of a simplicial map:

$$f_k(A_0...A_k) := f(A_0)...f(A_i)...f(A_k). \qquad (2)$$

Theorem 6.23 (Algebraic Continuity Condition). If $f : K \to L$ is a simplicial map, then its chain maps satisfy for $k = 1, 2, ...$:

$$\partial_k^L f_k = f_{k-1} \partial_k^K.$$

Proof. The proof is by examination. To prove the identity, we evaluate its left-hand side and its right-hand side for

$$s = A_0 A_1 ... A_k \in C_k(K).$$

First, the right-hand side. We use (1),

$$\begin{aligned} f_{k-1}(\partial_k(s)) &= f_{k-1}\Big(\sum_{i=0}^{k}(-1)^i A_0...[A_i]...A_k\Big), && \text{then by linearity} \\ &= \sum_{i=0}^{k}(-1)^i f_{k-1}(A_0...[A_i]...A_k), && \text{then by (2)} \\ &= \sum_{i=0}^{k}(-1)^i f(A_0)...[f(A_i)]...f(A_k). \end{aligned}$$

Next, the left-hand side. We use (2),

$$\begin{aligned} \partial_k(f_k(s)) &= \partial_k(f(A_0)...f(A_k)), && \text{then by (1)} \\ &= \sum_{i=0}^{k}(-1)^i f(A_0)...[f(A_i)]...f(A_k). \end{aligned}$$

Now, as we have proven the identity for all basis elements, simplices, of the vector space, $C_k(K)$, then the two linear operator coincide. ■

Corollary 6.24. The theorem holds for chains over $\mathbf{Z}_2$.

Exercise 6.25. Prove the corollary (a) by rewriting the above proof, (b) directly from the theorem. Hint: think of an appropriate function $\mathbf{Z} \to \mathbf{Z}_2$.

6.4 Homology maps

Thus, a simplicial map induces chain maps between the two chain complexes, but what about the *homology*? What is the effect of the map on the homology classes?

Algebraically, this is what we are after. We already have a linear operator

$$f_k : C_k(K) \to C_k(L)$$

with the above property. And now we need to define somehow another linear operator, the *quotient operator*,

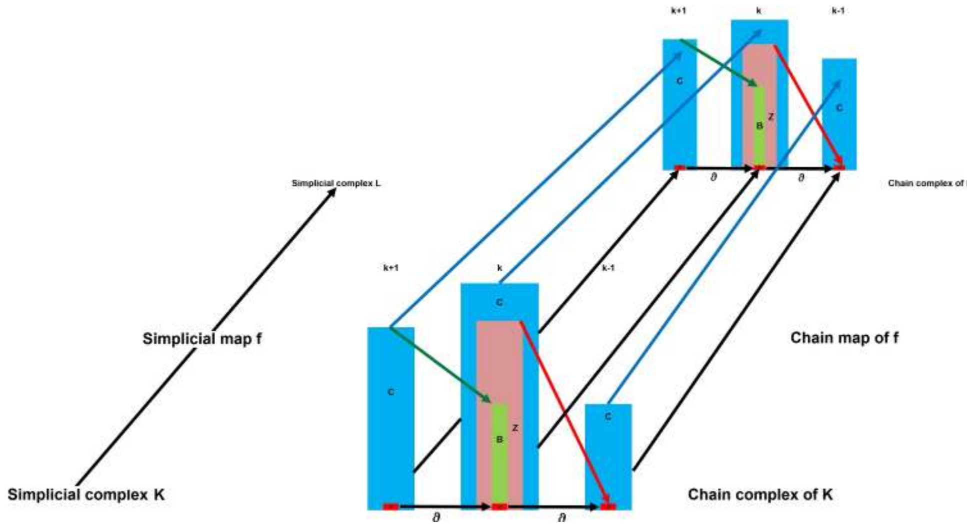
$$? : H_k(K) = \frac{Z_k(K)}{B_k(K)} \to H_k(L) = \frac{Z_k(L)}{B_k(L)},$$

where

$$\begin{aligned} Z_k(K) &:= \ker \partial_k^K, \quad B_k(K) := \operatorname{Im} \partial_{k+1}^K, \\ Z_k(L) &:= \ker \partial_k^L, \quad B_k(L) := \operatorname{Im} \partial_{k+1}^L, \end{aligned}$$

are the groups of cycles and the groups of boundaries of these complexes.

The way the new maps have linked the two chain complexes is informally illustrated by this diagram:



We need to see the interaction between these subspaces. We already know that our quotient operator is well-defined provided these conditions are met:

- (1) $f_k(Z_k(K)) \subset Z_k(L)$,
- (2) $f_k(B_k(K)) \subset B_k(L)$.

We will use the algebraic continuity condition: $\partial f = f\partial$. In particular, both squares of this diagram are commutative:

$$\begin{array}{ccccc}
C_{k+1}(K) & \xrightarrow{\partial_{k+1}} & C_k(K) & \xrightarrow{\partial_k} & C_{k-1}(K) \\
\downarrow f_{k+1} & & \downarrow f_k & & \downarrow f_{k-1} \\
C_{k+1}(L) & \xrightarrow{\partial_{k+1}} & C_k(L) & \xrightarrow{\partial_k} & C_{k-1}(L)
\end{array}$$

The proofs below are typical for algebraic topology.

Corollary 6.26. Chain maps take cycles to cycles:

$$f_k(Z_k(K)) \subset Z_k(L).$$

Proof. Suppose $x \in Z_k(K)$, then $\partial_k(x) = 0$. Then we trace x in the right square of the diagram:

$$\begin{aligned}
\partial_k f_k(x) &= f_{k-1}\partial_k(x) \\
&= f_{k-1}(0) \\
&= 0.
\end{aligned}$$

Hence $f_k(x)$ is a cycle too. ■

Corollary 6.27. Chain maps take boundaries to boundaries:

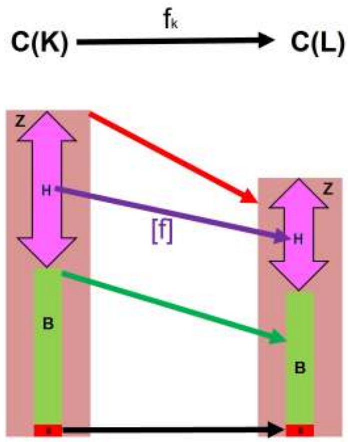
$$f_k(B_k(K)) \subset B_k(L).$$

Proof. Suppose $x \in B_k(K)$, then $x = \partial_{k+1}(u)$ for some $u \in C_{k+1}(K)$. Then we trace x back in the left square of the diagram:

$$\begin{aligned}
f_k(x) &= f_k\partial_{k+1}(u) \\
&= \partial_{k+1} f_{k+1}(u).
\end{aligned}$$

Hence $f_k(x)$ is a boundary too. ■

The two results are summarized below:



Thus, we are able to restrict f_k to cycles:

$$f_k : Z_k(K) \to Z_k(L),$$

or to boundaries:

$$f_k : B_k(K) \to B_k(L).$$

Therefore, the definition below is legitimate.

Definition 6.28. Given chain maps

$$f_k : C_k(K) \to C_k(L),\ k = 0, 1, 2, ...$$

(possibly induced by a simplicial map $f : K \to L$), the *homology maps induced by* $\{f_k\}$,

$$[f_k] : H_k(K) \to H_k(L),\ k = 0, 1, 2, ...,$$

are the linear operators given by

$$[f_k]([x]) := [f_k(x)].$$

Notation: The brackets are often omitted:

$$f_k : H_k(K) \to H_k(L),$$

or an alternative, common notation is used:

$$f_* : H_k(K) \to H_k(L).$$

Exercise 6.29. Devise a simplicial map the realization of which wraps a circle around another circle n times. Compute its homology.

6.5 Computing homology maps

Even with the most trivial examples we should pretend that we don't know the answer...

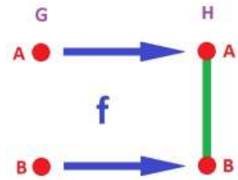
Example 6.30 (segment). We start with these simplicial complexes (graphs):

$$G = \{A, B\},\ H = \{A, B, AB\},$$

and

$$f(A) = A,\ f(B) = B.$$

Let's consider the inclusion:



Then on the chain level we have:

$$\begin{array}{lll} C_0(G) =< A, B >, & C_0(H) =< A, B >, & \\ f_0(A) = A, & f_0(B) = B & \Longrightarrow f_0 = \text{Id}; \\ C_1(G) = 0, & C_1(H) =< AB > & \\ & & \Longrightarrow f_1 = 0. \end{array}$$

Meanwhile, $\partial_0 = 0$ for G and for H we have:

$$\partial_1 = [-1, 1]^T.$$

Therefore, on the homology level we have:

$$\begin{array}{llll} H_0(H) := \frac{Z_0(H)}{B_0(H)} & = \frac{C_0(H)}{\partial_1(<AB>)} & = \frac{<A,B>}{<B-A>} & =< [A + B] >, \\ H_0(H) := \frac{Z_0(H)}{B_0(H)} & = \frac{C_0(H)}{\partial_1(<AB>)} & = \frac{<A,B>}{<B-A>} & =< [A + B] >, \\ H_1(G) := \frac{Z_1(G)}{B_1(G)} & = \frac{0}{0} & & = 0, \\ H_1(H) := \frac{Z_1(H)}{B_1(H)} & = \frac{0}{0} & & = 0. \end{array}$$

Then, by definition,

- $[f_0]([A]) := [f_0(A)] = [A], [f_0]([B]) = [f_0(B)] = [B] \Longrightarrow f_0 = [1, 1]^T$;
- $[f_1] = 0$. □

Exercise 6.31. Modify the computation for the case when there is no AB.

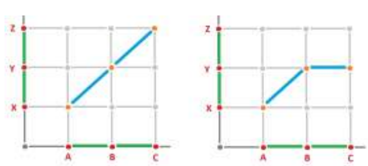
Example 6.32 (two segments). Given these two two-edge simplicial complexes:

- 1. $K = \{A, B, C, AB, BC\}$,
- 2. $L = \{X, Y, Z, XY, YZ\}$.

Consider these simplicial maps:

- 1. $f(A) = X,\ f(B) = Y,\ f(C) = Z,\ f(AB) = XY,\ f(BC) = YZ$;
- 2. $f(A) = X,\ f(B) = Y,\ f(C) = Y,\ f(AB) = XY,\ f(BC) = Y$.

They are given below:



We think of these identities as functions evaluated on the generators of the two chain groups:

- 1. $C_0(G) =< A, B, C >,\ C_1(G) =< AB, BC >$;
- 2. $C_0(H) =< X, Y, Z >,\ C_1(H) =< XY, YZ >$.

These functions induce these two linear operators:

$$f_0 : C_0(G) \to C_0(H),\ f_1 : C_1(G) \to C_1(H),$$

as follows.

The first function gives these values:

- $f_0(A) = X,\ f_0(B) = Y,\ f_0(C) = Z$,

- $f_1(AB) = XY,\ f_1(BC) = YZ$.

Those, written coordinate-wise, produce the columns of its matrices:

$$f_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathrm{Id},\ f_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathrm{Id}\,.$$

The second produces:

- $f_0(A) = X,\ f_0(B) = Y,\ f_0(C) = Y,$
- $f_1(AB) = XY,\ f_1(BC) = 0,$

which leads to:

$$f_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},\ f_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The first linear operator is the identity and the second can be thought of as a projection. □

Exercise 6.33. Modify the computation for the case of an increasing and then decreasing function.

Now homology...

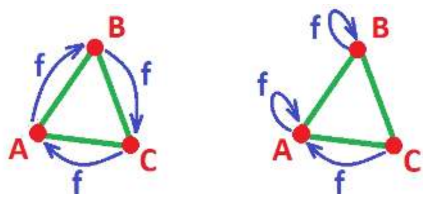
Example 6.34 (hollow triangle). Suppose:

$$G = H := \{A, B, C, AB, BC, CA\},$$

and

$$f(A) = B,\ f(B) = C,\ f(C) = A.$$

Here is a rotated triangle:



The homology maps are computed as follows:

$$Z_1(G) = Z_1(H) = < AB + BC + CA >\,.$$

Now,

$$f_1(AB + BC + CA) = f_1(AB) + f_1(BC) + f_1(CA) = BC + CA + AB = AB + BC + CA.$$

Therefore, $f_1 : Z_1(G) \to Z_1(H)$ is the identity and so is the homology map $[f_1] : H_1(G) \to H_1(H)$. Conclusion: the hole is preserved.

Alternatively, we collapse the triangle onto one of its edges:

$$f(A) = A,\ f(B) = B,\ f(C) = A.$$

Then

$$f_1(AB + BC + CA) = f_1(AB) + f_1(BC) + f_1(CA) = AB + BA + 0 = 0.$$

So, the map is zero. Conclusion: collapsing of an edge causes the hole collapse too. □

Exercise 6.35. Modify the computation for the case (a) a reflection and (b) a collapse to a vertex.

6.6 How to classify simplicial maps

From the fact that vertices are taken to vertices we derive this simple conclusion.

Proposition 6.36. If complexes K, L are edge-connected, any simplicial map between them induces the identity on the 0th homology group:

$$[f_0] = \mathrm{Id} : H_0(K) = \mathbf{R} \to H_0(L) = \mathbf{R}.$$

Therefore, not every linear operator between the homology groups can be "realized" as the homology map of a simplicial map.

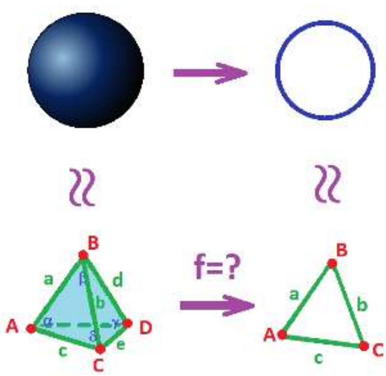
Exercise 6.37. Suppose K has n edge-components and L has m. List all possible 0th homology maps between them.

What follows is a typical homological argument.

Example 6.38. With our knowledge of the homology groups of the circle $\mathbf{S}^1$ and of the sphere $\mathbf{S}^2$:

$$\begin{array}{lll} H_0(\mathbf{S}^1) = \mathbf{R}, & H_1(\mathbf{S}^1) = \mathbf{R}, & H_2(\mathbf{S}^1) = 0; \\ H_0(\mathbf{S}^2) = \mathbf{R}, & H_1(\mathbf{S}^2) = 0, & H_2(\mathbf{S}^2) = \mathbf{R}; \end{array}$$

we can now answer the question whether it's possible to transform voids into tunnels:



It is not! The answer is understood in the sense that there is no map

$$f : \mathbf{S}^2 \to \mathbf{S}^1$$

that preserves the homology class of the void of the sphere. The reason is simple, the homomorphism

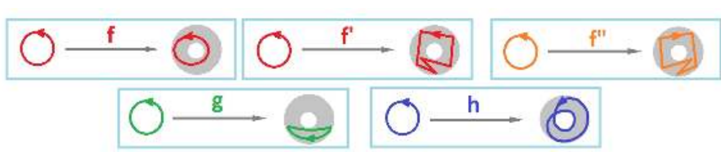
$$[f_2] : H_2(\mathbf{S}^2) = \mathbf{R} \to H_2(\mathbf{S}^1) = 0$$

can only be trivial. In particular, if s is the 2-homology class representing the void in the sphere, then $[f_2](s) = 0$. The void *must* collapse!

The answer applies to any other choice of triangulations of the two spaces. □

Exercise 6.39. What happens to the hole of the circle if it is mapped to the sphere?

Exercise 6.40. (a) Interpret these maps as maps of $\mathbf{S}^1$ to $\mathbf{S}^1$:



and represent those as realizations of simplicial maps. Hint: the last one will need special attention. (b) Confirm by computation that their homology maps are:

$$[f]=\mathrm{Id},\ [f']=-\mathrm{Id},\ [g]=0,\ [h]=2\cdot\mathrm{Id}\,.$$

(c) Make sense of this alternative answer:

$$[f]=-\mathrm{Id},\ [f']=\mathrm{Id},\ [g]=0,\ [h]=-2\cdot\mathrm{Id}\,.$$

Theorem 6.41. The homology map induced by an isomorphic chain map (and, moreover, by a simplicial isomorphism) is an isomorphism. The homology map induced by the identity chain map (and, moreover, by the identity simplicial map) is the identity operator.

Theorem 6.42. The homology map induced by an acyclic chain map, $f_k=0,\ k=1,2,...$, (and, moreover, by a constant simplicial map) is trivial; i.e., $[f_k]=0$.

Theorem 6.43. The homology map induced by the composition
- $g_kf_k : C_k(K)\to C_k(M),\ k=0,1,...,$

of chain maps
- $f_k : C_k(K)\to C_k(L),\ k=0,1,...,$ and
- $g_k : C_k(L)\to C_k(M),\ k=0,1,...,$

(and, moreover, by the composition of simplicial maps f,g) is the composition of the homology maps induced by these chain maps:

$$[g_kf_k]=[g_k][f_k],$$

and, moreover:

$$[(gf)_k]=[g_k][f_k].$$

Exercise 6.44. Prove these theorems.

6.7 Realizations

We already know that an abstract simplicial complex K can be realized as a topological space. The way to construct it is by treating the list of vertices and simplices of K as a blueprint of a *geometric* complex $|K|\subset\mathbf{R}^N$. It is built from the *geometric* n-simplices (corresponding to the simplices of K), i.e., the convex hulls of $n+1$ points

$$A_0A_1...A_n=\mathrm{conv}\{A_0,A_1,...,A_n\}$$

in general position. In the simplex, every point x is represented as a convex combination of the vertices:

$$x=\sum_i s_iA_i,$$

with the barycentric coordinates of x serving as coefficients.

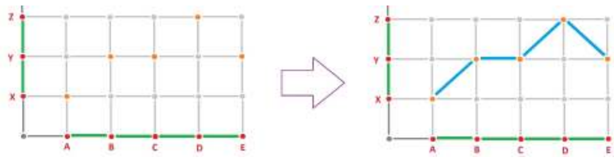
Now we want to learn how to realize simplicial maps – as continuous functions!

Of course, we want $f:K\to L$ to be realized as $|f|:|K|\to|L|$.

The 1-dimensional case is the case of graphs. Let's review what we know about realizations of maps of graphs. Suppose two complexes are realized in $\mathbf{R}$ and suppose we are provided with only the values of the vertices of the simplicial map:
- $f(A)=X,\ f(B)=Y,\ f(C)=Y,\ f(D)=Z,\ f(E)=Y.$

Consequently, we have
- $|f|(A)=X,\ |f|(B)=Y,\ |f|(C)=Y,\ |f|(D)=Z,\ |f|(E)=Y.$

These points are plotted in orange below:

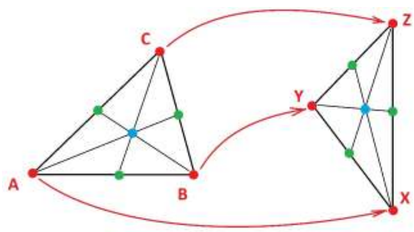


We reconstruct the rest of the map by connecting these points by straight segments. The formulas are familiar, such as this:

$$|f|(tA+(1-t)B)=tX+(1-t)Y,\ t\in[0,1].$$

The idea becomes clear: *we use the barycentric coordinates of each point in a simplex as the barycentric coordinates of the image of that point.*

It is similar in dimension 2:

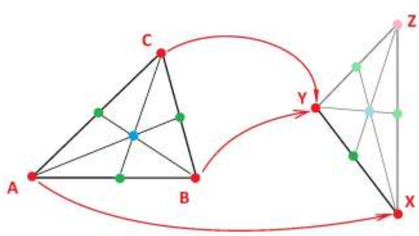


In particular, we can see:

$$\begin{aligned}|f|\left(\tfrac{1}{2}A+\tfrac{1}{2}B\right) &= \tfrac{1}{2}X+\tfrac{1}{2}Y,\\ |f|\left(\tfrac{1}{3}A+\tfrac{1}{3}B+\tfrac{1}{3}C\right) &= \tfrac{1}{3}X+\tfrac{1}{3}Y+\tfrac{1}{3}Z.\end{aligned}$$

Exercise 6.45. Show that this approach to realizing simplicial maps by linearity doesn't work for cubical complexes because the cubical analog of the barycentric coordinates doesn't provide a unique representation of a point in the square. Hint: the 4 vertices of a square can't be in general position.

Collapses are realized the same way – linearly – for simplicial complexes:



In particular, consider:

$$|f|\left(\tfrac{1}{3}A+\tfrac{1}{3}B+\tfrac{1}{3}C\right)=\tfrac{1}{3}X+\tfrac{1}{3}Y+\tfrac{1}{3}Y=\tfrac{1}{3}X+\tfrac{2}{3}Y.$$

Exercise 6.46. Plot the images of the green points and the blue point.

This approach won't work for cubical complexes. As an example, a function $g:[0,1]\times[0,1]\to[0,1]$ with

$$g(0,0)=g(1,1)=0,\ g(1,0)=g(0,1)=1,$$

can't be linear.

Exercise 6.47. Prove it.

Now the general case.

We start with a single simplex. Given a simplicial map between two (abstract) simplices

$$f : \sigma \to \tau,$$

here's how we define a map that will serve as its realization

$$|f| : |\sigma| \to |\tau|.$$

Suppose the dimensions are n and $m \leq n$ respectively and

$$\sigma = A_0...A_n, \ \tau = B_0...B_m.$$

Suppose $A_0, ..., A_n \in \mathbf{R}^N$ and $B_0, ..., B_m \in \mathbf{R}^N$ also denote the vertices in the realizations of these simplices $|\sigma|$ and $|\tau|$. Now suppose a point $x \in |\sigma|$ is given by its barycentric coordinates:

$$x = \sum_i s_i A_i.$$

The coefficients are unique. The map f is already realized on vertices:

$$|f|(A_i) = B_{j_i}, \ i = 0, ..., n.$$

Definition 6.48. Given complexes K and L and a simplicial map $f : K \to L$, we have a map

$$g = |f| : |K| \to |L|$$

constructed on vertices, as described above. Then a *geometric realization* of a simplicial maps f is an piece-wise extension defined by

$$|f|(x) := \sum_i s_i |f|(A_i).$$

Some of the terms $|f|(A_i)$ here may be equal but this is still a convex hull combination of $B_0...B_m$. We don't claim then (unless σ is cloned) that the coefficients s_i are the barycentric coordinates of $|f|(x)$ in $|\tau|$, but rather in a face of $|\tau|$.

Exercise 6.49. Prove that $|f|(x) \in |\tau|$.

The new map is *linear* on each simplex as

$$|f|(x) = |f|\Big(\sum_i s_i A_i\Big) = \sum_i s_i |f|(A_i).$$

Exercise 6.50. Describe how this map can be constructed skeleton by skeleton.

Theorem 6.51. A realization of a simplicial map is well-defined and continuous.

Exercise 6.52. Prove the theorem. Hint: show that if two simplices share a face, the two formulas for the function restricted to this face match.

In the first example above, the 2-simplex is cloned and in the second, it is collapsed. The data structure of abstract simplicial complexes and maps and the geometry of their realizations match, as follows:

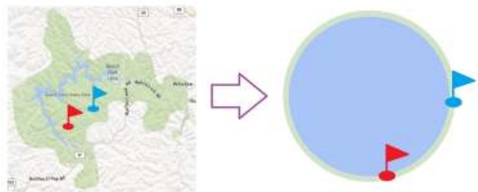
Proposition 6.53. Under a simplicial map $f : K \to L$, a simplex $s \in K$ is cloned if and only if its realization $|s|$ is mapped homeomorphically under a realization $|f|$ of f.

The topological issues, as opposed to the geometrical issues, of the realizations of simplicial maps will be discussed later under *cell maps*.

6.8 Social choice: no compromise

Let's recall the problem of social choice for two hikers.

We need to develop a procedure for finding a fair compromise on the location of the camp on the shore of a lake:



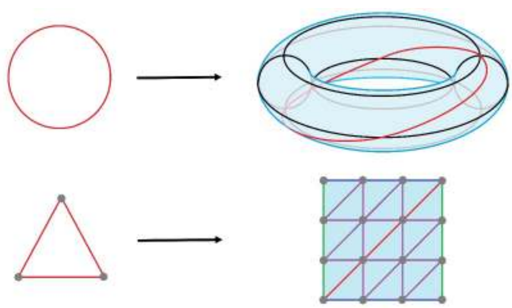
Social choice problem: is there a map

$$f : \mathbf{S}^1 \times \mathbf{S}^1 \to \mathbf{S}^1$$

that satisfies these two conditions:
- 1. $f(x, y) = f(y, x)$, and
- 2. $f(x, x) = x$?

We already know that some obvious solutions fail.

We recognize the product of two circles $\mathbf{S}^1 \times \mathbf{S}^1$ as simply the torus and its diagonal as just another circle:



For now, we will limit ourselves to the *simplicial case*:
- the circle is triangulated;
- the torus is triangulated;
- its *diagonal* Δ of the torus is triangulated identically to the circle;
- the *diagonal map* $\delta : \mathbf{S}^1 \to \Delta$ given by $\delta(x) = (x, x)$ is the identity simplicial map;
- the choice function is a simplicial map.

Let's translate this topological problem into an algebraic one via *homology.*

We know that f, if it exists, induces a homomorphism

$$[f_1] : H_1(\mathbf{S}^1) \to H_1(\mathbf{S}^1)$$

on the homology groups:

$$H_1(\mathbf{S}^1) = \mathbf{Z},\ H_1(\mathbf{S}^1 \times \mathbf{S}^1) = \mathbf{Z} \oplus \mathbf{Z}.$$

The problem then becomes: is there a homomorphism

$$g : \mathbf{Z} \oplus \mathbf{Z} \to \mathbf{Z},$$

whether it comes from some simplicial map f or not, that satisfies

- 1. $g(x,y) = g(y,x)$, and
- 2. $g(x,x) = x$?

Exercise 6.54. Show that these two conditions indeed follow from the two conditions above.

Exercise 6.55. Restate and solve the problem for the case of real coefficients.

Let's restate these two problems in terms of commutative diagrams.

Now the problem becomes, is it possible to complete this commutative diagram of simplicial complexes and simplicial maps, with some f?

$$\begin{array}{ccc} \mathbf{S}^1 & \xrightarrow{\ \delta\ } & \mathbf{S}^1 \times \mathbf{S}^1 \\ & {}^{\text{Id}}\searrow & \downarrow{\scriptstyle f=?} \\ & & \mathbf{S}^1 \end{array}$$

The diagonal arrow is the identity according to the second condition. Applying homology to the above diagram yields a commutative diagram of groups and homomorphisms:

$$\begin{array}{ccc} \mathbf{Z} & \xrightarrow{\ \delta_*=[1,1]^T\ } & \mathbf{Z} \oplus \mathbf{Z} \\ & {}^{\text{Id}}\searrow & \downarrow{\scriptstyle g=?} \\ & & \mathbf{Z} \end{array}$$

The problem becomes, is it possible to complete this diagram, with some g?

Now we use the fact that g has to be symmetric. Let's choose a specific generator of the 1st homology group:

$$H_1(\mathbf{S}^1) = \mathbf{Z} =< 1 >$$

and identify its value under g:

$$c := g(1,0) \in H_1(\mathbf{S}^1) = \mathbf{Z}.$$

Then

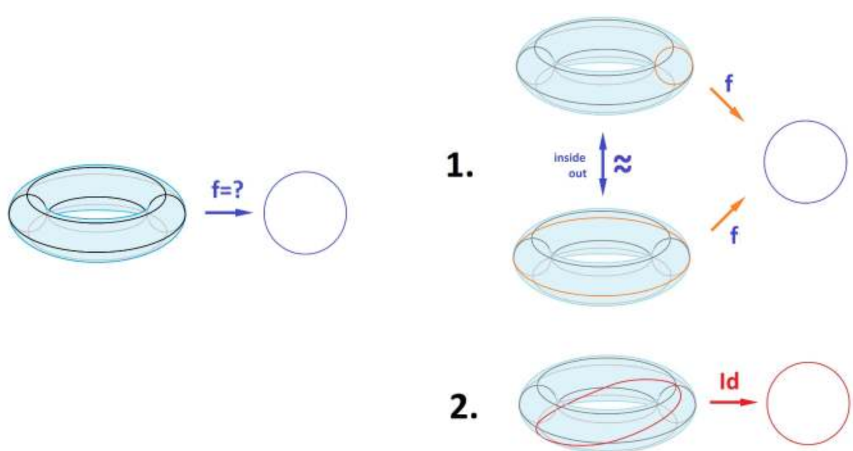
$$1 = \text{Id}(1) = g[\delta_1](1) = g(1,1) = g(1,0) + g(0,1) = 2g(1,0) = 2c.$$

A contradiction!

We have proven that there is *no fair compromise* for our social choice problem.

Exercise 6.56. Compare the outcome to that for homology over $\mathbf{R}$. What about $\mathbf{Z}_2$?

The mathematical version of the problem and the two conditions are illustrated below:



Exercise 6.57. State and solve homologically the problem of m hikers and the forest of n convex path-components.

Exercise 6.58. What if the two hikers are also to choose the time of the day to leave? Describe and analyze the new set of possible choices. What if they are also to choose the way to hang their hammock in the camp?

Exercise 6.59. Suppose another pair of hikers has stopped in the middle of the forest and has to decide on the direction to follow. Discuss.

It is a common sense observation that excluding some options might make compromise impossible. Furthermore, a compromise-producing rule might also become impossible. We will see broader results later.

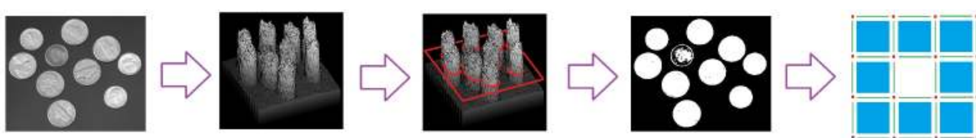
The fact that the analysis is limited to *simplicial maps* seems very unfortunate. Of course we should be able to freely choose the values for f from the continuum of locations on the shore of the lake! Later, we will extend the homology theory to topological spaces and maps.

---

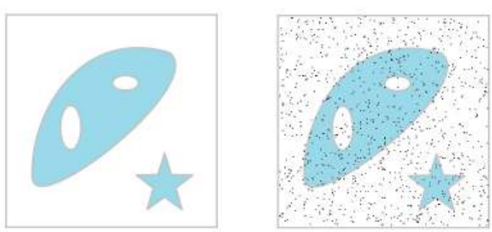
7 Parametric complexes

7.1 Topology under uncertainty

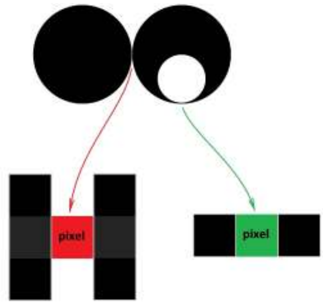
So far, homology has been used to describe the topology of abstract objects. As a result, all homology classes receive equal attention. Meanwhile, converting such real-life data as digital images into simplicial or cubical complexes requires extra steps that may lead to uncertainty and a loss of information.



Above, we see how subtle variations of gray may wind up discarded during such a procedure. In addition, the topology of digital images and other data may be "corrupted" by noise:



The so-called "salt-and-pepper" noise is normally dealt with by removing the path-components that are smaller than, say, a couple of pixels. The implicit assumption for this to work is that pixels are *really* small... And yet, misplacing just one could dramatically change the topology of the image:



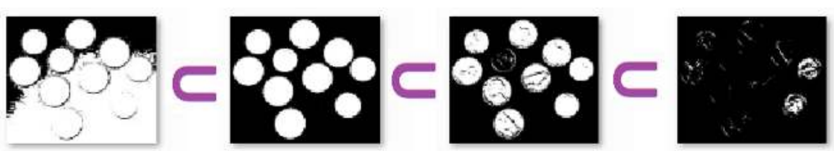
The example shows that adding the red pixel merges two components while adding the green one creates a hole.

We will learn how to rank homology classes according to their relative importance, called *persistence*, and then attempt to discard some of them as noise.

Since it's not known beforehand what is or is not noise in the dataset, we need to capture all homology classes including those that may be deemed noise later. We will introduce an algebraic structure that contains, without duplication, all these classes. Each of them is associated with its persistence and can be removed when the threshold for acceptable noise is set.

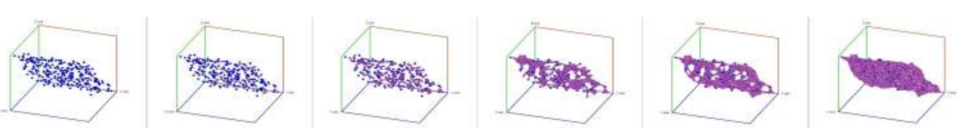
Let's consider the two basic real-life sources of datasets.

Example 7.1 (gray scale images). A *gray scale image* is a real-valued function $f : R \to \mathbf{R}_+$ defined on a rectangle Q. Then, given a threshold r, its lower level set $f^{-1}((-\infty, r))$ can be thought of as a binary image on Q. Each black pixel of this image is treated as a square cell in the plane. This process is called *thresholding*, which is also applicable to the n-dimensional case. Naturally, these 2-dimensional cells are combined with their edges (1-cells) and vertices (0-cells). The result is a cubical complex K for each r.



$\square$

Example 7.2 (point clouds). A *point cloud* is a finite set S in some Euclidean space of dimension d. Given a threshold r, we deem any two points that lie within r from each other to be "close". In that case, this pair of points is connected by an edge. Further, if three points are "close" to each other – in the sense that the diameter of this triangle is less than r – we add a face spanned by these points. If there are four, we add a tetrahedron, and, finally, any $d + 1$ "close" points create a d-cell. The process is called the *Vietoris-Rips construction*. The result is a simplicial complex K for each r.



$\square$

Next, we would like to quantify the topology of what's behind this data – with homology groups. Instead of using a single threshold and studying a single cell complex, one considers all thresholds and the corresponding complexes. Since increasing threshold r enlarges the corresponding complex, we have a sequence of complexes:

$$K^1 \hookrightarrow K^2 \hookrightarrow K^3 \hookrightarrow ... \hookrightarrow K^s,$$

with the arrows representing the inclusions:

$$i^n : K^n \hookrightarrow K^{n+1}.$$

This sequence **denoted** by

$$\{K^n\} = \{K^n, i^n : \ n = 1, 2, 3, ..., s-1\}$$

is called a *filtration*.

Next, each of the inclusions generates its homology map:

$$i^n_* : H(K^n) \to H(K^{n+1}).$$

As a result, we have a sequence of homology groups linked by these homomorphisms:

$$H(K^1) \to H(K^2) \to \ ... \ \to H(K^s) \to 0,$$

with 0 added for convenience. These homomorphisms record how the homology changes (or doesn't) as this "parametric" space grows at each step. For example, a component appears, grows, and then merges with another one; or a hole forms, shrinks, and then is filled. In either case, we think of these are the same homology class and we refer to these events as its *birth* and *death*.

In order to evaluate the relative importance of an element of one of these groups, the *persistence of a homology class* is defined as the number of steps in the homology sequence it takes for the class to end at 0. In other words,

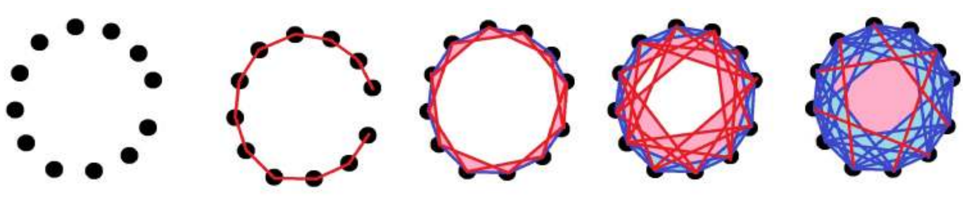
$$persistence = death - birth.$$

Indirectly, this number reflects the robustness of the homology class with respect to the changing parameter.

Exercise 7.3. Describe the homology of the filtrations in the above examples.

7.2 Persistence of homology classes

While constructing a simplicial complex from a point cloud, we gradually add cells of various dimensions in hope of being able to discover the topology of the object. For example, what comes out of a point cloud of a circle might look like this:

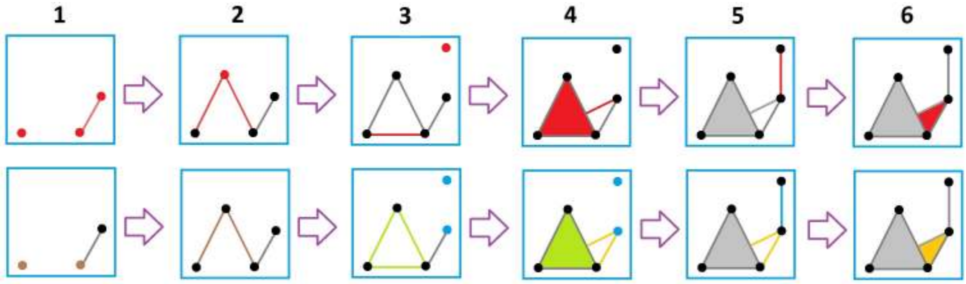


But how do we know which one of these complexes is the "true" representation of the circle without actually looking at it? In fact, how would a computer see a circle behind this data?

A simpler question is, how can we tell which cycle in which of these complexes captures the hole of the circle, without us looking at it to confirm?

To try to answer this question, we look at the *whole* sequence and how it develops. The idea is to measure and compare the "life-spans" of these cycles.

Example 7.4 (homology of filtration). Let's take a simpler example and provide complete computations:



The top row is the sequence of the complexes of the filtration, which looks like a sequence of frames in a movie. In each of these "frames", the new cells are marked in red. In the bottom row, the new cycles are given in various colors:

- The brown 0-cycle appears in the 1st frame and dies in the 2nd.
- The blue 0-cycle appears in the 3rd frame and dies in the 5th.
- The green 1-cycle appears in the 3rd frame and dies in the 4th.
- The orange 1-cycle appears in the 4th frame and dies in the 6th.

The importance of a homology class is then evaluated by the length of its life-span: how long it takes for the class to disappear as it is mapped to the next frame. Algebraically, the persistence is equal to how many applications of these homology maps it takes for the homology class to become 0.

The values are given below.

- The brown 0-class appears in the 1st frame and dies in the 2nd: persistence 1.
- The blue 0-class appears in the 3rd frame and dies in the 5th: persistence 2.
- The green 1-class appears in the 3rd frame and dies in the 4th: persistence 1.
- The orange 1-class appears in the 4th frame and dies in the 6th: persistence 2.

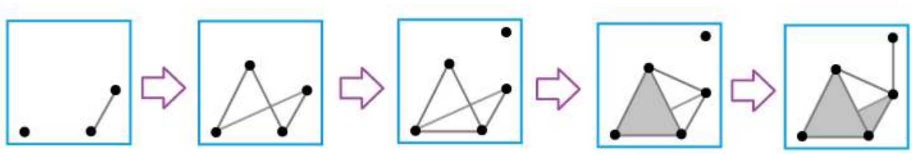
Now, in order to settle on a particular topology, we choose a *threshold for persistence.* If the choice is 1, we keep all these cycles: two 0-cycles and two 1-cycles. If, however, the choice is 2, there is only one of each. Which of the two is the "correct" one remains a judgement call.

Instead of looking at one class at a time, we list the homology groups (over $\mathbf{R}$) and the homology maps of the whole filtration:

$$\begin{array}{lccccccccccc}
 & H_n(K^1) & \xrightarrow{i_*^1} & H_n(K^2) & \xrightarrow{i_*^2} & H_n(K^3) & \xrightarrow{i_*^3} & H_n(K^4) & \xrightarrow{i_*^4} & H_n(K^5) & \longrightarrow 0 \\
n=0: & \mathbf{R}\oplus\mathbf{R} & \xrightarrow{[1,1]} & \mathbf{R} & \xrightarrow{[1,0]^T} & \mathbf{R}\oplus\mathbf{R} & \xrightarrow{\text{Id}} & \mathbf{R}\oplus\mathbf{R} & \xrightarrow{[1,1]} & \mathbf{R} & \xrightarrow{0} 0 \\
n=1: & 0 & \xrightarrow{0} & 0 & \xrightarrow{0} & \mathbf{R} & \xrightarrow{[0,1]^T} & \mathbf{R} & \xrightarrow{\text{Id}} & \mathbf{R} & \xrightarrow{0} 0
\end{array}$$
□

Exercise 7.5. Find the compositions of the homology maps of the inclusions and compare them to the values of the persistence.

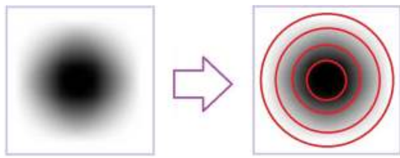
Exercise 7.6. Find the homology maps of the inclusions of this filtration:



Hint:

$$\begin{array}{lccccccccccc}
\dim = 0: & \mathbf{R}\oplus\mathbf{R} & \to & \mathbf{R} & \to & \mathbf{R}\oplus\mathbf{R} & \to & \mathbf{R}\oplus\mathbf{R} & \to & \mathbf{R} & \to & 0\\
\dim = 1: & 0 & \to & ? & \to & ? & \to & ? & \to & ? & \to & 0
\end{array}$$

Example 7.7 (blurring). Let's analyze this image of a blurred disk:

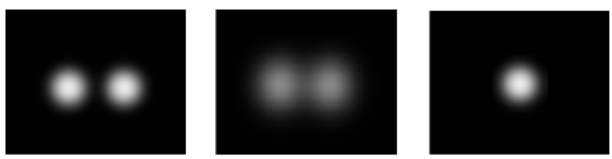


We use all 255 levels of gray as thresholds. Then each of the complexes is a disk and the homology groups of the filtration are very simple:

$$\dim = 0: \quad \mathbf{R} \ \to \ \mathbf{R} \ \to \ \mathbf{R} \ \to \ ... \ \to \ \mathbf{R} \ \to \ 0.$$

Moreover, the inclusions generate the identity maps on homology: $i_*^n = \mathrm{Id}$. So, if we assume that the center of the circle is pure black (value 0) and the outside is pure white (value 255), then the persistence of the smallest circle is 255. □

Exercise 7.8. A car or a motorcycle? Use thresholding to create a filtration, find its homology groups and the persistence of all homology classes for each of these images of headlights:



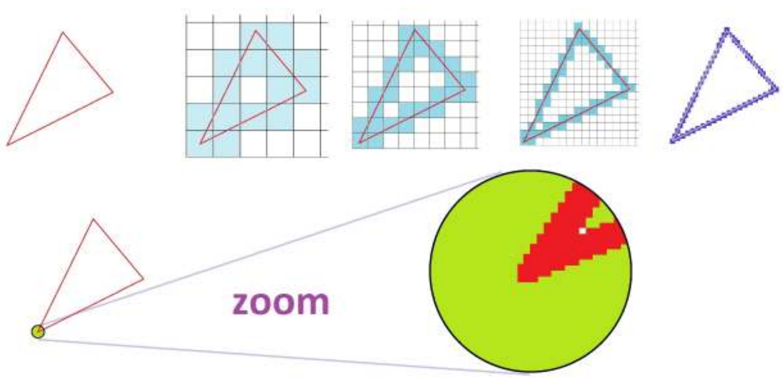
Let's consider, in contrast, how filtrations and persistence appears in the case of *binary* images.

Suppose a binary picture of a real-life object is taken with a higher and higher resolution. The result is a sequence of cubical complexes. At step n, one assigns a given pixel to the complex whenever the image of the object crosses it. As a result we have a *decreasing* sequence of complexes:

$$K^s \supset K^{s-1} \supset ... \supset K^1.$$

It is natural to think that each topological feature in the space being discretized is represented in at least one of the elements of the filtration (frame). The example below shows that the converse isn't always true.

Example 7.9 (artifacts). Here we discretize an ideal triangle (left) via grids with smaller and smaller squares:



In each case, the resulting binary image has the same problem: there is an extra, one-pixel hole that's not present in the original!

Or is it? The last image on the right shows the left bottom corner of the *original* zoomed in, and a disconnected pixel (hole) is visible. One can reproduce this effect with MS Paint or other similar software.

Thus, improving the resolution won't eliminate the spurious hole.

Now, these four images – going from right to left – form a filtration:

$$\begin{array}{rccccccccc}
 & H_1(K^1) & \xrightarrow{i^1_*} & H_1(K^2) & \xrightarrow{i^2_*} & H_1(K^3) & \xrightarrow{i^3_*} & H_1(K^4) & \longrightarrow & 0 \\
n=1: & \mathbf{R}\oplus\mathbf{R} & \xrightarrow{\mathrm{Id}\oplus 0} & \mathbf{R}\oplus\mathbf{R} & \xrightarrow{\mathrm{Id}\oplus 0} & \mathbf{R}\oplus\mathbf{R} & \xrightarrow{\mathrm{Id}\oplus 0} & \mathbf{R} & \xrightarrow{0} & 0
\end{array}$$

As the homology groups are identical, the extra hole might appear – incorrectly – persistent. Examining the homology maps reveals that, at every step, the homology class of the spurious hole goes to zero under the homology map of the inclusion. Therefore, the persistence of each of these holes is equal to the lowest possible value of 1! □

Exercise 7.10. What geometric figure on the plane will always have, under this kind of discretization, an extra component – no matter how much the resolution is improved?

Now, the algebra behind the idea of persistence of a homology class of the ith complex K^i – as a part of the filtration – becomes clear. The non-persistent ones go to 0. Then, together, they form the *kernel* of the homology map of the inclusion. The persistent ones are "what's left". Initially, we can think of this part as the orthogonal complement of the kernel. So,

- the 0-persistent homology group $H^0_n(K^i)$ of complex K^i is simply its homology group;
- the 1-persistent homology group $H^1_n(K^i)$ is the orthogonal complement of the kernel of the homology map of the inclusion of the complex into the next one;
- the 2-persistent homology group $H^2_n(K^i)$ is the orthogonal complement of the kernel of the homology map of the composition of the two inclusions;
- etc.

To be sure, the orthogonal complement can be understood as a quotient.

Definition 7.11. The *p-persistent homology group* of the ith element K^i of filtration $\{K^n\}$ is defined to be

$$H^p(K^i) := H(K^i)/\ker i^{i,i+p}_*,$$

where $i^{i,i+p} : K^i \to K^{i+p}$ is the inclusion.

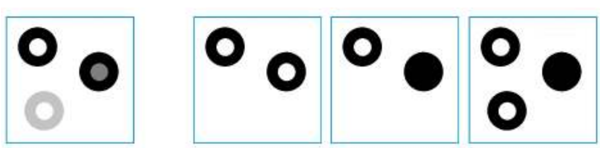
Exercise 7.12. Compute the persistent homology groups of each of the filtrations depicted in the subsection.

Can we combine these groups $H^p_n(K^1), H^p_n(K^2), ..., H^p_n(K^s)$ into one?

7.3 The homology of a gray scale image

Putting the persistence issue aside for now, we will try to understand the meaning of the homology of a gray scale image, taken "as is".

For a complete analysis, let's take this simple image on the left:



We assume that only these four colors – white, light gray, dark gray, and black – are in play.

First the image is thresholded. The lower level sets of the gray scale function of the image form a filtration: a sequence of three binary images, i.e., cubical complexes:

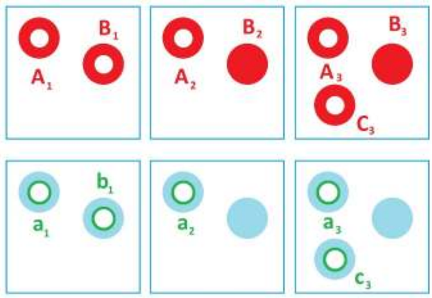
$$K^1 \hookrightarrow K^2 \hookrightarrow K^3,$$

with the arrows representing the inclusions i^1, i^2. Here, white means empty. For completeness sake, one can add a fully white image in the beginning and fully black in the end of the sequence.

We go clockwise starting at the upper left corner and use the following notation, for $i = 1, 2, 3$:

- A_i, B_i, C_i are the 0-homology classes that represent the components of K^i; and
- a_i, b_i, c_i are the 1-homology classes that represent the holes.

The homology classes of these images also form sequences – one for either dimension 0 and 1.



Then i^1_*, i^2_* are the two homology maps of the inclusions while $i^3_* = 0$ is included for convenience. These homomorphisms map the generators, as follows:

$$\begin{array}{ccccccc}
 & i^1_* & & i^2_* & & i^3_* & \\
A_1 & \mapsto & A_2 & \mapsto & A_3 & \mapsto & 0 \\
B_1 & \mapsto & B_2 & \mapsto & B_3 & \mapsto & 0 \\
 & & & & C_3 & \mapsto & 0 \\
a_1 & \mapsto & a_2 & \mapsto & a_3 & \mapsto & 0 \\
b_1 & \mapsto & 0 & & & & \\
 & & & & c_3 & \mapsto & 0
\end{array}$$

In order to avoid double counting, we need to count only the homology generators that don't reappear in the next homology group. We adopt the opposite but a more algebraically convenient approach: capture the homology classes taken to 0 by these homomorphisms.

Finally, this is the simplest way to combine these classes:

$$\begin{array}{ll}
H_0(\{K^i\}) & =< A_3, B_3, C_3 >, \\
H_1(\{K^i\}) & =< b_1, a_3, c_3 > .
\end{array}$$

These groups would serve as the homology group of the original, gray-scale image. The data shows that image has three components and three holes, as expected.

These classes generate the kernels of $i^1_*,\ i^2_*,\ i^3_*$. Then, the 0th and 1st homology groups of the frames of the gray scale image,

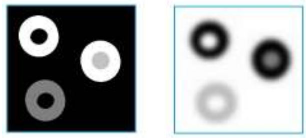
$$\begin{array}{ccccccc}
\mathbf{R}^2 & \to & \mathbf{R}^2 & \to & \mathbf{R}^3 & \to & 0, \\
\mathbf{R}^2 & \to & \mathbf{R}^1 & \to & \mathbf{R}^2 & \to & 0,
\end{array}$$

produce its homology groups:

$$\begin{array}{llll}
H_0(\{K^i\}) & = \ker[i^1_0] \oplus \ker[i^2_0] \oplus \ker[i^3_0] & = 0 \oplus 0 \oplus \mathbf{R}^3 & = \mathbf{R}^3, \\
H_1(\{K^i\}) & = \ker[i^1_1] \oplus \ker[i^2_1] \oplus \ker[i^3_1] & = \mathbf{R} \oplus 0 \oplus \mathbf{R}^2 & = \mathbf{R}^3.
\end{array}$$

Our conclusion is: *the homology group of a gray scale image is the direct sum of the kernels of the homology maps of the inclusions of its frames.*

Exercise 7.13. Compute the homology group of (a) the negative and (b) the blurred version of the above image:



This approach is simple but it has a drawback. Consider this sequence of homology classes

$$a_1 \mapsto a_2 \mapsto a_3 \mapsto 0.$$

They are meant to represent the same feature of the original image, but the one that we chose to use in the homology of the image defined above is the last one, a_3. But this class represents the *least* prominent feature of the three! This issue becomes more obvious if we consider a blurred version of the image with all 256 levels of gray present. Then we have a filtration of 256 complexes and a sequence of 256 homology classes:

$$a_1 \mapsto a_2 \mapsto ... \mapsto a_{255} \mapsto a_{256} \mapsto 0.$$

The last non-zero element is to appear in $H(\{K^i\})$ and its contrast is just 1. Such a contrast is simply invisible to the human eye. The best choice to capture the hole would be the class with the *highest* contrast, i.e., a_1.

What's behind the contrast is the *life-span*: the shortest for a_{256} and the longest for a_1. Following this analogy, instead of capturing the classes about to die, this time we concentrate on the ones that are just born. Then we have the same homology groups but with better generators:

$$H(\{K^i\}) =< A_1, B_1, C_2, a_1, b_1, c_3 > .$$

While the idea of "death" is straightforward (the class goes to 0), that of "birth" isn't. Certainly we need to exclude ones that are already alive, i.e., the ones present in the last complex. Those form the *image* of i^1_*! One can think of those left as the orthogonal complement in the linear algebra environment, $(\operatorname{Im} i^n_*)^\perp$, or we can exclude the image via quotient: $H(K^n)/\operatorname{Im} i^n_*$.

Special kinds of quotients like these appear in group theory. For a given homomorphism $f : G \to H$ of abelian groups, define the *cokernel* of f by

$$\operatorname{coker} f = H/\operatorname{Im} f.$$

Proposition 7.14. The following are exact sequences, with inclusions and projections unmarked:

$$0 \to \operatorname{Im} f \to H \to \operatorname{coker} f \to 0,$$

and

$$0 \to \ker f \to G \xrightarrow{\quad f \quad} H \to \operatorname{coker} f \to 0.$$

Exercise 7.15. Prove the proposition.

As before, homology generates a system of groups and homomorphisms:

$$0 \xrightarrow{i^0_*} H(K^1) \xrightarrow{i^1_*} H(K^2) \xrightarrow{i^2_*} ... \xrightarrow{i^{s-1}_*} H(K^s),$$

but this time we add 0 in the beginning, for convenience, with $i^0_* = 0$.

Thus, the alternative approach is: the *k-homology group of the filtration* $\{K^n\}$ is the product of the cokernels of the inclusions:

$$H(\{K^n\}) := \operatorname{coker} i_*^0 \oplus \operatorname{coker} i_*^1 \oplus ... \oplus \operatorname{coker} i_*^{s-1}.$$

Exercise 7.16. Prove that the two definitions produce isomorphic vector spaces for homology over $\mathbf{R}$.

7.4 Homology groups of filtrations

We have learned how to capture homology classes of a gray-scale image – represented by a filtration – without double counting. We now apply this approach to an arbitrary filtration:

$$K^1 \hookrightarrow K^2 \hookrightarrow K^3 \hookrightarrow ... \hookrightarrow K^s.$$

denoted by $\{K^n\}$. Here $K^1, K^2, ..., K^s$ are simplicial or cubical complexes and the arrows represent the inclusions $i^n : K^n \hookrightarrow K^{n+1}$. Next, homology generates a *direct system* of groups and homomorphisms:

$$H(K^1) \to H(K^2) \to ... \to H(K^s) \to 0.$$

We denote this direct system by

$$\{H(K^n)\} = \{H(K^n), i_*^n : \ n = 1, 2, ..., s\}.$$

The zero is added in the end for convenience.

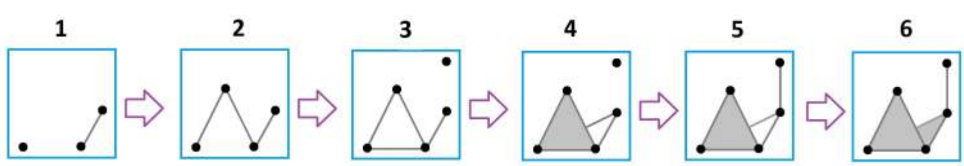
Our goal is to define a single structure that captures all homology classes in the whole filtration without double counting. We choose to concentrate on the kernels of the homology maps of the inclusions. The rationale is that if $x \in H(K^n)$, $y \in H(K^{n+1})$, $y = i_*^n(x)$, and there is no other x satisfying this condition, then x and y may be thought of as representing the same homology class of the geometric figure behind the filtration.

Definition 7.17. The *homology group of filtration* $\{K^n\}$ is defined as the direct sum of the kernels of the inclusions:

$$H(\{K^n\}) := \ker i_*^1 \oplus \ker i_*^2 \oplus ... \oplus \ker i_*^s.$$

Here, from each group, we take only the elements that are about to die. Since each dies only once, there is no double-counting. Since everyone will die eventually, every homology class is accounted for. Therefore, each appears once and only once.

Example 7.18. Let's consider this filtration:



Following the definition, its homology group is

	$\ker i_*^1$	$\oplus$	$\ker i_*^2$	$\oplus$	$\ker i_*^3$	$\oplus$	$\ker i_*^4$	$\oplus$	$\ker i_*^5$	$\oplus$	$\ker i_*^6$
$H_0 =$	$\mathbf{R}$	$\oplus$	0	$\oplus$	0	$\oplus$	$\mathbf{R}$	$\oplus$	0	$\oplus$	$\mathbf{R}$
$H_1 =$	0	$\oplus$	0	$\oplus$	$\mathbf{R}$	$\oplus$	0	$\oplus$	$\mathbf{R}$	$\oplus$	0

□

Exercise 7.19. Compute the homology groups for the filtrations from the previous subsections, produced by: (a) the Vietoris-Rips construction of the circle, and (b) the increasing resolution of the triangle.

Let's state a few simple facts about this group.

Proposition 7.20. If i^n_* is an isomorphism for each $n = 1, 2, ..., s-1$, then

$$H(\{K^n\}) = H(K^1).$$

Proposition 7.21. If i^n_* is a monomorphism for each $n = 1, 2, ..., s-1$, then

$$H(\{K^n\}) = H(K^s).$$

Proposition 7.22. Suppose $\{K^n, i^n\}$ and $\{L^n, j^n\}$ are filtrations. Then,

$$H(\{K^n \sqcup L^n\}) = H(\{K^n\}) \oplus H(\{L^n\}).$$

Exercise 7.23. Prove these propositions.

Exercise 7.24. Illustrate the propositions with examples of (a) thresholding of gray scale images, and (b) the Vietoris-Rips construction for point clouds.

Note: Any direct system of complexes – not just a filtration – will generate a direct system of groups. The above definition of the homology group will still make sense.

7.5 Maps of filtrations

What about maps? Suppose

$$\{K^n, i^n : \ n = 1, 2, ..., s\} \text{ and } \{L^n, j^n : \ n = 1, 2, ..., s\}$$

are filtrations of complexes, with the same number of elements. Suppose also we have a sequence of simplicial (or cubical) maps

$$f^n : K^n \to L^n, \ n = 1, 2, ..., s.$$

Since each K^n is a subset of K^s, each f^n is simply the restriction of f^s to K^n:

$$f^n = f^s \Big|_{K^n}.$$

They are well-defined as long as

$$f^s(K^n) \subset L^n.$$

Then such a sequence is called a *filtration map*. We denote this map by

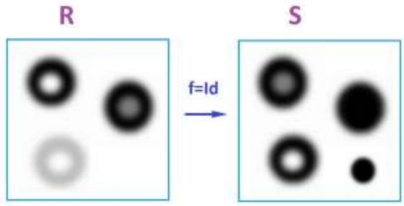
$$\{f^n\} : \{K^n\} \to \{L^n\}.$$

Example 7.25. What kind of function between two gray scale images generates a filtration map? Let's consider an arbitrary function $f : Q \to S$ from the rectangle Q of the first image to the rectangle S of the second. Is this a filtration map? We know that

$$K^n = p^{-1}(r_n, \infty), \ L^n = q^{-1}(s_n, \infty),$$

where $p : Q \to \mathbf{R}, q : S \to \mathbf{R}$ are the gray scale functions of the two images and $r_n, s_n \in \mathbf{R}$ are some thresholds. Then the condition $f(K^n) \subset L^n$ means that $qf(p^{-1}(r_n)) \geq s_n$.

If we are dealing with actual digital images, the thresholds for both are the same: $r_n = s_n = n = 0, 1, ..., 256$. In that case, the last condition means that f must make pixels *darker*:



□

Exercise 7.26. What kind of function between two point clouds generates a filtration map under the Vietoris-Rips construction?

Alternatively and equivalently, we require that these maps $f^n : K^n \to L^n$ commute with the inclusions:

$$j^n f^n = f^{n+1} i^n, \ n = 1, 2, ..., s-1.$$

In other words, the diagram is commutative:

$$\begin{array}{ccccccc}
... & \to & K^n & \xrightarrow{i^n} & K^{n+1} & \to & ... \\
... & & \downarrow f^n & & \downarrow f^{n+1} & & ... \\
... & \to & L^n & \xrightarrow{j^n} & L^{n+1} & \to & ...
\end{array}$$

The advantage of this definition is that it applies to any direct system of spaces. In addition, we have another commutative diagram:

$$\begin{array}{ccccccc}
... & \to & H(K^n) & \xrightarrow{i^n_*} & H(K^{n+1}) & \to & ... \\
... & & \downarrow f^n_* & & \downarrow f^{n+1}_* & & ... \\
... & \to & H(L^n) & \xrightarrow{j^n_*} & H(L^{n+1}) & \to & ...
\end{array}$$

We are now in a familiar position; we can enhance our understanding of this new concept by taking a broader, category theory view.

Exercise 7.27. Prove that filtrations form a category with filtration maps as morphisms.

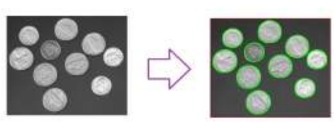
Exercise 7.28. Prove that homology is a functor from the category of filtrations to the abelian groups.

Exercise 7.29. Define the homology map $\{f_n\}_* : H(\{K^n\}) \to H(\{L^n\})$ of a filtration map $\{f_n\} : \{K^n\} \to \{L^n\}$. Compute the homology map of the filtration map in the last example.

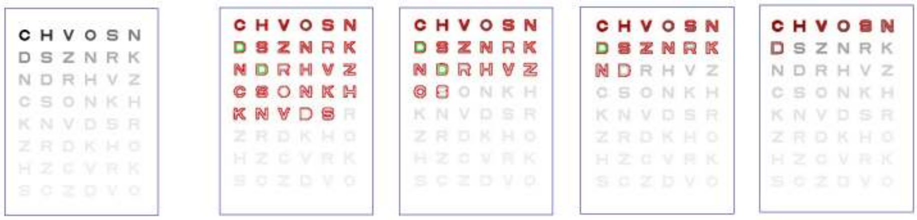
Next, we will further develop our new construction to allow us to "purify" the homology classes of low importance out of the homology groups.

7.6 The "sharp" homology classes of a gray scale image

In analysis of digital images, one may need to measure the importance of topological features in terms of their contrast and discard those with low contrast as irrelevant details or simply *noise*:

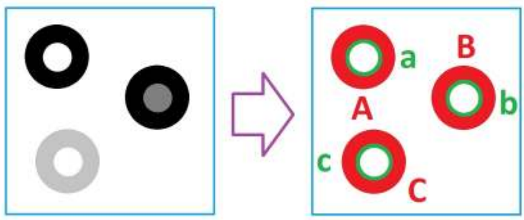


The images below show how the number of objects detected declines as the contrast threshold increases: $0, 30, 50, 70, 125$:



Let's apply this simple idea to the homology of gray scale images of any dimension.

Let's take a look at the image of the rings again:



One may observe that the contrast of some of the features in the original image is lower than others. These are, in comparison to the surrounding area:

- the third ring,
- the hole in the second ring, and
- the hole in the third ring.

It is conceivable that we would classify these less prominent features as minor and choose to ignore them.

It is understood that the "contrast" of an object is the difference between the highest gray level adjacent to the object and the lowest gray level within the object. It is equal to the length of the object's life-span, i.e., the persistence, and as such it is fully applicable to homology classes of all dimensions.

We use the algebra of homology classes again:

$$\begin{array}{ccccccc}
 & i_*^1 & & i_*^2 & & i_*^3 & \\
A_1 & \mapsto & A_2 & \mapsto & A_3 & \mapsto & 0 \\
B_1 & \mapsto & B_2 & \mapsto & B_3 & \mapsto & 0 \\
 & & & & C_3 & \to & 0 \\
a_1 & \mapsto & a_2 & \mapsto & a_3 & \mapsto & 0 \\
b_1 & \mapsto & 0 & & & & \\
 & & & & c_3 & \mapsto & 0
\end{array}$$

Then we find the contrast as the lengths of the life-spans of these homology classes, $P(x)$:

$$\begin{array}{lll}
P(A_1) = 3, & P(A_2) = 2, & P(A_3) = 1, \\
P(B_1) = 3, & P(B_2) = 2, & P(B_3) = 1, \\
 & P(C_2) = 2, & P(C_3) = 1, \\
P(a_1) = 3, & P(a_2) = 2, & P(a_3) = 1, \\
P(b_1) = 1, & & \\
 & & P(c_3) = 1.
\end{array}$$

We define more *inclusions*:

$$i^{m,n} := i^n i^{n-1} \ldots i^m : K^m \hookrightarrow K^n,$$

for $m \leq n$, in order to measure these life-spans.

Let's suppose that we are only interested in features with contrast of at least 3 and consider the rest noise. It is easy to see the homology classes with persistence of 3 or higher among the generators:

$$A_1, B_1, a_1.$$

Unfortunately, because the persistence can decrease under algebraic operations, the elements of high persistence do not, in general, form a subgroup of the respective homology group of the filtration. This is why we should, first, consider the classes with *low* persistence, i.e., the noise. In particular, the classes in $H(K^1)$ of persistence 2 or lower do form a group. It is, in fact, the kernel of $i_*^{2,3}$:

$$\begin{array}{llllll} 0 & = \ker i_*^{1,3}, & < A_2, B_2, C_2 > & = \ker i_*^{2,3}, & < A_3, B_3, C_3 > & = \ker i_*^{3,3}, \\ < b_1 > & = \ker i_*^{1,3}, & < a_2 > & = \ker i_*^{2,3}, & < a_3, c_3 > & = \ker i_*^{3,3}. \end{array}$$

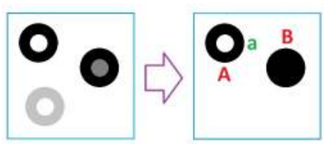
We now "remove" this noise from the homology groups of the filtration by taking their quotients modulo these kernels:

$$\begin{array}{ll} < A_1, B_1 > /0 & =< A_1, B_1 >, \\ < A_2, B_2, C_2 > / < A_2, B_2, C_2 > & = 0, \\ < A_3, B_3, C_3 > / < A_3, B_3, C_3 > & = 0, \\ < a_1, b_1 > / < b_1 > & =< a_1 >, \\ < a_2 > / < a_2 > & = 0, \\ < a_3, c_3 > / < a_3, c_3 > & = 0. \end{array}$$

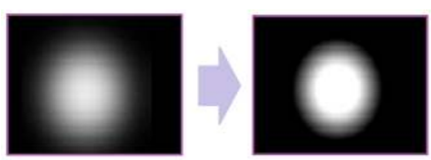
The end result is the 3-*persistent homology groups of the image*:

$$\begin{array}{lll} H_0^3(\{K^i\}) & =< A_1, B_1 > /0 & =< A_1, B_1 >, \\ H_1^3(\{K^i\}) & =< a_1, b_1 > / < b_1 > & =< a_1 > . \end{array}$$

Observe that the output is identical to the homology of a single complex, i.e., a binary image, with two components and one hole:



We have replaced the original with its "flattened" version. This step is illustrated below for the blurred circle:



Exercise 7.30. Similarly interpret "sharpening" as seen in image enhancement applications.

Let's confirm that the persistent homology group of a gray-scale image represents only its *topology*. First, the obvious:

• if we shrunk or stretched one of these rings, its persistence and, therefore, its place in the homology group wouldn't change.

But there is more:

• the holes in the second and third rings have the same persistence (the contrast) and, therefore, occupy the same position in the homology group regardless of their birth dates (the gray

level);
Therefore, these persistence homology groups are independent of any measurements.

In contrast, in the Vietoris-Rips construction, the persistence, i.e., the number *death − birth*, of a homology class does contain information about the *size* of the cycle representing this class. We saw an example with a set of points arranged in a circle that produced a 1-cycle with twice as large birth, death, and persistence than the same set shrunk by a factor of 2. One can then define the persistence via *division* instead, as *death/birth*, in order to have the desired property of scale independence. The same result can be achieved by an appropriate reparametrization of the filtration.

Exercise 7.31. Compute the p-persistent homology groups of (a) the blurred version and (b) the negative of the image of rings, for various values of p.

Exercise 7.32. Show that the p-persistence homology groups remain unchanged when the image is "degraded" by randomly changing the gray levels of the pixels as long as each change is less than p.

7.7 Persistent homology groups of filtrations

In the general context of filtrations, the measure of importance of a homology class is its persistence which is the length of its life-span within the direct system of homology of the filtration:

$$H(K^1) \xrightarrow{\ i_*^1\ } H(K^2) \xrightarrow{\ i_*^2\ } \dots \xrightarrow{\ i_*^{s-1}\ } H(K^s) \xrightarrow{\ 0\ } 0.$$

Definition 7.33. We say that *the persistence* $P(x)$ of $x \in H(K^n),\ n = 1, 2, \dots$, *is equal to* p if
- $i_*^{n,n+p}(x) = 0$, and
- $i_*^{n,n+p-1}(x) \neq 0$.

Our interest is in the "robust" homology classes, i.e., the ones with high persistence. However, the set of all of these classes is not a group as it doesn't even contain 0. That is why we will deal with the "noise" first.

Definition 7.34. Given a positive integer p, the *p-nonpersistent homology group* $\mathring{H}^p(\{K^n\})$ of filtration $\{K^n\}$ is the group of all elements of $H(\{K^n\})$ with persistence less than p.

This means that for every $x \in H(K^n),\ n = 1, 2, \dots$, we have:

$$x \in \mathring{H}^p(\{K^n\}) \Longleftrightarrow i_*^{n,n+p-1}(x) = 0.$$

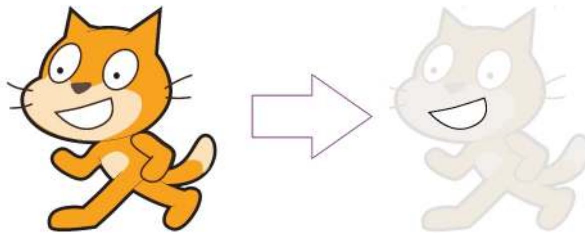
Next, we remove the noise from the homology group. Just as in the noiseless case we started with, we define a single structure to capture all (robust) homology classes.

Definition 7.35. The *p-persistent homology group of* $\{K^n\}$ is defined as

$$H^p(\{K^n\}) := H(\{K^n\}) / \mathring{H}^p(\{K^n\}).$$

The meaning of this definition is that, *if the difference between two homology classes is deemed noise, they ought to be equivalent.*

Exercise 7.36. Show that we can never end up with the "Cheshire cat of topology" by removing a component as noise but keeping a hole in it:



We next explore how this definition applies to each complex of the filtration. Suppose a positive integer p is given. For each complex K^n, we can think of its p-nonpersistent homology group $\mathring{H}^p(K^n)$ as the group of all elements of $H(K^n)$ with persistence less than p:

$$\mathring{H}^p(K^n) := \ker i_*^{n,n+p-1}.$$

Suppose

$$x \in \mathring{H}^p(K^n) \subset H(K^n).$$

Let $y := i_*^{n,n+1}(x) \in H(K^{n+1})$. What is its persistence? Let's compute:

$$\begin{array}{lll} i_*^{n+1,n+1+p}(y) & = & i_*^{n+1,n+1+p}\Big(i_*^{n,n+1}(x)\Big) \\ & = & i_*^{n,n+1+p}(x) = i_*^{n+p,n+p+1}\Big(i_*^{n,n+p}(x)\Big) \\ & = & i_*^{n+p,n+p+1}(0) = 0. \end{array}$$

Hence $y \in \ker i_*^{n+1,n+1+p} = \mathring{H}^p(K^{n+1})$. We have proved that the homomorphism

$$i_*^{n,n+1} : \mathring{H}^p(K^n) \to \mathring{H}^p(K^{n+1}),$$

as the restriction of the homology map of the inclusion, is well-defined.

Therefore, we have a new direct system:

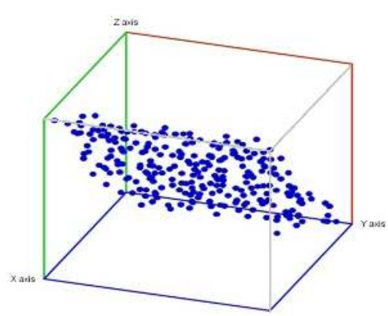
$$\mathring{H}(K^1) \xrightarrow{\ i_*^1\ } \mathring{H}(K^2) \xrightarrow{\ i_*^2\ } \ldots \xrightarrow{\ i_*^{s-1}\ } \mathring{H}(K^s) \xrightarrow{\ 0\ } 0.$$

Exercise 7.37. Prove these identities:

$$\begin{array}{llllll} \mathring{H}^p(\{K^n\}) & = \mathring{H}^p(K^1) & \oplus & \ldots & \oplus & \mathring{H}^p(K^s), \\ H^p(\{K^n\}) & = \Big(H(K^1)/\mathring{H}^p(K^1)\Big) & \oplus & \ldots & \oplus & \Big(H(K^1)/\mathring{H}^p(K^s)\Big). \end{array}$$

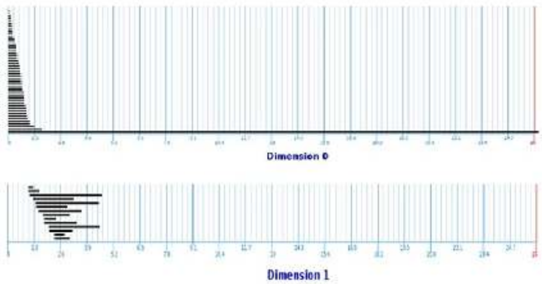
Exercise 7.38. State and prove the properties of $H^p(\{K^n\})$ analogous to the ones of $H(\{K^n\})$.

Example 7.39. We consider next an example of computing the persistent homology of a point cloud with homology software called JPlex. Here, noise has been added to a point cloud of the *plane* in space:



Let's see if we can recover the topology of the underlying data.

A filtration is built via the Vietoris-Rips construction, then JPlex provides us with the life-spans of the cycles:

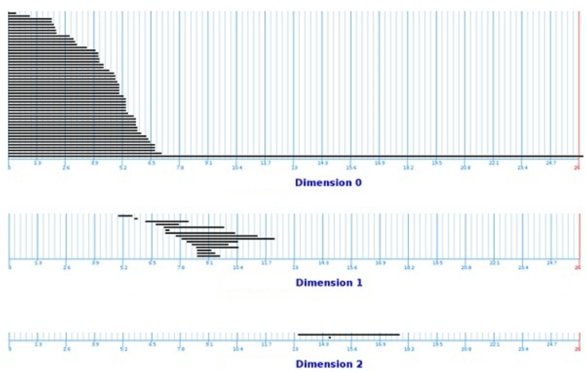


The data shows the following:

- There is a 0-cycle with a much longer life-span than others. So, the set is probably *path-connected*.
- There is no single 1-cycle with significantly longer life-span. So the set is probably *simply connected*.

In addition to these global topological characteristics of the complex, homology theory can also provide local information. For every small patch P of K, we can define and compute the "local" homology by, essentially, collapsing the complement of P to a single point. Then, if K is a manifold, its *dimension* is equal to n provided this homology is non-trivial only in dimension n.

These are the life-spans of these cycles of our filtration:



This is what we see:

- In dimension 0, one cycle persists longer that other 0-cycles.
- In dimension 1, no single cycle has a particularly long life-span compared to other 1-cycles.
- In dimension 2, one cycle has a significant life-span.

Following this data, we determine the dimension of our data set. We discard the low persistence homology classes, and what's left gives us the *persistent Betti numbers* as the dimensions of the persistent homology groups:

$$\beta_0 = 1,\ \beta_1 = 0,\ \beta_2 = 0, ...$$

Therefore, the dataset has a single component and no other topological features. Next, the local Betti numbers are:

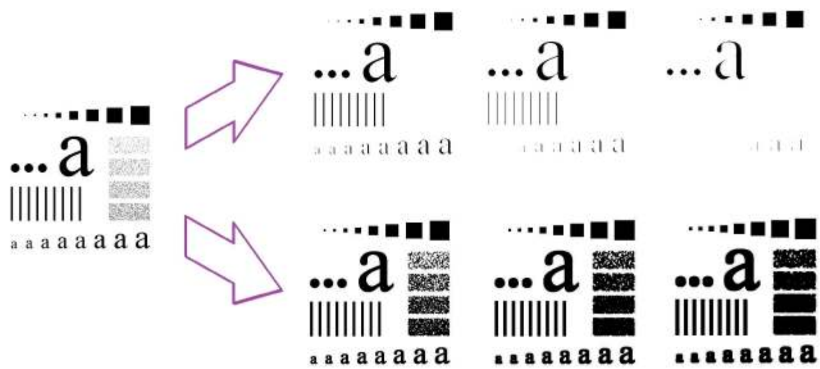
$$\beta_0 = 1,\ \beta_1 = 0,\ \beta_2 = 1,\ \beta_3 = 0, ...$$

The result confirms that the data set is *two-dimensional*, i.e., a surface. □

7.8 More examples of parametric spaces

We have seen three main instances of parametric complexes in this section:
• the Vietoris-Rips construction for a point cloud,
• the thresholding method for a gray scale image, and
• the increasing resolution method for a binary image. Gray scale images are an especially convenient target for demonstrating the power of persistence as a way to handle uncertainty. In terms of complexity, they lie between binary and color images.

There is no uncertainty in *binary images* similar to the one associated with the presence of gray levels as each pixel is either black or white. Yet, we have already seen that switching the value of a single pixel could dramatically change the topology of the image. When this happens, we can say that the topological features that have appeared or disappeared aren't *robust*. Another way to create filtrations comes from the co-called "morphological operations":
• *dilation* makes black each pixel adjacent to black, while
• *erosion* makes white each pixel adjacent to white.



Thus, dilation and erosion expand and shrink objects in a binary image, respectively. Then the robustness of the topological features of the image can be measured in terms of how many dilations and erosions it takes to make it disappear. It is clear that repeating dilations will produce a filtration while erosion produces a reversed filtration. Therefore, the robustness described above is just another example of *persistence*!

These are the types of questions we would answer:
• how many erosions does it take to split a component into two or more or make it disappear?
• how many dilations does it take to create a hole in a component or merge two components?
• etc.

These are some typical observations:
• a component separated from others will have higher persistence under dilation that one with close neighbors; or
• a round component will have higher persistence under erosion than an elongated one.

Exercise 7.40. Make the above statements precise and prove them.

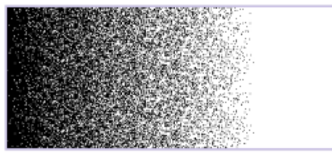
Exercise 7.41. Define filtrations and persistence for a continuous parameter, $t \in [0, 1]$. What would dilation/erosion filtrations look like?

Exercise 7.42. Suppose four individuals, A, B, C, D, visited four stores: Kroger, Macy's, JCPenney, and Walmart. Suppose they also rated the stores with numbers between 1 and 4:

	K	M	J	W
$A:$	1	2	3	4
$B:$	1	1	1	1
$C:$	4	4	1	1
$D:$	3	2	2	1

Construct a filtration of simplicial complexes based on these ratings and compute the persistence.

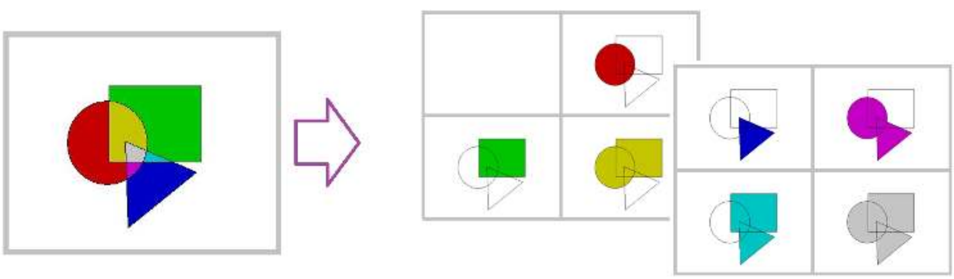
Exercise 7.43. Starting with a point cloud, define a filtration based on the *density* of its points:



7.9 Multiple parameters

The color effect in *color images* is produced by combining different levels of the "primary colors": red, green, and blue (RGB). Just as the gray in gray scale images, the values of each primary color runs from 0 to 255. So, in an RGB image, every pixel is assigned 3 integers. One can think of a color image as a 3-vector attached to each pixel, or as 3 tables of integers, or 3 gray scale images, etc. However, what we have learned from our analysis of gray scale images is that we should break this parametric image into *binary* images.

Example 7.44. To illustrate how, we present a simplified case when there are only 2 levels of each of the three primary colors. The image is represented by an array of 8 binary images (they are colored for convenience):



This array is a "triple filtration" with inclusions that go: from left to right, from top to bottom, and from the lower level to the top level. Then we conclude that, without double-counting, there is only one component in the image. This seemingly counter-intuitive conclusion is justified as follows: yes, there are three clearly visible objects in the image, but they *overlap*! □

In the general case, we apply a procedure similar to thresholding described above. The "color space" is a $255 \times 255 \times 255$ array with each entry corresponding to a particular color. Here is a strip from this space:

Therefore, thresholding a color image will produce a $255 \times 255 \times 255$ array of binary images.

Exercise 7.45. Give a precise description of this thresholding procedure.

The following property of this array is important: when the value of any of the primary color increases, the binary image grows. As a result, the array is also equipped with inclusions connecting these binary images. Together these complexes and maps form a 3-dimensional commutative diagram and so do their homology groups and homology maps.

One layer of this filtration, as well as its homology, is shown below with all arrows representing inclusions. Both diagrams are commutative:

$$
\begin{array}{ccccccc}
K^{0,0} & \xrightarrow{i} & K^{1,0} & \xrightarrow{i} & K^{2,0} & \xrightarrow{i} & ... \\
\downarrow i & & \downarrow i & & \downarrow i & & ... \\
K^{0,1} & \xrightarrow{i} & K^{1,1} & \xrightarrow{i} & K^{2,1} & \xrightarrow{i} & ... \\
\downarrow i & & \downarrow i & & \downarrow i & & ... \\
K^{0,2} & \xrightarrow{i} & K^{1,2} & \xrightarrow{i} & K^{2,2} & \xrightarrow{i} & ... \\
\downarrow i & & \downarrow i & & \downarrow i & & ... \\
.. & .. & .. & .. & .. & .. & ...
\end{array}
\qquad
\begin{array}{ccccccc}
H(K^{0,0}) & \xrightarrow{i_*} & H(K^{1,0}) & \xrightarrow{i_*} & H(K^{2,0}) & \xrightarrow{i_*} & ... \\
\downarrow i_* & & \downarrow i_* & & \downarrow i_* & & \\
H(K^{0,1}) & \xrightarrow{i_*} & H(K^{1,1}) & \xrightarrow{i_*} & H(K^{2,1}) & \xrightarrow{i_*} & ... \\
\downarrow i_* & & \downarrow i_* & & \downarrow i_* & & \\
H(K^{0,2}) & \xrightarrow{i_*} & H(K^{1,2}) & \xrightarrow{i_*} & H(K^{2,2}) & \xrightarrow{i_*} & ... \\
\downarrow i_* & & \downarrow i_* & & \downarrow i_* & & \\
.. & .. & .. & .. & .. & & ...
\end{array}
$$

Exercise 7.46. Draw the rest of the diagram.

We can call this a *multi-parameter filtration*.

What is the homology group of such an image? As before, we look at the kernels of the homology maps of the inclusions. We want to capture for each element of the filtration its homology classes that are about to die (the zero group is added at the end as before). There are three inclusions for each and we take the intersection of these three kernels in order to make sure we are including only the homology classes that won't reappear elsewhere. Finally, *the homology group of the color image* is the direct sum of these intersections over all elements of the filtration.

Exercise 7.47. Find the homology group of this color image and the one above:

Exercise 7.48. Prove that a gray scale image analyzed as a color image will have the same homology group as before.

Exercise 7.49. Besides color images, show how multi-parameter filtrations may come from combinations of previous constructions: the erosion-dilation of gray scale images and the Vietoris-Rips construction for gray scale images. Suggest your own.

It is just as easy, however, to develop a homology theory in an even more general setting. A *direct system* $\{K^n\} = \{K^n : n \in Q\}$ is a collection of sets indexed by a partially ordered set Q so that if $n < m$ then there is a function

$$i^{nm} : K^n \to K^m.$$

When these sets are complexes and the functions are simplicial or cubical maps, homology yields a new direct system, one of groups and homomorphisms:

$$i_*^{nm} : H(K^n) \to H(K^m), \; n < m.$$

For convenience, we add a maximal element $X = 0$ to this direct system. As before, the goal is to capture all homology classes of the filtration, without repetition. We define the *homology group of the direct system* $\{K^n\}$ as

$$H(\{K^n\}) := \bigoplus_n \bigcap_{m>n} \ker i_*^{nm}.$$

The analogs of the results about the single parameter filtrations hold.

Exercise 7.50. Show that this definition generalizes the ones given in this section.

Exercise 7.51. Define persistence in the multi-parameter setting. Hint: beware of the "Cheshire cat"!

Exercise 7.52. Compare the above construction to the following. Given a direct system of groups and homomorphisms:

$$\{A^n\} = \{A^n, j^{nm} : A^n \to A^m | \ n, m \in Q, n < m\},$$

the *direct limit* of this system is the group

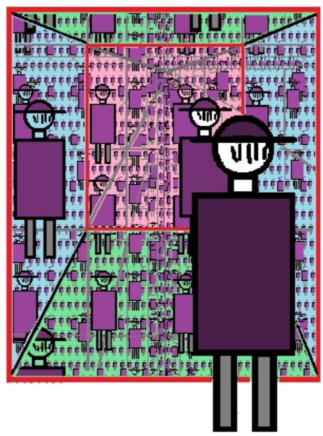
$$\lim_{\to}\{A^n\} := \bigoplus_n A^n /_{\sim},$$

where

$$x_n \sim x_m \Longleftrightarrow j^{nq}(x_n) = j^{mq}(x_m), \text{ for some } q \in Q.$$

Chapter IV

Spaces



1 Compacta

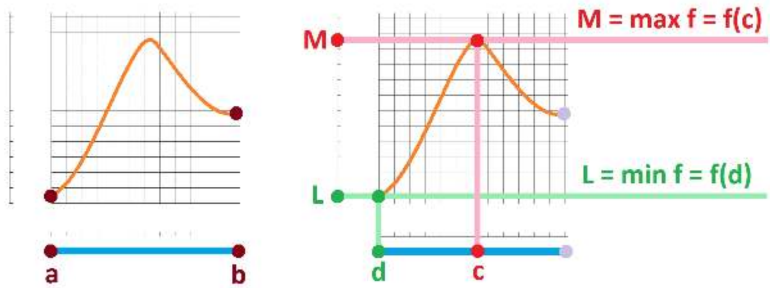
1.1 Open covers and accumulation points

Previously, we have proved one of the two main topological theorems of elementary calculus, the Intermediate Value Theorem: a continuous function $f : [a, b] \to \mathbf{R}$ attains all the values between $f(a)$ and $f(b)$. It follows from this theorem that the image of a path-connected space (under a continuous map) is a path-connected.

The second main topological theorem of Calculus 1 is:

Theorem 1.1 (Extreme Value Theorem). Continuous functions attain their extreme values (the maximum and minimum) on closed bounded intervals. In other words, if f is continuous on $[a, b]$, then there are c, d in $[a, b]$ such that

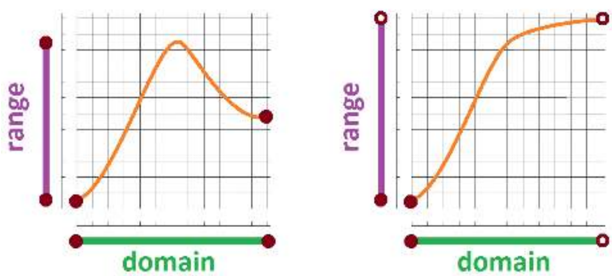
$$\begin{aligned} f(c) = M &= \max_{x\in[a,b]} f(x), \\ f(d) = L &= \min_{x\in[a,b]} f(x). \end{aligned}$$



Exercise 1.2. Generalize the theorem to the case of $f : [a, b] \cup [c, d] \to \mathbf{R}$.

Our immediate goal is to prove the theorem while developing some profound topological tools and ideas.

In a calculus course, one is supplied with examples of how the above theorem fails if we replace the domain interval $[a, b]$ with other kinds of intervals: $(a, b), [a, b), (a, b], [a, \infty)$, etc. Here, on the surface, two separate conditions about the interval are violated: these intervals are either not closed or not bounded. We will seek a single condition on the domain of the function that would guarantee that the theorem holds. As these counterexamples show, it must have something to do with the convergence of the sequence $\{f(x_n)\}$ as the sequence $\{x_n\}$ is approaching the complement of the domain:



Since, as we know, sequences aren't an adequate tool for studying topological spaces in general, we will look at all infinite subsets. To understand what's going on, we will rely on the following familiar concept. A point x in a topological space X is called an accumulation point of subset A of X if for any neighborhood U of x

$$U \cap (A \setminus \{x\}) \neq \emptyset.$$

Then, what makes a difference is that every sequence in $[a, b]$ appears to have an accumulation point – the condition that fails for the all other intervals.

Suppose A is a subset of a topological space X and let's assume that A has no accumulation points. Then, for any point $x \in X$, there is an open neighborhood U_x of x such that

$$U_x \cap (A \setminus \{x\}) = \emptyset.$$

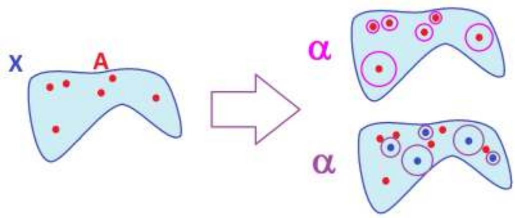
Let's consider the collection of all of these sets:

$$\alpha = \{U_x : x \in X\}.$$

What we have here is an *open cover* of X. By choice, this cover satisfies a certain special property. This property can be rewritten as a combination of two cases:

- 1. if $x \in A$ then $U_x \cap A = \{x\}$,
- 2. if $x \notin A$ then $U_x \cap A = \emptyset$.

Note that the former case means that x is an isolated point of A.



The two cases indicate that each element of α contains, at most, one element of A. Therefore, set A can't have more elements than α:

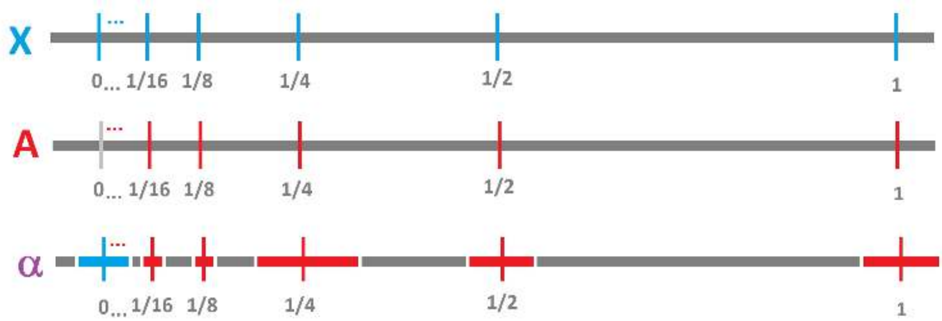
$$\#A \leq \#\alpha.$$

Let's consider a specific example that illustrates the consequences of this construction.

Example 1.3 (dimension 1). Let's choose

$$X := \{0, 1, 1/2, 1/3, ...\}, \quad A := \{1, 1/2, 1/3, ...\},$$

with the topology acquired from $\mathbf{R}$:



We keep the assumption that A has no accumulation points in X. Then, as we know, there is an open cover α satisfying the above properties. Because the topology of X is Euclidean, let's assume that the elements of α are the intersections of open finite intervals, such as $(a-\varepsilon, a+\varepsilon)$, with X. Then, for any $n = 1, 2, 3, ...$, there is an open interval $U_n = (1/n - \varepsilon_n, 1/n + \varepsilon_n)$ with a single point of intersection with A:

$$U_n \cap A = \{1/n\}.$$

Indeed, we just choose:

$$\varepsilon_n := \frac{1}{n(n+1)}.$$

Furthermore, there is one more interval $U = (\varepsilon, \varepsilon)$, the one that satisfies:

$$U \cap A = \emptyset.$$

Of course, we realize that this is impossible:

$$n > [1/\varepsilon] + 1 \Longrightarrow 1/n \in U,$$

and, therefore, the assumption was wrong. However, we will ignore this obvious conclusion and explore instead what this means for the open cover α. We just discovered that this *infinite* cover

$$\alpha := \{U, U_1, U_2, U_3, ...\}$$

has a *finite* subcover

$$\alpha' := \{U, U_1, U_2, U_3, ..., U_n\}.$$

Since there is, at most, one point of A per element of α, this new collection can't cover the whole A because A is infinite. A contradiction. □

The crucial step in this construction was our ability to choose a finite subcover. This motivates the following definition:

Definition 1.4. A topological space X is called *compact* (or a compactum) if every open cover contains a finite subcover.

We can now continue with the general argument and prove the following important theorem.

Theorem 1.5 (Bolzano-Weierstrass Theorem). In a compact space, every infinite subset has an accumulation point.

Proof. As before, we have a compact topological space X and a subset A. We have already constructed an open cover

$$\alpha = \{U_x : \; x \in X\}$$

such that

- there is, at most, one element of A in an element of α.

If we now choose a finite subcover

$$\alpha' = \{U_{x_i} : \; x_i \in X, \; i = 1, 2, ..., n\},$$

it will satisfy the same property. But this property implies that

$$A \subset \{x_i : \; i = 1, 2, ..., n\}.$$

This set is finite while A is infinite, a contradiction. ■

Thus, only the closed intervals are compact.

The following version of the Bolzano-Weierstrass Theorem is also very important.

Corollary 1.6. In a compact space, every sequence has a convergent subsequence.

Exercise 1.7. Prove the corollary for (a) metric spaces, and (b) all topological spaces. Hint: think of the sequence as a subset A and consider two cases: A is infinite and A is finite.

1.2 The definitions

The issue of compactness has two sides. First, as above, we look at the compactness of topological spaces:

Definition 1.8. For any topological space X, a collection of open sets α is called an *open cover* if

$$X = \cup \alpha.$$

Then, X is called a *compact topological space* if every open cover has a finite subcover.

Exercise 1.9. Provide a definition of compactness in terms of closed sets.

Alternatively, we define compactness of subsets:

Definition 1.10. For any subset X of a topological space Y, a collection α of open (in Y) sets is called an *open cover* if

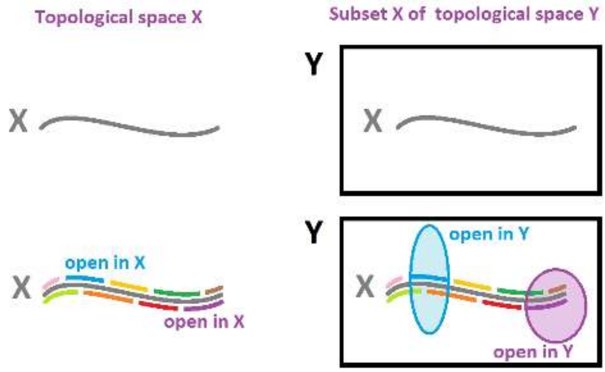
$$X \subset \cup \alpha.$$

Then, X is called a *compact subset* if every open cover has a finite subcover.

These definitions are equivalent in the following sense:

Proposition 1.11. If X is a compact subset of space Y, then X is compact, in relative topology.

Exercise 1.12. Prove the proposition. Hint:



The nature of the first definition – given entirely in terms of open sets – reveals that compactness is a topological invariant: it is preserved by homeomorphisms. Moreover, the image of a compact space under a continuous function is compact.

Exercise 1.13. Show that the preimage of a compact space doesn't have to be compact.

Theorem 1.14. Compactness is preserved by continuous functions; i.e., if $f : X \to Y$ is a map and X is compact then $f(X)$ is also compact.

Proof. Suppose α is an open cover of $f(X)$. We need to find its finite subcover. The idea is, with f, to take its elements to X, make the collection finite, and then bring what's left back to Y. The elements of α are open sets in Y. Then, since f is continuous, for each $U \in \alpha$, its preimage $f^{-1}(U)$ under f is open in X. Therefore,

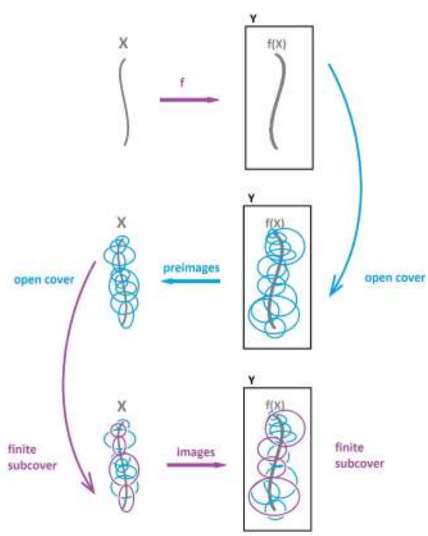
$$\beta := \{f^{-1}(U) : \ U \in \alpha\}$$

is an open cover of X. Suppose β' is its finite subcover. Then

$$\alpha' := \{U \in \alpha : \ f^{-1}(U) \in \beta'\}$$

is a finite subcover of α. ■

The proof is illustrated below:



Proposition 1.15. (1) All finite topological spaces are compact. (2) Anti-discrete is a compact topology.

Exercise 1.16. Prove the proposition.

The following result significantly simplifies the task of proving compactness.

Theorem 1.17. Suppose X is a topological space. Then X is compact if and only if there is a basis β of its topology so that every cover of X by members of β has a finite subcover.

Proof. $[\Leftarrow]$ Suppose γ is an open cover of X. Then, we represent every element of γ as the union of elements of β. This gives us a cover of X by the elements of β. Take its finite subcover, α. Finally, for every element of α, choose a single element of γ that contains it. These sets form a finite subcover of γ. ■

Exercise 1.18. Show that every element of γ can be represented as the union of elements of β

We would like to apply the idea of compactness and the results we have proven to calculus and, especially, the Extreme Value Theorem. For that, we need to take a better look at *intervals.*

1.3 Compactness of intervals

First, the real line $\mathbf{R} = (-\infty, \infty)$ is not compact: it suffices to consider the cover by all intervals of a given length, ε.

A similar argument will prove that no ray, $(a, \infty), [a, \infty)$, is compact.

Next, let's take a *finite* open interval $I = (a, b)$ in $\mathbf{R}$. Let's choose the cover that consists of all open ε-intervals:

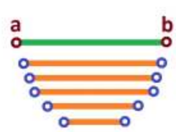
$$\alpha := \{(p, q) \cap I : \ p - q = \varepsilon\}.$$

Then finding a finite subcover is easy:

$$\alpha' := \{(0, \varepsilon), (\tfrac{1}{2}\varepsilon, (1 + \tfrac{1}{2})\varepsilon), (\varepsilon, 2\varepsilon), ...\}.$$

There will be $[2/\varepsilon] + 1$ intervals.

However, the following cover of (a, b) (an expanding sequence of intervals) does not have a finite subcover:



We choose:

$$\alpha := \{(p_n, q_n) : \ n = 1, 2, 3, ...\},$$

with

$$p_n = a + \tfrac{1}{n}, \ q_n = b - \tfrac{1}{n}.$$

The same argument applies to $[a, b)$ and $(a, b]$.

Thus, *open and half-open intervals aren't compact*!

But closed bounded intervals are. Why is this important?

We already know that the image of an interval under a continuous function is an interval.

Exercise 1.19. Prove this statement. Hint: use path-connectedness.

Now we also discover that the image of a *closed bounded* interval is also a closed bounded interval:

$$f\Big([a,b]\Big) = [c,d],\ a<b, c<d.$$

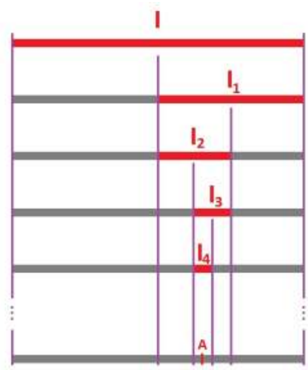
The Extreme Value Theorem will follow from this simple formula.

Exercise 1.20. Prove this statement.

In turn, it implies that the sup-norm is well-defined.

Theorem 1.21 (Heine-Borel Theorem). Every interval $[a,b]$ is compact.

Proof. The proof is by contradiction. Suppose we are given an open cover α of $[a,b]$ that doesn't have a finite subcover. The *idea of the proof* is that, since $[a,b]$ is "bad", then so is one of its halves, and a half of that half too, etc., until only a single point is left, which can't be "bad":



The plan is to construct a "nested" sequence of intervals

$$I=[a,b]\supset I_1=[a_1,b_1]\supset I_2=[a_2,b_2]\supset ...,$$

so that they have only one point in common.

We start with the following two observations:
- 1. α is an open cover for all elements of the sequence $I=[a,b]$ as well as for all of its subsets.
- 2. if α has no finite subcover for an interval $J\subset I$, then α has no finite subcover for at least one of its halves.

These two facts allow us to construct the sequence inductively.

Consider the halves of the interval $I=[a,b]$:

$$\left[a,\frac{a+b}{2}\right],\ \left[\frac{a+b}{2},b\right].$$

For at least one of them, there is no finite subcover for α. Call this interval $I_1:=[a_1,b_1]$. Next, we consider the halves of this new interval:

$$\left[a_1,\frac{a_1+b_1}{2}\right],\ \left[\frac{a_1+b_1}{2},b_1\right].$$

For at least one of them, there is no finite subcover for α. Call this interval $I_2 := [a_2, b_2]$. We continue on with this process and the result is a sequence of intervals

$$I = [a, b] \supset I_1 = [a_1, b_1] \supset I_2 = [a_2, b_2] \supset ...,$$

that satisfies these two properties:

$$a \leq a_1 \leq ... \leq a_n \leq ... \leq b_n \leq ... \leq b_1 \leq b,$$

and

$$|b_{n+1} - a_{n+1}| = \frac{1}{2}|b_n - a_n| = \frac{1}{2^n}|b - a|.$$

Now we invoke the *Completeness Property of Real Numbers*: every bounded increasing (decreasing) sequence converges to its least upper bound (the greatest lower bound). Therefore, the two sequences converge:

$$\lim_{n\to\infty} a_n = A, \lim_{n\to\infty} b_n = B.$$

The second property above implies that

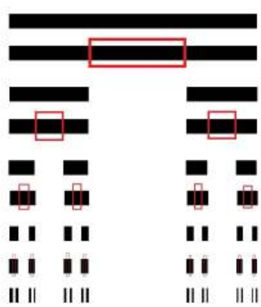
$$\lim_{n\to\infty} |a_n - b_n| = 0.$$

We then conclude that $A = B$.

Now, for this new single-point set, α must have a finite subcover. In fact, the subcover needs only one element: $A \in U \in \alpha$. But due to $a_n \to A$ and $b_n \to A$, we have $a_n \in U$ and $b_n \in U$ for large enough n. Then, $I_n = [a_n, b_n] \subset U$; it's covered by this, finite, subcover $\{U\}$ of α! This contradicts our assumption. ■

Exercise 1.22. What tacit assumption about U did we make? Fix the proof.

Example 1.23 (Cantor set). Suppose $I_0 = [0, 1]$ is the closed unit interval. Let I_1 be the result of deleting the middle third $(1/3, 2/3)$, from I_0, leaving us with $[0, 1/3] \cup [2/3, 1]$. Next, let I_2 be the result of deleting the middle third from the intervals of I_1, etc.



We define for $n > 0$:

$$I_n := I_{n-1} \setminus \bigcup_{k=0}^{n-1} \Big(\frac{1+3k}{3^n}, \frac{2+3k}{3^n}\Big).$$

Then we have a decreasing sequence of compact subsets of $\mathbf{R}$. Finally, we define the *Cantor set* by

$$C := \bigcap_n I_n.$$

Even though the total length of I_n declines to 0, the Cantor set is uncountable. □

Exercise 1.24. (a) Prove that the Cantor set is compact. (b) Prove that every point of C is a limit point of C.

1.4 Compactness in finite and infinite dimensions

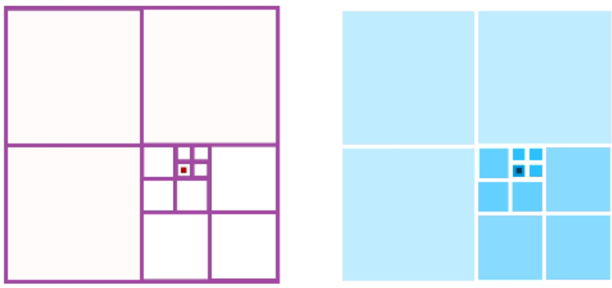
A generalization of the Heine-Borel Theorem is as follows.

Theorem 1.25. A closed bounded subset of a Euclidean space is compact.

It also follows that *all cells are compact*!

Exercise 1.26. Show that even though the combination of closedness and boundedness (of a subset of a Euclidean space) is a topological property, neither of the two alone is.

Before we address the general case, let's consider *rectangles* in the plane. Can we generalize the proof in the last subsection? The picture below suggests that we can:



Exercise 1.27. By following the proof of the Heine-Borel Theorem, prove that a closed rectangle in the plane is compact. Hint: instead of nested intervals, consider nested rectangles.

Later, we will show that compactness is preserved under *products*. That will prove compactness of the closed squares, cubes, ..., n-cubes. Then the above theorem will easily follow if we can transition to subsets. Of course, not all subsets of a compact space are compact, but the closed ones are.

Theorem 1.28. A closed subset of a compact space is compact.

Proof. Suppose Y is compact and $X \subset Y$ is closed. Suppose α is an open cover of X. We want to find a subcover of α using the compactness of Y. The only problem is that α doesn't cover the whole Y! Then the *idea* is to follow these steps:

- 1. augment α with extra open sets to construct an open cover β of the whole Y, then
- 2. find its finite subcover β', and finally
- 3. extract from β' a subcover of X by dropping the extra elements we added:

$$\alpha' := \beta' \cap \alpha.$$

The first step can be carried out in a number of ways; however, we need to make sure that step 3 will give us a collection that still covers X. This outcome is impossible to ensure unless we choose to add only the elements that don't intersect X. With that idea, the choice becomes obvious: we need only one set and it's the complement of X. This set is open because X is closed. So

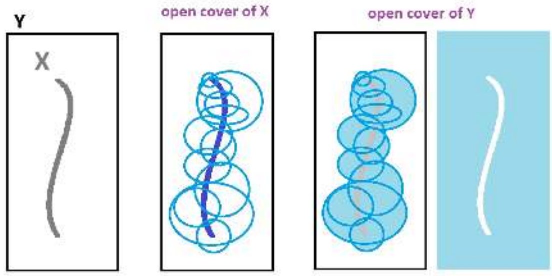
$$\beta := \alpha \cup \{Y \setminus X\}.$$

As Y is compact, β has a finite subcover, β'. Finally, we choose:

$$\alpha' := \beta' \setminus \{Y \setminus X\}.$$

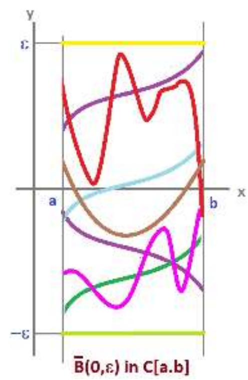
■

The proof is illustrated below:



Then any closed bounded subset is a subset of a closed cube in $\mathbf{R}^n$ and, therefore, is compact.

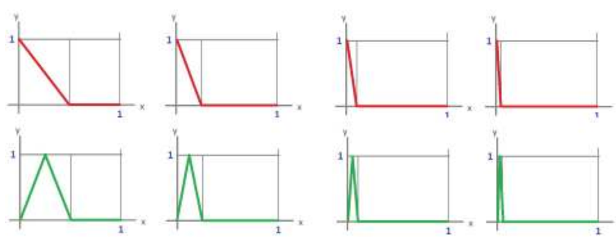
Example 1.29. This is no longer true for *infinite dimensional spaces.* As a counterexample, let's consider a closed ball in the space of functions with sup-norm:



Here

$$\bar{B}(0,\varepsilon)=\{f\in C[a,b]:\ \max_{x\in[a,b]}|f(x)|\le\varepsilon\},\ \varepsilon>0.$$

Now, consider the following two sequences of continuous functions in $\bar{B}(0,1)$:



Even though the values of the functions at any point become closer and closer to zero, there is no convergence of functions themselves. □

Exercise 1.30. Provide formulas for these functions. Show that the first sequence diverges and the second converges, point-wise. Yet, the distance to the limit, the zero function, remains the same. Conclude that there is no uniform convergence and that $\bar{B}(0,\varepsilon)$ isn't compact in $C[a,b]$. What about $C_p[a,b]$?

Exercise 1.31. Provide a similar argument to show that the sequence:

$$(1,0,0,0,...),(0,1,0,0,...),(0,0,1,0,...),...,$$

has no convergent subsequence. Hint: what's the topology?

Exercise 1.32. Prove the *Arzela-Ascoli Theorem*: a closed bounded subset X of $C[a.b]$ is compact if, in addition, it is *equicontinuous*: for each $c \in [a,b]$ and every $\varepsilon > 0$, there is a δ such that

$$|x-c| < \delta \Longrightarrow |f(x) - f(c)| < \varepsilon, \ \forall f \in X.$$

Hint: to understand the concept, limit the set to differentiable functions with the derivatives between, say, -1 and 1.

Exercise 1.33. Suppose X is a compact space. Define the analogs $C(X)$ and $C_p(X)$ of the spaces of real-valued continuous functions $C[a,b]$ and $C_p[a,b]$ and examine the issue of compactness.

Definition 1.34. A space is called *locally compact* if its every point has a neighborhood the closure of which is compact.

Plainly, $\mathbf{R}^n$ is locally compact.

Exercise 1.35. Is either discrete or anti-discrete topology locally compact?

Proposition 1.36. $C[a,b]$ is not locally compact.

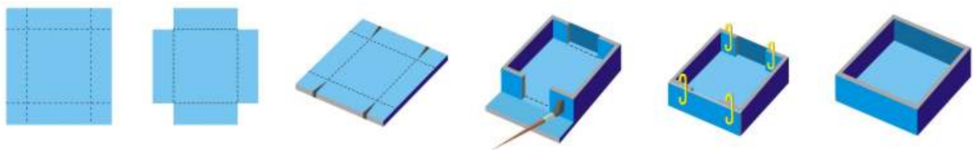
Exercise 1.37. Prove the proposition.

---

2 Quotients

2.1 Gluing things together

We can build new things from old by *gluing*:



In fact, we can build a lot of topologically different things with nothing but sheets of paper and a glue-stick:

What is behind this gluing metaphor is an equivalence relation. Its axioms will then make practical sense. The Reflexivity Axiom, $A \sim A$, is: every spot of the sheet is glued to itself. The Symmetry Axiom, $A \sim B \Longrightarrow B \sim A$, becomes: a drop of glue holds either of the two sheets equally well.

Most of the time, we will attach the sheets edge-to-edge, without overlap. In that case, *welding* may be a better metaphor:

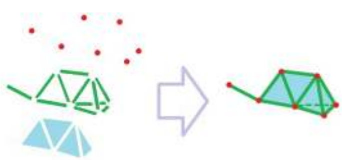
Then the Transitivity Axiom, $A \sim B, B \sim C \Longrightarrow A \sim C$, means: the two seams fully merge and become indistinguishable.

Thinking of a *zipper* is also appropriate:

Topologically, zipping a jacket turns a *square* with two holes into a *cylinder* with two holes, or, even better, a disk with *two* holes turns into a disk with *three* holes:

Making balloon animals is another example:

Meanwhile, we have already seen gluing when we constructed topological spaces from *cells* as realizations of simplicial complexes:



These realizations, however, were placed within a specific Euclidean space $\mathbf{R}^N$. We will see that this is unnecessary.

2.2 Quotient sets

Before we consider the topological issues, let's make clear what happens to the underlying *sets* first.

We pick two examples of somewhat different nature.

Example 2.1 (circle). The easiest way to construct the circle $\mathbf{S}^1$ is to take the closed segment $[0,1]$ and glue the endpoints together. The gluing procedure is expressed by the following equivalence relation on $X = [0,1]$:

- 1. $0 \sim 1$, and
- 2. $x \sim x$ for all $x \in X$.

The second condition is required by the Reflexivity Axiom and will be assumed implicitly in all examples. Then we record the equivalence relation simply as

$$0 \sim 1.$$

We use the following **notation** for the quotient set:

$$X/_{\sim} := \{[x] : x \in X\}.$$

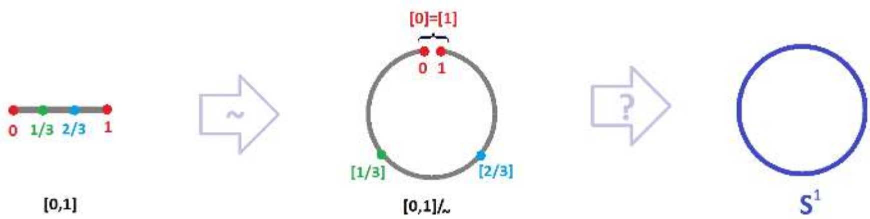
As we know, it is simply the set of all equivalence classes of this equivalence relation:

$$\mathbf{S}^1 := [0,1]/_{\sim}.$$

where

$$[x] := \{y \in X : y \sim x\} = \begin{cases} \{0,1\} & \text{if } x = 0,1; \\ \{x\} & \text{if } x \in (0,1). \end{cases}$$

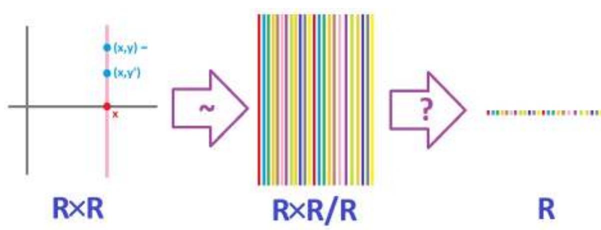
The image below is just an illustration of what these equivalence classes look like:



Only one of them is non-trivial. □

Example 2.2 (plane). The next example is a familiar one from linear algebra. We choose

- $X = \mathbf{R}^2$ and
- $(x,y) \sim (x,y')$ for all $x,y,y' \in \mathbf{R}$.



The equivalence classes are the vertical lines. Instead of using the gluing metaphor, we say that each of these lines *collapse* to a point. The reason is that each of the vertical lines corresponds to its point of intersection with the x-axis. Hence, the quotient set corresponds, in this sense, to the real line. Algebraically, we need to find a map:

$$q : X/_{\sim} = \{\{(x,y) : y \in \mathbf{R}\} : \ x \in \mathbf{R}\} \to \{x : \ x \in \mathbf{R}\} = \mathbf{R}.$$ □

Definition 2.3. The function that takes each point to its equivalence class is called the *identification function*:

- $q : X \to X/_{\sim}$ given by

- $q(x) := [x]$.

It may be called the "gluing", or "attaching", map in a topological context.

When we recognize the quotient as a familiar set, we may try to present the identification map explicitly. In the first example, the identification function

$$q : [0,1] \to \mathbf{S}^1$$

is given by

$$q(t) := (\cos(\pi t), \sin(\pi t)),\ t \in [0,1].$$

In the second example, q may be thought of as the projection:

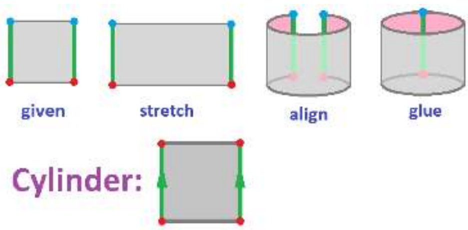
$$q : \mathbf{R}^2 \to \mathbf{R},$$

given by

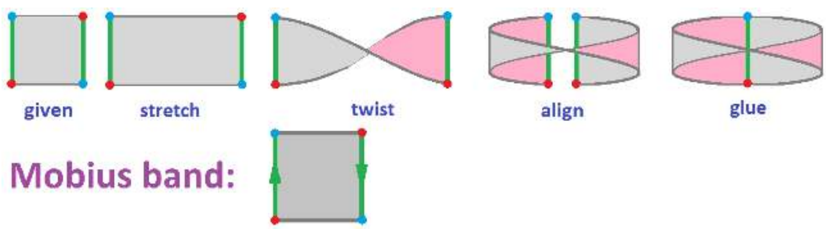
$$q(x,y) := x,\ x,y \in \mathbf{R}.$$

Two harder examples of gluing follow.

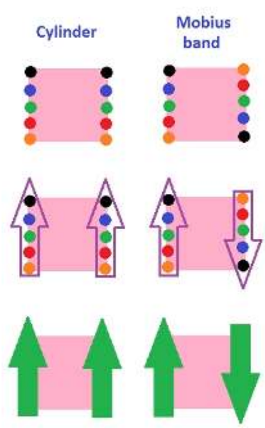
Example 2.4 (surfaces). One can glue the two opposite edges of the square to create a cylinder:



If you twist the edge before gluing, you get the Möbius band:



Now, in order to properly interpret this literal gluing in terms of quotients, we need to recognize that, mathematically, only *points* can be identified to each other. In other words, points are glued *pairwise*:



Once this is understood, we can sometimes see a pattern of how the points are glued to each other and think of this process as gluing two *whole edges*, as long as they are oriented properly! These orientations of the edges are shown with arrows that have to be aligned before being glued.

We have here:

$$X = [0,1] \times [0,1] = \{(x,y): \ x \in [0,1], y \in [0,1]\}.$$

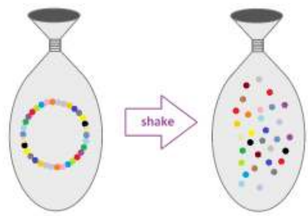
And the equivalence relation for the cylinder is given by:
- $(x,y) \sim (u,v)$ if $y = v$ and $x, u = 0$ or 1,

or simply:
- $(0,y) \sim (1,y)$.

The equivalence relation for the Möbius band is given by:
- $(x,y) \sim (u,v)$ if $y = 1-v$ and $x, u = 0$ or 1,

or simply:
- $(0,y) \sim (1,1-y)$. □

With all this talk about creating a circle, a torus, etc., we shouldn't fool ourselves into thinking that these are anything more than clues to possible outcomes. Without a topology as a way to specify the relation of proximity, the quotient set will remain a *bag of points*:



2.3 Quotient spaces and identification maps

Even though we start with X assumed to be a topological space, its quotient is, so far, *just a set.* We can't simply assume that the quotient also has "some" topology without losing its link to X.

What we need is a standard way of imparting or inducing a topology on the quotient set $X/_\sim$ from that of X. The idea is similar to the way we chose a topology, *the* relative topology, on a subset of X based on that of X. The crucial difference between these two situations is that the functions we use for this purpose point in the opposite directions:
- for a subset, the inclusion $i_A : A \to X$ is a function *to* X, while
- for a quotient set, the identification function $q : X \to X/_\sim$ is a function *from* X.

In either case, it is the continuity of the new function that we desire.

Example 2.5. Note that the identification functions in the examples above are familiar. In the second example, q is the *projection*:

$$q : \mathbf{R}^2 \to \mathbf{R}$$

given by

$$q(x,y) := x.$$

And it is continuous... provided $\mathbf{R}$ is equipped with the Euclidean topology, of course. It's easy to verify that the preimage of an open set is open:

$$q^{-1}\Big((a,b)\Big) = (a,b) \times \mathbf{R}.$$

In the first example, the identification function f is the gluing map:

$$q : [0,1] \to \mathbf{S}^1,$$

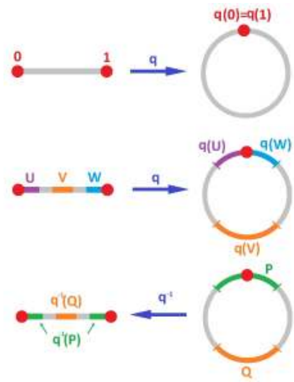
given by

$$q(t) := (\cos(\pi t), \sin(\pi t)), \ \forall t \in [0,1].$$

Once again, this function will be continuous provided we supply the circle with an appropriate topology. Naturally, we want the topology of the new topological space to match the relative topology acquired from the inclusion $\mathbf{S}^1 \subset \mathbf{R}^2$. The standard basis of this topology of the plane consists of disks; therefore, the basis of the topology of the circle should consist of the arcs:

$$A(a,b) := \{(\cos(\pi t), \sin(\pi t)) : \ t \in (a,b)\},$$

where (a,b) is any interval in $\mathbf{R}, a < b$.



Clearly, if $0 \leq a, b \leq 1$, we have

$$q^{-1}\Big(A(a,b)\Big) = (a,b),$$

which is open. What if $0 < a < 1$ while $1 < b < 2$? Then

$$q^{-1}\Big(A(a,b)\Big) = (a,1] \cup [0,b).$$

It is also open relative to $[0,1]$. Once again, we have a continuous function. □

Definition 2.6. Given a topological space X and an equivalence relation $\sim$ on X, the *quotient space* $X/\sim$ is a topological space on the quotient set $X/_{\sim}$ such that

$$U \text{ is open in } X/_{\sim} \text{ if and only if } q^{-1}(U) \text{ is open in } X,$$

where q is the identification function.

To put this differently, if τ is the topology of X then

$$\{q^{-1}(U) : \ U \in \tau\}$$

is the topology of $X/_{\sim}$.

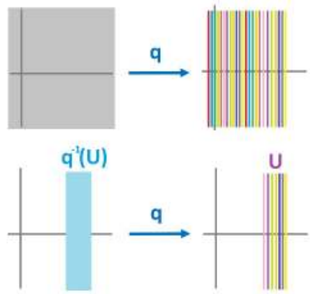
Exercise 2.7. Prove that this collection of sets is indeed a topology.

Exercise 2.8. What if we replace above "topology" with "a basis of neighborhoods"?

As the identification function is now continuous, we may call it the *identification map.*

In the example of the circle above, the preimage of an arc is either an open interval or the union of two half-open intervals at the endpoints.

Example 2.9 (plane). The equivalence classes are the vertical lines and, therefore, the preimage of any subset $U \subset X/_{\sim}$ is made of vertical lines. This set is an open vertical band when the lines in U cut an open interval on the x-axis:



Therefore, the topology on the quotient set is that of $\mathbf{R}$. □

Exercise 2.10. Consider the two equivalence relations for $\mathbf{R}^2$ below: $(x, y) \sim (a, b)$ if

- (a) $x^2 + y = a^2 + b$, or
- (b) $x^2 + y^2 = a^2 + b^2$.

Identify these quotient spaces $\mathbf{R}^2/_\sim$ as familiar topological spaces.

As we have seen, the quotient construction may add non-trivial homology classes. It's true for all dimensions except 0:

Theorem 2.11. If X is path-connected then so is $X/_\sim$, for any equivalence relation $\sim$.

Proof. Since the identification function $q : X \to X/_\sim$ is both continuous and onto, the new space $X/_\sim = \operatorname{Im} q$ is path-connected as the image of a path-connected topological space. ■

Theorem 2.12. If X is compact then so is $X/_\sim$, for any equivalence relation $\sim$.

Exercise 2.13. Prove the theorem.

We discover then that, instead of thinking in terms of the topological space of equivalence classes, it may be more convenient to consider a familiar topological space as a candidate and then prove that they are homeomorphic. In fact,

any function defined on a topological space can be thought of as an identification map.

We just need to find an appropriate topology for the target space of this function. Indeed, suppose $f : X \to Y$ is an onto function from a topological space X to a set Y. Then we define

$$a \sim b \Longleftrightarrow a, b \in f^{-1}(y), y \in Y,$$

or

$$a \sim b \Longleftrightarrow f(a) = f(b).$$

This is an equivalence relation on X and its equivalence classes are the preimages of points under f. Then the topology on Y is created by the definition from the last subsection.

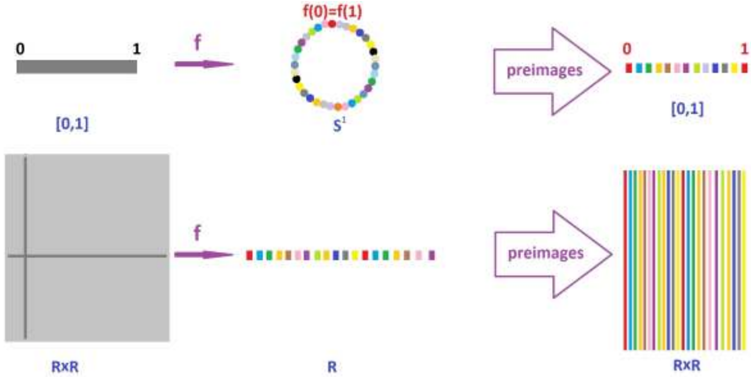
Example 2.14. Let's choose $f = q$ in the two examples of maps from the last subsection,

$$f : [0, 1] \to \mathbf{S}^1$$

and

$$f : \mathbf{R} \times \mathbf{R} \to \mathbf{R}.$$

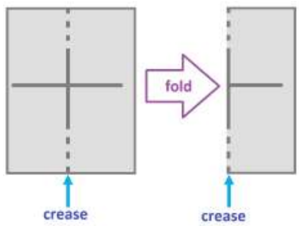
The idea is illustrated below:



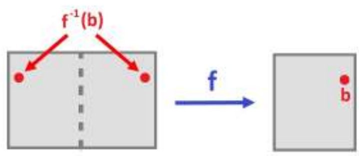
□

Example 2.15 (folding). Fold the xy-plane,

$$f(x,y) := (|x|, y).$$



Now, the folding of the rectangle $[-1,1] \times [0,1]$ is given by the same formula and is continuous as a restriction of f.



□

Not every continuous function, however, is an identification map. The following theorem presents the necessary extra conditions.

Exercise 2.16. Given an identification map $f : X \to Y$ and a subset $A \subset X$, show that the restriction $f\Big|_A : A \to f(A)$ doesn't have to be an identification map.

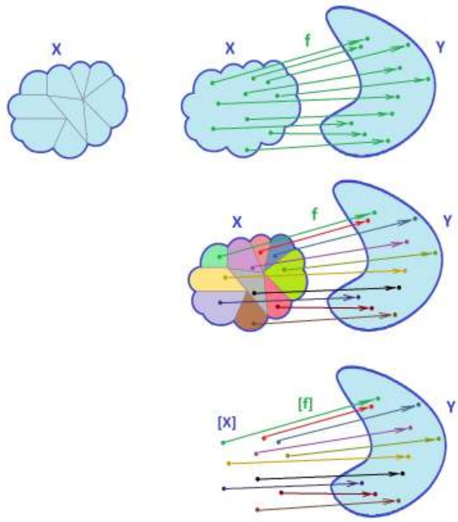
Theorem 2.17. Let $q : X \to X/_\sim$ be the identification map. Suppose Y is a topological space and $f : X \to Y$ is a map that is constant on each equivalence class of X. Then there is a map $f' : X/_\sim \to Y$ such that $f'q = f$. In other words, there is a map for the bottom arrow to make this diagram commutative:

$$\begin{array}{ccc} X & & \\ \downarrow{\scriptstyle q} & \searrow^{f} & \\ X/_\sim & \xrightarrow{f'} & Y \end{array}$$

Exercise 2.18. Prove the theorem.

What happens to *maps* when one or both of the spaces are subjected to the quotient construction?

Suppose first that just the domain of a map $f : X \to Y$, is equipped with an equivalence relation. Then quotient map $[f] : X/_\sim \to Y$ of f is given by $[f]([x]) := f(x)$.



Of course, the new map is well-defined only if f takes each equivalence class to a single point; i.e.,

$$x \sim x' \Longrightarrow f(x) = f(x').$$

Theorem 2.19. Let $f : X \to Y$ be an onto map. If

- f maps open sets of X to open sets in Y, or
- f maps closed sets of X to closed sets in Y,

then f is an identification map.

Theorem 2.20. The quotient map of $f : X \to Y$, when defined, is continuous.

Exercise 2.21. Prove the theorems.

Exercise 2.22. What if, this time, the target space Y has an equivalence relation too? Analyze the possibility of a map $[f] : X \to Y/_\sim$.

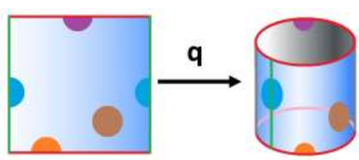
The general case of a map from a quotient space to a quotient space is familiar from algebra. Given a map $f : X \to Y$, its *quotient map* $[f] : X/_\sim \to Y/_\sim$ is given by

$$[f]([x]) := [f(x)].$$

Exercise 2.23. (a) When is $[f]$ well-defined? (b) Prove that it is continuous.

2.4 Examples

Example 2.24 (cylinder). Let's consider the cylinder construction and examine what happens to the topology. The results are similar to the example of the circle: the preimage of an open disk under the identification map is either an open disk or the union of two half-disks at the edge.

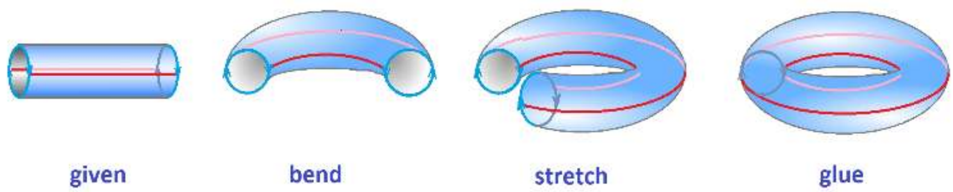


□

Example 2.25 (torus). One can construct the torus $\mathbf{T}^2$ from the cylinder by gluing the top to the bottom:

$$(x,0) \sim (x,1).$$

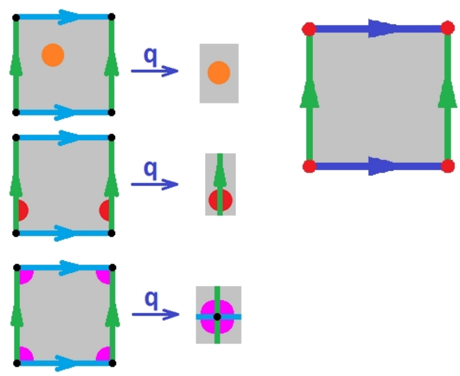
This is how it is visualized:



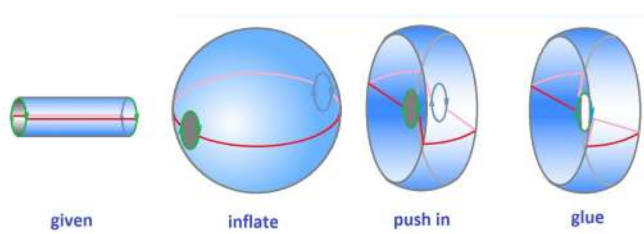
Meanwhile, the equivalence relation for the torus built directly from the square is as follows:

$$(0,y) \sim (1,y) \text{ and } (x,0) \sim (x,1).$$

As the torus is a quotient of the square, one can easily see the three types of neighborhoods created by the gluing:



In the 3d world, there is a different way to glue the edge to itself:

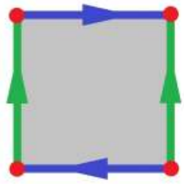


Because the outcome is supposed to be the same, we discover a self-homeomorphism of the torus which is *non-trivial* in the sense that a longitude is mapped to a latitude. Such a homeomorphism would, in a sense, turn the torus inside out. □

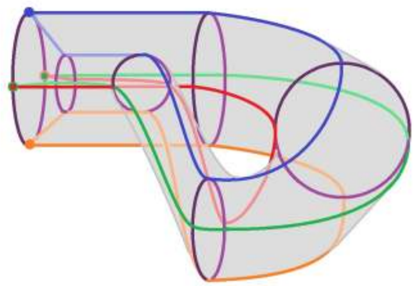
Example 2.26 (Klein bottle). One can get the Klein bottle $\mathbf{K}^2$ from the cylinder by gluing the top to the bottom in reverse:

$$(x,0) \sim (1-x,1).$$

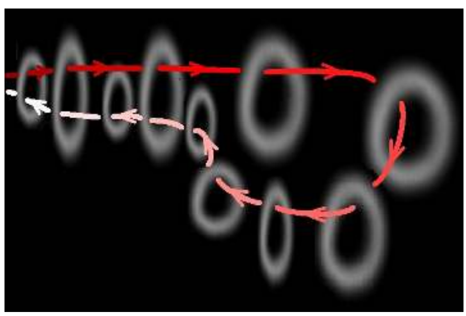
The horizontal arrows point in the opposite directions:



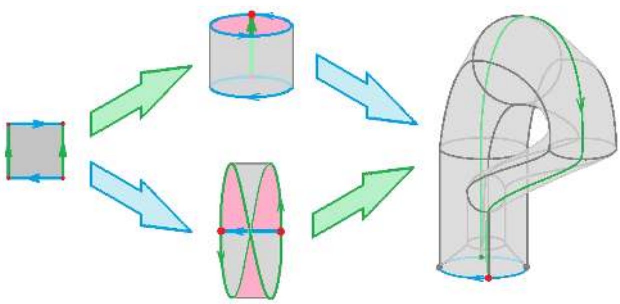
To bring them together, we need to “cut” through the cylinder’s side. This is how it is visualized:



To understand that there is really no self-intersection, one can think of the Klein bottle as a circle moving through space. As an illustration, imagine a smoke ring that leaves your mouth, floats forward, turns around, shrinks, and then floats back in:

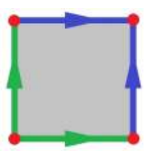


Just as with the torus, there are two ways to construct the Klein bottle from the square:

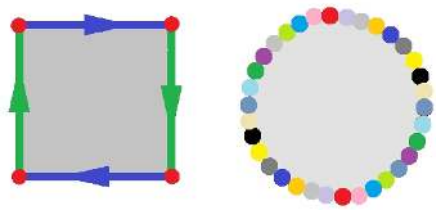


□

Exercise 2.27. Identify this space:



Example 2.28 (projective plane). We make the *projective plane* $\mathbf{P}^2$ from this square too:



One can understand it as if the diametrically opposite, or "antipodal", points on the boundary of the disk are identified. □

Exercise 2.29. Show that we can, alternatively, start with $\mathbf{R}^2 \setminus \{0, 0\}$ and choose the lines through the origin to be the equivalence classes.

Example 2.30 (complex projective plane). It is similar, but the reals replaced with the complex numbers:

$$\mathbf{CP}^n := \mathbf{S}^{2n+1}/_\sim,$$

where

$$z \sim z' \iff z = e^{i\theta} z', \ \exists \theta.$$

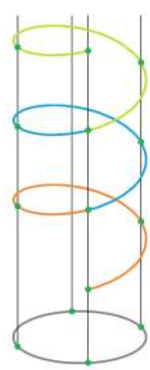
and

$$\mathbf{S}^{2n+1} = \{z \in \mathbf{C}^{n+1} : \ ||z||^2 = 1\}. \qquad \square$$

Exercise 2.31. The projective plane $\mathbf{P}^2$ contains the Möbius band $\mathbf{M}^2$. Find it.

Exercise 2.32. What happens if we identify the antipodal points on the circle $\mathbf{S}^1$?

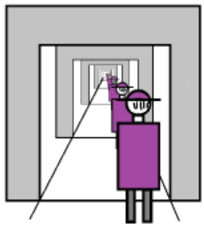
Example 2.33. There are many ways to create a circle. An insightful way is to make it from the line. One just winds the helix, which is $\mathbf{R}$ topologically, around the circle, $\mathbf{S}^1$:



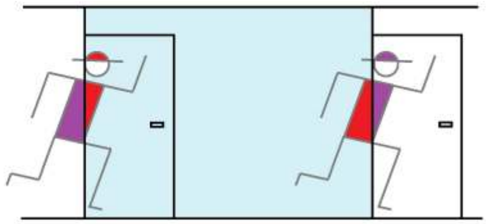
Then the identification map $q : \mathbf{R} \to \mathbf{S}^1$ may be thought of as the restriction of the projection of the 3-space to the plane, or it is given explicitly by

$$q(t) := (\cos(\pi t), \sin(\pi t)), \ t \in \mathbf{R}. \qquad \square$$

Example 2.34. Suppose we have a room with two doors and suppose as you exit through one door you enter through the other (it's the same door!). If you look through this doorway, this is what you see:



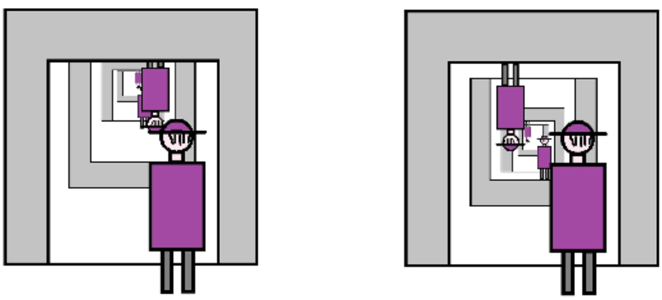
The reason is that, in this universe, light travels in circles... If you run fast enough, an outside observer might see (parts of) you at two different places at the same time:



To produce this effect, we identify the front wall with the back wall. □

Exercise 2.35. Describe this situation as a quotient of the cube.

Exercise 2.36. What if, as you exit one door, you enter the other – but upside down? Which of the two below is the correct view?



Exercise 2.37. What if the front wall is turned 90 degrees before it is attached to the back wall? Sketch what you'd see.

Exercise 2.38. If you've seen two mirrors hung on the opposite walls of the room, you know what they show: you see your face as well as the back of your head, with this pattern repeated indefinitely. Use quotients to achieve this effect without mirrors. Hint: you need a *compact* space.

Exercise 2.39. Repeat the last exercise for the second, web-cam, image above.

Exercise 2.40. What if the room had n walls? What if we have mirrors on all four walls of the room? Five walls, n walls? Four walls, ceiling, and floor?

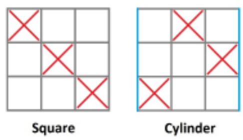
Exercise 2.41. Interpret the Pasting Lemma about the continuity of a piecewise defined function $f : X \to Y$

$$f(x) = \begin{cases} f_A(x) & \text{if } x \in A, \\ f_B(x) & \text{if } x \in B, \end{cases}$$

in terms of the quotient construction.

Exercise 2.42. Suppose our sphere $\mathbf{S}^2$ is the unit sphere in $\mathbf{R}^3$. Define a map $f : \mathbf{S}^2 \to \mathbf{R}^4$ by $f(x, y, z) := (x^2 - y^2, xy, xz, yz)$. Use f to construct an embedding of the projective plane $\mathbf{P}^2$ to $\mathbf{R}^4$.

Exercise 2.43. Explore the winning combinations and the strategies for playing tic-tac-toe on the main surfaces:



Note: See the files online.

Exercise 2.44. Which space do we obtain if we collapse the boundary of the Möbius band to a point? Hint: the boundary is a circle.

Exercise 2.45. Describe the space resulting from the ring if we identify antipodal pairs of points on the outer circle and also identify the antipodal pairs of points on the inner circle.

2.5 Topology of pairs of spaces

One can eliminate a homology class from a space by making it a boundary: add a path between two points, a disk over a hole, etc. Now, can we eliminate it by "removing" rather than "adding" things?

The idea is to simply collapse an appropriate subset, such as a circle surrounding the hole, to a point. We have seen collapses implemented by means of the quotient construction in algebra and, in fact, we used it to introduce homology. Now, we shall see that the idea is also applicable in topology.

If a subset is collapsed to a point, there is only one non-trivial equivalence class!

Notation: As we use the same notation as in algebra, the difference in the meaning needs to be emphasized:

- algebra: X/A is given by the equivalence relation: $x \sim y \Longleftrightarrow x - y \in A$;
- topology: X/A is given by the equivalence relation: $x \sim y \Longleftrightarrow x, y \in A$.

Exercise 2.46. The above example of projecting the plane onto the x-axis can be understood algebraically as the plane modulo the y-axis. Sketch this quotient if understood topologically.

With this simple equivalence relation, we can build a lot.

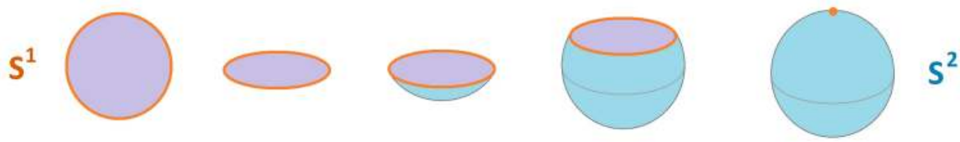
Example 2.47. The circle is still made from the segment, modulo its boundary:

$$[0, 1]/_{\sim} = [0, 1]/\{0, 1\} = \mathbf{I}/\partial\mathbf{I} \approx \mathbf{S}^1.$$

The sphere is made from the disk:

$$\mathbf{B}^2/\mathbf{S}^1 \approx \mathbf{S}^2,$$

as illustrated below:



This is a disk modulo its boundary. □

Exercise 2.48. Prove this is a homeomorphism. Hint: use the stereographic projection.

More generally, one can construct the n-sphere from the n-ball by collapsing its boundary to a point:

$$\mathbf{S}^n = \mathbf{B}^n/\mathbf{S}^{n-1}.$$

Note that the n-sphere is constructed topologically, not as a subset of $\mathbf{R}^{n+1}$.

Definition 2.49. Given X and a subset A, the equivalence relation defined by

$$x \sim y, \ \forall x, y \in A,$$

produces a *quotient set* denoted by

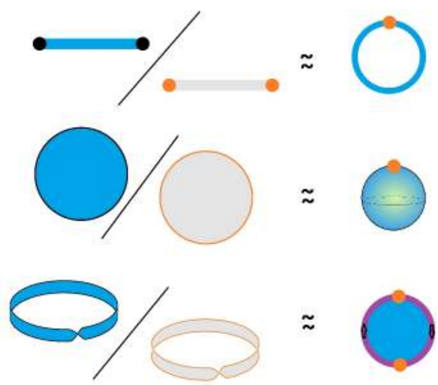
$$X/A.$$

It is read "X modulo A".

Once again, we need to keep in mind that this is just a special kind of a quotient space. The equivalence classes are

- $\{x\}$ for $x \in X \setminus A$, and
- A.

Example 2.50. A few simple examples of how collapses are used to build familiar spaces:



□

Exercise 2.51. Sketch the torus with a longitude, and then a latitude, collapsed.

Exercise 2.52. What is $\mathbf{R}^2/\mathbf{Z}^2$?

One can use this new idea to develop topology of *pairs* (X, A), where X is a topological space and A is its subset.

Notation: The topology of the pair is that of X with A collapsed:

$$(X, A) := X/A,$$

and, further, its homology is

$$H(X, A) := H(X/A).$$

We can interpret all the previous development in light of this new approach by assuming that

$$X = (X, \emptyset).$$

The main advantage of this approach is the ease of dealing with *maps*. Suppose there is a map between two spaces

$$f : X \to Y,$$

both of which are supplied with equivalence relations. Recall that one can construct the quotient map $[f]$ of f as a map of quotients

$$[f] : X/_{\sim} \to Y/_{\sim}$$

by setting

$$[f]([a]) := [f(a)].$$

Clearly, it is possible only if each equivalence class on X is mapped into an equivalence class on Y:

$$f([a]) \subset [f(a)].$$

In other words, we have to require the following:

$$a \sim b \Longrightarrow f(a) \sim f(b).$$

Now, when these two equivalence relations are simply collapses, X/A and Y/B, this requirement becomes much simpler:

$$a \in A \Longrightarrow f(a) \in B,$$

or

$$f(A) \subset B.$$

Another way to put it is that the restriction of f to A with target space B

$$f\Big|_A : A \to B$$

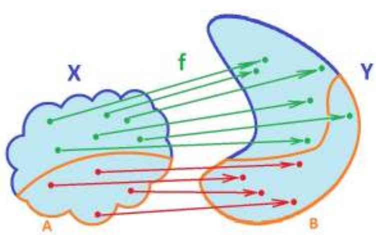
needs to be well-defined. In other words, this diagram is commutative:

$$\begin{array}{ccc} A & \xrightarrow{f|_A} & B \\ \downarrow{\scriptstyle i_A} & \searrow & \downarrow{\scriptstyle i_B} \\ X & \xrightarrow{f} & Y, \end{array}$$

where i_A, i_B are the inclusions.

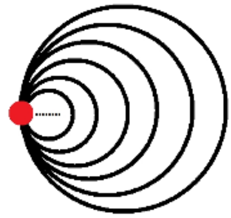
Our interpretation is that f is now a *map of pairs*:

$$f : (X, A) \to (Y, B).$$



Exercise 2.53. Prove that the composition of two maps of pairs is also a map of pairs. Is this a category? Define the chain maps and the homology maps of maps of pairs.

Example 2.54 (Hawaiian earring). The *Hawaiian earring* is the subset of the plane that consists of infinitely many circles with just one point in common:



Specifically, the nth circle is centered at $\left(\frac{1}{n}, 0\right)$ with radius $\frac{1}{n}$, $n = 1, 2, 3, ...$ $\square$

Exercise 2.55. Show that the Hawaiian earring X is path-connected.

Definition 2.56. The *one-point union* of two spaces X, Y is defined to be

$$X \vee Y := \Big(X \sqcup Y\Big)/\{x_0, y_0\},$$

for some $x_0 \in X, y_0 \in Y$.

Exercise 2.57. Under what circumstances is the one-point union independent of the choice of the points?

Exercise 2.58. Let $Y := \mathbf{R}/\mathbf{Z}$ be the quotient of the real line with all integer points identified. It can be understood as an "infinite bouquet of circles". Show that the Hawaiian earring X is *not* homeomorphic to Y.

Exercise 2.59. Explain the topology of $\mathbf{R}/\mathbf{Q}$.

---

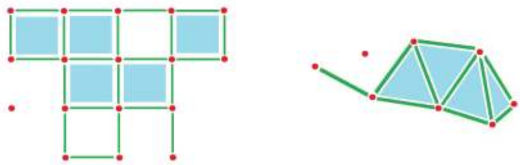
3 Cell complexes

3.1 Gluing cells together

With such a topological tool as the quotient available to us, we can construct anything imaginable starting from almost anything. We would like, however, to learn how to build things from the elementary pieces, *cells*, and do it in a gradual and orderly manner. The point is to be able to construct and compute its homology, and do it in a gradual and orderly manner.

One can see examples of things created from elementary pieces appear in real life, such as these balls sewn (or otherwise formed) from patches of leather:

We already have seen two ways to construct topological spaces from cells. Cubical sets are unions of cubes of various dimensions and the realizations of simplicial complexes are combinations of homeomorphic images of simplices of various dimensions:



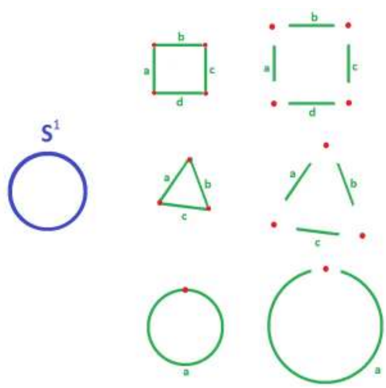
In either of these examples, we orderly glue edges to each other at their endpoints. But, those endpoints are 0-cells. Furthermore, we glue 2-cells to each other along their edges. And, those edges are 1-cells. And so on.

The main difference is in the manner these cells appear. In the case of a cubical complex, cells are subsets of a given Euclidean space, while a simplicial complex is built from data and its cells can then be realized in a Euclidean space.

Exercise 3.1. Note that the cells of cubical complexes are homeomorphic to cubes and squares, while the cells of simplicial complexes are homeomorphic to simplices and triangles. Explain the difference. Hint: what?!

Another difference is the number of adjacent cells each cell has: 4 for a square and 3 for a triangle, which dictates the manner they are attached to each other. We would like to develop a unified approach to such gluing constructions. *Cell complexes* have cells of arbitrary shape with arbitrary number of neighbors.

Example 3.2 (circle). The simplest cubical complex we could use to represent the topological circle consists of four edges arranged in a square. But why should we need four pieces when we know that a circle can be built from a single segment with the endpoints glued together? In fact, the number one example of identification is gluing the endpoints of a segment to create a circle. The latter will have just one 1-cell in contrast to the realizations we have seen, i.e., 4 for a cubical complex and 3 for a simplicial complex:



How the homology is computed is revealing. For the first two cases, we already know the answers:

- cubical: $H_1 =< [a+b+c+d] >\cong \mathbf{R}$;
- simplicial: $H_1 =< [a+b+c] >\cong \mathbf{R}$.

The third case will be much simpler – with so few cells:

- 0-cells: A;
- 1-cells: a.

Certainly, this is not a cubical complex because we can't place these cells on a grid. Nonetheless, the possibility of homological analysis remains because the boundary operator still makes sense:

$$\partial a = A - A = 0.$$

Further, the chain complex is just this:

$$\partial = 0 : C_1 =< a > \to C_0 =< A > .$$

Then the homology H_1 is generated by the homology class of a:

$$H_1 =< [a] >\cong \mathbf{R}.$$

Thus, the result is the same in all the three cases, as expected:

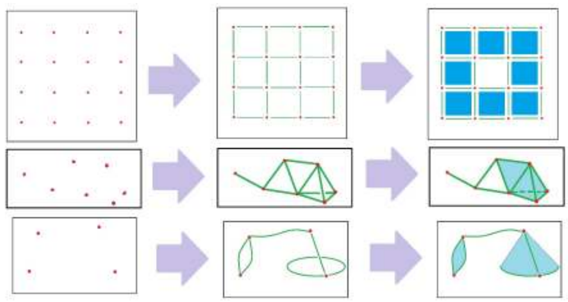
the 1st homology is generated by the 1-cycle that goes once around the hole, clockwise.

It is, in a way, the same cycle! Of course, the group can also be generated by the counterclockwise chain. □

Exercise 3.3. Provide a similar analysis for the sphere.

3.2 Examples and definitions

We are already familiar with gradual and orderly building – via skeleta – of cubical and simplicial complexes:



The only difference is that now we have more flexibility about the cells. Topologically they are *balls* of various dimensions:

Definition 3.4. A (closed) *n-cell* a is a topological space homeomorphic to the closed n-ball

$$\mathbf{B}^n := \{x \in \mathbf{R}^n : \ ||x|| \leq 1\}.$$

The image of the frontier of $\mathbf{B}^n$ in $\mathbf{R}^n$ under this homeomorphism $h : \mathbf{B}^n \to a$ is called the *boundary* of the cell

$$\partial a := h(\mathrm{Fr}(\mathbf{B}^n)).$$

Also, the image of the interior, and the complement of the boundary, is denoted by

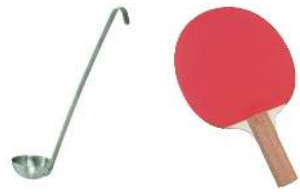
$$\dot{a} := a \setminus \partial a,$$

and may simply be called the "interior" of the cell.

Proposition 3.5. The definition of the boundary is topological in the sense that under homeomorphisms of cells, boundary is mapped to boundary.

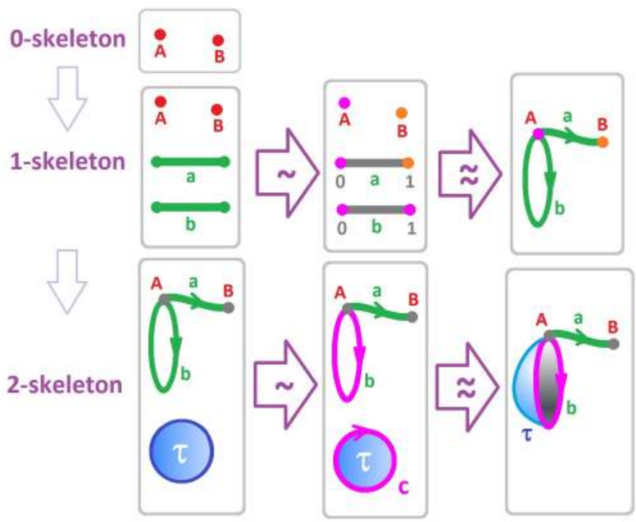
Exercise 3.6. Prove the proposition. Hint: note the difference between the topology of a small neighborhood of a point in the frontier and that of a point in the interior of the ball.

Example 3.7 (ladle). Let's consider a specific example. Suppose we want to build something that looks like a *ladle*, which is the same topologically as a Ping-Pong bat:



We want to build a simple topological space from the ground up, using nothing but cells attached to each other in the order of increasing dimensions. In our box, we have: the parts, the glue, the schematics, and a set of instructions of how to build it.

Here is the schematics:



Let's narrate the instructions. We start with the list K of all cells arranged according to their dimensions:

- dimension 0: A, B;
- dimension 1: a, b;
- dimension 2: τ.

These are the building blocks. At this point, they may be arranged in a number of ways.

Now, the two points A, B are united into one topological space – as the disjoint union. That's the 0-skeleton $K^{(0)}$ of K.

Next, we take this space $K^{(0)}$ and combine it, again as the disjoint union, with all 1-cells in K. To put them together, we introduce an equivalence relation on this set. But, to keep this process orderly, we limit ourselves to an equivalence relation between the vertices (i.e., the elements of $K^{(0)}$, and the boundaries of the 1-cells we are adding. In other words, we identify the endpoints of a and b to the points A and B. This can happen in several ways. We make our choice by specifying the *attaching map* for each 1-cell thought of as a segment $[0, 1]$:

$$f_a : \partial a \to K^{(0)}, \ f_b : b \to K^{(0)},$$

by specifying the values on the endpoints:

$$f_a(0) = A,\ f_a(1) = B,\ f_b(0) = A,\ f_b(1) = A.$$

We use these maps following the *attaching rule*:

$$x \sim y \Longleftrightarrow f_a(x) = y.$$

The result is the 1-skeleton $K^{(1)}$.

The rule we have followed is to choose

an equivalence relation on the last skeleton combined with the boundaries of the new cells.

Next, we take this space $K^{(1)}$ and combine it, again as the disjoint union, with all 2-cells in K. To put them together, we introduce an equivalence relation following the rule above. For this dimension, we identify the edge of τ with a 1-cell a, point by point. This can happen in several ways and we make our choice by specifying the attaching map for the 2-cell:

$$f_\tau : \tau \to K^{(1)}.$$

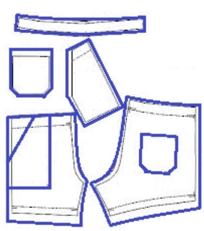
We only need to specify the values on the boundary and we assume that $f_\tau : \partial\tau \to b$ is a homeomorphism. We again use the attaching rule:

$$x \sim y \Longleftrightarrow f_\tau(x) = y.$$

The result is the 2-skeleton $K^{(2)}$, which happens to be the realization of the whole K. □

Exercise 3.8. Present cell complex structures for these two arrangements:


Exercise 3.9. (a) Present a cell complex structure for the sewing pattern below (shorts). (b) Do the same for a pair of breakaway pants.



The formal definition of cell complex does not rely on illustrations. It is inductive.

Definition 3.10. Suppose we have a finite collection of (unrelated to each other) cells K. Suppose C^n denotes the set of all n-cells in K. Next, the 0-skeleton $K^{(0)}$ is defined as the disjoint union of all 0-cells:

$$K^{(0)} = \bigsqcup C^0.$$

Now suppose that the n-skeleton $K^{(n)}$ has been constructed. Then the $(n+1)$-skeleton $K^{(n+1)}$ is constructed as follows. Suppose for each $(n+1)$-cell a, there is an *attaching map*

$$f_a : \partial a \to K^{(n)}.$$

Then the $(n+1)$-skeleton $K^{(n+1)}$ is defined as the quotient of the disjoint union of the n-skeleton $K^{(n)}$ and all the $(n+1)$-cells:

$$K^{(n+1)} = (K^{(n)} \bigsqcup C^{n+1})/_{\sim},$$

with $\sim$ defined by the following condition:

$$x \sim f_a(x) \text{ for all } x \in \partial a \text{ and all } a \in C^{n+1}.$$

The attaching maps must satisfy an extra condition:

$$\text{the image } f_a(\partial a) \text{ is the union of a collection of } n\text{-cells of } K.$$

The resulting topological space is called a *realization* $|K|$ *of cell complex* K. Often, we will use "cell complex K" for both.

Furthermore, we can assume without loss of generality that the attaching map is extended to

$$f_a : a \to K^{(n)},$$

with the only requirement:

- $f_a\Big|_{\dot{a}}$ is a homeomorphism.

In fact, this map may be seen as a restriction of the identification map. Further, with the inclusions of the skeleta into the complex provided, the attaching map can also be thought of as

$$f_a : a \to |K|.$$

Thus, a cell complex K is a combination of the following:

- 1. the collection of cells K,
- 2. the skeleta $K^{(0)} \subset K^{(1)} \subset \ ... \ \subset K^{(N)} = |K|$, and
- 3. the attaching maps $f_a : a \to K^{(n)}$, for each a an n-cell in K.

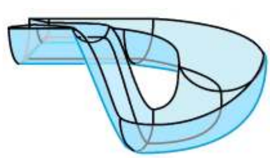
If this was a building, these three would be the inventory, the blueprints, and the daily work orders.

Exercise 3.11. Provide an illustration similar to the one above for a topological space that looks like (a) an air-balloon with a thread hanging, (b) a candy apple.

Exercise 3.12. Find the cell complex representation of the figure eight.

Exercise 3.13. Represent the sphere as a cell complex with *two* 2-cells, list all cells, and describe/sketch the gluing maps.

Exercise 3.14. (a) Find the cell complex representation of the surface acquired by cutting the Klein bottle in half. (b) Use the representation to identify the surface.



Example 3.15 (complex projective plane). First, let

$$\mathbf{S}^{2n+1} := \{z \in \mathbf{C}^{n+1} : \ ||z||^2 = 1\}.$$

Recall:

$$\mathbf{CP}^n := \mathbf{S}^{2n+1}/_{\sim},$$

where

$$z \sim z' \iff z = e^{i\theta} z'.$$

Then the map

$$f_n : \mathbf{B}^{2n} = \{z \in \mathbf{C}^n\} : ||z|| \leq 1\} \to \mathbf{S}^{2n+1},$$

given by

$$f_n(z) := \left(z, \sqrt{1 - ||z||^2}\right),$$

is the attaching map that makes $\mathbf{CP}^n$ into a cell complex.

When $n = 1$, all points are equivalent. Therefore, $\mathbf{CP}^1$ is a point. The identification map $\mathbf{S}^3 \to \mathbf{CP}^1 = \mathbf{S}^2$ is called the *Hopf map*. □

Exercise 3.16. Recall that the (real) projective space is a quotient of the sphere:

$$\mathbf{P}^n := \mathbf{S}^n/_{\sim},$$

where $x \sim -x$. Let $q : \mathbf{S}^n \to \mathbf{P}^n$ be the identification map of the equivalence relation. What space is produced by attaching $\mathbf{B}^{n+1}$ to $\mathbf{P}^n$ by means of q?

3.3 The topology

Let's observe that U is open in cell complex K if and only if for any n-cell $\sigma \in K$, the set $f_\sigma^{-1}(U)$ is open. This condition is equivalent to the following:

- $U \cap \sigma = U \cap f_\sigma(\mathbf{B}^n)$ is open.

The same is true with "open" replaced with "closed".

Proposition 3.17. Cell complex K has the so-called *weak* topology; i.e., U is open in K when $U \cap K^{(n)}$ is open for each n.

Exercise 3.18. Prove the proposition.

Exercise 3.19. Prove that a point is closed.

So, a cell complex is a topological space built from simple topological spaces, cells, via the quotient construction. This approach allows us to ignore the question of whether or not it fits into some Euclidean space and rely on the fact that all properties we discuss are purely topological.

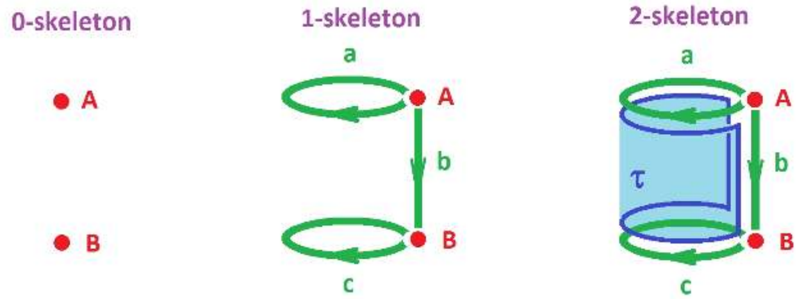
Theorem 3.20. A cell complex is compact.

Proof. It is easy to see that a finite union of compact spaces is compact. So, the initial union X of the cells of the complex is compact. Now $K = X/_{\sim}$, hence K is compact as a quotient of a compact space (indeed, it's the image of a compact space under a continuous map). ■

Exercise 3.21. Below is an object made of two squares connected by four (clockwise) twisted strips. (a) Find its cell complex representation. (b) To what is it homeomorphic? (c) What if some of the strips were twisted the opposite way? (d) Answer these questions for two triangles connected by three twisted strips.

Exercise 3.22. Explain how cubical and simplicial complexes can be understood as cell complexes.

Example 3.23 (cylinder). Let's explore more thoroughly the cell complex structure of the cylinder by explicitly presenting both the skeleta and the gluing maps.



The formulas for the gluing maps for the 1-skeleton are simple. Suppose,

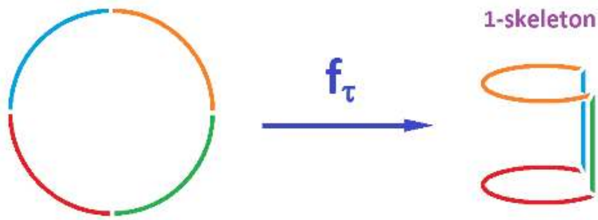
- $a = b = c := [0, 1]$,

then

- $f_a(0) := A,\ f_a(1) := A$;
- $f_b(0) := A,\ f_b(1) := B$;
- $f_c(0) := B,\ f_c(1) := B$.

For the 2-skeleton, we'll use a diagram to present the only gluing map f_τ. Since τ is a disk, we only need to show where in $K^{(1)}$ the points on its boundary $\mathbf{S}^1$ are taken. Below we show the preimages of the 1-cells in $K^{(1)}$:

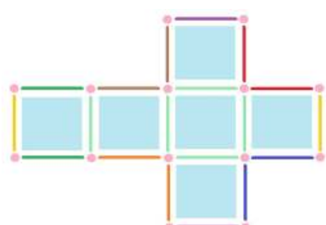
- the preimages of the 0-cells A, B are two points, and
- the preimages of the three 1-cells a, b, c are one arc, two arcs, one arc, respectively.



□

Exercise 3.24. Find other ways to attach a 2-cell to this 1-skeleton.

Exercise 3.25. Build a box by identifying some of the edges of this template:

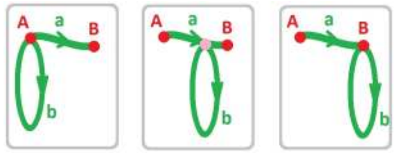


3.4 Transition to algebra

The way we attach the cells to each other is purposeful. Remember we want to be able to compute the boundary operator in just as orderly a fashion as we build the cell complex. This is the reason why we never attach k-cells to other k-cells but – along their boundaries – to the preexisting $(k-1)$-cells:

the image $f_a(\partial a)$ of an n cell is the union of a collection of $(n-1)$-cells of K.

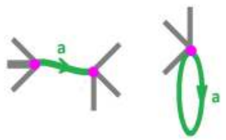
Note that, without this extra constraint, the result may look very different from what we expect:

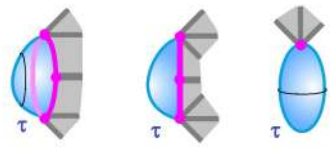


The middle option is unacceptable.

Let's illustrate the acceptable attaching maps.

A 1-cell may be attached to 0-cells as a rope or as a noose:





Meanwhile, a 2-cell may be attached to 1-cells as a soap film, a zip-lock bag, or an air-balloon.

For either dimension, the first of these attaching maps is a homeomorphism and the last is constant. Generally, a cell can be attached along its boundary in a number of ways that include collapses to lower dimensional cells.

The point of the extra constraint on attaching maps is to allow us to proceed directly to algebra as shown below. If the (topological) boundary of an $(n+1)$-cell τ is the union of several n-cells $a, b, c, \ldots$:

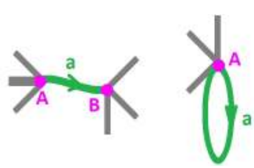
$$\partial \tau = a \cup b \cup c \cup \ldots,$$

then the boundary operator evaluated at τ is some linear combination of these cells:

$$\partial_{n+1}(\tau) = \pm a \pm b \pm c \pm \ldots$$

What are the signs? They are determined by the *orientation* of the cell τ as it is placed in the cell complex and attached to its n-cells. Let's consider this matching in lower dimensions.

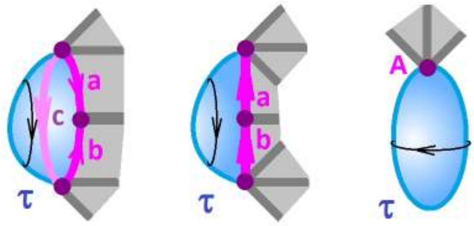
In dimension 1, the meaning of orientation is simple. It is the direction of the 1-cell as we think of it as a parametric curve. Then the boundary is the last vertex it is attached to minus the first vertex.



Here, if the (topological) boundary of the 1-cell a is identified, by the attaching maps, with the *union* of two 0-cells A, B (or just A), while the (algebraic) boundary of a is the *sum* (or a linear combination) of A, B:

$$\begin{array}{lll} f_a(\partial a) & = \{A, B\} & \rightsquigarrow \quad \partial_1(a) = B - A; \\ f_a(\partial a) & = \{A\} & \rightsquigarrow \quad \partial_1(a) = A - A = 0. \end{array}$$

For a 2-cell, a direction is chosen for its circular boundary, clockwise or counterclockwise. As we move along the boundary following this arrow, we match the direction to that of each 1-cell we encounter:

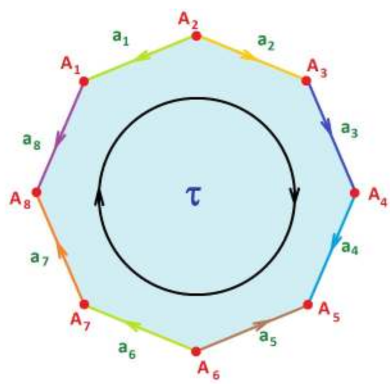


Here, we have three cases:

$$\begin{array}{llcll} f_\tau(\partial\tau) & = a\cup b\cup c & \leadsto & \partial_2\tau & = a-b-c;\\ f_\tau(\partial\tau) & = a\cup b & \leadsto & \partial_2\tau & = -a-b+a+b=0;\\ f_\tau(\partial\tau) & = A & \leadsto & \partial_2\tau & = 0.\end{array}$$

Initially, we can understand the orientation of a cell as an ordering of its vertices, just as we did for simplicial complexes.

Example 3.26. Let's evaluate the boundary operator for this 2-cell, with the orientations of 1-cells randomly assigned.



We have:

$$\partial\tau = -a_1 + a_2 + a_3 + a_4 - a_5 + a_6 + a_7 - a_8 .$$

Further,

$$\begin{array}{lll} \partial(\partial\tau) & = & \partial(-a_1 + a_2 + a_3 + a_4 - a_5 + a_6 + a_7 - a_8)\\ & = & -\partial a_1 + \partial a_2 + \partial a_3 + \partial a_4 - \partial a_5 + \partial a_6 + \partial a_7 - \partial a_8\\ & = & -(A_1 - A_2) + (A_3 - A_2) + (A_4 - A_3) + (A_5 - A_4)\\ & & -(A_5 - A_6) + (A_7 - A_6) + (A_8 - A_7) - (A_8 - A_1)\\ & = & 0.\end{array}$$

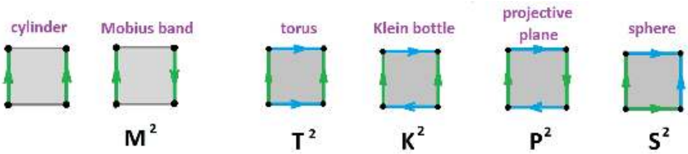
□

Theorem 3.27. The kth homology of a cell complex is fully determined by its $(k+1)$-skeleton.

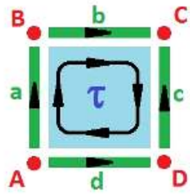
Exercise 3.28. Prove the theorem.

3.5 What we can make from a square

Let's see what we can construct from the square by gluing one or two pairs of its opposite edges. We will compute the homology of these surfaces:



Example 3.29 (square). Here is the simplest cell complex representation of the square (even though the orientations can be arbitrarily reversed):

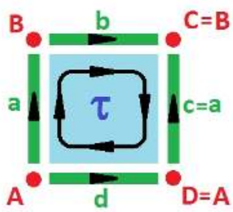


The complex K of the square is:

- 0-cells: A, B, C, D;
- 1-cells: a, b, c, d;
- 2-cells: τ;
- boundary operator: $\partial\tau = a + b + c - d$; $\partial a = B - A, \partial b = C - B$, etc. □

Example 3.30 (cylinder). We can construct the cylinder C by gluing two opposite edges with the following equivalence relation: $(0, y) \sim (1, y)$. The result of this equivalence relation of *points* can be seen as equivalence of *cells*:

$$a \sim c;\ \ A \sim D,\ \ B \sim C.$$



We still have our collection of cells (with some of them identified as before) and only the boundary operator is different:

- $\partial\tau = a + b + (-a) - d = b - d$;
- $\partial a = B - A,\ \partial b = B - B = 0,\ \partial d = A - A = 0.$

The chain complex is

$$\begin{array}{lccccc}
 & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \\
 & <\tau> & \xrightarrow{?} & <a, b, d> & \xrightarrow{?} & <A, B> \\
 & \tau & \mapsto & b - d & & \\
 & & & a & \mapsto & B - A \\
 & & & b & \mapsto & 0 \\
 & & & d & \mapsto & 0 \\
\text{kernels :} & Z_2 = 0 & & Z_1 =< b, d > & & Z_0 =< A, B > \\
\text{images :} & B_2 = 0 & & B_1 =< b - d > & & B_0 =< B - A > \\
\text{quotients :} & H_2 = 0 & & H_1 =< [b] = [d] >\cong \mathbf{Z} & & H_0 =< [A] = [B] >\cong \mathbf{Z}
\end{array}$$

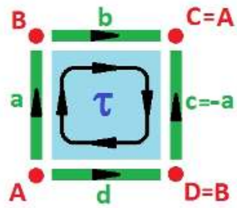
Here, "kernels" are the kernels of the maps to their right, "images" are the images of the maps to their left.

So, the homology is identical to that of the circle! □

Example 3.31 (Möbius band). In order to build the Möbius band $\mathbf{M}^2$, we use the equivalence relation: $(0, y) \sim (1, 1 - y)$. Once again, we can interpret the gluing as equivalence of *cells*, a and c. But this time they are attached to each other with c upside down. It makes sense then to interpret this as equivalence of cells but with a flip of the sign:

$$c \sim -a.$$

Here $-c$ represents edge c with the opposite orientation:



In other words, this is an equivalence of *chains*. Further,

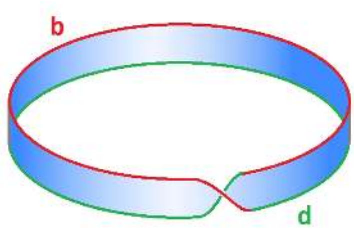
$$A \sim D, \ B \sim C.$$

The boundary operator is:
- $\partial\tau = a + b - (-a) - d = 2a + b - d$;
- $\partial a = B - A, \ \partial b = A - B, \ \partial d = B - A.$

The chain complex is

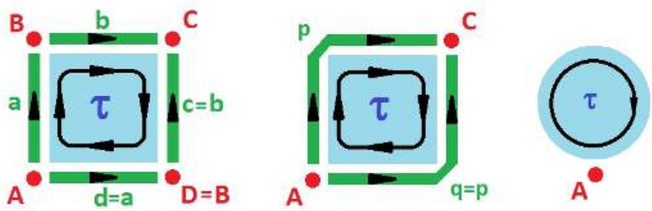
$$\begin{array}{lccccc}
 & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \\
 & < \tau > & \xrightarrow{?} & < a, b, d > & \xrightarrow{?} & < A, B > \\
 & \tau & \mapsto & 2a + b - d & & \\
 & & & a & \mapsto & B - A \\
 & & & b & \mapsto & A - B \\
 & & & d & \mapsto & B - A \\
\text{kernels :} & Z_2 = 0 & & Z_1 =< a + b, a - d > & & Z_0 =< A, B > \\
\text{images :} & B_2 = 0 & & B_1 =< 2a + b - d > & & B_0 =< B - A > \\
\text{quotients :} & H_2 = 0 & & H_1 =< [b - d] >\cong \mathbf{Z} & & H_0 =< [A] = [B] >\cong \mathbf{Z}
\end{array}$$

The bad news is that the homology is the same as that of the cylinder, which means that it doesn't detect the twist of the band. The good news is that the algebra reveals something that we may have missed: the hole in the Möbius band is captured by chain $b - d$, which is the *whole* edge of the band:



□

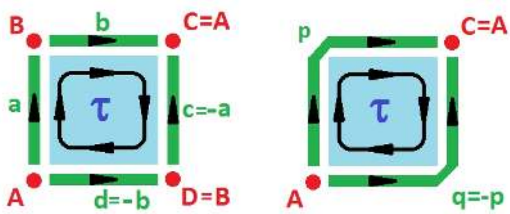
Example 3.32 (sphere). We can either build the sphere as a quotient of the square or, easier, we just combine these pairs of the consecutive edges:



Then we have only two edges left. Even better, the last option gives us a complex with no 1-cells whatsoever! □

Exercise 3.33. Compute the homology for all three.

Example 3.34 (projective plane). The projective plane comes from a certain quotient of the disk, $\mathbf{B}^2/_\sim$, where $u \sim -u$ is limited to the boundary of the disk. It can also be seen as a quotient of the square:



As we see, the edge of the disk is glued to itself, with a twist. Its algebraic representation is almost the same as before: $p \sim -p$. In fact, the presence of a dimension 2 homology class, $2p \sim 0$, causes the homology group to have *torsion*!

Let's compute the homology:

- 2-cells: τ with $\partial\tau = 2p$;
- 1-cells: p with $\partial p = 0$;
- 0-cells: A with $\partial A = 0$.

Then the 1st homology group is computed three ways for these three different rings:

$$\begin{array}{llll} H_1(\mathbf{P}^2;\mathbf{Z}) & =< p > / < 2p > & = \mathbf{Z}/2\mathbf{Z} & \cong \mathbf{Z}_2, \\ H_1(\mathbf{P}^2;\mathbf{Z}_2) & =< p > / < 2p > & =< p > /0 & \cong \mathbf{Z}_2, \\ H_1(\mathbf{P}^2;\mathbf{R}) & = \operatorname{span}(p)/\operatorname{span}(2p) & = \operatorname{span}(p)/\operatorname{span}(p) & \cong 0. \end{array}$$

So, the integer homology detects the twist. So does the binary homology (but it can't tell it from a hole). The real homology doesn't detect the twist. What happens is that, because of the twist of the 2-cell, the image of ∂_2 is generated by $2p$ in all three cases. But the algebra that follows is different. Over $\mathbf{R}$, the set of all real multiples of $2p$ coincides with that of p: $\operatorname{span}(2p) = \operatorname{span}(p)$. As a result, the information about the twist is lost. (The real numbers are for measuring, not counting!)

Another way to see what causes the difference is below:

$$\begin{array}{lllll} \mathbf{Z}: & \partial\tau = 2p & & \Longrightarrow & \text{so what?} \\ \mathbf{Z}_2: & \partial\tau = 2p & \Longrightarrow \ \partial\tau = 0 & \Longrightarrow & \tau \text{ is a cycle, so what?} \\ \mathbf{R}: & \partial\tau = 2p & \Longrightarrow \ \partial\left(\frac{1}{2}\tau\right) = p & \Longrightarrow & p \text{ is a boundary!} \end{array}$$

And here's the rest of the integral homology:

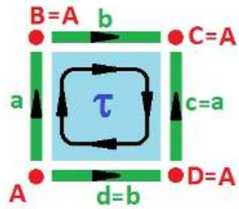
$$H_2(\mathbf{P}^2) = 0, \ H_0(\mathbf{P}^2) = \mathbf{Z}.$$

□

Example 3.35 (torus). What if after creating the cylinder by identifying a and c we then identify b and d? Like this:

$$c \sim a,\ d \sim -b.$$

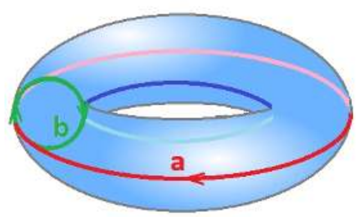
The result is the torus $\mathbf{T}^2$:



Note how all the corners of the square come together in one. Then the chain complex has very few cells to deal with:

$$\begin{array}{lccccc}
 & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \\
 & <\tau> & \xrightarrow{?} & <a,b> & \xrightarrow{?} & <A> \\
 & \tau & \mapsto & 0 & & \\
 & & & a & \mapsto & 0 \\
 & & & b & \mapsto & 0 \\
\text{kernels :} & Z_2 = <\tau> & & Z_1 = <a,b> & & Z_0 = <A> \\
\text{images :} & B_2 = 0 & & B_1 = 0 & & B_0 = 0 \\
\text{quotients :} & H_2 = <\tau> & & H_1 = <a,b> & & H_0 = <[A]>
\end{array}$$

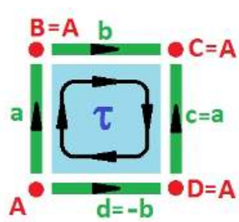
Accordingly, the two tunnels in the torus are captured by a longitude and a latitude:



□

Example 3.36 (Klein bottle). What if we flip one of the edges before gluing? Like this:

$$c \sim a,\ d \sim -b.$$



The corners once again come together in one and the chain complex has very few cells to deal with:

$$\begin{array}{lcccccc}
 & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \\
 & < \tau > & \xrightarrow{?} & < a, b > & \xrightarrow{?} & < A > \\
 & \tau & \mapsto & 2a & & \\
 & & & a & \mapsto & 0 \\
 & & & b & \mapsto & 0 \\
\text{kernels :} & Z_2 = 0 & & Z_1 =< a, b > & & Z_0 =< A > \\
\text{images :} & B_2 = 0 & & B_1 =< 2a > & & B_0 = 0 \\
\text{quotients :} & H_2 = 0 & & H_1 = \{a, b | 2a = 0\} & & H_0 =< [A] >
\end{array}$$

Thus, unlike the two tunnels in the torus, this time there is only one and the other topological feature is a *twist*. □

Exercise 3.37. What if we flip both? Like this:

$$a \sim -c, \ b \sim -d.$$

Exercise 3.38. Rewrite the computations of the homology of all six surfaces, over $R = \mathbf{Z}_p$.

Exercise 3.39. Compute the integral homology, providing all the details, of the sphere with a membrane inside. Indicate the generators.

Let's summarize the results of these computations:

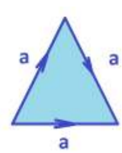
$R = \mathbf{Z}$	$\mathbf{S}^2$	$\mathbf{P}^2$	$\mathbf{T}^2$	$\mathbf{K}^2$
H_0	$\mathbf{Z}$	$\mathbf{Z}$	$\mathbf{Z}$	$\mathbf{Z}$
H_1	0	$\mathbf{Z}_2$	$\mathbf{Z} \times \mathbf{Z}$	$\mathbf{Z}_2 \times \mathbf{Z}$
H_2	$\mathbf{Z}$	0	$\mathbf{Z}$	0

Examining this table tells us a lot about the *topological features* we considered at the very beginning of this study.

- The path-components are clearly visible for all spaces.
- The tunnels are present too.
- The twists are new (1-dimensional) topological features.
- The voids are here.
- There are no voids in the projective plane and the Klein bottle. The reason is that the twist makes the two sides of the surface – inside and outside – indistinguishable and then there is no inside or outside.

Exercise 3.40. Recreate this table for $R = \mathbf{R}$ and $R = \mathbf{Z}_2$.

Exercise 3.41. Sketch the *dunce hat* and compute its homology:

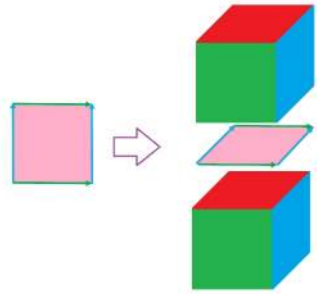


Exercise 3.42. Find a cell complex with the following homology groups:

$$H_0 = \mathbf{Z}, \ H_1 = \mathbf{Z}_2, \ H_2 = \mathbf{Z} \oplus \mathbf{Z}.$$

Exercise 3.43. In each of the above examples, determine whether this surface separates the space; i.e., whether it is one-sided or two-sided.

This is how we can construct 3-dimensional complexes. We start with a cube. First, we glue its top and bottom, with each point identified to the one directly underneath. Then we cut the cube into two by a horizontal square.



Then we consider the relations listed above to identify the edges of this square, just as we did for the torus, etc. These relations tell us how the opposite faces of the cube are glued together.

Exercise 3.44. Compute the homology produced by this construction, when the square is glued to itself according to each of the above examples.

3.6 The nth homology of the n-dimensional balls and spheres

Cell complexes use very few cells in comparison to cubical and simplicial complexes but the trade-off is the complexity of some of the concepts we will need. In order to develop a complete homology theory for cell complexes, we have to add one missing part: we would need to figure out the meaning of *orientation* of cells, beyond dimension 2. With that, we would be able to define the boundary operator $\partial : C_n \to C_{n-1}$ and to prove the property that makes all of this work: $\partial\partial = 0$. We will omit these steps and rely, informally, on what we have already developed.

The issue of orientation has been previously developed in full for the following cases:
- K is a cubical complex, and
- K is a simplicial complex.

In addition, the issue is simply moot when
- K is any cell complex but the ring of coefficients is $R = \mathbf{Z}_2$.

For the general cell complexes and arbitrary coefficients, there are two simple enough cases that we can handle. They appear in the highest dimension $n = \dim K$. For an n-cell σ, we have:
- $\partial_n(\sigma) = 0$ when there are no $(n-1)$-cells in $\partial\sigma$;
- $\partial_n(\sigma) = \pm s$ when s is the $(n-1)$-cell in K that forms the boundary $\partial\sigma$ of σ, and the sign is "+" when there are no other n-cells adjacent to s.

We use these two cases to compute the homology of the n-ball $\mathbf{B}^n$ and the $(n-1)$-sphere $\mathbf{S}^{n-1}$, $n > 2$.

Example 3.45 (balls). We represent $\mathbf{B}^n$ as a cell complex as follows.

Cells:
- n-cells: σ,
- $(n-1)$-cells: a,
- 0-cells: A.

The boundary operator:
- $\partial\sigma = a$,
- $\partial a = 0$,
- $\partial A = 0$.

The chain complex and the computation of homology are below:

$$\begin{array}{lccccccccc}
 & C_n & \xrightarrow{\partial} & C_{n-1} & \xrightarrow{\partial} & C_{n-2} & \to \dots \to & C_0 \\
 & <\sigma> & \xrightarrow{\cong} & <a> & \xrightarrow{0} & 0 & \to \dots \to & <A> \\
 & \sigma & \mapsto & a & \mapsto & 0 & & A \\
\text{kernels :} & Z_n = 0 & & Z_{n-1} =<a> & & Z_{n-2} = 0 & \dots & Z_0 =<A> \\
\text{images :} & B_n = 0 & & B_{n-1} =<a> & & B_{n-2} = 0 & \dots & B_0 = 0 \\
\text{quotients :} & H_n = 0 & & H_{n-1} = 0 & & H_{n-2} = 0 & \dots & H_0 =<A> & \square
\end{array}$$

Example 3.46 (spheres). The sphere $\mathbf{S}^{n-1}$ has the same cell complex representation as the ball except the n-cell is missing.

The cells:
- $(n-1)$-cells: a,
- 0-cells: A.

The boundary operator:
- $\partial a = 0$,
- $\partial A = 0$.

The chain complex and homology are:

$$\begin{array}{lccccccccc}
 & C_n & \xrightarrow{\partial} & C_{n-1} & \xrightarrow{\partial} & C_{n-2} & \to \dots \to & C_0 \\
 & 0 & \xrightarrow{0} & <a> & \xrightarrow{0} & 0 & \to \dots \to & <A> \\
 & & & a & \mapsto & 0 & & A \\
\text{kernels :} & Z_n = 0 & & Z_{n-1} =<a> & & Z_{n-2} = 0 & \dots & Z_0 =<A> \\
\text{images :} & B_n = 0 & & B_{n-1} = 0 & & B_{n-2} = 0 & \dots & B_0 = 0 \\
\text{quotients :} & H_n = 0 & & H_{n-1} =<a> & & H_{n-2} = 0 & \dots & H_0 =<A> & \square
\end{array}$$

In higher dimensions, the algebra tells us about things that we can't see. For example, when the boundary of an n-cell is glued to a single point, a new topological feature is created, an n-dimensional void.

The following summarizes our results.

Theorem 3.47.
- $H_k(\mathbf{B}^n) = 0$ for $k \neq 0$, and $H_k(\mathbf{B}^n) = \mathbf{R}$, $n = 0, 1, 2, \dots$;
- $H_k(\mathbf{S}^n) = 0$ for $k \neq 0, n$, and $H_0(\mathbf{S}^n) = H_n(\mathbf{S}^n) = \mathbf{R}$, $n = 2, 3, \dots$

Exercise 3.48. Solve the problem for $n = 1, 2$.

3.7 Quotients of chain complexes

Next, we explore how to compute the chain complex and the homology groups of a cell complex in the way it is built – gradually.

Example 3.49 (circle). Recall how we computed the homology of the circle. Its chain complex is

$$\partial = 0 \colon C_1 =<a> \to C_0 =<A> .$$

Then it follows that

$$H_1 =<a> \cong \mathbf{Z},\ H_0 =<A> \cong \mathbf{Z}.$$

The cell complex is built from the segment via the quotient construction, point-wise or cell-wise. But we have also seen examples when these identifications are interpreted algebraically, chain-wise.

Let's follow this alternative route. We assume that we initially have the (cubical) complex K of the segment:

- 0-cells: A, B,
- 1-cells: a,
- boundary operator: $\partial a = B - A$.

Its chain complex is

$$\partial : C_1(K) =< a >\to C_0(K) =< A, B > .$$

Now to build the circle, we identify A and B, $A \sim B$, but this time not as points but as 0-chains. The outcome of this construction is not the quotient of topological space $|K|$ or even of cell complex K but the quotient of the chain group $C_0(K)$! The result is easy to present and compute algebraically:

$$C_0(K)/_{\sim} =< A, B > / < A - B >=< A, B | A = B >\cong< A > .$$

The new boundary operator is the quotient of the old and it is trivial:

$$\partial' = q\partial = 0 : C_1(K) =< a >\to C_0(K)/_{\sim} =< A > .$$

Here q is the identification function of this equivalence relation. It immediately follows that

$$H_0 =< A > .$$

Meanwhile, since $C_1 =< a >$, we have the homology:

$$H_1 = \ker \partial_1 =< a >\cong \mathbf{R}.$$

That the hole!

The results are the same as that of the original analysis. □

Things are more complicated in higher dimensions.

Example 3.50 (cylinder). Let's make the cylinder from the square, algebraically.

The complex K of the square is:

- 0-cells: A, B, C, D;
- 1-cells: a, b, c, d;
- 2-cells: τ;
- boundary operator: $\partial(\tau) = a + b - c - d$; $\partial(a) = B - A$, $\partial(b) = C - B$, $\partial(c) = C - D$, etc.

We construct a chain complex for the cylinder by identifying a pair of opposite 1-cells:

$$a \sim c.$$

This produces the quotients of the chain groups of K since a and c are their generators.

Now the quotients of the boundary operators between these quotients are to be considered:

$$[\partial_n] : C_n(K)/_{\sim} \to C_{n-1}(K)/_{\sim}.$$

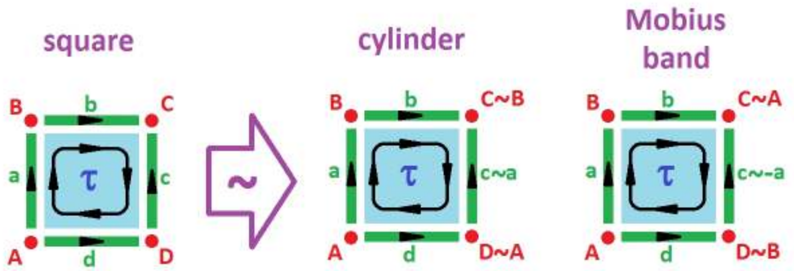
As was discussed previously, the quotient of an operator is well-defined only if the operator maps equivalence classes to equivalence classes; i.e., this diagram commutes:

$$\begin{array}{ccc}
C_n(K) & \xrightarrow{\partial_n} & C_{n-1}(K) \\
\downarrow q & & \downarrow q \\
C_n(K)/_{\sim} & \xrightarrow{[\partial_n]} & C_{n-1}(K)/_{\sim}
\end{array}$$

Here, q is the identification function. So, for $n = 1$, the equivalence of edges forces the equivalence of vertices:

$$A \sim D, B \sim C.$$

The relations are illustrated below:



We compute the quotients of the chain groups:

$$\begin{array}{l} C_2(K)/_{\sim} =< \tau >; \\ C_1(K)/_{\sim} =< a,b,c,d > / < a-c >=< [a],[b],[d] >; \\ C_0(K)/_{\sim} =< A,B,C,D > / < A-D, B-C >=< [A],[B] > . \end{array}$$

Then we compute the values of the quotients of the boundary operators:

$$\begin{array}{llll} [\partial_2](\tau) & = [\partial_2(\tau)] & = [a+b-c-d] & = [b-d]; \\ [\partial_1]([a]) & = [\partial_1(a)] & & = [B-A]; \\ [\partial_1]([b]) & = [\partial_1(b)] & = [C-B] & = 0; \\ [\partial_1]([d]) & = [\partial_1(d)] & = [D-A] & = 0. \end{array}$$

Now, these operators and groups form a chain complex and we compute its homology:

$$\begin{array}{lccccc} & < \tau > & \xrightarrow{[\partial_2]} & < [a],[b],[d] > & \xrightarrow{[\partial_1]} & < [A],[B] > \\ \text{kernels}: & Z_2 = 0 & & Z_1 =< [b],[d] > & & Z_0 =< [A],[B] > \\ \text{images}: & B_2 = 0 & & B_1 =< [b-d] > & & B_0 =< [B-A] > \\ \text{quotients}: & H_2 = 0 & & H_1 =< [b]=[d] >\cong \mathbf{Z} & & H_0 =< [A]=[B] >\cong \mathbf{Z} \end{array}$$

The results are the same as before. □

Exercise 3.51. Verify the claims made in the above example:
- the quotient operators are well-defined;
- the square diagram above is commutative for all n;
- the sequence of quotient groups is a chain complex.

Exercise 3.52. Provide such a homology computation for the Möbius band.

The approach is applicable to any cell complex K. As we build the complex K skeleton-by-skeleton:

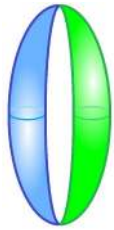
$$K^{(0)} \rightsquigarrow K^{(1)} \rightsquigarrow ... \rightsquigarrow K^{(N)} = K,$$

we also build the chain complex for K step-by-step as follows.

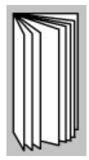
$$C(K^{(0)}) \rightsquigarrow C(K^{(1)}) \rightsquigarrow ... \rightsquigarrow C(K^{(N)}) = C(K).$$

For each $n = 1, 2, , N$, the attaching maps yield equivalence relations. Each n-cell σ to be added to $K^{(n)}$ is treated as a generator to be added to $C(K^{(n)})$ because its boundary $(n-1)$-cells are already present. Then the current chain groups (and the boundary operators) of K are modified via quotient modulo subgroups generated by these boundary cells of σ.

Exercise 3.53. Use this approach to compute the homology of the *double banana*:



Exercise 3.54. Use this approach to compute the homology of the "book with n pages":



---

4 Triangulations

4.1 Simplicial vs. cell complexes

Let's compare:

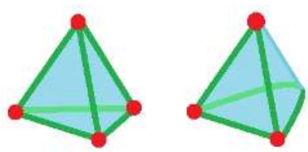
• *simplicial complexes*: cells are homeomorphic to points, segments, triangles, tetrahedra, ..., n-simplices;

• *cell complexes*: cells are homeomorphic to points, closed segments, disks, balls, ..., closed n-balls.

Since these are homeomorphic, the difference lies elsewhere.

Note: The cells in a cubical complex may be thought of as *open*, i.e., homeomorphic to open balls, while the cells in cell (and simplicial) complexes are *closed*, i.e., homeomorphic to closed balls. The reason is that a cubical complex may be built as the union of a collection of subsets of a Euclidean space, while a cell complex is built via the quotient construction, which always requires some points to be shared.

As we know, simplicial complexes are cell complexes with certain constrains on their cells and the cells' faces. In particular, an n-simplex τ in a simplicial complex K has $n+1$ faces, $\sigma < \tau$, each of which is an $(n-1)$-simplex, illustrated on the left:



This is why, for a cell complex to be a simplicial complex, it has to satisfy the following:

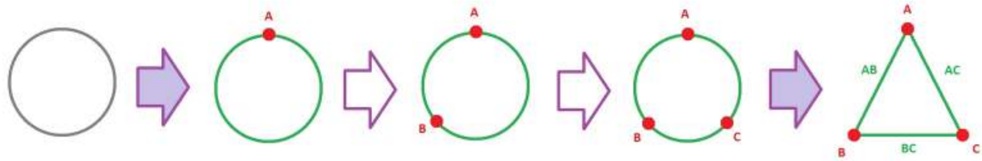
$$\text{the boundary of an } n\text{-cell consists of exactly } n+1 \ (n-1)\text{-cells.}$$

To see how this condition may be violated, consider the cell complex on the right and simply count the boundary cells:

• the 3-cell has only 3 faces;

- the bottom 2-cell has only 2 edges.

In addition, consider the simplest cell complex representation of the circle (one 0-cell and one 1-cell), below. The 1-cell has only one endpoint:



This isn't the only way things can go wrong though. If we add an extra vertex that cuts the 1-cell in two, the above condition is satisfied. Yet, this still isn't a simplicial complex, as the two new cells share *two* endpoints. In other words, two vertices A,B are connected by two distinct edges: $AB \neq AB$. Finally, cutting one of the edges one more time produces a familiar simplicial complex.

Any cell complex can be turned into a simplicial complex in this fashion. The construction is called *subdivision* understood as cutting the cells into smaller cells until certain conditions are met. Notice that the subdivision procedure is reminiscent of the subdivision of intervals in the definition of the definite integral via the Riemann sums.

Theorem 4.1. A cell complex K is a simplicial complex if for each of its cells $\tau\in K$, there is such a homeomorphism $h_\tau:\tau\to S_\tau$ of τ to a geometric simplex S_τ that

- (1) the complex contains all faces of each simplex:

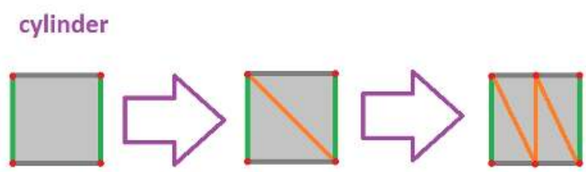
$$\tau\in K, s<S_\tau \Longrightarrow h_\tau^{-1}(s)\in K,$$

- (2) two simplices can only share a single face:

$$\tau,\sigma\in K \Longrightarrow h_\tau(\tau\cap\sigma)<h_\tau(\tau).$$

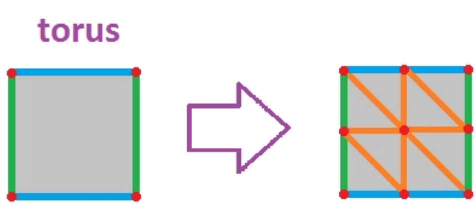
Definition 4.2. A representation of a cell complex K as a simplicial complex K', i.e., $|K|\approx|K'|$, is called its *triangulation*.

Example 4.3 (cylinder). The familiar representation of the cylinder isn't a triangulation simply because the 2-cell is a square not a triangle.



As we see, cutting it in half diagonally doesn't make it a triangulation because there is an edge still glued to itself. Adding more edges does the job. □

Example 4.4. A triangulation of the torus is given:



□

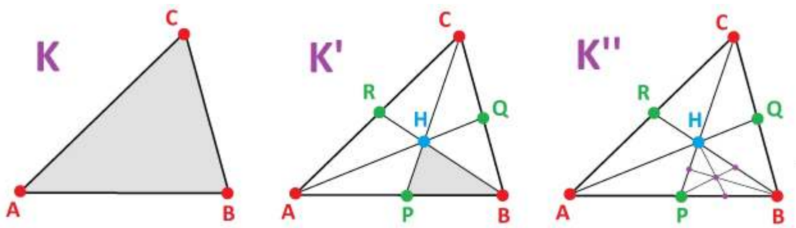
Exercise 4.5. Find a triangulation for the rest of the six main surfaces.

Of course, a given cell complex can have many triangulations. The complexes don't have to be *isomorphic*: there is no bijective simplicial map between them.

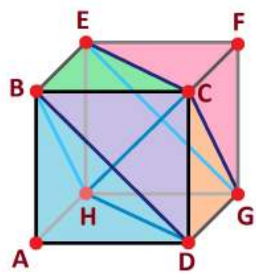
Instead of this *ad hoc* way of discovering triangulations, we can apply a method that always works. This method is a generalization of the barycentric subdivision that can be applicable to cell complexes. First, we want to cut every n-cell into a collection of n-simplices. For each cell $\tau \in K$,

- we remove τ and all of its boundary cells (except for the vertices) from K;
- we add a vertex V_τ to K, and then
- we create the new cells spanned by V_τ and each boundary cell a of τ.

The result is a new cell complex K'.



If we can be sure that there are no identifications of the boundary cells of ABC (in K), the K' would have the structure of a simplicial complex. Yet, it is possible that this new triangular cell is still attached to itself in the original complex K; for example, it's possible that $P = Q$. In that case, we have two edges, $PH = PQ$, connecting the same two vertices. The second barycentric subdivision K'' (of all cells) solves this problem.



Exercise 4.6. (a) List the simplices in the above triangulation of the cube. (b) Provide a sketch and the list of the simplices for the triangulation that starts with adding a vertex in the middle of the cube.

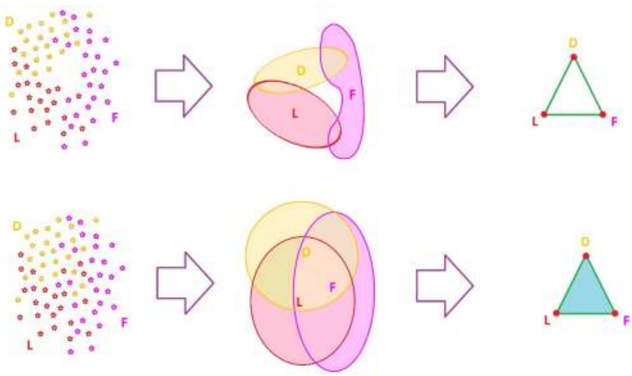
As we have seen, cell complexes provide more *economical* (i.e., with fewer cells) representations of some familiar objects than simplicial complexes. In light of this fact, if we need to compute the homology of a specific cell complex, we would never try to triangulate it. Yet, simplicial complexes are *simpler* in the way the cells are attached to each other. (In fact, every simplex is just a list of its vertices.) As a result, the proofs are simpler too.

Any representation of a topological space as a realization of a simplicial complex is also called a *triangulation*. Triangulable spaces are called *polyhedra*. Since we assumed that all our simplicial complexes are finite, below all polyhedra are assumed to be *compact*.

4.2 How to triangulate topological spaces

Cell complexes are "compact" representations of topological spaces, convenient for computing homology. But how do we find this representation when there isn't even a cell complex but only a topological space, i.e., a set with a specified collection of open subsets? What kind of topological space is a polyhedron?

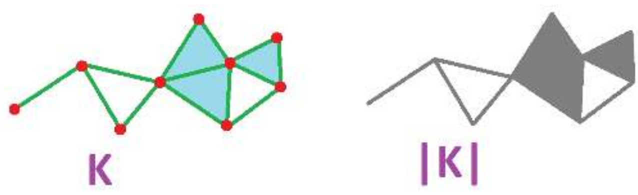
Example 4.7 (garden). Let's recall how, in the very beginning of the book, we were able to understand the layout of a field of flowers of three kinds, say, D, L, and F, based only on the smells. The conclusion was, if we can detect only the three types, each of the three pairs, and no triples, then there must be a bare patch.



In other words, the homology of the field is that of the circle. □

The implicit assumptions that make this work are:
- the patches cover the whole field,
- the patches are open, and
- both the patches and their intersections are acyclic.

With that idea in mind, let's consider the topology of realizations of simplicial complexes:

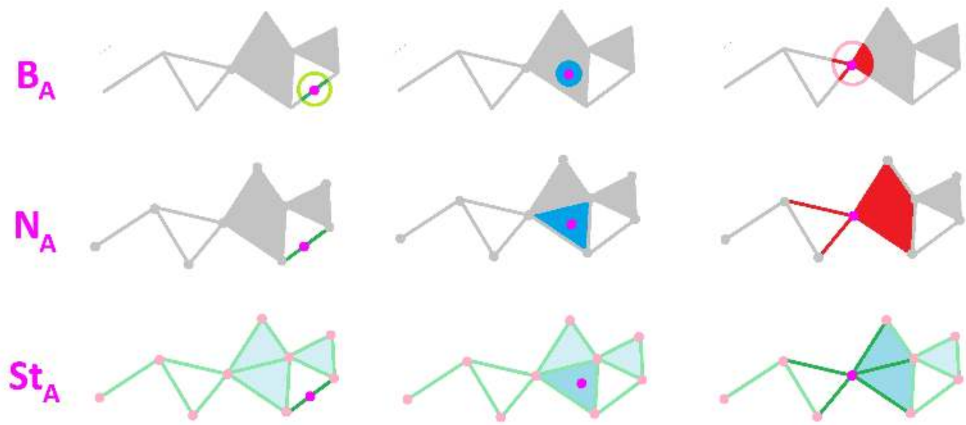


The topology of a realization comes from the topology of each simplex modulo the gluing procedure. What makes each complex special is the way the simplices are attached to each other. Meanwhile, the gluing never affects the topology of the simplex itself.

In particular, every point in an *open* cell has a Euclidean neighborhood relative to that cell:

Proposition 4.8. For every point $A \in \dot{\sigma}$ with $\dim \sigma = n$, there is a neighborhood B_A relative to σ such that $B_A \approx \mathbf{R}^n$.

Beyond the interior of a simplex, the neighborhoods are more complex but still made of the Euclidean neighborhoods of the adjacent cells. But what if we aren't interested in these "small" open sets but in "large" open sets? We choose the latter to be unions of the interiors of simplices:



Definition 4.9. Given a simplicial complex K and a vertex A in K, the *star* of A in K is the collection of all simplices in K that contain A:

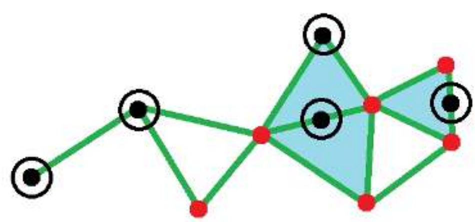
$$\mathrm{St}_A = \mathrm{St}_A K := \{\sigma \in K : A \in \sigma\}.$$

The *open star* is the union of the interiors of all these cells:

$$N_A = N_A K := \bigcup\{\dot{\sigma} : \sigma \in \mathrm{St}_A\}.$$

(Remember, $\dot{A} = A$.)

Exercise 4.10. The star of *any* point in $|K|$ is defined the same way. Sketch the stars of the circled points in the image below:



Exercise 4.11. If we add to St_A the boundary cells of all of its elements, the result is a simplicial complex (a "closed star"). Prove that it is acyclic.

Even though every open cell is open in its closed cell, it doesn't have to be open in $|K|$. We can see that in the above examples. We can't then jump to the conclusion that the open star N_A is open. But it is.

Theorem 4.12. N_A is an open subset of $|K|$ for any vertex A in K.

Proof. The complement of the open star is the union of finitely many closed cells. Indeed:

$$K \setminus \mathrm{St}_A = \{\sigma \in K : A \notin \sigma\}.$$

Hence,

$$|K| \setminus N_A = \bigcup\{\sigma : \sigma \notin \mathrm{St}_A\}.$$

But closed cells are closed as continuous images of compact spaces. Therefore, the complement is closed. ■

In other words, N_A is a neighborhood of A, which justifies the notation.

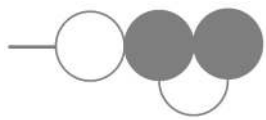
Exercise 4.13. Show that the result will still hold even if K isn't finite (such as $\mathbb{R}$), provided an infinite cell complex is properly defined. Hint: think "locally finite".

We use the above statement to define of the topology of $|K|$.

This is the key step.

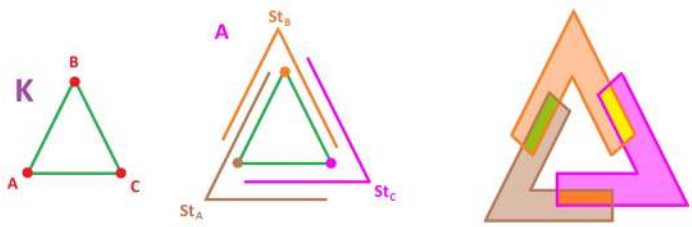
Corollary 4.14. $\{N_A : A \in K^{(0)}\}$ is an open cover of $|K|$.

Now, we want to learn how to construct a simplicial complex from this open cover of $|K|$, and do it in such a way that the result is the same as the original complex K. Keep in mind that in its realization the structure of the simplicial complex is lost, as in the image below, which is homeomorphic to the one above.

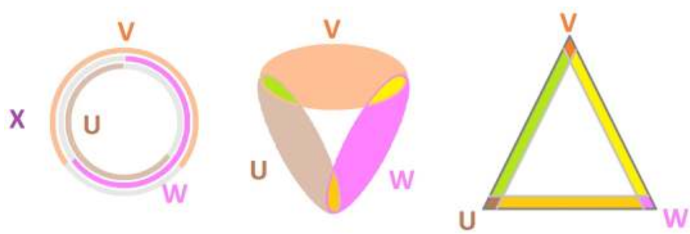


The example below suggests a way to approach the problem.

Example 4.15 (circle). Let's consider the circle as a simple enough example. We start with the simplest triangulation of the circle – a simplicial complex K with three edges and three vertices. Then the stars of vertices consist of two edges each. It's an open cover.



Now, suppose X is the (topological) circle. A homeomorphism of $|K|$ and X creates an open cover of X that consists of three overlapping open arcs:



Let $\gamma := \{U, V, W\}$ be this open cover of X. These sets came from the stars of the three vertices of K, i.e., $\text{St}_A, \text{St}_A, \text{St}_A$, respectively. The cover seems vaguely *circular*, but how do we detect that by simply looking at the data?

The sets are now devoid of all geometry or topology. Therefore, the only thing they may have in common is their *intersections*.

Let's make this more specific. These open sets have non-empty *pairwise* intersections but the only *triple* intersection is empty. Why is this important? Notice the pattern:

- $U \cap V \neq \emptyset \iff N_A \cap N_B \neq \emptyset \iff A, B$ are connected by an edge;
- $V \cap W \neq \emptyset \iff N_B \cap N_C \neq \emptyset \iff B, C$ are connected by an edge;
- $W \cap U \neq \emptyset \iff N_C \cap N_A \neq \emptyset \iff C, A$ are connected by an edge;
- $U \cap V \cap W = \emptyset \iff N_A \cap N_B \cap N_C = \emptyset \iff A, B, C$ are *not* connected by a face.

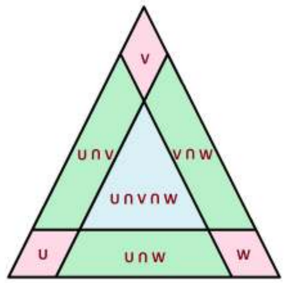
Following this insight, we forget for now about complex K and construct a new simplicial complex L based entirely on how the elements of this cover γ intersect.

We let

$$L^{(0)} := \{U, V, W\}.$$

These are the vertices of complex L. The 1-simplices of L come from the following rule: for every pair $G, H \in \gamma$,

$$GH \in K \iff G \cap H \neq \emptyset.$$



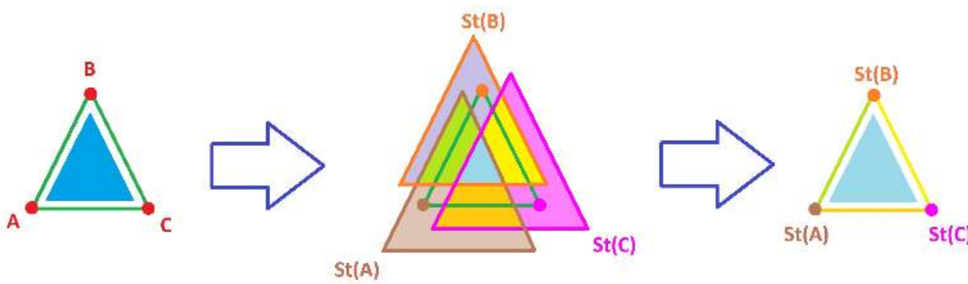
Further, the 2-simplices of L come from the following rule: for every triple $G, H, J \in \gamma$,

$$GHJ \in K \iff G \cap H \cap J \neq \emptyset.$$

There aren't any.

We ended where we started! Indeed, L is *isomorphic* to K. □

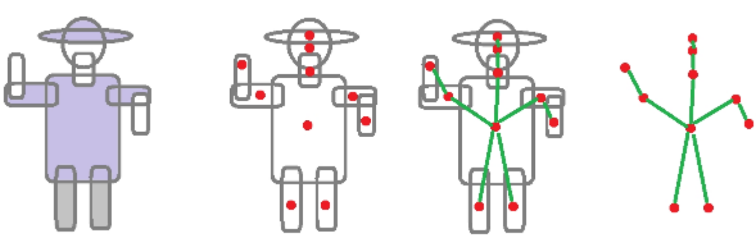
Example 4.16. This construction applied to the solid triangle is illustrated below:



□

Exercise 4.17. Carry out this construction for spaces representing these letters: C, H, P, T.

We can now follow this logic to create simplicial complexes from *any* open cover of *any* topological space. But, since the word "skeleton" is already taken, we are forced to use "nerve" for the output of this construction:



Definition 4.18. Given any collection of subsets α of a set X, the *nerve* of α is the simplicial complex $\mathcal{N}_\alpha$ with:

- vertices corresponding to the elements of the cover, and
- n-simplices corresponding to every non-empty intersection of $n+1$ elements of the cover.

In other words, the abstract simplicial complex $\mathcal{N}_\gamma$ is defined on the set α by the rule: given $A_0, A_1, ..., A_n \in \alpha$, we have

$$\sigma = A_0 A_1 ... A_n \in \mathcal{N}_\alpha \text{ if and only if } A_0 \cap A_1 \cap \ ... \ \cap A_n \neq \emptyset.$$

We will also refer to the realization of $\mathcal{N}_\alpha$ as the nerve.

4.3 How good are these triangulations?

Let's match this definition to what happens in all simplicial complexes.

Theorem 4.19 (Star Lemma). Suppose $A_0, A_1, ..., A_n$ are vertices in complex K. Then these vertices form a simplex

$$\sigma = A_0 A_1 ... A_n$$

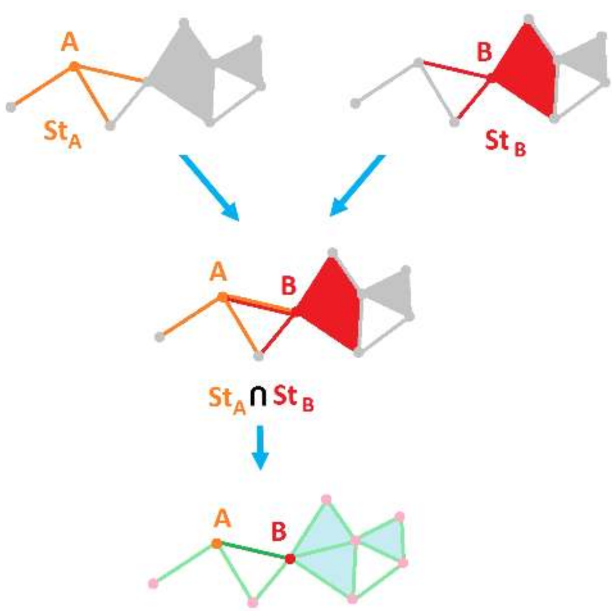
in K if and only if

$$\bigcap_{k=0}^{n} N_{A_k} \neq \emptyset.$$

Proof. To get started, for 2-simplices we have:

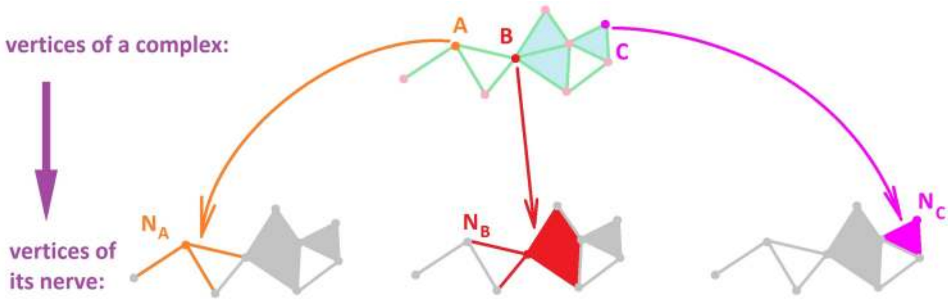
$$\begin{aligned} N_A \cap N_A &= \{\sigma \in K : A \in \sigma\} \cap \{\sigma \in K : B \in \sigma\} \\ &= \{\sigma \in K : A \in \sigma, B \in \sigma\}. \end{aligned}$$

This set is non-empty if and only if K contains AB. ■



Exercise 4.20. Provide the rest of the proof.

The construction of the simplicial complex of the nerve is fully represented by this "nerve map":



Theorem 4.21. The nerve N of the cover of a simplicial complex K that is comprised of the stars of all vertices of K is an abstract simplicial complex isomorphic to K via the simplicial map $h : K \to N$ defined by

$$h(A) := \mathrm{St}_A\,.$$

Proof. We follow the construction of the cover. First, list all the elements of the cover; they correspond to the vertices trivially:

$$\mathrm{St}_A \longleftrightarrow A.$$

The rest follows from the lemma. ■

Exercise 4.22. Show that the nerve of a cubical complex a cubical complex doesn't have to be a cubical complex?

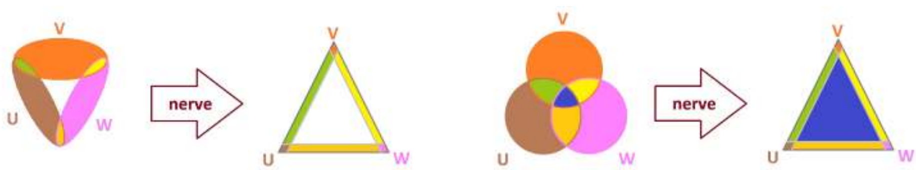
The examples demonstrate that the realization of the nerve of a cover doesn't have to be homeomorphic to the space:

$$\mathcal{N}_\alpha \not\approx X,$$

but maybe their homology groups coincide:

$$H(\mathcal{N}_\alpha) \cong H(X)?$$

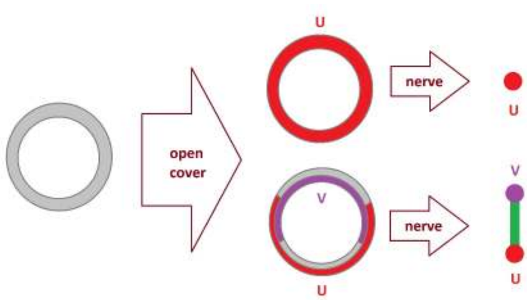
Example 4.23. We have seen some examples that support this idea:



The complexes are, respectively, for the circle and the disk:

$$\{U, V, W, UV, VW, WU\},\ \{U, V, W, UV, VW, WU, UVW\},$$

and their homology groups match those of the original spaces. Examples below show what can go wrong, for the circle:



The complexes are, respectively:

$$\{U\},\ \{U, V, UV\},$$

and their homology groups don't match those of the circle. □

Exercise 4.24. Find more examples, both positive and negative.

What makes the difference for the last two examples?

In the first example, the only element of the cover isn't homeomorphic to the segment as one would expect, but to the circle. In the second example, the two elements of the cover are homeomorphic to the segment but their intersection isn't as it's not path-connected. In either case, the issue is that these sets aren't *acyclic*.

Exercise 4.25. Consider the cover of the sphere $\mathbf{S}^2$ by two open sets obtained by removing the north pole and then the south pole:

$$\alpha := \{\mathbf{S}^2 \setminus \{N\}, \mathbf{S}^2 \setminus \{S\}\},$$

and show how to fix it so that the nerve is homeomorphic to the sphere.

The issue is resolved by the following theorem that we will accept without proof (see Rotman, *An Introduction to Algebraic Topology*, p. 154).

Theorem 4.26 (Nerve Theorem). Let K be a (finite) simplicial complex and α be a cover of $|K|$ that consists of the closed stars of its barycentric subdivision. Suppose the finite intersections of the elements of α are acyclic. Then the realization of the nerve of α has homology isomorphic to that of K:

$$H(\mathcal{N}_\alpha) \cong H(K).$$

Exercise 4.27. Prove the claim about the homology of the flower field in the above example.

Exercise 4.28. Find an example of a complex for which the conclusion of the theorem fails only because the *triple* intersections aren't acyclic.

Either of these two theorems can be seen as a sequence of transitions that ends where it starts. We represent them schematically.

The first theorem states:

- simplicial complex $K \longrightarrow$ collection σ of stars of vertices of $K \longrightarrow$ open cover α of $|K| \longrightarrow$ nerve $\mathcal{N}_\alpha$ of $\alpha \xrightarrow{\cong}$ simplicial complex K;

Meanwhile, the second theorem approaches but falls short of the following ideal:

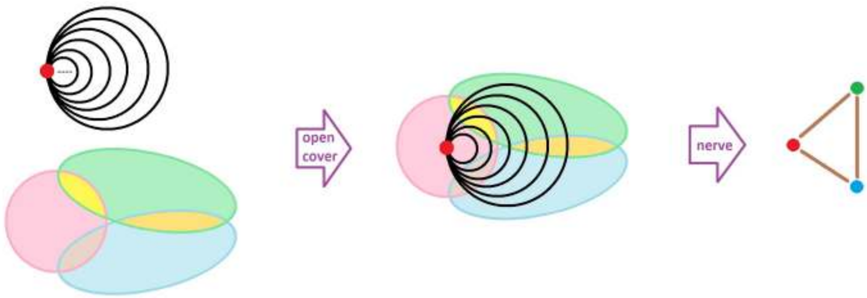
- topological space $X \longrightarrow$ open cover α of $X \longrightarrow$ nerve $\mathcal{N}_\alpha$ of $\alpha \longrightarrow$ realization $|\mathcal{N}_\alpha|$ of $\mathcal{N}_\alpha \xrightarrow{\approx}$ topological space X.

Exercise 4.29. Test this plan for: a point, two points, figure eight, the sphere, the torus, the Möbius band.

The examples and the theorems suggest that "refining" the open cover improves the chance that this plan will work. However, nothing will help the spaces that can't be represented as realizations of *finite* complexes.

Example 4.30 (Hawaiian earring). Recall that the Hawaiian earring is the subset of the plane that consists of infinitely many circles with just one point in common: nth circle is centered at $(1/n, 0)$ with radius $1/n$, $n = 1, 2, 3, ...$, in $\mathbf{R}^2$.

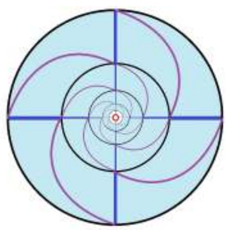
Let's consider possible covers and their nerves. If we choose three open sets small enough, we can capture the first circle:



We can also get the second too, with smaller open sets. But there are infinitely many of them! In general, one of the open sets, W, must contain $(0,0)$. Then, for large enough N, W will contain every nth circle with $n > N$. Therefore the holes in these smaller circles can't be captured by the nerve.

This failure to triangulate is not caused by just having infinitely many circles. Compare this space to the same set with an alternative topology. This alternative, the "bouquet of circles" $\mathbf{R}/\mathbf{Z}$, *can* be easily triangulated as long as we are allowed to use infinitely many simplices. □

Example 4.31. An attempt to triangulate the punctured disk, $\mathbf{B}^2 \setminus \{(0,1)\}$, will cause a similar problem:



□

Exercise 4.32. For a simplicial complex K and a simplex C in K define the star of C as the union of the interiors of all simplices in K that contain C. Prove or disprove: The set of all stars in a finite complex form a basis of topology.

Exercise 4.33. Find an open cover of the n-sphere $\mathbf{S}^2$ the nerve of which is homeomorphic to $\mathbf{S}^n$.

4.4 Social choice: ranking

We will consider the topology of the space of choices. In this subsection, we consider a simple setting of *ranking*, i.e., ordering multiple alternatives. The alternatives may be candidates running for elections, movie recommendations, and other votes of personal preferences.

Suppose the number n of alternatives/candidates is fixed:

$$A = A_n := \{1, 2, ..., n\}.$$

Every voter x has a preference of alternative $j \in A$ over alternative $i \in A$ expressed as follows:

$$i <_x j.$$

Then a complete *ranking vote* of x is a strict ordering of the alternatives. It may look like this:

$$1 <_x 2 <_x 3 <_x ... <_x n.$$

The totality of all possible votes is the set of these orderings

$$O_n := \{(i_1, i_2, ..., i_n) : \ i_k \in A_n, i_k \neq i_l\}.$$

This is the *space of choices* of our social choice problem.

Exercise 4.34. Show that O_n has the group structure of $\mathcal{S}_n$, the group of permutations of n elements.

In particular, for $n = 3$ there are 6 orderings:

$$O_3 = \{123, 132, 231, 321, 213, 312\}.$$

This order of elements of O_n is completely arbitrary and so would be any other. Is there a *meaningful* way to arrange them on the plane? Maybe a circle?.. What we are looking for is a *topology* for O_n!

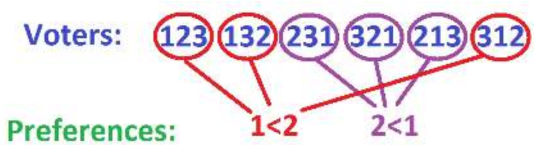
To find an appropriate one, we start with an observation. Let's fix two alternatives $i, j \in A$, $i < j$, and note that if voter x votes $i <_x j$ then we expect a "close enough" to x voter y to also choose $i <_y j$. We write:

$$i <_x j \Longrightarrow i <_y j.$$

This is just a way to informally express the idea that voting preferences should vary from person to person in a *continuous* fashion. The aggregate of all choices is then a continuous function F from the space of voters to the space of pairwise preferences. We will assume that there is exactly one voter for each ordering so that the space of voters is O_n, with the topology we are to choose. Then we have

$$F_{ij} : O_n \to Y := \{(i, j), (j, i) : \ i, j = 1, 2, ..., n, \ i \neq j\}.$$

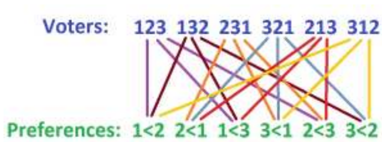
For $n = 3$, this function is illustrated below:



Since Y is discrete, the points in this space are open and, therefore, the preimages of points under F_{ij} are open too. There are only two:

$$\begin{aligned} U_{ij} &:= F_{ij}^{-1}\Big((i, j)\Big) &= \{x \in O_n : \ i <_x j\}, \\ U_{ji} &:= F_{ij}^{-1}\Big((j, i)\Big) &= \{x \in O_n : \ j <_x i\}. \end{aligned}$$

The complexity comes from the fact that there are many pairs (i, j), $i < j$, each of them produces such a pair of sets, and either of these sets is supposed to be open under F_{ij}:



We pool this information by defining the set of all *pairwise rankings*,

$$\alpha := \{U_{ij} : \ i, j \in A, i \neq j\}.$$

It is an open cover of O_n.

Exercise 4.35. Sketch α for $n = 3$.

Exercise 4.36. (a) Show that α is not a basis of neighborhoods. (b) Describe the basis of neighborhoods of this topology by taking finite intersections of the elements of α (that's why α is called a "pre-basis").

Definition 4.37. For a given n, the nerve $R_n := \mathcal{N}_\alpha$ of the cover α of pairwise rankings is a simplicial complex called the *complex of rankings* (or complex of rankings).

Notice that that while we've built simplicial complexes from data before, it is very different this time. Even though it is the orderings we care about, we didn't make them the *vertices*, as usual, of the complex but rather its *simplices*, in fact its n-simplices!

Example 4.38. For $n = 2$, there are only two choices and only two orderings, each an open set:

$$\{12\} = U_{12},\ \{21\} = U_{21}.$$

They don't intersect. Therefore,

$$R_2 = \mathcal{N}_\alpha = \{U_{12}, U_{21}\}$$

is the complex of two disconnected vertices. □

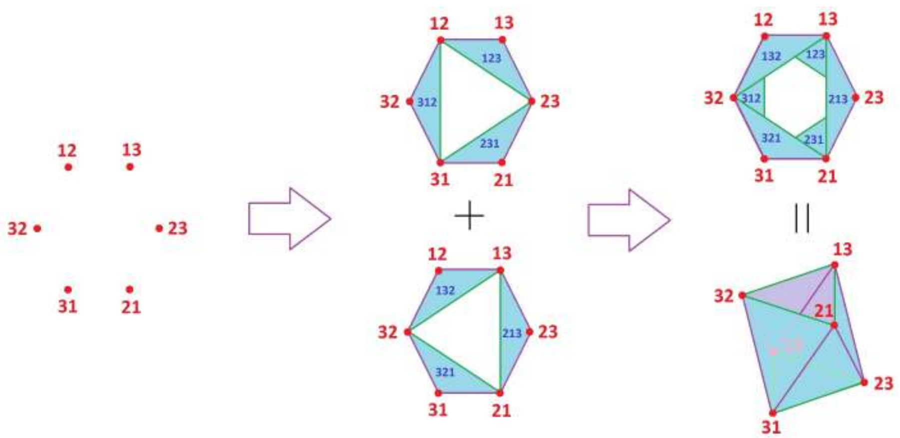
Example 4.39. For $n = 3$, we have

$$O_3 = \{123, 132, 231, 321, 213, 312\}.$$

The cover α consist of these six sets:

$$\begin{array}{ll} U_{12} = \{123, 132, 312\}, & U_{21} = \{213, 231, 321\}, \\ U_{23} = \{231, 213, 123\}, & U_{32} = \{321, 312, 132\}, \\ U_{13} = \{132, 123, 213\}, & U_{31} = \{312, 321, 231\}. \end{array}$$

These are the vertices of the nerve $\mathcal{N}_\alpha$ and they are marked below with simply "12","21", etc.:



Then, we add edges to the vertices by looking at the pair-wise intersections: $U_{ij} \cap U_{km} \neq \emptyset \Leftrightarrow \{i, j\} \cap \{k, m\} \neq \emptyset$. Then, we add faces to the complex $R_3 = \mathcal{N}_\alpha$ by looking at the triple intersections. We follow this idea: are there any voters who vote simultaneously $1 < 2, 1 < 3, 2 < 3$? Yes, they vote: $1 < 2 < 3$. Thus, we have a 2-simplex 123 connecting vertices $12, 13, 23$. But, are there any voters who vote simultaneously $1 < 2, 2 < 3, 3 < 1$? No, because this would be a circular, impossible vote: $1 < 2 < 3 < 1$. Therefore, there is no 2-simplex connecting vertices $12, 13, 23$. The end result is an octahedron with two, opposite to each other, faces missing. □

Exercise 4.40. Sketch a star of R_4.

Let's summarize that we know so far.

Theorem 4.41.

• (1) The simplices of highest dimension of $\mathcal{N}_\alpha$ correspond to the orderings, i.e., the elements of O_n.

• (2) The complex $R_n = \mathcal{N}_\alpha$ is the union of these simplices (as subcomplexes of the simplex of all possible votes).

• (3) The dimension of this complex is

$$\dim R_n = \frac{(n-2)(n+1)}{2}.$$

Proof. (3) Let's first observe that, since there are only n alternatives, there are at most $\frac{n(n-1)}{2}$ pairwise combinations with no repetitions, unless we reverse the order. Now, suppose there is a simplex σ with

$$\dim \sigma = \frac{(n-2)(n+1)}{2} + 1 = \frac{n(n-1)}{2}.$$

This means that σ has more vertices than available pairs. Therefore, there are at least two vertices of σ that are marked with the same pair of alternatives but reversed: ij and ji. Such a simplex would represent an impossible vote. That contradicts the way R_n is constructed. ■

Exercise 4.42. Prove parts (1) and (2).

One can guess that

$$H_1(R_3) = \mathbf{Z}.$$

Exercise 4.43. Prove this statement by a direct computation.

We have a more general result:

Theorem 4.44. The homology of the complex of orderings of n elements is that of the $(n-2)$-sphere:

$$H(R_n) = H(\mathbf{S}^{n-2}).$$

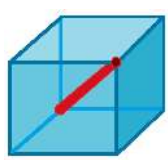
The idea of the proof is to find a familiar space and then show that it has an open cover with the nerve isomorphic to $R_n = \mathcal{N}_\alpha$. We choose as such space the n-dimensional Euclidean space with the diagonal cut out:

$$M := \mathbf{R}^n \setminus \Delta,$$

where

$$\Delta := \{(u_1, ..., u_n) \in \mathbf{R}^n : \ u_i = u_j\}.$$

It is illustrated below for dimension 3:



Consider the cover of M

$$\beta := \{\mathbf{R}^n_{ij} : \ 1 \le i \ne j \le n\}$$

that consists of all half-spaces:

$$\mathbf{R}^n_{ij} := \{(u_1, ..., u_n) \in \mathbf{R}^n : \ u_i < u_j\}.$$

Exercise 4.45. Prove that

$$\mathcal{N}_\beta \cong \mathcal{N}_\alpha,$$

and then finish the proof of the theorem by applying the Nerve Theorem from the last subsection.

4.5 The Simplicial Extension Theorem

Recall that a simplicial map

$$g : K \to L$$

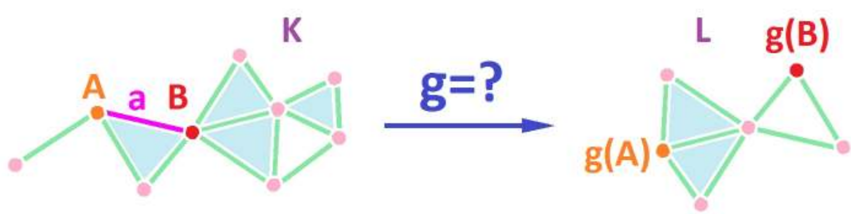
is fully determined by its values on the vertices:

$$g : K^{(0)} \to L^{(0)}.$$

What if we have nothing but the latter? Can we always extend g to the 1-simplices of K? If $AB \in K$, where A, B are vertices in K, we set:

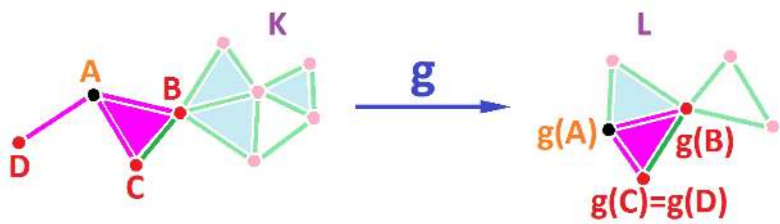
$$g(AB) := g(A)g(B).$$

We just need to make sure that there is an edge between $g(A)$ and $g(B)$:



We have to require that
- A and B are adjacent in K if and only if $g(A)$ and $g(B)$ are adjacent in L.

If we understand "adjacent" as "close", this condition mimics *continuity* as seen below.



For the higher dimensions, we set:

$$g(A_0 ... A_n) := g(A_0)...g(A_n),$$

with possible repetitions. The new map,

$$g : K \to L,$$

given by this formula is called the *simplicial extension* of g.

Proposition 4.46. Given a map $g : K^{(0)} \to L^{(0)}$ of vertices of two simplicial complexes. Then its simplicial extension is well-defined provided:

$$A_0 ... A_n \in K \Longrightarrow g(A_0)...g(A_n) \in L$$

(vertices may appear more than once).

A more subtle way to verify this condition follows from the result below.

Theorem 4.47 (Simplicial Extension Theorem). Suppose function

$$f : |K| \to |L|$$

maps vertices to vertices:

$$f(K^{(0)}) \subset L^{(0)}.$$

Then the simplicial extension g of f restricted to vertices, $f\big|_{K^{(0)}}$, is well-defined provided

$$f(N_A) \subset N_{f(A)},$$

for every vertex A in K.

Proof. The Star Lemma states that if we have a list, with possible repetitions, of vertices in a simplicial complex, they form a simplex of the complex if and only if the intersection of the stars of all these vertices is non-empty. We use this fact for both K and L below.

Let g be the restriction of f to the vertices. By the proposition, we only need to prove that

$$A_0...A_n \in K \Longrightarrow g(A_0)...g(A_n) \in L.$$

Now, by the Star Lemma, this is equivalent to:

$$\bigcap_{k=0}^{n} N_{A_k} \neq \emptyset \Longrightarrow \bigcap_{k=0}^{n} N_{f(A_k)} \neq \emptyset.$$

But, by the assumption, the image under f of the set on the left is contained in the set on the right, as seen below:

$$f\Big(\bigcap_{k=0}^{n} N_{A_k}\Big) \subset \bigcap_{k=0}^{n} f(N_{A_k}) \subset \bigcap_{k=0}^{n} N_{f(A_k)}.$$

■

Exercise 4.48. Prove the inclusion used above, i.e., for any family of sets $\{S_\alpha\}$ and any function f, we have:

$$f\Big(\bigcap_{\alpha} S_\alpha\Big) \subset \bigcap_{\alpha} f(S_\alpha).$$

Corollary 4.49. A function $g : K \to L$ is a simplicial map if and only if it satisfies the following conditions, for every vertex A in K,

- $g(A)$ is a vertex in L, and
- $g(\text{St}_A) \subset \text{St}_{g(A)}$.

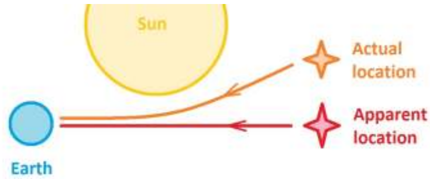
Exercise 4.50. Prove the corollary.

---

5 Manifolds

5.1 What is the topology of the physical Universe?

The space of locations of the Newtonian physics is the 3-dimensional Euclidean space, $\mathbf{R}^3$. We understand the topology of this space quite well. Most important is that it's acyclic. We know, however, from modern physics that the universe may be *curved.* For example, the observation that the light from a star passing the sun deviates from a straight line may be considered as evidence in support of this idea:

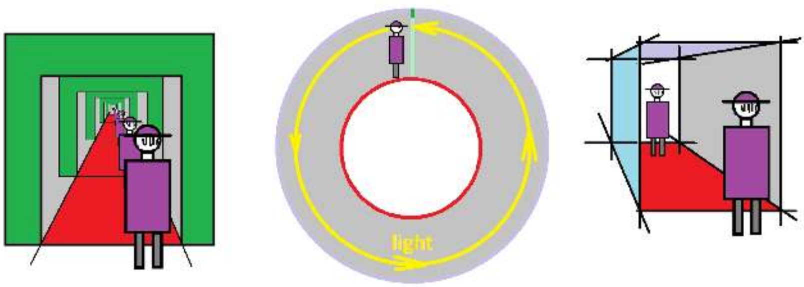


But if the universe curves, it might close on itself! Such a space then may have topological features and non-trivial homology groups. Our conclusion is that, locally, the universe looks like $\mathbf{R}^3$, but perhaps not globally.

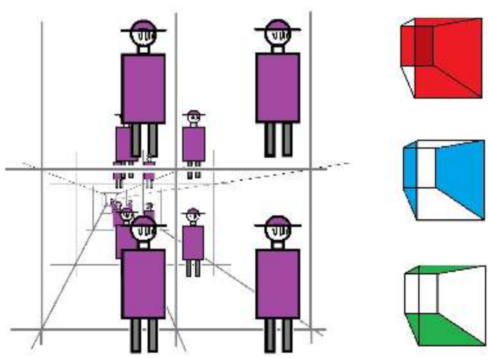
These "locally Euclidean" spaces are called *manifolds.*

It may be hard to understand or visualize such a space but, fortunately, we have already made a step in that direction.

Recall that identifying the front and back walls of a cubical room creates the enfilade effect as light circles this "universe" and comes back to us:



Now, there are still walls present and the light can't go in those directions. Therefore, this can't be a model of the universe. The idea is to take the same cube and glue *all* three pairs of the opposite faces, one to the other:



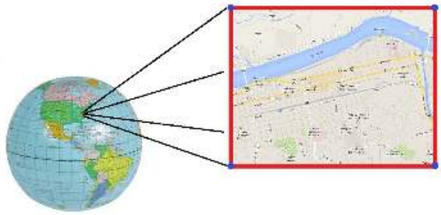
Exercise 5.1. What does the picture on the page of the chapter show?

This time, there are no walls and the light can freely propagate in all directions. The person can also walk in all directions and his experience will be completely Euclidean. Meanwhile, he sees enfilades of rooms – with himself in each of them – in all six directions.

Exercise 5.2. Compute the homology of the cube with opposite faces identified and compare it to that of the 3-sphere.

Note: Why don't we see these copies of ourselves, or the Earth, or the sun, everywhere we look? On a cosmic scale, it takes so long for the light to come back to us that what we see is a distant past. Even the sun was too young to be recognizable or maybe it simply didn't exist – there and then – yet.

Physics in dimension 2 is similar. Let's consider the surface of the Earth as a model. As we look around, this surface appears perfectly flat, just like the plane, $\mathbf{R}^2$:



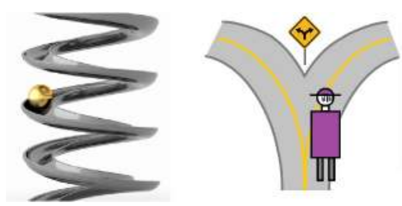
How do we even know it's not flat? Over the centuries, there have been a few experiments...

Experiment 1: Eratosthenes computed the difference between the lengths of shadows of the same object at two different locations at the same time of the day. The surface is proven to be curved!

Experiment 2: Magellan showed that one can come back without turning back. That can't be a plane!

Experiment 3: Gauss tried to check if the sum of the angles of a very large triangle (based on three mountain peaks) is equal to 180^o. Even though his result was inconclusive, the idea was sound. Indeed, this sum is larger than 180^o on a sphere! Just consider the triangle formed by two meridians and the equator:

The 1-dimensional physics is also conceivable – as a study of the motion of a car driven along a road or a ball rolling in a ball machine. The difference is that the motion of the ball, unlike that of a car on a road, is governed only by the laws of physics and, as a result, the ball can's make turns at will. This is the reason why, as a model of the universe, these roads and grooves can have no forks or intersections.



This means that they are always, everywhere, simple curves, i.e., curves that locally look like straight lines, $\mathbf{R}^1$. Once again, we arrive to the idea of a locally Euclidean space. The main examples are the line and the circle.

5.2 The locally Euclidean property

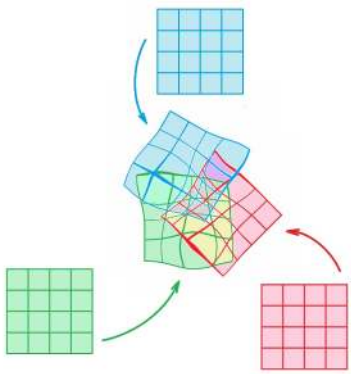
One can think of infinitely many maps that represent a piece of the surface of the Earth. That's why, for the precise definition of the concept we are after, we approach the issue from the opposite direction. We are given a (possibly finite) collection of maps and charts and we think of pieces of paper as if they are plastered all over the surface of the Earth. The result may look like a quilt:

So, the space is locally homeomorphic to the Euclidean space of dimension 2.

Definition 5.3. A topological spaces X is called *locally Euclidean of dimension* n if for every $x \in X$ there is an open set U such that $x \in U$ and there is a homeomorphism $h : \mathbf{R}^n \to U$. These homeomorphisms are called *charts* and any combination of charts that covers the whole X is called an *atlas*.

Each of these charts creates a local coordinate system for X. Such a system provides a locus for the n-dimensional linear algebra and *calculus*.

Of course, "homeomorphic to $\mathbf{R}^n$" can be replaced with "homeomorphic to an open n-ball", or "box", etc.:



We start with dimension 1, and here the charts are simply parametric curves:

$$r : (a, b) \to X.$$

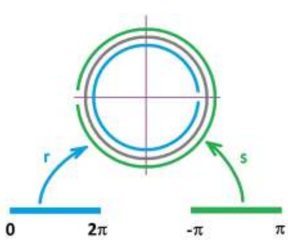
In order to ensure that these provide us with homeomorphisms, we need to require the curves to have no self-intersections; i.e., $r(t) \neq r(s)$ for all $a < s < t < b$.

Example 5.4 (circle). The circle is a very familiar curve and it is normally parametrized with a single function:

$$r(t) = (\cos t, \sin t), \ \ t \in [0, 2\pi].$$

However, $r(0) = r(2\pi)$, which means that this isn't a homeomorphism. We'll need at least two charts such as these:

$$\begin{array}{ll} r(t) = (\cos t, \sin t), & t \in (0, 2\pi), \\ s(t) = (\cos t, \sin t), & t \in (-\pi, \pi). \end{array}$$

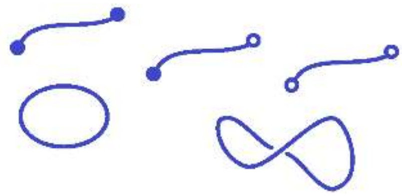


□

Exercise 5.5. Suggest other atlases for the circle.

Exercise 5.6. Prove that the figure eight on the plane is not locally Euclidean.

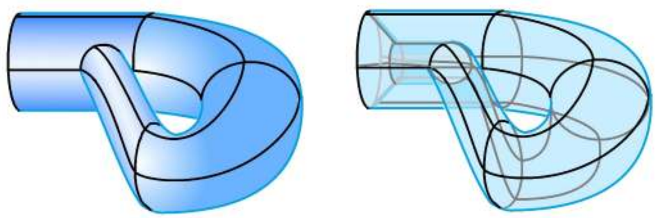
Example 5.7 (dimension 1). Consider these examples of dimension 1 locally Euclidean spaces. They may be closed, open, and half-open; finite and infinite; and the disjoint unions of these:



□

Example 5.8 (dimension 2). Some of the familiar spaces of dimension 2 can be parametrized by a single function:
- the sphere (parametrized by the spherical coordinates),
- the infinite cylinder (parametrized by the cylindrical coordinates), and
- the torus (suggest a parametrization).

As these functions aren't one-to-one, they can't be charts.



□

Exercise 5.9. Sketch an atlas for each of these spaces. Can you make it smaller?

Example 5.10 (dimension 0). There are 0-dimensional examples too:
- a point,
- n points,
- $\mathbf{N} \subset \mathbf{R}^1$,
- $\{\frac{1}{n} : n = 1, 2, ...\}$. □

Exercise 5.11. Prove that $\{0\} \cup \{\frac{1}{n} : n = 1, 2, ...\}$ is not locally Euclidean.

Exercise 5.12. Prove that in $\mathbf{R}^N$, the following spaces are locally Euclidean:
- open subsets;
- linear subspaces.

What are their dimensions?

5.3 Two locally Euclidean monstrosities

Can all locally Euclidean spaces serve as adequate models of the physical space? The answer is an emphatic *No*. Some of the locally well-behaved spaces exhibit strange global patterns.

For the first example, we examine how we handle sequences in $\mathbf{R}^n$.

The standard basis γ of $\mathbf{R}^n$ is the set of all open ε-balls. What if we are interested only in convergence of sequences to a *particular point* $a \in \mathbf{R}^n$? Then we would simply consider the standard "local" basis at a, which is the set of all open ε-balls around a. We have:

$$x_k \to a \text{ if and only if for any } \varepsilon > 0 \text{ there is an } N \text{ such that } x_k \in B(a, \varepsilon) \text{ for } k > N.$$

What's important, however, is that choosing only balls with radii equal to the reciprocals of integers gives us a *countable* local basis:

$$\gamma_a = \{B(a, \tfrac{1}{k}) : \ k = 1, 2, 3, ...\}.$$

Limiting our attention to just this collection still allows us to capture sequences convergent to a:

$$x_k \to a \text{ if and only if for any } m > 0 \text{ there is an } N \text{ such that } x_k \in B(a, \tfrac{1}{m}) \text{ for } k > N.$$

Exercise 5.13. Prove the above statement.

The reason why this still works is the following *local* refining property:

Definition 5.14. If for every open subset W of X with $a \in W$ there is a $U \in \gamma_a$ such that $U \subset W$, then such family is called a *local basis* of X at a.

What about general topological spaces? Recall that a basis of neighborhoods of a topological space X is any family γ of subsets of X that satisfies these two axioms:

- (B1) Covering: $\cup\gamma = X$;
- (B2) Refining: for any $U, V \in \gamma$ and any $x \in U \cap V$, there is a $W \in \gamma$ with $x \in W$ such that $W \subset U \cap V$.

A basis determines what sets are open in X. Next, for every $a \in X$, we can "localize" the basis γ by setting:

$$\gamma_a := \{U \in \gamma : a \in U\}.$$

Exercise 5.15. Prove that for any (global) basis γ, its localization γ_a is a local basis.

Definition 5.16. A spaces that has a countable local basis at every point is called *first-countable.*

Exercise 5.17. Prove that every metric space is first-countable.

It is obvious that every locally Euclidean space is also first-countable.

However, the Euclidean space has an even nicer feature: a countable "global" basis. All it takes is to combine the local countable bases to the ones centered at points with rational coordinates:

$$\gamma' := \bigcup \{\gamma_x : \ x = (x_1, ..., x_n) \in \mathbf{R}^n, x_i \in \mathbf{Q}\}.$$

Exercise 5.18. Prove that γ' is a basis.

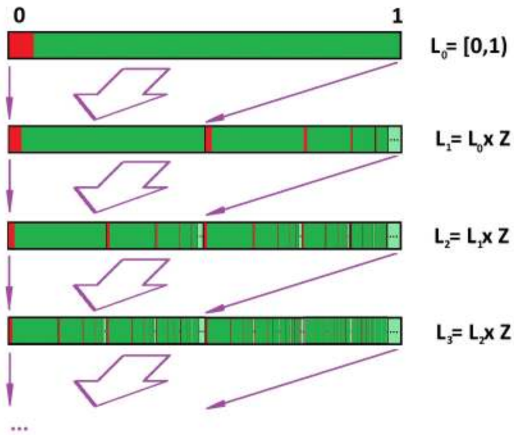
Definition 5.19. A space with a countable basis of neighborhoods is called *second-countable.*

Exercise 5.20. Prove that every second-countable space is first-countable. Give a simple example of a space that is first-countable but not second-countable.

Note: A second-countable space is well-behaved in the following sense. It contains a countable, *dense* subset; that is, there exists a sequence $\{x_n : \ n = 1, 2, ...\}$ of elements of the space such that every non-empty open subset of the space contains at least one element of the sequence. Such spaces are called *separable.*

Example 5.21 (long line). The *long line* is a topological space similar to the real line $\mathbf{R}$, but a lot "longer". While the former consists of a countable number of line segments $[0, 1)$ attached end-to-end, the latter is constructed from an uncountable number of such segments. This is how it is built in a step-by-step fashion:

• first we choose $L_0 := [0,1]$;
• then we create countably many copies of L_0 and attach to each other end-to-end; informally, $L_1 := L_0 \times \mathbf{Z}$;
• then we create countably many copies of L_1 and attach to each other end-to-end; $L_2 := L_1 \times \mathbf{Z}$;
• and so on:



The last step is:

$$L_\infty := \bigcup_n L_n \setminus \{0\}.$$

Any basis of the topology of this space must contain at least one open set from each $[0,1)$, and, therefore, it's uncountable. □

Exercise 5.22. Define this space via the order topology on $\mathbf{R}_+ \times [0,1)$ based on the lexicographical order on the set: $(a,b) \leq (a',b')$ if and only if $a < a'$ or $(a = a'$ and $b \leq b')$.

The key difference between the long line and the Euclidean space is given in the following exercise.

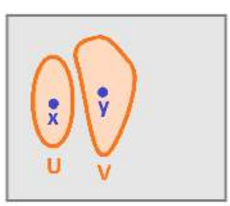
Exercise 5.23. Prove that (a) the long line isn't compact; yet, (b) every sequence has a convergent subsequence. (c) What about "uncountable" sequences?

Exercise 5.24. Prove that sequentially pasting together countably many copies of $[0,1)$ gives a space still homeomorphic to $[0,1)$.

The properties of the long line give an example of what we *don't* want our model of the physical space to be. This is why in our quest for an intrinsic, purely topological description of this space, we impose an extra condition: second-countability.

Next, any Euclidean space X satisfies the following condition (more in the next subsection).

Definition 5.25. A space is called *Hausdorff* if
• any two distinct points have neighborhoods that don't intersect.



In the Euclidean space, it is simple: if x, y are these points in X, we choose:

$$U := B(x,r),\ V := B(y,r),\ r := ||x-y||/2.$$

Example 5.26 (line with two origins). Consider the space called the *line with two origins.* It is the quotient of two copies of the real line:

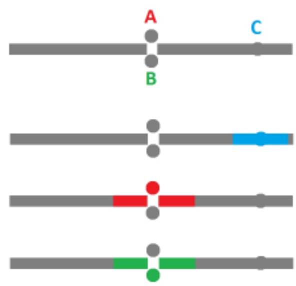
$$\mathbf{R} \times \{a\} \sqcup \mathbf{R} \times \{b\},$$

under the following equivalence relation:

$$(x,a) \sim (x,b) \text{ if } x \neq 0.$$

The points of this space are:
- $C = [(r,a)] = [(r,b)]$ for each real number $r \neq 0$,
- $A = [(0,a)]$, and
- $B = [(0,b)]$.

In this space, every point has an interval neighborhood:



But also any neighborhood of A intersects any neighborhood of B; that's not Hausdorff! □

This space is so unlike the Euclidean space (globally) that it can't be a model of the physical space and that's why we require an extra condition, Hausdorff.

Thus, we have two pathological examples of locally Euclidean spaces:
- the long line is Hausdorff but not second-countable, while
- the line with two origins is second-countable but not Hausdorff.

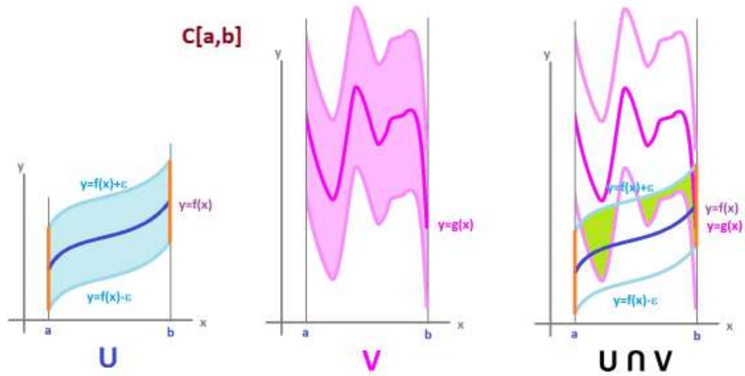
Exercise 5.27. Prove that
- the long line is Hausdorff, and
- the line with two origins is second-countable.

5.4 The separation axioms

The Hausdorff property is an example of a "separation axiom". We do actually *separate* the two points from each other by means of disjoint open sets:
- for any $x, y \in X, x \neq y$, there are open sets U, V such that $x \in U, y \in V$ and $U \cap V = \emptyset$.

Example 5.28. In addition to the Euclidean space, there are a few simple examples of this property. In the anti-discrete space, we always have $U = V = X$ for non-empty sets, so it's not Hausdorff. In the discrete space, one can always choose $U := \{x\},\ V := \{y\}$, so it's Hausdorff. The real line $\mathbf{R}$ equipped with the ray topology $\{(p, \infty) : p \in \mathbf{R}\}$ isn't Hausdorff because any two rays intersect. □

A less trivial example is that of the space of functions with sup-norm.



Exercise 5.29. Prove that $C[a, b]$ is Hausdorff. What about $C_p[a, b]$? Or $C(X)$ for some topological space X?

Theorem 5.30. If X is Hausdorff, then $\{x\}$ is closed for any $x \in X$.

Proof. The idea is to separate x from every other point in X. So, if $y \neq x$, then, by definition, there are open sets U, V such that $x \in U, y \in V_y$ and $U \cap V_y = \emptyset$. We don't care about U, but we do care that x does not belong to any V_y. Then we observe:

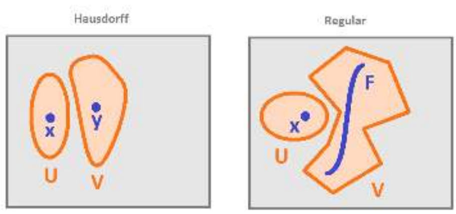
$$X \setminus \{x\} = \bigcup \{V_y : y \in X \setminus \{x\}\}.$$

This is an open set and, therefore, its complement, $\{x\}$, is closed. ■

Exercise 5.31. Prove that X is Hausdorff if and only if its diagonal $\Delta(X) := \{(x, x) : \ x \in X\}$ is closed in $X \times X$.

A higher degree of separation than Hausdorff is when a point and a *closed set* are separated by neighborhoods:

- for any closed set F and a point x that does not belong to X, there are open sets U, V such that $x \in U, F \subset V$ and $U \cap V = \emptyset$.



If, in addition, every singleton $\{x\}$ in X is closed, X is called *regular*.

Clearly, a discrete space is regular and an anti-discrete is not.

The following is obvious.

Theorem 5.32. Every regular space is Hausdorff.

Let's consider the possibility of the converse.

Suppose X is Hausdorff, F is a closed subset of X, $x \in X \setminus F$. The *idea* is to separate x from each y in F and then use these open sets to separate x from the whole F.

Without jumping to conclusions, let's write down carefully exactly what we have:

- for any $y \in F$, there are open sets U_y, V_y such that $x \in U_Y$, $y \in V_y$ and $U_y \cap V_y = \emptyset$.

How do we build, from these, the two open sets U, V that separate x and F?

The first idea is to take the intersections of these sets:

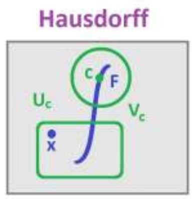
$$U := \bigcap_{y \in F} U_y, \ V := \bigcap_{y \in F} V_y.$$

Unfortunately, V won't have to include the whole F.

The second idea is to try the unions:

$$U := \bigcup_{y \in F} U_y, \ V := \bigcup_{y \in F} V_y.$$

Unfortunately, U might intersect F:



The third idea is to take the intersection for U and the union for V. This is something that might work.

Lemma 5.33. Given two collections of subsets indexed by the same set:

$$\alpha = \{U_y : y \in F\}, \ \beta = \{V_y : y \in F\},$$

suppose they are pairwise disjoint:

$$U_y \cap V_y = \emptyset, \ \forall y \in F.$$

Then the sets

$$U = \cap\alpha, \ V = \cup\beta,$$

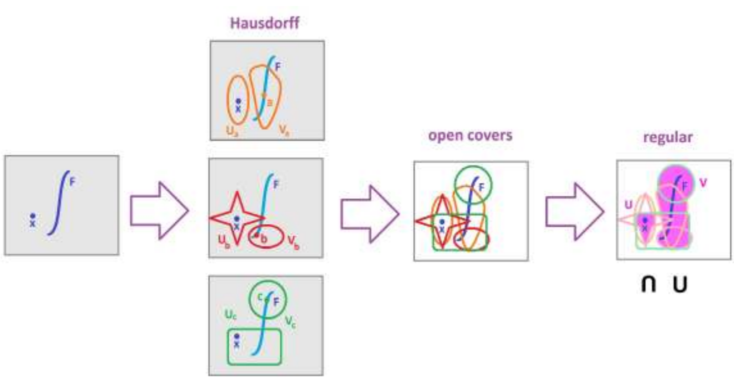
are disjoint too.

Proof. Suppose $z \in U \cap V$. Then

- $z \in U_y$ for all $y \in F$, and
- $z \in V_c$ for some $c \in F$.

Then, $z \in U_c \cap V_c$, a contradiction. ■

Here is an outline of the proof of the theorem:



We are in trouble when we realize that U doesn't have to be *open*!

Fortunately, an extra assumption of compactness allows us to salvage the proof.

Theorem 5.34. A compact space is Hausdorff is and only if it is regular.

Proof. [$\Longrightarrow$] Suppose F is a closed subset of X, $x \in X \setminus F$. Since X is Hausdorff, we have, for any $y \in F$, open sets U_y, V_y such that $x \in U_Y$, $y \in V_y$ and $U_y \cap V_y = \emptyset$. Since X is compact and F is closed, F is compact too. Hence the open cover

$$\gamma := \{V_y : y \in F\}$$

of F has a finite subcover, γ'. This means that we can write:

$$\gamma' = \{V_y : y \in F'\},$$

where F' is some finite subset of F. Now define

$$U := \bigcap\{U_y : y \in F'\},\ V := \bigcup\{U_y : y \in F'\}.$$

Both sets are open. And, by the lemma, they are disjoint. ■

This proof is a typical use of compactness...

We just used the following fact: a closed subset of a compact space is compact. A partial converse is contained in the following result.

Theorem 5.35. Suppose X is Hausdorff. Then a subset A of X is compact if and only if it is closed.

Exercise 5.36. Prove this theorem.

These kinds of extra conditions make it easier to prove topological equivalence by removing the need to verify the requirement that the inverse of the function is to be continuous.

Theorem 5.37. Suppose X is compact and Y is Hausdorff. If $f : X \to Y$ is continuous, one-to-one, and onto, then f is a homeomorphism.

Proof. Suppose A is a closed subset of X. Then A is compact. But the image of a compact space under a continuous function is a compact too. So, $f(A)$ is a compact subset of Y. Since Y is Hausdorff, this implies that $f(A)$ is closed in Y. Thus, we have proven that the image of any closed set under f is closed. Hence the preimage of any closed set under the inverse function f^{-1} of f is closed. Therefore, f^{-1} is continuous. ■

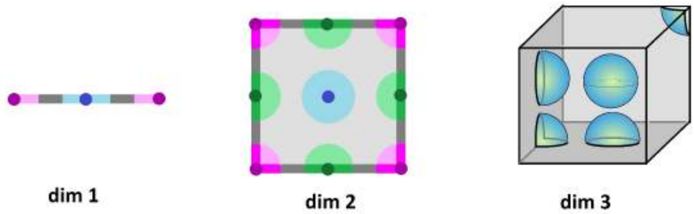
Even without being locally Euclidean, a space with such conditions can be seen as a metric space. We accept the following theorem without proof (see Munkres, *Topology. A First Course*, p. 216).

Theorem 5.38 (Urysohn's Metrization Theorem). Every second-countable, regular space is metrizable; i.e., there is a metric that generates its topology.

5.5 Manifolds and manifolds with boundary

Definition 5.39. An *n-dimensional manifold* is defined as a topological space that is
- locally Euclidean of dimension n,
- second-countable, and
- Hausdorff.

We've used *cells* as building blocks a lot now, are they manifolds?



No, some of the points – the ones on the boundary of these cells – aren't homeomorphic to Euclidean spaces.

The 1-cell is the segment $[0,1]$. Here, some of the neighborhoods are open intervals, $V \approx \mathbf{R}$, and others are half-open intervals, $U \approx \mathbf{R}_+$ (the half-line).

The same happens in all dimensions. For an n-cell σ, the neighborhoods of the boundary points $\partial\sigma$ are homeomorphic to the half-ball, or to $\mathbf{R}^n_+ = \{(x_1,...,x_n):\ x_1 \geq 0\}$, the closed half-space.

Proposition 5.40. Since an n-cell is homeomorphic to the closed n-ball $\mathbf{B}^n$, its boundary is homeomorphic to the $(n-1)$-sphere:

$$\partial \mathbf{B}^n \approx \mathbf{S}^{n-1}.$$

Exercise 5.41. Provide the details to prove the proposition.

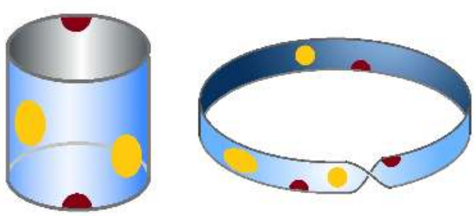
In dimension 2, manifolds are called *surfaces*. There are many surfaces around us including the actual manifolds:

But unlike the surface of the Earth or a balloon, these have edges called *boundaries*.

Definition 5.42. A second-countable, Hausdorff space M is called an *n-manifold with boundary* (which can be empty) if for every point $a \in M$, there is an open set (a chart) U such that $a \in U$ and:
- $p : \mathbf{R}^n \to U$ is a homeomorphism, or
- $p : \mathbf{R}^n_+ \to U$ is a homeomorphism.

The former points compose the *interior* $\dot{M}$ and the latter compose the *boundary* ∂M of M. We will sometimes use the **notation** that puts the dimension of the manifold as a superscript:

$$M = M^n.$$

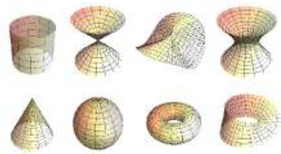


Note that, if we start with a manifold with non-empty boundary, such as a closed disk, we can remove points from its boundary one by one and acquire a new manifold at every step. In the end of this process, we have a manifold with empty boundary, such as an open disk. All of these manifolds are endowed with the same topological features and, therefore, have no independent interest. This is the reason why it is common to concentrate on *compact manifolds* exclusively.

Example 5.43. In addition to cells, these are compact manifolds with non-empty boundaries that we have seen:

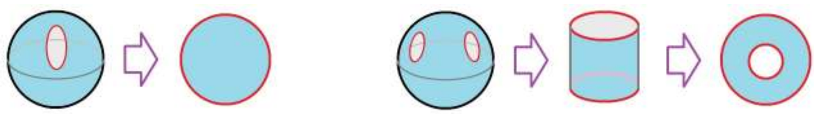
- the cylinder,
- the Möbius band,
- the cube,
- the cube with one pair of opposite walls identified, etc.

Also, we can produce numerous new examples of manifolds by taking any of these or any of the surfaces below and puncture holes:

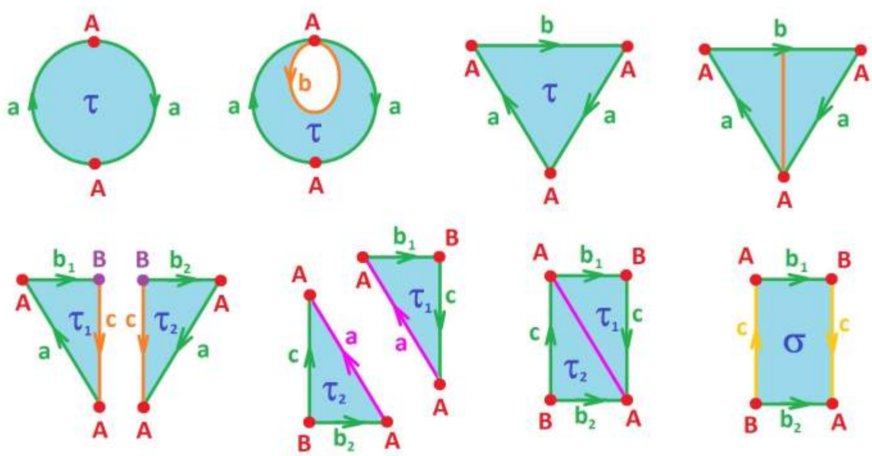


Exercise 5.44. One of the surfaces depicted above isn't a manifold. Which one? What about the book with n pages?

We can deform the sphere with a hole or two (but not more) into something familiar:



Example 5.45. We cut a hole in the middle of the projective plane and represent the result as a cell complex with a single 2-cell. What is it?



□

Exercise 5.46. Describe the steps in the above construction.

The n-dimensional analog of puncturing a hole in a surface is removing a neighborhood homeomorphic to $\mathbf{R}^n$ (a void) from an n-manifold.

The following is an important result.

Theorem 5.47. The boundary of a compact n-manifold with boundary M is an compact $(n-1)$-manifold with empty boundary.

Proof. Each point x in the boundary ∂M of the n-manifold M has a neighborhood homeomorphic to the half-space $\mathbf{R}^n_+ := \{(x_1,...,x_n) : \ x_n \geq 0\}$. In fact, x belongs to the image N_x of the boundary of this set, which is the hyperplane,

$$N_x \approx \mathbf{R}^n_0 := \{(x_1,...,x_n) : \ x_1 = 0\} \approx \mathbf{R}^{n-1}.$$

Therefore, $\{N_x : \ x \in \partial M\}$ is an atlas of ∂M. ■

Exercise 5.48. (a) Provide details for the above proof. Hint: prove that the boundary ∂M of a compact manifold is a closed subset. (b) Prove that the interior $\dot{M}$ is a manifold (without boundary).

Theorem 5.49. The disjoint union of two n-manifolds is an n manifold.

Because of this theorem, we concentrate on *path-connected manifolds.*

Let's apply these results to lower-dimensional manifolds.

Theorem 5.50. The only path-connected, compact 1-manifold is the circle, $\mathbf{S}^1$.

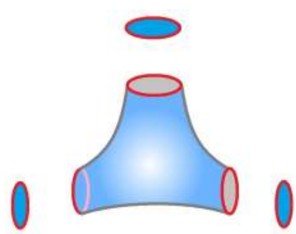
Also, as ∂M^1 is just a collection of points, we have the following result.

Theorem 5.51. The only path-connected, compact 1-manifold with non-empty boundary is the closed interval, $\mathbf{I}$.

While there is only one such 1-manifold, there is a variety of surfaces. The boundaries of the holes are circles, hence, so are the boundaries of these surfaces. Are all surfaces with boundary like that? We know this is the case for the cylinder and even the Möbius band.

Theorem 5.52. The boundary of a compact surface M^2 is:
- a closed subset of $M^2 \Longrightarrow$
- a compact 1-manifold N^1 without boundary $\Longrightarrow$
- the disjoint union of a finite number of circles $\mathbf{S}^1$.

It follows that a surface with boundary can be easily turned into a surface without boundary by attaching a few disks:



Exercise 5.53. What happens if you attach the disk to the boundary of the Möbius band? Prove via a diagram.

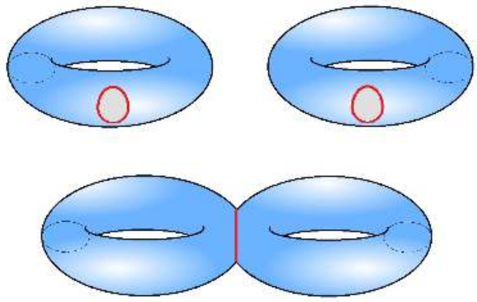
Exercise 5.54. A spider crawls on a surface and leaves a string of its web behind. How can it find out the topology of the surface?

Exercise 5.55. One can build 3-manifolds by "thickening" graphs. Explain exactly how it is done.

5.6 The connected sum of surfaces

Any two surfaces can be attached to each other by puncturing holes in them and then gluing them together along the new, circular edges. The result is called the *connected sum* $S_1 \# S_2$ of surfaces S_1, S_2.

Example 5.56. This is how you construct the *double torus* by attaching two tori to each other in this fashion:



□

More precisely, suppose N_i is a subset of S_i for $i = 1, 2$ homeomorphic to a disk, then its boundary is homeomorphic to a circle:

$$N_i \approx \mathbf{B}^2, \ \partial N_i \approx \mathbf{S}^1.$$

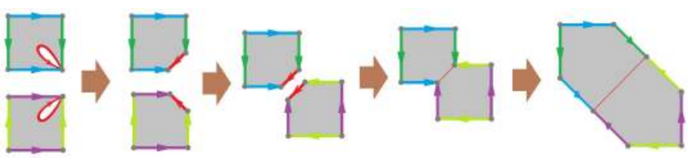
Then we define

$$S_1 \# S_2 := \Big[(S_1 \setminus N_1) \sqcup (S_2 \setminus N_2) \Big] /_{\sim},$$

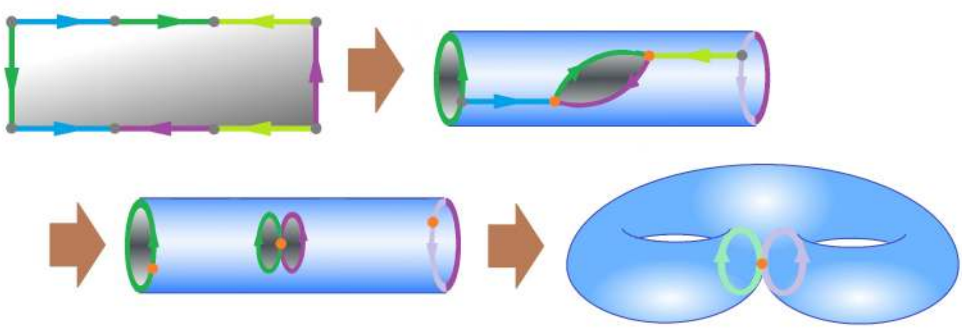
with the equivalence relation generated by a homeomorphism

$$h : \partial N_1 \to \partial N_2.$$

We can illustrate the process via these gluing diagrams:



One can also interpret the diagram by gluing along the edges:



Exercise 5.57. Prove $S \# \mathbf{S}^2 = S$.

Exercise 5.58. What is $\mathbf{P}^2 \# \mathbf{P}^2$?

Exercise 5.59. Prove

$$H_1(\mathbf{T}^2 \# \mathbf{T}^2) = \mathbf{R}^3.$$

Exercise 5.60. Suppose K_1, K_2 are simplicial complexes the realizations of which are surfaces:

$$|K_i| = S_i, \ i = 1, 2.$$

• (a) Define a complex K such that $|K| = S_1 \# S_2$ (in other words, find a triangulation of the connected sum in terms of those for the two surfaces). Hint: don't cut cells.
• (b) Express the homology of K in terms of the homology of K_1, K_2:

$$H_1(K) = H_1(K_1) \ ? \ H_1(K_2).$$

The construction allows one to classify all surfaces as in the theorem below that we will accept without proof (see Kinsey, *Topology of Surfaces*, p. 81).

Theorem 5.61 (Classification of surfaces). (1) A compact path-connected surface is homeomorphic to
• the sphere $\mathbf{S}^2$, or
• the connected sum of n tori $\mathbf{T}^2$, or
• the connected sum of n projective planes $\mathbf{P}^2$.
(2) These options are not homeomorphic.

Therefore, the options are only:

$$\mathbf{S}^2, \ n\mathbf{T}^2, \ n\mathbf{P}^2.$$

Exercise 5.62. Classify:
• $\mathbf{S}^2 \# \mathbf{S}^2$,
• $\mathbf{P}^2 \# \mathbf{K}^2$,
• $\mathbf{T}^2 \# \mathbf{K}^2$,
• $\mathbf{K}^2 \# \mathbf{K}^2$.

Exercise 5.63. (a) Define the 1-dimensional analog of the connected sum. (b) Classify all 1-manifolds with an analog of the above theorem.

Let's consider what happens to the Euler characteristic when we form the connected sum of two tori. Based on the above diagram, cutting out the disks has the following consequences:
• no new faces, net effect 0;
• one new, shared edge, net effect -1;
• one existing vertex becomes shared, net effect -1.
Accordingly, the net effect is -2.

This argument applies to all surfaces:

Theorem 5.64. For two surfaces S and T,

$$\chi(S \# T) = \chi(S) + \chi(T) - 2.$$

Exercise 5.65. Use the formula to compute the Euler characteristic of the n-torus.

5.7 Triangulations of manifolds

Next, we study manifolds represented by data; given a simplicial complex K, what are the necessary and sufficient conditions that its realization $|K|$ is an n-manifold M^n (without boundary)? When this is the case, we will call K a *combinatorial n-manifold.*

First, let's take care of the point-set topological issues. Since the realization is the union of finitely many n-balls and, therefore, second-countable, then so is $|K|$. For Hausdorff, two points are either in the same cell, which is itself Hausdorff, or in two disjoint stars, which serve as the required disjoint neighborhoods.

Exercise 5.66. Provide the details of this proof and state the theorem.

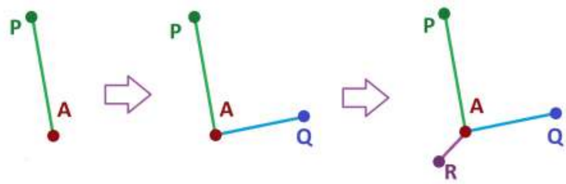
So, we only need to deal with the question whether $|K|$ is locally Euclidean: does every point in M^n have a neighborhood homeomorphic to $\mathbf{R}^n$?

The first observation is simple.

Theorem 5.67. If $|K| \approx M^n$ then $\dim K = n$.

Now, suppose we have an n-dimensional simplicial complex K. How do we know that K is a combinatorial n-manifold? Can we answer this question just by examining the list of simplices of K? The answer is Yes.

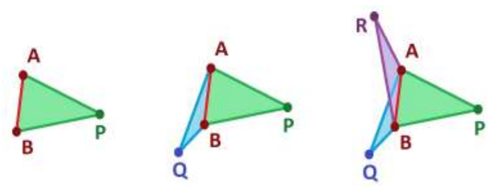
We start with dimension $n = 1$. We could use the two "global" results given previously, but instead we start from the definition. First, by the above theorem, K has to be a graph. Furthermore, the star of a vertex with more than one adjacent edge isn't homeomorphic to the open interval; that's why we require:
- each vertex is shared by exactly two edges.



The condition can be restated in terms of vertices only.

Theorem 5.68. A graph K is a combinatorial 1-manifold if and only if the following condition is satisfied:
- (A1) for any vertex $A \in K$, there are exactly two vertices $P, Q \in K$ with $AP, AQ \in K$.

Next, dimension $n = 2$. The first requirement a complex has to meet to be a surface is similar to the one above:
- each edge is shared by exactly two faces.



Once again, this condition can fail in three ways:
- an edge isn't contained in any face,
- an edge is contained in exactly one face,
- an edge is shared by more than two faces.

In the first case, we might have a curve. In the second, possibly, a surface with boundary.

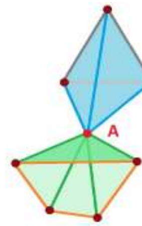
Exercise 5.69. To justify the above statement, prove that three copies of $\mathbf{R}^n_+$ glued together by the hyperplane $\{x_1 = 0\}$ is not homeomorphic to $\mathbf{R}^n$. Hint: same as idea as for "T" $\not\approx$ "I".

Now, we restate this condition *combinatorially*, i.e., for a 2-dimensional simplicial complex K given by a list vertices and simplices as sets of vertices:

$$K = \{A, B, C, ..., AB, BC, ..., ABC, ...\}.$$

In terms of these sets, every face is a triple, such as ABC, and every edge is a double, such as AB. The above condition can then be rephrased in terms K:
- (A2) for any edge $AB \in K$, there are exactly two vertices $P, Q \in K$ with $ABP, ABQ \in K$.

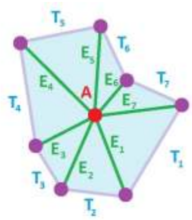
This condition does not guarantee that the realization of K is a surface, as the image below shows.



As we can see, condition (A2) is satisfied here for all *edges* but the *vertex* in the middle does not have a neighborhood homeomorphic to the disk.

Exercise 5.70. To justify the above statement, prove that two copies of $\mathbf{R}^n$ glued together by the origin 0 is not homeomorphic to $\mathbf{R}^n$.

Then we have to require that the faces around each vertex are arranged in, and can be listed as, a circular fashion to form a disk-like subcomplex.



Given a vertex A the faces (triangles) that contain A must form a "cycle":

$$T_1 \text{ glued to } T_2 \text{ glued to } T_2 \ ... \ T_m \text{ glued to } T_1.$$

We can state this as a condition on faces:
- For each vertex A, suppose
  - ◇ all edges that contain A are $E_1, E_2, ..., E_k$, and
  - ◇ all triangles that contain A are $T_1, T_2, ..., T_m$.
- Then $k = m$ and the two sets can be reindexed in such a way that:
  - ◇ E_1 is shared by T_1, T_2;
  - ◇ E_2 is shared by T_2, T_3;
  - ◇ ...
  - ◇ E_m is shared by T_m, T_1.

Of course, we recognize this collection of simplices as the star of vertex A.

Restated in terms of subsets of $K = \{A, B, C, ...\}$, the above condition will take the following form:

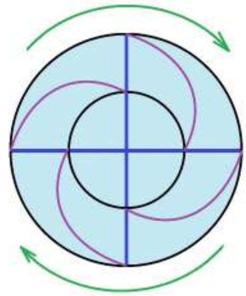
- (B2) for each vertex $A \in K$, the set of adjacent vertices $\{B \in K : AB \in K\}$ can be represented as $\{B_1, B_1, ..., B_m\}$ with

$$AB_1B_2, AB_2B_3, ..., AB_mB_1 \in K.$$

We have proved the following.

Theorem 5.71. A 2-dimensional simplicial complex K is a combinatorial 2-manifold if and only if it satisfies conditions (A2) and (B2).

Exercise 5.72. Is this a triangulation of the projective plane?



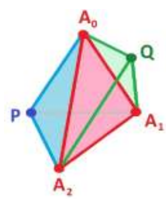
Exercise 5.73. Prove an analog of the above theorem for manifolds with boundary.

Exercise 5.74. When is a cubical complex in $\mathbb{R}^3$ a surface?

The reason why we impose these conditions on a complex is clear now: we have to check that the star of every simplex of every dimension forms an n-ball. The difficulty is that there will be n such conditions and we will have to set them up one at a time for each dimension from 0 to n. Let's consider just one of them.

In light of our discussion of conditions (A1) and (A2), the easiest way things can go wrong is when three pieces that are already homeomorphic to the n-ball are glued to each other. In other words, if an $(n-1)$-simplex is shared by three distinct n-simplices, this is not an n-manifold.

Condition (A3) is illustrated below:



Unlike the theorem above for $n = 2$, the next one gives us only a necessary condition.

Theorem 5.75. If K is a combinatorial n-manifold then the following condition holds:

- (An) if $A_0A_1...A_{n-1} \in K$ then there are exactly two vertices $P, Q \in K$ with

$$A_0A_1...A_{n-1}P \in K \quad \text{and} \quad A_0A_1...A_{n-1}Q \in K.$$

What conditions (A1),(A2), ...,(An) have in common is not the dimensions of these cells but rather their *codimensions*:

$$\operatorname{codim} \sigma := n - \dim \sigma.$$

These conditions are about cells of codimension 1.

Exercise 5.76. Suggest condition (B3) and illustrate it with a sketch. Hint: what's the codimension?

Exercise 5.77. Give an example of a simplicial complex that satisfies conditions (A3) and (B3) but is not a combinatorial 3-manifold.

5.8 Homology of curves and surfaces

From the discussion of the local issues in the last subsection, we move on to the global issues.

Suppose K is a combinatorial n-manifold which is path-connected. What do we know about the homology of K?

Let's list the results that we have established so far. Since there are no cells of dimension above n in the manifold, we conclude that

$$H_k(K) = 0, \ k > n.$$

Then we only need to consider $H_k(K)$ for $k = 0, 1, ..., n$. We group the results as follows.

Group 1, the 0 dimension:
- $H_0(K) = \mathbf{Z}$ for any K.

Group 2, the intermediate dimensions:
- $H_1(\mathbf{S}^2) = 0$;
- $H_1(\mathbf{T}^2) = \mathbf{Z} \times \mathbf{Z}$;
- $H_1(\mathbf{K}^2) = \mathbf{Z} \times \mathbf{Z}_2$;
- $H_k(\mathbf{S}^n) = 0$ for $0 < k < n$.

Group 3, the nth dimension:
- $H_1(\mathbf{S}^1) = \mathbf{Z}$;
- $H_2(\mathbf{S}^2) = H_2(\mathbf{T}^2) = \mathbf{Z}$;
- $H_2(\mathbf{K}^2) = H_2(\mathbf{P}^2) = 0$;
- $H_n(\mathbf{S}^n) = \mathbf{Z}$.

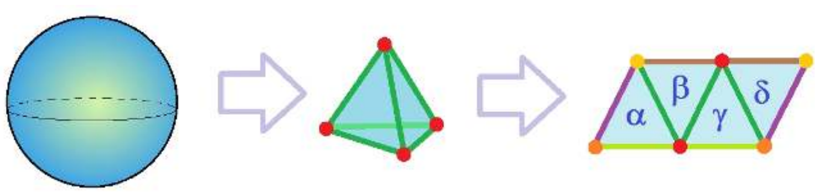
Note: Let's be clear here and make an important point: the formulas in Groups 2 and 3 are the result of computations of homology for several specific representations of these topological spaces as simplicial and cell complexes. Meanwhile, the formula in Group 1 was proved as a theorem independent of representations. Our goal in this subsection is a similar theorem for Group 3.

We see only two outcomes in Group 3: 0 and $\mathbf{Z}$. Let's try to explain why this is to be expected.

As we know, $H_2(K)$ is determined by whether $|K|$ has a void in it. In other words, whether this surface *holds air* inside. If it does, there is one side of the surface that touches the air while the other doesn't. Such surfaces are called *orientable*; for example, the sphere and the torus. Otherwise, a surface is called *non-orientable*. In that case the surface doesn't separate inside from the outside; for example, the Klein bottle and the projective plane. The reason is that, as we have seen, these surfaces contain the (one-sided) Möbius band.

Below we will concentrate on the orientable case.

Let's consider the sphere. In the case of its simplest *cell* complex, the void is captured by the single 2-cell, as a cycle. In the case of its simplest *simplicial* complex, it is the sum of all of its four 2-cells:



Hence the idea: the 2nd homology is generated by the sum of all 2-cells of the sphere:

$$H_2 = < \alpha + \beta + \gamma + \delta > .$$

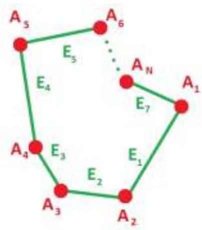
For the general case, we conjecture that the nth homology group of an n-manifold may be generated by the sum of all of its n-cells. The question is though, what should be signs (and the orientations) of these cells in the sum?

To see how we can handle this issue, we start with dimension $n = 1$. We will work combinatorially, with the simplicial complex K the realization of which is the 1-manifold M, which is supposed to be the circle.

Suppose we have an edge $A_1A_2 \in K$. Then, by condition (A1) from the last subsection applied to vertex A_2, there is exactly one vertex $A_3 \in K$ such that $A_2A_3 \in K$. We do the same for vertex A_3 next and then continue by induction. Since K is finite, this process will have to start to repeat itself. Also, since K is path-connected, the list of its vertices will then be exhausted. Therefore, there is an integer N such that

- $A_1, ..., A_N$ is the list of all vertices of K, and
- $A_1A_2, ..., A_{N-1}A_N, A_NA_1 \in K$.

Then we have a circular sequence of edges and vertices:



Let's consider the all-including 1-chain:

$$\tau := A_1A_2 + ... + A_{N-1}A_N + A_NA_1.$$

Is this what we want? Compute:

$$\begin{array}{ll}\partial\tau & = \partial(A_1A_2) + ... + \partial(A_{N-1}A_N) + \partial(A_NA_1)\\ & = (A_2 - A_1) + (A_3 - A_2) + ... + (A_N - A_{N-1}) + (A_1 - A_N)\\ & = 0.\end{array}$$

It's a cycle! What happened here? All vertices in the boundary of τ cancelled because each appeared twice, with $+$ and with $-$.

For convenience, let's add $A_{N+1} = A_1$ to the list; then we have

$$\tau = \sum_{i=1}^{N} s_iA_iA_{i+1}.$$

Now, something we could have overlooked, if we hadn't been expecting to see it, is that choosing $A_1A_2, ..., A_{N-1}A_N,\ A_NA_1 \in K$ amounts to choosing an *orientation* of all 1-cells in K, which becomes then an *oriented simplicial complex*. This isn't, however, just any set of orientations of simplices; they are chosen in such a way that they follow this "cancellation rule":

- two edges that share a vertex induce *opposite* orientations on that vertex.

Theorem 5.78. If K is a combinatorial path-connected 1-manifold, then

$$H_1(K) =< \tau >= \mathbf{Z}.$$

Proof. What's left to be proven? We need to find the rest of the 1-cycles. Suppose σ is one. Then,

$$\sigma = \sum_{i=1}^{N} s_iA_iA_{i+1},$$

for some $s_i \in \mathbf{Z}$. Then,

$$\begin{array}{ll}\partial\sigma & = \sum_{i=1}^{N} s_i\partial(A_iA_{i+1}),\\ & = \sum_{i=1}^{N} s_i(A_{i+1} - A_i)\\ & = \sum_{i=1}^{N} (s_{i+1} - s_i)A_i\\ & = 0?\end{array}$$

Due to $A_i \neq 0$, in order to produce zero, the coefficients have to vanish: $s_i = s_{i+1}$. Therefore, $s_i = s$ for all i. Hence

$$\begin{aligned}\sigma &= \sum_{i=1}^{N} sA_iA_{i+1}, \\ &= s\sum_{i=1}^{N+1} A_iA_{i+1} \\ &= s\tau \in < \tau > .\end{aligned}$$

■

The result is certainly familiar, but this is the first time it is proven to be independent of the realization.

Now, we are after a generator, **denoted** by O_K, of the nth homology group of a combinatorial n-manifold K:

$$H_n(K) =< O_K > .$$

When this group is non-zero, O_K is called the *fundamental class* of K.

What possible lessons have we learned from the case of dimension 1?

• (1) The candidate for the fundamental class is the sum of all n-cells in K, provided they are oriented according to the "cancellation rule":

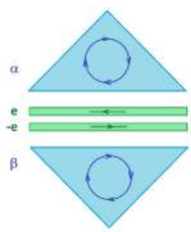
• (2) Two n-cells that share an $(n-1)$-cell induce opposite orientations on that cell. Such an orientation is called *compatible* and the procedure for finding one is based on the following:

• (3) Selecting an orientation for a single n-cell dictates the orientations of all the rest. If a compatible orientations exists for all cells, the manifold is called *orientable*.

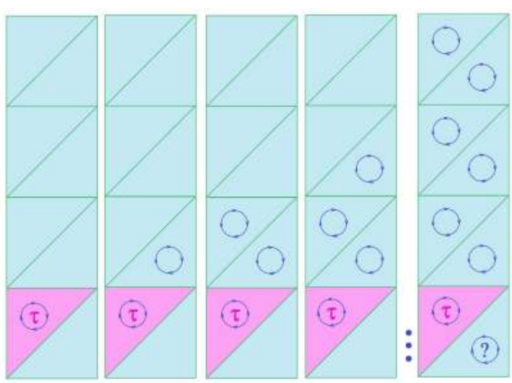
Exercise 5.79. Use the above approach to prove $H_n(\mathbf{S}^n;\ \mathbf{Z}_2) = \mathbf{Z}_2$.

We next apply this approach to the case $n = 2$.

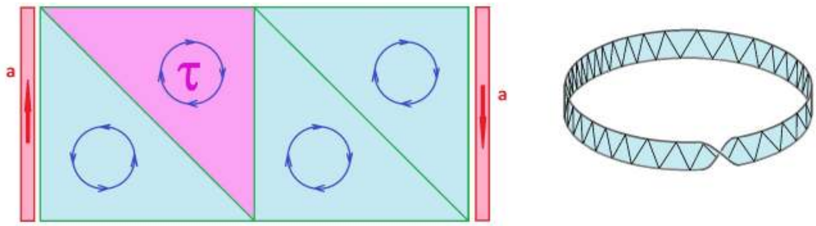
We can always choose compatible orientations, as long as we deal with one edge at a time:



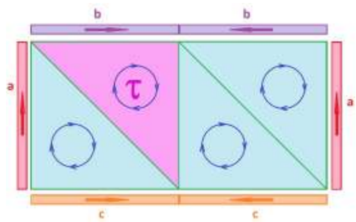
We proceed as follows. Start with an oriented face τ. Next, τ has an edge d. This edge must be shared with a face λ. Then λ is given the orientation compatible with that of τ. Next, λ has an edge f shared with another face which has to have the compatible orientation with λ, etc. At every step, we move from a face to an adjacent face and every time the orientation of the next face is "forced". This may continue for a while:



But what happens if we make a full circle and come back to τ? It's possible that the orientation to be imposed on τ from the last face will be opposite of what we have already. The last line above suggests how this can happen: the strip of faces we follow may have finished a full flip when we come back. In other words, we have the Möbius band inside:



Meanwhile, this four-cell triangulation of the sphere is acceptable:



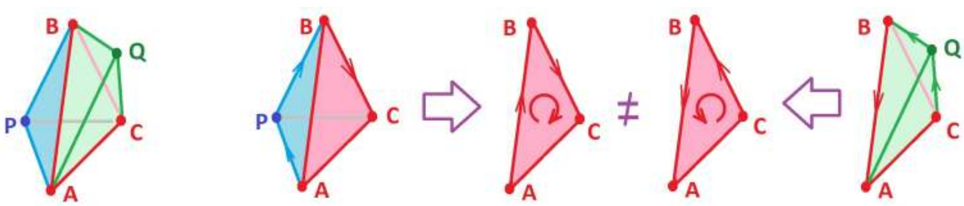
Exercise 5.80. Show that there is no compatible orientation of the Möbius band $\mathbf{M}^2$, regardless of the triangulation. Prove that a combinatorial surface is orientable if and only if it doesn't contain $\mathbf{M}^2$ as a subcomplex.

Exercise 5.81. Prove that for a non-orientable path-connected compact combinatorial surface K, the second homology group is trivial, $H_2(K) = 0$. Hint: consider again the propagation of coefficients of faces in the decomposition of a 2-cycle.

5.9 The nth homology of n-manifolds

The two low dimensions we have considered so far are special:
- dimension 1: all manifolds are orientable;
- dimension 2: non-orientability is detected by the presence of the Möbius band.

Let's recall that, in general, an orientation of a simplex is a particular way of listing its vertices. The given orientation of a simplex is passed along to its faces by dropping one of the vertices from the ordered list:



Once orientability is understood, the proof is very similar to the case of dimension 1. Let's just follow the algebra.

As a combinatorial n-manifold K has no $(n+1)$-cells, its nth boundary group is trivial:

$$B_n(K) = \partial(C_{n+1}(K)) = \partial(0) = 0.$$

Therefore, we only need to compute the n-cycle group:

$$H_n(K) = Z_n(K)/B_n(K) = Z_n(K)/0 = Z_n(K).$$

Suppose C is a non-zero n-cycle and

$$C := \sum_s d_s s, \ d_s \in \mathbf{Z},$$

with summation over all n-cells s in K. What do we know about these numbers?

The boundary ∂C of C is zero:

$$\partial C = 0,$$

or

$$\partial C = \sum_s d_s \partial s = 0.$$

Since this is the sum of all $(n-1)$-cells in an n-manifold, we can use condition (An): every $(n-1)$-cell is shared by exactly two n-cells. Therefore, every $(n-1)$-cell e appears exactly twice in the above sum. Because $\partial C = 0$, every $(n-1)$-cell e has to cancel. We have the following:

Lemma 5.82. If $(n-1)$-cell e is shared by n-cells σ and τ then

$$d_\sigma = \pm d_\tau.$$

So far, this number could be zero. Now, since K is path-connected, one can get from s to any other face by following edge-face-edge-face... sequence. By applying the lemma inductively, we obtain the following result.

Lemma 5.83. For any two n-cells σ and τ in K, we have

$$d_\sigma = \pm d_\tau.$$

Which sign? We need to look at the orientations of these cells. Suppose n-cells σ, τ share $(n-1)$-cell e. Then they have orientations that generate the *opposite* orientations on e. Otherwise,

$$\begin{array}{ll} C & = \sigma + \tau + \text{ terms w/o } \sigma, \tau \Longrightarrow \\ \partial C & = e + e + \text{ terms w/o } e \\ & \neq 0, \end{array}$$

because $e \neq 0$. These n-cell are then *compatibly oriented.*

Lemma 5.84. Suppose C is a non-zero n-cycle in an orientable path-connected combinatorial n-manifold K and

$$C = \sum_s d_s s, \ d_s \in \mathbf{Z},$$

with the summation over all n-cells s in K. Then, for any two faces σ and τ in K, we have

$$d_\sigma = d_\tau.$$

It follows that

$$C = d \sum_s s,$$

for some real d. We have proven the following important result.

Theorem 5.85. Suppose K is an orientable path-connected combinatorial n-manifold. Then the nth homology group of K is cyclic:

$$H_n(K) =< O_K >\cong \mathbf{Z},$$

generated by the *fundamental class*, which is the sum of all n-cells in K compatibly oriented:

$$O_K := \sum_s s.$$

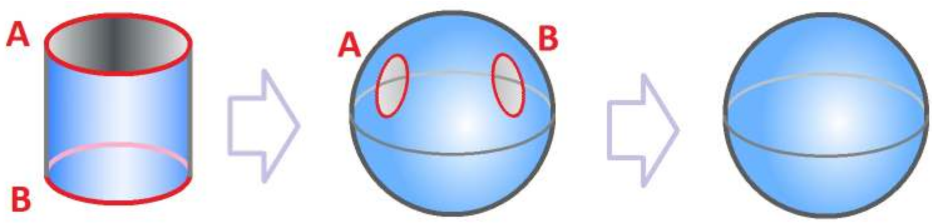
Theorem 5.86. Suppose K is a non-orientable path-connected combinatorial n-manifold. Then the nth homology group of K is trivial:

$$H_n(K) = 0.$$

Exercise 5.87. Prove the theorem. Hint: use the lemmas above.

5.10 Homology relative to the boundary

But what if K has *boundary*? We know that the formula doesn't hold anymore; after all, the cylinder can't hold air... We can make it hold air if we patch its holes! Patching the two holes turns the cylinder into a surface without boundary. This surface, the sphere, *can* hold air:



To appreciate this observation, compare it to patching the hole of the Möbius band. The result is the projective plane that still *can't* hold air. These two examples show that the homology of the "patched" version of the manifold is just as interesting as that of the original! The nth homology of the former is, in fact, *more* interesting because it tells us when the manifold is orientable.

The patching of a surface can be accomplished, topologically, by adding disks to cover the holes or by collapsing each hole to a point, one at a time.

Exercise 5.88. Suggest a procedure for patching holes in a combinatorial surface.

Patching isn't as simple for a general combinatorial n-manifold K. Fortunately, we know of an alternative approach: collapsing. We need to collapse to a point each path-component of the boundary ∂K. This collapse yields an equivalence relation on K. But, is this still a simplicial complex?

Exercise 5.89. Give an example when it's not.

Instead of dealing with this issue, we follow a purely algebraic path. We saw previously how to use the quotients of the chain groups:

$$C_k(K, \partial K) := C_k(K)/_\sim.$$

The homology is then computed via the quotients of the boundary operators:

$$[\partial_k] : C_k(K, \partial K) \to C_{k-1}(K, \partial K).$$

The result is called the *relative homology with respect to the boundary* and is denoted by $H(K,\partial K)$.

Example 5.90. Let's compute $H(\mathbf{B}^2,\mathbf{S}^1)$, the homology of the disk relative to its boundary.

The complex K of the disk is:
- 0-cell: A;
- 1-cell: a;
- 2-cell: τ;
- boundary operator: $\partial(\tau)=a;\partial(a)=0$.

We create the *relative chain complex* by identifying the 1-cell with 0. Then the quotients of the chain groups of K are:

$$\begin{array}{lll} C_2(K,\partial K) & = C_2(K)/_\sim & =<\tau>; \\ C_1(K,\partial K) & = C_1(K)/_\sim & =<a>/<a>=0; \\ C_0(K,\partial K) & = C_0(K)/_\sim & =<A>. \end{array}$$

Here is what relative 1-cycles look like:

The quotients of the boundary operators and groups form a chain complex and we compute its homology as follows:

$$\begin{array}{lccccc} & <\tau> & \xrightarrow{[\partial_2]=0} & 0 & \xrightarrow{[\partial_1]=0} & <A> \\ \text{kernels}: & Z_2=<\tau> & & Z_1=0 & & Z_0=<A> \\ \text{images}: & B_2=0 & & B_1=0 & & B_0=0 \\ \text{quotients}: & H_2=<\tau>\cong\mathbf{Z} & & H_1=0 & & H_0=<A>\cong\mathbf{Z} \end{array}$$

Both the chain complex and the homology groups coincide with those of the sphere, just as we expected. □

Exercise 5.91. Provide such a homology computation for the Möbius band.

Exercise 5.92. Prove that $\{C(K,\partial K),[\partial]\}$ is a chain complex.

---

6 Products

6.1 How products are built

The idea of the product may be traced to the image of a stack, which is a simple arrangement of multiple copies of X:



More complex outcomes result from attaching to every point of X a copy of Y:



Example 6.1 (vector spaces). As an example from linear algebra, what does the following identity mean?

$$\mathbf{R} \times \mathbf{R} = \mathbf{R}^2?$$

We can think of it as if a copy of the y-axis is attached to every point on the x-axis. Or, we can think in terms of *products of sets*:

$$x \in \mathbf{R}, y \in \mathbf{R} \Longrightarrow (x, y) \in \mathbf{R} \times \mathbf{R}.$$

□

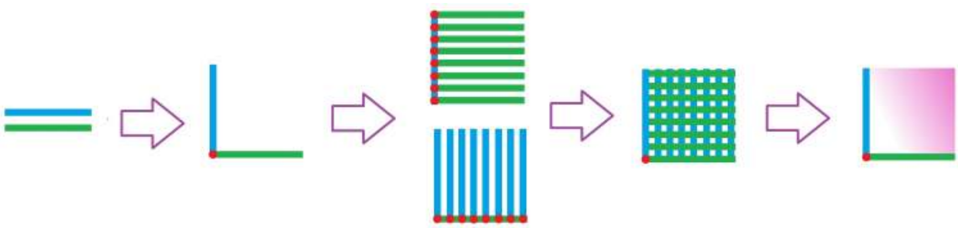
Generally, for any two sets X and Y, their *product set* is defined as the set of ordered pairs taken from X and Y:

$$X \times Y := \{(x, y) : x \in X, y \in Y\}.$$

Now, it is important to keep in mind that these three sets are just *bags of points*. How do these points form something tangible? Before we specify the topology, let's consider a few examples of visualization of products.

Example 6.2 (square). Let

$$[0, 1] \times [0, 1] := \mathbf{I} \times \mathbf{I} = \mathbf{I}^2$$



You can see how a copy of $Y = \mathbf{I}$ is attached to every point of $X = \mathbf{I}$, and vice versa. □

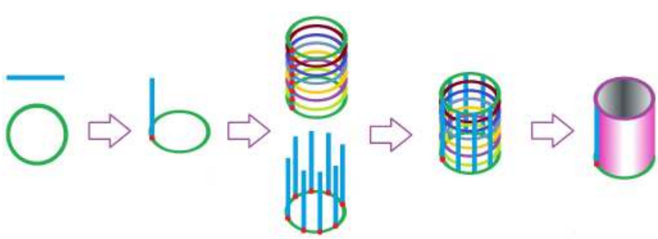
The building blocks here are just subsets of $\mathbf{R}$ and the construction simply follows that of $\mathbf{R} \times \mathbf{R}$.

Exercise 6.3. Provide a similar sketch for the cube:

$$\mathbf{I} \times \mathbf{I} \times \mathbf{I} = \mathbf{I} \times \mathbf{I}^2 = \mathbf{I}^3.$$

Example 6.4 (cylinder). Consider

$$\mathbf{S}^1 \times \mathbf{I}.$$



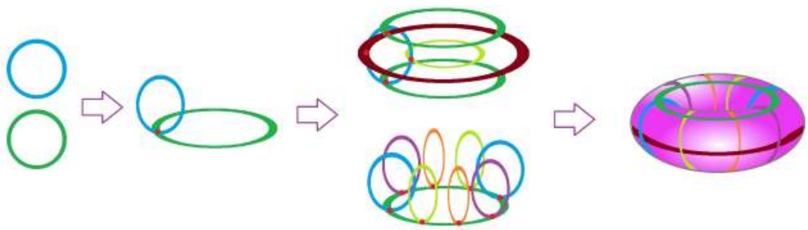
To build the cylinder this way, we place the circle $\mathbf{S}^1$ on the plane $\mathbf{R}^2$ and then attach a copy of $[0,1]$, vertically to each of its points. □

In all of these cases, the product fits into $\mathbf{R}^2$ or $\mathbf{R}^3$ and is easy to visualize. What if both sets are in $\mathbf{R}^2$?

Example 6.5 (torus). A similar procedure for the torus:

$$\mathbf{S}^1 \times \mathbf{S}^1 = \mathbf{T}^2$$

is impossible. As both copies of $\mathbf{S}^1$ lie in $\mathbf{R}^2$, the product would lie in $\mathbf{R}^4$. One, however, might think of small circles attached, vertically but with a turn, to each point of the large circle on the plane.

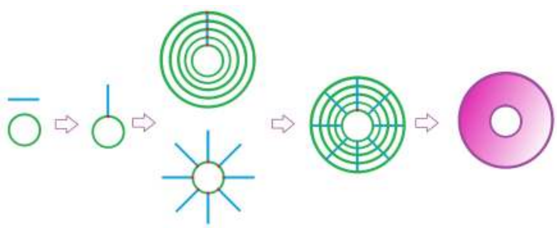


□

Example 6.6 (thickening). First, let's observe that we can understand the product

$$\mathbf{S}^1 \times \mathbf{I}$$

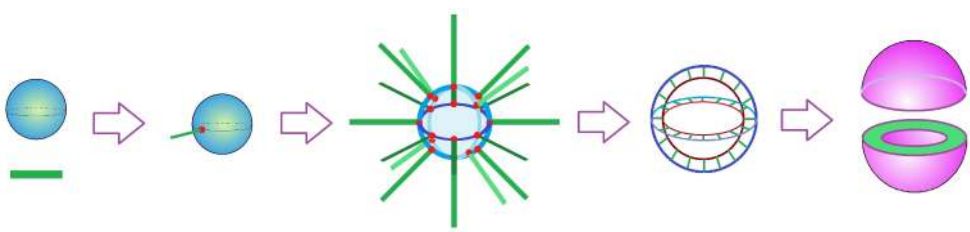
not as a cylinder but as a ring (annulus), if we attach the copies of Y in a different manner:



The idea is that the product of a space with the segment $\mathbf{I}$ means "thickening" of the space. As an example, the product

$$\mathbf{S}^2 \times \mathbf{I}$$

is a thickened sphere:



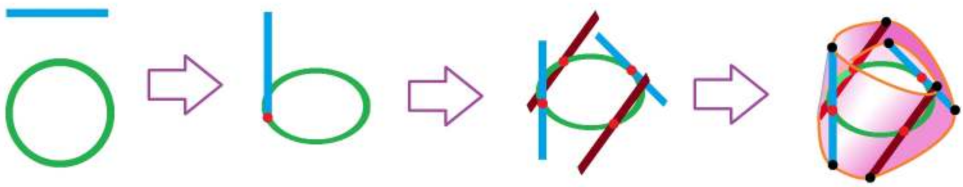
□

We have followed the rule:

- attach a copy of Y to every point of X (or vice versa).

Now, the ways these copies are attached aren't identical! We do see how this gluing is changing as we move in X from point to point. The change, however, is *continuous*. Is continuity enough to pinpoint what the topology of the product is? The example below shows that the answer is No.

Example 6.7. Just as with the cylinder, we are attaching segments to the circle but with a gradual rotation. The result is, of course, the Möbius band.



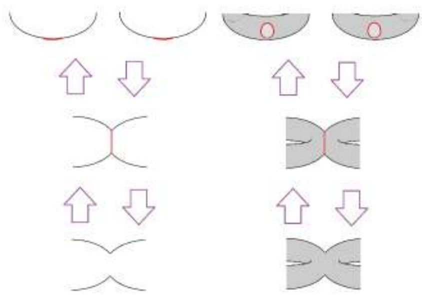
□

Exercise 6.8. Is the above rule violated in this construction?

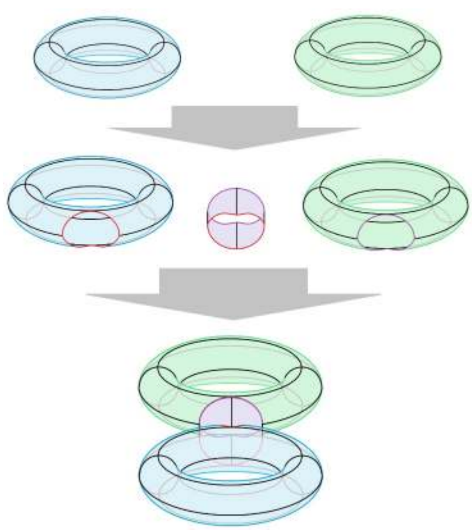
Example 6.9 (surgery). Recall that the connected sum takes two surfaces and attaches them to each other in two steps:

- punch a hole in either, and then
- glue the edges of the hole to each other.

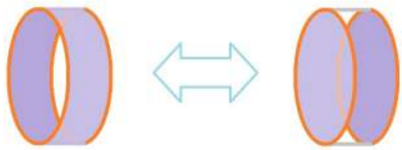
Below is an illustration of this construction along with its 1-dimensional analog:



Alternatively, we can look at this construction as one applied to a single manifold, either made of two pieces or not. Now, one can see this modification of the manifold as a replacement of its part:



The construction is called *surgery*. For surfaces, the replacement is of two disks with a cylinder. Note that together these three surfaces form the boundary of a solid cylinder, $\mathbf{B}^2 \times \mathbf{I}$.



Then we can see the surgery as a replacement of a certain product with another product

$$\partial \mathbf{I} \times \mathbf{B}^2 \longleftrightarrow \mathbf{I} \times \partial \mathbf{B}^2.$$

Such an operation is possible because the boundaries of these two sets are homeomorphic. □

In an $(n+m-1)$-manifold the surgery carries out this replacement:

$$\partial \mathbf{B}^m \times \mathbf{B}^n \longleftrightarrow \mathbf{B}^m \times \partial \mathbf{B}^n.$$

Exercise 6.10. Provide illustrations for surgeries with $n+m \leq 4$.

6.2 Products of spaces

What is the meaning of this identity:

$$\mathbf{R} \times \mathbf{R} = \mathbf{R}^2?$$

In linear algebra, the question arises because vector operations make sense with or without coordinates. One then needs to demonstrate how we can define the algebraic operations on the product set $\mathbf{R} \times \mathbf{R}$ in terms of the operations on either copy of $\mathbf{R}$, so that we have an isomorphism:

$$\mathbf{R} \times \mathbf{R} \cong \mathbf{R}^2.$$

Note that we have been using an alternative notation for vector spaces and groups:

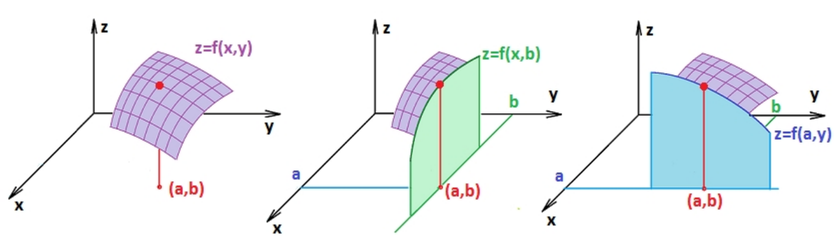
$$\mathbf{R} \oplus \mathbf{R} \cong \mathbf{R}^2.$$

Following this lead, we would like to demonstrate how we can define the topology on the product set $\mathbf{R} \times \mathbf{R}$ in terms of the topology on either copy of $\mathbf{R}$, so that we have an homeomorphism:

$$\mathbf{R} \times \mathbf{R} \approx \mathbf{R}^2.$$

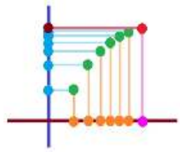
The question is: what is the relation between the topology of the xy-plane and the topologies of the x- and y-axes? The question appears indirectly in elementary calculus:

- $f(\cdot,\cdot)$ is continuous if and only if $f(\cdot,b)$ and $f(a,\cdot)$ are continuous for each a, b.

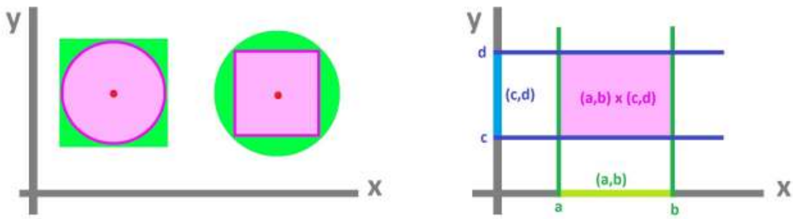


However, this isn't the definition of continuity of a function of two variables (that would make it dependent on the choice of the Cartesian system in $\mathbf{R}^2$) but rather a theorem. Its proof reduces to the question of convergence:

- $(x_n, y_n) \to (a, b) \iff x_n \to a$ and $y_n \to b$.



But, the convergence in $\mathbf{R}^2$ is based on the topology generated by disks, while that of $\mathbf{R}$ is based on intervals. The solution was demonstrated previously: the Euclidean topology of the plane coincides with the topology generated by *rectangles*:



And those rectangles are simply products of the intervals that come from the two copies of $\mathbf{R}$. Taking all pairwise products of intervals from the first copy of $\mathbf{R}$ and all from the second copy gives us the right topology on $\mathbf{R}^2$.

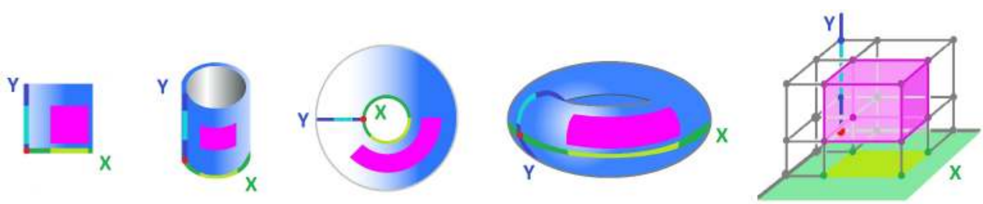
This analysis motivates the definition.

Definition 6.11. Given two topological spaces X and Y. Then the *product* $X \times Y$ of X and Y is a topological space defined on the product set $X \times Y$ with the following basis of neighborhoods:

$$\{U \times V : U \text{ open in } X,\ V \text{open in } Y\}.$$

The basis generates a topology called the *product topology*.

Pairwise products of the standard bases of the spaces we saw in the last subsection are shown below:



Now, as always, is this well-defined? In other words, is the collection defined this way always a basis?

Theorem 6.12. Given spaces X and Y with bases τ and σ respectively, the collection

$$\gamma := \{V \times W : V \in \tau, W \in \sigma\}$$

is a basis on $X \times Y$.

Proof. We need to verify the two axioms of basis.

For (B1), consider

$$\begin{aligned} \bigcup \gamma &= \bigcup \{V \times W : V \in \tau,\ W \in \sigma\} \\ &= \bigcup \Big(\{V : V \in \tau\} \times \{W : W \in \sigma\} \Big) \\ &= X \times Y, \end{aligned}$$

since both τ and σ satisfy (B1).

For (B2), suppose we have two neighborhoods A, A' in $X \times Y$ and a point u that belongs to their intersection:

$$\begin{aligned} A &= V \times W, & V \in \tau, W \in \sigma, \\ A' &= V' \times W', & V' \in \tau, W' \in \sigma, \\ u &= (x, y) \in A \cap A'. \end{aligned}$$

The last assumption implies that

$$x \in V \cap V',\ \ y \in W \cap W'.$$

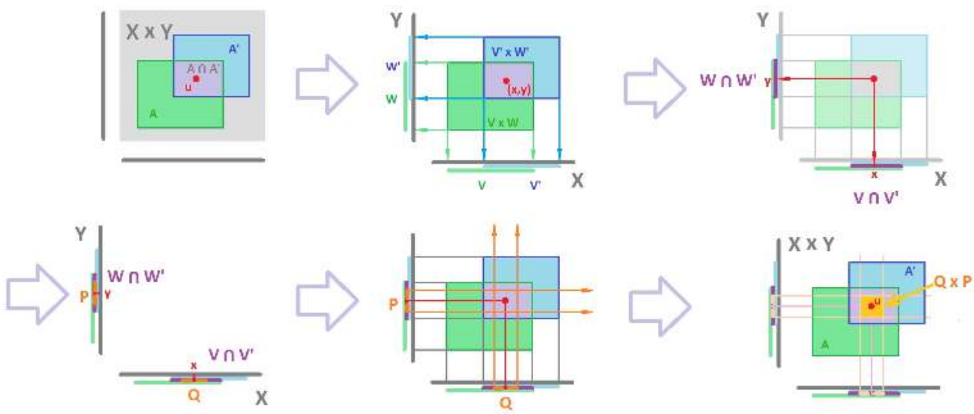
Then, from the fact that τ and σ satisfy (B2) we conclude:
- there is a $Q \in \tau$ such that $x \in Q \subset V \cap V'$, and
- there is a $P \in \sigma$ such that $y \in P \subset W \cap W'$.

Hence,

$$u = (x, y) \in Q \times P \subset V \times W \cap V' \times W' = A \times A'.$$

■

The construction followed this outline:



It is crucial to demonstrate that the product topology matches the implied topology of the spaces we build in the last subsection.

Exercise 6.13. Prove the following homeomorphisms:
- $\Big(n\text{-point set}\Big) \times \Big(m\text{-point set}\Big) \approx \Big(nm\text{-point set}\Big)$;
- $\mathbf{R} \times \{x\} \approx \mathbf{R}$;
- $\mathbf{R}^n \times \mathbf{R}^m \approx \mathbf{R}^{n+m}$;
- $\mathbf{I} \times \mathbf{I} \approx \mathbf{I}^2 =$ square;
- $\mathbf{I}^n \times \mathbf{I}^m \approx \mathbf{I}^{n+m}$;
- $\mathbf{S}^1 \times \mathbf{I} \approx$ the cylinder;
- $\mathbf{S}^1 \times \mathbf{S}^1 \approx \mathbf{T}^2 =$ the torus.

Exercise 6.14. Prove that the following is equivalent to the original definition:
- given two topological spaces X and Y with bases σ and τ respectively, the product $X \times Y$ is a topological space with the following basis:

$$\{U \times V : U \in \sigma,\ V \in \tau\}.$$

Exercise 6.15. Prove that the metric d of a metric space (X, d) is continuous as a function $d : X \times X \to \mathbf{R}$ on the product space.

Exercise 6.16. Prove:

$$X \times Y \approx Y \times X.$$

Exercise 6.17. Prove that the product of an n-manifold and an m-manifold is an $(n+m)$-manifold.

6.3 Properties

Many topological properties are preserved under products.

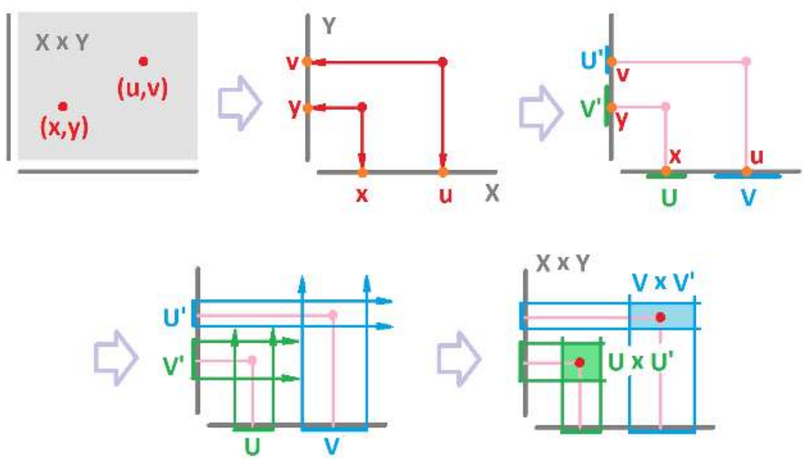
Theorem 6.18. If both X and Y are Hausdorff then so is $X \times Y$.

Proof. Suppose we have two distinct points (x, y) and (u, v) and we need to separate them by two disjoint neighborhoods.

There are only two possibilities.
- (1) x and u are distinct, then they can be separated by U and V in X.
- (2) y and v are distinct, then they can be separated by U' and V' in Y.

Then we have:
- if both (1) and (2) hold, (x, y) and (u, v) are separated by $U \times U'$ and $V \times V'$;
- if it's (1) but not (2), they are separated by $U \times Y$ and $V \times Y$;
- if it's (2) but not (1), they are separated by $X \times U'$ and $X \times V'$. ■

The construction followed this outline:



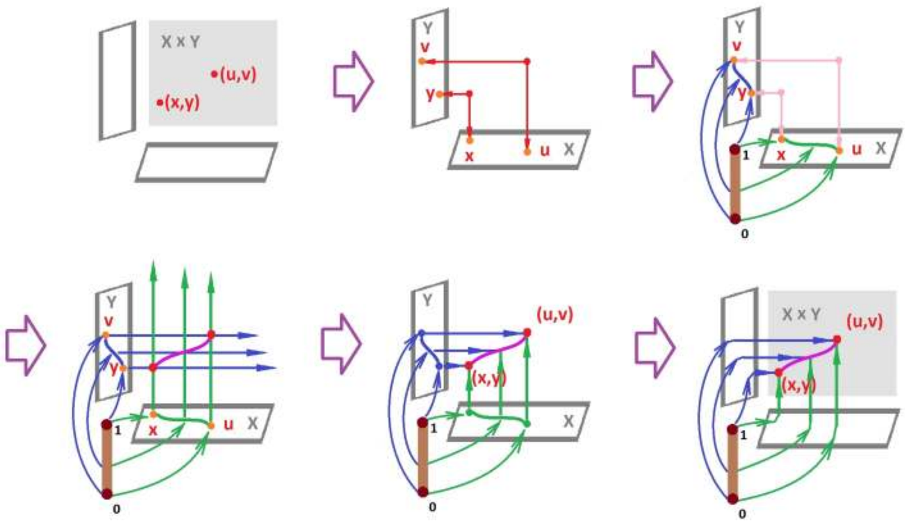
Theorem 6.19. If both X and Y are path-connected then so is $X \times Y$.

Proof. Suppose we have two points (u, v) and (x, y) in $X \times Y$, then we need to find a path from one to the other. From the path-connectedness of X and Y, it follows that there are continuous functions
- $f : [0, 1] \to X$ with $f(0) = u, f(1) = x$, and
- $g : [0, 1] \to Y$ with $g(0) = v, g(1) = y$.

Then the function
- $q : [0, 1] \to X \times Y$ defined by $q(t) := (f(t), g(t))$

gives us the path. ■

The construction followed this outline:



Exercise 6.20. Prove that q is continuous.

More general is the following.

Definition 6.21. Given two functions,

$$f : Z \to X,\ \ g : Z \to Y,$$

the *product function*

$$f \times g : Z \to X \times Y$$

is given by

$$(f \times g)(z) := (f(z), g(z)).$$

Theorem 6.22. If both f, g are continuous then so is their product $f \times g$.

Exercise 6.23. Prove the theorem.

Exercise 6.24. Even more general is the product of maps $f : A \to X,\ g : B \to Y$. Define it, show that it includes the last definition, and prove its continuity.

Recall that the graph of a map $f : X \to Y$ is a subset of $X \times Y$:

$$\operatorname{Graph} f := \{(x, y) \in X \times Y : y = f(x)\}.$$

Exercise 6.25. Prove that

$$\operatorname{Graph} f \approx X.$$

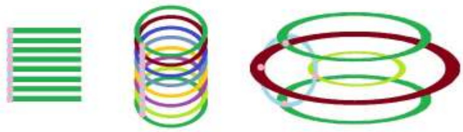
In particular, the graph of the identity map $\operatorname{Id}_X : X \to X$ coincides with the *diagonal*:

$$\operatorname{Graph} \operatorname{Id}_X = \Delta(X) := \{(x, y) \in X \times X :\ y = x\}.$$

Next, is compactness preserved under products?

Exercise 6.26. Prove that the answer is Yes for compact subsets X, Y of Euclidean spaces. Hint: you'll have to prove first that the product of two closed subsets is closed.

At least we know that $X \times Y$ is "made of" compact spaces:



Indeed,

$$X \times Y = \bigcup_{y \in Y} X_y,$$

where, for each $y \in Y$, the set

$$X_y := X \times \{y\} \subset X \times Y$$

is homeomorphic to X, which is compact.

Exercise 6.27. Show that this doesn't imply that $X \times Y$ is compact, unless Y is finite.

What about the general case?

We start over. We will try to apply the same approach as above: split the problem to those for X and Y, solve both separately, combine the solutions to form a solution for the product. This time, the problem is that of finding a finite subcover of an open cover. Unfortunately, the process of creating a basis of $X \times Y$ from those of X and Y, as shown above, isn't reversible. For example, the intervals in $\mathbf{R}$ and the disks in $\mathbf{R}^2$ produce solid cylinders in $\mathbf{R}^3$ not balls. So, for the plan to work, we need to show that compactness holds even if we only deal with open covers of a particular kind.

Theorem 6.28. A topological space X is compact if and only if there is a basis β of its topology so that every cover of X by members of β has a finite subcover.

Exercise 6.29. Prove the theorem.

We start over. Suppose β_X and β_Y are bases of X and Y, respectively. According to the last theorem, we only need to prove that every cover of $X \times Y$ by members of the basis

$$\beta_{X \times Y} := \beta_X \times \beta_Y = \{U \times V : \ U \in \beta_X, \ V \in \beta_Y\}$$

has a finite subcover. Suppose γ is such a cover. Now, every of its elements is a product and we have

$$\gamma = \{U \times V : \ U \in \gamma_X \subset \beta_X, \ V \in \gamma_Y \subset \beta_Y\},$$

for some covers γ_X, γ_Y. Now we find a finite subcover in either of them, say, γ'_X, γ'_Y, and let

$$\gamma' := \gamma_X \times \gamma_Y,$$

which is a finite subcover of γ.

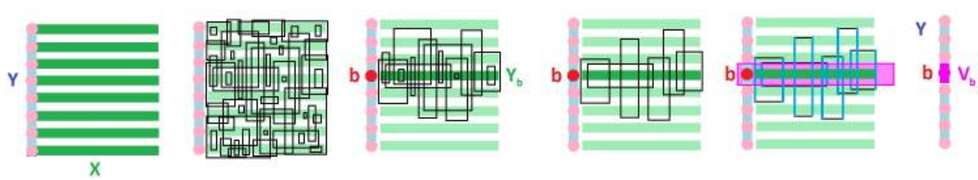
Exercise 6.30. Find a flaw in this argument.

We start over.

Theorem 6.31. If both X and Y are compact then so is $X \times Y$.

Proof. The *idea* is to use the compactness of every X_b, $b \in Y$, to come up with an open neighborhood $V_b \subset Y$ of b, and then use the compactness of Y.

The first stage is illustrated below:



Suppose β_X and β_Y are bases of X and Y, respectively. Suppose γ is a cover of $X \times Y$ the elements of which are products of the elements of the two bases:

$$\gamma = \{W = U \times V\},\ U \in \beta_X, V \in \beta_Y.$$

Choose any $b \in Y$. Then γ is a cover of Y_b. Since Y_b is compact, there is a finite subcover γ_b of γ. We then let

$$V_b := \bigcap \gamma_b.$$

As a finite intersection of open sets, V_b is open.

In the second stage, we first consider an open cover of Y:

$$\alpha := \{V_b : b \in Y\}.$$

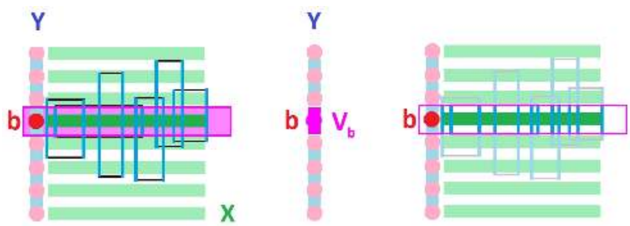
Then, since Y is compact, we can find its finite subcover:

$$\alpha' = \{V_b : b \in F\},$$

where F is some finite subset of Y. Finally, we set:

$$\gamma' := \{U \times V \in \gamma :\ U \times V \in \gamma_b, b \in F\}.$$ ■

Exercise 6.32. Prove that γ' is a finite subcover of γ. Hint:



Further, one can easily define and study products of finitely many spaces but products of an infinite number of spaces are also possible.

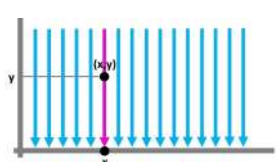
6.4 The projections

Just as the relative topology comes with the inclusion map and the quotient topology comes with the identification map, the product topology also comes with a new map – the projection.

Example 6.33. Let's start with the simple projection of the xy-plane on the x-axis,

$$p : \mathbf{R}^2 \to \mathbf{R}.$$

It is given by $p(x, y) = x$:



Let's make a couple of observations:

- for each b, the function $p(x, b) = x$ is continuous as the identity;
- for each a, the function $p(a, y) = a$ is continuous as a constant function.

Therefore, as we know from calculus, p is continuous. □

Definition 6.34. Suppose X, Y are two topological spaces. The *projections*

$$p_X : X \times Y \to X,\ p_Y : X \times Y \to Y,$$

of $X \times Y$ on X and Y respectively are defined by

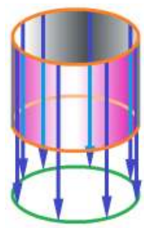
$$p_X(x, y) := x,\ p_Y(x, y) := y.$$

Theorem 6.35. The projections are continuous.

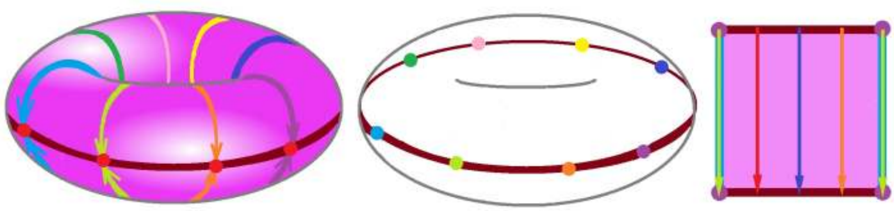
Proof. Suppose V is open in X. Then, $p_X^{-1}(V) = V \times Y$, which is open in $X \times Y$ in the product topology, since Y is open in Y. ■

Exercise 6.36. What is the linear algebra analog of this statement?

Example 6.37. The projection of the cylinder on the circle "looks" continuous – one can imagine an old chimney collapsing to the ground:



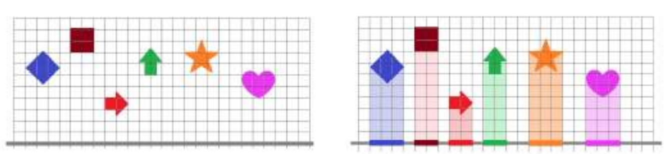
The meaning of the projection of torus $p : \mathbf{T}^2 \to \mathbf{S}^1$ is not as obvious because it might seem that in order to map it onto the equator you'd have to tear it:



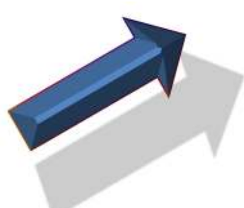
□

Exercise 6.38. Sketch the projection of the torus onto one of its meridians.

The projection of the square $\mathbf{I}^2 = [0, 1] \times [0, 1]$ to the x-axis is continuous as the restriction of the map p above. In fact, any subset in the plane can be projected to the x-axis:



This is the reason why the restrictions of the projection are also called by the same name. The idea is suggested by that of a *shadow*, such as the one of this arrow:



Exercise 6.39. Describe projections with a commutative diagram. Hint: you'll need to use inclusions.

Theorem 6.40.

$$(X \times Y)/_Y \approx X.$$

Exercise 6.41. Prove the theorem.

A more general concept is that of a *self-projection*, which is any self-map $P : X \to X$ that satisfies $PP = P$.

Exercise 6.42. Assuming that a self-projection p is a realization of a cell map P, what can you way about $\det P_*$?

Exercise 6.43. Suggest a non-constant self-projection of the Möbius band.

6.5 Products of complexes

If we are able to decompose a topological space into the product of two others, $Z = X \times Y$, we expect X and Y to be simpler than Z and to help us understand Z better. An example is the torus as the product of two copies of the circle, which reveals the two tunnels. To make this idea more precise, we need to apply it to spaces that are realizations of complexes. The hope is that our data structure will follow the topology:

$$|K \times L| \approx |K| \times |L|,$$

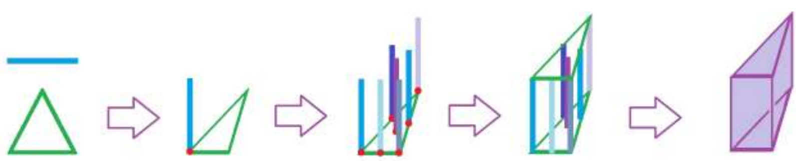
and we may be able to express the homology groups of $K \times L$ in terms of those of K and L. To start on this path, we need to define the product of two complexes.

The construction should be very simple: for two complexes K, L, let

$$K \times L := \{a \times b : \ a \in K, b \in L\}.$$

These are pairwise products of every cell of K and every cell in L, of all (possible different) dimensions! What's left is to define the product of two cells.

Simplicial complexes have proven to be the easiest to deal with, until now. The problem we encounter is at the very beginning: the product of two simplices isn't a simplex! It's a prism:



Then, we can't study the product $K \times L$ of simplicial complexes without further triangulation.

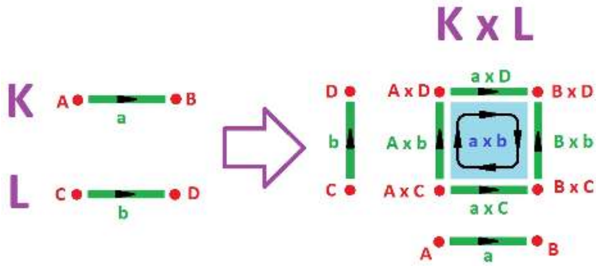
Exercise 6.44. Triangulate the prism.

Fortunately, *cubical complexes* have no such problem because the product of two cubes is a cube! In fact, the product of an n-cube and an m-cube is an $(n + m)$-cube:

$$\mathbf{I}^n \times \mathbf{I}^m = \mathbf{I}^{n+m}.$$

In other words, if a and b are n- and m-cells respectively, then $a \times b$ is an $(n + m)$-cell.

Example 6.45 (segment times segment). Suppose we have two copies of the complex that represents the segment:



Then K has 3 cells:

- 0-cells: A and B,
- 1-cell: a;

and L has 3 cells too:

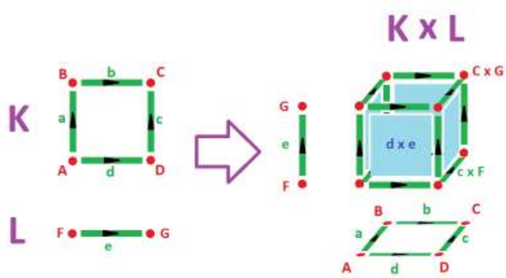
- 0-cells: C and D,
- 1-cell: b.

Now, to find the product of K and L, we take a cell x from K and a cell y from L, and add their product $x \times y$ to the list. Then $K \times L$ has $3 \times 3 = 9$ cells:

- 0-cells: $A \times B$, $A \times C$, $B \times C$, $B \times D$;
- 1-cells: $A \times b$, $B \times b$, $a \times C$, $a \times D$;
- 2-cells: $a \times b$.

What we have is the complex of the square. □

Example 6.46 (hollow square times segment). More complicated is the product of the complexes of a hollow square and a segment:



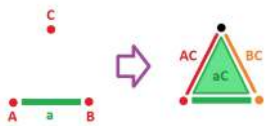
We have for $K \times L$:

- 0-cells: 8,
- 1-cells: 8,
- 2-cells: 4.

This is the finite analog of the topological product of the circle and the segment, $\mathbf{S}^1 \times \mathbf{I}$. The result is the cylinder, as expected. □

Exercise 6.47. Finish the example.

Back to simplicial complexes. It is in fact easy to build new simplices from old! Consider how adding a new vertex instantly gives as several:



It is simply a matter of adding a new vertex to each list, which defines a certain operation of simplices:

- $A \cdot C = AC$;
- $B \cdot C = BC$;
- $a \cdot C = AB \cdot C = ABC = aC$.

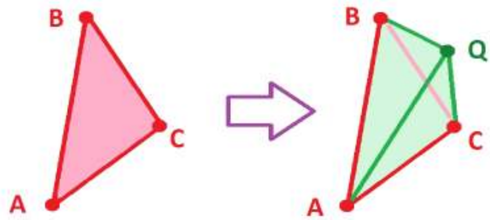
How do we build an $(n+1)$-simplex from an n-simplex? Suppose we have n geometrically independent (in general position) points $A_0, ..., A_n \in \mathbf{R}^N$ that represent n-simplex

$$\sigma := A_0 ... A_n.$$

Suppose another vertex $A_{n+1} = Q$ is also available. If the vertex is geometrically independent of the rest, we have a desired $(n+1)$-simplex

$$\tau := A_0 ... A_n A_{n+1}.$$

This is called the *cone construction*.



Meanwhile, on the data side, we are simply adding a new element to the list of vertices of the original simplex, as above.

The construction suggests a simple idea of how to define "products" of simplices – just combine their lists of vertices!

Definition 6.48. The *join* of an n-simplex $\sigma = A_0 ... A_n$ and an m-simplex $\tau = B_0 ... B_m$ is an $(n+m+1)$-simplex

$$\sigma \cdot \tau := A_0 ... A_n B_0 ... B_m.$$

For example, the 3-simplex is the join of two 2-simplices:

This operation isn't commutative because the order affects the orientation of the new simplex.

Exercise 6.49. Prove:

$$\sigma \cdot \tau = (-1)^{(\dim \sigma + 1)(\dim \tau + 1)} \tau \cdot \sigma.$$

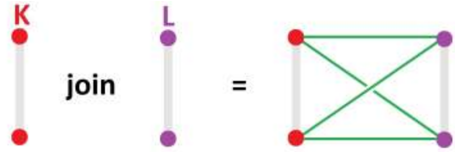
Exercise 6.50. Describe the geometric realization of the join by expressing $|a \cdot b|$ in terms of $|a|, |b|$.

Definition 6.51. The *join of two simplicial complexes* is defined as the set of all pairwise joins of the simplices (including the empty simplex) of the two complexes:

$$K \cdot L = \{a \cdot b : \ a \in K, b \in L\}.$$

Exercise 6.52. Prove that $K \cdot L$ is a simplicial complex.

Example 6.53. The circle is the join of two 2-point complexes:



□

Exercise 6.54. Represent these simplicial complexes as joins:
- the segment, the disk, the n-ball;
- the sphere, the n-sphere.

6.6 Chains in products

What effect does forming the product of two *cubical complexes* have on the homology? Is the homology group of the product equal to the product of the homology groups? The idea seems to be confirmed by our most popular example of product, the torus:

$$H_1(\mathbf{T}^2) = H_1(\mathbf{S}^1 \times \mathbf{S}^1) \cong H_1(\mathbf{S}^1) \times H_1(\mathbf{S}^1).$$

We need to understand what happens to the chain groups first, in the cubical case.

According to the last subsection, the *cells* of K are "cross-multiplied" with those of L:
- an i-cell in K and a j-cell in L are combined to create an $(i+j)$-cell in $K \times L$.

As linear combinations of cells, the *chains* of K are also "cross-multiplied" with those of L:
- an i-chain in K and a j-chain in L are combined to create an $(i+j)$-chain in $K \times L$.

Consequently, we won't try to compute the kth chain groups $C_k(K \times L)$ of the product from the groups $C_k(K)$ and $C_k(L)$ of chains of the *same* dimension. Instead, we'll look at the *complementary* dimensions. In other words,
- $C_k(K \times L)$ is found from the pairwise products of the elements of $C_i(K)$ and the elements of $C_j(L)$ for all pairs (i,j) with $i+j=k$.

Taken together, these correspondences create a function called the *cross product*:

$$\times : C_i(K) \times C_j(L) \to C_{i+j}(K \times L),$$

given by

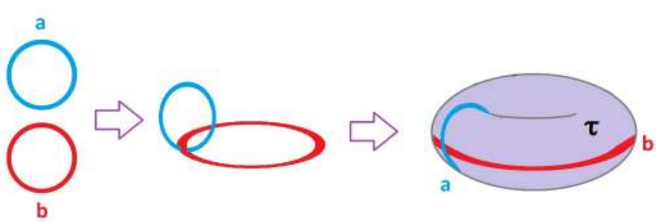
$$(x, y) \mapsto x \times y.$$

Proposition 6.55. The cross product is
- *bilinear*; i.e., it is linear on either of the two arguments; and
- *natural*; i.e., for any cubical maps $f : K \to K'$, $g : L \to L'$, we have

$$(f,g)_\Delta(a \times b) = f_\Delta(a) \times g_\Delta(b).$$

Note that this idea will later apply to homology: the kth homology $H_k(K \times L)$ is expressed as the sum of pairwise combinations of $H_k(K)$ and $H_{i-k}(L)$ for all i (the Künneth formula).

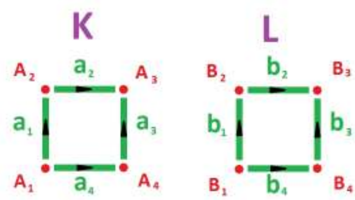
Example 6.56 (torus). Let's consider the torus again.



The torus itself, a 2-manifold, is the product of the two circles, 1-manifolds, and their 1st homology groups are generated by the 1-cycles, the fundamental classes, a and b. Now, we conjecture that the 2-cycle τ of the void of the torus, the fundamental class of this 2-manifold, is constructed as the product of these two 1-classes a, b.

This reasoning applies to the other homology classes of the torus. The longitude (or the latitude) of the torus, a 1-manifold, is the product of a circle, 1-manifold, and a point, 0-manifold. We conjecture that the 1-cycle represented by the longitude (latitude) is constructed as the product of the 1-cycle of the first circle and the 0-cycle of the second circle. If this is true, the identity we started with was just a coincidence!

Let's work out this example algebraically...



We list the chain groups of these two complexes and their generators:

$$\begin{array}{ll} C_0(K) =< A_1, A_2, A_3, A_4 >, & C_1(K) =< a_1, a_2, a_3, a_4 >; \\ C_0(L) =< B_1, B_2, B_3, B_4 >, & C_1(L) =< b_1, b_2, b_3, b_4 > . \end{array}$$

Then, for the product

$$M = K \times L,$$

we have:

$$\begin{array}{lll} C_0(M) =< A_i \times B_j & : i, j = 1, 2, 3, 4 >; \\ C_1(M) =< A_i \times b_j, \ a_i \times B_j & : i, j = 1, 2, 3, 4 >; \\ C_2(M) =< a_i \times b_j & : i, j = 1, 2, 3, 4 > . \end{array}$$

□

As we know, the boundary operator of a cubical complex can be defined on the cubes as products of cubes of lower dimension according to this Leibniz-type formula:

$$\partial^M (a \times b) = \partial^K a \times b + (-1)^{\dim a} a \times \partial^L b.$$

This formula is extended to the chains and then serves as the boundary operator

$$\partial^M : C_k(M) \to C_{k-1}(M)$$

of the product complex $M := K \times L$. We also know that

$$\partial^M \partial^M = 0.$$

Example 6.57 (torus). To confirm this idea, let's compute the boundary operator for the torus M:

$$\begin{array}{lll} \partial(A_i \times B_j) & = \partial A_i \times B_j + A_i \times \partial B_j & = 0; \\ \partial(A_i \times b_j) & = \partial A_i \times b_j + A_i \times \partial b_j & = A_i \times \partial b_j; \\ \partial(a_i \times B_j) & = \partial a_i \times B_j + a_i \times \partial B_j & = \partial a_i \times B_j; \\ \partial(a_i \times b_j) & = \partial a_i \times b_j + a_i \times \partial b_j. & \end{array}$$

What we see is this:

$$\begin{array}{ll} \ker \partial^M & = \ker \partial^L \times \ker \partial^L, \\ \operatorname{Im} \partial^M & = \operatorname{Im} \partial^L \times \operatorname{Im} \partial^L. \end{array}$$

□

Exercise 6.58. Confirm these identities for the torus example.

Exercise 6.59. Consider the projections:

$$p_K : K \times L \to K,\ p_L : K \times L \to L,$$

and find their chain maps:

$$C(K \times L) \to C(K),\ C(K \times L) \to C(L).$$

Under products, the chains are easy to handle but, as we shall see, the homology classes aren't...

There is an analog of the Leibniz-like formula above for the joins.

Proposition 6.60. For two simplicial complexes $\{K, \partial^K\}$, $\{L, \partial^L\}$, the boundary operator of their join $k \cdot L$ is given by

$$\partial^M (a \cdot b) = \partial^K a \cdot b + (-1)^{n+1} a \cdot \partial^L b.$$

Exercise 6.61. Prove the proposition.

6.7 The Universe: 3-sphere, 3-torus, or something else?

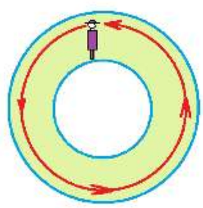
Our everyday experience suggests that the universe is 3-dimensional and, probably, a manifold. In addition, we will assume that this manifold is orientable, compact, with empty boundary.

But first let's try to understand the "lower-dimensional universes"...

If the universe is 1-dimensional, it's a topological circle. To understand how it feels to live in such a world, we thicken this circle:

$$\mathbf{S}^1 \times \mathbf{B}^2.$$

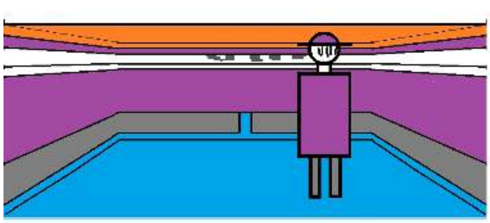
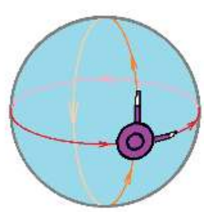
Since the light makes round trips around the circle, this is what an observer would see:

If the universe is 2-dimensional, there are many choices. Once again, we thicken this surface to be able to fit into it:

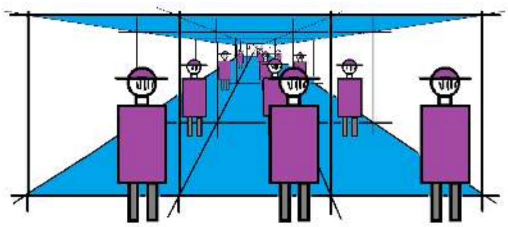
$$M^2 \times \mathbf{I}.$$

The simplest choice for M is the sphere $\mathbf{S}^2$. Here the light comes back no matter in what direction it leaves. This is what an observer would see:

If there is nothing else in the universe, you'd see your own, panoramic image stretched around the walls with the back of your head in front of you.

The second simplest choice for M is the torus $\mathbf{T}^2$. Here the light goes through the wall on the left and reappears from the right. Or it goes through the wall in front and reappears from the back. This is what an observer would see:

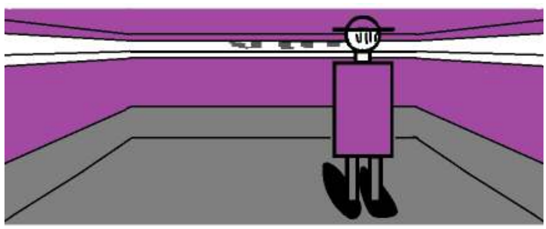


If there is nothing else in the universe, you'd see your own image in all four directions repeated infinitely many times. Through these observations, we recognize the product of two 1-dimensional universes either one seen an enfilade of rooms.

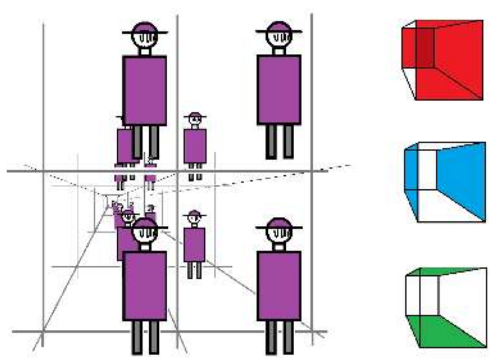
The triple torus will have the same affect with three pairs of walls having this property, etc.

What about the real, 3-dimensional universe?

There are even more choices. The 3-sphere will be similar to the 2-sphere. One will see the 3d image of himself stretched over all walls and the ceiling and the floor:



The next option is the one we have seen before: the box with the opposite faces identified:



It is similar to the case of $\mathbf{T}^2$ with one's image repeated in six directions instead of four. And, one will see three enfilades of rooms instead of two. This suggests that the space is the product of three copies of the circle, which is the 3-dimensional analog of the torus:

$$\mathbf{T}^3 = \mathbf{S}^1 \times \mathbf{S}^1 \times \mathbf{S}^1 .$$

Exercise 6.62. Prove this statement.

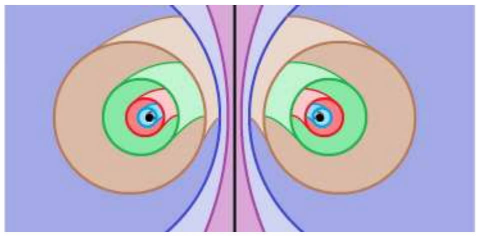
Even though these two spaces *look* different from the inside, how do we know that they *are* different? Just consider the homology groups:

$$\begin{array}{llllll} H(\mathbf{S}^3) & = \{\mathbf{Z}, & 0, & 0, & \mathbf{Z}, & 0, ...\}, \\ H(\mathbf{T}^3) & = \{\mathbf{Z}, & \mathbf{Z}\oplus\mathbf{Z}\oplus\mathbf{Z}, & \mathbf{Z}\oplus\mathbf{Z}\oplus\mathbf{Z}, & \mathbf{Z}, & 0, ...\}. \end{array}$$

Exercise 6.63. Suggest a plan how to find out.

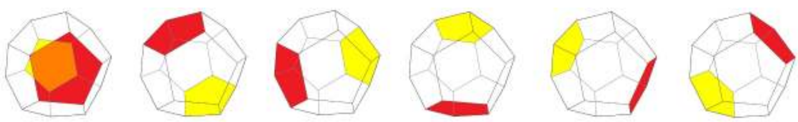
Exercise 6.64. Can you build the universe from a tetrahedron?

This is how $\mathbf{R}^3$ can be decomposed into infinitely many tori:



Exercise 6.65. Imagine a universe made of two solid tori glued to each other along their boundaries. What is it homeomorphic to?

Another candidate for the model of the universe is the *Poincaré homology sphere*. The name is justified by the fact that, even though it is not homeomorphic to the 3-sphere, their homology groups coincide. This space is built from a regular dodecahedron by gluing together its opposite faces, the 6 pairs of pentagons:

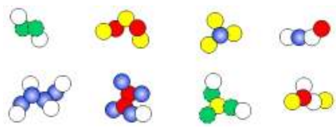


Each face is turned 32 degrees for gluing and this is what an observer would see:

Another issue is, does the universe have to be orientable? Can one come back as a mirror image of himself?

Exercise 6.66. Devise a universe where such a trip is possible.

Exercise 6.67. From these examples of molecules, choose ones that would survive intact such an around the (non-orientable) world trip:



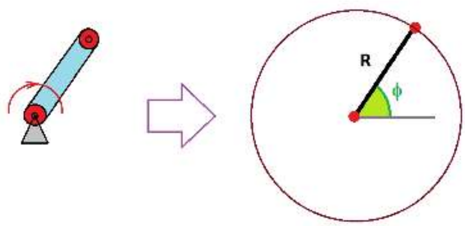
6.8 Configuration spaces

We will consider a few simple designs for a robotic arm.

Example 6.68 (rotating rod). Suppose the robotic arm has a single joint, i.e., just a rotating rod. What is the set of all possible positions of its end? It's the circle $\mathbf{S}^1$ parametrized in the standard way:

$$x = R\cos\phi,\ \ y = 0, z = R\sin\phi,$$

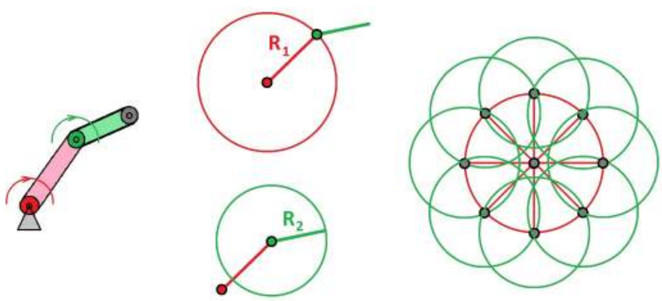
with ϕ the angle of the rod and R is the (fixed) length of the rod.



□

Exercise 6.69. Show that, if this is a *ball joint* instead, the set of possibilities is the sphere $\mathbf{S}^2$.

Example 6.70 (two-joint arm). Further, the set of possible positions of a two-joint arm is more complicated:



The locations can be parametrized, but not uniquely, by

$$x = R_1\cos\phi_1 + R_2\cos\phi_2,\ \ y = 0,\ \ z = R_1\sin\phi_1 + R_2\sin\phi_2,$$

with ϕ_1, ϕ_2 the angles of the two arms and R_1, R_2 are the lengths of the arms. □

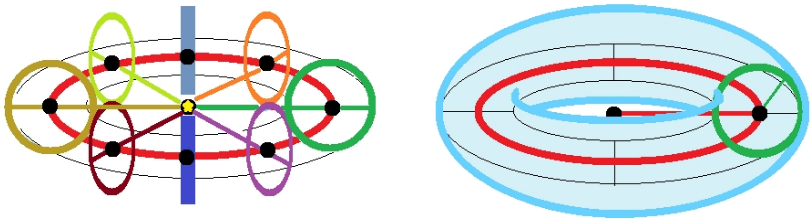
Exercise 6.71. Prove that it's either the disk or the annulus.

Meanwhile, the space of *states* of the arm is simply the set of all possible values of the angles ϕ_1, ϕ_2. That's the torus:

$$\mathbf{T}^2 = \mathbf{S}^1 \times \mathbf{S}^1.$$

This is the reason why we separate the two:
- the *operational space* as the set of all positions reachable by a robot's end in space, and
- the *configuration space* as the set of all possible combinations of the positions of the joints.

Example 6.72. For the two-joint arm, we could make these two homeomorphic to each other if we make the axes of the joints (first red, second green) perpendicular to each other (with the latter shorter than the former):



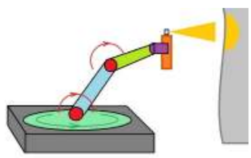
□

This can't happen with 3 or more joints as there will be $n \geq 3$ parameters. The configuration space of an n-joint arm (called also a *linkage*) is the n-torus:

$$\mathbf{T}^n = \mathbf{S}^1 \times ... \times \mathbf{S}^1,$$

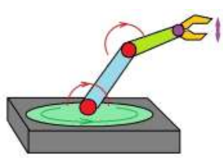
which won't fit (can't be embedded) into our 3-dimensional Euclidean space.

Example 6.73. Furthermore, even if the arm has delivered the end of the arm to its intended *location*, the task at hand, such as spray-painting, may require a particular *direction*:



This would add an extra term $\times \mathbf{S}^2$ to the operational space. □

In general, such a setup might be much more complicated with a number of other features.



Exercise 6.74. Find the configuration space of this robotic arm. Hint: don't forget the fingers!

So, the positions of the joints are used as the parameters of the state of the robot.

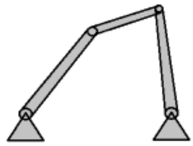
If the motion of the joints is independent (and we ignore the possible self-intersections), *the configuration space is the product of the configuration spaces of the joints.*

For example, a telescoping arm with a joint has the cylinder, $C = [0, 1] \times \mathbf{S}^1$, as the configuration space.

These examples suggest that configuration spaces (but not operational spaces) should be manifolds.

Exercise 6.75. What is the configuration space of a pendulum attached to a spring?

Exercise 6.76. What is the configuration space of the three-joint arm the end of which is fixed? Hint: the answer will depend on the relative lengths of the rods.



The robot's *forward and inverse kinematics equations* define functions from its configuration space to the operational space. These functions are then used for motion planning.

In physics, an example of a configuration space is *the state space of* n *particles.* In the simplest setting, it is

$$C=(\mathbf{R}^3)^n.$$

Every configuration corresponds to a single point in this space.

The answer changes, however, if we drop the implicit assumption that the particles are *distinguishable.* In that case, the configuration space is

$$C=(\mathbf{R}^3)^n/_\sim,$$

where $\sim$ is an equivalence relation derived from the identification of each pair of particles.

Exercise 6.77. What is the configuration space of two and three particle systems with identical particles?

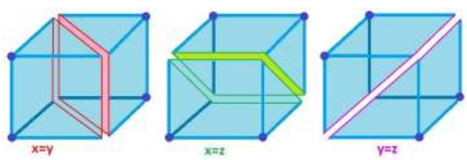
Another issue is, can two particles occupy the same location? If the answer is No, we need to exclude the diagonals from the configuration space:

$$C=(\mathbf{R}^3)^n\setminus D,$$

where

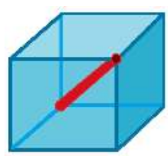
$$D=\{(u_1,...,u_n)\in(\mathbf{R}^3)^n:\ \exists i\neq j, u_i=u_j\}.$$

This idea is used to plan for *collision avoidance* in robotics by constructing such a "safe" configuration space. On a line segment, this configuration space of three robots will be a cube with three planes cut out:



Exercise 6.78. How many path-connected components does this space have?

If two can pass each other (but not three), the configuration space is a cube with the diagonal drilled out:



It is connected!

Exercise 6.79. Find the safe configuration space for two robots on a line, the circle, the tripod. What if they are connected by a rod? a rope?

In general, if the n particles move within a manifold M of dimension m, the configuration space is also a manifold but of dimension mn.

If molecular *bonds* are present, things are more complicated. Suppose, for example, we have two atoms in space. Then their configuration space is $(\mathbf{R}^3)^2$. Now suppose this is a two-atom molecule with two (different) atoms. Then its state is described, independently, by

- the location of its center of mass (or either of the two atoms) as a point of reference, and
- the orientation of one of the atoms with respect to that point (the distance is fixed).

Therefore, the configuration space is $\mathbf{R}^3 \times \mathbf{S}^2$.

Exercise 6.80. What if this is a two-atom molecule with two *identical* atoms?

One can also take into account the velocity (or the momentum) of each particle moving within manifold M. These velocities combined give us what we call the tangent bundle TM of M. The totality of all possible positions and velocities is called the *phase space* of the system.

Exercise 6.81. Show that the phase space of a pendulum is the cylinder.

6.9 Homology of products: the Künneth formula

What happens to the homology when we take the product of two, in the simplest case cubical, complexes? In other words, given two complexes K and L, express $H(K \times L)$ entirely in terms of $H(K)$ and $H(L)$.

It would be naive to assume that the answer is "it's the product of the homology groups", because we have already seen that taking the direct sums of the homology groups of the same dimension doesn't produce the desired results:

$$H_2(\mathbf{T}^2) = H_2(\mathbf{S}^1 \times \mathbf{S}^1) = \mathbf{R} \neq H_2(\mathbf{S}^1) \oplus H_2(\mathbf{S}^1) = 0 \oplus 0 = 0.$$

As discussed above, we need instead to look at the *complementary* dimensions. After all, the product of an n-cube in K and an m-cube in L in an $(n+m)$-cube in $K \times L$. For example, to find all 2-cubes in $K \times L$ one has to look at all the products of 1-cubes in K with 1-cubes in L as well as 0- with 2-cubes. More generally, we look at the algebraic decompositions of the dimension k of the cubes we are considering:

- $k = 1 = 1 + 0 = 0 + 1$;
- $k = 2 = 2 + 0 = 1 + 1 = 0 + 2$;
- $k = 3 = 3 + 0 = 2 + 1 = 1 + 2 = 0 + 3$,
- ...
- $k = k + 0 = (k-1) + 1 = (k-2) + 2 = ... = 0 + k$.

There are exactly $k+1$ such decompositions of k.

This idea of decomposing the dimension k applies to the chains next. Each decomposition of k corresponds to a component of $C_k(K \times L)$:

$$C_k(K) \ \& \ C_0(L), \quad C_{k-1}(K) \ \& \ C_1(L), \quad ... \quad , C_0(K) \ \& \ C_k(L).$$

It follows then that the kth homology is the sum of the combinations:

$$H_k(K) \ \& \ H_0(L), \quad H_{k-1}(K) \ \& \ H_1(L), \quad ... \quad , H_0(K) \ \& \ H_k(L).$$

Now, how exactly do we combine each of these pairs,

$$V = H_i(K), \ W = H_j(L), \ i + j = k?$$

After all, we know that the product won't work.

We provide the definition of this new operation for two arbitrary vector spaces V and W over field R.

First, we consider the *product set* $V \times W$ of the vector spaces as sets (rather than vector spaces), so that it consists of *all* pairs (v, w) with $v \in V$ and $w \in W$.

Second, we define the *free vector space* $< V \times W >$ of this space as the vector space of all formal linear combinations of the elements of $V \times W$. In other words, $V \times W$ serves as its basis because the relations among the elements of V, W are lost in the new vector space.

Third, we consider a certain *quotient vector space* of $< V \times W >$ as follows.

We consider the subspace Z of $< V \times W >$ generated by the following elements:

$$\begin{array}{l} (v_1, w) + (v_2, w) - (v_1 + v_2, w), \\ (v, w_1) + (v, w_2) - (v, w_1 + w_2), \\ c \cdot (v, w) - (cv, w), \\ c \cdot (v, w) - (v, cw), \end{array}$$

where

$$v, v_1, v_2 \in V, \ \ w, w_1, w_2 \in W, \ \ c \in R.$$

To simplify the notation, we drop the coefficient if it is equal to 1; i.e., (v, w) stands for $1 \cdot (v, w) \in < V \times W >$.

Then the *tensor product* of V and W is defined to be

$$V \otimes W := < V \times W > /Z.$$

Also, the tensor product of two vectors $v \in V$ and $w \in W$ is the equivalence class(coset) of (v, w):

$$v \otimes w := (v, w) + Z \in V \otimes W.$$

The elements of this form are called *elementary tensors.*

Exercise 6.82. Prove that, if B_V, B_W are bases of V, W, then

$$\{v \otimes w : \ v \in B_V, \ w \in B_W\}$$

is a basis of $V \otimes W$. Hint: not all of the elements of the tensor product are elementary tensors.

It follows that these equations hold in $V \otimes W$:

$$\begin{array}{ll} (v_1 + v_2) \otimes w & = v_1 \otimes w + v_2 \otimes w; \\ v \otimes (w_1 + w_2) & = v \otimes w_1 + v \otimes w_2; \\ cv \otimes w & = v \otimes cw = c(v \otimes w). \end{array}$$

Exercise 6.83. Prove that $R \otimes V = V$ over field R.

The main result is stated below without proof (see Bredon, *Topology and Geometry*, p. 320).

Theorem 6.84 (Künneth Formula). For two cell complexes K, L, the homology over $\mathbf{R}$ is given by

$$H_k(K \times L) \cong \bigoplus_i H_k(K) \otimes H_{i-k}(L).$$

For future computations, we'll rely only on these two main properties of the tensor product:

Proposition 6.85.

- $\mathbf{R} \otimes \mathbf{R} = \mathbf{R}$,
- $\mathbf{R} \otimes 0 = \mathbf{R}$.

Example 6.86. Let's consider the torus and compute the sum of all tensor products of the homology groups of dimensions that add up to k, for each k.

First, the ones that add up to 0:

$$\begin{aligned} H_0(\mathbf{T}^2) &= H_0(\mathbf{S}^1 \times \mathbf{S}^1) \\ &= H_0(\mathbf{S}^1) \otimes H_0(\mathbf{S}^1) \\ &= \mathbf{R} \otimes \mathbf{R} = \mathbf{R}. \end{aligned}$$

Those that add up to 1:

$$\begin{aligned} H_1(\mathbf{T}^2) &= H_1(\mathbf{S}^1 \times \mathbf{S}^1) \\ &= H_1(\mathbf{S}^1) \otimes H_0(\mathbf{S}^1) \ \oplus \ H_0(\mathbf{S}^1) \otimes H_1(\mathbf{S}^1) \\ &= \mathbf{R} \otimes \mathbf{R} \quad \oplus \quad \mathbf{R} \otimes \mathbf{R} \\ &= \mathbf{R} \quad \oplus \quad \mathbf{R} = \mathbf{R}^2. \end{aligned}$$

Those that add up to 2:

$$\begin{aligned} H_2(\mathbf{T}^2) &= H_2(\mathbf{S}^1 \times \mathbf{S}^1) \\ &= H_2(\mathbf{S}^1) \otimes H_0(\mathbf{S}^1) \ \oplus \ H_1(\mathbf{S}^1) \otimes H_1(\mathbf{S}^1) \ \oplus \ H_0(\mathbf{S}^1) \otimes H_2(\mathbf{S}^1) \\ &= 0 \otimes \mathbf{R} \quad \oplus \quad \mathbf{R} \otimes \mathbf{R} \quad \oplus \quad \mathbf{R} \otimes 0 \\ &= 0 \quad \oplus \quad \mathbf{R} \quad \oplus \quad 0 = \mathbf{R}. \end{aligned}$$

The results match our previous computations. □

Exercise 6.87. Use the formula to compute the homology of the 3-torus.

Exercise 6.88. Sometimes the "naive product formula" does hold. Derive it for H_1 from the theorem:

$$H_1(K \times L) \cong H_1(K) \oplus H_1(L).$$

The tensor product is also defined for two modules over any ring R. Then the things are made (even) more complicated by the presence of the torsion. Fortunately, $H_p(K \times L)$ is only affected by the torsion of $H_k(K), H_k(L)$ for $k < p$. The result is the following generalization of the last formula.

Theorem 6.89 (Naive Product Formula). If complexes K, L are path-connected and, for integral homology, we have

$$H_k(K) = H_k(L) = 0, \ k = 1, 2, ..., p-1,$$

then

$$H_p(K \times L) \cong H_p(K) \oplus H_p(L).$$

The isomorphism $x \times y \mapsto (x, y)$ is natural.

Exercise 6.90. Prove that

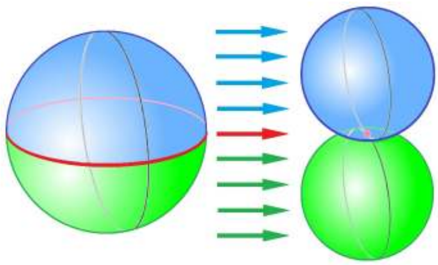
$$H_1(\mathbf{T}^n) = \mathbf{Z}^n.$$

Exercise 6.91. Prove that

$$H_p\Big((\mathbf{S}^p)^n \Big) = \mathbf{Z}^n.$$

Chapter V

Maps



1 Homotopy

1.1 Deforming spaces vs. deforming maps

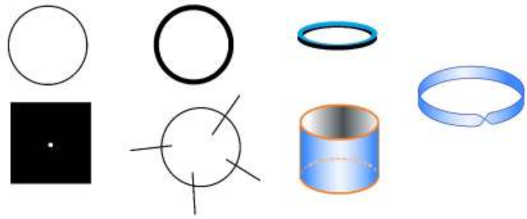
What do the figures below have in common?

The answer we have been giving is: they all have one hole. However, there is a profound reason *why* they must all have one hole. These space are homeomorphic!

The reasoning, still not fully justified, is transparent:

$$X \approx Y \Longrightarrow H(X) \cong H(Y).$$

Now, let's choose another collection. This time, the spaces aren't homeomorphic, but do they have anything in common?



The answer is the same: they all have one hole. But, once again, maybe there is a profound reason *why* they all have one hole.

Is there a relation between two topological spaces, besides homeomorphism, that ensures that they would have the same count of topological features? We will discover an equivalence relation that produces the same result for a much broader class of spaces than topologically equivalent:

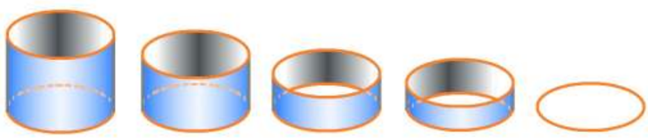
$$X \sim Y \Longrightarrow H(X) \cong H(Y).$$

Informally, we say that one space is "deformed into" the other.

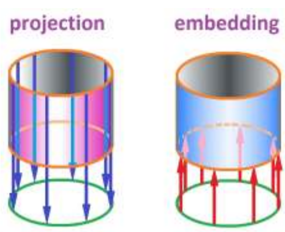
Let's try to understand the actual mathematics behind these words, with the help of this juxtaposition:

the *cylinder* vs. the *circle*.

The first meaning of the word "deformation" is a transformation that is *gradual*. Unlike a homeomorphism, which is meant to be instantaneous, this transformation is stretched over time through a continuum of intermediate states:



Let X be the cylinder and Y the circle. Let's take a look at the *maps* that naturally connect these two spaces. The first is the projection $p : X \to Y$ of the cylinder along its axis and the second is the embedding $e : Y \to X$ of the circle into the cylinder as one of its boundary circles:



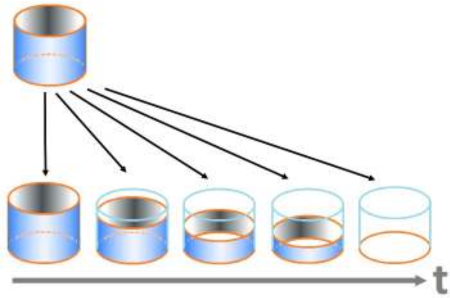
Both preserve the hole even though neither one is a homeomorphism.

Let's assume that X is the unit cylinder in $\mathbf{R}^3$ and Y is the unit circle in $\mathbf{R}^2 \subset \mathbf{R}^3$. Then the formulas for these functions are simple:

$$p(x, y, z) = (x, y), \ e(x, y) = (x, y, 0).$$

Let's next consider the compositions of these maps:
- $pe : Y \to Y$ is the identity Id_Y;
- $ep : X \to X$ is the collapse (and also a self-projection) of the cylinder on its bottom edge.

The latter, even though not the identity, is related to Id_X. This relation is seen through the continuum of maps that connects the two maps; we choose map $h_t : X \to X$ to *shrink* the cylinder – within itself – to height $t \in [0, 1]$. The images of these maps are shown below:



This is our main idea: we should interpret deformations of spaces via deformations of maps.

The formulas for these maps are simple:

$$h_t(x, y, z) = (x, y, tz),\ t \in [0, 1].$$

And, it is easy to confirm that we have what we need:

$$h_0 = ep,\ h_1 = \mathrm{Id}_X\,.$$

Looking closer at these maps, we realize that

- h_t is continuous for *each* t, but also
- h_t are continuous, as a whole, *over* t.

Therefore, the transformation can be combined into a single map

$$H(t, x, y, z) = h_t(x, y, z),$$

continuous with respect to the product topology.

The precise interpretation of this analysis is given by the two definitions below.

Definition 1.1. Two maps $f_0, f_1 : X \to Y$ are called *homotopic* if there is a map

$$F : [0, 1] \times X \to Y$$

such that

$$F(0, x) = f_0(x),\ F(1, x) = f_1(x),$$

for all $x \in X$. Such a map F is called a *homotopy* between f_0 and f_1. For this relation we use the following **notation**:

$$F : f_0 \simeq f_1,$$

or simply:

$$f_0 \simeq f_1.$$

Definition 1.2. Suppose that X and Y are topological spaces and $f : X \to Y,\ g : X \to Y$ are maps, and fg and gf are homotopic to the identity maps on Y and X respectively:

$$fg \simeq \mathrm{Id}_Y,\ gf \simeq \mathrm{Id}_X,$$

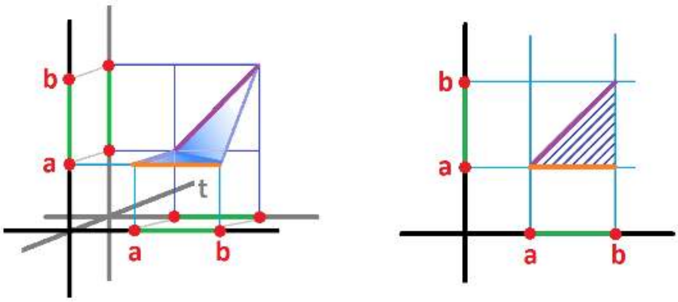
then f is called a *homotopy equivalence*. In this case, X and Y are called *homotopy equivalent*.

We set the latter concept aside for now and study properties of homotopy.

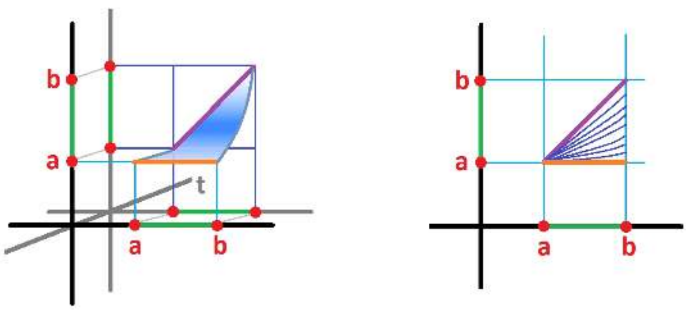
1.2 Properties and examples

It is often hard to visualize a homotopy via its graph, unless the dimensions of the spaces are low. Below we have the graph of a homotopy (blue) between a constant map (orange) and the identity (purple) of an interval:

$$c, \mathrm{Id} : [a, b] \to [a, b].$$



The diagram on the left demonstrates that $c \simeq \mathrm{Id}$ by showing the homotopy as a surface that “connects” these two maps. On the right, we also provide a more common way to illustrate a homotopy – by plotting the “intermediate” maps. Those are, of course, simply the vertical cross-sections of this surface. The homotopy above is piecewise linear and the one below is differentiable:



Theorem 1.3. Homotopy is an equivalence relation: for two given topological spaces X and Y, the space $C(X, Y)$ of maps from X to Y is partitioned into equivalence classes:

$$[f] := \{g : X \to Y,\ g \simeq f\}.$$

Proof. The axioms of equivalence relations can be understood in terms of what happens to these gradual transformations.

1. Reflexivity: do nothing. For $F : f \simeq f$, choose $H(t, x) = f(x)$.

2. Symmetry: reverse time. Given $H : f \simeq g$, choose $F(t, x) = H(1 - t, x)$ for $F : g \simeq f$.

3. Transitivity: make it two-stage. We need to demonstrate this:

$$F : f \simeq g,\ G : g \simeq h \Longrightarrow H = ? : f \simeq h.$$

We simply carry out these two processes consecutively but, since it has to be within the same time frame, twice as fast. We define:

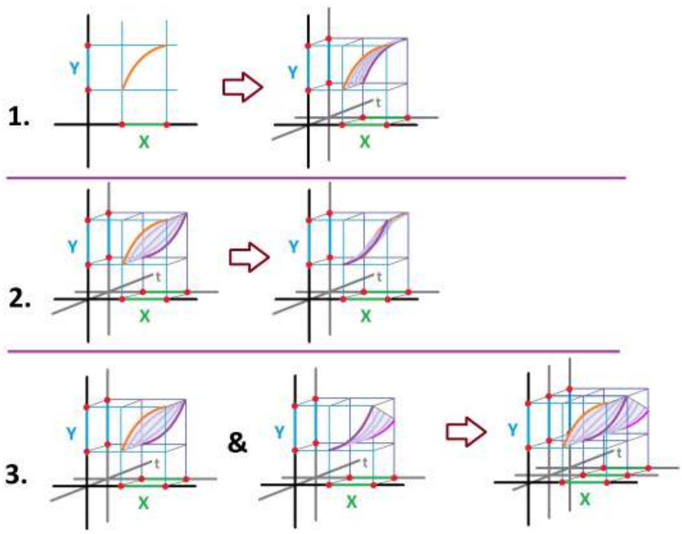
$$H(t, x) := \begin{cases} F(2t, x) & \text{if } 0 \le t \le 1/2 \\ G(2t - 1, x) & \text{if } 1/2 \le t \le 1. \end{cases}$$

■

Referring to the proof, these three new homotopies are called respectively:

- 1. a *constant homotopy*,
- 2. the *inverse of a homotopy*, and
- 3. the *concatenation of two homotopies*.

They are illustrated below:

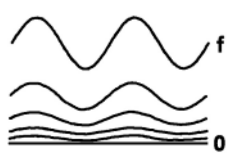


Exercise 1.4. Provide the missing details of the proof.

Notation: We will denote this quotient set of homotopy classes as follows:

$$[X, Y] := C(X, Y)/_{\simeq}.$$

Example 1.5. Sometimes things are simple. Any map $f : X \to \mathbf{R}$ can be "transformed" into any other. In fact, a simpler idea is to push the graph of a given function f to the x-axis:



We simply put:

$$f_t(x) := tf(x).$$

Then, we use the fact that an equivalence relation yields a partition into equivalence classes, to conclude that

$$[\mathbf{R}, \mathbf{R}] = \{[0]\}.$$

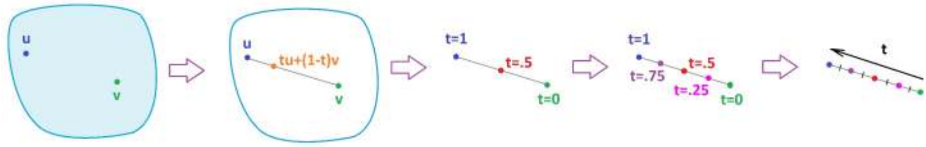
In other words, all maps are homotopic.

We can still have an explicit formula for a homotopy between two functions f, g:

$$F(t, x) = (1 - t)f(x) + tg(x). \qquad \square$$

Exercise 1.6. Prove the continuity of F.

The same exact formula describes how one "gradually" slides $f(x)$ toward $g(x)$ when $f, g : X \to Y$ are maps, X is any topological space, and Y is a *convex* subset of $\mathbf{R}^n$. The last condition guarantees that all convex combinations make sense:



This is called the *straight-line homotopy*.

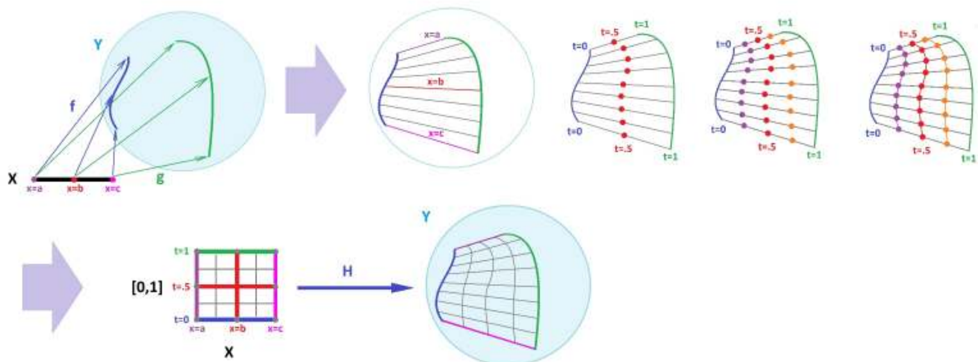
A more general setting for this construction is the following.

Definition 1.7. A vector space V over $\mathbf{R}$ is called a *topological vector space* if it is equipped with a topology with respect to which its vector operations are continuous:
- addition: $V \times V \to V$, and
- scalar multiplication: $\mathbf{R} \times V \to V$.

Exercise 1.8. Prove that these are topological vector spaces: $\mathbf{R}^n$, $C[a,b]$. Hint: it has something to do with the product topology.

Proposition 1.9. If Y is a convex subset of a topological vector space, then all maps to Y are homotopic: $\#[X,Y]=1$.

Exercise 1.10. Prove the proposition. Hint:



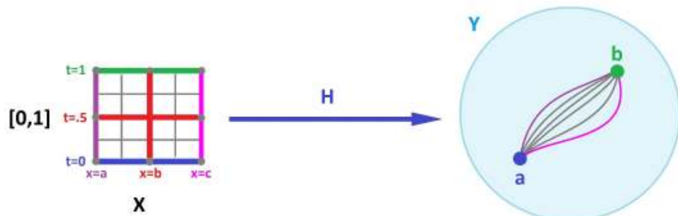
What if Y is the disjoint union of m convex sets in $\mathbf{R}^n$? Will we have:

$$\#[X,Y]=m?$$

Yes, but only if X is path-connected!

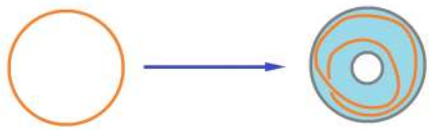
Exercise 1.11. Prove this statement and demonstrate that it fails without the path-connectedness requirement.

Exercise 1.12. Prove that if Y isn't path-connected then $\#[X,Y]>1$. Hint: consider constant maps:

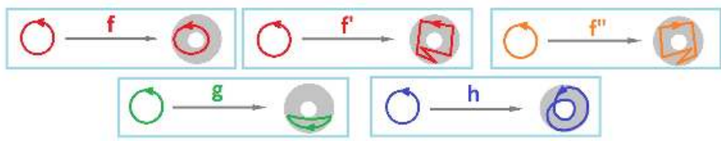


Exercise 1.13. What if X has k path-components?

Example 1.14. What if the target space isn't acyclic? What can we say about the homotopy classes of maps from the circle to the ring, or another circle?

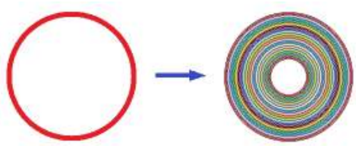


The circles can be moved and stretched as if they were rubber bands:



We will need to develop our theory before we can classify these maps according to their homotopy classes.

But sometimes the homotopy is easy to specify. For example, one can gradually stretch this circle:



To get from f_0 (the smaller circle) to f_0 (the large circle), one goes through intermediate steps – circles indexed by values between 0 and 1:

$$f_0, f_{1/4}, f_{1/2}, f_{3/4}, f_1.$$

Generally:

$$f_t(x) := (1+t)f_0(x).$$

It is clear that the right-hand side continuously depends on both t and x. □

Exercise 1.15. Suppose that two maps $f, g : X \to \mathbf{S}^1$ never take antipodal to each other points: $f(x) \neq -g(x), \ \forall x \in X$. Prove that f, g are homotopic.

To summarize, *homotopy is a continuous transformation of a continuous function.*

Homotopies help us tame the monster of a space-filling curve.

Theorem 1.16. Every map $f : [0,1] \to S$ of the interval to a surface, is homotopic to a map that isn't onto.

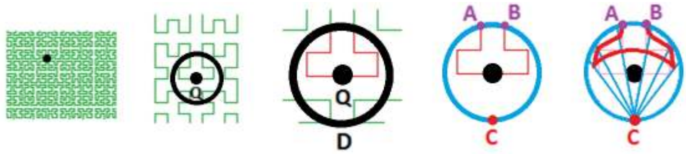
Proof. Suppose Q is the point on S that we want to avoid and $f(0) \neq Q, \ f(1) \neq Q$. Let D be a small disk neighborhood around Q. Then $f^{-1}(D) \subset (0,1)$ is open, and, therefore, is the disjoint union of open intervals. Pick one of them, (a,b). Then we have:

$$\begin{array}{ll} A := f(a) & \in \partial D, \\ B := f(b) & \in \partial D, \\ f\Big((a,b)\Big) & \subset D. \end{array}$$

Choose a point

$$C \in \partial D \setminus \{A, B\}.$$

Now, $f\Big([a,b]\Big)$ is compact and, therefore, closed. Then there is a neighborhood of C disjoint from the path $f\Big([a,b]\Big)$. We construct a homotopy that pushes this path away from C towards the opposite side of ∂D. If we push far enough, the point $Q \in D$ will no longer lie on the path.



This homotopy is "relative": all the changes to f are limited to the values $x \in (a,b)$. This allows us to construct such a homotopy independently for every interval (a,b) in the decomposition of $f^{-1}(D)$. Then, the fact that the condition $f(a)=A,\ f(b)=B$ is preserved under this homotopy allows us to stitch these "local" homotopies together one by one. The resulting homotopy F pushes every piece of the path close to Q away from Q.

This construction only works when there are finitely many such open intervals. ■

Exercise 1.17. Show that the proof, as written, fails when we attempt to stitch together infinitely many "local" homotopies to construct $F:[0,1]\times(0,1)\to S$. Hint: start with

- $f^{-1}(D)=\bigcup_n (a_n,b_n),\ f(a_n)=A_n,\ f(b_n)=B_n,$
- $A_n\to A,\ B_n\to B=A,\ F(1,t_n)=-A,\ t_n\in(a_n,b_n).$

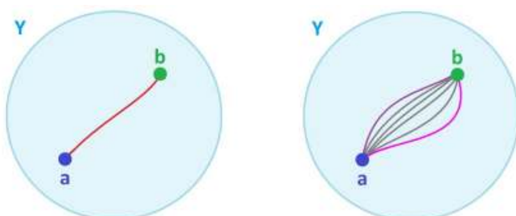
Exercise 1.18. Fix the proof. Hint: choose only the intervals on which the path actually passes through Q.

Exercise 1.19. Provide an explicit formula for one of those "local homotopies".

Exercise 1.20. Generalize the theorem as much as you can.

1.3 Types of connectedness

Let's take a look at the familiar illustrations. We have path-connectedness on the left and a homotopy between two constant maps on the right:



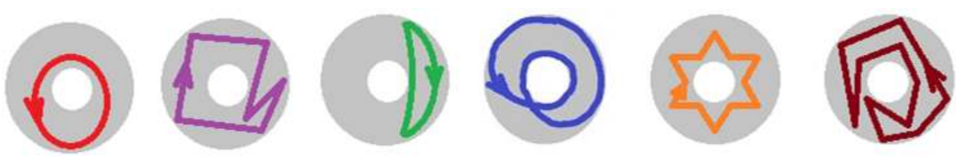
The picture reveals that the former is a special case of the latter.

Indeed, suppose in Y there is a path between two points a,b as a function $p:[0,1]\to Y$ with $p(0)=a, p(1)=b$. Then the function $H:[0,1]\times X$ given by $H(t,x)=p(t)$ is a homotopy between these two constant maps.

Exercise 1.21. Sketch the graph of a homotopy between two constant maps defined on $[a,b]$.

To summarize, in a path-connected space all constant maps are homotopic.

Let's carry out a similar analysis for simple-connectedness. This condition was defined informally as follows: "every closed path can be deformed to a point".



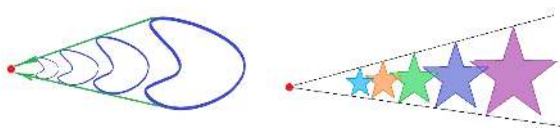
By now we know exactly what "deformed" means. Let's translate:

Informal:	$\longrightarrow$	Formal:
"A closed path in Y...	$\longrightarrow$	"A map $f : \mathbf{S}^1 \to Y$...
can be deformed to...	$\longrightarrow$	is homotopic to...
a point."	$\longrightarrow$	a constant function."

Exercise 1.22. Indicate which of the following spaces are simply connected:

- disk: $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 < 1\}$;
- circle with pinched center: $\{(x, y) \in \mathbf{R}^2 : 0 < x^2 + y^2 < 1\}$;
- ball with pinched center: $\{(x, y, z) \in \mathbf{R}^3 : 0 < x^2 + y^2 + z^2 < 1\}$;
- ring: $\{(x, y) \in \mathbf{R}^2 : 1/2 < x^2 + y^2 < 1\}$;
- thick sphere: $\{(x, y, z) \in \mathbf{R}^3 : 1/2 < x^2 + y^2 + z^2 < 1\}$;
- the doughnut (solid torus): $\mathbf{S}^1 \times \mathbf{D}^2$.

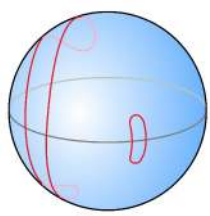
Recall that the plane $\mathbf{R}^2$ is simply connected because every loop can be deformed to a point via a straight line homotopy:



More general is the following result.

Theorem 1.23. Any convex subset of $\mathbf{R}^n$ is simply connected.

And so are all spaces homeomorphic to convex sets. There are many others.



Theorem 1.24. The n-sphere $\mathbf{S}^n$, $n \geq 2$, is simply connected.

Proof. *Idea for $n = 2$.* We assume that the loop p isn't space-filling, so that there is a point Q in its complement, $Q \in \mathbf{S}^2 \setminus \operatorname{Im} p$. But $\mathbf{S}^2 \setminus \{Q\}$ is homeomorphic to the plane. It also contains the loop p. Then p can be contracted to a point within this set. ∎

Exercise 1.25. Provide an explicit formula for the homotopy.

It is often more challenging to prove that a space is *not* simply connected. We will accept the following without proof (see Kinsey, *Topology of Surfaces*, p. 203).

Theorem 1.26. The circle $\mathbf{S}^1$ is not simply connected.

Corollary 1.27. The plane with a point taken out, $\mathbf{R}^2 \setminus \{(0, 0)\}$, is not simply connected.

Exercise 1.28. Prove the corollary.

Simple-connectedness helps us classify manifolds.

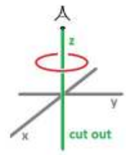
Theorem 1.29 (Poincaré Conjecture). If M is a simply connected compact path-connected 3-manifold without boundary, then M is homeomorphic to the 3-sphere.

The Poincaré homology sphere serves as an example that shows that simple-connectedness can't be replaced with $H_1 = 0$ for the theorem to hold.

Exercise 1.30. Prove that the sphere with a point taken out, $\mathbf{S}^2 \setminus \{N\}$, is simply connected.

Exercise 1.31. Prove that the sphere with two points taken out, $\mathbf{S}^2 \setminus \{N, S\}$, is not simply connected.

Exercise 1.32. Is the 3-space with a line, such as the z-axis, taken out simply connected? Hint: imagine looking down on the xy-plane:



Exercise 1.33. Is the 3-space with a point taken out, $\mathbf{R}^3 \setminus \{(0,0,0)\}$, simply connected?

This analysis demonstrates that we can study the topology of a space X *indirectly*, by studying the set of homotopy classes of maps $[Q, X]$ from a collection of wisely chosen topological spaces Q to X. Typically, we use the n-spheres $Q = \mathbf{S}^n,\ n = 0, 1, 2, \ldots$. These sets are called the *homotopy groups* and denoted by

$$\pi_n(X) := [\mathbf{S}^n, X].$$

Then, similarly to the homology groups $H_n(X)$, the sets $\pi_n(X)$ capture the topological features of X:

- $\pi_0(X)$ for cuts,
- $\pi_1(X)$ for tunnels,
- $\pi_2(X)$ for voids, etc.

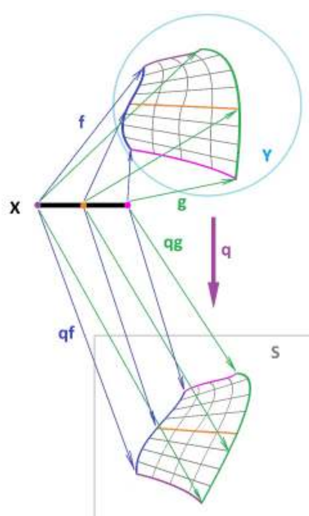
In particular, we showed above that $\pi_1(\mathbf{S}^n)$ is trivial for $n \geq 2$.

1.4 Homotopy theory

We will seek a structure for the set of homotopy classes of maps.

First, what happens to two homotopic maps when they are composed with other maps? Are the compositions also homotopic?

In particular, if $f \simeq d : X \to Y$ and $q : X \to S$ is a map, are $qf, qg : x \to S$ homotopic too? As the picture below suggests, the answer is Yes:



The picture illustrates the right upper part of the following diagram:

$$\begin{array}{ccc}
X & \xrightarrow{f\simeq g} & Y \\
\uparrow p & \searrow & \downarrow q \\
R & \xrightarrow{pfq\simeq pgq} & S
\end{array}$$

The two parts of the result below refers to the two triangles of this diagram.

Theorem 1.34. Homotopy, as an equivalence of maps, is preserved under compositions, as follows.

- (1) $F : f \simeq g \Longrightarrow H : qf \simeq qg$, where $H(t, x) := qF(t, x),\ x \in X, t \in [0, 1]$;
- (2) $F : f \simeq g \Longrightarrow H : fp \simeq gp$, where $H(t, z) := F(t, p(z)),\ z \in R, t \in [0, 1]$.

In either case, the new homotopy is the composition of the old homotopy and the new map.

Exercise 1.35. Finish the proof.

So, a map takes – via compositions – every pair of homotopic maps to a pair of maps that is also homotopic. Therefore, the map takes every homotopy class of maps to another homotopy class. Then, for any topological spaces X, Y, Q and any map

$$h : X \to Y,$$

the quotient map

$$[h] : [Q, X] \to [Q, Y]$$

given by

$$[h]([f]) := [hf]$$

is well-defined.

In particular, $h : X \to Y$ generates a function on the homotopy groups:

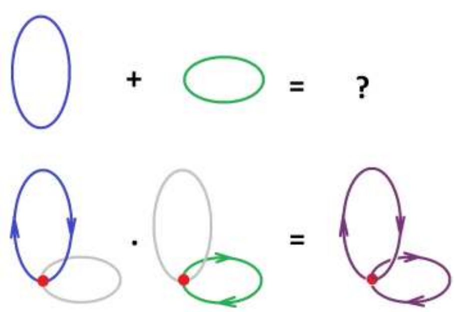
$$[h_n] : \pi_n(X) \to \pi_n(Y),\ n = 0, 1, 2, ...$$

The outcome is very similar to the way h generates the homology maps:

$$[h_n] : H_n(X) \to H_n(Y),\ n = 0, 1, 2, ...$$

as quotients of the maps of chains. However, in comparison, where is the algebra in these homotopy *groups*?

We need to define an algebraic operation on the homotopy classes. To begin with, given two homotopy classes of maps $f, g : \mathbf{S}^1 \to X$, i.e., two loops, what is the meaning of their "sum", or their "product? Under the homology theory approach, we'd deal with a formal sum of the loops. But such a "double" loop can't be a real loop, i.e., a map $f \cdot g : \mathbf{S}^1 \to X$.



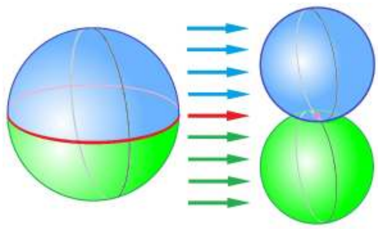
Unless, the two loops have a point in common! Indeed, if $f(c) = g(c)$ for some $c \in \mathbf{S}^1$, then one can "concatenate" f, g to create a new loop.

Exercise 1.36. Provide a formula for this construction.

Note that we will be speaking of products and use the multiplicative notation for this group-to-be because, in general, it's not abelian.

Exercise 1.37. Give an example of two loops with $a \cdot b \not\simeq b \cdot a$.

It's similar for $n = 2$. This map of the sphere to the kissing spheres is seen as the product two maps of spheres:



To make sure that this algebra works, we choose the following setting.

Definition 1.38. Every topological space X comes with a selected *base point* $x_0 \in X$. This space is denoted by (X, x_0) and is called a *pointed space*. Further, every map $f : (X, x_0) \to (Y, y_0)$ between two pointed spaces is assumed to take the base point to the base point: $f(x_0) = y_0$. These are called *pointed maps*.

Exercise 1.39. Prove that pointed spaces and pointed maps form a category.

This is, of course, just a special case of maps of pairs.

Now, the concatenation of two pointed maps $f, g : (\mathbf{S}^n, u) \to (X, x_0)$ always makes sense as a new pointed map $f \cdot g : (\mathbf{S}^n, u) \to (X, x_0)$.

What about their homotopy classes? They are supposed to be elements of the homotopy groups. As usual, the product of the quotients is defined as the quotient of the product: $[f] \cdot [g] := [f \cdot g]$ for $f, g \in \pi_n(X, x_0)$. We will accept without proof (see Bredon, *Topology and Geometry*, p. 443) that this operation is well-defined and satisfies the axioms of group. The classes are with respect to the *pointed homotopies*, i.e., ones that remain fixed for the base points, as follows. Given two pointed maps $f_0, f_1 : (X, x_0) \to (Y, y_0)$, a pointed homotopy is a homotopy $F : [0,1] \times X \to Y$ of f_0, f_1 which, in addition, is constant at x_0: for all $t \in [0,1]$ we have

$$F(t, x_0) = f_0(x_0) = f_1(x_0).$$

Exercise 1.40. Prove that any pointed loop on a surface is homotopic to a loop that isn't onto. Derive that $\pi_1(\mathbf{S}^2) = 0$.

That's a *group* structure for the set of homotopy classes. Is there a *topology* structure too?

Recall that we have two interpretations of a homotopy F between $f, g : X \to Y$. First, it's a continuous function $\mathbf{I} \times X \to Y$:

$$F \in C(\mathbf{I} \times X, Y),$$

where $\mathbf{I} = [0,1]$ is the interval. Second, it's a continuously parametrized (by $\mathbf{I}$) collection of continuous functions $X \to Y$:

$$F \in C(\mathbf{I}, C(X, Y)).$$

We conclude that

$$C(\mathbf{I} \times X, Y) = C(\mathbf{I}, C(X, Y)).$$

This is a particular case of the continuous version of the *exponential identity of functions*, as follows.

Proposition 1.41. For polyhedra A, B, C, we have

$$C(A \times B, C) = C(A, C(B, C)).$$

Exercise 1.42. Prove the proposition.

We can take this construction one step further if we equip every set of continuous functions $C(X, Y)$ with a topology. Then we can argue that this isn't just two sets with a bijection between them but a homeomorphism:

$$C(A \times B, C) \approx C(A, C(B, C)).$$

Exercise 1.43. Suggest an appropriate choice of topology for this set. Hint: start with $A = B = C = \mathbf{I}$.

From this point of view, *a homotopy is a path between two maps in the function space.*

Exercise 1.44. What path-connected function spaces do you know?

1.5 Homotopy equivalence

Let's review what we've come up with in our quest for a relation among topological spaces that is "looser" than topological equivalence but still respectful of the topological features that we have been studying.

Spaces X and Y are called *homotopy equivalent*, or are of the same *homotopy type*, $X \simeq Y$, if there are maps

$$f : X \to Y \text{ and } g : Y \to X$$

such that

$$fg \simeq \text{Id}_Y \text{ and } gf \simeq \text{Id}_X .$$

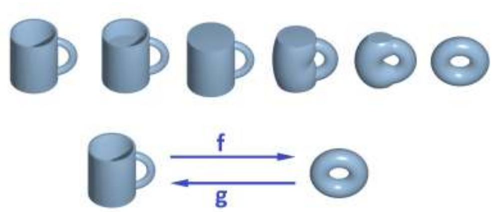
In that case, the **notation** is:

$$f : X \simeq Y,\ g : Y \simeq X,$$

or simply:

$$X \simeq Y.$$

Homotopy equivalence of two spaces is commonly described as a "deformation" and illustrated with a sequence of images that transforms one into the other:



One should be warned that in this very common example, the fact that the two spaces are also homeomorphic is irrelevant and may lead to confusion. If we think of these two transformations as if preformed over a period of time, this is the difference:

- topological equivalence: continuous over space, *incremental* and reversible over time;
- homotopy equivalence: continuous over space, *continuous* and reversible over time.

Some of the examples of homotopies above also suggest examples of homotopy equivalent spaces. For instance, with all maps homotopic, the n-ball is homotopy equivalent to the point:

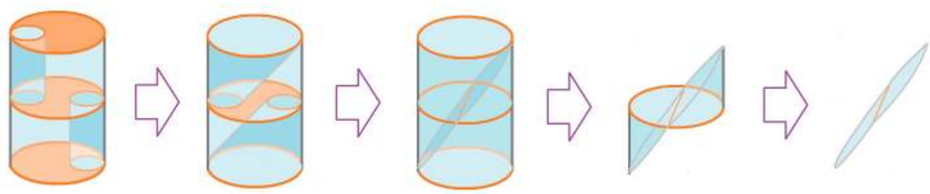
$$\mathbf{B}^n \simeq \{0\}.$$

Indeed, we can just contract the ball to its center. Note here that, if done incrementally, this contraction becomes a collapse and can't be reversed.

Exercise 1.45. Prove that any space *homeomorphic* to a convex subset of a Euclidean space is homotopy equivalent to a point.

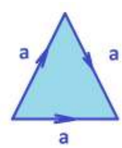
Definition 1.46. A topological space X is called *contractible* if X is homotopy equivalent to a point; i.e., $X \simeq \{x_0\}$.

Example 1.47. A less obvious example of a contractible space is the "two-room house". Here we have two rooms each taking the whole floor. There is access to the first floor through a tube from the top and to the second floor from the bottom:

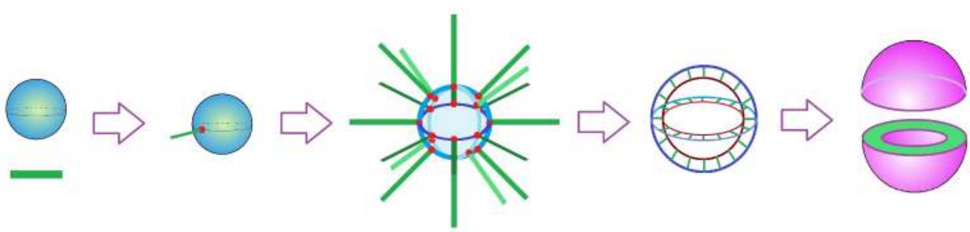


These are the steps of the deformation:
- we expand the entries into the building at the top and bottom to the size of the whole circle, which turns the two tubes into funnels;
- then we expand the entries into the rooms from the funnels to the half of the whole circle creating a sloped ellipse;
- finally, we contract the walls and we are left with only the ellipse. □

Exercise 1.48. Prove that the dunce hat is contractible:



Further, our intuition suggests that "thickening" of a topological space doesn't introduce new topological features. We can construct this new space by taking an ε-neighborhood of a realization of complex K in some Euclidean space or, more precisely, by taking the product of K with the interval:



We have even stronger statement of homotopy equivalence.

Theorem 1.49. For any topological space Y, we have

$$Y \times \mathbf{I} \simeq Y.$$

Proof. The plan of the proof comes from our analysis of the case of circle vs. cylinder. Let

$$X := \mathbf{I} \times Y.$$

Consider the two maps that naturally appear. The first is the projection

$$p : X \to Y,\ p(t, x) = x,$$

of X onto Y as its "bottom" and the second is the embedding

$$e : Y \to X,\ e(x) = (0, x)$$

of Y into X as its "bottom".

Let's consider their compositions:

- $pe : Y \to Y = \mathrm{Id}_Y$;
- $ep : X \to X$ is the collapse:

$$ep(t, x) = (0, x).$$

For the latter, we define the homotopy

$$H : [0,1] \times [0,1] \times Y \to [0,1] \times Y$$

by

$$H(t, s, x) := (ts, x).$$

It is easy to confirm that:

$$H(0, \cdot) = ep,\ H(1, \cdot) = \mathrm{Id}_X .$$

Next, we prove that this is a continuous map with respect to the product topology. Suppose U is an element of the standard basis of the product topology of $[0,1] \times Y$; i.e., $U = (a, b) \times V$, where $a < b$ and V is open in Y. Then

$$\begin{aligned} H^{-1}(U) &= H^{-1}((a,b) \times V) \\ &= \{(t,s) \in [0,1] \times [0,1] : a < ts < b\} \times V \end{aligned}$$

The first component of this set is a region between two hyperbolas $s = a/t, s = b/t$. It is open and so is $H^{-1}(U)$. ■

The main properties are below. The first two are obvious:

Theorem 1.50.

- Homotopy invariance is a topological property.
- A homeomorphism is a homotopy equivalence.

Later we will also prove the stronger version of the invariance of homology:

- Homology groups are preserved under homotopy equivalence.

Theorem 1.51. Homotopy equivalence is an equivalence relation for topological spaces.

Proof. 1. Reflexivity: $f, g : X \simeq X$ with $f = g := \mathrm{Id}_X$.

2. Symmetry: if $f, g : X \simeq Y$ then $g, f : Y \simeq X$.

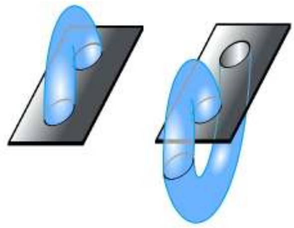
3. Transitivity: if $f, g : X \simeq Y$ and $p, q : Y \simeq Z$ then $pf, gq : X \simeq Z$. Indeed, we have:

$$\begin{aligned} (pf)(gq) = p(fg)q &\simeq p \,\mathrm{Id}_X\, q &= pq \simeq \mathrm{Id}_Z, \\ (gq)(pf) = g(qp)f &\simeq g \,\mathrm{Id}_Z\, f &= gf \simeq \mathrm{Id}_X, \end{aligned}$$

by the theorem from the last subsection. ■

Once again, homotopy helps us classify manifolds.

Theorem 1.52 (Generalized Poincaré Conjecture). Every compact n-manifold without boundary which is homotopy equivalent to the n-sphere is homeomorphic to the n-sphere.



Exercise 1.53. (a) Prove that the above spaces are homotopy equivalent but not homeomorphic. (b) Consider other possible arrangements of the square and a handle and find out whether they are homeomorphic or homotopy equivalent. (c) What about the Möbius band with a handle? (d) What about the sphere with a handle? Hint: don't try to keep the sphere in the 3-space.

Exercise 1.54. Prove that $\mathbf{S}^1 \times \mathbf{S}^1$ and $\mathbf{S}^3 \vee \mathbf{S}^1$ are not homotopy equivalent.

Exercise 1.55. Show that $\mathbf{S}^2 \vee \mathbf{S}^1 \simeq \mathbf{S}^2 \cup A$, where A is the line segment from the north to south pole.

Exercise 1.56. The following spaces are homotopy equivalent to some familiar ones: (a) $\mathbf{S}^2 \cup A$, where A is disk bounded by the equator, (b) $\mathbf{T}^2 \cup D_1 \cup D_2$, where D_1 is a disk bounded by the inner equator and D_2 is a disk bounded by a meridian. Find those spaces.

Exercise 1.57. With the amoeba-like abilities, this person can unlink his fingers without unlocking them. What will happen to the shirt?

Reminder: Homotopy is a relation among *maps* while homotopy equivalence is a relation among *spaces*.

1.6 Homotopy in calculus

The importance of these concepts can be seen in calculus, as follows:

Recall a *vector field* in the plane is a function $V : \mathbf{R}^2 \supset D \to \mathbf{R}^2$ for some open set D. It is called *conservative* if it is the gradient of a scalar function:

$$V = \operatorname{grad} f.$$

Such a vector field may represent the vertical velocity of a flow on a surface ($z = f(x, y)$) under gravity or a force of a physical systems in which energy is conserved.

We know the following theorem from calculus.

Theorem 1.58. Suppose we have a vector field $V = (P, Q)$ defined on an open set $D \subset \mathbf{R}^2$. Suppose V is *irrotational*; i.e., $P_y = Q_x$. Then V is conservative provided D is simply connected.

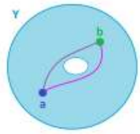
It is then easy to prove then that the line integral along a closed path is 0.

But what if the region isn't simply connected? The theorem doesn't apply anymore but, with the tools presented in this section, we can salvage it and even generalize.

Theorem 1.59. Suppose we have an irrotational vector field $V = (P, Q)$ defined on an open set $D \subset \mathbf{R}^2$. Then the line integral along any closed path (i.e., given by a map $p : \mathbf{S}^1 \to D$) homotopic to a point (i.e., given by a constant map) is 0.

The further conclusion is that line integrals are *path-independent*; i.e., the choice of integration path from a point to another does not affect the integral.

But does path-independence still apply if we can get from point a to point b in two "topologically different" ways?



Exercise 1.60. We can see several homotopy classes of paths from a to b with fixed endpoints. Derive from the theorem the fact that the line integrals are constant within each of these homotopy classes.

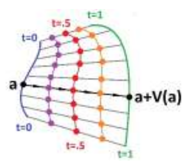
Recall next the Euler map of a continuous vector field in $\mathbf{R}^n$ (ODE)

$$Q_t : \mathbf{R}^n \to \mathbf{R}^n$$

is given by

$$Q_t(a) = a + tV(a).$$

Naturally, any restriction of $Q_t\Big|_D$, $D \subset \mathbf{R}^n$, is also continuous. Suppose D is given and we use Q_t for this restriction. Suppose also that D is "invariant"; i.e., $Q_t(D) \subset D$. Then $Q_t : D \to D$ is well-defined. Now we observe that it is continuous with respect to *either* of the variables. We realize that this is a homotopy.



But not just any; we have:

$$Q_0(a) = a, \ \forall a \in R.$$

Then we have the following.

Theorem 1.61. Suppose open set D is invariant. Then the Euler map $Q_t : D \to D, \ t > 0$, generated by a continuous vector field is homotopic to the identity,

$$Q_t \simeq \mathrm{Id}_D .$$

Therefore, as we shall see later, the possible behaviors of dynamical systems defined on a space is restricted by its topology.

Exercise 1.62. Prove the corollary.

Theorem 1.63 (Fundamental Theorem of Algebra). Every non-constant (complex) polynomial has a root.

Proof. Choose the polynomial p of degree n to have the leading coefficient 1. Suppose $p(z) \neq 0$ for all $z \in \mathbf{C}$. Define a map, for each $t > 0$,

$$p_t : \mathbf{S}^1 \to \mathbf{S}^1, \quad p_t(z) := \frac{p(tz)}{|p(tz)|}.$$

Then $p_t \simeq p_0$ for all $t > 0$ with p_0 a constant map. This contradicts the fact that $p_t \simeq z^n$ for large enough t. ■

Exercise 1.64. Provide the details of the proof.

1.7 Is there a constant homotopy between constant maps?

We know that path-connectedness implies that every two constant maps are homotopic and homotopic via a "constant" homotopy, i.e., one with every intermediate function f_t constant. Suppose $f_0, f_1 : X \to Y$ are two constant maps. If they are homotopic, does this mean that there is a constant homotopy between them?

You can fit the short version of the solution, if you know it, in this little box:

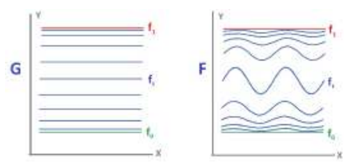
What if, however, we don't see the solution yet? Let's try to imagine how we would *discover* the solution following the idea: "Teach a man to fish..."

Let's first rewrite the problem in a more concrete way:
- given $f_0, f_1 : X \to Y$, that satisfy
- $f_0(x) = c \in Y,\ f_1(x) = d \in Y$ for all $x \in X$;
- also given a homotopy $F : f_0 \simeq f_1$, where $F : [0,1] \times X \to Y$, with $F(t, \cdot) = f_t$.

Is there a homotopy $G : f_0 \simeq f_1$, where $G : [0,1] \times X \to Y$, with $G(0, \cdot) = f_0, G(1, \cdot) = f_1$ and $G(t, x) = g(t)$ for all t, x?

We will attempt to simplify or even "trivialize" the setup until the solution becomes obvious, but not too obvious.

The simplest setup appears to be a map from an interval to an interval. Below, we show a constant, G, and a non-constant, F, homotopies between two constant maps:



Clearly, this is too narrow as *any* two maps are homotopic under these circumstances (Y is convex). The *target space is too simple*!

In order to avoid this obviousness, we
- try to imagine that Y is slightly bent, and then
- try to build G *directly* from F.

But, really, how do we straighten out these waves?

The fact that this question seems too challenging indicates that the *domain space is too complex*!

But what is the simplest domain? A single point!

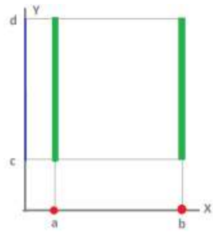
In this case, *all* homotopies are constant. The *domain space is too simple*!

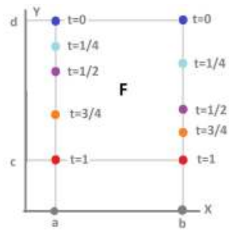
What is the next simplest domain? Two points!

Let's make the setup more concrete:
- 1. Given $f_0, f_1 : \{a, b\} \to [c, d]$ (bent),
- 2. $f_0(a) = f_0(b) = c,\ f_1(a) = f_1(b) = d$;

- 3. also given a homotopy $F : f_0 \simeq f_1$, where $F : [0,1] \times \{a,b\} \to [c,d]$, with $F(t,\cdot) = f_t$.
- 4. Find a homotopy $G : f_0 \simeq f_1$, where $G : [0,1] \times \{a,b\} \to [c,d]$, with $G(0,\cdot) = f_0$, $G(1,\cdot) = f_1$ and $G(t,a) = G(t,b)$ for all t.

The setup is as simple as possible. As the picture on the left shows, *everything* is happening within these two green segments.





Meanwhile, the picture on the right is a possible homotopy F and we can see that it might be non-trivial: within either segment, d is "moved" to c but at a different pace.

The setup is now very simple but the solution isn't obvious yet. That's what we want for our analysis!

Items 1 and 2 tell us that there are just *two values*, c and d, here. Item 3 indicates that there are just *two functions*, $F(\cdot,a)$ and $F(\cdot,b)$ here, and item 4 asks for just *one function*, $G(\cdot,a)$, same as $G(\cdot,b)$.

How do we get one function from two?

It is tempting to try to combine them algebraically (e.g., via a convex combination) but remember, $[c,d]$ is bent and there is no algebra there.

So, we need to construct $G(\cdot,a) = G(\cdot,b)$ from $F(\cdot,a)$ and $F(\cdot,b)$. What to do?..

Exercise 1.65. Fill the box:

1.8 Homotopy equivalence via cell collapses

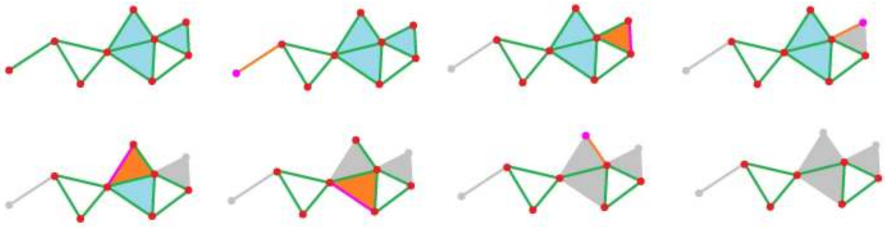
The idea of homotopy equivalence allows one to try to *simplify* a topological space before approaching computation of its homology. This idea seems good on paper but, practically, how does one simplify a given topological space?

A stronger version of the Nerve Theorem, which we accept without proof (see Alexandroff and Hopf, *Topology*) gives us a starting point.

Theorem 1.66 (Nerve Theorem). Let K be a (finite) simplicial complex and S an open cover of its realization $|K|$. Suppose the finite intersections of the elements of S are contractible. Then the realization of the nerve of S is homotopy equivalent to $|K|$.

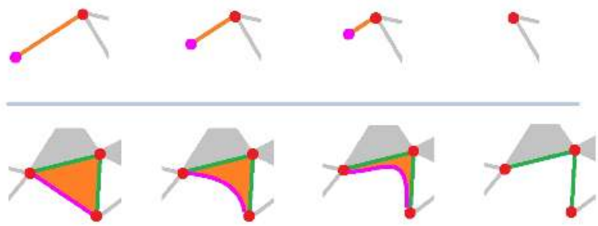
The question becomes, how does one simplify a given simplicial complex K while staying within its homotopy equivalence class?

The answer is, step-by-step. We shrink cells one at a time:



As we see, we have reduced the original complex to one with just seven edges. Its homology is obvious.

Now, more specifically, we remove cell σ and one of its boundary cells a at every step, by gradually pulling σ toward the closure of $\partial\sigma \setminus a$. Two examples, $\dim\sigma = 1$ and $\dim\sigma = 2$, are given below:



This step, defined below, is called an *elementary collapse* of K.

Of course, some cells can't be collapsed depending on their place in the complex:

The collapsible n-cells are marked in orange, $n = 1, 2$. As we see, for an n-cell σ to be collapsible, one of its boundary $(n-1)$-cells, a, has to be a "free" cell in the sense that it is not a part of the boundary of any other n-cell. The free cells are marked in purple. The rest of the n-cells aren't collapsible because each of them is either fully surrounded by other n-cells or is in the boundary of an $(n+1)$-cell.

Let's make this construction more precise.

The shrinking is understood as a homotopy equivalence – of the cell and, if done correctly, of the whole complex. We deform the chosen cell to a part of its boundary and do that in such a way that the rest of the complex remains intact!

The goal is accomplished via a more general type of homotopy than the pointed homotopy.

Definition 1.67. Given two maps $f_0, f_1 : X \to Y$ and a subset $A \subset X$, the maps are *homotopic relative to* A if there is a homotopy $F : [0,1] \times X \to Y$ of f_0, f_1 which, in addition, is constant on A: for all $t \in [0,1]$ we have

$$F(t,a) = f_0(a) = f_1(a)$$

for all $a \in A$. We use the following **notation**:

$$f_0 \simeq f_1 \quad \text{rel } A.$$

Theorem 1.68. Suppose σ is a geometric n-simplex and a is one of its $(n-1)$-faces. Then σ is homotopy equivalent to the union of the rest of its faces relative to this union:

$$\sigma \simeq \partial\sigma \setminus \dot{a} \quad \text{rel } \partial\sigma \setminus \dot{a}.$$

Exercise 1.69. Provide a formula for this homotopy equivalence. Hint: use barycentric coordinates.

Definition 1.70. Given a simplicial complex K, suppose σ is one of its n-simplices. Then σ is called a *maximal simplex* if it is not a boundary simplex of any $(n+1)$-simplex. A boundary $(n-1)$-simplex a of σ is called a *free face* of σ if a is not a boundary simplex of any other n-simplex.

In other words, a is a maximal simplex in $K \setminus \{\sigma\}$.

Proposition 1.71. If σ is a maximal simplex in complex K and a is its free face, then $K^1 = K \setminus \{\sigma, a\}$ is also a simplicial complex.

Exercise 1.72. Prove the proposition.

This removal of two cells from a simplicial complex is called an *elementary collapse* and the step is recorded with the following **notation**:

$$K \searrow K^1 := K \setminus \{\sigma, a\}.$$

Corollary 1.73. The homotopy class of a realization of a simplicial complex is preserved under elementary collapses.

Definition 1.74. A simplicial complex K is called *collapsible* if there is a sequence of elementary collapses of K that ends with a single vertex V; i.e.,

$$K \searrow K^1 \searrow K^2 \ ... \ \searrow K^N = \{V\}.$$

Corollary 1.75. The realization of a collapsible complex is contractible.

Exercise 1.76. Show that the complexes of the dunce hat and the two-room house are examples of contractible but not collapsible complexes.

As it is presented in the language of *cells*, the above analysis equally applies to both *simplicial* and *cubical complexes*.

Exercise 1.77. Prove the theorem above for cubical complexes.

Exercise 1.78. Define the "non-elementary" cell collapse that occurs when, for example, a triangle has *two* free edges and the total of *four* cells are removed. Repeat the above analysis. Hint: it is the vertex that is free.

Exercise 1.79. Define *elementary expansions* as the "inverses" of elementary collapses and repeat the above analysis:

$$\{V\} = K^0 \nearrow K^1 \nearrow K^2 \ ... \ \nearrow K^N.$$

Exercise 1.80. Prove that the torus with a point taken out is homotopy equivalent to the figure eight: $\mathbf{T}^2 \setminus \{u\} \simeq \mathbf{S}^1 \vee \mathbf{S}^1$.

1.9 Invariance of homology under cell collapses

The real benefit of this construction comes from our realization that the homology can't change. In fact, later we will show that the homology groups are always preserved under homotopy equivalence. With cell collapses, the conclusion is intuitively plausible as it seems impossible that new topological features could appear or the existing features could disappear.

Let's thoroughly analyze this construction. Suppose we have a simplicial complex K and a collapse:

$$K \searrow K'.$$

We suppose that from K an n-cell σ and its free face a are removed.

We need to find an isomorphism on the homology groups of these two complexes. We don't want to build it from scratch but instead find a map that induces it. Naturally, we choose the *inclusion*

$$i : K' \hookrightarrow K.$$

As a simplicial map, it induces a chain map:

$$i_\Delta : C(K') \to C(K),$$

which is also an inclusion. This chain map generates a homology map:

$$i_* : H(K') \to H(K).$$

We expect it to be an isomorphism.

Exercise 1.81. Define the projection $P : C(K) \to C(K')$. What homology map does it generate?

For brevity, we denote:

$$\partial := \partial^K,\ \partial' := \partial^{K'}.$$

This is what the chain complexes and the chain maps look like:

$$\begin{array}{ccccccc}
... & \xrightarrow{\partial_{n+1}} & C_n(K) & \xrightarrow{\partial_n} & C_{n-1}(K) & \xrightarrow{\partial_{n-1}} & ...\\
... & & \uparrow{i_n} & & \uparrow{i_{n-1}} & & ...\\
... & \xrightarrow{\partial'_{n+1}} & C_n(K') & \xrightarrow{\partial'_n} & C_{n-1}(K') & \xrightarrow{\partial'_{n-1}} & ...
\end{array}$$

Only a part of this chain map diagram is affected by the presence or absence of these two extra cells. The rest has identical rows and columns:

$$\begin{array}{ll}
C_k(K) = C_k(K'), & \forall k \neq n, n-1,\\
\partial_k = \partial'_k, & \forall k \neq n+1, n, n-1,\\
i_k = Id_{C_k(K')}, & \forall k \neq n, n-1.
\end{array}$$

Let's compute the specific values for all elements of the diagram.

These are the chain groups and the chain maps:

$$\begin{array}{ll}
C_n(K) & = C_n(K') \oplus < \sigma >,\\
C_{n-1}(K) & = C_{n-1}(K') \oplus < a >,\\
i_n & = Id_{C_n(K')} \oplus 0,\\
i_{n-1} & = Id_{C_{n-1}(K')} \oplus 0.
\end{array}$$

Now, we need to express ∂_k in terms of ∂'_k. Then, with the boundary operators given by their matrices, we can handle cycles and boundaries in a *purely algebraic* way...

The matrix of the boundary operator for dimension $n+1$ has the last row equal to $\partial^{-1}\sigma$. It has all zeros because σ is a maximal cell:

$$\partial_{n+1} = \partial'_{n+1} \oplus 0 = \left[\begin{array}{c} \partial^1_{n+1} \\ \hline 0, 0, \quad ... \quad , 0 \end{array} \right] : C_{n+1}(K') \to C_n(K') \oplus < \sigma > .$$

Then

$$\ker \partial_{n+1} = \ker \partial'_{n+1},\ \operatorname{Im} \partial_{n+1} = \operatorname{Im} \partial'_{n+1} \oplus 0.$$

The matrix of the boundary operator for dimension n has last row equal to $\partial^{-1}a$. It has only one 1 because a is a part of the boundary of only one cell. That cell is σ and the last column, which is non-zero, is $\partial\sigma$ with 1s corresponding to its faces:

$$\partial_n = \left[\begin{array}{c|c} & \pm 1 \\ & \pm 1 \\ \partial'_n & 0 \\ & \vdots \\ & 0 \\ \hline 0,\quad 0,\quad \dots\quad ,0 & \pm 1 \end{array}\right] : C_n(K') \oplus < \sigma > \to C_{n-1}(K') \oplus < a > .$$

Then

$$\ker \partial_n = \ker \partial'_n \oplus 0,\ \operatorname{Im} \partial_n = \operatorname{Im} \partial'_n \oplus \partial\sigma.$$

For the boundary operator for dimension $n-1$, the matrix's last column is ∂a, which is non-zero, with 1s corresponding to a's faces:

$$\partial_{n-1} = \left[\begin{array}{c|c} & \pm 1 \\ & \pm 1 \\ \partial'_{n-1} & 0 \\ & \vdots \\ & 0 \end{array}\right] : C_{n-1}(K') \oplus < a > \to C_{n-2}(K').$$

Then

$$\ker \partial_{n-1} = \ker \partial'_{n-1} \oplus 0,\ \operatorname{Im} \partial_{n-1} = \operatorname{Im} \partial'_{n-1}.$$

Finally, the moment of truth...

$$H_{n+1}(K) := \frac{\ker \partial_{n+1}}{\operatorname{Im} \partial_{n+2}} = \frac{\ker \partial'_{n+1}}{\operatorname{Im} \partial'_{n+2}} =: H_{n+1}(K');$$

$$H_n(K) := \frac{\ker \partial_n}{\operatorname{Im} \partial_{n+1}} = \frac{\ker \partial'_n \oplus 0}{\operatorname{Im} \partial'_{n+1} \oplus 0} \cong \frac{\ker \partial'_n}{\operatorname{Im} \partial'_{n+1}} =: H_n(K');$$

$$H_{n-1}(K) := \frac{\ker \partial_{n-1}}{\operatorname{Im} \partial_n} = \frac{\ker \partial'_{n-1} \oplus 0}{\operatorname{Im} \partial'_n} \cong \frac{\ker \partial'_{n-1}}{\operatorname{Im} \partial'_n} =: H_{n-1}(K').$$

The rest of the homology groups are unaffected by the two extra cells:

$$H_k(K) \cong H_k(K'),\ \forall k \neq n, n-1, n-2.$$

We have proven the following.

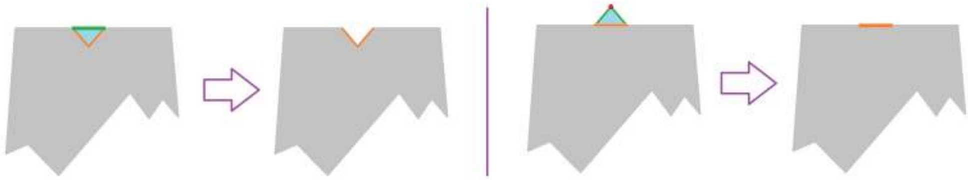
Theorem 1.82. If $K \searrow K'$ then

$$H(K) \cong H(K'),$$

under the homology map induced by the inclusion $i : K' \hookrightarrow K$.

Exercise 1.83. Provide details for the last part of the proof.

All these lengthy computations were needed to demonstrate something that may seem obvious, that these spaces have the same homology:



Exercise 1.84. Suppose K is a 2-dimensional simplicial complex. Suppose σ is a 2-cell of K that has two free edges, a, b, with vertex A between them. Let $K' := K \setminus \{\sigma, a, b, A\}$. Use the above approach to prove that $H(K) \cong H(K')$.

Exercise 1.85. Under what conditions is an elementary collapse a homeomorphism?

---

2 Cell maps

2.1 The definition

Here, we revisit the following issue:

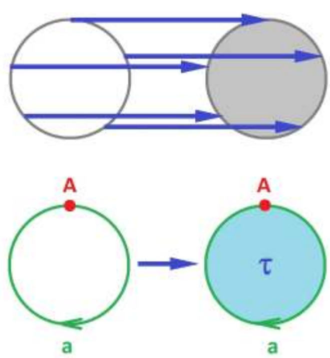
How does a continuous function change the topology of the space?

A narrower version of this question is:

How is the homology affected by maps?

Below, we will see how the theory of simplicial maps and their homology is extended to general cell complexes.

Example 2.1 (inclusion). As a quick example, consider the inclusion f of the circle into the disk as its boundary:



Here we have:

$$f : K = \mathbf{S}^1 \to L = \mathbf{B}^2.$$

After representing the spaces as cell complexes, we examine to what cell in L each cell in K is taken by f:

- $f(A) = A$,

- $f(a) = a$.

Further, we compute the chain maps

$$f_i : C_i(K) \to C_i(L),\ i = 0, 1,$$

as linear operators by defining them on the generators of these vector spaces:

- $f_0(A) = A$,
- $f_1(a) = a$.

Finally, we compute the homology maps as quotients of these linear operators. These maps reveal what f does to each homology class. Because the 0th homology groups of K and L are generated by the same cell A, this must be an isomorphism:

$$[f_0] : H_0(K) =< A > \to H_0(L) =< A > .$$

We also have:

$$[f_1] : H_1(K) =< a > \to H_1(L) = 0;$$

therefore, this operator is 0. □

The outline of the homology theory we've learned so far is this:

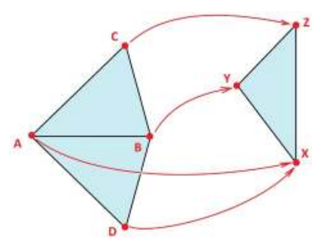
topological spaces $\longrightarrow$ cell complexes $\longrightarrow$ chain complexes $\longrightarrow$ homology groups

Now, if we add maps to this setup, the procedure will also include the following:

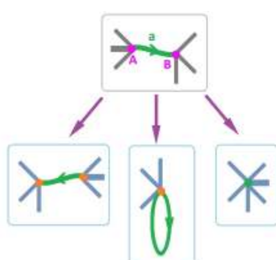
maps $\longrightarrow$ cell maps $\longrightarrow$ chain maps $\longrightarrow$ homology maps.

We would like to map cell complexes to each other in a manner that lends itself to the homological analysis outlined above.

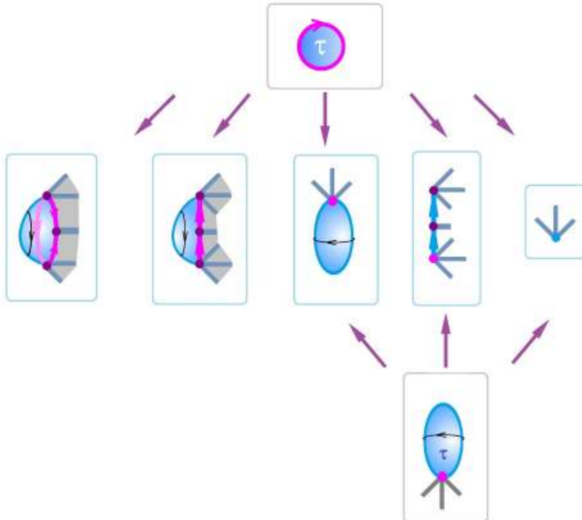
We take the lead from simplicial maps: every n-cell s is either cloned, $f(s) \approx s$, or collapsed, $f(s)$ is an k-cell with $k < n$.



The result is used as a template to define cell maps. The difference is that, first, if $f(s)$ is attached to itself along its boundary but s isn't, cloning doesn't produce a homeomorphism – unless restricted to the interior $\dot{s}$ of s. The construction is illustrated for dimension 1 below:



The illustration of this construction for dimension 2 also shows that, in contrast to the simplicial case, a collapsed cell may be stretched over several lower dimensional cells:



Definition 2.2. Given two cell complexes K and L, a continuous map

$$f : |K| \to |L|$$

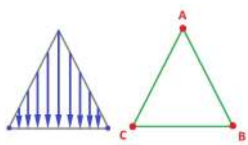
of their realizations is called a *cell map* (or cellular map) if for every k-cell s in K either

- 1. $f(s)$ is a k-cell in L and $f(\dot{s}) \approx \dot{s}$ under f; or
- 2. $f(s) \subset L^{(k-1)}$, where $L^{(k-1)}$ is the $(k-1)$-skeleton of L.

Exercise 2.3. List all possible ways complex $K = \{A, a, \alpha\}$ can be mapped to another cell complex.

Exercise 2.4. Represent the rotations of a circle through $\pi, \pi/2, \pi/3, \pi/4$ as cell maps.

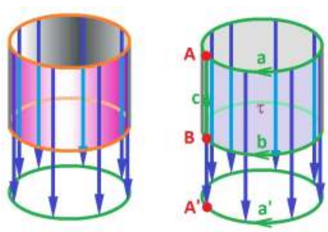
Example 2.5. This projection of a triangle on a line segment is not a cell map:



□

Exercise 2.6. Choose a different cell complex to represent Y above in such a way that the projection is then a cell map. How about the rotation of a circle through $\sqrt{2}\pi$?

Example 2.7 (projection). Let's consider the projection of the cylinder on the circle:



We have:

- $f(A) = f(B) = A'$, cloned;
- $f(a) = f(b) = a'$, cloned;
- $f(c) = A'$, collapsed;
- $f(\tau) = a'$, collapsed.

□

Note: the continuity requirement in the definition is crucial. Otherwise, one could, for example, cut the torus into a cylinder, twist it, and then glue it into the Klein bottle while preserving all cells intact.

2.2 Examples of cubical maps

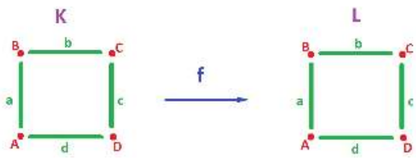
The next step after defining a cell map is to construct a function that records what happens to the cells.

Let's consider maps of the "cubical circle" to itself

$$f : X = \mathbf{S}^1 \to Y = \mathbf{S}^1.$$

We represent X and Y as two identical cubical complexes K and L and then find an appropriate representation $g : K \to L$ for each f in terms of their cells. More precisely, we are after

$$g = f_\Delta : C(K) \to C(L).$$



We will try to find several possible functions g under the following condition:

- (A) g maps each cell in K to a cell in L of the same dimension, otherwise it's 0.

The condition corresponds to the clone/collapse condition for a cell map f.

We make up a few examples.

Example 2.8 (identity).

$$\begin{array}{llll} g(A) = A, & g(B) = B, & g(C) = C, & g(D) = D, \\ g(a) = a, & g(b) = b, & g(c) = c, & g(d) = d. \end{array}$$

□

Example 2.9 (constant).

$$\begin{array}{llll} g(A) = A, & g(B) = A, & g(C) = A, & g(D) = A, \\ g(a) = 0, & g(b) = 0, & g(c) = 0, & g(d) = 0. \end{array}$$

All 1-cells collapse. Unlike the case of a general constant map, there are only 4 such maps for these cubical complexes. □

Example 2.10 (flip).

$$\begin{array}{llll} g(A) = D, & g(B) = C, & g(C) = B, & g(D) = A, \\ g(a) = c, & g(b) = b, & g(c) = a, & g(d) = d. \end{array}$$

This is a vertical flip; there are also the horizontal and diagonal flips, a total of 4. Only these four axes allow condition (A) to be satisfied. □

Example 2.11 (rotation).

$$\begin{array}{lll} \partial(f_1(AB)) & = \partial(0) & = 0; \\ f_0(\partial(AB)) & = f_0(A+B) = f_0(A) + f_0(B) = X + X & = 0. \end{array}$$

□

Exercise 2.12. Complete the example.

Next, let's try these values for our vertices:

$$g(A) = A, \ g(B) = C, \ ...$$

This is trouble. Even though we can find a cell for $g(a)$, it can't be AC because it's not in L. Therefore, $g(a)$ won't be aligned with its endpoints. As a result, g breaks apart. To prevent this from happening, we need to require that the endpoints of the image in L of any edge in K are the images of the endpoints of the edge.

Furthermore, we want to ensure the cells of *all* dimensions remain attached after g is applied and we require:

- (B) g takes boundary to boundary.

Algebraically, we arrive to the familiar *algebraic continuity* condition:

$$\partial g = g \partial.$$

Exercise 2.13. Verify this condition for the examples above.

Observe now that g is defined on complex K but its values aren't technically all in L. There are also 0s. They aren't cells, but rather *chains*. Recall that, even though g is defined on cells of K only, it can be extended to all chains, by linearity:

$$g(A + B) = g(A) + g(B), ...$$

Thus, condition (A) simply means that g maps k-chains to k-chains. More precisely, g is a collection of functions (a chain map):

$$g_k : C_k(K) \to C_k(L), \ k = 0, 1, 2, ...$$

For brevity we use the following notation:

$$g : C(K) \to C(L).$$

Example 2.14 (projection).

$$\begin{array}{llll} g(A) = A, & g(B) = A, & g(C) = D, & g(D) = A, \\ g(a) = 0, & g(b) = d, & g(c) = 0, & g(d) = d. \end{array}$$

Let's verify condition (B):

$$\begin{array}{l} \partial g(A) = \partial 0 = 0, \\ g\partial(A) = g(0) = 0. \end{array}$$

Same for the rest of 0-cells.

$$\begin{array}{l} \partial g(a) = \partial(0) = 0, \\ g\partial(a) = g(A + B) = g(A) + g(B) = A + A = 0. \end{array}$$

Same for c.

$$\begin{array}{l} \partial g(b) = \partial(d) = A + D, \\ g\partial(b) = g(B + C) = g(B) + g(C) = A + D. \end{array}$$

Same for d. □

Exercise 2.15. Try the "diagonal fold": A goes to C, while C, B and D stay.

In each of these examples, an idea of a map f of the circle/square was present first, then f was realized as a chain map g.

Notation: The chain map of f us denoted by

$$g = f_\Delta : C(K) \to C(L).$$

Let's make sure that this idea makes sense by reversing this construction. This time, we suppose instead that we already have a chain map

$$g : C(K) \to C(L),$$

what is a possible "realization" of g:

$$f = |g| : |K| \to |L|?$$

The idea is simple: if we know where each vertex goes under f, we can construct the rest of f using linearity, i.e., *interpolation.*

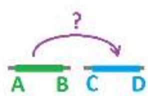
Example 2.16 (interpolation). A simple example first. Suppose

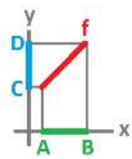
$$K = \{A, B, a : \ \partial a = B - A\}, \ L = \{C, D, b : \ \partial b = D - C\}$$

are two complexes representing the closed intervals. Define a chain map:

$$g(A) = C, \ g(B) = D, \ g(a) = b.$$

If the first two identities is all we know, we can still construct a continuous function $f : |K| \to |L|$ such that $f_\Delta = g$. The third identity will be taken care of by condition (B).





If we include $|K|$ and $|L|$ as subsets of the x-axis and the y-axis respectively, the solution becomes obvious:

$$f(x) := C + \tfrac{D-C}{B-A} \cdot (x - A).$$

This approach allows us to give a single formula for realizations of all chain operators:

$$f(x) := g(A) + \tfrac{g(B)-g(A)}{B-A} \cdot (x - A).$$

For example, suppose we have a constant map:

$$g(A) = C, \ g(B) = C, \ g(a) = 0.$$

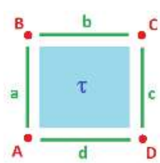
Then

$$f(x) = C + \tfrac{C-C}{B-A} \cdot (x - A) = C.$$

□

Of course, this repeats exactly the construction of the geometric realization of a simplicial map. There is no difference as long as we stay within dimension 1, as in all the examples above. For dimensions above 1, we can think by analogy.

Example 2.17. Let's consider a chain map g of the complex K representing the solid square.



Knowing the values of g on the 0-cells of K gives us the values of $f = |g|$ at those points. How do we extend it to the rest of $|K|$?

An arbitrary point u in $|K|$ is represented as a convex combination of A, B, C, D:

$$u = sA + tB + pC + qD, \text{ with } s + t + p + q = 1.$$

Then we define $f(u)$ to be

$$f(u) := sf(A) + tf(B) + pf(C) + qf(D).$$

Accordingly, $f(u)$ is a convex combination of $f(A), f(B), f(C), f(D)$. But all of these are vertices of $|L|$, hence $f(u) \in |L|$. □

Example 2.18 (projection). Let's consider the projection:

$$\begin{array}{llll} g(A) = A, & g(B) = A, & g(C) = D, & g(D) = A, \\ g(a) = 0, & g(b) = d, & g(c) = 0, & g(d) = d, \\ g(\tau) = 0. & & & \end{array}$$

Then,

$$\begin{array}{ll} f(u) & = sf(A) + tf(B) + pf(C) + qf(D) \\ & = sA + tA + pD + qD \\ & = (s + t)A + (p + q)D. \end{array}$$

Due to $(s + t) + (p + q) = 1$, we conclude that $f(u)$ belongs to the interval AD. □

Exercise 2.19. Are these maps well-defined? Hint: the construction for simplicial maps is based on barycentric *coordinates*.

2.3 Modules

Before we proceed to build the homology theory of maps, we review what it takes to have *arbitrary ring of coefficients* R.

Recall that

- with $R = \mathbf{Z}$, the chain groups and the homology groups are abelian groups,
- with $R = \mathbf{R}$ (or other fields), the chain groups and the homology groups are vector spaces,

and now

- with an arbitrary R, the chain groups and the homology groups are *modules*.

Informally,

modules are vector spaces over rings.

The following definitions and results can be found in the standard literature such as Hungerford, *Algebra* (Chapter IV).

Definition 2.20. Given a commutative ring R with the multiplicative identity 1_R, a (commutative) *R-module* M consists of an abelian group $(M, +)$ and a scalar product operation $R \times M \to M$ such that for all $r, s \in R$ and $x, y \in M$, we have:

- $r(x + y) = rx + ry$,
- $(r + s)x = rx + sx$,
- $(rs)x = r(sx)$,
- $1_R x = x$.

The scalar multiplication can be written on the left or right.

If R is a field, an R-module is a vector space.

The rest of the definitions are virtually identical to the ones for vector spaces.

A subgroup N of M is a *submodule* if it is closed under scalar multiplication: for any $n \in N$ and any $r \in R$, we have $rn \in N$.

A group homomorphism $f : M \to N$ is a (module) *homomorphism* (or a linear operator) if it preserves the scalar multiplication: for any $m, n \in M$ and any $r, s \in R$, we have $f(rm + sn) = rf(m) + sf(n)$.

A bijective module homomorphism is an (module) *isomorphism*, and the two modules are called *isomorphic*.

Exercise 2.21. Prove that this is a category.

The *kernel* of a module homomorphism $f : M \to N$ is the submodule of M consisting of all elements that are taken to zero by f. The isomorphism theorems of group theory are still valid.

A module M is called *finitely generated* if there exist finitely many elements $v_1, v_2, ..., v_n \in M$ such that every element of M is a linear combination of these elements (with coefficients in R).

A module M is called *free* if it has a basis. This condition is equivalent to: M is isomorphic to a direct sum of copies of the ring R. Every submodule L of such a module is a *summand*; i.e.,

$$M = L \oplus N,$$

for some other submodule N of M.

Of course, $\mathbf{Z}^n$ is free and finitely generated. This module is our primary interest because that's what a chain group over the integers has been every time. It behaves very similarly to $\mathbf{R}^n$ and the main differences lie in these two related areas.

First, the quotients may have *torsion*, such as in $\mathbf{Z}/2\mathbf{Z} \cong \mathbf{Z}_2$. We have seen this happen in our computations of the homology groups.

Second, some operators invertible over $\mathbf{R}$ may be singular over $\mathbf{Z}$. Take $f(x) = 2x$ as an example.

We will refer to finitely generated free modules as simply *modules*.

2.4 The topological step

The topological setup above is now translated into algebra. From a cell map, we construct maps on the chain groups of the two cell complexes.

Definition 2.22. Given a cell map $f : |K| \to |L|$, the kth *chain map generated by map* f,

$$f_k : C_k(K) \to C_k(L),$$

is defined on the generators as follows. For each k-cell s in K,

- 1. if s is cloned by f, define $f_k(s) := \pm f(s)$, with the sign determined by the orientation of the cell $f(s)$ in L induced by f;
- 2. if s is collapsed by f, define $f_k(s) := 0$.

Also,

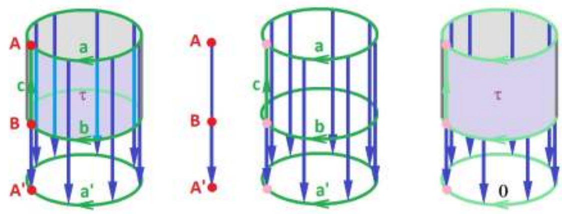
$$f_\Delta = \{f_i : i = 0, 1, ...\} : C(K) \to C(L)$$

is the (total) *chain map* generated by f.

In the items 1 and 2 of the definition, the left-hand sides are the *values* of s under f_k, while the right hand sides are the *images* of s under f.

Note: The notion of orientation was fully developed for both simplicial and cubical complexes. For the general case of cell complexes, we just point out that for dimensions 0 and 1 the notion applies without change, while in dimension 2 the orientation of a cell is simply the direction of a trip around its boundary. Also, we can avoid dealing with the issue of orientation by resigning, with a certain penalty, to the algebra over $R = \mathbf{Z}_2$.

Example 2.23 (projection). Let's consider the projection of the cylinder on the circle again:



From the images of the cells under f listed above, we conclude:

- $f_0(A) = f_0(B) = A'$;
- $f_1(a) = f_1(b) = a';\ f_1(c) = 0$;
- $f_2(\tau) = 0$.

Now we need to work out the operators:

- $f_0 :\ C_0(K) =< A, B >\to C_0(L) =< A' >$;
- $f_1 :\ C_1(K) =< a, b, c >\to C_1(L) =< a' >$;
- $f_2 :\ C_2(K) =< \tau >\to C_2(L) = 0$.

From the values of the operators on the basis elements, we conclude:

- $f_0 = [1, 1]$;
- $f_1 = [1, 1, 0]$;
- $f_2 = 0$. □

Recall also that a chain map as a combination of maps between chain complexes, in addition, has to take boundary to boundary, or, more precisely, has to commute with the boundary operator.

Theorem 2.24. If f is a cell map, then f_Δ is a chain map:

$$\partial_k f_k = f_{k-1}\partial_k,\ \forall k.$$

Proof. Done for simplicial maps. ■

Example 2.25. To continue the above example, the diagram has to commute:

$$\begin{array}{r|ccccccccc}
 & 0 & \xrightarrow{\partial_3} & C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & 0 \\
\hline
C(K): & 0 & \xrightarrow{0} & \mathbf{R} & \xrightarrow{\partial_2} & \mathbf{R}^3 & \xrightarrow{\partial_1} & \mathbf{R}^2 & \xrightarrow{0} & 0 \\
f_\Delta: & \downarrow{0} & & \downarrow{f_2} & & \downarrow{f_1} & & \downarrow{f_0} & & \downarrow{0} \\
C(L): & 0 & \xrightarrow{0} & 0 & \xrightarrow{\partial_2} & \mathbf{R} & \xrightarrow{\partial_1} & \mathbf{R} & \xrightarrow{0} & 0
\end{array}$$

Let's verify that the identity is satisfied in each square above. First we list the boundary operators for the two chain complexes:

$$\begin{array}{lll}
\partial^K_2(\tau) = a - b, & \partial^K_1(a) = \partial^K_1(b) = 0, \partial^K_1(c) = A - B, & \partial^K_0(A) = \partial^K_0(B) = 0; \\
\partial^L_2 = 0, & \partial^L_1(a') = 0, & \partial^L_0(A') = 0.
\end{array}$$

Now we go through the diagram from left to right.

$$\begin{array}{l|l|l|l}
\partial_2 f_2(\tau) = f_1\partial_2(\tau) & \partial_1 f_1(a) = f_0\partial_1(a) & \partial_1 f_1(b) = f_0\partial_1(b) & \partial_1 f_1(c) = f_0\partial_1(c) \\
\partial_2(0) = f_1(a - b) & \partial_1(a') = f_0(0) & \partial_1(a') = f_0(0) & \partial_1(0) = f_0(B - A) = A' - A' \\
0 = a' - a' = 0,\ \text{OK} & 0 = 0,\ \text{OK} & 0 = 0,\ \text{OK} & 0 = 0,\ \text{OK}
\end{array}$$

So, f_Δ is indeed a chain map. □

Exercise 2.26. Represent the cylinder as a complex with two 2-cells, find an appropriate cell complex for the circle and an appropriate cell map for the projection, compute the chain map, and confirm that the diagram commutes.

Note: a broader definition of a cell map requires only: $f(K^n) \subset L^n$. It can be proven that the two definitions are equivalent, up to homotopy.

2.5 The algebraic step

For a cell map, we have defined – by topological means – its chain maps as the homomorphisms between the chain groups of the complexes. At this point, we can ignore the origin of these new maps and proceed to homology in a purely algebraic manner. Fortunately, this part was fully developed for simplicial complexes and maps and, being algebraic, the development is *identical* for cell complexes. A quick review follows.

First, we suppose that we have two *chain complexes*, i.e., combinations of modules and homomorphisms between these modules, called the *boundary operators*:

$$\begin{array}{lll} M := \{M_i, \partial_i^M : & M_i \to M_{i-1} : & i = 0, 1, ...\}, \\ N := \{N_i, \partial_i^N : & N_i \to N_{i-1} : & i = 0, 1, ...\}, \end{array}$$

with

$$M_{-1} = N_{-1} = 0.$$

As chain complexes, they are to satisfy the "double boundary identity":

$$\begin{array}{ll} \partial_i^M \partial_{i+1}^M = 0, & i = 0, 1, 2, ...; \\ \partial_i^N \partial_{i+1}^N = 0, & i = 0, 1, 2, ... \end{array}$$

The compact form of this condition is, for both:

$$\partial\partial = 0.$$

Second, we suppose that we have a *chain map* as a combination of homomorphisms between the corresponding items of the two chain complexes:

$$f_\Delta = \{f_i : \ M_i \to N_i : \ i = 0, 1, ...\} : M \to N.$$

As a chain map, it is to satisfy the "algebraic continuity condition":

$$\partial_i^M f_i = f_{i-1} \partial_i^N, \ i = 0, 1, ...$$

The compact form of this condition is:

$$f\partial = \partial f.$$

In other words, the diagram commutes:

$$\begin{array}{ccc} M_{i+1} & \xrightarrow{\partial_{i+1}^M} & M_i \\ \downarrow f_{i+1} & \searrow & \downarrow f_i \\ N_{i+1} & \xrightarrow{\partial_{i+1}^N} & N_i \end{array}$$

This combination of modules and homomorphisms forms a diagram with the two chain complexes occupying the two rows and the chain map connecting them by the vertical arrows, item by item:

$$\begin{array}{ccccccccccc}
... & \to & M_{k+1} & \xrightarrow{\partial^M_{k+1}} & M_k & \xrightarrow{\partial^M_k} & M_{k-1} & \to & ... & \to & 0\\
... & & \downarrow{f_{k+1}} & & \downarrow{f_k} & & \downarrow{f_{k-1}} & & ... & & \\
... & \to & N_{k+1} & \xrightarrow{\partial^N_{k+1}} & N_k & \xrightarrow{\partial^N_k} & N_{k-1} & \to & ... & \to & 0
\end{array}$$

Each square commutes.

Next, we define the *homology groups of the chain complex* for either of them, as these quotient modules:

$$\begin{array}{ll}
H_i(M) & = \frac{\ker \partial^M_i}{\operatorname{Im} \partial^M_{i+1}}, \quad i = 0, 1, ...;\\
H_i(N) & = \frac{\ker \partial^N_i}{\operatorname{Im} \partial^N_{i+1}}, \quad i = 0, 1, ...
\end{array}$$

And we use the following compact **notation**:

$$\begin{array}{ll}
H(M) & = \{H_i(M) : \ i = 0, 1, ...\},\\
H(N) & = \{H_i(N) : \ i = 0, 1, ...\}.
\end{array}$$

Finally, the *homology map induced by the chain map*

$$f_\Delta = \{f_0, f_1, ...\} : M \to N,$$

is defined as the combination:

$$f_* = \{[f_0], [f_1], ...\} : H(M) \to H(N),$$

of the quotient maps of f_i:

$$[f_i] : H_i(M) \to H_i(N), \ i = 0, 1, ...$$

In other words, each one is a homomorphism defined by

$$[f_i]([x]) := [f_i(x)].$$

The homology map is well-defined as guaranteed by the two properties above.

Meanwhile, under this equivalence relation the quotients of the boundary operators are trivial:

$$[\partial^M_i] = 0, \ [\partial^N_i] = 0, \ i = 0, 1, ...$$

Exercise 2.27. Prove these identities.

Then, under the quotients, the above diagram of chain complexes and chain maps is distilled into this:

$$\begin{array}{cccccccccccc}
... & \to & H_{k+1}(M) & \xrightarrow{0} & H_k(M) & \xrightarrow{0} & H_{k-1}(M) & \to & ... & \to & 0 & \qquad H(M)\\
... & & \downarrow{[f_{k+1}]} & & \downarrow{[f_k]} & & \downarrow{[f_{k-1}]} & & ... & & & \qquad \downarrow{f_*}\\
... & \to & H_{k+1}(N) & \xrightarrow{0} & H_k(N) & \xrightarrow{0} & H_{k-1}(N) & \to & ... & \to & 0 & \qquad H(N)
\end{array}$$

Still technically commutative, the diagram is just a collection of columns. They are combined into our very compact notation:

$$f_* : H(M) \to H(N).$$

Example 2.28. From the last subsection, we have the following chain maps:

$$\begin{array}{llllll}
f_0: & M_0 =< A, B > & \to & N_0 =< A' >, & f_0(A) = f_0(B) = A'; \\
f_1: & M_1 =< a, b, c > & \to & N_1 =< a' >, & f_1(a) = f_1(b) = a', f_1(c) = 0; \\
f_2: & M_2 =< \tau > & \to & N_2 = 0, & f_2(\tau) = 0.
\end{array}$$

From the algebraic point of view, the nature of these generators is irrelevant... Now, the diagram is:

$$\begin{array}{r|ccccccccc}
 & 0 & \xrightarrow{\partial_3} & C_2 & \xrightarrow{\partial_2} & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\partial_0} & 0 \\
\hline
M: & 0 & \xrightarrow{0} & < \tau > & \xrightarrow{[1,-1,0]^T} & < a, b, c > & \xrightarrow{\left[\begin{smallmatrix} 0 & 0 & -1 \\ 0 & 0 & 1 \end{smallmatrix}\right]} & < A, B > & \xrightarrow{0} & 0 \\
f_\Delta: & \downarrow{\scriptstyle 0} & & \downarrow{\scriptstyle 0} & & \downarrow{\scriptstyle [1,1,0]} & & \downarrow{\scriptstyle [1,1]} & & \downarrow{\scriptstyle 0} \\
N: & 0 & \xrightarrow{0} & 0 & \xrightarrow{0} & < a' > & \xrightarrow{0} & < A' > & \xrightarrow{0} & 0
\end{array}$$

Compute the homology:

$$\begin{array}{r|ccc}
 & H_2 & H_1 & H_0 \\
\hline
M: & 0 & < [a] = [b] > & < [A] = [B] > \\
f_*: & \downarrow{\scriptstyle 0} & \downarrow{\scriptstyle [1]} & \downarrow{\scriptstyle [1]} \\
N: & 0 & < [a'] > & < [A'] >
\end{array}$$

□

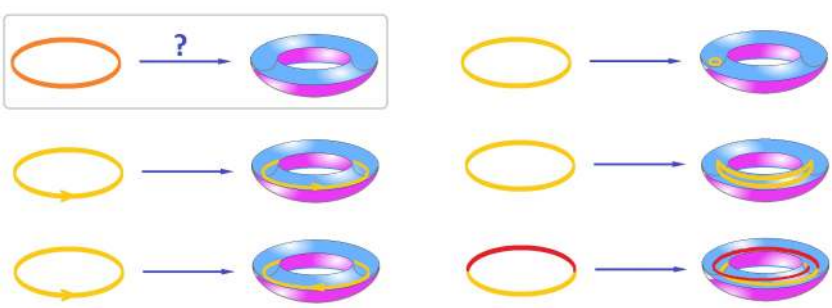
Exercise 2.29. For the chain complexes M, N above, suggest three examples of chain maps $g : N \to N$ and compute their homology maps. Make sure that all of them are different from the homology map above. Hint: make no references to cell maps.

2.6 Examples of homology maps

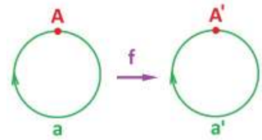
What happens to homology classes under continuous functions?

Let's consider a few examples with some typical maps that exhibit common behavior.

We start with maps of the circle to itself. Such a map can be thought of as a circular rope, X, being fitted into a circular groove, Y. You can carefully transport the rope to the groove without disturbing its shape to get the identity map, or you can compress it into a tight knot to get the constant map, etc.:



To build cell map representations for these maps, we use the following cell complexes of the circles:



Example 2.30 (constant). If $f : X \to Y$ is a constant map, all k-homology classes of X with $k > 0$ collapse or, algebraically speaking, are mapped to 0 by f_*. Meanwhile, all 0-homology classes of X are mapped to the same 0-homology class of Y. So, it's the identity for any two path-connected spaces. □

Example 2.31 (identity). If f is the identity map, we have

$$f_*(A) = A',$$
$$f_*(a) = a'.$$ □

Example 2.32 (flip). Suppose f is a flip (a reflection about the y-axis) of the circle. Then

$$f_*(a) = -a'.$$ □

Example 2.33 (turn). You can also turn (or rotate) the rope before placing it into the groove. The resulting map is very similar to the identity regardless of the degree of the turn. Indeed, we have:

$$f_*(a) = a'.$$

Even though the map is simple and its homology interpretation is clear, this isn't a cell map! □

Example 2.34 (wrap). If you wind the rope twice before placing it into the groove, you get:

$$f_*(a) = 2a'.$$

Once again, this isn't a cell map! □

We will need to subdivide the complex(es) to deal with this issue...

Example 2.35 (inclusion). In the all examples above we have

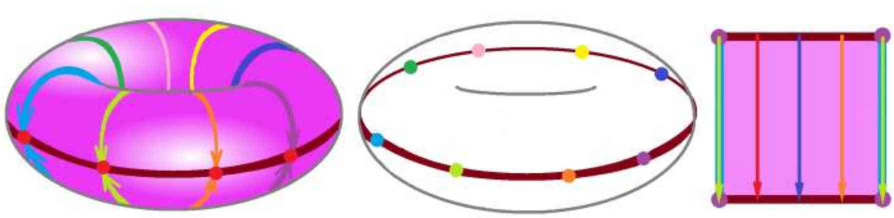
$$f_*(A) = A'.$$

Let's consider an example that illustrates what else can happen to 0-classes. Consider the inclusion of the two endpoints of a segment into the segment. Then $f : X = \{A, B\} \to Y = [A, B]$ is given by $f(A) = A$, $f(B) = B$. Now, even though A and B aren't homologous in X, their images under f are, in Y. So, $f(A) \sim f(B)$. In other words,

$$f_*([A]) = f_*([B]) = [A] = [B].$$

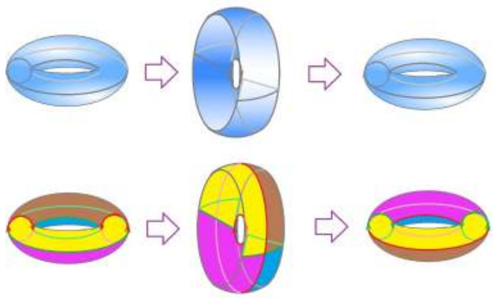
Algebraically, $[A] - [B]$ is mapped to 0. □

Example 2.36 (collapse). A more advanced example of collapse is that of the torus to a circle, $f : X = \mathbf{T}^2 \to Y = \mathbf{S}^1$.



We choose one longitude L (in red) and then move every point of the torus to its nearest point on L. □

Example 2.37 (inversion). And here's turning the torus inside out:



□

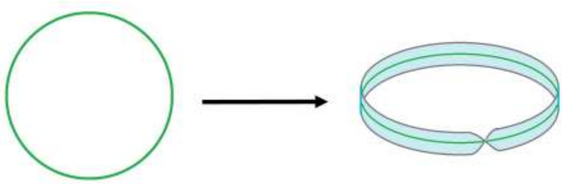
Exercise 2.38. Consider the following self-maps of the torus $\mathbf{T}^2$:
- (a) collapsing it to a meridian,
- (b) collapsing it to the equator (above),
- (c) collapsing a meridian to a point,
- (d) gluing the outer equator to the inner equator,
- (e) turning it inside out.

For each of those,
- describe the cell structure of the complexes,
- represent the map as a cell map,
- compute the chain maps of this map, and
- compute the homology maps of these chain maps.

Example 2.39 (embedding). We consider maps from the circle to the Möbius band:

$$f : \mathbf{S}^1 \to \mathbf{M}^2.$$

First, we map the circle to the median circle of the band:



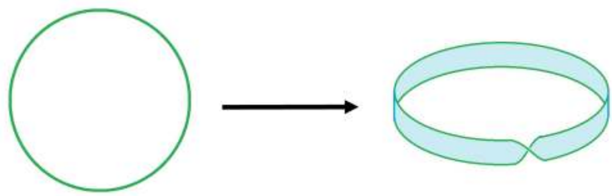
As you can see, we are *forced* to subdivide the band's square into two to turn this map into a cell map. Then we have:

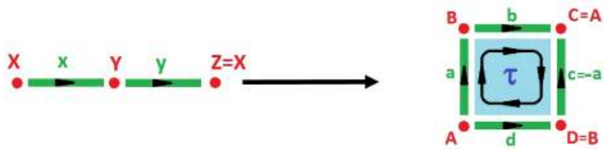
$$f(X) = f(Y) = M \text{ and } f(x) = m.$$

Then, due to $f_1(x) = m$, the 1st homology map is the identity:

$$f_* = \mathrm{Id} : H_1(K) = \mathbf{Z} \to H_1(L) = \mathbf{Z}.$$

Second, we map the circle to the edge of the band:





As you can see, we have to subdivide the circle's edge into two edges to turn this map into a cell map. Then we have:

$$f(X) = B,\ f(Y) = A \text{ and } f(x) = b,\ f(y) = d.$$

Then, due to $f_1(x + y) = b + d$, one can see that the 1st homology map is *not* the identity:

$$f_* = 2 \cdot \text{Id} : H_1(K) = \mathbf{Z} \to H_1(L) = \mathbf{Z}.$$

□

Exercise 2.40. Provide details of the homology computations in the last example.

Exercise 2.41. Consider a few possible maps for each of these and compute their homology maps:
- embeddings of the circle into the torus;
- self-maps of the figure eight;
- embeddings of the circle to the sphere.

2.7 Homology theory

Let's review how the homology theory of maps is built one more time:

A cell map produces its chain map, which in turn produces its homology map.

That's the whole theory.

Now, algebraically:

$$\begin{array}{rccccc} f: & |K| & \to & |L| & \rightsquigarrow \\ f_\Delta: & C(K) & \to & C(L) & \rightsquigarrow \\ f_*: & H(K) & \to & H(L). & \end{array}$$

We recognize each of these three settings as *categories*:
- cell complexes and cell maps,
- chain complexes and chain maps,
- modules and homomorphisms.

Indeed, in each of them, we can always complete this diagram with a composition:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {}_{gf}\searrow & \downarrow g \\ & & Z \end{array}$$

Let's, once again, consider these two steps separately.

First, the topological step. It relies on these familiar theorems.

Theorem 2.42. The identity cell map induces the identity chain map:

$$\left(\mathrm{Id}_{|K|}\right)_i = \mathrm{Id}_{C_i(K)}\,.$$

Theorem 2.43. The chain map of the composition of two cell maps is the composition of their chain maps:

$$(gf)_i = g_i f_i.$$

Second, the algebraic step.

Theorem 2.44. The identity chain map induces the identity homology map:

$$\left(\mathrm{Id}_{C(K)}\right)_* = \mathrm{Id}_{H(K)}\,.$$

Theorem 2.45. The homology map of the composition of two chain maps is the composition of their homology maps:

$$(gf)_* = g_* f_*.$$

Now, we combine the two steps omitting the intermediate state.

The following **notation** we have been using implicitly: when the chain complex $M = C(K)$ comes from a cell complex K, we replace the cumbersome $H(C(K))$ with $H(K)$ (as well as $H(C(K,K'))$ with $H(K,K')$).

The following is obvious.

Theorem 2.46. The identity cell map induces the identity homology map:

$$\left(\mathrm{Id}_{|K|}\right)_* = \mathrm{Id}_{H(K)}\,.$$

Theorem 2.47. The homology map of the composition of two cell maps is the composition of their homology maps

$$(gf)_* = g_* f_*.$$

The main result is below.

Theorem 2.48. Suppose K and L are cell complexes, if a map

$$f : |K| \to |L|$$

is a cell map and a homeomorphism, and

$$f^{-1} : |L| \to |K|$$

is a cell map too, then the homology map

$$f_* : H_k(K) \to H_k(L)$$

is an isomorphism for all k.

Proof. From the definition of inverse function,

$$\begin{aligned} ff^{-1} &= \mathrm{Id}_{|L|}, \\ f^{-1}f &= \mathrm{Id}_{|K|}\,. \end{aligned}$$

From the above theorems, it follows that

$$\begin{aligned} f_*(f^{-1})_* &= \mathrm{Id}_{H(L)}, \\ (f^{-1})_* f_* &= \mathrm{Id}_{H(K)}. \end{aligned} \qquad \blacksquare$$

The theorems prove that the effect of either of these two steps:

- from the category of cell complexes to the category of chain complexes; and then
- from the category of chain complexes to the category of modules,

is that objects are mapped to objects and morphisms to morphisms – in such a way that the compositions are preserved. Inevitably, the combination of these two steps has the same effect.

2.8 Functors

The results in the last subsection have a far-reaching generalization.

Definition 2.49. A *functor* $\mathscr{F}$ from category $\mathscr{C}$ to category $\mathscr{D}$ consists of two functions:

- the first associates to each object X in $\mathscr{C}$ an object $\mathscr{F}(X)$ in $\mathscr{D}$,
- the second associates to each morphism $f : X \to Y$ in $\mathscr{C}$ a morphism $\mathscr{F}(f) : \mathscr{F}(X) \to \mathscr{F}(Y)$ in $\mathscr{D}$.

In other words, a functor is a combination of the following. First:

$$\mathscr{F} : \mathrm{Obj}(\mathscr{C}) \to \mathrm{Obj}(\mathscr{D}).$$

And second, if $\mathscr{F}(X) = U,\ \mathscr{F}(Y) = V$, we have

$$\mathscr{F} = \mathscr{F}_{X,Y} : \mathrm{Hom}_{\mathscr{C}}(X,Y) \to \mathrm{Hom}_{\mathscr{D}}(U,V).$$

We assume that the following two conditions hold:

- (identity) $\mathscr{F}(\mathrm{Id}_X) = \mathrm{Id}_{\mathscr{F}(X)}$, for every object X in $\mathscr{C}$;
- (compositions) $\mathscr{F}(gf) = \mathscr{F}(g)\mathscr{F}(f)$, for all morphisms $f : X \to Y$, and $g : Y \to Z$.

The latter condition can be illustrated with a commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {}_{gf}\searrow & \downarrow{g} \\ & & Z \end{array} \quad \leadsto \quad \begin{array}{ccc} \mathscr{F}(X) & \xrightarrow{\mathscr{F}(f)} & \mathscr{F}(Y) \\ & {}_{\mathscr{F}(gf)}\searrow & \downarrow{\mathscr{F}(g)} \\ & & \mathscr{F}(Z) \end{array}$$

Exercise 2.50. Define (a) the identity functor, (b) a constant functor.

A slightly more complex is a *forgetful functor*: it is a functor that strips all structure off a given category and leaves us with the category of sets. For example, for groups we have:

$$\mathscr{F}((G,*)) = G.$$

Exercise 2.51. Prove that the composition of two functors is a functor.

Back to homology. Above, we proved the following result.

Theorem 2.52. Homology is a functor from cell complexes and maps to abelian groups and homomorphisms.

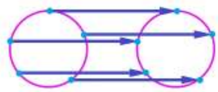
Exercise 2.53. Explain in what sense a chain complex is a functor.

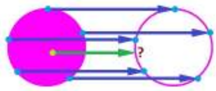
Exercise 2.54. Outline the homology theory for relative cell complexes and cell maps of pairs. Hint: $C(K,K') = C(K)/C(K')$.

This is how the purely "functorial" aspects of homology theory are used.

Example 2.55 (extension). Suppose we would like to answer the following question:
- Can a soap bubble contract to the ring without tearing?

We recast this question as an example of the *Extension Problem*:
- Can we extend the map of the circle onto itself to the whole disk?





In other words, is there a continuous F to complete the first diagram below so that it is commutative? Here, we translate our initial topological diagram – with cell complexes and cell maps – into *algebraic* diagram – with groups and homomorphisms:

$$\begin{array}{ccccccccccc}
\mathbf{S}^{n-1} & \hookrightarrow & \mathbf{B}^n & & H_{n-1}(\mathbf{S}^{n-1}) & \to & H_{n-1}(\mathbf{B}^n) & & \mathbf{Z} & \to & 0\\
& \searrow^{\mathrm{Id}} & \downarrow^{F=?} & \rightsquigarrow & & \searrow^{\mathrm{Id}} & \downarrow^{?} & = & & \searrow^{\mathrm{Id}} & \downarrow^{?}\\
& & \mathbf{S}^{n-1} & & & & H_{n-1}(\mathbf{S}^{n-1}) & & & & \mathbf{Z}
\end{array}$$

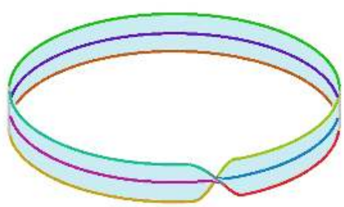
The latter is easy to evaluate and the result is the third diagram. That diagram is impossible to complete! □

Exercise 2.56. Provide such analysis for the extension problem(s) for: the identity map of the circle to the torus.

Exercise 2.57. For each $n \in \mathbf{Z}$, construct a map from the circle to the Möbius band:

$$f : K = \mathbf{S}^1 \to L = \mathbf{M}^2,$$

following the pattern we saw in the last subsection:



and confirm that its 1st homology map is the multiplication by n:

$$f_* = n \cdot \mathrm{Id} : H_1(K) = \mathbf{Z} \to H_1(L) = \mathbf{Z}.$$

Exercise 2.58. List possible maps for each of these based on the possible homology maps:
- embeddings of the circle into the torus;
- self-maps of the figure eight;
- embeddings of the circle to the sphere.

2.9 New maps from old

In this section, we have been using tools for building new topological spaces from old. The main tools are the product and the quotient. Both come with a method to construct appropriate new maps. We will review these maps and construct their combinations.

For two spaces X, Y, their product is a topological space defined on the product set $X \times Y$ with the product topology. If there are maps

$$f : X \to Y,\ f' : X' \to Y',$$

the *product map*

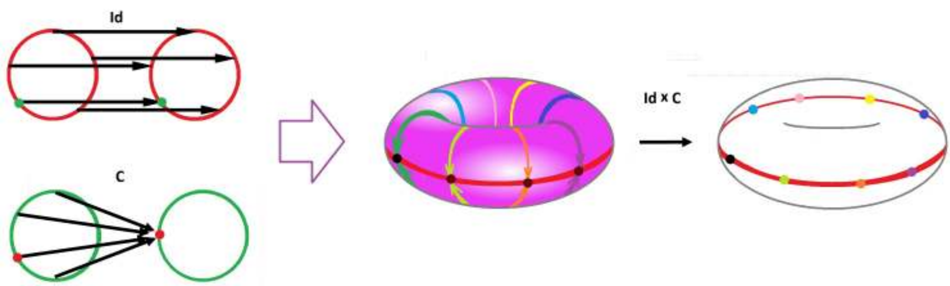
$$f \times f' : X \times X' \to Y \times Y',$$

of these maps is, naturally,

$$\Big(f \times f'\Big)(x, y) := \Big(f(x), f(y)\Big).$$

It is always well-defined.

Example 2.59. Suppose X, Y both are the circle. Then the product of the identity and a constant map collapses the torus on the equator:



□

Exercise 2.60. Represent, when possible, as product maps the following maps of the torus: constant, collapse to a meridian, collapse to the bottom half, collapse to the diagonal, rotation about its axis, reflection about a vertical plane.

Exercise 2.61. For a Q^m an m-cube, suppose we have an identification map

$$f_m : Q^m \to \mathbf{S}^m,\ f(\partial Q^m) = \{a\},\ a \in \mathbf{S}^m.$$

Describe $f_m \times f_n$.

Exercise 2.62. Show that fixing one of the spaces creates a simple functor, such as

$$\mathscr{F}(X) := \mathbf{I} \times X,\ \mathscr{F}(f) := \mathrm{Id} \times f.$$

Next, given a space X and an equivalence relation on it, the quotient is a topological space defined on the quotient set $X/_\sim$ with the quotient topology. For any map

$$f : X \to Y,$$

its *quotient map*

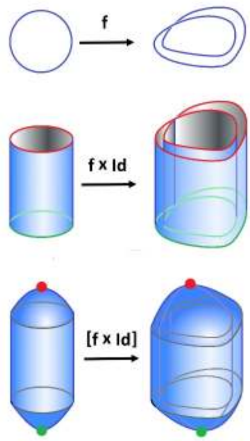
$$[f] : X/_\sim \to Y/_\sim,$$

is, naturally,

$$[f]([x]) := [f(x)].$$

It is well-defined whenever we have $x \sim y \Longrightarrow f(x) \sim f(y)$.

Example 2.63 (wrap). Suppose f is the "double wrap" of the circle. Then $f \times \mathrm{Id}$ is the "double wrap" of the cylinder.



Suppose the equivalence relation collapses the top and the bottom edges of the cylinder, separately. Then $[f \times \mathrm{Id}]$ is the "double wrap" of the resulting sphere. □

Unless both of the spaces are non-trivial, the product won't create any new topological features; indeed:

$$\mathbf{I} \times X \simeq X.$$

The flip side is that we can use products to produce new building blocks from old, such as cubes:

$$\mathbf{I} \times Q^n = Q^{n+1}.$$

The simplicial analog of this construction is the *cone* which is simply adding an extra vertex to the simplex:

$$\mathrm{C}\,\sigma^n = \sigma^{n+1}.$$

Both increase the dimension by 1!

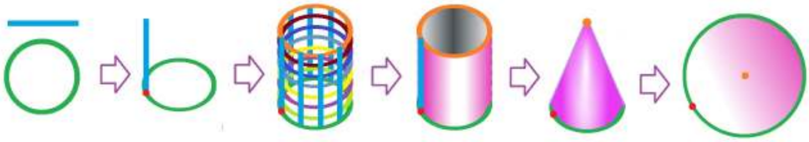
There is also the cone construction for topological spaces. Given a space X, first form the product $\mathbf{I} \times X$, then collapse the top to a point:

$$\mathrm{C}\,X := \Big(\mathbf{I} \times X\Big)/_{\{1\}\times X}\ .$$

The result is contractible:

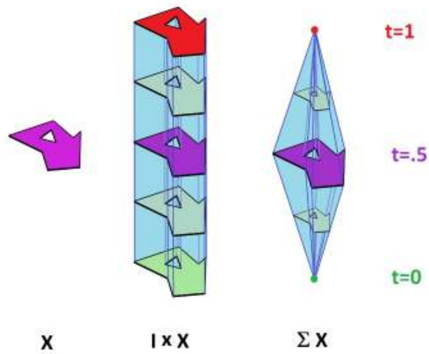
$$\mathrm{C}\,X \simeq \{p\}.$$

For instance, the cone of the circle is simply the disk.



However, *two* such cones glued together will produce the sphere!

This is how. First form the product $\mathbf{I} \times X$. Nothing new here. Then create the quotient by collapsing the top to a point and the bottom to a point.



Definition 2.64. Given a topological space X, define the *suspension* of X to be

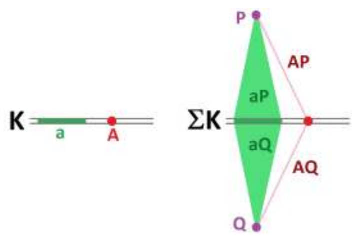
$$\Sigma X := \Big(\mathbf{I} \times X\Big)/_{\sim},$$

with the equivalence relation given by

$$(0, x) \sim (0, y),\ (1, x) \sim (1, y).$$

Without the second part, this is the cone of X.

Exercise 2.65. Given a triangulation of X, construct a triangulation of ΣX. Hint: use cones.



Exercise 2.66. Represent the suspension as a join.

The goal is to use the suspension to create new spaces with new, higher dimensional, topological features. Now, these features don't appear from nothing; just compare what happens to the homology classes:

- $\Sigma \mathbf{S}^1 \approx \mathbf{S}^2$, a non-trivial 2-class is produced from a non-trivial 1-class; but
- $\Sigma \mathbf{B}^1 \approx \mathbf{B}^2$, trivial 1-classes don't produce non-trivial 2-classes.

For the spheres, things are simple:

Proposition 2.67.

$$\Sigma \mathbf{S}^n \approx \mathbf{S}^{n+1},\ n = 0, 1, 2, \ldots.$$

Exercise 2.68. Prove the proposition.

Let's express the homology of the suspension ΣX in terms of the homology of X. The proposition implies:

$$H_{n+1}(\Sigma \mathbf{S}^n) \cong H_n(\mathbf{S}^n).$$

The case of $n = 0$ is an exception though:

$$\mathbf{S}^0 = \{p, q\} \Longrightarrow H_0(\mathbf{S}^0) = \mathbf{Z} \oplus \mathbf{Z}.$$

A convenient way of treating this exception is given by the reduced homology.

Theorem 2.69. For a cell complex K, we have

$$\tilde{H}_{n+1}(\Sigma K) \cong \tilde{H}_n(K),\ n = 0, 1, 2, ...$$

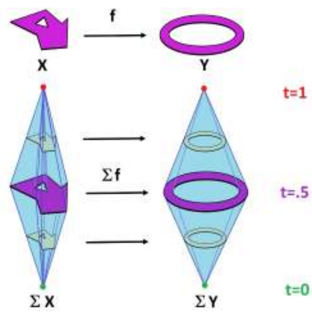
Exercise 2.70. Prove the theorem for simplicial complexes. Hint: write the two chain complexes and suggest a new kind of chain map between them, one that *raises* the dimension.

Definition 2.71. Given a map $f : X \to Y$, there is a map between their suspensions,

$$\Sigma f : \Sigma X \to \Sigma Y,$$

called the *suspension map* of f, defined by

$$(\Sigma f)([t, x]) := [t, f(x)].$$



Then, for a map f of the circles, its suspension is built by repeating f on the equator and all other longitudes. In particular, the "double wrap" of the sphere is the suspension of the "double wrap" of the circle, as shown above.

Exercise 2.72. Prove that the suspension map is well-defined.

Exercise 2.73. Prove that Σ is a functor from the category of topological spaces into itself.

Let's express the homology map of the suspension map Σf in terms of the homology map of f.

Theorem 2.74. Given a cell map $f : K \to L$, the following diagram is commutative for every $n = 0, 1, ...$:

$$\begin{array}{ccc} \tilde{H}_n(K) & \xrightarrow{f_*} & \tilde{H}_n(L) \\ \downarrow \cong & & \downarrow \cong \\ \tilde{H}_{n+1}(\Sigma K) & \xrightarrow{(\Sigma f)_*} & \tilde{H}_{n+1}(\Sigma L) \end{array}$$

Exercise 2.75. Prove the theorem for simplicial maps.

---

3 Maps of polyhedra

3.1 Maps vs. cell maps

Previously, we proved that if complex K^1 is obtained from complex K via a sequence of elementary collapses, then

$$H(K) \cong H(K^1).$$

(In spite of its length, the proof was straightforward.) However, the result, as important as it is, is a very limited instance of the invariance of homology. We explore next what we can achieve in this area if we take an *indirect* route and examine seemingly unrelated issues.

In order to build a complete theory of homology, our long-term goal now is to prove that the homology groups of two homeomorphic, or even homotopy equivalent, spaces are isomorphic – for the triangulable spaces we call *polyhedra*.

But these two concepts rely entirely on the general concept of *map*; meanwhile, this is the area where our theory remains most inadequate. Even if we are dealing with cell complexes, a map $f : |K| \to |L|$ between their realizations doesn't have to be a cell map. As a result we are unable to define its chain map $f_\Delta : C(K) \to C(L)$ and unable to define or compute its homology map $f_* : H(K) \to H(L)$.

The idea of our new approach is to *replace a map with a cell map*, somehow:

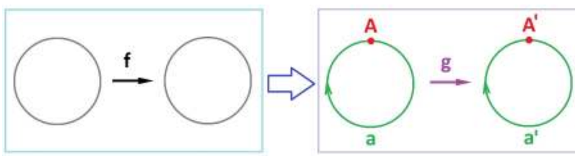
$$f : |K| \to |L| \quad \rightsquigarrow \quad g : K \to L,$$

and then simply declare that the homology map of the former is that of the latter:

$$f_* := g_* : H(K) \to H(L).$$

Naturally, we have to do it in such a way that this resulting homology map is well-defined. This is a non-trivial task.

Example 3.1 (self-maps of circle). We have already seen several examples of maps $f : \mathbf{S}^1 \to \mathbf{S}^1$ of the circle to itself that aren't cell maps:



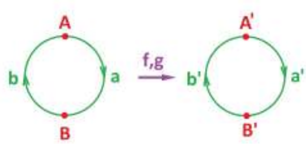
Suppose the circle is given by the simplest cell complex with just two cells A, a. Let's list *all* maps that can be represented as cell maps:
- *the* constant map: $f(u) = A \rightsquigarrow g(A) = A',\ g(a) = A$;
- the identity map: $f(u) = u \rightsquigarrow g(A) = A',\ g(a) = a'$;
- the flip around the y-axis: $f(x, y) = (-x, y) \rightsquigarrow g(A) = A',\ g(a) = -a'$.

These are examples of simple maps that *aren't* cell maps:
- any other constant map;
- any rotation;
- any double-wrapping of the loop (or triple-, etc.);
- the flip around any other axis.

And, we can add to the list anything more complicated than these!

Let's see if we can do anything about this problem.

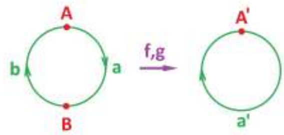
One approach is to try to *refine* the given cell complexes. For example, cutting each 1-cell in half allows us to accommodate a rotation through 180 degrees:



Indeed, we can set:

$$g(A) = g(B) := A',\ g(a) := b',\ g(b) := a'.$$

What about "doubling"? Cutting the only 1-cell of the domain complex (but not in the target!) in half helps:



Indeed, we can set:

$$g(A) = g(B) := A',\ g(a) = g(b) := a'.$$

Can we accommodate the rotation through 180 degrees by subdividing only the domain? Yes:

$$g(A) = g(B) := A',\ g(a) := a',\ g(b) := A'.$$

Since b collapses, the representation isn't ideal; however, it does capture what happens to the 1-cycle. □

Exercise 3.2. Provide cell map representations of the maps listed above with as few subdivisions as possible.

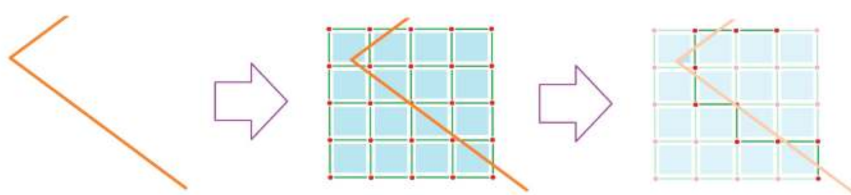
Exercise 3.3. Demonstrate that the rotation of the circle through $\sqrt{2}\pi$ cannot be represented as a cell map no matter how many subdivisions one makes.

3.2 Cell approximations of maps

There are many more maps that might be hard or impossible to represent as cell maps, such as these maps from the circle to the ring:

Indeed, if we zoom in on this ring and it is equipped with a cubical complex structure, we can see that the diagonal lines can't be represented exactly within this grid, no matter how many times we refine it. There is another way however...

The idea is to *approximate* these maps:



The meaning of approximation is the same as in other fields: every element of a broad class of functions is substituted – with some accuracy – with an element of a smaller class of "nicer" functions. For example, differential functions are approximated with polynomials (Taylor) or trigonometric polynomials (Fourier). The end result of such an approximation is a *sequence* of polynomials converging to the function. This time, the goal is to

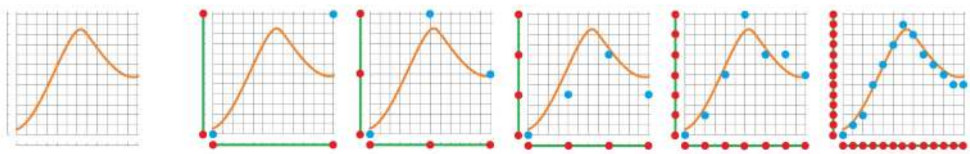
- approximate continuous functions (maps) with cell maps.

But how do we measure how close two maps are within a complex? Because we want the solution independent of the realization, no actual measurements will be allowed. Instead, we start with the simple idea of using "cell distance" as our one and only measurement of proximity.

We can say that a cell map $g : K \to L$ *approximates* a map $f : |K| \to |L|$ at vertex A of K if its value $g(A)$ is within the cell distance from $f(A)$; i.e., there is a cell $\sigma \in L$ such that

$$f(A), g(A) \in \sigma.$$

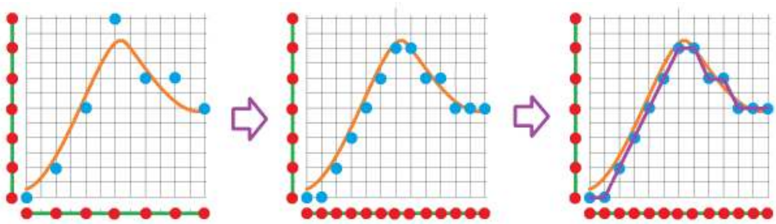
Judging by the above examples, this constraint won't make the functions very close. The only way to allow for cell maps to approximate f better and better and generate a convergent sequence is to *refine the complexes*. Below we can see how, under multiple subdivisions, cell maps start to get closer to the original:



Even though these approximations seem to work well, they are incomplete as they are defined only on the vertices. We need to extend them to the edges.

We discover that we *cannot* extend the "cell maps" in 3rd, 5th, and 6th graphs above to 1-cells. The reason is the same: as x changes by 1 cell, y changes by 2 or more.

The idea of how to get around this problem is suggested by the continuity of f. This property will ensure that a small enough increment of x will produce an increment of y as small as we like. Then the plan is to subdivide but *subdivide only the domain*:

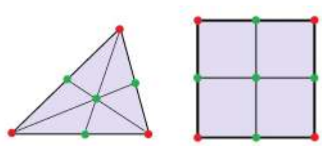


Exercise 3.4. Show that the approach works for all examples above.

We will prove that this positive outcome is guaranteed – after a sufficient number of subdivisions.

Without refining the target space, repeating this approximation doesn't produce a sequence g_n convergent to the original map f. For such an approximation to have any value at all, g_n has to become "good enough" for large enough n. We will demonstrate that it is good enough *topologically*: $g_n \simeq f$!

From this point on, we limit ourselves in this section to *simplicial and cubical complexes* and we refer to them as simply "complexes" and to their elements as "cells". The reason is that either class has a standard *refinement scheme*: the barycentric subdivision for simplicial and the "central" subdivision for cubical complexes:



In either case, each cell acquires a new vertex in the middle. For a given complex k, the nth subdivision of this kind will be within this section denoted by K^n (not to be confused with $K^{(n)}$, which is the nth skeleton).

Exercise 3.5. Given a cubical complex K, list all cells of K^1.

Exercise 3.6. Prove that the sizes of the cells converge to 0 as $n \to \infty$, for both types of complexes.

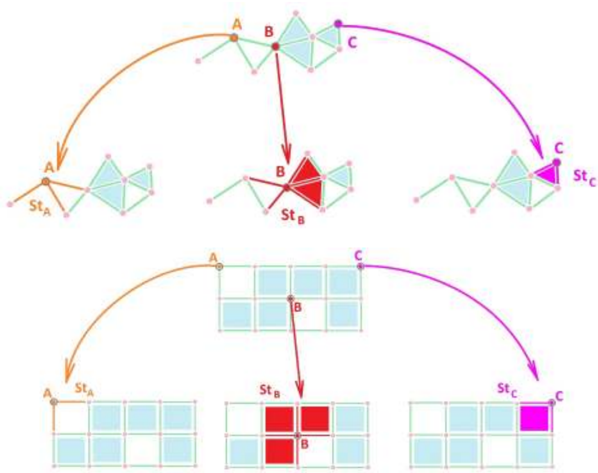
In addition to vertices mapped by f and g within a cell from each other, we need to ensure a similar behavior for all other cells. Suppose for a moment that the values of f and g coincide on all vertices of K. Then any cell a adjacent to a vertex A must be mapped to a cell adjacent to vertex $f(A) = g(A)$. Here we recognize a familiar situation: we are talking about a and $f(a)$ located within the *star* of the corresponding vertex!

Recall that given a complex K and a vertex A in K, the star of A in K is the collection of all cells in K that contain A:

$$\mathrm{St}_A = \mathrm{St}_A(K) := \{\sigma \in K : A \in \sigma\} \subset K,$$

while the open star is the union of the interiors of all these cells:

$$N_A = N_A(K) = \bigcup \{\dot{\sigma} : \sigma \in St_A(K)\} \subset |K|.$$



We recall also this simple result about simplicial complexes, which also holds for cubical complexes.

Proposition 3.7. The set of all open stars of all vertices of complex K forms an open cover of its realization $|K|$.

Exercise 3.8. Prove that the set of all open stars of all vertices of all subdivisions of complex K forms a basis of the topology of its realization $|K|$.

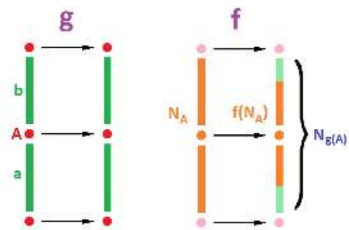
We are ready now to state the main definition.

Definition 3.9. Suppose $f : |K| \to |L|$ is a map between realizations of two complexes. Then a cell map $g : K \to L$ is called a *cell approximation* of f if

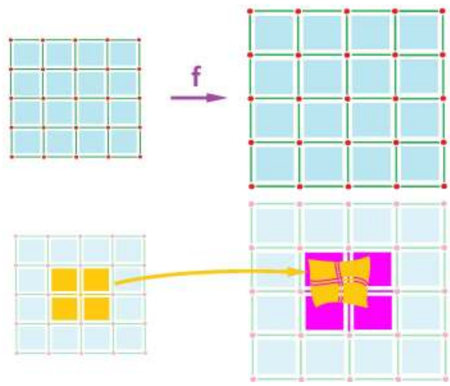
$$f(N_A) \subset N_{g(A)},$$

for any vertex A in K.

The condition is illustrated for dimension 1:



And for dimension 2:



Observe here that, if we assume again that f and g coincide on the vertices, the above condition is exactly what we have in the definition of *continuity*:

$$f(N_x) \subset N_{f(x)},$$

if N_a stands simply for a neighborhood of point a.

Proposition 3.10. Cell approximation is preserved under compositions. In other words, given

$$p : |K| \to |L|,\ q : |L| \to |M|,$$

maps between realizations of three complexes, if

$$g : K \to L,\ h : L \to M$$

are cell approximations of p, q respectively, then hg is a cell approximation of qp.

Exercise 3.11. Prove the proposition. Hint: compare to compositions of continuous functions.

3.3 The Simplicial Approximation Theorem

We need to prove that such an approximation always exists.

We start, for a given map $f : |K| \to |L|$, building a cell approximation $g : K^m \to L$, for some m.

We begin at the ground level, the vertices. The value $y = f(A)$ of f at a vertex A of K doesn't have to be a vertex in L. We need to pick one of the nearest, i.e., in the cell of L that contains A inside of it:

- $g(A) = V$ for some V, a vertex of cell $\sigma \in L$ such that $f(A) \in \dot{\sigma}$.

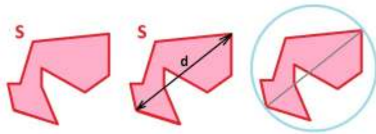
This cell is called the *carrier*, carr(y), of y.

Exercise 3.12. Prove that the carrier of y is the smallest (closed) cell that contains it.

Exercise 3.13. Prove that g is a cell approximation of f if and only if

$$\text{carr}(g(x)) \subset \text{carr}(f(x)).$$

Would all or some of these vertices satisfy the definition? Not when the cells of K are "large". That's why we need to subdivide K to reduce the size of the cells in a uniform fashion. We have measured the sizes of sets topologically in terms of open covers, i.e., whether the set is included in one of the elements of the cover. In a metric space, it's simpler:



Definition 3.14. The *diameter* of a subset S of a metric space (X, d) is defined to be

$$\operatorname{diam}(S) := \sup\{d(x, y) : x, y \in S\}.$$

Exercise 3.15. In what sense does the word "diameter" apply?

Exercise 3.16. Prove that for a compact S, this supremum is "attained": $\operatorname{diam}(S) = d(x, y)$ for some $x, y \in S$.

Every simplicial complex K has a geometric realization $|K| \subset \mathbf{R}^N$, while every cubical complex *is* a collection of cubes in $\mathbf{R}^N$. Therefore, for any cell σ in K, we have:

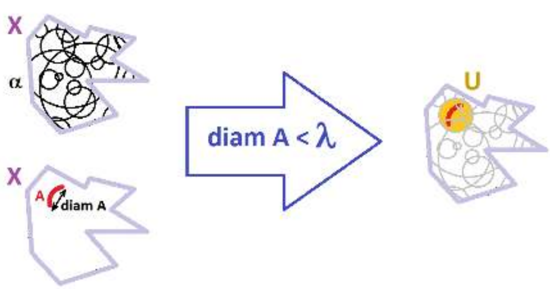
$$\exists x, y \in \sigma, \operatorname{diam}(\sigma) = ||x - y||.$$

Exercise 3.17. (a) Prove this fact directly, without using the last exercise. (b) Prove that these x, y belong to the boundary of σ.

Exercise 3.18. Find the diameter of an n-cube.

Now, we compare the idea of measuring sizes of sets via open covers and via the diameters.

Lemma 3.19 (Lebesgue's Lemma). Suppose (X, d) is a compact metric space and α is an open cover of X. Then every subset with small enough diameter is contained in a single element of the cover; i.e., there is a number $\lambda > 0$ such that if $S \subset X$ and $\operatorname{diam}(S) < \lambda$ then $S \subset U$ for some $U \in \alpha$.



Proof. We prove by contradiction.

Suppose for every $\lambda_n = \frac{1}{n}$, $n = 1, 2, ...$, there is a subset $S_n \subset X$ with
- (1) $\operatorname{diam}(S_n) < \frac{1}{n}$, and
- (2) $S_n \not\subset U$, $\forall U \in \alpha$.

Choose a single point in each set:

$$x_n \in S_n, \ n = 1, 2, ...$$

Then, by compactness, this sequence has a limit point, say $a \in X$. Since α is a cover, there is a $V \in \alpha$ such that $a \in V$, and, since it's open, there is a $B(a, \varepsilon) \subset V$ for some $\varepsilon > 0$. Now choose

m large enough so that

- (a) $d(a, x_m) < \frac{\varepsilon}{2}$, and
- (b) $1/m < \frac{\varepsilon}{2}$.

Using conditions (1) and (a), we conclude:

$$\operatorname{diam} S_m < \frac{1}{m} < \frac{\varepsilon}{2}.$$

It follows that

$$S_m \subset B(a, \varepsilon) \subset V.$$

This contradicts condition (2). ■

Exercise 3.20. Prove the last step of the proof. Hint: use condition (b) and the Triangle Inequality.

Definition 3.21. The *mesh* of a complex K is defined to be

$$\operatorname{mesh}(K) := \max_{\sigma \in K} \operatorname{diam}(\sigma).$$

We will use a variation of this common definition:

$$\operatorname{Mesh}(K) := \max_{A \in K} \operatorname{diam}(N_A).$$

Exercise 3.22. Find the mesh of a cubical complex in $\mathbb{R}^N$.

Lemma 3.23. For a complex K, $\operatorname{Mesh}(K) \leq 2 \operatorname{mesh}(K)$.

Lemma 3.24. For any complex K, $\operatorname{Mesh}(K^n) \to 0$ as $n \to \infty$.

Exercise 3.25. Prove the two lemmas.

Now, we need to refine the cell structure of complex K to such a degree that its stars would be mapped by f into the stars of L. We need to have for any vertex A in K:

$$f(N_A) \subset N_{g(A)}.$$

In other words, we require that

$$N_A \subset f^{-1}(N_V),\ V = g(A).$$

The sets such as the one on the right form an open cover of $|K|$:

$$\alpha := \{f^{-1}(N_V) : V \in L\}.$$

We now apply Lebesgue's Lemma to $X = |K|$ and this open cover. It follows that there is a number $\lambda > 0$ such that if $S \subset |K|$ and $\operatorname{diam} S < \lambda$ then $S \subset U$ for some $U \in \alpha$, or

$$S \subset f^{-1}(N_V).$$

Thanks to the second lemma above, we can choose m large enough so that

$$\operatorname{Mesh}(K^m) < \lambda,$$

or

$$\operatorname{diam}(N_A(K^m)) < \lambda.$$

Due to $|K^m| = |K| = X$, we conclude that

$$N_A(K^m) \subset f^{-1}(N_V(L)).$$

We summarize the outcome below.

Theorem 3.26. Given a map $f : |K| \to |L|$ between realizations of two complexes. Then there is such an M that for every vertex A in K^m with $m > M$, there is a vertex V_A in L such that

$$f(N_A(K^m)) \subset N_{V_A}(L).$$

Exercise 3.27. As a generalization of cubical and simplicial complexes, define a cell complex with a "refinement scheme" in such a way that the theorem holds.

Thus, we've made the first step in constructing $G : K^m \to L$. It is defined on all vertices:

$$g(A) := V_A.$$

Next, we need to extend g to all cells:

$$g(A_0...A_n) := g(A_0)...g(A_n).$$

Theorem 3.28 (Simplicial Approximation Theorem). Given a map

$$f : |K| \to |L|$$

between realizations of two simplicial complexes. Then there is a simplicial approximation of f, i.e., a simplicial map

$$g : K^m \to L,$$

for some m, such that

$$f(N_A(K^m)) \subset N_{g(A)}(L).$$

Proof. The inclusion is proven in the last theorem. Then the conclusion follows from the Simplicial Extension Theorem (subsection IV.4.5). ■

We derived the Simplicial Extension Theorem from the Star Lemma. For cubical complexes the lemma doesn't hold; indeed, two vertices may be the opposite corners of a square. An alternative way to approximate maps, including cubical maps, is discussed later.

Exercise 3.29. Sketch a simplicial approximation of a map for $\dim K = 1$ and $m = 1$.

Exercise 3.30. Suppose $g : \sigma^m \to \sigma$ is a simplicial approximation of the identity map $\text{Id} : |\sigma| \to |\sigma|$ of the complex of the n-simplex σ. Prove that the number of simplices cloned by g is odd. Hint: what happens to the boundary simplices of σ?

3.4 Simplicial approximations are homotopic

At this point, we can finally define, as planned, the *homology maps* of an arbitrary map $f : X \to Y$ between two polyhedra.

Definition 3.31. Suppose X, Y are realizations of two simplicial complexes K, L:

$$X = |K|, \ Y = |L|,$$

and suppose that there is a simplicial approximation $g : K^m \to L$, for some m, of f with its chain maps

$$g_k : C_k(K^m) \to C_k(L), \ k = 0, 1, 2, ...$$

well-defined. Then we let, as a matter of definition, the *chain maps of map* f to be

$$f_k := g_k, \ k = 0, 1, 2, ...,$$

and, further, the *homology maps of map* f to be

$$[f_k] := [g_k], \ k = 0, 1, 2, ...$$

To be sure, these homomorphisms are well-defined only if we can show that the homology maps produced by another simplicial approximation $h : K^n \to L$ of f coincide with it,

$$[g_k] = [h_k].$$

The chain maps for these two cell maps may of course be different; in fact, even the number of subdivisions may be different, $n \neq m$. Then, just to get started with the two homomorphisms, we need to show that the groups are the same:

$$H_k(K^m) \cong H_k(K^n) =: H_k(K).$$

Then, logically, another step should precede this proof. What is the meaning of the homology of a polyhedron?

Definition 3.32. Given a polyhedron X, we choose a triangulation K of X, i.e., some simplicial complex K with $X = |K|$, with well-defined homology groups $H_k(K)$ of K. Then, as a matter of definition, we let the *homology groups of polyhedron* X to be

$$H_k(X) := H_k(K), \ k = 0, 1, 2, ...$$

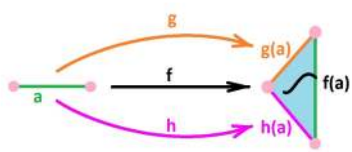
To be sure, these groups are well-defined only if we can show that

$$H_k(K') \cong H_k(K)$$

for any other triangulation K' of X. We call this the *invariance of homology.*

Both conclusions will be proven as by-products of what follows.

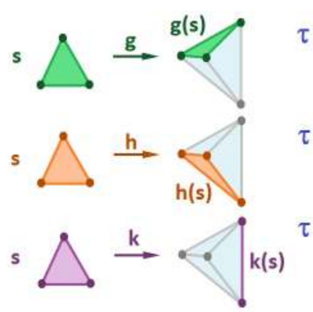
What do two simplicial approximations g and h of map f have in common? They are close to f and, therefore, close to each other:



Indeed, the Simplicial Approximation Theorem implies the following.

Proposition 3.33. For every simplex $\sigma \in K^m$, there is a simplex $\tau \in L$ such that

$$g(\sigma), h(\sigma) \subset \tau.$$



It immediately follows that, if we choose geometric realizations $|g|, |h| : X \to Y$ of these simplicial maps, they will be straight-line homotopic:

$$F(t, x) = (1 - t)|g|(x) + t|h|(x).$$

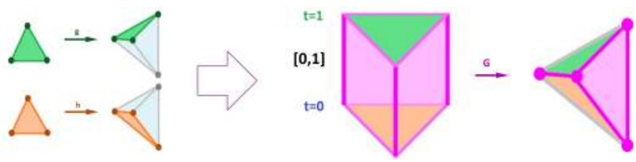
Unfortunately, a homotopy is just another *map*:

$$F : \mathbf{I} \times X \to Y,\ F(0, x) = |h|(x),\ F(1, x) = |g|(x),\ \forall x \in X.$$

Instead, we would like to use the fact that both g and h are simplicial maps and construct a *simplicial homotopy*; i.e., a simplicial map that satisfies:

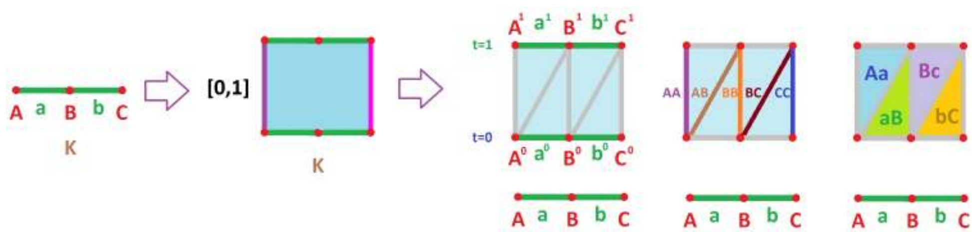
$$G : \mathbf{I} \times K \to L, G(\{0\} \times \sigma) = h(\sigma), G(\{1\} \times \sigma) = g(\sigma),\ \forall \sigma \in K,$$

with $\mathbf{I} \times K$ somehow triangulated. If such a map exists, we can say that g and h are *simplicially homotopic*.



Exercise 3.34. Under what circumstances is a straight line homotopy of (realizations) of two simplicial maps a (realization) of a simplicial homotopy?

We start with a triangulation. Naturally, we want to reuse the simplices of K. As we see in the 1-dimensional example below, each simplex of K appears twice – in the bottom $\{0\} \times K$ and the top $\{1\} \times K$ of the product, but also as the bases of several cones:



More generally, for each vertex A in K, we define two vertices A^0 and A^1 as the counterparts of the vertices A in $\{0\} \times K$ and $\{1\} \times K$ respectively:

$$A^i := \{i\} \times A,\ i = 0, 1.$$

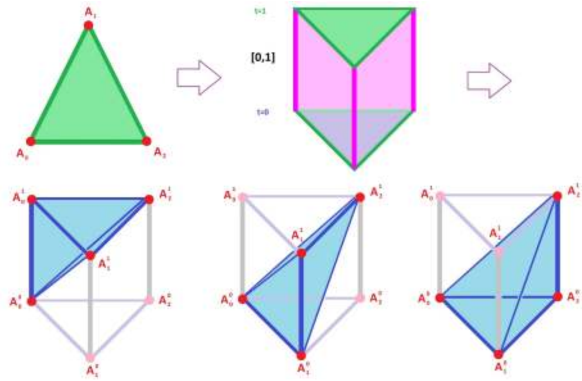
Then, for every simplex

$$\sigma = A_0 ... A_n \in K$$

and every $k = 0, 1, 2, ..., n$, we define a typical $(n+1)$-simplex in the triangulation of the product to be

$$\sigma_k := A_0^0 ... A_k^0 A_k^1 ... A_n^1.$$

Exercise 3.35. Show that this simplex is a certain join. Hint:



We notice now that, even though σ is a simplex, $\mathbf{I}\times\sigma$ isn't. However, it is a *chain*.

Exercise 3.36. Prove that

$$\mathbf{I}\times\sigma=\sum_k(-1)^k\sigma_k,$$

for some simplices σ_k in K.

As a simplicial map, G is defined on all vertices (top and bottom) first:

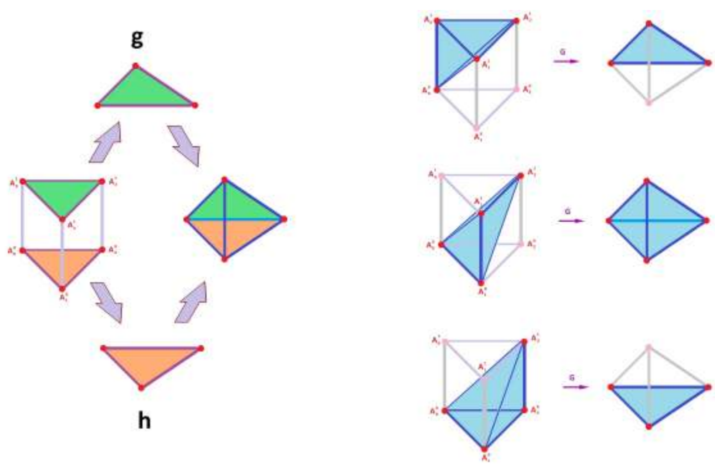
$$G(A^0):=h(A),\quad G(A^1):=g(A),$$

for every vertex A in K. Then we have the value of G for each new simplex σ_k above given by

$$G(A_0^0...A_k^0A_k^1...A_n^1)=h(A_0)...h(A_k)g(A_k)...g(A_n),$$

and, certainly, for the old ones:

$$G(A_0^0...A_n^0)=h(A_0)...h(A_n),\quad G(A_0^1...A_n^1)=g(A_0)...g(A_n).$$



Exercise 3.37. Prove that G is well-defined; i.e., its values are indeed simplices in L. Hint: use the Star Lemma and the proposition above.

Exercise 3.38. Why can't we simply use for G a simplicial approximation of the straight-line homotopy F?

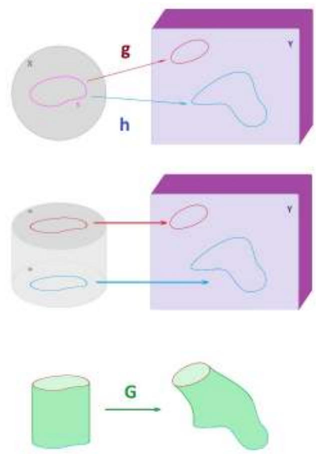
We have proven the following.

Theorem 3.39. All simplicial approximations of a given map are simplicially homotopic.

Exercise 3.40. How does one construct a cubical homotopy of two cubical maps? Hint: the product of two cubes is a cube.

3.5 Homology maps of homotopic maps

Why do two homotopic maps produce the same homology map? It suffices to take a look at what they do to a given *cycle*. The conclusion is very transparent: the combination of its images under these two maps bounds its image under the homotopy, illustrated below:



That is why these two images (they are cycles too!) are homologous.

Exercise 3.41. Sketch this construction for the case of a 2-cycle.

More precisely, here's the setup for this idea. We assume that $\mathbf{I} \times K$ is triangulated, so that $\mathbf{I} \times s$ is a specific $(k+1)$-chain for every k-cycle s in K. Suppose we have two simplicial maps

$$f, g : K \to L$$

and suppose

$$G : \mathbf{I} \times K \to L$$

is a simplicial homotopy between them:

$$G : g \simeq h : K \to L.$$

Suppose now that s is a k-cycle in K. Then the above observation can be seen as two ways of writing a certain k-cycle in L:

- the combination of the images of s under g, h properly oriented, i.e., chain $g(s) - h(s)$; or
- the boundary of the image under G of a certain $(k+1)$-chain, i.e., $\partial G(\mathbf{I} \times s)$.

This identity has to be written in terms of chain maps.

Proposition 3.42. Suppose chain maps $g_\Delta, h_\Delta : C(K) \to C(L)$ and $G_\Delta : C(\mathbf{I} \times K) \to C(L)$ satisfy:

$$g_k(s) - h_k(s) = \partial_{k+1} G_{k+1}(\mathbf{I} \times s),$$

for a k-cycle s in K. Then

$$[g_k](s) = [h_k](s).$$

Proof. When this identity is satisfied, the two cycles in the left-hand side are homologous, by definition; i.e.,

$$g_k(s) \sim h_k(s),$$

is written as

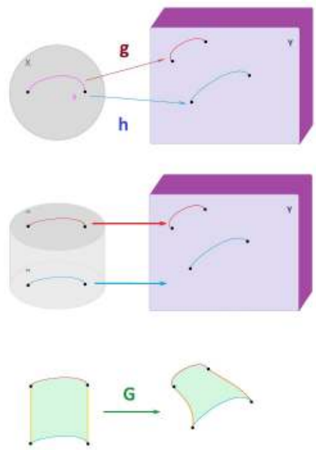
$$g_k([s]) = [g_k(s)] = h_k([s]) = [h_k(s)] \in H(L).$$

■

If the identity is satisfied for *all* cycles s in K, we have simply the identity of the maps:

$$g_* = h_* : H(K) \to H(L).$$

Our next step is to try to understand this identity in terms of chains rather than cycles. Let's compare the above diagram that starts with a 1-cycle to the diagram below that starts with a 1-chain:



As we see, the presence of a homotopy doesn't mean that these two chains form a boundary anymore. In fact, we can see that the boundary of the surface $G(\mathbf{I} \times s)$ has, in addition to $g(s)$ and $h(s)$, two extra edges. What are they?

Those two edges come from the two endpoints of the 1-chain s. But that's its boundary!

In general, we will have an extra term in the right-hand side, as below:

$$g_k(s) - h_k(s) = \partial_{k+1} G_{k+1}(\mathbf{I} \times s) + G_k(\mathbf{I} \times \partial_k s).$$

Exercise 3.43. Sketch this construction for the case of a 2-chain.

We proceed to make this identity fully algebraic. Both sides of the identity are functions of chains. In fact, we can interpret the expressions on the right as functions of the k-chain s. We let

$$F_k(s) := G_{k+1}(\mathbf{I} \times s).$$

Then the identity becomes

$$g(s) - h(s) = \partial F_k(s) + F_{k-1}(\partial s).$$

Unlike a chain map, this new "function of chains"

$$F_k : M_k \to N_{k+1},\ \ k = 0, 1, ...$$

raises the dimension of the chain. What happens is visible in the above diagram as one proceeds from a 1-chain s at the top left to the 2-chain at the bottom right.

At this point, we resign to the "chain level" entirely. We have:

- two chain complexes M, N, that may or may not be the chain complexes of some simplicial complexes;
- two chain maps $p, q : M \to N$, that may or may not be the chain maps of some simplicial maps, and
- a function of chains $F = \{F_k : M_k \to N_{k+1},\ \ k = 0, 1, ...\}$, a collection of homomorphisms,

that may or may not come from the chain map of some simplicial homotopy.
And, they follow a certain algebraic relation that ensures they generate the same homology maps $p_*, q_* : H(M) \to H(N)$.

Definition 3.44. The function of chains F is called a *chain homotopy* of two chain maps $p, q : M \to N$, and they are called *chain homotopic*, if the following is satisfied:

$$p - q = \partial F + F\partial.$$

We can use the same **notation**:

$$F : p \simeq q.$$

Exercise 3.45. Prove that chain homotopy defines an equivalence relation on chain maps.

Let's illustrate the definition of chain homotopy with a diagram. Below we have the standard setup of two chain complexes combined with two, instead of just one, chain maps; as a result, we have below two commutative diagrams in one:

$$\begin{array}{ccccccccccc}
... & \to & M_{k+1} & \xrightarrow{\partial^M_{k+1}} & M_k & \xrightarrow{\partial^M_k} & M_{k-1} & \to & ... & \to & 0 \\
... & & p_{k+1}\downarrow\downarrow q_{k+1} & & p_k\downarrow\downarrow q_k & & p_{k-1}\downarrow\downarrow q_{k-1} & & ... & & \\
... & \to & N_{k+1} & \xrightarrow{\partial^N_{k+1}} & N_k & \xrightarrow{\partial^N_k} & N_{k-1} & \to & ... & \to & 0
\end{array}$$

How does F fit into this? If we add diagonal arrows, $\swarrow$, for each F_k, we break each of these commutative squares into two *non-commutative* triangles. Let's extract only the relevant part:

$$\begin{array}{ccccc}
 & & M_k & \xrightarrow{\partial^M_k} & M_{k-1} \\
 & ^{F_k}\swarrow \ \neq & \downarrow g_k - h_k & \neq \ \swarrow_{F_{k-1}} & \\
N_{k+1} & \xrightarrow{\partial^N_{k+1}} & N_k & &
\end{array}$$

This diagram has to be accompanied with an explanation: the sum of the left path and the right path coincides with the vertical path.

We have proven the following.

Theorem 3.46. If two chain maps are chain homotopic, then their homology maps coincide:

$$p \simeq q : M \to N \Longrightarrow p_* = q_* : H(M) \to H(N).$$

In order to transition from this algebra back to the topology, we, naturally, use simplicial complexes and simplicial maps:

$$M = C(K),\ N = C(M),\ p = g_\Delta,\ q = h_\Delta.$$

The result is the following.

Proposition 3.47. If two simplicial maps $g, h : K \to L$ are simplicially homotopic, then their chain maps $g_\Delta, h_\Delta : C(K) \to C(L)$ are chain homotopic.

Corollary 3.48. The homology maps of two simplicially homotopic simplicial maps coincide.

Our final conclusion makes homology theory complete.

Corollary 3.49. The homology map of any map between polyhedra is well-defined.

Exercise 3.50. Prove these statements.

3.6 The up-to-homotopy homology theory

Homology takes the local information – in the form of cells and their adjacency – and converts it into the global information in the form of homology classes. In the last section, we developed a homology theory for the category of cell complexes and cell maps. The sentence below summarizes the new features of this theory.

Any map, and not just a cell map,

$$f : X \to Y$$

between two polyhedra produces – via a triangulation $X = |K|, Y = |L|$ and a simplicial approximation – its chain map

$$f_\Delta : C(K) \to C(L),$$

which in turn produces its homology map

$$f_* : H(K) \to H(L).$$

It might seem that what we have built is a homology theory with spaces still limited to polyhedra except this time the maps can be arbitrary. There is much more than that here. Let's review.

First, the topological step. It relies on these familiar theorems.

Theorem 3.51. The identity map induces the identity chain map:

$$\left(\mathrm{Id}_{|K|}\right)_\Delta = \mathrm{Id}_{C(K)}\,.$$

Theorem 3.52. The chain map of the composition of two maps is the composition of their chain maps:

$$(gf)_\Delta = g_\Delta f_\Delta .$$

Exercise 3.53. Prove that the polyhedra and maps form a category. Do the above theorems give us a functor?

Next the algebraic step (no change here!).

Theorem 3.54. The identity chain map induces the identity homology map:

$$\left(\mathrm{Id}_{C(K)}\right)_* = \mathrm{Id}_{H(K)}.$$

Theorem 3.55. The homology map of the composition of two chain maps is the composition of their homology maps:

$$(pq)_* = p_* q_* .$$

The main result is below.

Theorem 3.56 (Invariance of homology). Homology is independent of triangulation; i.e., if a map

$$f : |K| \xrightarrow{\ \approx\ } |L|$$

between realizations of simplicial complexes K and L is a homeomorphism then its homology map

$$f_* : H(K) \xrightarrow{\ \cong\ } H(L)$$

is an isomorphism.

Exercise 3.57. Prove this theorem.

The theorem allows us to combine the topological and the algebraic steps.

Definition 3.58. The *homology of a polyhedron* X is that of any triangulation K of X:

$$H(X) := H(K).$$

Corollary 3.59. The homology group of polyhedra is well-defined.

The two pairs of theorems above combined produce the following pair.

Theorem 3.60. The identity map induces the identity homology map:

$$\Big(\mathrm{Id}_X\Big)_* = \mathrm{Id}_{H(X)}\,.$$

Theorem 3.61. The homology operator of the composition of two maps is the composition of their homology maps:

$$(gf)_* = g_* f_*.$$

Thus, we have moved away from the combinatorial view of homology as it was developed for simplicial complexes and simplicial maps. We now look at homology in a topological way – homology now is a functor H from the category of polyhedra, which is just a special class of topological spaces, and arbitrary maps to the category of abelian groups.

Even though this new point of view is a significant accomplishment, it is the role played by *homotopy* that truly makes a difference. We state the following theorem without proof (see Rotman, *An Introduction to Algebraic Topology*, p. 72).

Theorem 3.62. For polyhedra, homology maps of homotopic maps coincide:

$$f \simeq g : X \to Y \Longrightarrow f_* = g_* : H(X) \to H(Y).$$

This theorem ensures a certain "robustness" of the homology of both maps (above) and spaces (below). Informally, this idea is captured by the expression "up to homotopy".

Corollary 3.63. Homotopy equivalent polyhedra have isomorphic homology.

Exercise 3.64. Suppose two holes are drilled through a cannon ball, $X := \mathbf{B}^3 \setminus \Big(L_1 \cup L_2\Big)$. Find the homology of X. Hint: there will be two cases.

Exercise 3.65. Prove that this is a category:
- the objects are the homotopy equivalence classes of polyhedra and
- the morphisms are the homotopy classes of maps.

Describe homology as a functor H from this category to abelian groups.

This up-to-homotopy point of view is so productive that it is desirable to develop its algebraic analog.

Theorem 3.66. Homology maps of chain homotopic chain maps coincide:

$$p \simeq q : M \to N \Longrightarrow p_* = q_* : H(M) \to H(N).$$

Definition 3.67. Two chain complexes M, N are called *chain homotopy equivalent* if there are chain maps $p : M \to N,\ q : N \to M$ such that $pq \simeq \mathrm{Id}_N,\ qp \simeq \mathrm{Id}_M$.

Exercise 3.68. Prove that this defines an equivalence relation on chain complexes.

Corollary 3.69. Chain homotopy equivalent chain complexes have isomorphic homology.

This formalization of the algebraic step (from chain complexes to homology) allows us to deal with various categories of topological spaces – polyhedra, cubical complexes, manifolds, etc. – in the identical fashion.

3.7 How to classify maps

Armed with this theory we can now attack the problem of homology of *all* maps, not just cell maps, from the circle to the circle. These maps are easier to visualize as maps to the ring:

Let's try to *classify* them.

Let
$$F := \{f : \mathbf{S}^1 \to \mathbf{S}^1\}$$
be the set of all such maps. Every such map will induce, as its 1st homology map, a homomorphism $f_* : \mathbf{Z} \to \mathbf{Z}$. The problem of classifying the set

$$H := \{h : \mathbf{Z} \to \mathbf{Z}\}$$

of homomorphisms is easy. In fact, we can list them in full based on the homomorphism's value on the generator $1 \in \mathbf{Z}$. We define $h_n : \mathbf{Z} \to \mathbf{Z}$ by requiring $h_n(1) = n$. Then

$$H = \{h_n : n \in \mathbf{Z}\}.$$

We have listed all possible homomorphisms and, therefore, all possible homology maps $f_* : \mathbf{Z} \to \mathbf{Z}$. So, F is broken into equivalence classes:

$$F_n := \{f \in F : f_*(1) = n\},\ F = \bigsqcup_n F_n.$$

Let's supply each of these classes with its most simple representative. We choose f_n to be the map that wraps the first circle n times around the second in a uniform fashion, counterclockwise. It is given by the formula: $f_n(t) = nt,\ t \in [0, 2\pi]$.

The analysis of the set of all maps of the n-sphere to itself is identical.

Exercise 3.70. Provide such analysis for maps from the circle to the figure eight.

Exercise 3.71. Justify the substitution made above of the circle with the ring.

But, are the elements of each F_n homotopic to each other? The affirmative answer compactly given by: $\pi_1(\mathbf{S}^1) = \mathbf{Z}$, goes beyond homology theory.

Recall next that, for X a topological space and A a subspace of X, a map $r : X \to A$ is called a retraction if
$$ri_A = \mathrm{Id}_A,$$
where i_A is the inclusion of A into X and Id_A is the identity map on A. This equation allows us to establish a relation between the homology groups of A and X:

$$r_*(i_A)_* = \mathrm{Id}_{H(A)}.$$

Moreover, we know that for the last identity to hold, all we need is a homotopy between these maps.

Definition 3.72. Let X be a topological space and A a subspace of X. A map $r : X \to A$ is called a *deformation retraction* of X to A and A is called a *deformation retract of* X if there is a homotopy relative to A between Id_X and r.

Exercise 3.73. Prove the following.

- (a) Any point in a convex set is its deformation retract.
- (b) The $(n-1)$-sphere $\mathbf{S}^{n-1}$ is a deformation retract of $\mathbf{R}^n \setminus \{0\}$.
- (c) The circle $\mathbf{S}^1$ is a deformation retract of the Möbius band $\mathbf{M}^2$.
- (d) The union of the equator and a meridian of the torus $\mathbf{T}^2$ is a deformation retract of the torus with a point removed.

Exercise 3.74. Find an example of a deformation retraction that isn't a retraction.

3.8 Means

Averaging numbers and even locations is a familiar task. For example, the arithmetic mean of m values is computed by this standard formula:

$$\text{the mean of } x_1, ..., x_m := \tfrac{1}{m}(x_1 + ... + x_m).$$

This formula might not work in a more complex setting. If these are just elements of a group, the group may have no division. If these are points in a subset of a vector space, the set may be non-convex. We take an axiomatic approach.

Definition 3.75. For a set X and integer $m > 1$, a function $f : X^m \to X$ is called a *mean* of degree m on X if it is

- symmetric:

$$f(s(u)) = f(u), \ \forall u \in X^m, \ \forall s \in \mathcal{S}_m;$$

and

- diagonal:

$$f(x, x, ..., x) = x, \ \forall x \in X.$$

For X an abelian group, a mean is *algebraic* if it is a homomorphism. For X a topological space, a mean is *topological* if it is a continuous map.

Theorem 3.76. There is an algebraic mean of degree m on an abelian group G if and only if G allows division by m. In that case, the mean is unique and is given by the standard formula above.

Proof. Suppose $g : G^m \to G$ is a mean.

Let a be any non-zero element of G; then let $b := g(a, 0, 0, ...) \in G$. We compute:

$$\begin{aligned} a &= g(a, a, ..., a) \\ &= g(a, 0, 0, ...) + g(0, a, 0, ..., 0) + ... + g(0, 0, ..., 0, a) \\ &= mb. \end{aligned}$$

Therefore, a is divisible by m.

Next, the uniqueness. In the following computation, we use the diagonality, the additivity, the symmetry, and then the additivity again:

$$\begin{aligned} x_1 + ... + x_m &= g(x_1 + ... + x_m, ..., x_1 + ... + x_m) \\ &= g(x_1, 0, ..., 0) + ... + g(0, ..., 0, x_1) \\ &\quad ... \\ &\quad + g(x_m, 0, ..., 0) + ... + g(0, ..., 0, x_m) \\ &= mg(x_1, 0, ..., 0) \\ &\quad ... \\ &\quad + mg(0, ..., 0, x_m) \\ &= mg(x_1, ..., x_m). \end{aligned}$$

■

Exercise 3.77. What can you say about the torsion part of a group G with a mean?

What happens to the theorem if we weaken the diagonality condition?

Theorem 3.78. Suppose G is a free abelian group and suppose $f : G^m \to G$ is a symmetric homomorphism such that, for some integer q, we have:

$$f(x, ..., x) = qx.$$

Then, $m \big| q$. In that case, f is a multiple of the sum.

Exercise 3.79. (a) Prove the theorem. (b) What if G has torsion?

Let $d : G \to G^m$ be the diagonal function:

$$d(x) := (x, ..., x).$$

Then the theorem states that we can't complete this diagram with a homomorphism that is also symmetric:

$$\begin{array}{ccc} G & \xrightarrow{d} & G^m \\ \text{Id} & \searrow & \downarrow{g=?} \\ & & G \end{array}$$

We would like to apply this theorem to the question of the existence of a topological mean, via homology.

Below we have a commutative diagram of polyhedra and maps to be completed (δ is the diagonal map) by finding a map f and, second, the same diagram with the homology, over $\mathbf{Z}$, applied:

$$\begin{array}{ccc} X & \xrightarrow{\delta} & X^m \\ \text{Id} & \searrow & \downarrow{f=?} \\ & & X \end{array} \quad \xrightarrow{H} \quad \begin{array}{ccc} H_p(X) & \xrightarrow{\delta_*=d} & H_p(X^m) \\ \text{Id} & \searrow & \downarrow{f_*=g=?} \\ & & H_p(X) \end{array}$$

At this point, we observe that the non-triviality of these groups could make it impossible to complete the latter and, therefore, the former diagram.

The "naive product formula" (subsection IV.6.9) for homology implies the following:

Lemma 3.80. Suppose X is path-connected and p is a positive integer. Suppose also:
- $H_k(X) = 0, \ k = 1, 2, ..., p-1$;
- $G := H_p(X) \neq 0$.

Then,

$$H_p(X^m) = G^m.$$

This lemma, in turn, implies the following:

Lemma 3.81.
- $d := \delta_* : G \to G^m$ is the diagonal function.
- $g := f_* : G^m \to G$ is symmetric.

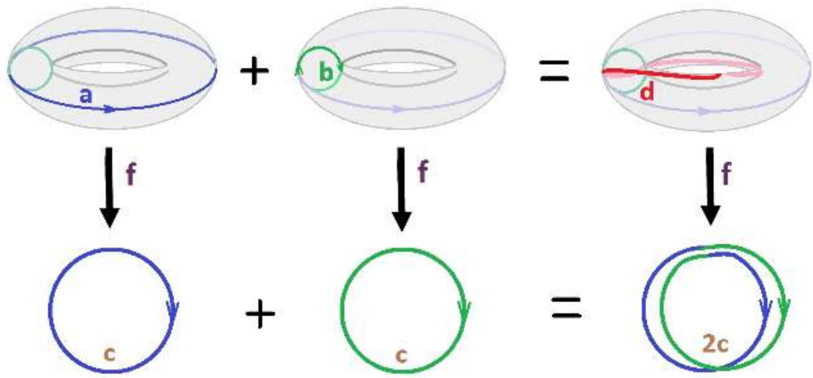
Exercise 3.82. Prove the lemma. Hint: watch the generators of G.

Unless X is acyclic, there is such a p. As a result, the last, algebraic, diagram becomes identical to the one in the last theorem, which leads to a contradiction. The main result is the following:

Theorem 3.83. Suppose X is path-connected and has torsion-free integral homology. If there is a topological mean on X then X is acyclic.

In other words, there can be no *extension* of the identity map of the diagonal of the product to the rest of it, in a symmetric manner. Whether such an extension always exists when X is acyclic lies beyond the scope of this book.

This is how the final contradiction is illustrated:



Exercise 3.84. Provide the missing details of the proof. Hint: use induction.

Exercise 3.85. Show that the proof fails if we choose G to be (a) $H_0(X)$, or (b) $C_i(X)$. Hint: a special case is when either i or X is small.

Exercise 3.86. Strengthen the theorem for the case of X a surface.

Exercise 3.87. What other maps of the diagonal lead to a contradiction?

If $\{G_i, \partial\}$ is a chain complex then so is $\{G_i^m, D\}$, with the boundary operator given by

$$D(x_1, ..., x_m) := (\partial x_1, ..., \partial x_m).$$

Theorem 3.88. Suppose $f_i : G_i^m \to G_i,\ i = 0, 1, ...,$ is a chain map. If each f_i is a mean then so is each homology map $[f_i] : H_i(\{G_i^m\}) \to H_i(\{G_i\}),\ i = 0, 1, ...$

Exercise 3.89. (a) Prove the theorem. (b) What if we relax the diagonality condition as above?

Both of these theorems will find applications in the theory of social choice.

3.9 Social choice: no impartiality

Recall that we have proved that the problem of two hikers has no solution in the case of a forest with a lake: there is no *cell map* $f : \mathbf{S}^1 \times \mathbf{S}^1 \to \mathbf{S}^1$ that satisfies

- $f(x, y) = f(y, x)$ and
- $f(x, x) = x$.

In light of the advances we've made, there is no such *map* whatsoever: there is no extension of the identity map of the diagonal of the torus to the rest of it in a symmetric manner.

These simple results already suggest that the reason why there is no such a solution is the non-triviality of the homology of the forest. We generalize this problem and discuss the new features.

This is the setup of the problem for the case of m voters/agents making their selections:

- the *space of choices*: a polyhedron W;
- the *choice* made by the kth voter (a "location vote"): a point $x_k \in W,\ k = 1, 2, ..., m$;
- the compromise decision: the *choice function* $f : W^m \to W$.

We say that f is a solution to the *social choice problem* if the following three conditions are satisfied:

- **Continuity (Stability Axiom):** The choice function $f : W^m \to W$ is continuous.
- **Symmetry (Anonymity Axiom):** The choice function is symmetric:

$$f(s(u)) = f(u),\ \forall u \in W^m,\ \forall s \in S_m.$$

- **Diagonality (Unanimity Axiom):** The choice function restricted to the diagonal is the identity:

$$f(x, x, ..., x) = x,\ \forall x \in W.$$

Therefore, this function is a topological mean on W. The necessary condition of existence of such a mean given in the last subsection yields the following important result.

Theorem 3.90 (Impossibility). Suppose the space of choices W is path-connected and has torsion-free homology. Then the social choice problem on W has no solution unless W is acyclic.

Example 3.91 (yeast colonies). The development of an yeast colony is cyclic: each stage is just a point on the circle $W = \mathbf{S}^1$. If two such colonies are to merge, what is the stage of the new one? The answer should be given by a function $f : \mathbf{S}^1 \times \mathbf{S}^1 \to \mathbf{S}^1$ that satisfies the axioms of symmetry and diagonality. Then the impossibility theorem indicates the discontinuous nature of such a merge. □

Exercise 3.92. State and solve homologically the problem of m hikers choosing a camp location in the forest.

Exercise 3.93. Suppose several countries are negotiating the placement of a jointly-owned satellite. Discuss.

Exercise 3.94. In light of the above analysis, what could be the meaning of a function

$$F : C(W \times W, W) \times C(W \times W, W) \to C(W \times W, W)?$$

We observe once again that excluding some options might make compromise as well as a compromise-producing rule impossible.

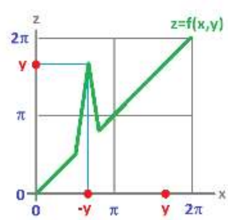
There is more to our theorem.

The up-to-homotopy approach provides a new insight now. The axioms – as constraints on the choice function that lead to a contradiction – can now be seen as weakened. Indeed, we don't have to require f to be symmetric: it suffices for f to be *homotopic* to such a function. Same applies to the diagonality requirement.

Example 3.95 (mother and child). To appreciate the difference, let's modify the two hikers problem. Suppose this time we have a mother and a small child choosing a place for a picnic in the park with a pond.

If the two are equal in this, there is no compromise as we know. Naturally, the mother can be a dictator. Can the child?

We still assume that the mother always decides on the location, $f(x, y) = x$. She does – unless of course the child starts to cry; then the decision becomes his. We can assume that this happens only when the mother's choice is very close to the diametrically opposite to that of the child's: $f(-y, y) = y$.



To ensure continuity, there must be a gradual transition between the two cases. It could be an arc of length $\varepsilon > 0$. For a small ε, the mother is a clear decision-maker. However, f is homotopic to h with $h(x, y) = y$. Therefore, even though not a dictator now, the child may gradually become one... □

Exercise 3.96. Provide f and the homotopy.

Definition 3.97. When the choice function $f : W^m \to W$ is homotopic to the ith projection h_i given by $h_i(x_1, ..., x_m) := x_i$, we call the ith agent a *homotopy dictator*.

Example 3.98. A simpler (and less dramatic) example of a homotopy dictator is the hiker who makes his choice of location in the forest and then allows the other person to move the site by, say, 20 feet or less. □

Exercise 3.99. State and prove a theorem to that effect.

To rephrase, a homotopy dictator is anyone who *can* become a full dictator gradually over time.

Plainly, if W is contractible, all choice functions are homotopic.

Proposition 3.100. If W is contractible, any social choice problem on W has a solution but everybody is a homotopy dictator.

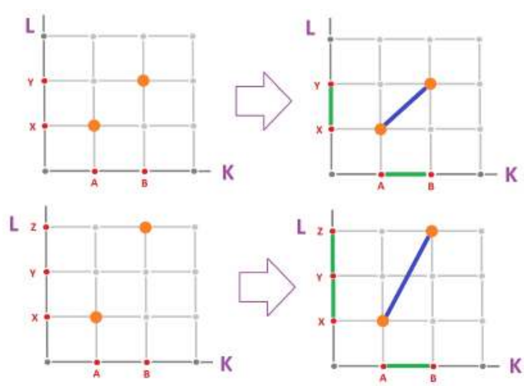
In this case, regardless of the starting point, anyone can gradually become a dictator.

3.10 The Chain Approximation Theorem

What about cubical complexes? No doubt, the homology theory we have developed is fully applicable; after all, the realizations of cubical complexes are polyhedra. However, we couldn't use the cubical structure to evaluate the homology unless we prove first that turning a cubical complex into a simplicial one preserves the homology groups. Below, we online an alternative approach.

The idea of the construction for a map $f : |K| \to |L|$ was to define a simplicial map $g : K^{(0)} \to L^{(0)}$ on the vertices first and then to extend it to the edges, faces, etc., $g : K^{(i)} \to L^{(i)},\ i = 1, 2, ..,$ simply using the convexity of the cells.

Both the problem we encounter and an idea of how to get around it are illustrated below:



The illustration shows how we may be able to create a cubical map:

$$g(AB) := XY,$$

by extending its values from vertices to edges. The extension succeeds (first illustration) when $X = g(A),\ Y = g(B)$ are the endpoints of an edge, XY, in L. When they aren't, the extension fails (second illustration); indeed, $XZ = g(A)g(B)$ isn't an edge in L. In that case, we may try to subdivide K as before. A different idea is to construct a *chain* for the value that we seek:

$$g(AB) := XY + YZ.$$

Thus, the general idea is then to construct a *chain map* that approximates f.

Let's review why we were unable to prove a cubical version of the approximation theorem along with the simplicial one.

First, the theorem that provides a continuity-type condition with respect to the cubical structure still holds:

Theorem 3.101. Given a map $f : |K| \to |L|$ of realizations of two complexes. Then there is such an M that for every vertex A in K^m with $m > M$ there is a vertex V_A in L such that

$$f(N_{K^m}(A)) \subset N_L(V_A).$$

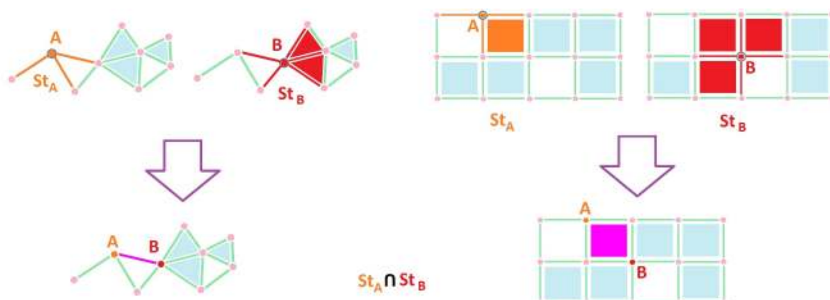
Second, just as with simplicial complexes, these vertices in L serve as the values of a "cubical approximation" g of f – defined, so far, on the vertices of K: $g_0(A) = V_A$. Then

$$f(N_{K^m}(A)) \subset N_L(g_0(A)).$$

The next step, however, is to extend g to edges, then faces, etc. of K. Unfortunately, the cubical analog of the Star Lemma that we need isn't strong enough for that.

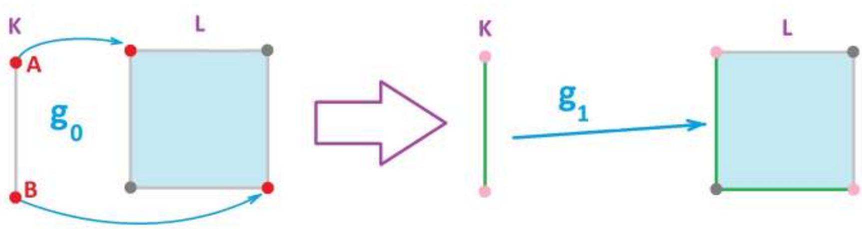
Lemma 3.102. Suppose $\{A_0, A_1, ..., A_n\}$ is a list, with possible repetitions, of vertices in a cubical complex. Then these vertices belong to the same cell in the complex if and only if

$$\bigcap_{k=0}^{n} N(A_k) \neq \emptyset.$$



Exercise 3.103. Prove the lemma.

In contrast to the case of simplicial complexes, here the condition doesn't guarantee that the vertices *form a cell* but only that they *are within the same cell*. Indeed, A_0A_1 may be the two opposite corners of a square. This is the reason why we can't guarantee that an extension of g from vertices to edges is always possible. For instance, even when AB is an edge in K, it is possible that $g_0(A)$ and $g_0(B)$ are the two opposite corners of a square in L:

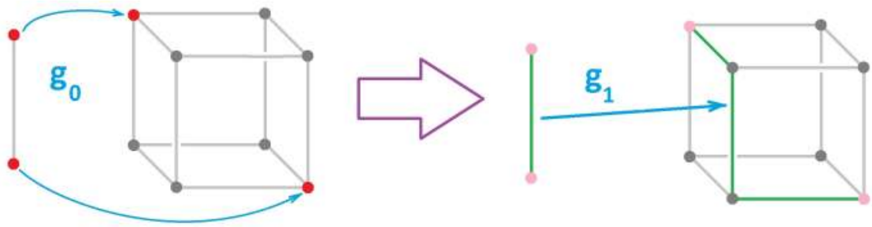


Then there is no single edge connecting them as in the case of a simplex. However, a *sequence of edges* with this property is possible. This sequence is the chain we are looking for.

Exercise 3.104. Prove that, for any map $g_0 : C_0(K^m) \to C_0(L)$ with

$$\{AB\} \in K^m \Longrightarrow g_0(A), g_0(B) \in \sigma \in L,$$

there is a map $g_1 : C_1(K^m) \to C_1(L)$ such that together they form a chain map in the sense that $\partial_1 g_1 = g_0 \partial_1$. Hint: the cube is edge-connected.



Let's make explicit the fact we have been using; it follows from the last lemma.

Lemma 3.105. Under the conditions of the theorem, if vertices $A_0, A_1, ..., A_n$ form a cell in K^m then the vertices $g_0(A_0), g_0(A_1), ..., g_0(A_n)$ lie within a single cell σ in L.

In the general case, the existence of such a chain follows from the acyclicity of the cube. Each step of our construction is justified by the following familiar result that we present as a refresher.

Lemma 3.106. Suppose complex Q is acyclic. Then Q satisfies the following equivalent conditions, for all $0 < i < k$:
- (a) $H_i(Q^{(k)}) = 0$;
- (b) every i-cycle is a boundary, in $Q^{(k)}$;
- (c) for every i-cycle a there is a $(i+1)$-cycle τ such that $a = \partial\tau$, in $Q^{(k)}$.

Theorem 3.107 (Chain Approximation Theorem). Given a map

$$f : |K| \to |L|$$

between realizations of two cubical complexes. Then there is a chain approximation of f, i.e., a chain map

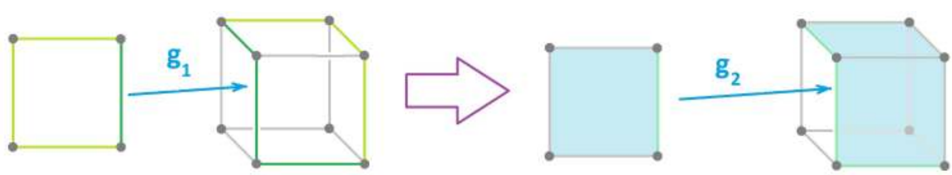
$$g = \{g_i\} : C(K^m) \to C(L),$$

for some m, such that

$$f(N_{K^m}(A)) \subset N_L(g_0(A)),$$

for any vertex A in K, and $|g(a)|$ lies within a cell of L for every cell a in K.

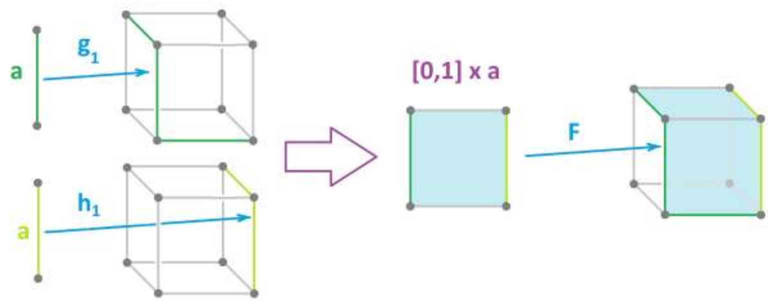
Proof. The proof is inductive. We build $g : C(K^m) \to C(L)$ step by step, moving from $g_i : C_i(K^m) \to C_i(L)$ to $g_{i+1} : C_{i+1}(K^m) \to C_{i+1}(L)$ in such a way that the result is a chain map: $\partial_{i+1} g_{i+1} = g_i \partial_{i+1}$. Since this extension happens within a single cell, which is acyclic, it is justified by the lemmas. ■



Exercise 3.108. Sketch an illustration for $i = 2$.

Exercise 3.109. Provide the details of the proof of the theorem.

Exercise 3.110. Follow the induction in the proof of the theorem to show that two chain approximations, as constructed, are chain homotopic. Hint:



The theorem fills a gap in the theory of *cubical homology*. Now we are in a position to say that the homology map of $f : |K| \to |L|$ is well-defined as the homology map of any of those chain approximations:

$$f_* := g_* : H(K) \to H(L).$$

The rest of the homology theory is constructed as before.

Let's set the approximation problem aside and can summarize the construction of the chains in the last part of the proof, as follows.

Theorem 3.111 (Chain Extension Theorem). Suppose K and L are cell complexes and suppose α is a collection of acyclic subcomplexes in L. Suppose $g_0 : K^{(0)} \to L^{(0)}$ is a map of vertices that satisfies the following condition:
• if vertices $A_0, ..., A_n$ form a cell in K then the vertices $g_0(A_0), ..., g_0(A_n)$ lie within an element of α.
Then there is such a *chain extension* of g_0, i.e., a chain map

$$g = \{g_i : \ i = 0, 1, 2, ...\} : C(K) \to C(L),$$

that it satisfies the following condition:
• $g(a)$ lies within an element of α for every cell a in K.
Moreover, any two chain maps that satisfy this condition are chain homotopic.

This theorem will find its use in the next section.

---

4 The Euler and Lefschetz numbers

4.1 The Euler characteristic

Recall the Euler Formula for a convex polyhedron:

$$\#\text{vertices} - \#\text{edges} + \#\text{faces} = 2.$$

A remarkable result, especially when you realize that it has nothing to do with convexity! For example, we can simply turn protrusions into indentations without changing this number:

We already know, and will prove below, that the nature of the formula is topological.

Definition 4.1. The *Euler characteristic* $\chi(K)$ of an n-dimensional cell complex K is the alternating sum of the number of cells in K for each dimension:

$$\chi(K) = \#0\text{-cells} - \#1\text{-cells} + \#2\text{-cells} - ... \pm \#n\text{-cells}.$$

Exercise 4.2. Compute the Euler characteristic of the circle, the cylinder, the Möbius band, the torus, and the Klein bottle.

To reveal the topological nature of this number, we will need to link it to homology. This task will require a more formal, algebraic approach which will prove to be very powerful when applied to the Euler characteristic and related concepts. In particular, counting cells in the definition is replaced with the following linear algebra.

Proposition 4.3. If $C_k(K)$ is the k-chain group of cell complex K, then

$$\#k\text{-cells} = \dim C_k(K).$$

Consequently,

$$\chi(K) = \sum_k (-1)^k \dim C_k(K).$$

We assume that all chains and homology below are over the *reals*.

We will need this well-known result.

Theorem 4.4 (Inclusion-Exclusion Formula). For sets $A, B \subset X$, we have:

$$\#(A \cup B) = \#A + \#B - \#(A \cap B).$$

What is the algebraic analog of this formula?

Proposition 4.5. For two subspaces U, V of a finitely dimensional vector space, we have

$$\dim < U \cup V >= \dim U + \dim V - \dim U \cap V.$$

Exercise 4.6. Prove the proposition. Hint: $\dim W = \#$ basis of W.

Theorem 4.7 (Inclusion-Exclusion Formula for Euler Characteristic). Suppose K, L, and $K \cap L$ are subcomplexes of the finite cell complex $K \cup L$. Then

$$\chi(K \cup L) = \chi(K) + \chi(L) - \chi(K \cap L).$$

Proof. By the above proposition, we have:

$$\begin{array}{c}
\dim < C_0(K) \cup C_0(L) >= \dim C_0(K) + \dim C_0(L) - \dim \big(C_0(K) \cap C_0(L)\big),\\
\dim < C_1(K) \cup C_1(L) >= \dim C_1(K) + \dim C_1(L) - \dim \big(C_1(K) \cap C_1(L)\big),\\
\vdots\\
\dim < C_n(K) \cup C_n(L) >= \dim C_n(K) + \dim C_n(L) - \dim \big(C_n(K) \cap C_n(L)\big).
\end{array}$$

For each $i = 0, 1, ..., n$, we have

$$\begin{array}{ll} < C_i(K) \cup C_i(L) > & = C_i(K \cup L), \\ C_i(K) \cap C_i(L) & = C_i(K \cap L). \end{array}$$

Then the alternating sum of the above equations gives us the formula. ■

The Euler characteristic behaves well under products.

Theorem 4.8 (Product Formula for Euler Characteristic). The Euler characteristic is multiplicative; i.e., for two finite cell complexes K and L, we have:

$$\chi(K \times L) = \chi(K) \cdot \chi(L).$$

Exercise 4.9. Prove the theorem. Hint: $\dim(U \oplus V) = \dim U + \dim V$.

Exercise 4.10. Compute the Euler characteristic of the n-dimensional torus $\mathbf{T}^n$.

4.2 The Euler-Poincaré Formula

The Euler characteristic of a cell complex is computed from this data:

$$c_k(K) := \dim C_k(K).$$

Notice that these numbers resemble (and in the simplest case are equal to) the Betti numbers of the cell complex:

$$\beta_k(K) := \dim H_k(K).$$

What if we evaluate an alternating sum of the latter numbers rather than the former?

Implicitly, we have already done this, in the Euler Formula. Indeed, we have:

$$\chi(\mathbf{S}^2) = 2,$$

which holds regardless of the triangulation. On the other, the alternating sum of the Betti numbers of the sphere

$$\beta_0(\mathbf{S}^2) = 1,\ \beta_1(\mathbf{S}^2) = 0,\ \beta_2(\mathbf{S}^2) = 1,\ \beta_3(\mathbf{S}^2) = 0, ...$$

also produces 2!

Is this a coincidence? Let's check out other complexes.

Example 4.11 (point). First, the point P:

$$\begin{array}{l} \chi(P) = 1 - 0 = 1, \\ \beta_0(P) - \beta_1(P) = 1 - 0 = 0. \end{array}$$

Once again, the Betti numbers add up to the Euler number. □

Example 4.12 (circle). Next, the circle C:

$$\begin{array}{l} \chi(C) = 1 - 1 + 0 = 0, \\ \beta_0(C) - \beta_1(C) + \beta_2(C) = 1 - 1 + 0 = 0. \end{array}$$ □

Example 4.13 (torus). Now the torus $\mathbf{T}^2$. We have:

- 0-cells: 1,
- 1-cells: 2,

- 2-cells: 1.

So,

$$\chi(\mathbf{T}^2) = 1 - 2 + 1 = 0.$$

Because this cell complex representation is so efficient, each of the cells also represents an element of a basis element of the corresponding homology group:

$$\begin{array}{ll} H_0(\mathbf{T}^2) = < A > & \Longrightarrow \beta_0(\mathbf{T}^2) = 1, \\ H_1(\mathbf{T}^2) = < a, b > & \Longrightarrow \beta_1(\mathbf{T}^2) = 2, \\ H_2(\mathbf{T}^2) = < \tau > & \Longrightarrow \beta_2(\mathbf{T}^2) = 1, \\ H_3(\mathbf{T}^2) = 0 & \Longrightarrow \beta_3(\mathbf{T}^2) = 0. \end{array}$$

The formula holds:

$$\chi(\mathbf{T}^2) = \beta_0(C) - \beta_1(C) + \beta_2(C) - ...$$ □

As we have been doing nothing but counting the cells, it's no surprise that the formula holds. Undoubtedly, it will always hold when there is *one cell per homology generator.* We need to show that this formula holds for all cell complexes.

Theorem 4.14 (Euler-Poincaré Formula). If K is a finite cell complex, then its Euler characteristic is equal to the alternating sum of the Betti numbers of each dimension:

$$\chi(K) = \sum_k (-1)^k \beta_k(K).$$

The proof will require a couple of facts from linear algebra of finite dimensional vector spaces.

Proposition 4.15.

- (1) If M, L are vector spaces and $A : M \to L$ is a linear operator, then

$$M/\ker A \simeq \operatorname{Im} A.$$

- (2) If Y is a subspace of vector space X, then

$$\dim X/Y = \dim X - \dim Y.$$

Exercise 4.16. Prove the proposition.

The proposition leads to the following conclusion.

Corollary 4.17. If M, L are vector spaces and $A : M \to L$ is a linear operator, then

$$\dim M - \dim \ker A = \dim \operatorname{Im} A.$$

Now, there are four vector spaces involved in the computation of the Betti numbers of complex K, for each $k = 0, 1, 2, ...$:

chain group:	C_k,	$c_k := \dim C_k$;
cycle group:	$Z_k = \ker \partial_k$,	$z_k := \dim Z_k$;
boundary group:	$B_k = \operatorname{Im} \partial_{k+1}$,	$b_k := \dim B_k$;
homology group:	$H_k = Z_k/B_k$,	$\beta_k := \dim H_k$,

where

$$\partial_k : C_k \to C_{k-1}$$

is the boundary operator.

Applying the corollary to the boundary operator, we obtain:

$$\dim C_k - \dim \ker \partial_k = \dim \operatorname{Im} \partial_k .$$

In other words, we have the following.

Lemma 4.18. $c_k - z_k = b_{k-1}$.

Next, from part (2) of the proposition, we have the following.

Lemma 4.19. $\beta_k = z_k - b_k$.

Finally, the **proof** of the theorem. Suppose n is the highest dimension of a cell in K. Then the right-hand side of the formula is computed as follows:

$$\begin{array}{lllll}
= & \beta_0 & -\beta_1 & +\beta_2 - ... & +(-1)^k \beta_n \\
 & & & \text{substitute from 2nd lemma} & \\
= & (z_0 - b_0) & -(z_1 - b_1) & +(z_2 - b_2) - ... & +(-1)^n (z_n - b_n) \\
 & & & & \\
= & z_0 & -b_0 - z_1 & +b_1 + z_2 - ... & +(-1)^n z_n - (-1)^n b_n \\
 & & & \text{substitute from 1st lemma} & \\
= & z_0 & -(c_1 - z_1) - z_1 & +(c_2 - z_2) + z_2 + ... & +(-1)^n z_n - (-1)^n (c_{n+1} - z_{n+1}) \\
 & & & \text{cancel the z's} & \\
= & z_0 & -c_1 & +c_2 - ... & +(-1)^{n+1} c_{n+1} - (-1)^n z_{n+1} \\
= & c_0 & -c_1 & +c_2 - ... & +0 \quad - \quad 0 \\
= & \chi(K). & & &
\end{array}$$

■

The invariance of homology then gives us the following.

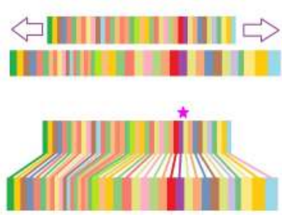
Corollary 4.20. The Euler characteristic is a topological invariant of polyhedra.

An interesting discovery but, because so much information is lost, the Euler characteristic is poor man's homology...

Exercise 4.21. Under what circumstances can the homology of X be derived from its Euler characteristic for (a) X a graph, (b) X a surface?

4.3 Fixed points

Imagine stretching a rubber band by moving one end to the right and the other to the left. After some experimenting one might discover that some point of the band will end up in its original position:



In fact, it doesn't matter if the stretching isn't uniform or if there is some shrinking, or even folding...

This problem is described mathematically as follows.

Fixed Point Problem: If X is a topological space and $f : X \to X$ is a self-map, does f have a *fixed point*, i.e., $x \in X$ such that $f(x) = x$?

The result above is, therefore, about a fixed point of a map $f : [a, b] \to [a, b]$.

Exercise 4.22. Prove this result by using the Intermediate Value Theorem.

Exercise 4.23. Demonstrate that both continuity of the map and the compactness of the interval are crucial.

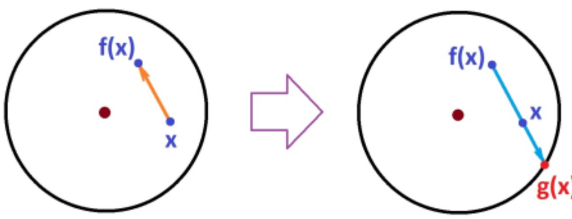
Exercise 4.24. Under what circumstances is the set of fixed points closed?

A similar problem about stretching a disk or a ball will require more advanced techniques.

Theorem 4.25 (Brouwer Fixed Point Theorem). Any self-map of a closed ball has a fixed point. In other words, if $f : \mathbf{B}^n \to \mathbf{B}^n$ is a map then there is a $x \in \mathbf{B}^n$ such that $f(x) = x$.

An informal interpretation of the 2-dimensional version of this theorem is: a page and its crumpled copy placed on it will have two identical words located one exactly over the other:

Exercise 4.26. Prove the theorem. Hint:



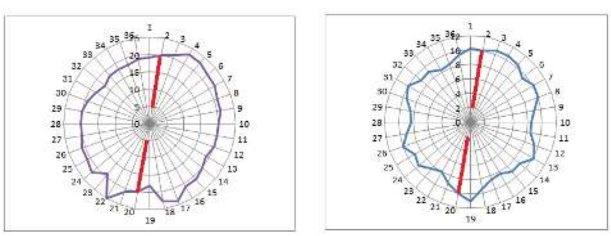
Plainly, the theorem holds if the ball is replaced with anything homeomorphic to it, such a simplex.

Though technically not a fixed point theorem, the following result is closely related to the last theorem.

Theorem 4.27 (Borsuk-Ulam Theorem). Every map $f : \mathbf{S}^n \to \mathbf{R}^n$ from an n-sphere to n-dimensional Euclidean space maps some pair of antipodal points to the same point:

$$f(x) = f(-x).$$

A common interpretation of the case $n = 1$ is that at any moment there is always a pair of antipodal points on the Earth's surface with exactly equal temperatures (or with equal barometric pressures, etc.).



Furthermore, an interpretation for $n = 2$ is that there is always a pair of antipodal points with exactly equal temperatures and equal barometric pressures, *simultaneously*.

Exercise 4.28. Explain how the theorem implies that at any moment there is always a pair of antipodal points on the Earth's surface with identical wind.

Exercise 4.29. Derive the case $n = 1$ of the theorem from the Intermediate Value Theorem.

Exercise 4.30. Prove the theorem. Hint: consider this map:

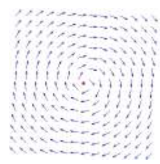
$$g(x) = \frac{f(x) - f(-x)}{||f(x) - f(-x)||}.$$

4.4 An equilibrium of a market economy

Suppose we have a map $F : S \to S$. Applied consecutively, it yields a *discrete dynamical system*:

$$X_{k+1} = F(X_k),$$

where $X_k \in S$ is the current state, at time k, of the system if it started at time $k = 0$ at the state $X_0 \in S$. Then the meaning of the *equilibrium* of the system is simple: $F(a) = a$. It's a fixed point!



A typical interpretation of such a system is a particle moving through a stream. We will consider instead the *dynamics of prices* in a simple market economy. Is it possible for them to stay constant? In other words, does every price dynamics allow an equilibrium?

Recall that we have n commodities freely traded with possible prices that form a *price vector* $p = (p_1, ..., p_n)$, at each moment of time. Then all positive, or even non-negative, prices are considered possible; i.e., the set of prices is

$$P := \mathbf{R}^n_+.$$

The dynamics of prices is represented by a continuous function $F : P \to P$. However, we can't conclude that there is a fixed point because P isn't compact!

Let's make some additional assumptions in order to be able to guarantee an equilibrium.

How can we make the set of prices compact? One approach is to normalize the prices:

$$p \mapsto \frac{p}{||p||},$$

under some norm so that the new prices form the n-simplex σ.

Exercise 4.31. Show that this is an equivalence relation. What assumptions would guarantee that the price function remains well-defined?

A more natural (if not more realistic) assumption is that the *money supply is fixed.* Suppose there are n commodities. We assume that this is a pure exchange economy; i.e., *there is no production* of goods (other than making up for possible consumption). Then there is only a fixed amount of each of the commodities:

$$\bar{c} := \sum_i c^i \in \mathbf{R}^n_+.$$

Then the total cost of all the commodities, i.e., the total wealth $W > 0$, is fixed too, for any combination of prices p:

$$\langle p, c\rangle = W.$$

Then the space of prices can be set to be

$$P := \{p \in \mathbf{R}^n_+ : \langle p, c\rangle = W\}.$$

Proposition 4.32. The set P of prices with a fixed money supply is a convex polyhedron.

Exercise 4.33. Prove the proposition. What kind of polyhedron is it?

Therefore, the price map $F : P \to P$ has a fixed point by the Brouwer's theorem.

Proposition 4.34. With a fixed money and commodity supply, a continuous price dynamics of an exchange economy has an equilibrium.

We have proven the existence of an equilibrium: there is such a combination of prices that, once acquired, will remain unchanged and all trading will happen in perfect harmony, forever... In reality, the outside events will be changing the function D itself.

Exercise 4.35. What difference would price controls have?

To be sure, this simple topological analysis won't tell us what this special combination of prices is, or how many of these are possible, or whether one of them will be reached (or even approached) from the current state.

Next, we consider the trading dynamics in this model. There are m "agents" in the market and ith agent owns c^i_j of the jth commodity. These amounts form the ith *commodity vector* $c^i \in \mathbf{R}^n_+$. The dynamics of the distribution of the commodities is represented by a continuous function $G : C \to C$, where

$$C := \{(c^1, ..., c^m) : \ c^i \in \mathbf{R}^n_+\} = \left(\mathbf{R}^n_+\right)^m.$$

Once again, we can't conclude that there is a fixed point because C isn't compact!

As before, we have to assume that there is no production or consumption. Then we can set instead:

$$C := \{(c^1, ..., c^m) : \ c^i \in \mathbf{R}^n_+, \ \sum_i c^i = \bar{c}\}.$$

Proposition 4.36. The set C of commodities with a fixed commodity supply is a convex polyhedron.

Exercise 4.37. Prove the proposition. What kind of polyhedron is it?

Exercise 4.38. What if the agents are allowed to borrow?

Once again, there is a fixed point by the Brouwer's theorem.

Proposition 4.39. With a fixed commodity supply, a continuous commodity dynamics of an exchange economy has an equilibrium.

In other words, there might be no trading at all!

We can combine these variables by defining the *state* of the system to be a combination of the prices and a distribution of the commodities. Then the *state space* is

$$S := P \times C,$$

and the state dynamics is given by a continuous function $D : S \to S$. Therefore, $S = P \times C$ is also a convex polyhedron and the map $D : S \to S$ has a fixed point by the Brouwer's theorem.

How can we incorporate other quantities into this model? Any new quantity can be added to this model by simply extending the state space S via the product construction. And there still be an equilibrium as long as:

- the next value of this quantity is determined by the circumstances not time,
- this dependence is continuous, and
- the range of its values is *a priori* bounded.

Exercise 4.40. What if we treat the money as just another commodity?

4.5 Chain maps of self-maps

We are to study fixed points of a self-map $f : |K| \to |K|$ by considering its homology map $f_* : H(K) \to H(K)$. We shouldn't forget, however, that the latter is built from the former via a simplicial approximation $g : K' \to K$ with K' an appropriate subdivision of K. Then g cannot possibly be a *self-map*!

The invariance of homology can be restated as follows.

Proposition 4.41. If K' is a subdivision of a simplicial complex K, then $H(K') \cong H(K)$.

In other words, two cycles are homologous in K' if and only if their "counterparts" in K are also homologous. What is the correspondence between the finer cycles of K' and the coarser cycles of K?

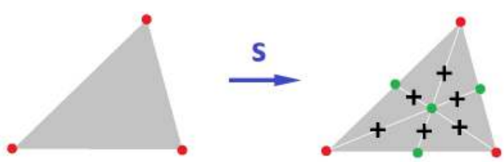
It is easy to find the counterpart in K' of a chain in K. We define the *subdivision map*

$$s : K \to K'$$

by setting the value of each k-simplex x in K to be

$$s(x) := \sum_i x_i,$$

with x subdivided into the collection of k-simplices $\{x_i\} \subset K'$ oriented compatibly with each other and x.



Proposition 4.42. The subdivision map extended linearly to chains,

$$s : C(K) \to C(K'),$$

is a chain map.

Exercise 4.43. Prove the proposition.

What about the opposite direction?

This time it is the domain that is subdivided. Fortunately, this is the familiar setting of simplicial approximation. A *simplicial approximation of the identity* can be constructed as follows.

Suppose, initially, that $K = A_0...A_n$ is an n-simplex and K' is its barycentric subdivision. We define a simplicial map $I : K' \to K$ in two steps. First, I is the identity on the vertices of K:

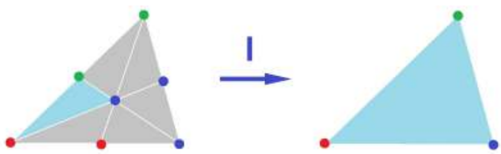
$$I(A_i) := A_i,\ i = 0, 1, ..., n.$$

There is only one vertex left, the barycenter A of K. Then let

$$I(A) := A_0$$

(or any other vertex). Furthermore, this rule is extended to the lower dimensional simplices of K:

- if vertex A in K' is the barycenter of simplex $B_0...B_m$ in K, let $I(A)$ to be one of these vertices.

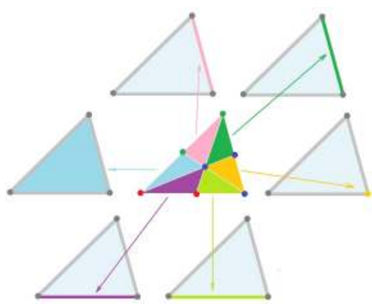


Within the rule, the vertices above are randomly assigned.

Exercise 4.44. Prove that this is indeed a simplicial map.

Lemma 4.45. There is exactly one n-simplex of K' that is cloned (to the whole K) by I.

Proof. Suppose $C_0 \in K'$ is the barycenter of the n-simplex K and $I(C_0) = B \in K$. Then there must be an $(n-1)$-subsimplex $\tau = B_0...B_{n-1}$ of K the vertices of which are taken by I to *other* vertices: $I(B_i) \neq B$. Next, let C_1 be the barycenter of simplex τ and repeat the process. Finally, we set $\sigma := C_0C_1...C_n$. ■



Exercise 4.46. (a) Prove the uniqueness. (b) Show how the construction and the proof applies to the general case of a simplicial complex K and its subdivision $K' = K^p$.

Theorem 4.47. The subdivision map generates the identity map on homology,

$$s_* = \text{Id} : H(K) \to H(K').$$

Proof. From the last lemma, it follows that $I_\Delta s = \text{Id} : C(K) \to C(K)$. ■

Accordingly, every self-map $f : |K| \to |K|$ does indeed have a matching *chain* self-map $sf : C(K') \to C(K')$ in the sense that $f_* = s_* f_* : H(K) \to H(K)$.

4.6 The Lefschetz number

The simplicial analog of a fixed point free map is a cell map that is fixed *cell* free. It is a self-map $g : K \to K$ of a cell complex K with

$$g(s) \neq s, \ \forall s \in K.$$

What does it mean algebraically? Let's consider the chain maps of g,

$$g_k : C_k(K) \to C_k(K), \ k = 0, 1, 2, ...$$

Then the above condition simply means that the *diagonal elements* of the matrix A of g_k are all zero! It follows, in particular, that the *trace* of this matrix, i.e., the sum of its diagonal elements:

$$\operatorname{Tr}(A) := \sum_i A_{ii},$$

is also zero. We develop this idea further and convert it fully from the chain level to the homology level by means of a generalization of the Euler-Poincaré formula.

Definition 4.48. The *Lefschetz number* $\Lambda(f)$ of a self-map $f : S \to S$ of polyhedron S is the alternating sum of the traces of the homology maps $[f_k] : H_k(S) \to H_k(S)$ of f, i.e.,

$$\Lambda(f) := \sum_k (-1)^k \operatorname{Tr}([f_k]).$$

As the alternating sum over the dimensions, the formula resembles the Euler characteristic. This is not a coincidence: the latter is equal to the Lefschetz number of the identity function of S:

$$\chi(S) = \Lambda(\mathrm{Id}_S).$$

This fact follows from the simple observation:

$$\operatorname{Tr}(\mathrm{Id} : \mathbf{R}^n \to \mathbf{R}^n) = n.$$

Exercise 4.49. (a) Compute the Lefschetz numbers of δP and $P\delta$, where $P : \mathbf{S}^2 \to \mathbf{S}^2$ is the projection and $\delta : \mathbf{S}^2 \to \mathbf{S}^2 \times \mathbf{S}^2$ is the diagonal map. (b) Replace $\mathbf{S}^2$ with M, a path-connected compact surface.

The main result is below.

Theorem 4.50 (Lefschetz Fixed Point Theorem). If a map $f : S \to S$ of a polyhedron S has non-zero Lefschetz number,

$$\Lambda(f) \neq 0,$$

then f has a fixed point, $f(a) = a$, and so does any map homotopic to f.

If the polyhedron S happens to be acyclic, we have

$$[f_0] = \mathrm{Id}, \ [f_1] = 0, \ [f_2] = 0, \ ...,$$

which yields $\Lambda(f) = 1 \neq 0$. The Brouwer Fixed Point Theorem follows.

Corollary 4.51. Suppose S is a polyhedron with $\chi(S) \neq 0$ (such as a sphere) and a map $f : S \to S$ is homotopic to the identity, $f \simeq \mathrm{Id}_S$. Then f has a fixed point.

The Lefschetz number only tells us that the set of fixed points is non-empty. The converse of course isn't true.

Exercise 4.52. Give a counter-example for the converse.

Example 4.53. Let f be the flip of the figure eight, $\mathbf{S}^1 \vee \mathbf{S}^1$, one loop onto the other. Then

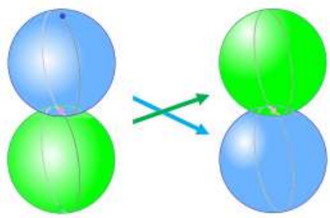
$$[f_0] = 1,\ [f_1] = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right],\ [f_2] = 0, ...$$

Hence,

$$\Lambda(f) = 1 - 0 = 1.$$

Not only there is a fixed point, as we can see, but also any map homotopic to f has a fixed point. $\square$

Example 4.54. We consider the flip of the "kissing spheres" space, $\mathbf{S}^2 \vee \mathbf{S}^2$, one sphere onto the other:



Then

$$[f_0] = 1,\ [f_1] = 0,\ [f_2] = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right],\ [f_3] = 0, ...$$

Hence,

$$\Lambda(f) = 1 - 0 + 0 = 1.$$ $\square$

Exercise 4.55. Consider the flip of the "double banana".

Exercise 4.56. Consider the other flip of the figure eight, the "kissing spheres", and the "double banana".

Exercise 4.57. Suggest a map f with $\Lambda(f) \neq 0$ but $\Lambda(f) = 0$ modulo 2.

The beginning of the **proof** of the Lefschetz's theorem is similar to that of the Euler-Poincaré Formula, except, we will manipulate the traces of linear operators rather than the dimensions of subspaces (after all, $\dim V = \operatorname{Tr}(\operatorname{Id}_V)$). We rely on the following fact from linear algebra.

Lemma 4.58. Suppose we have a commutative diagram of vector spaces and linear operators:

$$\begin{array}{ccccccccc} 0 & \to & U & \hookrightarrow & V & \to & W & \to & 0 \\ & & \downarrow f^U & & \downarrow f & & \downarrow f^W & & \\ 0 & \to & U & \hookrightarrow & V & \to & W & \to & 0 \end{array}$$

If the rows are exact rows, then we have:

$$\operatorname{Tr}(f) = \operatorname{Tr}(f^U) + \operatorname{Tr}(f^W).$$

Exercise 4.59. Prove the lemma.

Exercise 4.60. Show that the lemma implies the proposition in subsection 1.

Next, we apply this lemma, for each $k = 0, 1, 2, ...,$ to
- the chain map $f_k : C_k \to C_k$,

- its restriction f_k^Z to the subspace of cycles Z_k,
- its restriction f_k^B to the subspace of boundaries B_k, and
- the homology map $[f_k]$ induced by f.

Exercise 4.61. Apply the lemma to

$$U = Z_k \hookrightarrow V = C_k \xrightarrow{\partial_k} W = B_{k-1}.$$

Exercise 4.62. Apply the lemma to

$$U = B_k \hookrightarrow V = Z_k \xrightarrow{q} W = H_k,$$

where q is the quotient map.

The conclusions are:

$$\begin{array}{ll} \operatorname{Tr}(f_k) & = \operatorname{Tr}(f_k^Z) + \operatorname{Tr}(f_{k-1}^B), \\ \operatorname{Tr}(f_k^Z) & = \operatorname{Tr}([f_k]) + \operatorname{Tr}(f_k^B). \end{array}$$

Next, we use these two identities to modify the alternating sum on the left:

$$\sum_k (-1)^k \Big(\operatorname{Tr}(f_k) - \operatorname{Tr}(f_{k-1}^B) \Big) = \sum_k (-1)^k \Big(\operatorname{Tr}([f_k]) + \operatorname{Tr}(f_k^B) \Big).$$

The terms with f^B in the left-hand side and the right-hand side cancel. The result is this generalization of the Euler-Poincaré formula.

Theorem 4.63 (Hopf Trace Formula).

$$\sum_k (-1)^k \operatorname{Tr}(f_k) = \sum_k (-1)^k \operatorname{Tr}\big([f_k]\big).$$

The result is purely algebraic. In fact, the Hopf Trace Formula is similar to the Euler-Poincaré formula in the sense that its algebra is identical on the left and right but applied to two different sequences of linear operators. This observation justifies the definition of the Lefschetz number of an arbitrary sequence of linear (self-)operators $h = \{h_k : \ k = 0, 1, 2, ...\}$, as follows:

$$L(h) := \sum_k (-1)^k \operatorname{Tr}(h_k).$$

Then the trace formula becomes:

$$L(h) = L\big([h]\big),$$

for any chain map h. At this point, the latter is recognized as $\Lambda(f)$ if we choose $h := f_\Delta$.

Exercise 4.64. Show that L is a linear operator.

Now we finish the proof of the fixed point theorem.

Suppose map $f : S \to S$ has no fixed points. Since f is continuous and S is compact, there is an $\varepsilon > 0$ such that

$$d(x, f(x)) > \varepsilon$$

for all $x \in S$. Suppose $S = |K|$ for some simplicial complex K. We subdivide K enough times, p, so that $\operatorname{Mesh}(K^p) < \varepsilon$. Then

$$d(\sigma, f(\sigma)) > \operatorname{Mesh}(K^p)$$

for all $\sigma \in K^p$.

We rename K^p as K and proceed to the next step. According to the Simplicial Approximation Theorem, we can acquire a simplicial approximation $g : K^m \to K$ of f; i.e.,

$$f(N_A) \subset N_{g(A)}$$

for any vertex A. We do so for finer and finer subdivisions of K – until we have

$$\sigma \cap g(\sigma) = \emptyset$$

for all $\sigma \in K^m$. It follows that

$$sg_\Delta(\sigma) \neq \sigma,$$

where s is the subdivision chain map from the last subsection. Then,

$$L(sg_\Delta) := \sum_k (-1)^k \operatorname{Tr}(s_k g_k) = \sum_k (-1)^k 0 = 0.$$

By the Hopf Trace Formula, we have

$$\Lambda(f) := \Lambda(g_*) = L(sg_\Delta) = 0.$$

This finishes the proof of the Lefschetz's theorem. ■

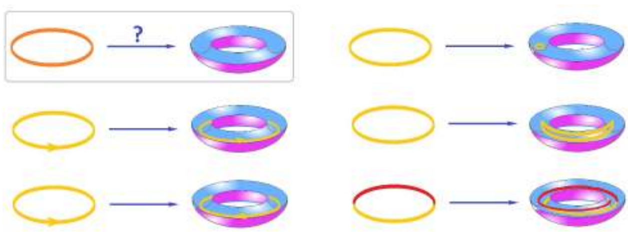
Exercise 4.65. Prove that $\Lambda(hfh^{-1}) = \Lambda(f)$ for any homeomorphism h.

4.7 The degree of a map

The degree is initially introduced for maps of the circle:

$$f : \mathbf{S}^1 \to \mathbf{S}^1,$$

as the number of times the first circle wraps around the second.



The degree is also defined for maps

$$f : \mathbf{S}^1 \to \mathbf{R}^2 \setminus \{0\},$$

as the "winding number". It is defined via the polar coordinate parametrization, $f = (r, \theta)$, measuring the change in the angle in terms of full turns, i.e.,

$$\deg f := \frac{\theta(2\pi) - \theta(0)}{2\pi}.$$

We, instead, approach the idea homologically.

Let's consider the homology map induced by our map:

$$[f_1] : H_1(\mathbf{S}^1) \to H_1(\mathbf{S}^1).$$

Since both of the groups are isomorphic to $\mathbf{Z}$, group theory tells us that such a homomorphism must be the multiplication by an integer. This integer is the degree of f.

The idea and the construction are now applied to a more general case.

Definition 4.66. Given a map between any two n-dimensional compact path-connected manifolds

$$f : M^n \to N^n,$$

the *degree* $\deg f$ of f is the integer that satisfies:

$$[f_n](x) = \deg f \cdot x,$$

where

$$[f_n] : H_n(M^n) \cong \mathbf{Z} \to H_n(N^n) \cong \mathbf{Z}.$$

is the homology map of f.

Exercise 4.67. Prove the following.

- The degree of the constant map is 0.
- The degree of the identity map is 1.
- The degree antipodal map of the n-sphere $\mathbf{S}^n \subset \mathbf{R}^{n+1}$ is $(-1)^{n+1}$.

Exercise 4.68. Any map $f : \mathbf{S}^n \to \mathbf{S}^n$ can be used as an attaching map of a cell complex. Suppose two such maps f, g of degrees k, m are used to attach two copies of $\mathbf{B}^{n+1}$ to $\mathbf{S}^n$. Compute the homology of the resulting cell complex for all k, m. Hint: start with $n = 1,\ k = 1,\ m = 2$.

The following theorems are derived from the properties of homology maps.

Theorem 4.69. The degree is multiplicative under compositions; i.e.,

$$\deg(fg) = \deg(f) \cdot \deg(g).$$

Theorem 4.70. The degree is homotopy invariant; i.e.,

$$f \simeq g \Longrightarrow \deg(f) = \deg(g).$$

Theorem 4.71. The degree of a homeomorphism, or even a homotopy equivalence, is ± 1.

Exercise 4.72. Prove the theorems.

We will accept the following without proof (see Bredon, *Topology and Geometry*, p. 187).

Theorem 4.73. The degree of a reflection of the n-sphere $\mathbf{S}^n \subset \mathbf{R}^{n+1}$ through an $(n+1)$-dimensional hyperplane containing 0, such as

$$f(x_1, x_2, ..., x_n, x_{n+1}) := (-x_1, x_2, ..., x_n, x_{n+1}),$$

is -1.

The following comes from an easy computation.

Theorem 4.74. For a map $f : \mathbf{S}^n \to \mathbf{S}^n$ of spheres, we have

$$\Lambda(f) = 1 + (-1)^{n+1} \deg f.$$

Corollary 4.75. A map $f : \mathbf{S}^n \to \mathbf{S}^n$ with $\deg f \neq (-1)^n$ has a fixed point.

It follows that a map of spheres not homotopic to a homeomorphism will have a fixed point.

The theorem below provides a partial solution of the problem of *surjectivity*.

Theorem 4.76. A map $f : \mathbf{S}^n \to \mathbf{S}^n$ of nonzero degree is onto.

Proof. If $a \in \mathbf{S}^n \setminus \operatorname{Im} f$, the image is contractible as a subset of a contractible set $\mathbf{S}^n \setminus \{a\}$, the latter fact proven via the stereographic projection. Therefore, f is homotopic to the constant map and $\deg f = 0$, a contradiction. ■

Notice that $\mathbf{T}^2 \setminus \{a\}$ isn't contractible and that's why the proof fails if applied to an arbitrary pair of manifolds. A more general result will be provided later.

Exercise 4.77. Provide details of the proof and then make both the proof and the statement of the theorem as general as possible.

Exercise 4.78. What is the converse of the theorem? Hint: don't try to prove it.

Let's recall how we used a homological argument to prove the insolvability of the *Extension Problem* for homeomorphisms of spheres. The result is generalized below.

Theorem 4.79. A map $f : \mathbf{S}^n = \partial \mathbf{B}^{n+1} \to \mathbf{S}^n$ of n-spheres of non-zero degree cannot be extended to a map on the whole ball, $F : \mathbf{B}^{n+1} \to \mathbf{S}^n$.

Proof. The fact follows from the commutativity of the following diagram:

$$\begin{array}{ccc} \mathbf{S}^n & \hookrightarrow & \mathbf{B}^{n+1} \\ & {}_{f}\searrow & \downarrow{}_{F=?} \\ & & \mathbf{S}^n \end{array}$$

■

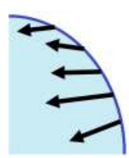
Exercise 4.80. Provide details of the proof and then make both the proof and the statement of the theorem as general as possible.

4.8 Zeros of vector fields

First, we shall see what the Brouwer Fixed Point Theorem tells us about zeros of vector fields.

Suppose a vector field V is defined on a subset $R \subset \mathbf{R}^n$ which is homeomorphic to $\mathbf{B}^m$, $m < n$. We know that V determines a map f but is it a self-map? If it is, the map will have a fixed point.

We know that we can follow the direction of $V(x)$ to define $f(x)$; that's the Euler map. We will say that V *points inside* R if for every $x \in \partial R$ there is an $\varepsilon > 0$ such that $\varepsilon V(x) \in R$.



Exercise 4.81. Prove that, in that case, there is a single $\varepsilon > 0$ such that $\varepsilon V(x) \in R$ for all $x \in R$.

Then $f(x) := \varepsilon V(x)$ is well-defined as a map $f : R \to R$. Therefore f has a fixed point, V has a zero, and, if V defines an ODE, there is a stationary point.

We can do more with the Lefschetz Fixed Point Theorem, without requiring that V points inside R.

We know that, if the Euler map $Q_t : R \to S$ of V is well-defined for some $S \subset \mathbf{R}^n$, and some $t > 0$, and R is a deformation retract of S, then we have:

$$rQ_t \simeq \mathrm{Id}_R,$$

where $r : S \to R$ is the retraction. We also know that

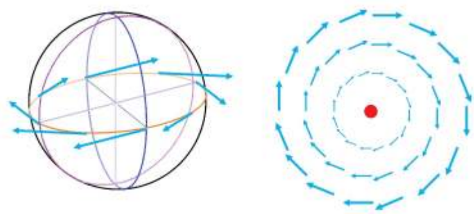
$$\Lambda(rQ_t) = \Lambda(\mathrm{Id}_R) = \chi(R).$$

Therefore, when $\chi(R) \neq 0$, map rQ_t has a fixed point by the corollary to the Lefschetz Fixed Point Theorem.

Exercise 4.82. Prove that, in this case, V has a zero.

Example 4.83. It is easy to design a non-zero vector field for the *circle*. In fact, we can have a unit vector field:

We run into problems if try this idea for a sphere. If we are to assign vectors of equal length to points of equal latitude, we realize that their lengths would have to diminish as we approach the pole to assure continuity:



□

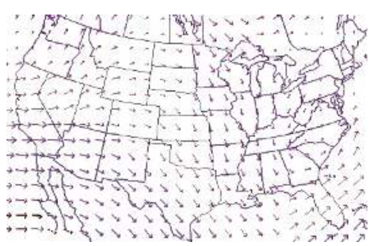
Theorem 4.84 (Hairy Ball Theorem). A continuous vector field tangent to the even-dimensional sphere has a zero.

For dimension 2, it reads: "you can't comb the hair on a coconut".

Exercise 4.85. What about the torus?

Exercise 4.86. Prove the theorem.

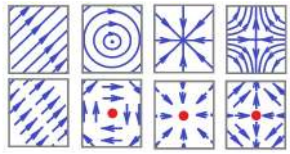
Example 4.87 (cyclone). Another commonly known interpretation of the theorem is the following. The atmosphere forms a layer around the Earth and the air moves within this layer.



Assuming that this layer is "thin", what we have is a sphere. Therefore, the vector field of the wind has a stationary point, which is a point where the wind speed is 0. That point might be a cyclone. □

Exercise 4.88. What if the layer of air is "thick"?

We would like to be able to distinguish the following behaviors of ODEs generated by vector fields:



We can consider only the behavior of the vector field V on the boundary of the neighborhood. In the 2-dimensional case, we can simply evaluate the *rotation* of V along this curve. It is 0 for the first one and 2π for the rest of them.

Exercise 4.89. Confirm these computations.

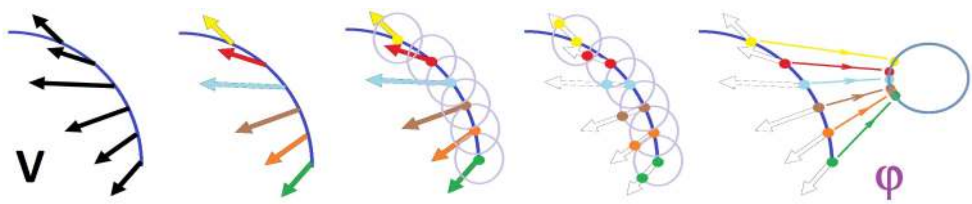
This approach to detecting and classifying zeros applies to every vector field with no zeros on the boundary. Problem solved!

What about higher dimensions? The boundary of an n-ball is an $(n-1)$-sphere, not a curve. What is the analog of the rotation then?

A convenient way to compute the rotation of a vector field $V : \mathbf{B}^2 \to \mathbf{R}^2$ with $V(u) \neq 0,\ \forall u \in \partial\mathbf{B}^2$, is to *normalize* it first:

$$\phi(u) := \frac{V(u)}{||V(u)||}.$$

The result is a new function, $\phi : \mathbf{S}^1 \to \mathbf{S}^1$:



Then the rotation of V along the boundary of the disk is simply $2\pi \cdot \deg \phi$!

Suppose D is a closed ball neighborhood in $\mathbf{R}^n$ and vector field V has no zeros on its boundary ∂D.

Definition 4.90. We define the *Gauss map* of V on D,

$$\phi_D : \partial D \to \mathbf{S}^{n-1},$$

by

$$\phi_D(z) := \frac{V(z)}{||V(z)||}.$$

Now, if V has no zeros on the whole D, the Gauss map can be extended to the whole D with the same formula. But by the last theorem from the last subsection, such an extension is only possible if $\deg \phi_D = 0$.

Definition 4.91. We define the *index of a vector field* V on D, $\text{Index}_D(V)$, to be the degree of the Gauss map:

$$\text{Index}_D(V) := \deg \phi_D.$$

Exercise 4.92. Prove that the index is independent of the choice of D.

We have the following.

Theorem 4.93. If $\text{Index}_D(V) \neq 0$ then vector field V has a zero in D.

There is a generalization that follows from the Lefschetz Fixed Point Theorem. We state it without proof (see Bredon, *Topology and Geometry*, p. 387).

Theorem 4.94 (Poincaré-Hopf Index Theorem). Suppose $Q \subset \mathbf{R}^n$ is a compact set which is also the closure of an open set. Suppose a vector field V is continuous on Q. Suppose also that V points in the outward direction along the boundary of Q. Suppose also that V has only isolated zeroes, all in the interior of Q. Then we have:

$$\sum_i \text{Index}_{D_i}(V) = \chi(Q),$$

with summation over neighborhoods of all the zeroes of V.

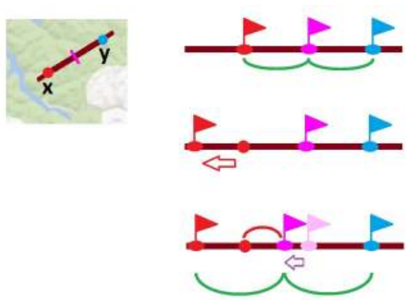
Therefore, if $\chi(S) \neq 0$, there is a stationary point.

4.9 Social choice: gaming the system

Let's recall story of two hikers that are about to go into the woods and they need to choose (or we may suggest to them) a procedure on how to decide where to set up their camp, as a fair compromise.

This procedure should be equally applicable to the future trips in case they decide to go to that forest again. The simplest answer is: take the middle point between the two locations, if possible. It is always possible when the forest is convex.

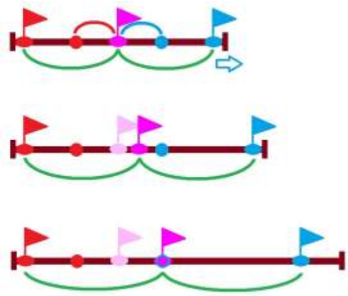
Now, what if one of the participants decides to manipulate the situation and, while his (ideal but possibly secret) preference point remains the same, declares his choice to be a point farther away from the other person's choice?



Then the mid-point, compromise location moves closer to the first person's preferred location and away from the second's!

Seeing this, the other person would move his declared choice in the opposite direction as well. They continue to respond to each other in this fashion and their search for compromise turns into a *hostile game*!

Example 4.95. To be able to fully analyze such a game, let's consider its 1-dimensional version: the choices are limited to a straight segment of a road in the forest. Suppose the players counteract each other's moves by moving backwards continuously. In that case, this game will go on until one of them reaches the end of the road (i.e., the edge of the forest). Then the other person continues to gain ground until the "compromise" location coincides with his preference or, otherwise, until he also reaches the end of the road.



The three possible outcomes share an interesting property: neither player can improve his position by a unilateral action! Such a state is called a *Nash equilibrium*.

Let's confirm this analysis with a quick computation.

Suppose the road is the interval $[0,1]$ and the preferred (ideal) locations of the two players are A_1 and A_2 respectively with $A_1 < A_2$. We assume that they will remain unchanged. In contrast, the players choose points x_1 and x_2. Either one makes his selection as he tries to decrease the distance from the preferred location to the compromise, midpoint location. In other words, A_i's goal is to

- minimize the "harm" $h_i := |A_i - \frac{1}{2}(x_1 + x_2)|$

(or to maximize some related utility). We have:

$$0 \le x_1 \le A_1 < A_2 \le x_2 \le 0.$$

If x_1 is chosen and fixed, the *response* of the second player would be to try to get the prefect hit and choose x_2 so that

$$\tfrac{1}{2}(x_1 + x_2) = A_2,$$

if possible. Then

$$x_2 = r_2(x_1) := \min\{1,\ 2A_2 - x_1\}.$$

Similarly, if x_2 is chosen and fixed, the response of the first player is

$$x_1 = r_1(x_2) := \max\{0,\ 2A_1 - x_2\}.$$

These maps are continuous!

The condition of equilibria is: $x_1 = r_1(x_2)$ and $x_2 = r_2(x_1)$. Now, we can't have both x_1 and x_2 within $(0,1)$; otherwise,

$$x_1 = 2A_1 - x_2,\ x_2 = 2A_2 - x_1,$$

which leads to a contradiction: $A_1 = A_2$. Therefore, let's assume that $x_1 := 0$. Then, $x_2 = 2A_2$ or 1, just as in the above illustration. □

We summarize this analysis with the following.

Definition 4.96. Given a pair of maps

$$r_1 : X_2 \to X_1,\ r_2 : X_1 \to X_2,$$

its *Nash equilibrium* is a fixed point of the *combined response function* $R : X_1 \times X_2 \to X_1 \times X_2$ given by

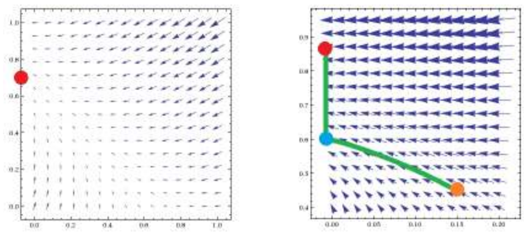
$$R(x_1, x_2) := (r_1(x_2), r_2(x_1)).$$

This function can also be defined to be

$$R := \tau(r_2 \times r_1),$$

where $\tau : X_1 \times X_2 \to X_2 \times X_1$ is the transposition: $\tau(x_1, x_2) = (x_2, x_1)$.

Example 4.97. This function can be illustrated as a vector field. In particular, for the example above with $A_1 = .15, A_2 = .45$, the vector field looks like this (left):



The path taken by the pair of players is also shown (right). □

The Brouwer Fixed Point Theorem implies the following.

Theorem 4.98. Suppose X is an acyclic polyhedron, then any pair of self-maps of X allows a Nash equilibrium.

The key to this theorem is not only the acyclicity of the space of choices but also the continuity of the response map. The continuity assumption is natural but the uniqueness of a possible response is not. For example, when the above game is played not on a segment of a road but in a forest, the response isn't unique (as one can deviate from the straight line) and the simple theorem above won't apply.

Exercise 4.99. Prove that the response is still unique provided the forest is convex.

Let's follow the example and assume that either player wants to maximize some *utility* of his own:

$$u_1 : X \times X \to \mathbf{R},\ u_2 : X \times X \to \mathbf{R}.$$

Then, in general, the maxima of these functions are attained for multiple values of x.

One way to avoid having to deal with set-valued maps (discussed in the next section) is thinking of this game as a dynamical system, as follows. Suppose X is a closed domain in $\mathbf{R}^n$ and the utilities u_1, u_2 are defined on some open neighborhood of X. Then to find a maximum, all we need to do is to *follow the gradient*, the gradient of either utility with respect to the other variable. The gradients of the utilities define a vector field V (just as illustrated above):

$$V_x = \operatorname{grad}_x u_1(x, y),\ V_y = \operatorname{grad}_y u_2(x, y).$$

Then the Poincaré-Hopf Index Theorem applied to V proves the following.

Theorem 4.100. Suppose the utilities are continuously differentiable on an open neighborhood of the compact closure of an open set, $Q \subset \mathbf{R}^n$, and suppose that the vector field V points in the outward direction along the boundary of Q. Then, provided $\chi(Q) \neq 0$, there is a point in Q critical for both utilities.

If the utilities are also concave down, this point is the maximum for both. We have an analog of a Nash equilibrium.

Exercise 4.101. What is the meaning of a *minimum* for both utilities?

Exercise 4.102. What can we say about how the game may be played around points without the concavity condition?

---

5 Set-valued maps

5.1 Coincidences of a pair of maps

The *surjectivity* question is, if $f : N \to M$ is a map, when can we guarantee that the image of f is the whole M:

$$\operatorname{Im} f = M?$$

Let's compare:

- fixed point problem: For $f : M \to M$, is there $a \in M$ with $f(a) = a$?
- surjectivity problem: For $f : N \to M$, given $b \in M$, is there an $a \in N$ with $f(a) = b$?

The two problems are united into the following.

Coincidence Problem: Given two maps $f, g : N \to M$, do they have a *coincidence*, i.e.,

$$a \in N : f(a) = g(a) = b \in M?$$

First, setting $N := M, g := \mathrm{Id}_M$ gives us the fixed point problem.

Second, setting $g(x) := b$, a constant map, gives us the surjectivity problem, if solved for all $b \in M$. Since all constant maps are homotopic (whenever M is path-connected), the positive answer – in our homological setting – for one value of b will give us all.

We are motivated by several related problems.

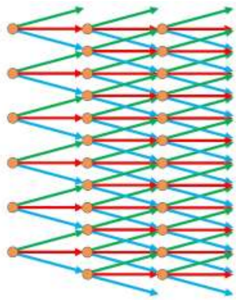
First, the surjectivity results in the last section rely on the degrees of the maps. Therefore, they apply only to a narrow class of maps, those between manifolds of the same dimension. The forward kinematic map of a control system (discussed below) is a map from a high-dimensional domain space N, with the dimension commonly equal to the number of degrees of freedom of the robot, to the 3-dimensional operating space.

More generally, let's consider *discrete dynamical systems*. In such a system, the next position, or *state*, $f(x)$, with $f : M \to M$ just a map, depends only on the current state, $x \in M$. What if there is a parameter? Suppose the next position $f(u, x)$ depends not only on the current one, $x \in M$, but also on the input, or *control*, $u \in U$. Then our map becomes:

$$f : N = U \times M \to M.$$

Such a parametrized dynamical system is called a *control system*.

Below, the control system has three inputs:



As an example, for a cruise control system, M is the space of all possible values of the car's speed and U is the engine's possible throttle positions as they determine how much power the engine generates. Then an equilibrium of such a system is a combination of a state $a \in M$ and a control $p \in U$ such that the state remains fixed:

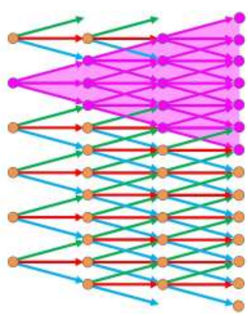
$$f(p, a) = a.$$

This point $(p, a) \in N$ is then a coincidence point of this map f and the projection $g : N = U \times M \to M$!

Exercise 5.1. Describe the state-control space of the tether-ball:

Meanwhile, the surjectivity of the map f ensures what is called the "reachability" of the control system.

What if we ignore how a particular control leads to a particular state? Then each current state will lead to numerous future states:



It is given then by a *set-valued map* $F : M \to M$; i.e., its values are subsets of M: $F(x) \subset M$. Such a map will also appear when, considering noise and other uncertainty, a discrete dynamical system is only known imprecisely.

5.2 Set-valued maps

For two given spaces X, Y, a *set-valued map* $F : X \to Y$ is a correspondence

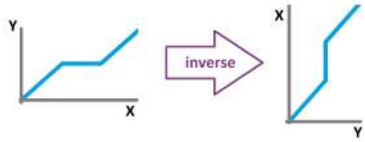
$$x \mapsto F(x) \subset Y.$$

It can also be thought of as a function to the power set of Y:

$$F : X \to 2^Y.$$

There are many subtle examples of the emergence of set-valued maps.

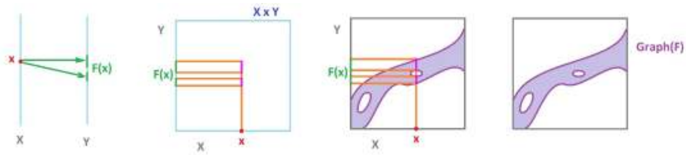
The *inverse* $f^{-1} : Y \to X$ of a function $f : X \to Y$ doesn't, of course, have to be a function.



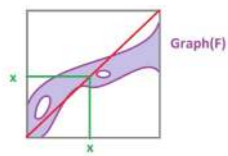
The inverse can be seen as a set-valued map $f^{-1} : Y \to X$ and, if f is continuous, this map will also have a certain continuity property. Since its graph is the *same* as that of f itself, it is, therefore, closed. It may have empty values when f isn't surjective.

In fact, any set-valued map $F : X \to Y$ can be understood entirely via its graph

$$\text{Graph}\, F = \{(x, y) :\ x \in X, y \in Y, y \in F(x)\} \subset X \times Y.$$

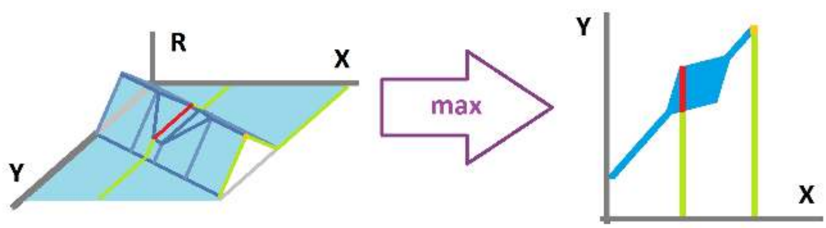


The *fixed point problem* is restated for a set-valued map $F : X \to X$: is there an $x \in X$ with $x \in F(x)$? Such a point would have to lie on the diagonal $\Delta(X)$ of $X \times X$:



Another area where set-values maps naturally appear is *optimization*. For any function of two variables $f : X \times Y \to \mathbf{R}$, one can define a set-valued map $f^{max} : X \to Y$ by

$$f^{max}(x) := \arg\max_{y \in Y} f(x, y).$$



Since the maximum value may be attained for several values of y, the value of f^{max} isn't necessarily a single point. This set may even be empty.

Proposition 5.2. Suppose Y is compact and $f : X \times Y \to \mathbf{R}$ is continuous. Then the graph of f^{max}, Graph f^{max}, is closed and its values, $f^{max}(x)$, are non-empty.

Exercise 5.3. (a) Prove the proposition. (b) State and prove the analog of the proposition with $\mathbf{R}$ replaced by a metric space.

A function $f : X \to \mathbf{R}$, with X a convex subset of $\mathbf{R}^N$, is called *convex* (or concave down) if

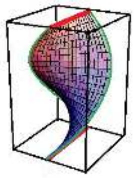
$$f(tx_1 + (1 - t)x_2) \geq tf(x_1) + (1 - t)f(x_2),\ \forall t \in [0, 1], x_1, x_2 \in X.$$

Exercise 5.4. Prove that any convex function is continuous on an open set.

Proposition 5.5. If $f(x_0, \cdot)$ is convex for each $x_0 \in X$, then f^{max} has convex values.

Exercise 5.6. Prove the proposition.

There are set-valued maps with properties impossible for their single-valued counterparts. For example, if we identify the top and bottom of the surface shown below, the resulting Möbius band will serve as the graph, which is closed, of a "set-valued retraction" of the disk to the circle:



This surface is illustrated by the DNA's double helix and the famous double staircase in the Vatican.

Exercise 5.7. Provide a formula for this map. Examine preimages of open sets under this map.

Even though the set-valued approach provides a very productive point of view, we would prefer a direct way to apply all of the topological machinery we have developed for single-valued maps. We interpret each set-valued map in terms of two single-valued maps, as follows.

Suppose N is the *graph of the set-valued map* $F : X \to Y$,

$$N := \operatorname{Graph} F,$$

and f, g are the two *projections of the graph*

$$f : N \to X,\ g : N \to Y.$$

When $X = Y$, the fixed point problem for F, once again, becomes a coincidence problem for f, g:

$$a \in F(a) \Longleftrightarrow b = (a, a) \in N \Longleftrightarrow f(b) = g(b) = a.$$

We can treat all set-valued maps as map of pairs. The *composition* $GF : X \to Z$ *of two set-values maps* $F : X \to Y,\ G : Y \to Z$, is defined by simply letting

$$GF(x) := G(F(x)).$$

This expression can now be interpreted in terms of the projections:

$$\begin{array}{ccccc}
\operatorname{Graph} GF & \longrightarrow & \operatorname{Graph} G & \xrightarrow{Q_Z} & Z \\
\downarrow & & \downarrow Q_Y & \nearrow_G & \\
\operatorname{Graph} F & \xrightarrow{P_Y} & Y & & \\
\downarrow P_X & \nearrow_F & & & \\
X & & & &
\end{array}$$

Exercise 5.8. Define the inverse of a set-valued map and interpret it in terms of map pairs.

Next, how do we define the homology map $F_* : H(X) \to H(Y)$ of a set-valued map $F : X \to Y$ between two polyhedra? Suppose $f : N \to X,\ g : N \to Y$ are the projections. Then F_* could be defined in terms of the homology map of this composition:

$$F_* := (fg^{-1})_* : H(X) \to H(Y),$$

but only if the inverse exists! It does, for example, when one of the projections is a homeomorphism, but in that case our set-valued map is single-valued anyway. Or, the homology map of the pair f, g, and that of F, could be defined via their homology maps:

$$F_* := f_*(g_*)^{-1} : H(X) \to H(Y),$$

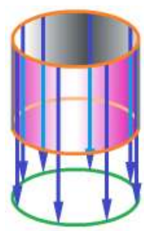
but only if the inverse exists! The theorem below gives us a sufficient condition when this is the case.

5.3 The Vietoris Mapping Theorem

Let's compare the homology maps of these two examples of projections.

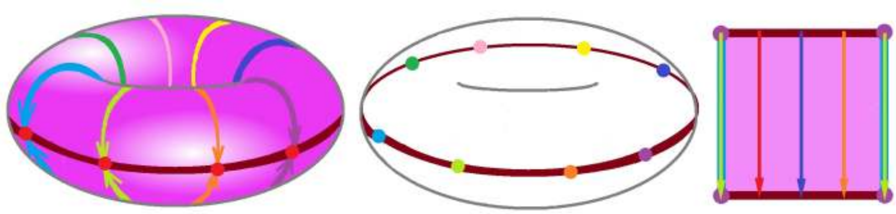
Example 5.9. First, the projection of the cylinder on the circle:

$$P : \mathbf{C}^2 = [0,1] \times \mathbf{S}^1 \to \mathbf{S}^1.$$



Second, the projection of the torus on the circle:

$$Q : \mathbf{T}^2 = \mathbf{S}^1 \times \mathbf{S}^1 \to \mathbf{S}^1.$$



The homology map of the first,

$$[P_1] : H_1(\mathbf{C}^2) = \mathbf{R} \to H_1(\mathbf{S}^1) = \mathbf{R},$$

is an isomorphism and that of the second,

$$[Q_1] : H_1(\mathbf{T}^2) = \mathbf{R}^2 \to H_1(\mathbf{S}^1) = \mathbf{R},$$

is another projection.

Why the difference?

Clearly, Q collapses the other circle to a point and, as one would expect, $[Q_1]$ "kills" its homology class. The effect of P on the homology classes of the cylinder is nil.

What are the signs that could help us predict such as outcome without actually computing the homology maps?

All we need to examine is the *preimages of points* (aka "fibers") of the two maps.

- For P, the preimages are segments. Therefore, they are homologically trivial, acyclic.
- For Q, they are circles and, therefore, not acyclic.

In the latter case, these cycles collapse to points and, therefore, their homology classes are lost. It remains to be shown that in the former case, there can be no such collapsing. □

The following important theorem is the answer.

Theorem 5.10 (Vietoris Mapping Theorem). Let $h : K \to L$ be a surjective simplicial map. Suppose that the preimages $h^{-1}(b)$ of simplices in L are acyclic subcomplexes of K. Then the homology map

$$h_* : H(K) \xrightarrow{\cong} H(L)$$

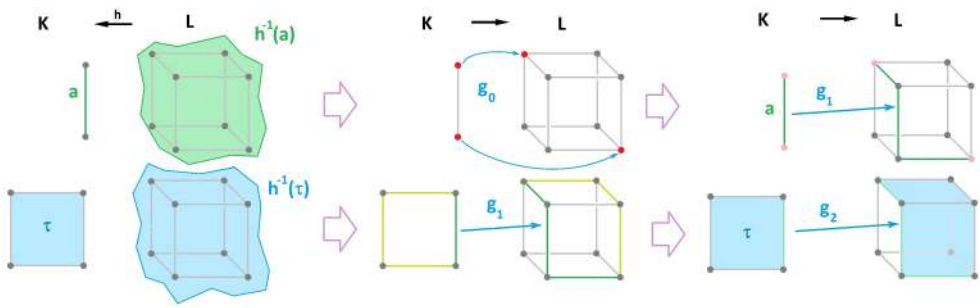
of h is an isomorphism.

The Chain Extension Theorem is restated for convenience below.

Lemma 5.11. Suppose K and L are simplicial complexes and Q_a is an acyclic subcomplex of L for every cell a in K. (1) Suppose $g_0 : K^{(0)} \to L^{(0)}$ is a map of vertices that satisfies the following condition:

$$a = A_0A_1...A_n \in K \Longrightarrow g_0(A_0), g_0(A_1), ..., g_0(A_n) \in Q_a.$$

Then there is such a chain map (an extension of g_0) $g = \{g_i\} : C(K) \to C(L)$ that $g(a) \subset Q_a$ for every cell a in K. (2) If two chain maps q and q' satisfy this condition, $q(a), q'(a) \subset Q_a, a \in K$, then q and q' are chain homotopic, $q \simeq q'$.



Proof. We construct $(h_*)^{-1}$ – without h^{-1}! We apply part (1) of the lemma to the pair:

$$g_0, \quad Q_a := h^{-1}(a).$$

Then g_0 has a chain extension, a chain map $g : C(K) \to C(L)$, with $g(a) \subset h^{-1}(a)$ for every cell a in K.

Next, we observe that $h_\Delta(g(a)) = a$. It follows that $h_\Delta g = \mathrm{Id}_{C(K)}$ and, furthermore, $h_*g_* = \mathrm{Id}_{H(K)}$.

Next, we apply part (2) of the lemma to:

$$Q_a := h^{-1}h(a), \quad q := gh_\Delta, \ q' := \mathrm{Id}_{C(L)}\,.$$

Then these two chain maps are chain homotopic: $gh_\Delta \simeq \mathrm{Id}_{C(L)}$. Therefore, $g_*h_* = \mathrm{Id}_{H(L)}$.

We have shown that h_* is invertible, which concludes the proof of the theorem. ■

Exercise 5.12. Prove that $h_\Delta(q(a)) = a$.

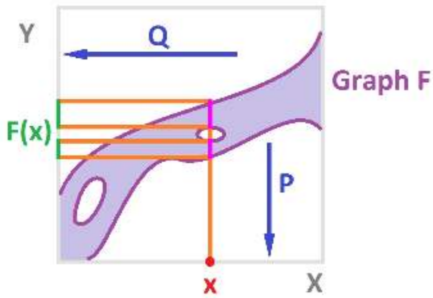
Exercise 5.13. State the up-to-homotopy version of this theorem. Is its converse true? Hint: think of degrees.

A stronger version of this theorem that doesn't rely on the simplicial structure is given without proof (see Spanier, *Algebraic Topology*, p. 344).

Theorem 5.14. The conclusion of the Vietoris Mapping Theorem holds for any map whose preimages (of points) are acyclic.

The ability to "reverse" a homology map is found useful when one studies set-valued maps.

Let $G := \text{Graph}F$ be the graph of a set-valued map $F : X \to Y$. It is a subset of $X \times Y$. Then there are two projections $P : G \to X$ and $Q : G \to Y$. These projections are continuous and, therefore, their homology maps P_*, Q_* are well-defined – provided all spaces involved, including the graph G, are polyhedra. Suppose F is *acyclic-valued*: $F(x)$ is an acyclic subset of Y for each $x \in X$.



Then the preimages of P, $P^{-1}(x) = \{x\} \times F(x)$, are also acyclic. By the theorem, P generates an isomorphism:

$$P_* : H(G) \xrightarrow{\cong} H(X).$$

Definition 5.15. Given a set-valued map $F : X \to Y$ between polyhedra, with acyclic images and polyhedral graph, the *homology map* $F_* : H(X) \to H(Y)$ of F is defined to be

$$F_* := Q_*(P_*)^{-1}.$$

So, the homology operator of such a *set*-valued map is *single*-valued!

Exercise 5.16. What happens to their homology maps under compositions of set-valued maps?

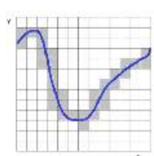
With the help of this construction, we can now reinterpret the *homology of maps*. Recall that we defined the homology map of a general, not necessarily cell map, $f : X \to Y$, with X, Y polyhedra, via a simplicial approximation of f and subdivisions of the domain X. What if, instead, we replace map f with a set-valued map? Specifically, what if the new map has whole cells as values? In other words, given a map $f : X \to Y$, define a set-valued map $F : X \to Y$ by setting $F(x)$ equal to the carrier of $f(x)$,

$$F(x) := \text{carr}(f(x)).$$

the cell to the interior of which $f(x)$ belongs. Then

$$f(x) \in F(x), \ \forall x \in X.$$

This is what the graph of a set-valued "approximation" F of f might look like:



Finally, we let $f_* := F_*$.

A more general case is that of a pair of arbitrary maps $P : G \to X$ and $Q : G \to Y$, where X, Y, G are polyhedra. Yet we still can define a set-valued map $F : X \to Y$ by $F(x) := QP^{-1}$. Suppose one of them, P, has acyclic preimages. Then this pair of maps,

$$X \xleftarrow{\ P\ } G \xrightarrow{\ Q\ } Y,$$

produces a pair of linear operators:

$$H(X) \xleftarrow{\ P_* \cong\ } H(G) \xrightarrow{\ Q_*\ } H(Y).$$

But unlike the former diagram, in the latter, the first arrow is reversible!

Definition 5.17. The *homology map of the pair* is defined as

$$Q_* P_*^{-1} : H(X) \to H(Y).$$

5.4 The Lefschetz number of a pair of maps

Now we turn to the study of coincidences of maps or, which is the same thing, fixed points of set-valued maps.

The homology map of the pair f, g – when one of them, g, has acyclic preimages – is defined as

$$f_* g_*^{-1} : H(M) \to H(M).$$

Then, we can define the *Lefschetz number of the pair of maps* as the Lefschetz number of this homomorphism, via the same trace formula:

$$\Lambda(f, g) := L(f_* g_*^{-1}) = \sum_n (-1)^n \mathrm{Tr}([f_n][g_n]^{-1}).$$

The two theorems below are stated without proof (see Saveliev, *Lefschetz coincidence theory for maps between spaces of different dimensions*, Topology and Its Applications, 116 (2001) 1, 137-152).

Theorem 5.18 (Lefschetz Coincidence Theorem). Suppose M and N are polyhedra and $f, g : N \to M$ are maps. Suppose also that g has acyclic preimages. Then, if

$$\Lambda(f, g) \neq 0,$$

then the pair has a coincidence.

The proof of the following important result relies on the Brouwer Fixed Point Theorem.

Corollary 5.19 (Kakutani Fixed Point Theorem). Let X be a non-empty, compact and convex subset of some Euclidean space. Let $F : X \to X$ be a set-valued map with a polyhedral graph and each $F(x)$ non-empty and convex for all $x \in X$. Then F has a fixed point: $a \in F(a)$.

Exercise 5.20. Show that this result generalizes the Brouwer Fixed Point Theorem.

Exercise 5.21. Derive the Kakutani Fixed Point Theorem from the Lefschetz Coincidence Theorem.

Another version of the Lefschetz Coincidence Theorem does not require acyclic preimages. We state only its corollary: the following *homological* solution to the surjectivity problem.

Theorem 5.22 (Surjectivity). Suppose N is a polyhedron, M is an orientable compact path-connected manifold, $\dim M = n$, and $f : N \to M$ is a map. If

$$[f_n] : H_n(N) \to H_n(M)$$

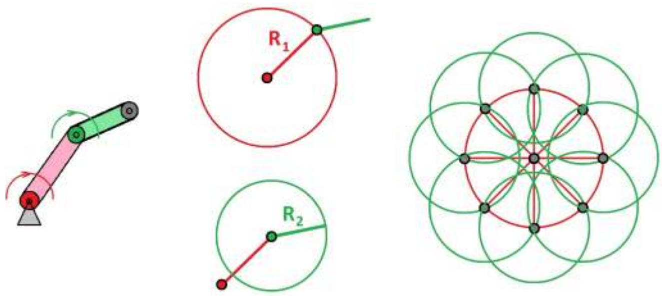
is nonzero, then every map homotopic to f is onto.

Exercise 5.23. Show that this result generalizes the Surjectivity Theorem for maps of non-zero degree.

5.5 Motion planning in robotics

Example 5.24 (two-joint arm). Recall that the configuration space of a two-joint robotic arm is the torus:

$$\mathbf{T}^2 = \mathbf{S}^1 \times \mathbf{S}^1.$$



Meanwhile, its operational space is parametrized by the locations of the rod's end:

$$x = R_1 \cos\phi_1 + R_2 \cos\phi_2,\ y = 0,\ z = R_1 \sin\phi_1 + R_2 \sin\phi_2.$$

Under the assumption $R_1 > R_2$, the operational space is the annulus. Since these functions are periodic, they can be seen as a single map from the torus to the annulus. □

It is an example of the *forward kinematics map* of a robot, i.e., the map from its configuration space to the operational space that associates a location of the end of the arm to each configuration.

Exercise 5.25. A telescoping arm with a joint has the cylinder, $[0,1] \times \mathbf{S}^1$, as its configuration space. Find its forward kinematics map.

The forward kinematics map is used for motion planning. Especially when the topology of the operation space is non-trivial, looking at the homology map of this map is a way to learn about the possible solutions.

Example 5.26. In the above example of a two-joint arm, the forward kinematics map is a map

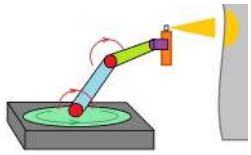
$$f : \mathbf{T}^2 \to A,$$

where A is the annulus. Whatever the exact map is, its homology map

$$f_1 : H_1(\mathbf{T}^2) = \mathbf{R} \times \mathbf{R} \to H_1(A) = \mathbf{R}$$

can be easily understood. It is the projection. Indeed, either of the generators of the 1st homology group $H_1(\mathbf{T}^2) = \mathbf{R}^2$ is the cycle that represents a single turn of the joint. Yet, only the first one is mapped to the generator of $H_1(A) = \mathbf{R}$. The image of the second generator doesn't go around the annulus and, therefore, its homology class collapses under f. This wouldn't be the case if $R_1 < R_2$, but then the homology of the operational space is trivial anyway. □

One can use this information to try to match the required operational space with the configuration possible space or vice versa. For example, one may ask whether a given robotic arm design is appropriate to spray paint on a topologically complex surface S:



The maps to consider are of this form:

$$f : \mathbf{S}^1 \times ... \times \mathbf{S}^1 \times \mathbf{S}^2 \to S.$$

The analysis of the possible homology maps between the homology groups of these manifolds will help to rule out some designs. As we want to be able to paint the whole surface, the question becomes, *does such a map have to be onto*?

Example 5.27. In the case of a two-joint arm, the map is

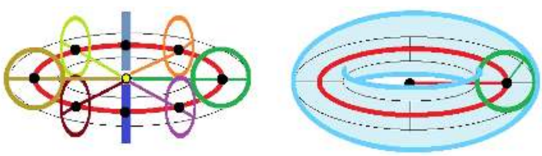
$$f : \mathbf{T}^2 \to S.$$

If, for example, $S = \mathbf{T}^2$, then, of course, we can choose the identity map and this map is onto. But what if our knowledge of the robot is imperfect? Then we need to deal with the *perturbations* of this map. Are all maps homotopic to it also surjective? The answer is Yes. □

Exercise 5.28. Prove that all maps homotopic to the identity map of the torus are surjective.

Example 5.29. A more typical object to paint is homeomorphic to the sphere $S = \mathbf{S}^2$. Then the answer is No, because *every* map $\mathbf{T}^2 \to \mathbf{S}^2$ is homotopic to the constant map! □

Example 5.30. Here is the configuration space of a 3d two-joint arm:



It has the two rods (first red, second green) perpendicular to each other. Then the configuration space and the operational space are homeomorphic via the obvious forward kinematics map. Its homology map is the identity. □

For more precise tasks, such as drawing or cutting, one may need the end of the arm to follow a particular path in the operational space. An example is the task of drawing a specified curve on a surface. Then one needs to find the values of the configuration parameters that will make this happen. This is called *the inverse kinematic problem.*

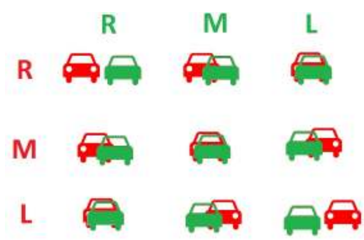
Specifically, given a continuous path $p : [0,1] \to A$, in the operational space, find a path $q : [0,1] \to C$, in the configuration space, such that $p = fq$, where f is the forward kinematic map. If such a solution exists, we say that the problem is "well-posed".

If, in addition, we assume that these paths are loops, the existence of a solution for this problem can studied via the maps of the fundamental group induced by f (or, indirectly, via the 1-homology classes of these loops).

Exercise 5.31. Consider the inverse kinematic problem for drawing a curve on the torus that may be (a) a meridian, (b) longitude, or (c) the diagonal. What if the robot has many joints but only one rotational joint?

5.6 Social choice: standstill in hostilities

Example 5.32 (the game of chicken). Let's imagine a road and two cars approaching from the opposite directions. Both have a continuum of strategies: from keeping all the way to the left to keeping all the way to the right. The goal of either of the two drivers is to minimize the damage and each strategy will produce a different outcome relative to this goal. There are two Nash equilibria: both keep right or both keep left.



One can see that the response maps – provided the road is wide enough – are *set-valued.* □

We generalize the setting for a hostile game from two to multiple players.

Suppose we have $m > 1$ players each with a *space of choices*:

$$X_i,\ i = 1, 2, ..., m.$$

The choice of the ith player is determined by the choices of the rest of the players. The ith *response map* is a (possibly set-valued) map

$$r_i : X_1 \times ... \times [X_i] \times ... \times X_m \to X_i,$$

with the bracket indicating that the ith item is excluded. The *combined response function* of the game is a (possibly set-valued) map

$$R : X_1 \times ... \times X_m \to X_1 \times ... \times X_m$$

given by

$$R(x_1, ..., x_m) := \Big(r_1([x_1], x_2, ..., x_m), ..., r_m(x_1, ..., x_{m-1}, [x_m]) \Big).$$

A *Nash equilibrium* of the game is a fixed point of the combined response function R.

We already know that the Brouwer Fixed Point Theorem implies that, for X an acyclic polyhedron, any pair of self-maps of X allows a Nash equilibrium. For the case when the response isn't unique, we have to apply the Kakutani Fixed Point Theorem instead.

Theorem 5.33. Suppose the choice spaces $X_i,\ i = 1, 2, ..., m$, are acyclic polyhedra and the response maps have polyhedral graphs and non-empty convex values. Then the game has a Nash equilibrium.

For an instance of this theorem, suppose that each player tries to maximize some *utility* of his own:

$$u_i : X_1 \times ... \times X_m \to \mathbf{R},\ i = 1, 2, ...m.$$

Then the response functions of this game are set-valued:

$$r_i(y) = \arg \max_{x_i \in X_i} u_i(x_1, ..., x_m).$$

Theorem 5.34. Suppose X is an convex polyhedron and the utilities are convex. Then the game above allows a Nash equilibrium.

Exercise 5.35. Prove the theorem.

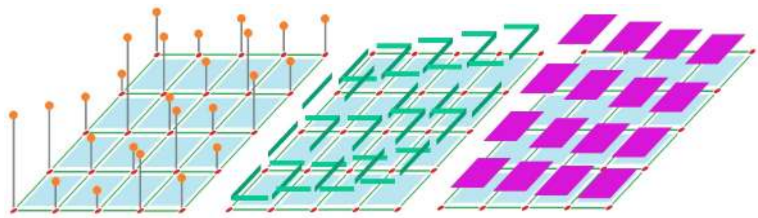
Exercise 5.36. What are the utilities of the two drivers in the above example of the game of chicken? Find its Nash equilibria. How does the game change as the road becomes wider, narrower? What if the road is a cylinder?

Exercise 5.37. Does haggling over a price have a Nash equilibrium?

The space of choices may be a simplicial complex with every simplex seen as the probability simplex. In that case, the choices are the mixed strategies.

Chapter VI

Forms



1 Discrete differential forms and cochains

1.1 Calculus and differential forms

Suppose I is a closed interval. In the analysis context, the definite integral – with the interval of integration fixed – is often thought of as a real-valued *function of the integrand*. This idea is revealed in the usual function notation:

$$G(f) := \int_I f(x)dx \in \mathbf{R}.$$

This point of view is understandable: after all, the Riemann integral is introduced in calculus as the limit of the Riemann sums of f. The student then discovers that this function is *linear*:

$$G(sf + tg) = sG(f) + tG(g),$$

with $s, t \in \mathbf{R}$. However, this notation might obscure another important property of integral, the *additivity*:

$$\int_{[a,b]\cup[c,d]} f(x)dx = \int_{[a,b]} f(x)dx + \int_{[c,d]} f(x)dx,$$

for $a < b \leq c < d$. We then realize that we can also look at the integral as a *function of the interval* – with the integrand fixed – as follows:

$$H(I) := \int_I f(x)dx.$$

In higher dimensions, the intervals are replaced with surfaces and solids while the expression $f(x)dx$ is replaced with $f(x,y)dxdy$ and $f(x,y,z)dxdyz$, etc. These "expressions" are called *differential forms* and each of them determines such a new function. That's why we further modify the notation as follows:

$$\omega(I)=\int_I \omega.$$

This is an indirect definition of a differential form of dimension 1 – it is a function of intervals. Moreover, it is a function of 1-*chains* such as $[a,b]+[c,d]$. We can see this idea in the new form of the additivity property:

$$\omega(I+J)=\omega(I)+\omega(J).$$

We recognize this function as a 1-*cochain*!

In light of this approach, let's take a look at the integral theorems of vector calculus. There are many of them and, with at least one for each dimension, maybe too many...

Let's proceed from dimension 3, look at the formulas, and see what they have in common.

Green's Theorem:

$$\iint_S \left(\frac{\partial q}{\partial x}-\frac{\partial p}{\partial y}\right) dA=\int_{\partial S} pdx+qdy.$$

Here, the integrals' domains are a solid and its boundary surface respectively.

Gauss' Theorem:

$$\iiint_R \operatorname{div} F dV=\iint_{\partial R} F\cdot N dA.$$

The domains of integration are a plane region and its boundary curve.

Fundamental Theorem of Calculus:

$$\int_{[a,b]} F'dx=F\Big|_a^b.$$

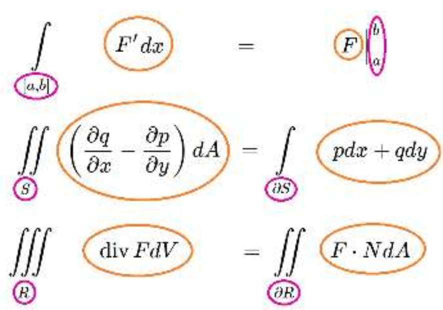
In the left-hand side, the integrand is $F'dx$. We think of the right-hand side as an integral too: the integrand is F. Then the domains of integration are a segment and its two endpoints:

$$[a,b] \text{ and } \{a,b\}=\partial[a,b].$$

What do these three have in common?

Setting aside possible connections between the integrands, the pattern of the *domains of integration* is clear. The relation is the same in all these formulas:

a region on the left and its boundary on the right.



Now, there *must* be some kind of a relation for the integrands too. The Fundamental Theorem of Calculus suggests a possible answer:

a function on the right and its derivative is on the left.

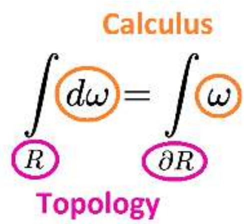
Clearly, for the other two theorems, this simple relation can't possibly apply. We can, however, make sense of this relation if we treat those integrands as differential forms. Then the form on the left is what we call the *exterior derivative* of the form on the right. Consequently, the theorem can be turned into a definition of this new form.

Thus, we have just *one* general formula which includes all three (and many more):

Stokes Theorem:

$$\int_R d\omega = \int_{\partial R} \omega.$$

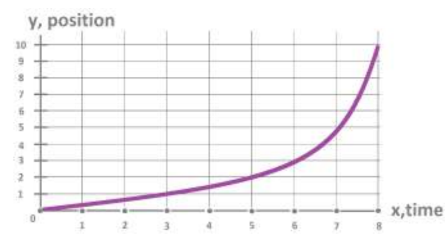
The relation between R and ∂R is a matter of *topology*. The relation between $d\omega$ and ω is a matter of *calculus*, the calculus of differential forms:



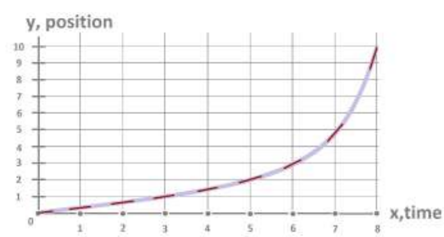
Furthermore, as we shall see, the transition from topology to calculus is just *algebra*!

1.2 How do we approximate continuous processes?

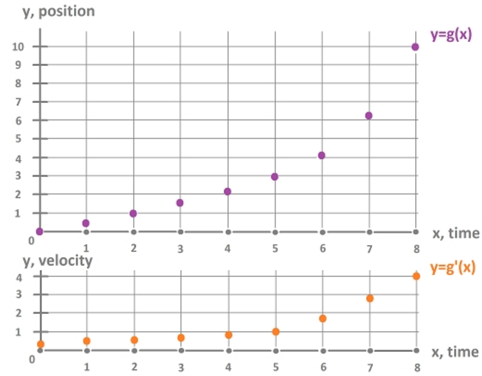
Suppose we have a function, f, representing the position, y, of your car as a function of time, x:



Let's suppose, however, that we know only a "sampled" version of this function; i.e., we only know its values at, say, $0, 1, 2, 4, 5, \ldots$. These numbers may come from looking occasionally at your odometer. Suppose also that the derivative, f', of this function – representing the velocity of your car – is also sampled:



You get those numbers from looking at your speedometer once in a while. Thus, while f and f' are unknown, we have instead two new functions: g and g', defined only at predetermined points of time:



We may call them *discrete functions*.

What do these two have to do with each other? Now, it would make sense if g' was the derivative of g in some sense. But with all the functions being discrete, there is no differentiation or the derivative in the conventional sense. Can we ever establish such a relation between two discrete functions?

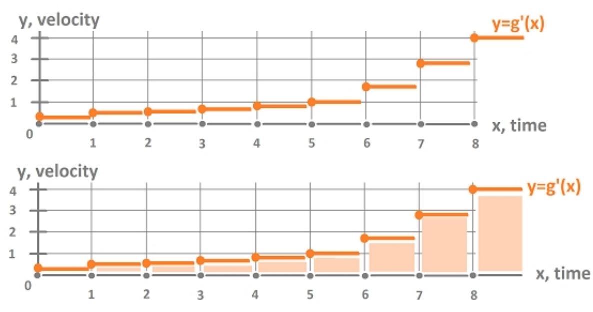
The integral provides may be the answer. The *Fundamental Theorem of Calculus* states:

$$f(8) - f(0) = \int_0^8 f'(x)dx.$$

Now, we would like see such a formula to link the two new functions together:

$$g(8) - g(0) \stackrel{?}{=} \int_0^8 g'(x)dx.$$

We rely on the conventional understanding of the Riemann integral as the area under the graph and that's why we turn g' into a *step-function*:



The result is as follows. The right-hand side is:

$$g(8) - g(0) = 10,$$

but the left-hand side is:

$$\int_0^1 g'(x)dx \approx 0.2 + 0.4 + 0.5 + 0.7 + 0.8 + 1 + 1.7 + 2.9 + 4 = 12.2.$$

Then the Fundamental Theorem of Calculus fails and, with it, one of *the laws of physics*!

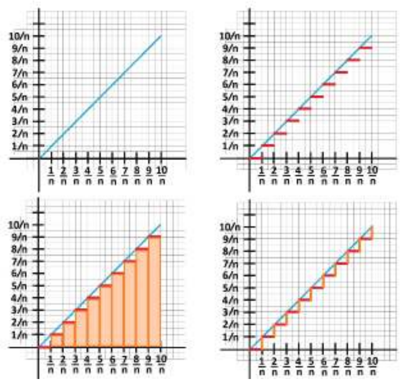
Of course, we can improve the sampling by taking more sample values (i.e., we look at the dials more often). Then the approximation improves too and, mathematically, the Riemann sums above converge to the Riemann integral, the actual displacement.

But what about the convergence of the approximations of the *velocity*? Suppose $f_n \to f$, does it mean that $f_n' \to f'$? We know that the answer is No. Just consider the famous Weierstrass function:

$$f(x) := \sum_{n=0}^{\infty} a^n cos(b^n \pi x),\ 0 < a < 1.$$

This series converges uniformly, hence f is continuous. But for $b > 1 + \frac{a}{2}\pi$, the limit function f is nowhere-differentiable.

Example 1.1. A related example is as follows. From calculus, we know how to use step-functions to approximate the definite integral of a function.



For example, for $f(x) = x,\ x \in [0,1]$, the (left-end) Riemann sum is

$$L_n := \sum_{i=0}^{n-1} \frac{i}{n}\frac{1}{n} = \frac{1}{n^2}\sum_{i=0}^{n-1} i = \frac{1}{n^2}\frac{n(n-1)}{2} \to \frac{1}{2},$$

as expected. In other words, the areas under the graphs of the approximating step-functions converge to the *area* under the graph of f. In contrast, an attempt to use the graphs of these functions to approximate the *length* of the graph fails, manifestly:

$$L := \sum_{i=0}^{n-1} \left(\frac{1}{n} + \frac{1}{n}\right) = \frac{2}{n}\sum_{i=0}^{n-1} 1 = \frac{2}{n}n = 2.$$ □

Exercise 1.2. Use other step-functions to approximate the length of the diagonal segment.

Exercise 1.3. Use step-functions to approximate the area and the length of the circle.

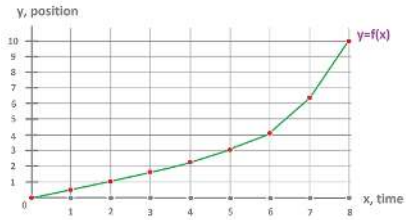
To summarize:
- No specific approximation satisfies the laws of calculus (or the laws of physics).
- Improving the approximation won't necessarily bring us closer to the process being approximated.

1.3 Cochains as building blocks of discrete calculus

We will introduce discrete structures that don't approximate but rather *mimic* the behavior of the continuous structures of calculus.

We already have a substitute for a function f – a discrete function g defined at predetermined points of the real line. What is its derivative?

We consider, instead of the tangent lines of f, the secant lines of f and, therefore, of g:



Then the values of g' are the slopes on the secant lines over these intervals; e.g., g' takes the values $g(1) - g(0),\ g(2) - g(1)$, etc. This step may be understood as "differentiation" of g. To confirm that this makes sense, we would like to "integrate" g' and recover g.

The key question is: *where* should g' be defined? For interval $[0, 1]$, which one is correct:

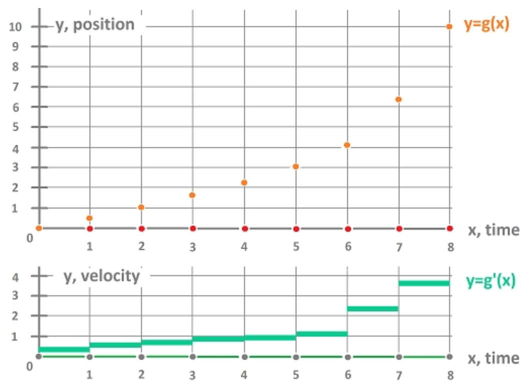
$$g'(0) = g(1) - g(0) \text{ or } g'(1) = g(1) - g(0)?$$

Neither!

We would like to think of the integral in the usual sense – as the area under the graph. That's why, instead, we prefer to assign this value to the points of the interval itself:

$$g'(x) = g(1) - g(0) \text{ for } x \in (0, 1),$$

This makes g' a step-function:



That we can integrate. For example:

$$g(1) - g(0) = \int_0^1 g'(x)dx,$$

and the results match! And so on. They all match on every interval $(m, m+1)$. Therefore, they match on the whole interval:

$$g(n) - g(0) = \int_0^n g'(x)dx.$$

In other words, the Fundamental Theorem of Calculus holds.

Exercise 1.4. Prove the last statement.

We should simplify things even further. The domain of the "derivative" g' is understood as the set

$$(0, 1) \cup (1, 2) \cup (2, 3) \cup \ \ldots.$$

We are able skip the values $g'(0) =?,\ g'(1) =?,\ \ldots$, because they don't affect the integrals. Now, it is more convenient to think of the domain of this function as the *collection* of intervals (closed or open) rather than their union. For example,

$$g'\Big|_{(0,1)} = .3$$

is replaced with

$$g'\Big([0,1]\Big) = .3.$$

So, g' is defined on the 1-cells. Furthermore, the integral becomes the sum:

$$\int_0^n g'(x)dx = g'\Big([0,1]\Big) + ... + g'\Big([n-1,n]\Big) = g'\Big([0,n]\Big).$$

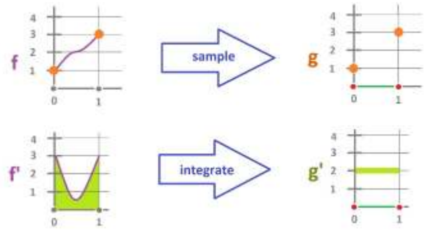
It is a function of *chains*.

The bonus insight is that g and g' are revealed to be of the same nature: they *both* are functions of chains, just of different dimensions:

	g	g'
domain:	$\{...,0,1,2,3,...\}$	$\{...,[0,1],[1,2],[2,3],...\}$
inputs:	points	closed intervals
dimension:	0	1

Accordingly, these two functions are defined on 0- and 1-cells respectively. Consequently, both can be extended linearly to functions of 0- and 1-chains. We recognize them as a 0-cochain and a 1-cochain! In the calculus context, we will call them *discrete differential forms of degree* 0 and 1, or simply 0- and 1-forms.

Thus, we acquire g by sampling f and then acquire g' by taking the differences of the values of g.



Alternatively, we can find g' independent of g by integrating f' on these intervals:

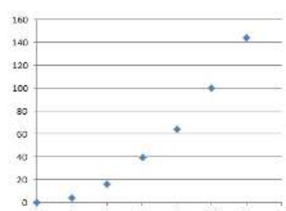
$$g\Big([a,b]\Big) := \int_{[a,b]} f'dx.$$

1.4 Calculus of data

Now, let's approach the issue from the opposite direction. What if we start with *discrete data*?

In real life, a function is given simply by a series of numbers. For example, this is what the "graph" of a function looks like in a common spreadsheet:

x	*y*
0	0
2	4
4	16
6	36
8	64
10	100
12	144

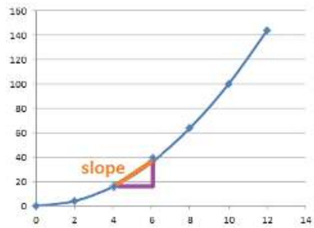


Can we do *calculus* with such functions?

The two primary components of calculus as we know it are the following two questions:

1) *What is the function's rate of change*?

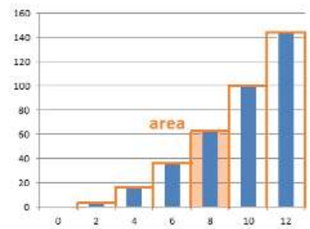
The derivative tells us the rate of change of data that continuously varies. Its discrete analog is the *difference*, the difference of values at two consecutive points. More precisely, if the function is represented as a collection of points on the plane, then its "derivative" – evaluated at one of the intervals rather than a point – is the slope of the line that connects the endpoints of the interval.



2) *What is the area under the graph of the function*?

The definite integral gives us the area under a continuous curve. Its discrete analog, as we just saw, is the *sum*, the sum of the values of the function for all points within the interval. Moreover, if this is a step-function, the "integral" is the sum of the areas of the rectangles determined by these steps, just as before.

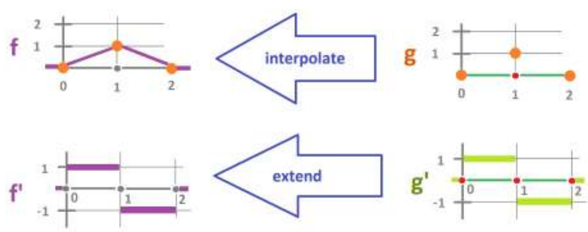
Even though the graph above is made of *dots*, with a couple of clicks, we can plot the same function with *bars*:



So, the same discrete function has been interpreted in two different ways:
- for differentiation – with *dots* plotted above the integer values, and
- for integration – with *rectangles* plotted over the segments.

We, again, arrive to the idea of 0-forms and 1-forms.

At this point, we can construct continuous counterparts of these discrete forms by interpolation:



To summarize, an incremental motion is represented by a 0-form. Its values provide the locations at every instant of the chosen sequence of moments of time while we think of the intermediate locations as unknown or unimportant. Meanwhile, a 1-form gives the incremental changes of locations.

Exercise 1.5. Show that, given an initial location, all locations can be found from a given 1-form.

Exercise 1.6. Given a 0- and a 1-form, what kind of continuous motion produces these forms *exactly*?

1.5 The spaces of cochains

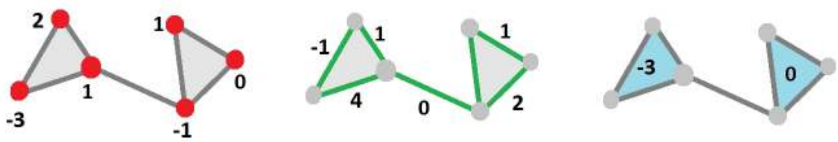
Below we start to develop the theory of cochains that culminates in cohomology theory. We focus on the case when:

- K is an oriented simplicial complex, and
- the ring of coefficients $R := \mathbf{R}$ is the reals.

Definition 1.7. A k-*cochain* on K is any linear function from the vector space of k-chains to R:

$$s : C_k(K) \to \mathbf{R}.$$

If we consider one cell at a time, it is simple: every such function is the multiplication by a real number. This number is the only thing that we need to record. We visualize 0-, 1-, and 2-cochains accordingly:



Thus,

- a k-cochain assigns a number, $r \in \mathbf{R}$, to each k-chain.

Proposition 1.8. The k-cochains on complex K form a vector space.

Notation: The vector space of k-cochains is denoted by

$$C^k = C^k(K) := \{k\text{-cochains in complex}K\}.$$

The notation coincides unfortunately with that for the space of differentiable functions.

There is a special correspondence between chains and cochains. For every k-cell c in K, there is a corresponding "elementary" cochain. We define the *dual* cochain $c^* : C_k \to \mathbf{R}$ of c by

$$c^*(t) = \begin{cases} 1 & \text{if } t = c, \\ 0 & \text{if } t \neq c. \end{cases}$$

Proposition 1.9. If the vector space of all k-chains C_k has as its standard basis the k-cells:

$$\gamma := \{c_1, ..., c_n\},$$

then the space of all k-cochains C^k has as its standard basis the *duals* of these cells:

$$\gamma^* := \{c_1^*, ..., c_n^*\}.$$

Proof. These functions are defined, following the duality idea, on the basis elements of γ by

$$c_i^*(c_j) := \delta_{ij}.$$

■

Exercise 1.10. Finish the proof.

Proposition 1.11. These two spaces are isomorphic in every dimension:

$$C_k \cong C^k,$$

with the isomorphism given by

$$\Phi(c_i) := c^i.$$

Then the dual of *any* chain makes sense:

$$\Big(\sum_i r_i c_i\Big)^* := \sum_i r_i c^i,$$

and it satisfies the duality property below.

Theorem 1.12 (Duality). For any non-zero chain a, we have

$$a^*(a) = 1.$$

Exercise 1.13. Prove the theorem.

Proposition 1.14. The function $x \mapsto x^*$ generates an isomorphism between the vector spaces of k-chains and k-cochains.

Exercise 1.15. Prove the proposition.

Suppose now we are to evaluate an arbitrary cochain at an arbitrary chain:

$$s(a) = ?$$

These two are given by

$$s = \sum_i r_i c_i^*, \ a = \sum_i t_i c_i, \ \text{for some } r_i, t_i \in \mathbf{R},$$

where $\{c_i\}$ are the cells. Then the linearity and the duality properties produce the following:

$$\begin{array}{ll} s(a) & = \Big(\sum_i r_i c_i^*\Big)\Big(\sum_j t_j c_j\Big) \\ & = \sum_i \sum_j r_i t_j c_i^*(c_j) \\ & = \sum_i \sum_j r_i t_j \delta_{ij} \\ & = \sum_i r_i t_i \in \mathbf{R}. \end{array}$$

This coordinate-wise multiplication suggests a convenient **notation:** the value of a cochain s at a given chain a is denoted by

$$s(a) = \langle s, a \rangle,$$

when they are written as coordinate vectors in terms of their respective bases.

1.6 The coboundary operator

Recall that the boundary operator relates $(k+1)$-chains and k-chains and, therefore, $(k+1)$-cochains and k-cochains. What happens is seen in the following commutative diagram:

$$\begin{array}{rcccc} a \in & C_{k+1}(K) & \xrightarrow{\ \ s\ \ } & \mathbf{R} & \ni r \\ & \downarrow{\scriptstyle \partial_{k+1}} & & || & || \\ b \in & C_k(K) & \xrightarrow{\ \ t\ \ } & \mathbf{R} & \ni r \end{array}$$

If a is any $(k+1)$-chain (in the first row), then $b := \partial_{k+1} a$ is its *boundary* (in the second row). Now, if t is a k-cochain t (in the second row), we can define its *coboundary* (in the first row) as a $(k+1)$-cochain s given by its values:

$$s(a) := r = t(b) = t\partial_{k+1}(a),$$

for all a; or simply:

$$s := t\partial_{k+1}.$$

This observation justifies the following.

Definition 1.16. The kth *coboundary operator* of K is the linear operator:

$$\partial^k : C^k \to C^{k+1},$$

defined by

$$\partial^k(t) := t\partial_{k+1},$$

for any k-cochain t in K.

We use **notation** ∂^* for the coboundary operator when the dimension of the cochain is clear. We will also use d in the context of differential forms.

The dot product notation reveals duality in the above definition:

$$\boxed{\langle \partial^* t, a \rangle = \langle t, \partial a \rangle}$$

Let's apply this concept in the familiar context first: the graphs. The only boundary operator that matters is

$$\partial_1 : C_1(K) \to C_0(K).$$

Therefore the only coboundary operator that matters is

$$\partial^0 : C^0(K) \to C^1(K).$$

Example 1.17 (circle). Let's compute the coboundary operator of the circle. Suppose the complex has three vertices A, B, C and their edges $a = AB,\ b = BC,\ c = CA$. Then the bases of C_0 and C_1 are respectively:

$$\{A, B, C\} \text{ and } \{a, b, c\}.$$

Then the bases of C^0 and C^1 are respectively:

$$\{A^*, B^*, C^*\} \text{ and } \{a^*, b^*, c^*\}.$$

Now the values of the coboundary operator:

$$\begin{array}{rlrl}
\partial^0(A^*)(a) := A^*(\partial_1 a) & = A^*(B - A) = & -1, & \\
\partial^0(A^*)(b) := A^*(\partial_1 b) & = A^*(C - B) = & 0, & \\
\partial^0(A^*)(c) := A^*(\partial_1 c) & = A^*(A - C) = & 1, & \leadsto \partial^0(A^*) = -a^* + c^*; \\
\partial^0(B^*)(a) := B^*(\partial_1 a) & = B^*(B - A) = & 1, & \\
\partial^0(B^*)(b) := B^*(\partial_1 b) & = B^*(C - B) = & -1, & \\
\partial^0(B^*)(c) := B^*(\partial_1 c) & = B^*(A - C) = & 0, & \leadsto \partial^0(B^*) = a^* - b^*; \\
\partial^0(C^*)(a) := C^*(\partial_1 a) & = C^*(B - A) = & 0, & \\
\partial^0(C^*)(b) := C^*(\partial_1 b) & = C^*(C - B) = & 1, & \\
\partial^0(C^*)(c) := C^*(\partial_1 c) & = C^*(A - C) = & -1, & \leadsto \partial^0(C^*) = b^* - c^*.
\end{array}$$

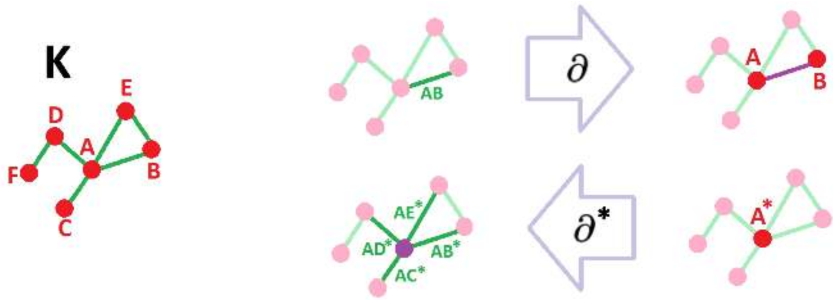
These nine numbers form the matrix of ∂^0, which is, as we see, the transpose of ∂_1:

$$\partial_1 = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}, \quad \partial^0 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

□

Exercise 1.18. Compute the coboundary operator of the figure eight.

The pattern of lowering vs. raising the dimension is illustrated below:



Theorem 1.19. For any vertex A in graph K, we have

$$\partial^* A^* = \sum_i AB_i^*,$$

with summation over all vertices B_i adjacent to A.

Exercise 1.20. Prove the theorem.

Exercise 1.21. Provide an illustration, similar to the one above, for a 2-dimensional complex K and ∂^1.

1.7 The cochain complex and cohomology

The duality between chains and cochains is used to define the isomorphism $\Phi : C_k \to C^k$ via $\Phi(c_i) := c^i$. We sum up the similarity of algebra:

- both chains and cochains are formal linear combinations of some geometric figures;
- the algebra in both is exactly the same: addition and scalar multiplication;
- the vector spaces have the same dimension.

The *difference* becomes visible as we add the (co)boundary operators to the picture. What is happening is seen in the following, non-commutative diagram:

$$\begin{array}{ccc} C_{k+1} & \xrightarrow{\partial_{k+1}} & C_k \\ \downarrow \Phi & \neq & \downarrow \Phi \\ C^{k+1} & \xleftarrow{\partial^k} & C^k \end{array}$$

Exercise 1.22. Give an example to show that the diagram is indeed non-commutative.

Combined, the chain groups $C_k,\ k = 0, 1, ...$, form the chain complex $\{C_*, \partial\}$:

$$0 \xrightarrow{\partial_{N+1}=0} C_N \xrightarrow{\partial_N} ... \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0=0} 0$$

Here, $N = \dim K$. Meanwhile, the cochain groups $C^k,\ k = 0, 1, ...$, form the *cochain complex* $\{C^*, \partial^*\}$:

$$0 \xleftarrow{\partial^N=0} C^N \xleftarrow{\partial^{N-1}} ... \xleftarrow{\partial^0} C^0 \xleftarrow{\partial^{-1}=0} 0$$

The operators here go in the opposite direction!

The result below shows that these two sequences of operators behave in a similar way. The formula comes from and looks identical to the familiar "the boundary of the boundary is zero" formula.

Theorem 1.23 (Double Coboundary Identity). For $k = 0, 1, ...$, we have

$$\partial^{k+1}\partial^k = 0.$$

Proof. For a given chain a, we use the definition of the coboundary operator:

$$\langle \partial^* \partial^* q, a \rangle = \langle \partial^* q, \partial a \rangle = \langle q, \partial \partial a \rangle = \langle q, 0 \rangle = 0. \qquad \blacksquare$$

The short version of the formula is below:

$$\boxed{\partial^* \partial^* = 0}$$

So, the cochain complex is a chain complex after all! The difference is only in the way we index its elements...

Therefore, we can simply repeat the definitions of homology theory, adding "co-" and raising the indices, as follows:

Definition 1.24.
- The elements of $Z^k := \ker \partial^k$ are called the *cocycles*, and
- the elements of $B^k := \operatorname{Im} \partial^{k-1}$ are called the *coboundaries*.

Suppose s is a coboundary, $s = \partial^* q$, then $\partial^* s = \partial^* \partial^* q = 0$. By the argument identical to the one for chain complexes, we prove the following.

Corollary 1.25. Every coboundary is a cocycle, or

$$B^k \subset Z^k, \ \forall k = 0, 1, ...$$

The complete collections of these vector spaces can be linked to each other, as in this diagram:

$$\begin{array}{ccccccc}
\xleftarrow{\partial^{k+1}} & 0 \in Z^{k+1} = \ker \partial^{k+1} & \xleftarrow{\partial^{k}} & 0 \in Z^{k} = \ker \partial^{k} & \xleftarrow{\partial^{k-1}} \\
 & \cap & & \cap & \\
\xleftarrow{\partial^{k+1}} & \operatorname{Im} \partial^{k+1} = B^{k+1} \subset C^{k+1} & \xleftarrow{\partial^{k}} & \operatorname{Im} \partial^{k-1} = B^{k} \subset C^{k} & \xleftarrow{\partial^{k-1}}
\end{array}$$

Definition 1.26. The kth *cohomology group* of K is the quotient of the cocycles over the coboundaries:

$$H^k = H^k(K) := Z^k / B^k = \ker \partial^{k+1} / \operatorname{Im} \partial^k.$$

Notation: When the dimension k is clear, we use H^* for cohomology.

Example 1.27. Let's compute the cohomology of the circle given, as in the last subsection, by the complex with three vertices A, B, C and there edges $a = AB, B = BC, c = CA$. We already have the coboundary operator:

$$\begin{array}{ll}
\partial^0(A^*) & = -a^* + c^*; \\
\partial^0(B^*) & = a^* - b^*; \\
\partial^0(C^*) & = b^* - c^*.
\end{array}$$

Its kernel is

$$Z^0 := \ker \partial^0 =< A^* + B^* + C^* > .$$

Hence,

$$H^0 := Z^0 / B^0 =< A^* + B^* + C^* > .$$

Next, the image is

$$B^1 := \operatorname{Im} \partial^0 =< -a^* + c^*, a^* - b^*, b^* - c^* > .$$

Hence,

$$H^1 := Z^1 / B^1 =< a^* >=< b^* >=< c^* > .$$

It is isomorphic to the homology group! we draw the same conclusion: there is one hole in the circle. □

Exercise 1.28. Compute the cohomology of the figure eight.

The example suggests that not just the chain and cochain groups are isomorphic but so are the homology and cohomology groups – under Φ – as vector spaces.

Theorem 1.29. For graphs, we have

$$H^* \cong H.$$

Exercise 1.30. Prove the theorem. Hint: follow the example.

Corollary 1.31. In an acyclic graph, every 1-cocycle is a 1-coboundary.

Example 1.32. Observe that in the last example Z^0 is generated by $A^* + B^* + C^*$. This means that the cocycles are multiples of this cochain, such as

$$s_r = r(A^* + B^* + C^*).$$

But

$$s_r(A) = r,\ s_r(B) = r,\ s_r(C) = r.$$

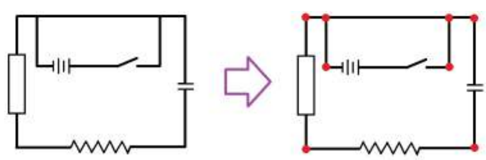
In other words, these functions are *constant*. □

Theorem 1.33. A 0-cochain is a cocycle in K if and only if it is constant on each path-component of K.

Proof. If $\partial^0 Q = 0$ then $\langle \partial^0 Q, a\rangle = 0$ for any 0-chain $a = AB$. Therefore, $\langle Q, B - A\rangle = 0$, or $Q(A) = Q(B)$. ■

Exercise 1.34. Provide the rest of the proof.

Example 1.35 (Kirchhoff's current law). Suppose we have an electrical circuit represented by a graph.

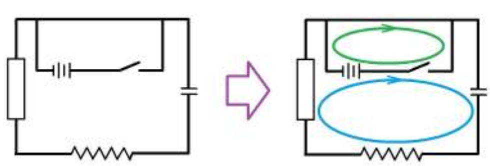


Then the conservation principle is satisfied: *the sum of currents flowing into any junction is equal to the sum of currents flowing out of the junction.* If the edges are oriented, we can think of the currents as 1-chains – over the reals. Moreover, a current is such a 1-cochain I that satisfies, for any vertex A:

$$\sum_k I(e_k) = 0,$$

with summation over all edges e_k adjacent to A. It follows that $I = \partial^* A^*$. It's a coboundary! □

Example 1.36 (Kirchhoff's voltage law). We, once again, consider an electrical circuit understood as a graph.



As the edges are oriented, we can think of the voltages as 1-cochains (over **R**). Moreover, the voltage V over any edge is the drop of the potential U over this edge. The potential is then a 0-cochain and $V = dU$. Therefore, $dU(c) = U(\partial c) = U(0) = 0$ for any 1-cycle c. So, the following conservation principle is satisfied: *the sum of all the voltages around a loop is equal to zero.* □

1.8 Calculus I, the discrete version

Below, we review these new concepts in the context of discrete calculus on $\mathbb{R}$. In this context, the cochains are called *differential forms*, or simply *forms*, while the coboundary operator is called the *exterior derivative* denoted by

$$d := \partial^*.$$

The reason for this terminology is revealed below.

The setting of calculus of one variable requires only 0- and 1-cochains. Therefore, the only complexes we need to consider are *graphs*.

Substituting $a = AB$ into the definition of the exterior derivative, we acquire the following:

Theorem 1.37 (Fundamental Theorem of Calculus for Graphs). Suppose we have a sequence of adjacent nodes in the graph:

$$A = A_0, ..., A_N = B,$$

and let a be the 1-chain represented by the sum of the edges of the sequence:

$$a := A_0A_1 + ... + A_{N-1}A_N.$$

Then, for any 0-form Q, we have

$$\langle dQ, a\rangle = \langle Q, B - A\rangle.$$

Exercise 1.38. State and prove the theorem for the case a an arbitrary 1-form.

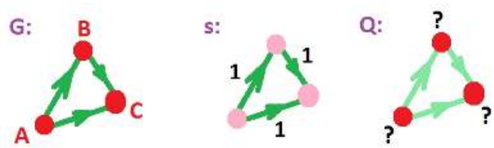
In the integral notation, the result looks familiar:

$$\int_A^B dQ = Q(B) - Q(A).$$

We know that dQ is the coboundary of A but we can also think of Q as an *antiderivative* of dQ.

There are often two issues to be considered about a new concept: existence and uniqueness.

First, given a 1-form s on graph G, does an antiderivative, i.e., a 0-form Q with $dQ = s$, always *exist*? The picture below shows why the answer is No:



Here, if we take the upper path, we have

$$Q(B) - Q(A) = 1, \; Q(C) - Q(B) = 1;$$

but if we take the lower path, we have

$$Q(C) - Q(A) = 1.$$

This is a contradiction! We are facing a *path-dependent* integral.

The idea from calculus is that path-dependence of line integrals can be recast as the non-triviality of integrals along closed loops, i.e., 1-cycles. Indeed, we can restate the above as follows:

$$\int_{[A,B]+[B,C]+[C,A]} dQ = \int_{[A,B]} dQ + \int_{[B,C]} dQ + \int_{[C,A]} dQ = 1 \neq 0.$$

The result should remind us of a similar situation with chains:
- not every chain is a boundary, or a cycle.

This time, we are speaking of cochains:
- not every cochain is a coboundary, or a cocycle.

There is more than just similarity here: if a is a non-zero 1-cycle, then a^* isn't a coboundary. Indeed, we have otherwise:

$$1 = \langle a^*, a \rangle = \langle \partial^* Q, a \rangle = \langle Q, \partial a \rangle = \langle Q, 0 \rangle = 0.$$

In other words, for all integrals over closed loops to be always 0, there should be no closed loops to begin with.

In the calculus context, coboundaries are also called *exact forms*.

Theorem 1.39. Every 1-form is exact on a graph if and only if the graph is a tree.

Exercise 1.40. Provide the rest of the proof.

Second, given a 1-form s on graph G, is an antiderivative, i.e., a 0-form Q with $dQ = s$, *unique*? The recognizable result below shows that the answer is "Yes, in a sense".

Theorem 1.41. If Q, Q' are two 0-forms with $dQ = dQ'$, then they differ by a constant on every path-component of the graph.

Exercise 1.42. Prove the theorem. Hint: consider $Q - Q'$.

A common interpretation of this result is, if two runners run with an identical but possibly variable velocity, the distance between them remains unchanged.

1.9 Social choice: ratings and comparisons

In several stages, we will consider the foundation of a basic *voting system*.
In a typical political election, the voter *ranks* the candidates with no numbers involved. We will concentrate on simpler, numeric voting methods that are meant to evaluate rather than just to rank. These methods require us to use our ring R again.

This is how, typically, votes are used for evaluation. The customer/voter assigns a number from $\{1, 2, ..., N\}$ to a product or service as a substitute of its perceived quality, such as:
- books,
- movies,
- hotels,
- various products,
- service providers,
- etc.

The choice is commonly visualized by 5 stars:

$$\star \quad \star\star \quad \star\star\star \quad \star\star\star\star \quad \star\star\star\star\star\ .$$

In political elections, it is simple:

$$Yes : \underline{\checkmark} \quad No : \underline{\quad}$$

As common it is to *rate* the alternatives, it is very uncommon to *compare* them – pairwise – to evaluate the perceived inequality of value. As we will see, the latter approach will help us resolve some of the issues with the former.

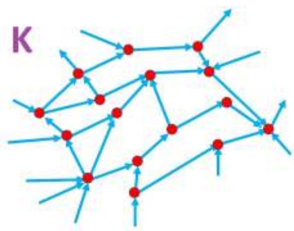
The general setup is as follows. There are n alternatives/candidates:

$$A := \{0, 1, 2, ..., n-1\} = \{A_0, A_1, A_2, ..., A_{n-1}\}.$$

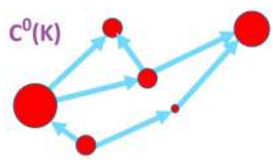
They are ordered. We also suppose that these candidates are the vertices of an oriented simplicial complex K. All candidates are subject to evaluation but the presence of edge AB in K reflects the fact that A and B are *comparable*.

At first, we will only consider what a single voter does; however, a single vote may come from aggregated votes.

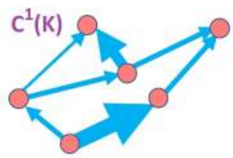
We initially assume that K is simply a (directed) *graph*:



A *rating vote* is a number $X \in R$ assigned to each candidate, as well as a combination of all these numbers. Therefore, this is a 0-cochain on K. The totality is $C^0 = C^0(K)$.



A *comparison vote* is a number $x \in R$ assigned to each pair of candidates (i.e., the edge between them), as well as a combination of all these numbers. Therefore, this is a 1-cochain on K. The totality is $C^1 = C^1(K)$.



That's all.

The differences, according to the order of vertices, over a given rating vote for each pair of candidates produces a comparison vote:

$$x(12) = \partial^0 X(12) := X(2) - X(1).$$

Combined, this is the 0-coboundary operator $\partial^0 : C^0 \to C^1$.

A *rating* comparison vote is one that comes from some rating vote of the candidates, as described. Therefore, this is a 1-coboundary. The totality is $B^1 = \operatorname{Im} \partial^0$.

A *non-circular* comparison vote is one that satisfies the condition: as one takes a circular route through the candidates/vertices of K, the preferences always add up to zero. Therefore, this is a 1-cocycle. The totality is $Z^1 = \ker \partial^1$.

We now use this terminology to restate what we have learned in the this section.

"Every coboundary is a cocycle" is restated as follows.

Proposition 1.43. Every rating comparison vote is non-circular.

"Every k-cocycle is a k-coboundary if $H^1(K) = 0$" is restated as follows.

Proposition 1.44. If K is a tree, every comparison vote is rating.

The cochain complex and these subspaces are shown below:

$$\begin{array}{ccccccc}
0 & \xleftarrow{\partial^2=0} & C^2 & \xleftarrow{\partial^1} & C^1 & \xleftarrow{\partial^0} & C^0\\
 & & \cup & & \cup & & ||\\
 & & 0 & \xleftarrow{0} & Z^1 & \longleftarrow & C^0\\
 & & || & & \cup & & ||\\
 & & 0 & \xleftarrow{0} & B^1 & \xleftarrow{onto} & C^0.
\end{array}$$

Even in the case of a single voter, who wins?

In the simplest case, there is only a single rating. Provided R is equipped with an order relation, we choose the winner to be the one (perhaps tied) with the largest rating.

What if the voter casts a comparison vote b instead? Even though what we really need is a rating vote, we don't have to discard b. The reason is that b may be a rating comparison vote: $b = \partial^0(c)$ for some rating c (i.e., b is a coboundary). Then the winner should be the one with the largest value of c. Based on the last proposition, we conclude that we can guarantee that there is such a c if and only if K is a tree, which is certainly atypical. Finally, if this is the case, the choice of the winner(s) is unambiguous.

Exercise 1.45. Prove the last statement.

Exercise 1.46. What if, in order to guarantee a winner, we simply disallow circular comparison votes? Hint: what is the new chain complex?

Exercise 1.47. What to do if the voter casts both a rating vote and a comparison vote?

Exercise 1.48. What if there are two voters?

The votes of many voters may be combined into a single vote over $\mathbf{Z}$ or $\mathbf{R}$. Aggregation methods are discussed later.

Exercise 1.49. Use cochains to represent exchange rates between world currencies. What about barter?

---

2 Calculus on cubical complexes

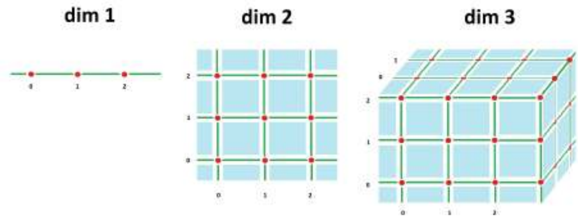
2.1 Visualizing cubical forms

In calculus, the quantities to be studied are typically *real numbers*. We choose our ring of coefficients to be $R = \mathbf{R}$.

Meanwhile, the locus is typically the Euclidean space $\mathbf{R}^n$. We choose for now to concentrate on the *cubical grid*, i.e., the infinite cubical complex acquired by dividing the Euclidean space into cubes, $\mathbb{R}^n$.

In $\mathbb{R}^1$, these pieces are: points and (closed) intervals,
- the 0-cells: ..., $-3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,$ ..., and
- the 1-cells: ..., $[-2,-1],\ [-1,0],\ [0,1],\ [1,2],$ ....

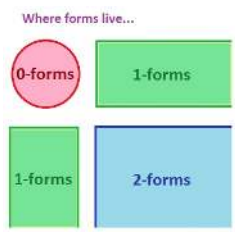
In $\mathbb{R}^2$, these parts are: points, intervals, and squares ("pixels"):



Moreover, in $\mathbb{R}^2$, we have these cells represented as products:
- 0-cells: $\{(0,0)\}, \{(0,1)\}, ...$;
- 1-cells: $[0,1] \times \{0\},\ \{0\} \times [0,1], ...$;
- 2-cells: $[0,1] \times [0,1], ...$.

In this section, we will use the calculus terminology: "differential forms" instead of "cochains".

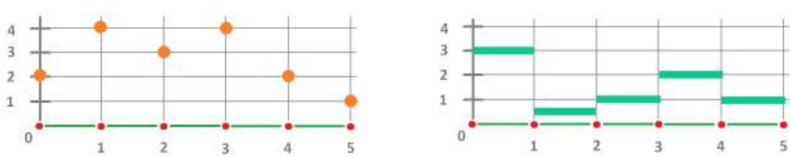
Recall that within each of these pieces, a form is unchanged; i.e., it's a single number.



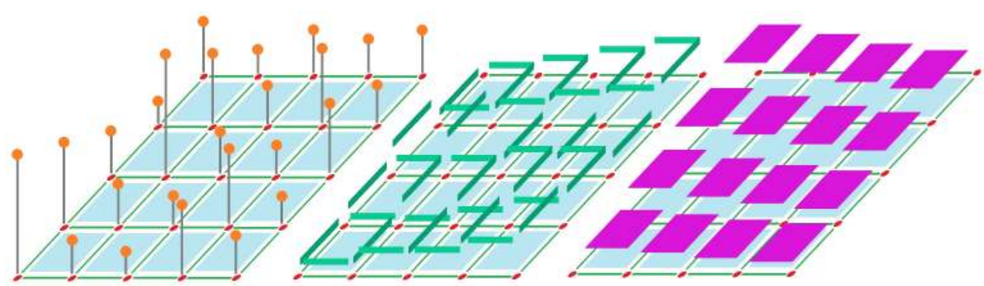
Then, the following is the simplest way to understand these forms.

Definition 2.1. A *cubical k-form* is a real-valued function defined on k-cells of $\mathbb{R}^n$.

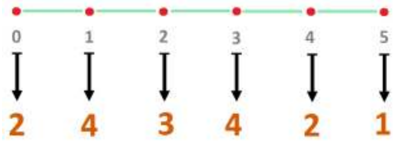
This is how we plot the graphs of forms in $\mathbb{R}^1$:

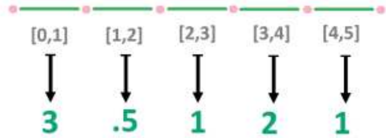


And these are 0-, 1-, and 2-forms in $\mathbb{R}^2$:



To emphasize the nature of a form as a function, we can use arrows:

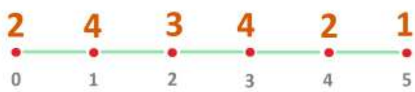




Here we have two forms:

- a 0-form with $0 \mapsto 2,\ 1 \mapsto 4,\ 2 \mapsto 3, \ldots$; and
- a 1-form with $[0,1] \mapsto 3,\ [1,2] \mapsto .5,\ [2,3] \mapsto 1, \ldots$.

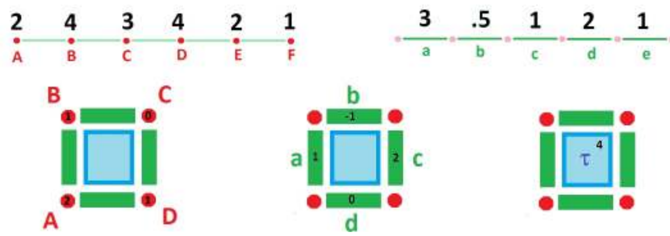
A more compact way to visualize is this:



Here we have two forms:

- a 0-form Q with $Q(0) = 2,\ Q(1) = 4,\ Q(2) = 3, \ldots$; and
- a 1-form s with $s\Big([0,1]\Big) = 3,\ s\Big([1,2]\Big) = .5,\ s\Big([2,3]\Big) = 1, \ldots$.

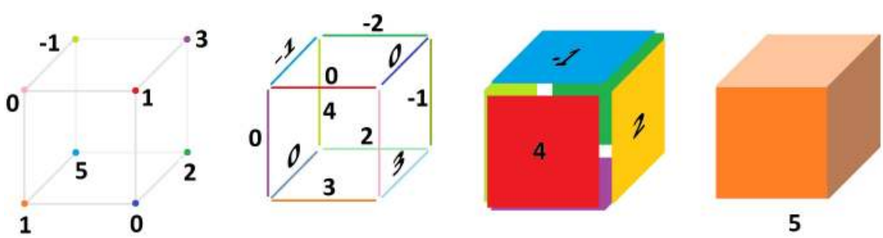
We can also use letters to label the cells, just as before. Each cell is then assigned *two* symbols: one is its name (a latter) and the other is the value of the form at that location (a number):



Here we have:

- $Q(A) = 2,\ Q(B) = 4,\ Q(C) = 3, \ldots$;
- $s(AB) = 3,\ s(BC) = .5,\ s(CD) = 1, \ldots$.

We can simply label the cells with numbers, as follows:



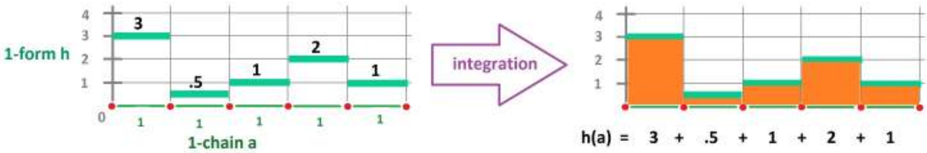
Exercise 2.2. Another way to visualize forms is with color. Implement this idea with a spreadsheet.

2.2 Forms as integrands

It is common for a student to overlook the distinction between chains and cochains/forms and to speak of the latter as linear combinations of cells. The confusion is understandable because they "look" identical. Frequently, one just assigns numbers to cells in a complex as we did above. The difference is that these numbers aren't the coefficients of the cells in some chain but the *values* of the 1-cochain on these cells. The idea becomes explicit when we think in calculus terms:

- forms are integrands, and
- chains are domains of integration.

In the simplest setting, we deal with the intervals in the complex of the real line $\mathbb{R}$. Then the form assigns a number to each interval to indicate the values to be integrated and the chain indicates how many times the interval will appear in the integral, typically once:



Here, we have:

$$\begin{array}{llllll} h(a) & = \int_a h & & & & \\ & = \int_{[0,1]} h & + \int_{[1,2]} h & + \int_{[2,3]} h & + \int_{[3,4]} h & + \int_{[4,5]} h \\ & = 3 & +.5 & +1 & +2 & +1. \end{array}$$

The simplest form of this kind is the form that assigns 1 to each interval in the complex $\mathbb{R}$. We call this form dx. Then any form h can be built from dx by multiplying – cell by cell – by a discrete function that takes values $3, .5, 1, 2, 1$ on these cells:

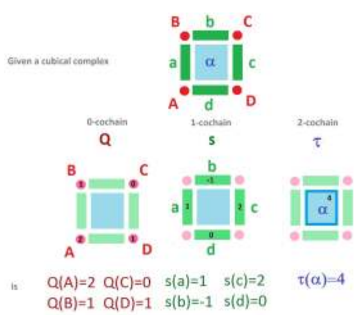
The main property of this new form is:

$$\int_{[A,B]} dx = B - A.$$

Exercise 2.3. What is the antiderivative of dx?

Exercise 2.4. Show that every 1-form in $\mathbb{R}^1$ is a "multiple" of dx: $h = Pdx$.

Next, $\mathbb{R}^2$:

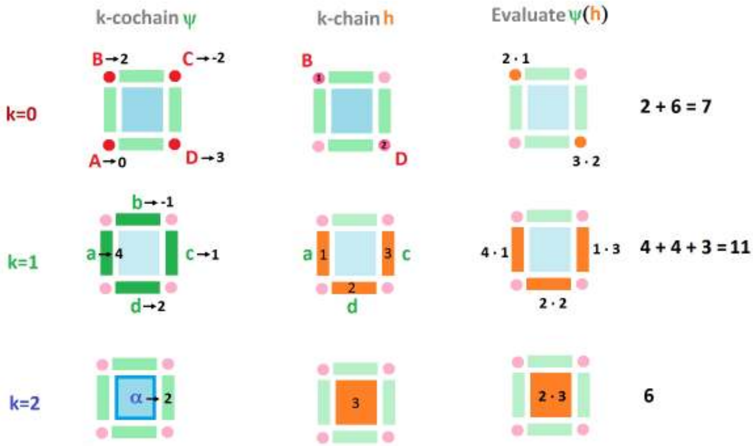


In the diagram,

- the names of the cells are given in the first row;
- the values of the form on these cells are given in the second row; and
- the algebraic representation of the forms is in the third.

The second row gives one a compact representation of the form when you don't want to name the cells.

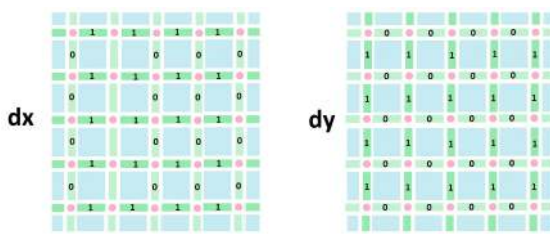
Discrete differential forms (cochains) are real-valued, linear functions defined on chains:



One should recognize the second line as a line integral:

$$\psi(h) = \int_h \psi.$$

What is dx in $\mathbb{R}^2$? Naturally, its values on the edges parallel to the x-axis are 1's and on the one parallel to the y-axis are 0's:



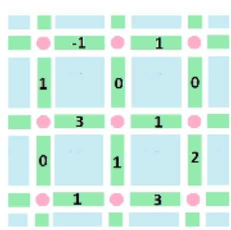
Of course, dy is the exact opposite. Algebraically, their representations are as follows:

- $dx\Big([m, m+1] \times \{n\}\Big) = 1,\ dx\Big(\{m\} \times [n, n+1]\Big) = 0;$
- $dy\Big([m, m+1] \times \{n\}\Big) = 0,\ dy\Big(\{m\} \times [n, n+1]\Big) = 1.$

Now we consider a general 1-form:

$$Pdx + Qdy,$$

where P, Q are discrete functions, not just numbers, that may vary from cell to cell. For example, this could be P:



Exercise 2.5. Show that every 1-form in $\mathbb{R}^2$ is such a "linear combination" of dx and dy.

At this point, we can integrate this form. As an example, suppose S is the chain that represents the 2×2 square in this picture going clockwise. The edges are oriented, as always, along the axes. Let's consider the line integral computed along this curve one cell at a time starting at the left lower corner:

$$\int_S Pdx = 0 \cdot 0 + 1 \cdot 0 + (-1) \cdot 1 + 1 \cdot 1 + 0 \cdot 0 + 2 \cdot 0 + 3 \cdot (-1) + 1 \cdot (-1).$$

We can also compute:

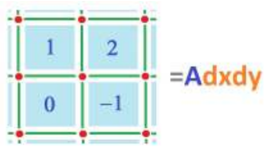
$$\int_S Pdy = 0 \cdot 1 + 1 \cdot 1 + (-1) \cdot 0 + 1 \cdot 0 + 0 \cdot (-1) + 2 \cdot (-1) + 3 \cdot 0 + 1 \cdot 0.$$

If Q is also provided, the integral

$$\int_S Pdx + Qdy$$

is a similar sum.

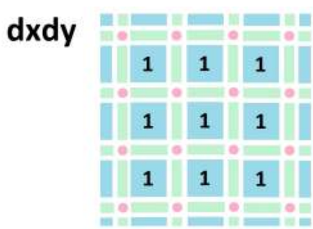
Next, we illustrate 2-forms in $\mathbb{R}^2$:



The double integral over this square, S, is

$$\int_S Adxdy = 1 + 2 + 0 - 1 = 2.$$

And we can understand $dx\,dy$ as a 2-form that takes the value of 1 on each cell:



Exercise 2.6. Evaluate $\int_S dxdy$, where S is an arbitrary collection of 2-cells.

2.3 The algebra of forms

We already know that the forms, as cochains, are organized into vector spaces, one for each degree/dimension. Let's review this first.

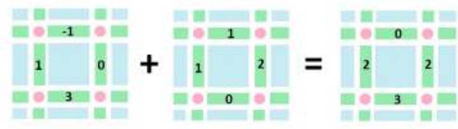
If p, q are two forms of the same degree k, it is easy to define algebraic operations on them.

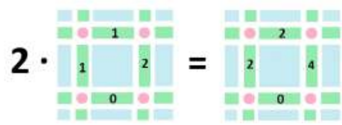
First, their *addition*. The sum $p + q$ is a form of degree k too and is computed as follows:

$$(p + q)(a) := p(a) + q(a),$$

for every k-cell a.

As an example, consider two 1-forms, p, q. Suppose these are their values defined on the 1-cells (in green):





Then $p+q$ is found by

$$1+1=2,\ -1+1=0,\ 0+2=2,\ 3+0=3,$$

as we compute the four values of the new form one cell at a time.

Next, *scalar multiplication* is also carried out cell by cell:

$$(\lambda p)(a) := \lambda p(a),\ \lambda \in \mathbf{R},$$

for every k-cell a.

We know that these operations satisfy the required properties: associativity, commutativity, distributivity, etc. Subsequently, we have a vector space:

$$C^k = C^k(\mathbb{R}^n),$$

the space of k-forms on the cubical grid of $\mathbf{R}^n$.

There is, however, an operation on forms that we haven't seen yet.

Can we make $dxdy$ from dx and dy? The answer is provided by the *wedge product* of forms:

$$dxdy = dx \wedge dy.$$

Here we have:
- a 1-form $dx \in C^1(\mathbb{R}_x)$ defined on the horizontal edges,
- a 1-form $dy \in C^1(\mathbb{R}_y)$ defined on the vertical edges, and
- a 2-form $dxdy \in C^2(\mathbb{R}^2)$ defined on the squares.

But squares are products of edges:

$$\alpha = a \times b.$$

Then we simply set:

$$(dx \wedge dy)(a \times b) := dx(a) \cdot dy(b).$$

What about $dydx$? To match what we know from calculus:

$$\int_\alpha dydx = -\int_\alpha dxdy,$$

we require *anti-commutativity* of cubical cochains under wedge products:

$$dy \wedge dx = -dx \wedge dy.$$

Now, suppose we have two *arbitrary* 1-forms p, q and we want to define their wedge product on the square $\alpha := a \times b$. We can't use the simplest definition:

$$(p \wedge q)(a \times b) \stackrel{?}{=} p(a) \cdot q(b),$$

as it fails to be anti-commutative:

$$(q \wedge p)(a \times b) = q(a) \cdot p(b) = p(b) \cdot q(a).$$

Since we need both of these terms:

$$p(a)q(b) \quad p(b)q(a),$$

let's combine them.

Definition 2.7. The *wedge product* of two 1-forms is a 2-form given by

$$(p \wedge q)(a \times b) := p(a)q(b) - p(b)q(a).$$

The minus sign is what gives us the *anti*-commutativity:

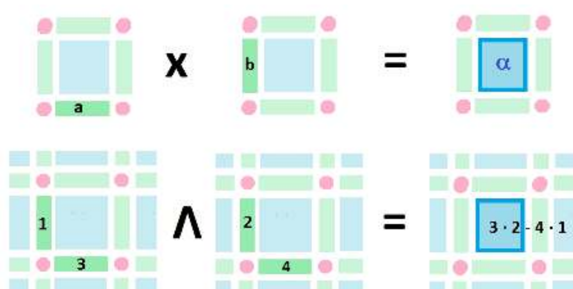
$$(p \wedge p)(a \times b) := q(a)p(b) - q(b)p(a) = -(p(a)q(b) - p(b)q(a)).$$

Proposition 2.8.

$$dx \wedge dx = 0, \ dy \wedge dy = 0.$$

Exercise 2.9. Prove the proposition.

Here is an illustration of the relation between the product of cubical chains and the wedge product of cubical forms:



The general definition is as follows.

Recall that, for our cubical grid $\mathbb{R}^n$, the cells are the cubes given as products:

$$Q = \prod_{k=1}^{n} A_k,$$

with each A_k either a vertex or an edge in the kth component of the space. We can derive the formula for the wedge product in terms of these components. If we omit the vertices, a $(p+q)$-cube can be rewritten as

$$Q = \prod_{i=1}^{p} I_i \times \prod_{i=p+1}^{p+q} I_i,$$

where I_i is its ith edge. The two summands are a p-cube and a q-cube respectively and can be the inputs of a p-form and a q-form respectively.

Definition 2.10. The *wedge product* of the a p-form and a q-form is a $(p+q)$-form given by its value on the $(p+q)$-cube, as follows:

$$(\varphi^p \wedge \psi^q)(Q) := \sum_s (-1)^{\pi(s)} \varphi^p \Big(\prod_{i=1}^{p} I_{s(i)} \Big) \cdot \psi^q \Big(\prod_{i=p+1}^{p+q} I_{s(i)} \Big),$$

with summation over all permutations $s \in \mathcal{S}_{p+q}$ with $\pi(s)$ the parity of s (the superscripts are the degrees of the forms).

Exercise 2.11. Verify that $Pdx = P \wedge dx$. Hint: what is the dimension of the space?

Proposition 2.12. The wedge product satisfies the *skew-commutativity*:

$$\varphi^m \wedge \psi^n = (-1)^{mn} \psi^n \wedge \varphi^m.$$

Under this formula, we have the anti-commutativity when $m = n = 1$, as above.

Exercise 2.13. Prove the proposition.

Unfortunately, the wedge product isn't associative!

Exercise 2.14. (a) Give an example of this:

$$\phi^1 \wedge (\psi^1 \wedge \theta^1) \neq (\phi^1 \wedge \psi^1) \wedge \theta^1.$$

(b) For what class of form is the wedge product associative?

The crucial difference between the linear operations and the wedge product is that the former two act *within* the space of k-forms:

$$+, \cdot : C^k \times C^k \to C^k;$$

while the latter acts *outside*:

$$\wedge : C^k \times C^m \to C^{k+m}.$$

We can make both operate within the same space if we define them on the *graded space* of all forms:

$$C^* := C^0 \oplus C^1 \oplus ...$$

2.4 Derivatives of functions vs. derivatives of forms

The difference is:

- the derivative of a function is the *rate of change*, while
- the exterior derivative of a 0-form is the *change*.

The functions we are dealing with are discrete. At their simplest, they are defined on the integers:

$$n = ... - 1, 0, 2, 3, 4, ...$$

They change abruptly and, consequently, the change is the *difference of values*:

$$f(n+1) - f(n).$$

The only question is *where to assign* this number as the value of some new function. What is the nature of the input of this function?

The illustration below suggests the answer:

The output should be assigned to the (oriented) edge that connects n to $n+1$:

$$[n, n+1] \mapsto f(n+1) - f(n).$$

Assigning this number to either of the two endpoints would violate the symmetry. If the input changes in the opposite way, so does the change of the output, as expected:

$$[n+1, n] = -[n, n+1] \mapsto f(n) - f(n+1).$$

Example 2.15. Let's look at this construction from the point of view of our study of motion. Suppose function p gives the position and suppose

- at time n hours we are at the 5 mile mark: $p(n) = 5$, and then
- at time $n+1$ hours we are at the 7 mile mark: $p(n+1) = 7$.

We don't know what exactly has happened during this hour but the simplest assumption would be that we have been walking at a constant speed of 2 miles per hour. Now, instead of our velocity function v assigning this value to each instant of time during this period, it is assigned to the *whole* interval:

$$v\Big([n, n+1]\Big) = 2.$$

This way, the elements of the domain of the velocity function are the *edges*. □

The relation between a discrete function and its change is illustrated below:

Definition 2.16. The *exterior derivative* of a discrete function f at $[n, n+1]$ is defined to be the number

$$(df)\Big([n, n+1]\Big) := f(n+1) - f(n).$$

On the whole domain of f, we have:

- f is defined on the 0-cells, and
- df is defined on 1-cells.

Definition 2.17. The *exterior derivative* df of a 0-form f is a 1-form df given on each interval by the above formula.

Let's contrast:

- the derivative of a function f at $x = a$ is a number $f'(a)$ assigned to point a, hence the derivative function f' of a function f on $\mathbf{R}$ is another function on $\mathbf{R}$, while
- the exterior derivative of a function f at $[n, n+1]$ is a number $df\Big([n, n+1]\Big)$ assigned to interval $[n, n+1]$, hence the exterior derivative df of a function f on $C^0(\mathbb{R})$ is a function on $C^1(\mathbb{R})$.

Furthermore, if the interval was of length h, we would see the obvious difference between the derivative and the exterior derivative:

$$\frac{f(a+h) - f(a)}{h} \text{ vs. } f(a+h) - f(a).$$

Unlike the former, the latter can be defined over *any ring* R.

Proposition 2.18. The exterior derivative is a linear operator

$$d : C^0(\mathbb{R}) \to C^1(\mathbb{R}).$$

Exercise 2.19. Prove the proposition.

Exercise 2.20. State and prove the analogs of the familiar theorems from calculus about the relation between the exterior derivative and: (a) monotonicity, and (b) extreme points.

Let's next approach this operator from the point of view of the *Fundamental Theorem of Calculus.*

Given $f \in C^0(\mathbb{R})$, we take our definition of $df \in C^1(\mathbb{R})$,

$$df\Big([a, a+1]\Big) := f(a+1) - f(a),$$

and simply rewrite it in the integral notation:

$$\int_{[a,a+1]} df = f(a+1) - f(a).$$

Observe that here $[a, a+1]$ is a 1-cell and the points $\{a\}, \{a+1\}$ make up the boundary of this cell. In fact,

$$\partial\Big([a, a+1]\Big) = \{a+1\} - \{a\},$$

taking into account the orientation of this edge.

What about integration over "longer" intervals? For $a, k \in \mathbf{Z}$, we have:

$$\begin{aligned}\int_{[a,a+k]} df &= \int_{[a,a+1]} df + \int_{[a+1,a+2]} df + ... + \int_{[a+k-1,a+k]} df \\ &= f(a+1) - f(a) + f(a+2) - f(a+1) + ... + f(a+k) - f(a+(k-1)) \\ &= f(a+k) - f(a).\end{aligned}$$

We have simply applied the definition repeatedly.

Thus, the Fundamental Theorem of Calculus still holds, in its "net change" form.

In fact, we can now rewrite it in a fully algebraic way. Indeed,

$$\sigma = [a, a+1] + [a+1, a+2] + ... + [a+k-1, a+k]$$

is a 1-chain and df is a 1-cochain. Then the above computation takes this form:

$$\begin{aligned}df(\sigma) &= df\Big([a, a+1] + [a+1, a+2] + ... + [a+k-1, a+k]\Big) \\ &= df\Big([a, a+1]\Big) + df\Big([a+1, a+2]\Big) + ... + df\Big([a+k-1, a+k]\Big) \\ &= f(a+k) - f(a) \\ &= f\Big(\partial[a, a+k]\Big).\end{aligned}$$

The resulting interaction of the operators of exterior derivative and boundary,

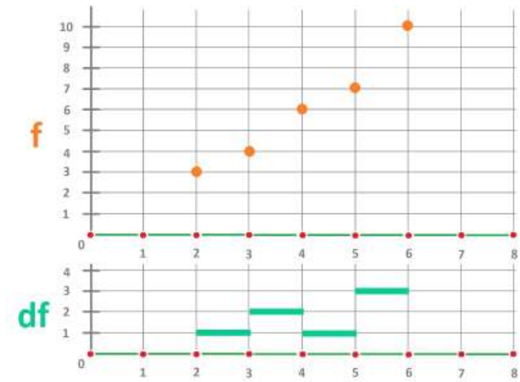
$$\boxed{df(\sigma) = f(\partial\sigma)}$$

is an instance of the (general) *Stokes Theorem.*

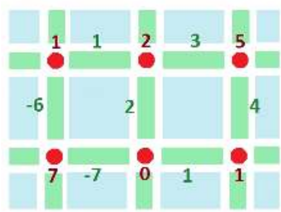
Exercise 2.21. Explain the relation between this formula and the formula used for integration by substitution, $df = f'dx$.

To summarize, *the exterior derivative of a* 0*-form is a* 1*-form computed as the difference of its values.*

Exercise 2.22. Show how the definition, and the theorem, is applicable to any cubical complex $R \subset \mathbb{R}$. Hint:

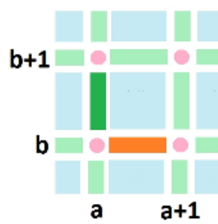


Next we consider the case of the space of dimension 2 and forms of degree 1. Given a 0-form f (in red), we compute its exterior derivative df (in green):



Once again, it is computed by taking differences.

Let's make this computation more specific. We consider the differences horizontally (orange) and vertically (green):



According to our definition, we have:

- (orange) $df\Big([a, a+1] \times \{b\}\Big) := f\Big(\{(a+1, b)\}\Big) - f\Big(\{(a, b)\}\Big)$
- (green) $df\Big(\{a\} \times [b, b+1]\Big) := f\Big(\{(a, b+1)\}\Big) - f\Big(\{(a, b)\}\Big)$.

Therefore, we have:

$$df = \langle \text{grad } f, dA \rangle,$$

where

$$dA := (dx, dy), \quad \text{grad } f := (d_x f, d_y f).$$

The notation is justified if we interpret the above as "partial exterior derivatives":

- $d_x f\Big([a, a+1] \times \{b\}\Big) := f(a+1, b) - f(a, b)$,
- $d_y f\Big(\{a\} \times [b, b+1]\Big) := f(a, b+1) - f(a, b)$.

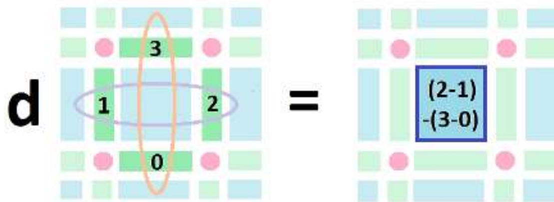
2.5 The exterior derivative of a 1-form

What about the higher degree forms?

Let's start with 1-forms in $\mathbb{R}^2$. The exterior derivative is meant to represent the *change* of the values of the form as we move around the space. This time, we have possible changes as we move in both horizontal and vertical directions. Then we will be able to express these quantities by a single number as a *combination of the changes*:

$$\text{the horizontal change } \pm \text{ the vertical change.}$$

If we concentrate on a single square, these differences are computed on the opposite edges of the square.



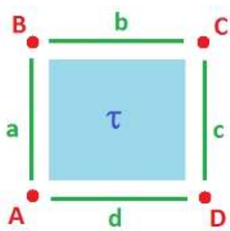
Just as in the last subsection, the question arises: where to assign this value? Conveniently, the resulting value can be given to the square itself.

We will justify the negative sign in the formula below.

With each 2-cell given a number in this fashion, the exterior derivative of a 1-form is a 2-form.

Exercise 2.23. Define and compute the exterior derivative of 1-forms in $\mathbb{R}$.

Let's consider the exterior derivative for a 1-form defined on the edges of this square oriented along the x- and y-axes:



Definition 2.24. The *exterior derivative* $d\varphi$ of a 1-form φ is defined by its value at each 2-cell τ as the difference of the changes of φ with respect to x and y along the edges of τ; i.e.,

$$d\varphi(\tau) = \Big(\varphi(c) - \varphi(a)\Big) - \Big(\varphi(b) - \varphi(d)\Big).$$

Why minus? Let's rearrange the terms:

$$d\varphi(\tau) = \varphi(d) + \varphi(c) - \varphi(b) - \varphi(a).$$

What we see is that we go *full circle* around τ, counterclockwise with the correct orientations. Of course, we recognize this as a line integral. We can read this formula as follows:

$$\int_\tau d\varphi = \int_{\partial\tau} \varphi.$$

Algebraically, it is simple:

$$d\varphi(\tau) = \varphi(d) + \varphi(c) + \varphi(-b) + \varphi(-a) = \varphi(d + c - b - a) = \varphi(\partial\tau).$$

Thus, the resulting interaction of the operators of exterior derivative and boundary takes the same form as for dimension 1 discussed above:

$$d\varphi = \varphi\partial.$$

Once again, it is an instance of the Stokes Theorem, which is used as the definition of d.

Let's represent our 1-form as

$$\varphi = Adx + Bdy,$$

where A, B are the *coefficient functions* of φ:

- A is the numbers assigned to the horizontal edges: $\varphi(b), \varphi(d)$, and
- B is the numbers assigned to the vertical edges: $\varphi(a), \varphi(c)$.

Now, if we think one axis at a time, we use the last subsection and conclude that

- A is a 0-form with respect to y and $dA = \big(\varphi(b) - \varphi(d)\big)dy$, and
- B is a 0-form with respect to x and $dB = \big(\varphi(c) - \varphi(a)\big)dx$.

Now, from the definition we have:

$$\begin{aligned} d\varphi &= \Big(\big(\varphi(c) - \varphi(a)\big) - \big(\varphi(b) - \varphi(d)\big)\Big)dxdy \\ &= \big(\varphi(c) - \varphi(a)\big)dxdy - \big(\varphi(b) - \varphi(d)\big)dxdy \\ &= \big(\varphi(c) - \varphi(a)\big)dxdy + \big(\varphi(b) - \varphi(d)\big)dydx \\ &= \Big(\big(\varphi(c) - \varphi(a)\big)dx\Big)dy + \Big(\big(\varphi(b) - \varphi(d)\big)dy\Big)dx \\ &= dBdy + dAdx. \end{aligned}$$

We have proven the following.

Theorem 2.25.

$$d(Adx + Bdy) = dA \wedge dx + dB \wedge dy.$$

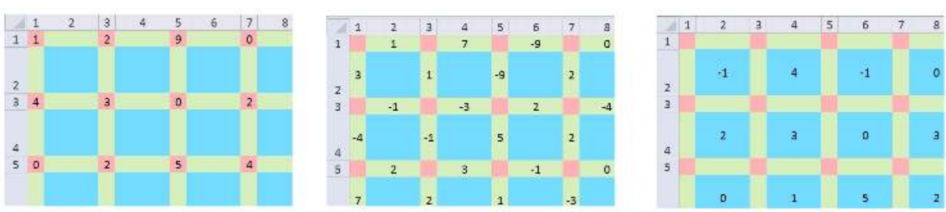
Exercise 2.26. Show how the result matches Green's Theorem.

In these two subsections, we see the same pattern: if $\varphi \in C^k$ then $d\varphi \in C^{k+1}$ and

- $d\varphi$ is obtained from φ by applying d to each of the coefficient functions involved in φ.

2.6 Representing cubical forms with a spreadsheet

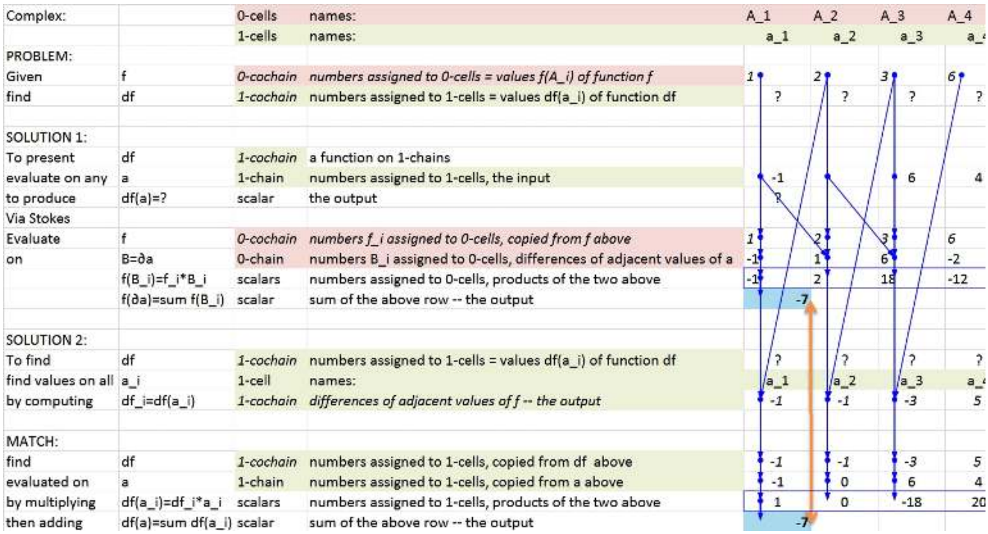
This is how 0-, 1-, and 2-forms are presented in a spreadsheet:



The difference of k-forms (as k-cochains) from k-chains is only that this time there are no blank k-cells!

The exterior derivative in dimensions 1 and 2 can be easily computed according to the formulas provided above. The only difference from the algebra we have seen is that here we have to present the results in terms of the coordinates with respect to the cells. They are listed at the top and on the left.

The case of $\mathbb{R}$ is explained below. The computation is shown on the right and explained on the left:



Complex:		0-cells	names:	A_1		A_2		A_3		A_4	
		1-cells	names:		a_1		a_2		a_3		a_
PROBLEM:											
Given	f	0-cochain	numbers assigned to 0-cells = values f(A_i) of function f	1		2		3		6	
find	df	1-cochain	numbers assigned to 1-cells = values df(a_i) of function df		?		?		?		?
SOLUTION 1:											
To present	df	1-cochain	a function on 1-chains								
evaluate on any	a	1-chain	numbers assigned to 1-cells, the input		-1				6		4
to produce	df(a)=?	scalar	the output		?						
Via Stokes											
Evaluate	f	0-cochain	numbers f_i assigned to 0-cells, copied from f above	1		2		3		6	
on	B=∂a	0-chain	numbers B_i assigned to 0-cells, differences of adjacent values of a	-1		1		6		-2	
	f(B_i)=f_i*B_i	scalars	numbers assigned to 0-cells, products of the two above	-1		2		18		-12	
	f(∂a)=sum f(B_i)	scalar	sum of the above row -- the output		-7						
SOLUTION 2:											
To find	df	1-cochain	numbers assigned to 1-cells = values df(a_i) of function df		?		?		?		?
find values on all	a_i	1-cell	names:		a_1		a_2		a_3		a_
by computing	df_i=df(a_i)	1-cochain	differences of adjacent values of f -- the output		-1		-1		-3		5
MATCH:											
find	df	1-cochain	numbers assigned to 1-cells, copied from df above		-1		-1		-3		5
evaluated on	a	1-chain	numbers assigned to 1-cells, copied from a above		-1		0		6		4
by multiplying	df(a_i)=df_i*a_i	scalars	numbers assigned to 1-cells, products of the two above		1		0		-18		20
then adding	df(a)=sum df(a_i)	scalar	sum of the above row -- the output		-7						

The Excel formulas are hidden but only these two need to be explained:

• first, "$B=\partial a$, 0-chain numbers B_i assigned to 0-cells, differences of adjacent values of a" is computed by

```
= R[-4]C - R[-4]C[-1]
```

• second, "$df\_i = df(a\_i)$, 1-cochain, differences of adjacent values of f – the output" is computed by

```
= R[-16]C - R[-16]C[1]
```

Thus, the exterior derivative is computed in two ways. We can see how the results match.

Note: See the files online.

Exercise 2.27. Create a spreadsheet for "antidifferentiation".

Exercise 2.28. (a) Create a spreadsheet that computes the exterior derivative of 1-forms in $\mathbb{R}^2$ directly. (b) Combine it with the spreadsheet for the boundary operator to confirm the Stokes Theorem.

2.7 A bird's-eye view of calculus

We have now access to a bird's-eye view of the topological part of discrete calculus, as follows.

Suppose we are given the cubical grid $\mathbb{R}^n$ of $\mathbf{R}^n$. On this complex, we have the vector spaces of k-chains C_k. Combined with the boundary operator ∂, they form the chain complex $\{C_*, \partial\}$ of K:

$$0 \xrightarrow{\partial_{n+1}=0} C_n \xrightarrow{\partial_n} \dots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0=0} 0.$$

The next layer is the cochain complex $\{C^*, d\}$, formed by the vector spaces of forms $C^k = (C_k)^*,\ k=0,1,\dots$:

$$0 \xleftarrow{d=0} C^n \xleftarrow{d} \dots \xleftarrow{d} C^0 \xleftarrow{d=0} 0.$$

Here d is the exterior derivative. The latter diagram is the "dualization" of the former as explained above:

$$d\varphi(x) = \varphi\partial(x).$$

The shortest version of this formula is as follows.

Theorem 2.29 (Stokes Theorem). The exterior derivative is the dual of the boundary operator:

$$\boxed{\partial^* = d}$$

Rather than using it as a theorem, we have used it as a formula that defines the exterior derivative.

The main properties of the exterior derivative follow.

Theorem 2.30. The operator $d : C^k \to C^{k+1}$ is linear.

Theorem 2.31 (Product Rule - Leibniz Rule)

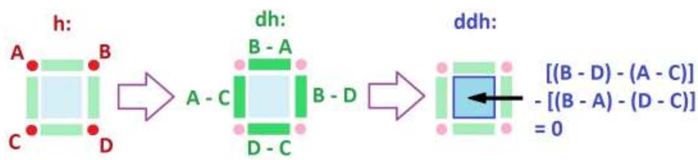
$$d(\varphi \wedge \psi) = d\varphi \wedge \psi + (-1)^k \varphi \wedge d\psi.$$

Exercise 2.32. Prove the theorem for dimension 2.

Theorem 2.33 (Double Derivative Identity). $dd : C^k \to C^{k+2}$ is zero.

Proof. We prove only $dd = 0 : C^0(\mathbb{R}^2) \to C^2(\mathbb{R}^2)$.

Suppose A, B, C, D are the values of a 0-form h at these vertices:



We compute the values of dh on these edges, as differences. We have:

$$-(B - A) + (C - D) + (B - C) - (A - D) = 0,$$

where the first two are vertical and the second two are horizontal. ∎

In general, the property follows from the *double boundary identity*.

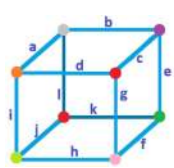
The proof indicates that the two mixed partial derivatives are equal:

$$\Phi_{xy} = \Phi_{yx},$$

just as in Clairaut's Theorem.

Exercise 2.34. Prove $dd : C^1(\mathbb{R}^3) \to C^3(\mathbb{R}^3)$ is zero.

Exercise 2.35. Compute $dd : C^1 \to C^3$ for the following form:



The actual (non-trivial) second derivative is discussed later.

2.8 Algebraic properties of the exterior derivative

We start with the properties of forms that hold on an arbitrary complex.

The exterior derivative is linear:

$$d(\alpha f + \beta g) = \alpha df + \beta dg,$$

for all forms f, g and all $\alpha, \beta \in \mathbf{R}$. Therefore, we have the following two familiar facts:

Theorem 2.36 (Sum Rule). For any two k-forms f, g, we have

$$d(f + g) = df + dg.$$

Theorem 2.37 (Constant Multiple Rule). For any k-form f and any $c \in \mathbf{R}$, we have:

$$d(cf) = c \cdot df.$$

For the next two, we limit the functions to the ones defined on the real line.

Theorem 2.38 (Power Formula). On complex $\mathbb{R}$, we have for any positive integer k:

$$d(x^{\underline{k}})\Big([n, n+1]\Big) = kn^{\underline{k-1}},$$

where

$$p^{\underline{q}} := p(p-1)(p-2)(p-3)...(p-q+1).$$

Theorem 2.39 (Exponent Formula). On complex $\mathbb{R}$, we have for any positive real b:

$$d(b^n)\Big([n, n+1]\Big) = (b-1)b^n.$$

Theorem 2.40 (Trig Formulas). On complex $\mathbb{R}$, we have for any real a:

$$\begin{aligned} d(\sin an)\Big([n, n+1]\Big) &= 2\sin\tfrac{a}{2}\cos a(n+\tfrac{1}{2}), \\ d(\cos an)\Big([n, n+1]\Big) &= -2\sin\tfrac{a}{2}\sin a(n+\tfrac{1}{2}). \end{aligned}$$

Theorem 2.41 (Product Rule). For any two 0-forms f, g on $\mathbb{R}$, we have

$$d(f \cdot g)\Big([n, n+1]\Big) = f(n+1)dg\Big([n, n+1]\Big) + df\Big([n, n+1]\Big)g(n).$$

Theorem 2.42 (Quotient Rule). For any two 0-forms f, g on $\mathbb{R}$, we have

$$d(f/g)([n, n+1]) = \frac{df\Big([n, n+1]\Big)g(n) - f(n)dg\Big([n, n+1]\Big)}{g(n)g(n+1)},$$

provided $g(n), g(n+1) \neq 0$.

Exercise 2.43. Prove these theorems.

Exercise 2.44. Derive the "Power Formula" for $k < 0$.

Exercise 2.45. Derive the "Log Formula": $d(\log_b an)\Big([n, n+1]\Big) =?$

Theorem 2.46 (Chain Rule 1). For an integer-valued 0-form g on complex K and any 0-form f on $\mathbb{R}$, we have
$$d(fg) = fdg.$$

Proof. Using the Stokes Theorem twice, with any 1-chain a, we have:
$$d(fg)(a) = (fg)(\partial a) = f(g(\partial a)) = f(dg)(a).$$ ■

Further, a cubical map $g : K \to \mathbb{R}$ generates a chain map
$$g_k : C_k(K) \to C_k(\mathbb{R}),\ k = 0, 1.$$

Then g_0 can be thought of as an integer-valued 0-form on K, and g_1 is, in a way, the derivative of g_0.

Theorem 2.47 (Chain Rule 2). For a cubical map $g : K \to \mathbb{R}$ and any 0-form f on $\mathbb{R}$, we have
$$d(fg_0) = dfg_1.$$

Proof. Using the Stokes Theorem once, with any 1-chain a, we have again:
$$d(fg_0)(a) = (fg_0)(\partial a) = f(g_0(\partial a)).$$

Now, using the algebraic continuity property $\partial g = g\partial$ and the Stokes Theorem, we conclude:
$$= f\partial(g_1(a)) = (df)(g_1(a)).$$ ■

In the last section, the idea of the derivative of a discrete function $f : \mathbf{Z} \to \mathbf{R}$ as a function of the *same* nature $f' : \mathbf{Z} \to \mathbf{R}$ was rejected on the grounds that this approach doesn't match our understanding of the integral as the area under the graph of f'. Certainly, there are other issues. Let's suppose the derivative is given by: $g'(x) = g(x+1) - g(x),\ x \in \mathbf{Z}$. Now, let's differentiate $h(x) = g(-x)$. We have: $h'(0) = h(1) - h(0) = g(-1) - g(0)$. On the other hand, $-g'(0) = -(g(1) - g(0)) = g(0) - g(1)$, no match! There is no chain rule in such a "calculus".

Exercise 2.48. Verify that there is such a match for the derivatives of these functions, if we see them as 1-cochains, and confirm both of the versions of the chain rule.

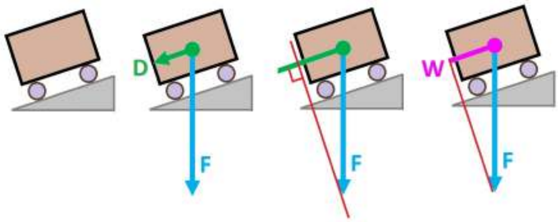
2.9 Tangent spaces

Suppose we want to find the *work* of a constant force along a straight path. As we know,
$$\text{work} \ = \ \text{force} \ \cdot \ \text{distance}.$$

This simple formula only works if we carefully take into account the *direction* of motion relative to the direction of the force F. For example, if you move forward and then back, the work breaks into two parts and they may cancel each other. The idea is that the work W may be positive or negative and we should speak of the *displacement* D rather than the distance. We then amend the formula:
$$W = F \cdot D.$$

Now, in the context of discrete calculus, the displacement D may be given by a single oriented edge in $\mathbb{R}$, or a combination of edges. It is a 1-*chain*. Furthermore, the force F defines W as a linear function of D. It is a 1-*form*!

The need for considering directions becomes clearer when the dimension of the space is 2 or higher. We use *vectors*. First, as we just saw, the work of the force is $W = \pm F \cdot D$ if $F||D$, and we have the plus sign when the two are collinear. Second, $W = 0$ if $F \perp D$. Therefore, only the *projection* of F on D matters when calculating the work and it is the projection when the length of D is 1.



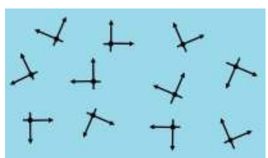
Then the work W of force F along vector D is defined to be

$$W := \langle F, D\rangle.$$

It is simply a (real-valued) linear function of the displacement.

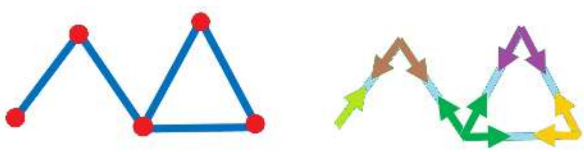
Our conclusion doesn't change: D is a 1-chain and F is a 1-form. Even though this idea allows us to continue our study, the example shows that it is impossible to limit ourselves to cubical complexes. Below, we make a step toward discrete calculus over general cell complexes.

On a plane, the force F may vary from location to location. Then the need to handle the displacement vectors, i.e., directions, arises, separately, at every point. The set of all possible directions at point $A \in V = \mathbf{R}^2$ form a vector space of the same dimension. It is V_A, a copy of V, attached to each point A:



Next, we apply this idea to cell complexes.

First, what is the set of all possible directions on a *graph*? We've come to understand the edges starting from a given vertex as independent directions. That's why we will need as many basis vectors as there are edges, at each point:



Of course, once we start talking about *oriented* cells, we know it's about *chains*, over R.

Definition 2.49. For each vertex A in a cell complex K, the *tangent space* at A of K is the set of 1-chains over ring R generated by the 1-dimensional star of the vertex A:

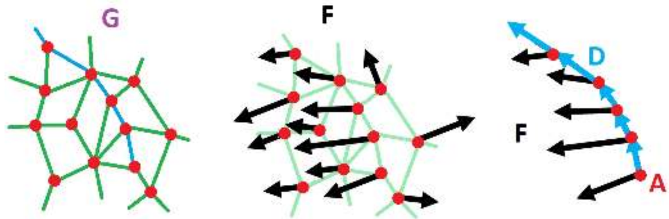
$$T_A = T_A(K) :=< \{AB \in K\} >\subset C_1(K).$$

Proposition 2.50. The tangent space $T_A(K)$ is a submodule of $C_1(K)$.

Definition 2.51. A *local* 1-*form* on K is a collection of linear functions for each of the tangent spaces,

$$\varphi_A : T_A \to R,\ A \in K^{(0)}.$$

The work of a force along edges of a cell complex is an example of such a function. Now, if we have a force in complex K that *varies* from point to point, the work along an edge – seen as the displacement – will depend on the location.



Proposition 2.52. Every 1-form (cochain) is a local 1-form.

We **denote** the set of all local 1-forms on K by $T^1(K)$, so that

$$C^1(K) \subset T^1(K).$$

Let's review the setup. First, we have the *space of locations* $X = K^{(0)}$, the set of all vertices of the cell complex K. Second, to each location $A \in X$, we associate the *space of directions* determined by the structure of K, specifically, by its edges. Now, while the directions at vertex $A \in K$ are given by the edges adjacent to A, we can also think of all 1-*chains* in the star of A as directions at A. They are subject to algebraic operations on chains and, therefore, form a module, T_A.

We now combine all the tangent spaces into one total tangent space. It contains all possible directions in each location: each tangent space T_A to every point A in K.

Definition 2.53. The *tangent bundle* of K is the disjoint union of all tangent spaces:

$$T(K) := \bigsqcup_{A \in K} \Big(\{A\} \times T_A \Big).$$

Then a local 1-form is seen as a function on the tangent bundle,

$$\varphi = \{\varphi_A : \ A \in K\} : T(K) \to R,$$

linear on each tangent space, defined by

$$\varphi(A, AB) := \varphi_A(AB).$$

In particular, the work associates to every location and every direction at that location, a quantity:

$$\varphi(A, AB) \in R.$$

The total work over a path in the complex is the *line integral* of φ over a 1-chain a in K. It is simply the sum of the values of the form at the edges in a:

$$\begin{array}{ll} a = & A_0A_1 + A_1A_2 + ... + A_{n-1}A_n \Longrightarrow \\ & \displaystyle\int_a \varphi := \varphi(A_0, A_0A_1) + \varphi(A_1, A_1A_2) + ... + \varphi(A_{n-1}, A_{n-1}A_n). \end{array}$$

Let's suppose, once again, that this number $\varphi(A, AB)$ is the work performed by a given force while moving from A to B along AB. We know that this work should be the negative of the one carried out while going in the opposite direction; i.e.,

$$\varphi(A, AB) = -\varphi(B, BA).$$

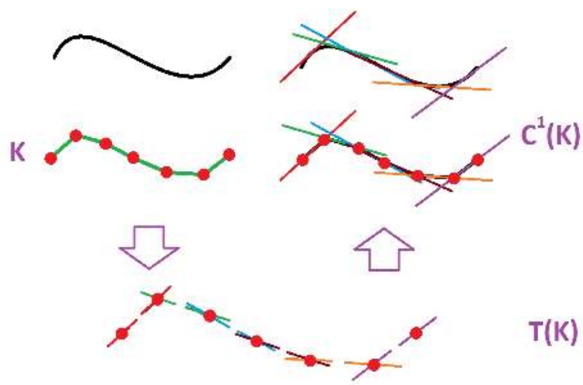
From the linearity of φ, it follows $\varphi(A, AB) = \varphi(B, AB)$. Then, the local form φ defined separately on each of the stars must have matched values on their overlaps. Therefore, it is well-defined as a linear map on $C_1(K)$.

Theorem 2.54. A local 1-form that satisfies the matching condition above is a 1-cochain of K.

This conclusion reveals a potential redundancy in the way we defined the space of all directions as the tangent bundle $T(K)$. We can then postulate that the direction from A to B is the opposite of the direction from B to A:

$$(A, AB) \sim -(B, BA).$$

This equivalence relation reduces the size of the tangent bundle via the quotient construction. Below, we can see the similarity between this new space and the space of tangents of a curve:



It looks as if the disconnected parts of $T(K)$, the tangent spaces, are glued together.

The fact that this equivalence relation preserves the operations on each tangent space implies the following.

Theorem 2.55.

$$T(K)/_{\sim} = C_1(K).$$

Exercise 2.56. Prove the theorem.

Then 1-forms are local 1-forms that are well-defined on $C_1(K)$. More precisely, the following diagram commutes for any $f \in C^1(K)$:

$$\begin{array}{ccc}
T(K) & \xrightarrow{f} & R \\
\downarrow{\scriptstyle p} & & || \\
C_1(K) & \xrightarrow{f} & R,
\end{array}$$

where p is the identification map. This justifies our focus on 1-forms as 1-cochains.

Exercise 2.57. Define the analog of the exterior derivative $d : C^0(K) \to T^1(K)$.

Higher order differential forms are *multi*-linear functions but this topic lies outside the scope of this book.

---

3 Cohomology

3.1 Vectors and covectors

What is the relation between number 2 and the "doubling" function $f(x) = 2 \cdot x$?

Linear algebra helps one appreciate this seemingly trivial relation. The answer is given by a linear operator,

$$\mathbf{R} \to \mathcal{L}(\mathbf{R}, \mathbf{R}),$$

from the reals to the vector space of all linear functions on the reals. In fact, it's an isomorphism!

More generally, suppose

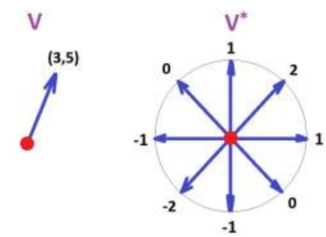
- R is a commutative ring, and
- V is a finitely generated free module over R.

Definition 3.1. We let the *dual module* of V to be

$$V^* := \{\alpha : V \to R,\ \alpha \text{ linear}\}.$$

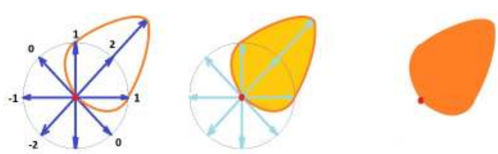
As the elements of V are called *vectors*, those of V^* are called *covectors*.

Example 3.2. An illustration of a vector in $v \in V = \mathbf{R}^2$ and a covector in $u \in V^*$ is given below:



Here, a vector is just a pair of numbers, while a covector is a match of each *unit* vector with a number. The linearity of the latter is visible. □

Exercise 3.3. Explain the alternative way a covector can be visualized as shown below. Hint: it resembles an oil spill.



Exercise 3.4. Prove that, if the spaces are finite-dimensional, we have

$$\dim \mathcal{L}(V, U) = \dim V \cdot \dim U.$$

Example 3.5. In the above example, it is easy to see a natural way of building a vector w from this covector u. Indeed, let's pick w such that

- the direction of w is that of the one that gives the largest value of the covector u (i.e., 2), and
- the magnitude of w is that value of u.

So the result is $w = (2, 2)$. Moreover, covector u can be reconstructed from this vector w. □

Exercise 3.6. What does this construction have to do with the norm of a linear operator?

In the spirit of this terminology, we might add "co" to any word to indicate its dual. Such is the relation between chains and cochains (forms). In that sense, 2 is a number and $2x$ is a "conumber", 2^*.

Exercise 3.7. What is a "comatrix"?

3.2 The dual basis

Proposition 3.8. The dual V^* of module V is also a module, with the operations ($\alpha, \beta \in V^*,\ r \in R$) given by

$$\begin{array}{ll}(\alpha+\beta)(v) & := \alpha(v)+\beta(v),\ v \in V;\\ (r\alpha)(w) & := r\alpha(w),\ w \in V.\end{array}$$

Exercise 3.9. Prove the proposition for $R = \mathbf{R}$. Hint: start by indicating what $0, -\alpha \in V^*$ are, and then refer to theorems of linear algebra.

Below we assume that V is finite-dimensional. Suppose also that we are given a basis $\{u_1, ..., u_n\}$ of V.

Definition 3.10. The *dual vector* $u_p^* \in V^*$ of vector $u_p \in V$ is defined by

$$u_p^*(u_i) := \delta_{ip},\ i = 1, ..., n;$$

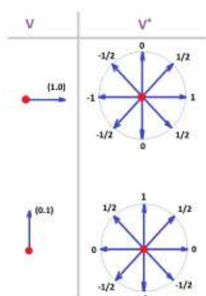
or

$$u_p^*(r_1u_1 + ... + r_nu_n) = r_p.$$

Exercise 3.11. Prove that $u_p^* \in V^*$.

Definition 3.12. The *dual* of the basis $\{u_1, ..., u_n\}$ of V is $\{u_1^*, ..., u_n^*\}$.

Example 3.13. The dual of the standard basis of $V = \mathbf{R}^2$ is shown below:



□

Let's prove that *the dual of a basis is a basis.* It takes two steps.

Proposition 3.14. The set $\{u_1^*, ..., u_n^*\}$ is linearly independent in V^*.

Proof. Suppose

$$s_1u_1^* + ... + s_nu_n^* = 0$$

for some $r_1, ..., r_k \in R$. This means that

$$s_1u_1^*(u) + ... + s_nu_n^*(u) = 0,$$

for all $u \in V$. For each $i = 1, ..., n$, we do the following. We choose $u := u_i$ and substitute it into the above equation:

$$s_1u_1^*(u_i) + ... + s_iu_i^*(u_i) + ... + s_nu_n^*(u_i) = 0.$$

Then we use $u_j^*(u_i) = \delta_{ij}$ to reduce the equation to:

$$s_1 0 + ... + s_i 1 + ... + s_n 0 = 0.$$

We conclude that $s_i = 0$. The statement of the proposition follows. ■

Proposition 3.15. The set $\{u_1^*, ..., u_n^*\}$ spans V^*.

Proof. Given $u^* \in V^*$, let's set $r_i := u^*(u_i) \in R,\ i = 1, ..., n$. Now define

$$v^* := r_1u_1^* + ... + r_nu_n^*.$$

Consider

$$v^*(u_i) = r_1u_1^*(u_i) + ... + r_nu_n^*(u_i) = r_i.$$

So the values of u^* and v^* match for all elements of the basis of V. Accordingly, $u^* = v^*$. ■

Exercise 3.16. Find the dual of $\mathbf{R}^2$ for two different choices of bases.

Corollary 3.17.

$$\dim V^* = \dim V = n.$$

Therefore, by the *Classification Theorem of Vector Spaces*, we have the following:

Corollary 3.18.

$$V^* \cong V.$$

Even though *a module is isomorphic to its dual*, the behaviors of the linear operators on these two spaces aren't "aligned", as we will show. Moreover, the isomorphism is dependent on the choice of basis.

The relation between a module and its dual is revealed if we look at vectors as column-vectors (as always) and covectors as row-vectors:

$$V = \left\{ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right\}, \quad V^* = \Big\{ y = [y_1, ..., y_n] \Big\}.$$

Then we can multiply the two as matrices:

$$y \cdot x = [y_1, ..., y_n] \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1y_1 + ... + x_ny_n.$$

As before, we utilize the similarity to the dot product and, for $x \in V, y \in V^*$, represent y evaluated at x as

$$\langle x, y \rangle := y(x).$$

This isn't the dot product or an inner product, which is symmetric. It is called a *pairing*:

$$\langle \cdot , \cdot \rangle : V^* \times V \to R,$$

which is linear with respect to either of the components.

Exercise 3.19. Show that the pairing is independent of a choice of basis.

Exercise 3.20. When V is *infinite*-dimensional with a fixed basis γ, its dual is defined as the set of all linear functions $\alpha : V \to R$ that are equal to zero for all but a finite number of elements of γ. (a) Prove the infinite-dimensional analogs of the results above. (b) Show how they fail to hold if we use the original definition.

3.3 The dual operators

Next, we need to understand what happens to a linear operator

$$A : V \to W$$

under duality. The answer is uncomplicated but also unexpected, as the corresponding dual operator goes in the opposite direction:

$$A^* : V^* \leftarrow W^*.$$

This isn't just because of the way we chose to define it:

$$A^*(f) := fA;$$

a dual counterpart of A can't be defined in any other way! Consider this diagram:

$$\begin{array}{ccc} V & \xrightarrow{A} & W \\ {\scriptstyle g\in V^*} & \searrow & \downarrow {\scriptstyle f\in W^*} \\ & & R, \end{array}$$

where R is our ring. If this is understood as a commutative diagram, the relation between f and g is given by the equation above. Therefore, we acquire g from f (by setting $g = fA$) and not vice versa.

Furthermore, the diagram also suggests that the reversal of the arrows has nothing to do with linearity. The issue is "functorial".

We restate the definition in our preferred notation.

Definition 3.21. Given a linear operator $A : V \to W$, its *dual operator* $A^* : W^* \to V^*$, is given by

$$\langle A^* g, v \rangle = \langle g, Av \rangle,$$

for every $g \in W^*, v \in V$.

The behaviors of an operator and its dual are matched but in a non-trivial (dual) manner.

Theorem 3.22.
- (a) If A is one-to-one, then A^* is onto.
- (b) If A is onto, then A^* is one-to-one.

Proof. To prove part (a), observe that $\operatorname{Im} A$, just as any submodule in the finite-dimensional case, is a summand:

$$W = \operatorname{Im} A \oplus N,$$

for some submodule N of W. Consider some $f \in V^*$. Now, there is a unique representation of every element $w \in W$ as $w = w' + n$ for some $w' \in \operatorname{Im} A$ and $n \in N$. Therefore, there is a unique representation of every element $w \in W$ as $w = A(v) + n$ for some $v \in V$ and $n \in N$, since A is one-to-one. Then we can define $g \in W^*$ by $\langle g, w \rangle := \langle f, v \rangle$. Finally, we have:

$$\langle A^* g, v \rangle = \langle g, Av \rangle = \langle g, w - n \rangle = \langle g, w \rangle - \langle g, n \rangle = \langle f, v \rangle + 0 = \langle f, v \rangle .$$

Hence, $A^* g = f$. ■

Exercise 3.23. Prove part (b).

Theorem 3.24. The matrix of A^* is the transpose of that of A:

$$A^* = A^T .$$

Exercise 3.25. Prove the theorem.

The compositions are preserved under the duality but in reverse:

Theorem 3.26.

$$(BA)^* = A^* B^* .$$

Proof. Consider:

$$\begin{array}{ccccc} V & \xrightarrow{A} & W & \xrightarrow{B} & U \\ {\scriptstyle g\in V^*} & \searrow & \Big\downarrow {\scriptstyle f\in W^*} & \swarrow & {\scriptstyle h\in U^*} \\ & & R. & & \end{array}$$

■

Exercise 3.27. Finish the proof.

As you see, the dual A^* behaves very much like, but not to be confused with, the inverse A^{-1}.

Exercise 3.28. When do we have $A^{-1} = A^T$?

The isomorphism between V and V^* is very straight-forward.

Definition 3.29. The *duality isomorphism* of the module V,

$$D_V : V \to V^* ,$$

is given by

$$D_V(u_i) := u_i^* ,$$

provided $\{u_i\}$ is a basis of V and $\{u_i^*\}$ is its dual.

In addition to the compositions, as we saw above, this isomorphism preserves the identity.

Theorem 3.30.

$$(\mathrm{Id}_V)^* = \mathrm{Id}_{V^*} .$$

Exercise 3.31. Prove the theorem.

Next, because of the reversed arrows, we can't say that this isomorphism "preserves linear operators". Therefore, the duality does not produce a *functor* as we know it but rather a new kind of functor discussed later in this section.

Now, for $A : V \to U$ a linear operator, the diagram below isn't commutative:

$$\begin{array}{ccc} V & \xrightarrow{A} & U \\ \downarrow{\scriptstyle D_V} & \neq & \downarrow{\scriptstyle D_U} \\ V^* & \xleftarrow{A^*} & U^* \end{array}$$

Exercise 3.32. Why not? Give an example.

Exercise 3.33. Show how a change of basis of V affects differently the coordinate representation of vectors in V and covectors in V^*.

However, the isomorphism with the *second dual*

$$V^{**} := (V^*)^*$$

given by

$$D_{V^*} D_V : V \cong (V^*)^*$$

does preserve linear operators, in the following sense.

Theorem 3.34. The following diagram is commutative:

$$\begin{array}{ccc} V & \xrightarrow{A} & U \\ \downarrow{\scriptstyle D_V} & & \downarrow{\scriptstyle D_U} \\ V^* & & U^* \\ \downarrow{\scriptstyle D_{V^*}} & & \downarrow{\scriptstyle D_{U^*}} \\ V^{**} & \xrightarrow{A^{**}} & U^{**} \end{array}$$

Exercise 3.35. (a) Prove the commutativity. (b) Demonstrate that the isomorphism is independent of the choice of basis of V.

Our **conclusion** is that we can think of the second dual (but not the first dual) of a module as the module itself:

$$V^{**} = V.$$

Same applies to the second duals of linear operators:

$$A^{**} = A.$$

Exercise 3.36. Confirm that if we replace the dot product above with any choice of inner product (to be considered later), the duality theory presented above remains valid.

3.4 The cochain complex

Recall that cohomology theory is the dual of homology theory and it starts with the concept of a k-*cochain* on a cell complex K. It is any linear function from the module of k-chains to R:

$$s : C_k(K) \to R.$$

Then the chains are the vectors and the cochains are the corresponding covectors.

We use the duality theory we have developed to define the module of k-cochains as the dual of the module of the k-chains:

$$C^k(K) := \Big(C_k(K)\Big)^*,$$

Further, the kth *coboundary operator* of K is the dual of the $(k+1)$st boundary operator:

$$\partial^k := \Big(\partial_k\Big)^* : C^k \to C^{k+1}.$$

It is given by the *Stokes Theorem*:

$$\langle \partial^k Q, a\rangle := \langle Q, \partial_{k+1} a\rangle,$$

for any $(k+1)$-chain a and any k-cochain Q in K.

Theorem 3.37. The matrix of the coboundary operator is the transpose of that of the boundary operator:

$$\partial^k = \Big(\partial_{k+1}\Big)^T.$$

Definition 3.38. The elements in $Z^k := \ker \partial^*$ are called *cocycles* and the elements of $B^k := \operatorname{Im} \partial^*$ are called *coboundaries*.

The following is a crucial result.

Theorem 3.39 (Double Coboundary Identity). Every coboundary is a cocycle; i.e., for $k = 0, 1, ...$, we have

$$\partial^{k+1}\partial^k = 0.$$

Proof. It follows from the fact that the coboundary operator is the dual of the boundary operator. Indeed,

$$\partial^*\partial^* = (\partial\partial)^* = 0^* = 0,$$

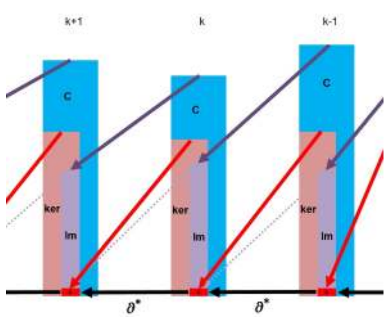
by the double boundary identity. ■

The cochain modules $C^k = C^k(K),\ k = 0, 1, ...,$ form the *cochain complex* $\{C^*, \partial^*\}$:

$$0 \xleftarrow{\ 0\ } C^N \xleftarrow{\ \partial^{N-1}\ } ... \xleftarrow{\ \partial^0\ } C^0 \xleftarrow{\ 0\ } 0.$$

According to the theorem, a cochain complex is a chain complex, just indexed in the opposite order.

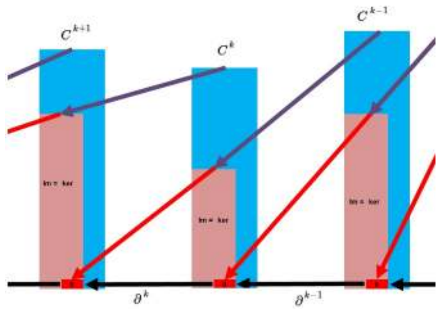
Our illustration of a cochain complex is identical to that of the chain complex but with the arrows reversed:



Recall that a cell complex K is called acyclic if its chain complex is an *exact sequence*:

$$\operatorname{Im} \partial_k = \ker \partial_{k-1}.$$

Naturally, if a cochain complex is exact as a chain complex, it is also called exact:



Exercise 3.40. (a) State the definition of an exact cochain complex in terms of cocycles and coboundaries. (b) Prove that $\{C^k(K)\}$ is exact if and only if $\{C_k(K)\}$ is exact.

To see the big picture, we align the chain complex and the cochain complex in one, non-commutative, diagram:

$$\begin{array}{ccccccc}
\dots & \xleftarrow{\partial^k} & C^k & \xleftarrow{\partial^{k-1}} & C^{k-1} & \xleftarrow{\partial^{k-1}} & \dots \\
\dots & & \downarrow\cong & \neq & \downarrow\cong & & \dots \\
\dots & \xrightarrow{\partial_{k+1}} & C_k & \xrightarrow{\partial_k} & C_{k-1} & \xrightarrow{\partial_k} & \dots
\end{array}$$

Exercise 3.41. (a) Give an example of a cell complex that demonstrates that the diagram doesn't have to be commutative. (b) What can you say about the chain complexes (and the cell complex) when this diagram *is* commutative?

3.5 Social choice: higher order voting

Recall that we previously considered n alternatives/candidates, $\{A, B, C, ...\}$, placed at the vertices of a simplicial complex K.

Let's consider some examples.

On the most basic level, voters *evaluate the alternatives.* For instance,

- voter a votes: candidate A is worth 1; translated:
- $a(A) = 1$.

Here, a 0-chain A is evaluated by a 0-cochain a. Then $a = A^*$.

Moreover, the voter may vote for 0-*chains* as linear combinations of the alternatives; for example, "the average of candidates A and B" is worth 1 means:

$$a\left(\frac{A+B}{2}\right) = 1.$$

In fact, we can understand this vote as a vote in support of a 50-50 lottery between A and B. Then,

$$a = \frac{1}{2}(A^* + B^*).$$

What if, furthermore, the voter would like to *compare* the candidates? For example,

- voter a votes: candidate A is worse, by 1, than candidate B; translated:
- $a(B) - a(A) = 1$.

Unfortunately, this vote might conflict with others of this kind as the vote may be circular:

$$a(B) - a(A) = 1,\ a(C) - a(B) = 1,\ a(A) - a(C) = 1.$$

The attempt to express this vote as an evaluation fails!

Exercise 3.42. Represent the following vote as a cochain:
- the average of candidates A and B is worse, by B, than candidate C.

When unable to evaluate, the voter may decide to only *compare pairs of alternatives* to each other. For example,
- voter a votes: candidate A is worse, by 1, than candidate B; translated:
- $a(AB) = 1$.

Fortunately, this vote won't conflict with others of this kind, even when the vote is circular:

$$a(AB) = 1,\ a(BC) = 1,\ a(CA) = 1.$$

Once again, the voter may also vote for 1-*chains* as linear combinations of the comparisons. For example, the voter may judge:
- the sum of the advantages of candidate B over A, C over B, and C over A is 1; translated:
- $a(AB) + a(BC) + a(CA) = 1$.

Unfortunately, this vote might conflict with others of this kind:

$$\begin{array}{ll} a(AB) + a(BC) + a(CA) = 1, & a(AD) + a(DB) + a(BA) = 1, \\ a(BD) + a(CC) + a(CB) = 1, & a(CD) + a(DA) + a(AC) = 1. \end{array}$$

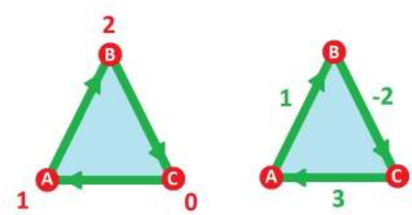
The attempt to express this vote as a (pair-wise) comparison fails!

When unable to compare pair-wise, the voter may decide to only *compare the triples of alternatives*. For example,
- the sum of the advantages of candidate B over A, C over B, and C over A is 1; translated:
- $a(ABC) = 1$.

Fortunately, this vote won't conflict with others of this kind.

And so on...

In the case of three candidates, there are three votes of degree 0 and three votes of degree 1:



Let's sum up. These are possible votes on these groups of candidates:
- A_0 is evaluated by a vote of degree 0: $a^0 \in C^0$ and $a^0(A_0) \in R$.
- A_0, A_1 are evaluated by a vote of degree 1: $a^1 \in C^1$ and $a^1(A_0A_1) \in R$.
- A_0, A_1, A_2 are evaluated by a vote of degree 2: $a^2 \in C^2$ and $a^2(A_0A_1A_2) \in R$.
- ...
- $A_0, A_1, ..., A_k$ are evaluated by a vote of degree k: $a^k \in C^k$ and $a^k(A_0A_1...A_k) \in R$.
- ...

Definition 3.43. A *vote* is any cochain in complex K:

$$a = a^0 + a^1 + a^2 + ...,\ a^i \in C^i(K).$$

Now, how do we make sense of the outcome of such a vote? *Who won*?

In the simplest case, we ignore the higher order votes, $a^1, a^2, ...$, and choose the *winner* to be the one with the highest rating:

$$winner := \arg \max_{i \in K^{(0)}} a^0(i).$$

But do we really have to discard the information about pairwise comparisons? Not necessarily: if a^1 is a rating comparison vote (i.e., a coboundary), we can use it to create a new set of ratings:

$$b^0 := (\partial^0)^{-1}(a^1).$$

We then find the candidate(s) with the highest value of

$$c^0 := a^0 + b^0$$

to become the winner.

Example 3.44. Suppose $a^0 = 1$ and

$$a^1(AB) = 1,\ a^1(BC) = -1,\ a^1(CA) = 0.$$

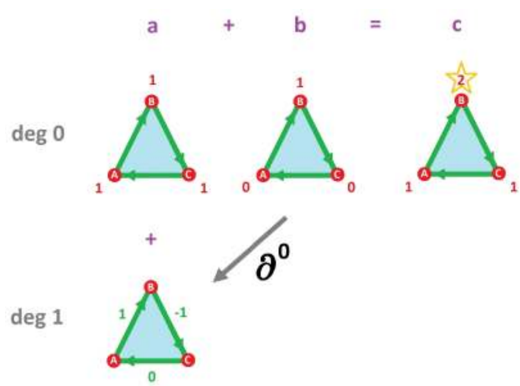
We choose:

$$b^0(A) = 0,\ b^0(B) = 1,\ b^0(C) = 0,$$

and observe that

$$\partial^0 b^0 = a^1.$$

Therefore, B is the winner!



□

Thus, the comparison vote helps to break the tie.

Exercise 3.45. Show that the definition is guaranteed to apply only when $K^{(1)}$ is a tree.

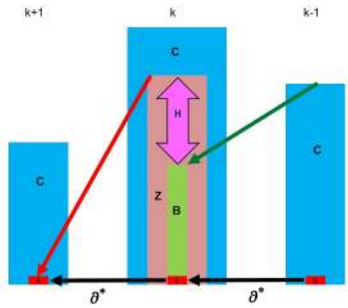
Exercise 3.46. Give other examples of how a^1 helps determine the winner when $a^0 = 0$.

The inability of the voter to assign a number to each *single* candidate to convey his perceived value (or quality or utility or rating) is what makes comparison of *pairs* necessary. By a similar logic, the inability of the voter to assign a number to each *pair* of candidates to convey their relative value is what could make comparison of *triples* necessary. Even when this is the case we still need to find a single winning candidate! Simply put, can $a^2 \neq 0$ help determine the winner when $a^0 = 0, a^1 = 0$? Unfortunately, there is no $b^0 \in C^0$ such that $\partial^1 \partial^0 b^0 = a^2$.

3.6 Cohomology

Definition 3.47. The kth *cohomology group* of K is the quotient of the cocycles over the coboundaries:

$$H^k = H^k(K) := Z^k/B^k = \ker \partial^{k+1} / \operatorname{Im} \partial^k.$$



It is then the homology of the "reversed" cochain complex.

Most of the theorems about homology have corresponding theorems about cohomology. Often, the latter can be derived from the former via duality. Sometimes it is unnecessary.

We will now discuss cohomology in the context of connectedness and simple connectedness.

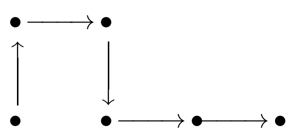
Proposition 3.48. The constant functions are 0-cocycles.

Proof. If $\varphi \in C^0(\mathbb{R})$, then $\partial^* \varphi\Big([a, a+1]\Big) = \varphi(a+1) - \varphi(a) = 0$. A similar argument applies to $C^0(\mathbb{R}^n)$. ∎

Proposition 3.49. The 0-cocycles are constant on a path-connected complex.

Proof. For $\mathbb{R}^1$, we have $\partial^* \varphi\Big([a, a+1]\Big) = 0$, or $\varphi(a+1) - \varphi(a) = 0$. Therefore, $\varphi(a+1) = \varphi(a)$; i.e., φ doesn't change from a vertex to the next. ∎

The general case is illustrated below:



Exercise 3.50. Prove that a complex is path-connected if and only if any two vertices can be connected by a sequence of adjacent edges. Use this fact to finish the proof of the proposition.

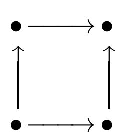
Corollary 3.51. $\dim \ker \partial^0 = \#$ of path-components of $|K|$.

We summarize the analysis as follows.

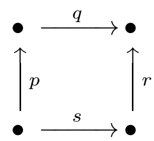
Theorem 3.52. $H^0(K) \cong R^m$, where m is the number of path-components of $|K|$.

Now, simple-connectedness.

Example 3.53 (circle). Consider a cubical representation of the circle:



Here, the arrows indicate the orientations of the edges – along the coordinate axes. A 1-cochain is just a combination of four numbers:

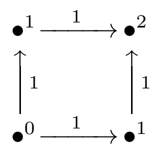


First, which of these cochains are cocycles? According to our definition, they should have "horizontal difference - vertical difference" equal to 0:

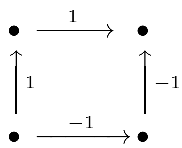
$$(r - p) - (q - s) = 0.$$

We could choose them all equal to 1.

Second, what 1-cochains are coboundaries? Here is a 0-cochain and its coboundary:



So, this 1-form is a coboundary. But this one isn't:



This fact is easy to prove by solving a little system of linear equations, or we can simply notice that the complete trip around the square adds up to 4 not 0.

Therefore, $H^1 \neq 0$. □

We accept the following without proof (see Bredon, *Topology and Geometry*, p. 172).

Theorem 3.54. If $|K|$ is simply connected, $H^1(K) = 0$.

These topological results for homology and cohomology match! Does it mean that homology and cohomology match in other dimensions? Consider first the fact that duality gives as the isomorphism:

$$C_k \cong C^k.$$

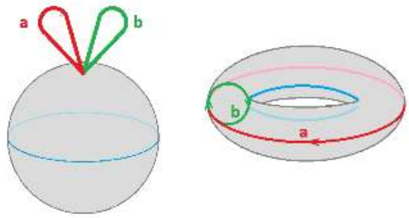
Second, the quotient construction of cohomology is identical to the one that defined homology. However, this doesn't mean that the resulting quotients are also isomorphic; in general, we have:

$$H_k \ncong H^k.$$

The main cause is that the quotient construction, over $\mathbf{Z}$, may produce *torsion* components for both homology and cohomology. Those components often don't match, as in the case of the Klein bottle considered in the next subsection.

A more subtle difference is that the cohomology isn't just a module; it also has a *graded ring* structure provided by the wedge product.

Example 3.55 (sphere with bows). To see the difference this algebra makes, consider these two spaces: the sphere with two bows attached and the torus:



Their homology groups coincide in all dimensions. The cohomology groups also coincide as vector spaces! The basis elements in dimension 1 behave differently under the wedge product. For the sphere with bows, we have:

$$[a^*] \wedge [b^*] = 0,$$

because there is nowhere for this 2-cochain to "reside". Meanwhile for the torus, we have:

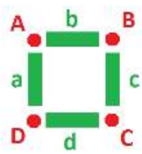
$$[a^*] \wedge [b^*] \neq 0.$$

□

3.7 Computing cohomology

In the last subsection, we used cochains to detect the hole in the circle. Now, we compute the cohomology without shortcuts – the above theorems – just the way a computer would.

Example 3.56 (circle). We go back to our cubical complex K and compute $H^1(K)$:



Consider the cochain complex of K:

$$C^0 \xrightarrow{\partial^0} C^1 \xrightarrow{\partial^1} C^2 = 0.$$

Observe that, due to $\partial^1 = 0$, we have $\ker \partial^1 = C^1$.

Let's list the bases of these vector spaces. Naturally, we start with the bases of the groups of chains C_k, i.e., the cells:

$$\{A, B, C, D\}, \ \{a, b, c, d\};$$

and then write their dual bases for the groups of cochains C^k:

$$\{A^*, B^*, C^*, D^*\}, \ \{a^*, b^*, c^*, d^*\};$$

They are given by

$$A^*(A) = \langle A^*, A \rangle = 1, ...$$

We find next the *formula* for ∂^0, a linear operator, which is its 4×4 matrix. For that, we just look at what happens to the basis elements:

$$A^* = [1,0,0,0]^T = \begin{array}{ccc} 1 & - & 0 \\ | & & | \\ 0 & - & 0 \end{array} \Longrightarrow \partial^0 A^* = \begin{array}{ccc} \bullet & -1 & \bullet \\ 1 & & 0 \\ \bullet & 0 & \bullet \end{array} = a^* - b^* = [1,-1,0,0]^T;$$

$$B^* = [0,1,0,0]^T = \begin{array}{ccc} 0 & - & 1 \\ | & & | \\ 0 & - & 0 \end{array} \Longrightarrow \partial^0 B^* = \begin{array}{ccc} \bullet & 1 & \bullet \\ 0 & & 1 \\ \bullet & 0 & \bullet \end{array} = b^* + c^* = [0,1,1,0]^T;$$

$$C^* = [0,0,1,0]^T = \begin{array}{ccc} 0 & - & 0 \\ | & & | \\ 0 & - & 1 \end{array} \Longrightarrow \partial^0 C^* = \begin{array}{ccc} \bullet & 0 & \bullet \\ 0 & & -1 \\ \bullet & 1 & \bullet \end{array} = -c^* + d^* = [0,0,-1,1]^T;$$

$$D^* = [0,0,0,1]^T = \begin{array}{ccc} 0 & - & 0 \\ | & & | \\ 1 & - & 0 \end{array} \Longrightarrow \partial^0 D^* = \begin{array}{ccc} \bullet & 0 & \bullet \\ -1 & & 0 \\ \bullet & -1 & \bullet \end{array} = -a^* - d^* = [-1,0,0,-1]^T.$$

The matrix of ∂^0 is formed by the column vectors listed above:

$$\partial^0 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

Now, from this data, we find the kernel and the image using the standard linear algebra.

The kernel is the solution set of the equation $\partial^0 v = 0$. It may be found by solving the corresponding system of linear equations with the coefficient matrix ∂^0 such as the Gaussian elimination (we can also notice that the rank of the matrix is 3 and conclude that the dimension of the kernel is 1). Therefore,

$$\dim H^0 = 1.$$

We have a single component!

The image is the set of u's with $\partial^0 v = u$. Once again, Gaussian elimination is an effective approach (we can also notice that the dimension of the image is the rank of the matrix, 3). In fact,

$$\operatorname{span}\left\{ \partial^0(A^*), \partial^0(B^*), \partial^0(C^*), \partial^0(D^*) \right\} = \operatorname{Im} \partial^0.$$

Hence,

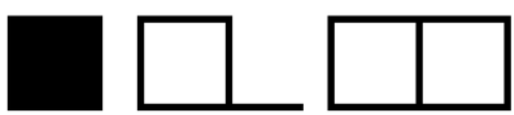
$$\dim H^1 = \dim \left(C^1 / \operatorname{Im} \partial^0 \right) = 4 - 3 = 1.$$

We have a single hole!

Another way to see that the columns aren't linearly independent is to compute $\det \partial^0 = 0$. □

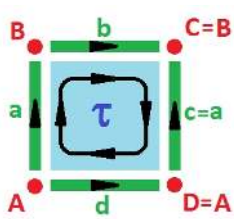
Exercise 3.57. Compute the cohomology of the T-shaped graph.

Exercise 3.58. Compute the cohomology of the following complexes: (a) the square, (b) the mouse, and (c) the figure 8, shown below.



Example 3.59 (cylinder). We build the cylinder C via an equivalence relation of cells of the cell complex:

$$a \sim c; \ A \sim D, \ B \sim C.$$



The chain complex is known, the cochain complex is derived, and the cohomology is computed accordingly:

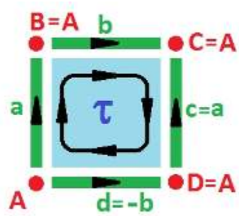
$$\begin{array}{llllll}
 & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \\
 & <\tau> & \xrightarrow{?} & <a,b,d> & \xrightarrow{?} & <A,B> \\
 & \tau & \mapsto & b-d & & \\
 & & & a & \mapsto & B-A \\
 & & & b & \mapsto & 0 \\
 & & & d & \mapsto & 0 \\
 & C^2 & \xleftarrow{\partial^*} & C^1 & \xleftarrow{\partial^*} & C^0 \\
 & <\tau^*> & \xleftarrow{?} & <a^*,b^*,d^*> & \xleftarrow{?} & <A^*,B^*> \\
 & 0 & \leftarrow\!\shortmid & a^* & & \\
 & \tau^* & \leftarrow\!\shortmid & b^* & & \\
 & -\tau^* & \leftarrow\!\shortmid & d^* & & \\
 & & & -a^* & \leftarrow\!\shortmid & A^* \\
 & & & a^* & \leftarrow\!\shortmid & B^* \\
\text{kernels :} & Z^2 =<\tau^*> & & Z^1 =<a^*, b^*-d^*> & & Z^0 =<A^*+B^*> \\
\text{images :} & B^2 =<\tau^*> & & B^1 =<a^*> & & B^0 =<0> \\
\text{quotients :} & H^2 = 0 & & H^1 =<[b^*]-[d^*]> \cong \mathbf{Z} & & H^0 =<[A^*]+[B^*]> \cong \mathbf{Z}
\end{array}$$

So, the cohomology is identical to that of the circle. □

In these examples, the cohomology is isomorphic to the homology. This isn't always the case.

Example 3.60 (Klein bottle). The equivalence relation on the complex of the square that gives $\mathbf{K}^2$ is:

$$c \sim a, \ d \sim -b.$$



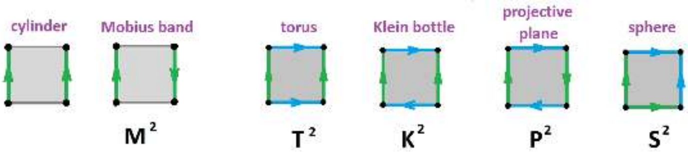
The chain complex is known, the cochain complex is derived, and the cohomology is computed accordingly:

$$\begin{array}{llllll}
 & C_2 & \xrightarrow{\partial} & C_1 & \xrightarrow{\partial} & C_0 \\
 & <\tau> & \xrightarrow{?} & <a,b> & \xrightarrow{?} & <A> \\
 & \tau & \mapsto & 2a & & \\
 & & & a & \mapsto & 0 \\
 & & & b & \mapsto & 0 \\
 & C^2 & \xleftarrow{\partial^*} & C^1 & \xleftarrow{\partial^*} & C^0 \\
 & <\tau^*> & \xleftarrow{?} & <a^*,b^*> & \xleftarrow{?} & <A^*> \\
 & 2\tau^* & \leftarrow\!\shortmid & a^* & & \\
 & 0 & \leftarrow\!\shortmid & b^* & & \\
 & & & 0 & \leftarrow\!\shortmid & A^* \\
\text{kernels :} & Z^2 =<\tau^*> & & Z^1 =<b^*> & & Z^0 =<A> \\
\text{images :} & B^2 =<2\tau^*> & & B^1 = 0 & & B^0 = 0 \\
\text{quotients :} & H^2 = \{\tau^* \mid 2\tau^* = 0\} \cong \mathbf{Z}_2 & & H^1 =<b^*> \cong \mathbf{Z} & & H^0 =<[A^*]> \cong \mathbf{Z}
\end{array}$$ □

Exercise 3.61. Show that $H^*(\mathbf{K}^2;\mathbf{R}) \cong H(\mathbf{K}^2;\mathbf{R})$.

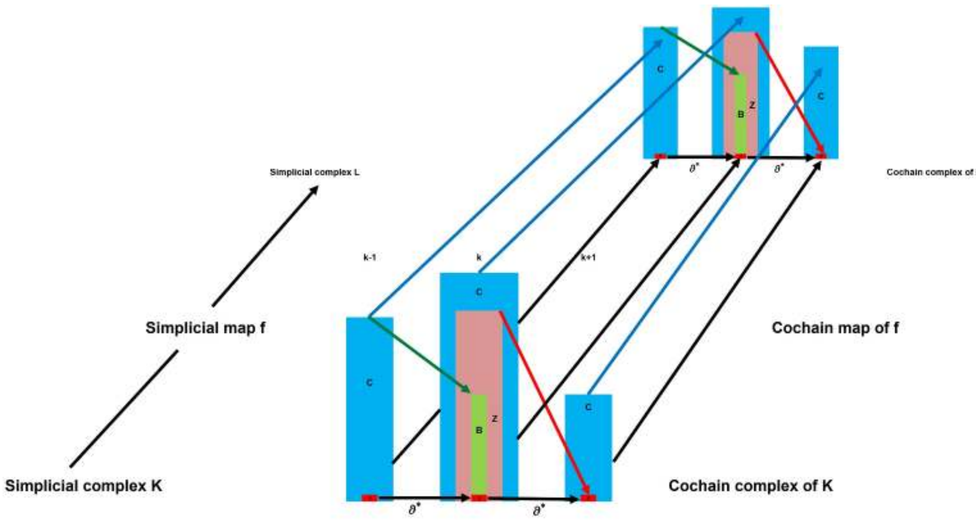
Exercise 3.62. Compute $H^*(\mathbf{K}^2;\mathbf{Z}_2)$.

Exercise 3.63. Compute the cohomology groups of the rest of these surfaces:



3.8 Homology maps vs. cohomology maps

Our illustration of cochain maps is identical to the one for chain maps but with the arrows reversed:



The idea is the same: *Homology and cohomology respect maps that respect boundaries.* The two constructions are also very similar and so are the results.

One can think of a function between two cell complexes

$$f : K \to L$$

as one that preserves cells; i.e., the image of a cell is also a cell, possibly of a lower dimension:

$$a \in K \Longrightarrow f(a) \in L.$$

If we expand this map by linearity, we have a map between (co)chain complexes:

$$f_k : C_k(K) \to C_k(L) \quad \text{and} \quad f^k : C^k(K) \leftarrow C^k(L).$$

What makes f_k and f^k continuous, in the algebraic sense, is that, in addition, they preserve (co)boundaries:

$$f_k(\partial a) = \partial f_k(a) \quad \text{and} \quad f^k(\partial^* s) = \partial^* f^k(s).$$

In other words, the (co)chain maps make these diagrams – the linked (co)chain complexes – commute:

$$\begin{array}{ccccccc}
\xleftarrow{\partial} & C_k(K) & \xleftarrow{\partial} & C_{k+1}(K) & \xleftarrow{\partial} \\
& \downarrow f_* & & \downarrow f_* & \\
\xleftarrow{\partial} & C_k(L) & \xleftarrow{\partial} & C_{k+1}(L) & \xleftarrow{\partial}
\end{array}
\quad \text{and} \quad
\begin{array}{ccccccc}
\xrightarrow{\partial^*} & C^k(K) & \xrightarrow{\partial^*} & C^{k+1}(K) & \xrightarrow{\partial^*} \\
& \uparrow f^* & & \uparrow f^* & \\
\xrightarrow{\partial^*} & C^k(L) & \xrightarrow{\partial^*} & C^{k+1}(L) & \xrightarrow{\partial^*}
\end{array}$$

They generate the (co)homology maps:

$$f_* : H_k(K) \to H_k(L) \quad \text{and} \quad f^* : H^k(K) \leftarrow H^k(L),$$

as the linear operators given by

$$f_*([x]) := [f_k(x)] \quad \text{and} \quad f^*([x]) := [f^k(x)],$$

where $[\cdot]$ stands for the (co)homology class.

Exercise 3.64. Prove that the latter is well-defined.

The following is obvious.

Theorem 3.65. The identity map induces the identity (co)homology map:

$$(\mathrm{Id}_{|K|})_* = \mathrm{Id}_{H(K)} \quad \text{and} \quad (\mathrm{Id}_{|K|})^* = \mathrm{Id}_{H^*(K)}\,.$$

This is what we derive from what we know about compositions of cell maps.

Theorem 3.66. The (co)homology map of the composition is the composition of the (co)homology maps

$$(gf)_* = g_* f_* \quad \text{and} \quad (gf)^* = f^* g^*.$$

Notice the change of order in the latter case!

This is the realm of category theory, explained later:

$$\begin{array}{ccccccccccc}
K & \xrightarrow{\quad f \quad} & L & & H(K) & \xrightarrow{\quad f_* \quad} & H(L) & & H^*(K) & \xleftarrow{\quad f^* \quad} & H^*(L) \\
& {}_{gf} \searrow & \downarrow{\scriptstyle g} & \rightsquigarrow & & {}_{g_* f_*} \searrow & \downarrow{\scriptstyle g_*} & \text{and} & & {}_{f^* g^*} \nwarrow & \uparrow{\scriptstyle g^*} \\
& & M & & & & H(M) & & & & H^*(M)
\end{array}$$

Theorem 3.67. Suppose K and L are cell complexes. If a map

$$f : |K| \to |L|$$

is a cell map and a homeomorphism, and

$$f^{-1} : |L| \to |K|$$

is a cell map too, then the (co)homology maps

$$f_* : H_k(K) \to H_k(L) \quad \text{and} \quad f^* : H^k(K) \leftarrow H^k(L),$$

are isomorphisms for all k.

Corollary 3.68. Under the conditions of the theorem, we have:

$$(f^{-1})_* = (f_*)^{-1} \quad \text{and} \quad (f^{-1})^* = (f^*)^{-1}.$$

Theorem 3.69. If two maps are homotopic, they induce the same (co)homology maps:

$$f \simeq g \implies f_* = g_* \quad \text{and} \quad f^* = g^*.$$

In the face of the isomorphisms of the groups and the matching behavior of the maps, let's not forget who came first:

$$\begin{array}{ccccc}
K & \rightsquigarrow & C(K) & \xrightarrow{D} & C^*(K) \\
& & \downarrow^{\sim} & & \downarrow^{\sim} \\
& & H(K) & \xleftrightarrow{?} & H^*(K).
\end{array}$$

Also, in spite of the fact that the cochain module is the dual of the chain module, we can't assume that the cohomology module is the dual of the homology module. In fact, we saw an example when they aren't isomorphic. However, the example was for the integer coefficients. What if these modules were vector spaces? We state the following without proof (see Bredon, *Topology and Geometry*, p. 282).

Theorem 3.70. If R is a field, the cohomology vector space is the dual of the homology vector space:

$$H(K;R)^* = H^*(K;R);$$

therefore, they are isomorphic:

$$H(K;R) \cong H^*(K;R).$$

Exercise 3.71. Define "cocohomology" and prove that it is isomorphic to homology.

3.9 Computing cohomology maps

We start with the most trivial examples and pretend that we don't know the answer...

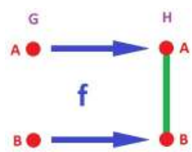
Example 3.72 (inclusion). Suppose:

$$G = \{A, B\},\ H = \{A, B, AB\},$$

and

$$f(A) = A,\ f(B) = B.$$

This is the inclusion:



On the (co)chain level we have:

$$\begin{array}{lll|lll}
C_0(G) =< A, B >, & C_0(H) =< A, B > & & \Longrightarrow C^0(G) =< A^*, B^* >, & C^0(H) =< A^*, B^* >, \\
f_0(A) = A, & f_0(B) = B & & \Longrightarrow f^0 = f_0^T = \text{Id}; & \\
C_1(G) = 0, & C_1(H) =< AB > & & \Longrightarrow C^1(G) = 0, & C^1(H) =< AB^* > \\
& & & \Longrightarrow f^1 = f_1^T = 0. &
\end{array}$$

Meanwhile, $\partial^0 = 0$ for G and for H we have:

$$\partial^0 = \partial_1^T = [-1, 1].$$

Therefore, on the cohomology level we have:

$$\begin{array}{llll}
H^0(G) := \frac{Z^0(G)}{B^0(G)} & = \frac{C^0(G)}{0} & =< [A^*], [B^*] >, \\
H^0(H) := \frac{Z^0(H)}{B^0(H)} & = \frac{\ker \partial^0}{0} & =< [A^* + B^*] >, \\
H^1(G) := \frac{Z^1(G)}{B^1(G)} & = \frac{0}{0} & = 0, \\
H^1(H) := \frac{Z^1(H)}{B^1(H)} & = \frac{0}{0} & = 0.
\end{array}$$

Then, for $f^k : H^k(H) \to H^k(G),\ k = 0, 1$, we compute from the definition:

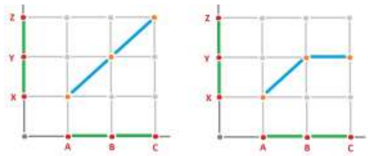
$$\begin{array}{lll}
[f^0]([A^*]) := [f^0(A^*)] = [A^*], & [f^0]([B^*]) := [f^0(B^*)] = [B^*] & \Longrightarrow [f^0] = [1, 1]^T; \\
f^1 = 0 & & \Longrightarrow [f^1] = 0.
\end{array}$$

The former identity indicates that A and B are separated if we reverse the effect of f. □

Exercise 3.73. Modify the computation for the case when there is no AB.

Exercise 3.74. Compute the cohomology maps for the following two two-edge simplicial complexes and these two simplicial maps:

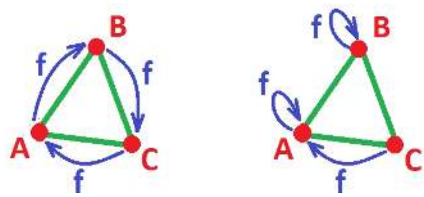
$$\begin{array}{ll} & K=\{A,B,C,AB,BC\},\ L=\{X,Y,Z,XY,YZ\};\\ \text{(a)} & f(A)=X,\ f(B)=Y,\ f(C)=Z,\ f(AB)=XY,\ f(BC)=YZ;\\ \text{(b)} & f(A)=X,\ f(B)=Y,\ f(C)=Y,\ f(AB)=XY,\ f(BC)=Y.\end{array}$$



Example 3.75 (rotation and collapse). Suppose we are given the following complexes and maps:

$$G=H:=\{A,B,C,AB,BC,CA\} \text{ and } f(A)=B,\ f(B)=C,\ f(C)=A.$$

Here is a rotated triangle (left):



The cohomology maps are computed as follows:

$$\begin{aligned} f^1(AB^*+BC^*+CA^*) &= f^1(AB^*)+f^1(BC^*)+f^1(CA^*)\\ &= CA^*+AB^*+BC^*\\ &= AB^*+BC^*+CA^*.\end{aligned}$$

Therefore, the cohomology map $[f^1]:H^1(H)\to H^1(G)$ is the identity. Conclusion: the hole is preserved.

Also (right) we collapse the triangle onto one of its edges:

$$f(A)=A,\ f(B)=B,\ f(C)=A.$$

Then,

$$\begin{aligned} f^1(AB^*+BC^*+CA^*) &= f^1(AB^*)+f^1(BC^*)+f_1(CA^*)\\ &= (AB^*+CB^*)+0+0\\ &= 2\partial^0(B^*).\end{aligned}$$

Therefore, the cohomology map is zero. Conclusion: collapsing of an edge causes the hole collapse too. □

Exercise 3.76. Provide the details of the computations.

Exercise 3.77. Modify the computation for the case (a) the map shown on the right, (b) a reflection and (c) a collapse to a vertex.

3.10 Functors, continued

The duality reverses arrows but preserves everything else. This idea deserves a functorial interpretation.

We generalize a familiar concept below.

Definition 3.78. A *functor* $\mathscr{F}$ from category $\mathscr{C}$ to category $\mathscr{D}$ consists of two functions:

$$\mathscr{F}: \mathrm{Obj}(\mathscr{C}) \to \mathrm{Obj}(\mathscr{D}),$$

and, if $\mathscr{F}(X) = U, \mathscr{F}(Y) = V$, we have, for a *covariant* functor:

$$\mathscr{F} = \mathscr{F}_{X,Y} : \mathrm{Hom}_{\mathscr{C}}(X,Y) \to \mathrm{Hom}_{\mathscr{D}}(U,V),$$

and for a *contravariant* functor:

$$\mathscr{F} = \mathscr{F}_{X,Y} : \mathrm{Hom}_{\mathscr{C}}(X,Y) \to \mathrm{Hom}_{\mathscr{D}}(V,U).$$

We assume that the functor preserves,
- the identity: $\mathscr{F}(\mathrm{Id}_X) = \mathrm{Id}_{\mathscr{F}(X)}$; and
- the compositions:
  - ⋄ covariant, $\mathscr{F}(gf) = \mathscr{F}(g)\mathscr{F}(f)$, or
  - ⋄ contravariant, $\mathscr{F}(gf) = \mathscr{F}(f)\mathscr{F}(g)$.

Exercise 3.79. Prove that duality is a contravariant functor. Hint: consider $\mathcal{L}(U,\cdot)$ and $\mathcal{L}(\cdot,V)$.

Exercise 3.80. What if we consider a composition of functors?

We will continue to refer to covariant functors as simply "functors" when there is no ambiguity.

The latter condition can be illustrated with these commutative diagrams:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ gf & \searrow & \downarrow g \\ & & Z \end{array} \quad \xrightarrow{\mathcal{F}} \quad \begin{array}{ccc} \mathscr{F}(X) & \xrightarrow{\mathscr{F}(f)} & \mathscr{F}(Y) \\ \mathscr{F}(gf) & \searrow & \downarrow \mathscr{F}(g) \\ & & \mathscr{F}(Z) \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathscr{F}(X) & \xleftarrow{\mathscr{F}(f)} & \mathscr{F}(Y) \\ \mathscr{F}(gf) & \nwarrow & \uparrow \mathscr{F}(g) \\ & & \mathscr{F}(Z) \end{array}$$

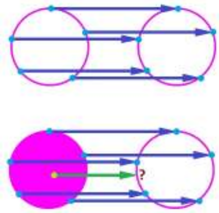
Previously we proved the following:

Theorem 3.81.
- Homology is a covariant functor from the polyhedra and their maps to the modules and their homomorphisms.
- Cohomology is a contravariant functor from the polyhedra and their maps to the modules and their homomorphisms.

Exercise 3.82. Show that homology theory (and cohomology theory) are functors from the relative cell complexes and the cell maps of pairs. Hint: $C^*(K,K') = C^*(K)/C^*(K')$.

Cohomology is just as good a functorial tool as homology. Below, we apply it to a problem previously solved with homology.

Example 3.83. We consider the question, "Can a soap bubble contract to the ring without tearing?" and recast it as an example of the Extension Problem, "Can we extend the map of a circle onto another circle to the whole disk?"



Can we find a continuous F to complete the first diagram below? With the cohomology functor, we answer: only if we can complete the last one.

$$\begin{array}{ccc} \mathbf{S}^{n-1} & \hookrightarrow & \mathbf{B}^n \\ & {}_{Id}\searrow & \uparrow F=? \\ & & \mathbf{S}^{n-1} \end{array} \quad \leadsto \quad \begin{array}{ccc} H^{n-1}(\mathbf{S}^{n-1}) & \leftarrow & H^{n-1}(\mathbf{B}^n) \\ & {}_{Id}\nwarrow & \uparrow ? \\ & & H^{n-1}(\mathbf{S}^{n-1}) \end{array} \quad = \quad \begin{array}{ccc} \mathbf{Z} & \xrightarrow{H^*} & 0 \\ & {}_{Id}\nwarrow & \uparrow ? \\ & & \mathbf{Z} \end{array}$$

Just as in the case of homology, we see that the last diagram is impossible to complete. □

Exercise 3.84. Provide such analysis for the extension problem(s) for: the identity map of the circle to the torus.

Exercise 3.85. List all possible maps for each of these description based on the possible cohomology maps:
- embeddings of the circle into the torus;
- self-maps of the figure eight;
- embeddings of the circle into the sphere.

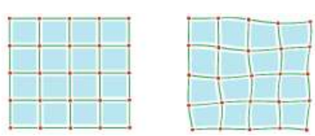
---

4 Metric tensor

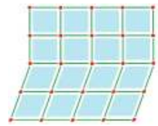
4.1 The need for measuring in calculus

We will learn how to introduce an arbitrary geometry into the (so far) purely topological setting of cell complexes.

It is much easier to separate topology from geometry in the framework of cell complexes (and discrete calculus) because we have built it from scratch! To begin with, it is obvious that only the way the cells are attached to each other affects the matrix of the boundary operator (and the exterior derivative):



It is then clear that the sizes or the shapes of the cells are topologically irrelevant. Note that, even when those are given, the geometry of the domain will remain unspecified unless the angles are also provided:



These two grids can be thought of as different (but homeomorphic) realizations of the same cell complex.

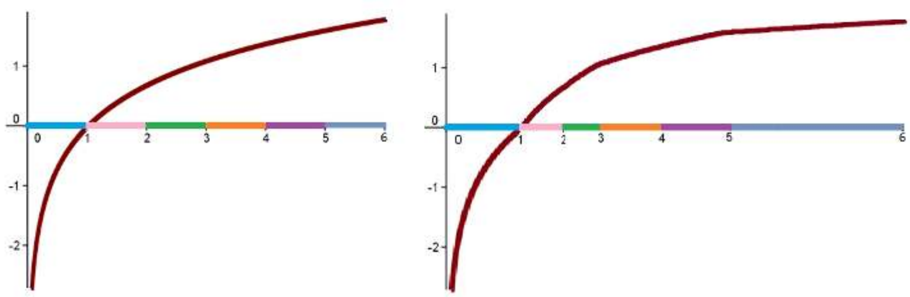
In calculus, we rely only on the topological and algebraic structures to develop the mathematics. There are two main exceptions. One needs the dot product and the norm for the following related concepts:

- the concavity,
- the arc-length, and
- the curvature. We'll need to introduce more structure – geometric in nature – to the complex.

Of course, the norm is explicitly present in the formulas for the last two. But first we will consider the role of geometry in the meaning of concavity.

The point where we start to need geometry is when we move from the first to the second derivative.

This is what we observe: shrinking, stretching, or deforming the x- or the y-axes won't change the monotonicity of a function but it will change its concavity. Below, we have deformed the x-axis and changed the concavity of the graph at $x = 1$:



Of course, it is sufficient to know the sign of the derivative to distinguish between increasing and decreasing behavior. Therefore, this behavior depends only on the topology of the spaces, in the above sense. In fact, the chain rule implies the following elementary result.

Proposition 4.1. Monotonicity is preserved under compositions with increasing functions; i.e., given a differentiable function $f : \mathbf{R} \to \mathbf{R}$, we have:

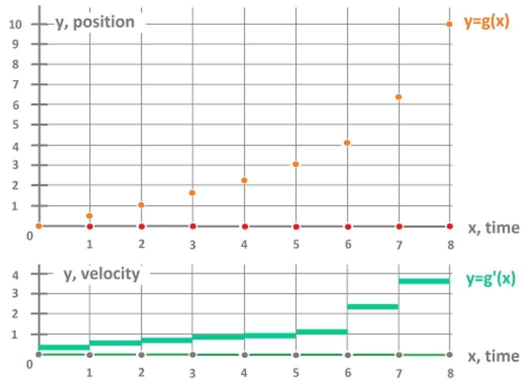
$$f \nearrow \iff fr \nearrow,$$

for any differentiable $r : \mathbf{R} \to \mathbf{R}$ with $r' > 0$.

Exercise 4.2. Give an example of f and r as above so that f and fr have different concavity.

Exercise 4.3. Restate the proposition in the context of discrete calculus.

Similarly, in the discrete calculus we are developing, the sign of the exterior derivative will tell us the difference between increasing and decreasing behavior. Concavity is an important concept that captures the shape of the graph of a function in calculus as well as the acceleration in physics. The upward concavity of the discrete function below is obvious, which indicates a positive acceleration:



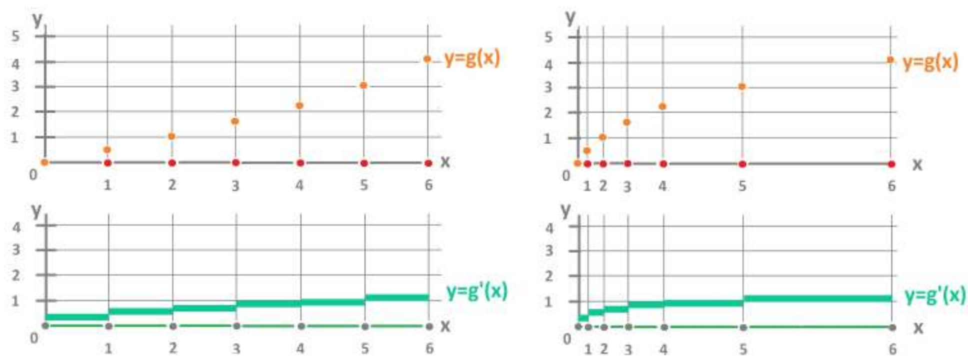
The concavity is computed and studied via the second derivative, the rate of change of the rate of change. But, with $dd = 0$, we don't have the second *exterior* derivative...

We start over and look at the *change of change*. This quantity can be seen algebraically as the increase of the exterior derivative (i.e., the change):

$$g(2) - g(1) < g(3) - g(2),\ g(3) - g(2) < g(4) - g(3),\ etc.$$

But do these inequalities imply concavity? Not without assuming that the intervals have equal lengths!

Below, on the same, topologically, cell complex, we have the same discrete function with, of course, the same exterior derivative:



...but with the opposite concavity!

4.2 Inner products

In order to develop a complete discrete calculus we need to be able compute:

- the lengths of vectors and
- the angles between vectors.

In linear algebra, we learn how an inner product adds geometry to a vector space. We choose a more general setting.

Suppose R is our ring of coefficients and suppose V is a finitely generated free module over R.

Definition 4.4. An *inner product* on V is a function that associates a number to each pair of vectors in V:

$$\langle\cdot,\cdot\rangle : \begin{cases} V \times V \to R, \\ (u,v) \mapsto \langle u,v\rangle, \end{cases}$$

that satisfies these properties:

- 1. $\diamond$ $\langle v,v\rangle \geq 0$ for any $v \in V$ – *non-degeneracy*,
  $\diamond$ $\langle v,v\rangle = 0$ if and only if $v = 0$ – *positive definiteness*;
- 2. $\langle u,v\rangle = \langle v,u\rangle$ – *symmetry* (commutativity);
- 3. $\langle ru,v\rangle = r\langle u,v\rangle$ – *homogeneity*;
- 4. $\langle u+u',v\rangle = \langle u,v\rangle + \langle u',v\rangle$ – *distributivity.*

Items 3 and 4 together make up *bilinearity.* Indeed, consider $p : V \times V \to R$ given by $p(u,v) = \langle u,v\rangle$. Then p is linear with respect to the first variable, and the second variable, separately:

- fix $v = b$, then $p(\cdot,b) : V \to R$ is linear;
- fix $u = a$, then $p(a,\cdot) : V \to R$ is linear.

Exercise 4.5. Show that this isn't the same as linearity.

It is easy to verify these axioms for the *dot product* defined on $V = \mathbf{R}^n$ or any other module with a fixed basis. For

$$u = (u_1,...,u_n),\ v = (v_1,...,v_n) \in \mathbf{R}^n,$$

define

$$\langle u,v\rangle := u_1v_1 + u_2v_2 + ... + u_nv_n.$$

Moreover, a *weighted dot product* on $\mathbf{R}^n$ is given by

$$\langle u,v\rangle := w_1u_1v_1 + w_2u_2v_2 + ... + w_nu_nv_n,$$

where $w_i \in \mathbf{R},\ i = 1,...,n$, are the positive "weights", is also an inner product.

Exercise 4.6. Prove the last statement.

A module equipped with an inner product is called an *inner product space.*

When ring R is a subring of the reals $\mathbf{R}$, such as the integers $\mathbf{Z}$, we can conduct some useful computations with some familiar formulas.

Definition 4.7. The *norm of a vector* v in an inner product space V is defined as

$$\|v\| := \sqrt{\langle v,v\rangle} \in \mathbf{R}.$$

This number measures the length of the vector.

Definition 4.8. The *angle between vectors* $u,v \neq 0$ in V is defined as

$$\cos\widehat{uv} := \frac{\langle u,v\rangle}{\|u\|\cdot\|v\|}.$$

The norm satisfies certain properties that we can also use as *axioms of a normed space.*

Definition 4.9. Given a vector space V, a *norm* on V is a function

$$\|\cdot\| : V \to R$$

that satisfies

- 1. $\diamond$ $\|v\| \geq 0$ for all $v \in V$,
  $\diamond$ $\|v\| = 0$ if and only if $v = 0$;
- 2. $\|rv\| = |r|\|v\|$ for all $v \in V, r \in R$;
- 3. $\|u+v\| \leq \|u\| + \|v\|$ for all $u,v \in V$.

Exercise 4.10. Prove the propositions below.

Proposition 4.11. A normed space is a metric space with the metric $d(x,y) := \|x-y\|$.

Proposition 4.12. A normed space is a topological vector space.

Exercise 4.13. Prove that the inner product is continuous in this metric space.

This is what we know from linear algebra.

Theorem 4.14. Any inner product $\langle \cdot, \cdot \rangle$ on an n-dimensional module V can be computed via matrix multiplication

$$\langle u, v \rangle = u^T Q v,$$

where Q is some positive definite, symmetric $n \times n$ matrix.

In particular, the dot product is represented by the identity matrix I_n, while the weighted dot product is represented by the diagonal matrix:

$$Q = \begin{pmatrix} w_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & w_n \end{pmatrix}.$$

Below we assume that $R = \mathbf{R}$.

To learn about the eigenvalues and eigenvectors of matrix Q of the inner product, we observe that

$$||v||^2 = v^T Q v = v^T \lambda v = \lambda ||v||^2.$$

We have proven the following:

Theorem 4.15. The eigenvalues of a positive-definite matrix are real and positive.

We know from linear algebra that if the eigenvalues are real and *distinct*, the matrix is diagonalizable. As it turns out, having distinct eigenvalues isn't required for a positive-definite matrix. We state the following without proof.

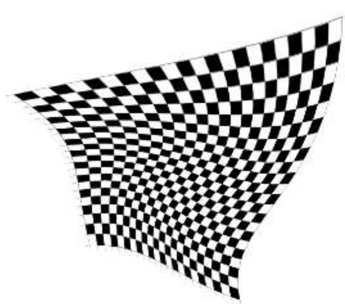
Theorem 4.16. Any inner product $\langle \cdot, \cdot \rangle$ on an n-dimensional vector space V can be represented, via a choice of basis, by a diagonal $n \times n$ matrix Q with positive numbers on the diagonal. In other words, every inner product is a weighted dot product, in some basis.

Also, since the determinant $\det Q$ of Q is the product of the eigenvalues of Q, it is also positive and we have the following:

Theorem 4.17. Matrix Q that represents an inner product is invertible.

4.3 The metric tensor

What if the geometry – as described above – varies from point to point in some space X? Then the inner product of two vectors will depend on their location.



In fact, we'll need to look at a different set of vectors at each location as the angle between two vectors is meaningless unless they have the same origin.

We already have such a construction! For each vertex A in a cell complex K, the *tangent space* at A of K is a submodule of $C_1(K)$ generated by the 1-dimensional star of A:

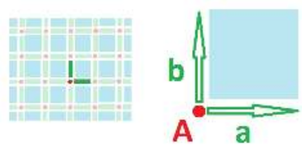
$$T_A(K) :=< \{AB \in K\} >\subset C_1(K).$$

However, the tangent bundle $T(K)$ is inadequate because this time we need to consider *two directions* at a time!

Since $V = T_A$ is a module, an inner product can be defined (so far locally). Then we have a collection of bilinear functions for each of the vertices,

$$\psi_A : T_A(K)^2 \to R,\ \ A \in K^{(0)}.$$

Thus the setup is as follows. We have the *space of locations* $X = K^{(0)}$, the set of all vertices of a cell complex K, and, second, to each location $A \in X$, we associate the *space of pairs of directions* $T_A(K)^2$.



We now collect these "double" tangent spaces into one.

Definition 4.18. The *tangent bundle of degree* 2 of K is defined to be

$$T^2(K) := \bigsqcup_{A\in K} \Big(\{A\} \times T_A(K)^2 \Big).$$

Then a locally defined inner product is seen as a function on this space,

$$\psi = \{\psi_A\} : T^2(K) \to R,$$

bilinear on each tangent space, defined by

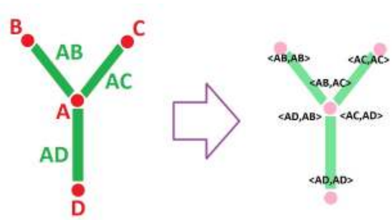
$$\psi(A, AB, AC) := \psi_A(AB, AC).$$

In other words, we have inner products "parametrized" by location.

Furthermore, we have a function that associates to each location, i.e., a vertex A in K, and each pair of directions at that location, i.e., edges $x = AB, y = AC$ adjacent to A with $B, C \neq A$, a number $\langle AB, AC\rangle$:

$$(A, AB, AC) \mapsto \langle AB, AC\rangle(A) \in R.$$

Note: we have used this notation for $f(a) = \langle f, a\rangle$, where f a form, but won't use in the rest of the book.



Definition 4.19. A *metric tensor* on a cell complex K is a collection of functions,

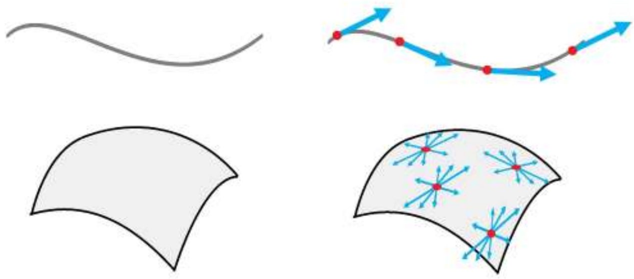
$$\psi = \{\psi_A : \ A \in X\},$$

such that each of them,

$$\psi_A : \{A\} \times T_A(K)^2 \to R,$$

satisfies the axioms of inner product, and also:

$$\psi_A(AB, AB) = \psi_B(BA, BA).$$



Using our notation, a metric tensor is such a function:

$$\langle \cdot, \cdot \rangle(\cdot) : T^2(K) \to R,$$

that, when restricted to any tangent space $T_A(K)$, $A \in K^{(0)}$, it produces a function

$$\langle \cdot, \cdot \rangle(A) : T_A(K)^2 \to R,$$

which is an inner product on $T_A(K)$. The last condition,

$$\langle AB, AB \rangle(A) = \langle BA, BA \rangle(B),$$

ensures that these inner products match!

Note: Compare to the standard setup: the location space X is an n-manifold; then $T_A X \cong \mathbf{R}^n$ is its tangent space, when $R = \mathbf{R}$. Then we have a function that associates a number to each location in X and each pair of vectors at that location:

$$\langle \cdot, \cdot \rangle(\cdot) : \begin{cases} X \times \mathbf{R}^n \times \mathbf{R}^n & \to \mathbf{R}, \\ (x, u, v) & \mapsto \langle u, v \rangle(x). \end{cases}$$

It is an inner product that depends continuously on location $x \in X$.

Exercise 4.20. State the axioms of inner product (non-degeneracy, positive-definiteness, symmetry, homogeneity, and distributivity) for the metric tensor.

The end result is similar to a combination of 1-forms, because we have a number assigned to each edge and to each pair of adjacent edges:

- $a \mapsto \langle a, a \rangle$;
- $(a, b) \mapsto \langle a, b \rangle$.

Then the metric tensor is given by the table of its values for each vertex A:

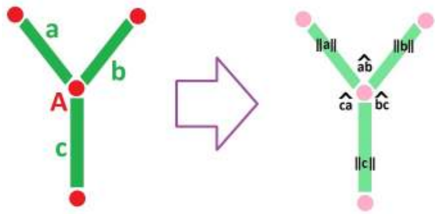
A	AB	AC	AD	...
AB	$\langle AB, AB \rangle$	$\langle AB, AC \rangle$	$\langle AB, AD \rangle$	...
AC	$\langle AC, AB \rangle$	$\langle AC, AC \rangle$	$\langle AC, AD \rangle$	...
AD	$\langle AD, AB \rangle$	$\langle AD, AC \rangle$	$\langle AD, AD \rangle$	...
...	...	...	...	...

Whenever our ring of coefficients happens to be a subring of the reals **R**, such as the integers **Z**, this data can be used to extract more usable information:

- $a \mapsto \|a\|$, and
- $(a, b) \mapsto \widehat{ab}$,

where

$$\cos \widehat{ab} = \frac{\langle a, b\rangle}{\|a\|\|b\|}.$$



All the information we need for measuring is contained in this symmetric matrix:

A	AB	AC	AD	...
AB	$\|AB\|$	$\widehat{BAC}$	$\widehat{BAD}$	...
AC	$\widehat{BAC}$	$\|AC\|$	$\widehat{CAD}$	...
AD	$\widehat{BAD}$	$\widehat{CAD}$	$\|AD\|$	...
...	...	...	...	...

Exercise 4.21. In the 1-dimensional cubical complex $\mathbb{R}$, are 01 and 10 parallel?

Proposition 4.22. From any matrix with positive entries on the diagonal, we can (re)construct a metric tensor by the formula:

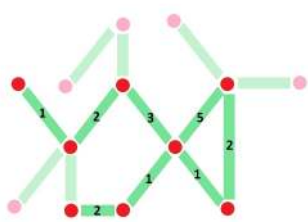
$$\langle AB, AC\rangle = \|AB\|\|AC\| \cos \widehat{BAC}.$$

With a metric tensor, we can define the two main geometric quantities of calculus in a very easy fashion. Suppose we have a *discrete curve* C in the complex K, i.e., a sequence of adjacent vertices and edges

$$C = \{A_0A_1, A_1A_2, ..., A_{N-1}A_N\} \subset K.$$

Definition 4.23. The *arc-length* of curve C is the sum of the lengths of its edges:

$$l_C := \|A_0A_1\| + \|A_1A_2\| + ... + \|A_{N-1}A_N\|.$$

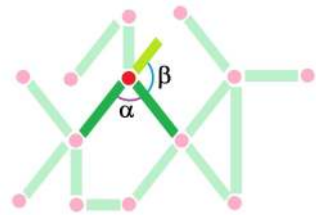


The arc-length is an example of a *line integral* of a 1-form ρ over a 1-chain a in complex K equipped with a metric tensor:

$$\begin{aligned} a = \ & A_0A_1 + A_1A_2 + ... + A_{n-1}A_n \Longrightarrow \\ & \int_a \rho := \rho(A_0, A_0A_1)\|A_0A_1\| + ... + \rho(A_{N-1}, A_{N-1}A_N)\|A_{N-1}A_N\|. \end{aligned}$$

In particular, if we have a curve made of rods represented by edges in our complex while ρ represents the linear density of each rod, the integral gives us the *weight* of the curve.

For the curvature, the angle $\alpha := \widehat{A_{s-1}A_sA_{s+1}}$ might appear a natural choice to represent it, but, since we are to capture the change of the direction, it should be the "outer" angle β:



Definition 4.24. The *curvature* of curve C at a given vertex A_s, $0 < s < N$, is the value of the angle

$$\kappa_C(A_s) := \pi - \widehat{A_{s-1}A_sA_{s+1}}.$$

As before, the result depends on our choice of a metric tensor.

The following is a simple but important observation.

Theorem 4.25. The set of vertices $K^{(0)}$ of a cell complex K equipped with a metric tensor is a metric space with its metric given by:

$$d(A,B) = \min\{l_C : C \text{ a curve from } A \text{ to } B\},$$

for any two vertices A, B.

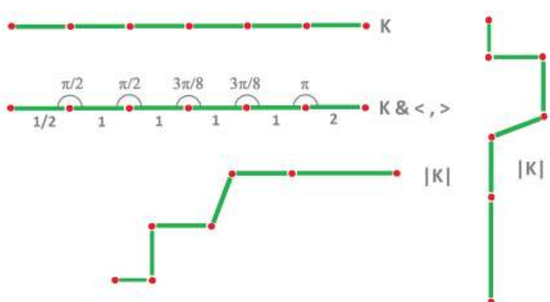
Exercise 4.26. (a) Prove the theorem. (b) What is the relation between the topology of $|K|$ and this metric space?

4.4 Metric tensors in dimension 1

Let's consider examples of how adding a metric tensor turns a cell complex, which is a discrete representation of a topological space, into something more *rigid*.

Example 4.27. In the illustration below you see the following:

- a cubical complex K, a *topological* entity: cells and their boundaries also made of cells;
- complex K combined with a choice of metric tensor $\langle\cdot,\cdot\rangle$, a *geometric* entity: lengths and angles;
- a realization of K in $\mathbf{R}$ as a segment;
- a realization of K in $\mathbf{R}^2$;
- another realization of K in $\mathbf{R}^2$.



The three realizations are homeomorphic but they differ in the way they geometrically fit into the plane, even though the last two have the same lengths and angles. □

Definition 4.28. A *geometric realization of dimension* n of a cell complex K equipped with a metric tensor is a realization $|K| \subset \mathbf{R}^n$ of K such that the Euclidean metric tensor of $\mathbf{R}^n$ matches this metric tensor; i.e.,

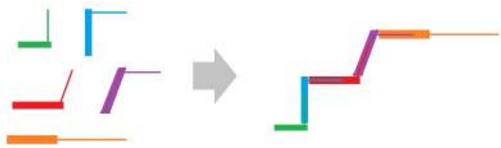
$$\langle a, b\rangle_K = \langle r(a), r(b)\rangle_{\mathbf{R}^n},$$

where r is the realization.

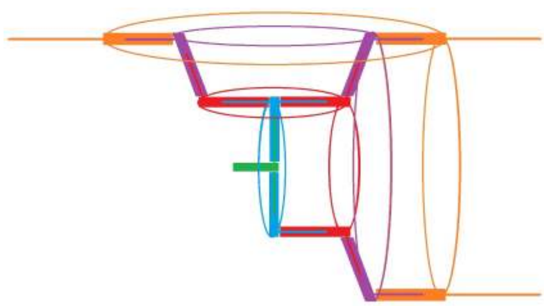
In the last example, we see a 1-dimensional cell complex with two different geometric realizations of dimension 2.

There is even more room for variability if we consider realizations in $\mathbf{R}^3$. Indeed, even with fixed angles, one can rotate the edges freely with respect to each other. To illustrate this idea, we consider a *mechanical interpretation* of a realization. We think of $|K|$ as if it is constructed of pieces in such a way that

- each edge is a tube with one opening and a rod rigidly attached to the other end, and
- at each vertex, a rod is inserted into the next tube.

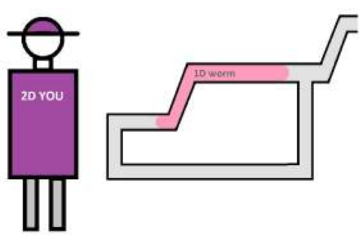


The rods are free to rotate – independent of each other – while still preserving the lengths and angles:



Just as since the very beginning we have studied intrinsic topological properties, now we are after only the *intrinsic geometric properties* of these objects, i.e., the ones that you can detect while staying inside the object.

Example 4.29. In a 1-dimensional complex, one can think of a worm moving through a tube. The worm can only feel the distance that it has passed and the angle (but not the direction) at which its body bends:



Meanwhile, we can see the whole thing by stepping *outside*, to the second dimension.

But, can the worm tell a left turn from a right? □

Exercise 4.30. Show that the ability to rotate these rods allows us to make the curve flat, i.e., realized in $\mathbf{R}^2$.

These observations underscore the fact that a metric tensor may not represent unambiguously the geometry of the realization.

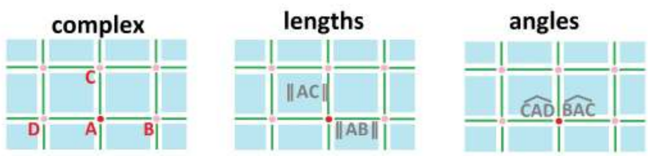
Exercise 4.31. (a) Provide metric tensors for a triangle, a square, and a rectangle. (b) Approximate a circle.

A metric tensor can also be thought of as a pair of maps:

- $AB \mapsto \|AB\|$;
- $(AB, AC) \mapsto \widehat{BAC}$.

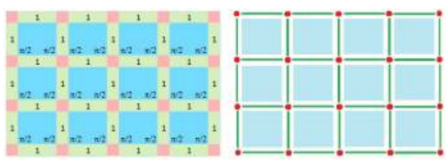
This means that for $\mathbf{R}^2$ there will be one (positive) number per edge and six per vertex. To simplify this a bit, we assume below that the opposite angles are equal. This is a type of local symmetry that lets us work with fewer parameters and yet allows for a sufficient variety of examples.

We can illustrate metric tensors by specifying the lengths and the angles:



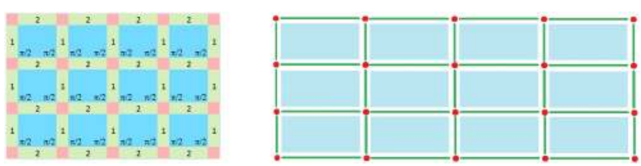
In the examples below, we put the metric data on the left over the complete cubical complex for the whole grid and then use it as a *blueprint* to construct its realization as a rigid structure, on the right. The 2-cells are ignored for now.

Example 4.32 (square grid). Here is the *standard metric cubical complex* $\mathbb{R}^2$:



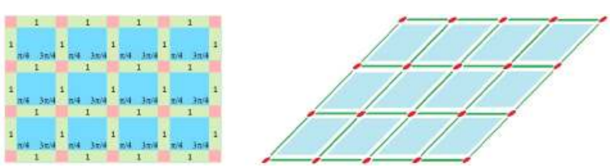
□

Example 4.33 (rectangular grid). Here is $\mathbb{R}^2$ horizontally stretched:



□

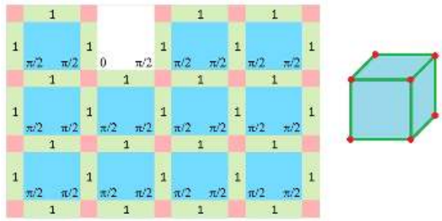
Example 4.34 (rhomboidal grid). Here is $\mathbb{R}^2$ skewed horizontally:



□

Exercise 4.35. Provide a realization of $\mathbb{R}^2$ with lengths: $1, 1, 1, ...$, and angles: $\pi/2, \pi/4, \pi/2, ...$.

Example 4.36 (cubical grid). We want a hollow cube build from the same squares:



Here we don't assume anymore that the opposite angles are equal, nor do we use the whole grid. □

Exercise 4.37. Provide the rest of the realization for the last example.

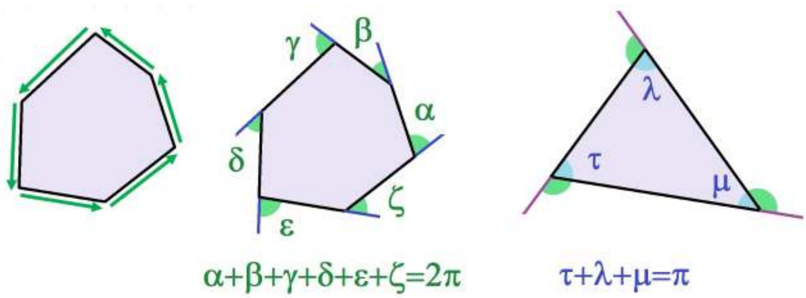
Suppose we have a discrete curve C in the complex K:

$$C := \{A_0A_1, A_1A_2, ..., A_{N-1}A_N\} \subset K,$$

that is closed: $A_N = A_0$. The *total curvature* of curve C is the sum of the curvatures at the vertices (except the endpoints):

$$\kappa_C := \sum_{s=1}^{N} \kappa_C(A_s) := \sum_{s=1}^{N} (\pi - \widehat{A_{s-1}A_sA_{s+1}}).$$

Theorem 4.38. Suppose a cubical or simplicial complex is realized in the plane and suppose a polygon is bounded by a closed discrete curve in this complex. Then the total curvature is equal to 2π.



Exercise 4.39. Prove the theorem.

In the case of a triangle, the sum of the *inner* angles is 3π minus the former. We have proved the following theorem familiar from middle school.

Corollary 4.40 (Sum of Angles of Triangle). It is equal to 180 degrees.

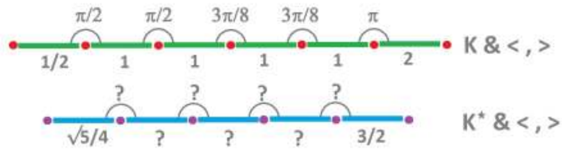
4.5 Metric complexes in dimension 1

Let's consider again the 1-dimensional complex K from the last subsection. With all of those rotations possible, how do we make this construction *more* rigid?

We start by considering another *mechanical interpretation* of a geometric realization $|K|$:

- we think of edges as *rods* and vertices as *hinges*.

Then there is a way to make the angles of the hinges fixed: we can connect the centers of the adjacent rods using an extra set of rods. These rods (connected by a new set of hinges) form a new complex **denoted** by $K^\star$.



Exercise 4.41. Assume that the vertices of the new complex are placed at the centers of the edges. (a) Find the rest of the lengths. (b) Find the angles too. (c) Find a metric tensor for $K^\star$.

The cells of the two complexes are matched:
- an edge a in K corresponds to a vertex $a^\star$ in $K^\star$; and
- a vertex A in K corresponds to an edge $A^\star$ in $K^\star$.

Furthermore, the correspondence (except for the endpoints) can be reversed!

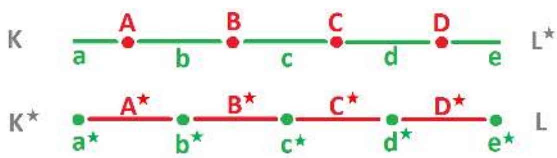
Then we have a one-to-one correspondence between the original, called *primal*, cells and the new, called *dual*, cells:

$$\begin{array}{rcl}\text{1-cell (primal)} & \longleftrightarrow & \text{0-cell (dual)}\\ \text{0-cell (primal)} & \longleftrightarrow & \text{1-cell (dual)}\end{array}$$

The combined correspondence is stated for $k = 0, 1$:

each primal k-cell corresponds to a dual $(1-k)$-cell, and vice versa.

Next, we assemble these new cells into a new complex:



Given a complex K, the set of all of the duals of the cells of K is the new complex $K^\star$.

Definition 4.42. Two 1-dimensional cell complexes K and $K^\star$ are called *Hodge-dual* of each other if there is a one-to-one correspondence between the k-cells of K and $(1-k)$-cells of $K^\star$, for $k = 0, 1$.

Notation: In order to avoid confusion between the familiar duality between vectors and covectors (and chains and cochains), known as the "Hom-duality", and the new, Hodge-duality, we will use the star $\star$ instead of the usual asterisk $*$. Both are contravariant functors.

Exercise 4.43. Define the boundary operator of the dual grid in terms of the boundary operator of the primal grid, in dimension 1.

What kinds of complexes are subject to this construction?

Suppose K is a graph and its vertex A has three adjacent edges AB, AC, AD. Then, in the dual complex $K^\star$, the edge $A^\star$ is supposed to join the vertices $AB^\star, AC^\star, AD^\star$. This is impossible, and therefore, such a complex has no dual.

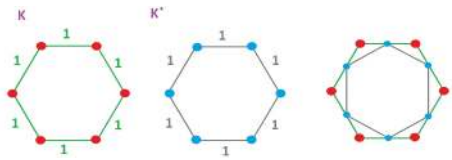
Then,
- both K and $K^\star$ have to be *curves*.

Furthermore, once constructed, $K^\star$ doesn't have to be a cell complex as some of the boundary cells may be missing. The simplest example is that the dual of a single vertex complex $K = \{A\}$ is a single edge "complex" $K^\star = \{A^\star\}$. The endpoints of this edge aren't present in $K^\star$. This is a typical example of the dual of a "closed" complex being an "open" complex. Such a complex won't have the boundary operator well-defined.

What's left? The complex K must be a complex representation of the infinite line or the circle:

$$\mathbb{R}^1 \text{ or } \mathbb{S}^1,$$

or the disjoint union of their copies.



In other words, they are 1-dimensional manifolds without boundary.

Proposition 4.44.

$$\left(\mathbb{R}^1\right)^\star = \mathbb{R}^1, \quad \left(\mathbb{S}^1\right)^\star = \mathbb{S}^1.$$

Exercise 4.45. Prove the proposition.

Up to this point, the construction of the dual $K^\star$ has been purely topological. The geometry comes into play when we add a metric tensor to the picture.

Definition 4.46. A *metric (cell) complex of dimension* 1 is a pair of a 1-dimensional cell complex K and its dual $K^\star$, either one equipped with a metric tensor.

Note that "dimension 1" also refers to the "ambient dimension" of the pair of complexes:

$$1 = \dim a + \dim a^\star.$$

Notation: Instead of $||AB||$, we use $|AB|$ for the length of 1-cell AB. The latter notation will also be used for the "volumes" of cells of all dimensions, including vertices.

Hodge duality is a correspondence between the *cells* of the two complexes; what about the *chains*?

As we have done many times, we could extend this correspondence, $a \mapsto a^\star$, by linearity from cells to chains. Because this is a bijection, the result is diagonal matrix with non-zero entries on the diagonal. The geometry of the complex is incorporated into these entries.

Definition 4.47. The (chain) *Hodge star operator* of a metric complex K is the pair of homomorphisms on chains of complementary dimensions:

$$\begin{array}{lll} \star: & C_k(K) \to C_{1-k}(K^\star), & k = 0, 1, \\ \star: & C_k(K^\star) \to C_{1-k}(K), & k = 0, 1, \end{array}$$

defined by

$$\star(a) := \frac{|a|}{|a^\star|} a^\star.$$

for any cell a, dual or primal, under the **convention**:

$$|A| = 1, \text{ for every vertex } A.$$

The choice of the coefficient of this operator is justified by the following crucial property.

Theorem 4.48 (Isometry). The Hodge star operator preserves lengths:

$$|\star(a)| = |a|.$$

Proposition 4.49. The two star operators are the inverses of each other;

$$\star\star = \operatorname{Id}.$$

Proposition 4.50. The matrix of the Hodge star operator $\star$ is diagonal with:

$$\star_{ii} = \frac{|a_i|}{|a_i^\star|},$$

where a_i is the ith cell of K.

4.6 Hodge duality of forms

What happens to the *forms* under the Hodge duality?

Since i-forms are defined on i-cells, $i = 0, 1$, the matching of the cells that we saw,

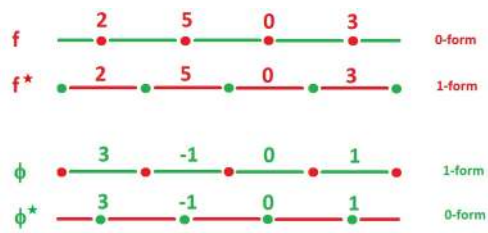
- primal i-cell $\Longleftrightarrow$ dual $(1-i)$-cell,

produces a matching of the forms,

- primal i-form $\Longleftrightarrow$ dual $(1-i)$-form.

For instance,

- If $f(A) = 2$ for some 0-cell A in K, there must be somewhere a 1-cell, say, $A^\star$ so we can have $f^\star(A^\star) = 2$ for a new form $f^\star$ in $K^\star$.
- If $\phi(a) = 3$ for some 1-cell a, there must be somewhere a 0-cell, say, $a^\star$ so we can have $\phi^\star(a^\star) = 3$.



To summarize,

- if f is a 0-form then $f^\star$ is a 1-form *with the same values* on the dual cells;
- if ϕ is a 1-form then $\phi^\star$ is a 0-form *with the same values* on the dual cells.

This is what is happening in dimension 1, illustrated with a spreadsheet:

Complex:	primal	0-cells	A_i	A_1	A_2	A_3	A_4
		1-cells	a_i	a_1	a_2	a_3	a_4
	dual	0-cells	dual of a_i	B_1	B_2	B_3	B_4
		1-cells	dual of A_i		b_2	b_3	b_4
PROBLEM:							
Given	f	*0-cochain, primal*	numbers assigned to primal 0-cells = values f(A_i) of function f	*1*	*2*	*3*	*6*
find	f*	*1-cochain, dual*	numbers assigned to dual 1-cells = values f*(b_i) of function f*		*2*	*3*	*6*
PROBLEM:							
Given	f	*1-cochain, primal*	numbers assigned to primal 1-cells = values f(a_i) of function f	*3*	*2*	*0*	*6*
find	f*	*0-cochain, dual*	numbers assigned to dual 0-cells = values f*(B_i) of function f*	*3*	*2*	*0*	*6*
PROBLEM:							
Given	f	*0-cochain, dual*	numbers assigned to primal 0-cells = values f(A_i) of function f	*1*	*2*	*3*	*6*
find	f*	*1-cochain, primal*	numbers assigned to dual 1-cells = values f*(b_i) of function f*		*2*	*3*	*6*
PROBLEM:							
Given	f	*1-cochain, dual*	numbers assigned to primal 1-cells = values f(a_i) of function f		*2*	*0*	*6*
find	f*	*0-cochain, primal*	numbers assigned to dual 0-cells = values f*(B_i) of function f*	*0*	*2*	*0*	*6*

Note: See the files online.

Algebraically, if f is a 0-form over the primal complex, we define a new form by

$$f^\star(a) := f(a^\star).$$

for any 1-cell a in the dual complex. And if ϕ is a 1-form over the primal complex, we define a new form by:

$$\phi^\star(A) := \phi(A^\star)$$

for any 0-cell A in the dual complex.

We put this together below.

Definition 4.51. For a metric complex K, the *Hodge-dual form* $\psi^\star$ of a k-form ψ over K is a $(1-k)$-form on $K^\star$ given by

$$\psi^\star(\sigma) := \psi(\sigma^\star),$$

where σ is a $(1-k)$-cell.

This formula will used as the definition of duality of form of all dimensions.

Naturally, the forms are extended from cells to chains by linearity making these diagrams commutative:

$$\begin{array}{ccc} C_k(K) & \xrightarrow{\ \star\ } & C_{1-k}(K^\star) \\ \downarrow \varphi & & \downarrow \varphi^\star \\ R & = & R \end{array} \qquad \begin{array}{ccc} C_k(K) & \xleftarrow{\ \star\ } & C_{1-k}(K^\star) \\ \downarrow \psi^\star & & \downarrow \psi \\ R & = & R \end{array}$$

Then the formula can be rewritten:

$$(\star\psi)(a) := \psi(\star a),$$

where $\star$ in the right-hand side is the chain Hodge star operator, for any primal/dual cochain ψ and any dual/primal chain a. We will use the same notation when there is no confusion.

Definition 4.52. The (cochain) *Hodge star operator* of a metric complex K is the pair of homomorphisms,

$$\begin{array}{lll} \star: & C^k(K) \to C^{1-k}(K^\star), & k = 0, 1, \\ \star: & C^k(K^\star) \to C^{1-k}(K), & k = 0, 1, \end{array}$$

defined by

$$\star(\psi) := \psi^\star.$$

In other words, the new, *cochain* Hodge star operator is the dual (Hom-dual, of course) of the old, *chain* Hodge star operator; i.e.,

$$\star = \star^*.$$

The diagonal entries of the matrix of the new operator are the reciprocals of those of the old.

Proposition 4.53. The matrix of the Hodge star operator is diagonal with

$$\star_{ii} = \frac{|a_i^\star|}{|a_i|},$$

where a_i is the ith cell of K.

Proof. From the last subsection, we know that the matrix of the chain Hodge star operator,

$$\star : C_k(K) \to C_{1-k}(K^\star),$$

is diagonal with

$$\star_{ii} = \frac{|a_i|}{|a_i^\star|},$$

where a_i is the ith cell of K. We also know that the matrix of the dual is the transpose of the original. Therefore, the matrix of the dual of the above operator, which is the cochain Hodge star operator,

$$\star^* : C^{1-k}(K^\star) \to C^k(K),$$

is given by the same formula. We now restate this conclusion for the other cochain Hodge star operator,

$$\star^* : C^k(K) \to C^{1-k}(K^\star).$$

We simply replace in the formula: a with $a^\star$ and k with $(1-k)$. ■

Exercise 4.54. Provide the formula for the exterior derivative of $K^\star$ in terms of that of K.

4.7 The first derivative

The geometry of the complex allows us to define the first derivative of a discrete function (i.e., a 0-form) as "the rate of change" instead of just "change" (the exterior derivative).

This is the definition from elementary calculus of the derivative of a real-valued function $g : \mathbf{R} \to \mathbf{R}$ at $u \in \mathbf{R}$:

$$g'(u) := \lim_{h\to 0} \frac{g(u+h) - g(u)}{h}.$$

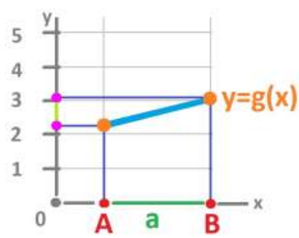
It is important to notice here that the definition remains equivalent when h is required to be *positive*.

Proposition 4.55. For a differentiable function g, we have

$$g'(u) = \lim_{u-v\to 0+} \frac{g(v) - g(u)}{v - u}.$$

Exercise 4.56. Prove the proposition.

Now, let's consider the setting of a 1-dimensional complex K:



We would like to know the rate of change of a 0-form over an interval, which is simply an edge in K. Then the discrete analog of the first derivative given above is

$$\frac{g(B) - g(A)}{|AB|}.$$

Let's take a closer look at this formula. The numerator is easily recognized as the exterior derivative of the 0-form g evaluated at the 1-cell $a = AB$. Meanwhile, the denominator is the *length* of this cell. In other words, we have:

$$g'(a^\star) := \frac{dg(a)}{|a|}.$$

This analysis justifies the definition of the derivative (not to be confused with the exterior derivative d) as follows.

Definition 4.57. The *first derivative* of a primal 0-form g on a metric cell complex K is a dual 0-form given by its values at the vertices of $K^\star$:

$$g'(P) := \frac{dg(P^\star)}{|P^\star|},\ \forall P \in K^\star.$$

The form g is called *differentiable* if this fraction exists in ring R.

Proposition 4.58. All forms are differentiable when we have

$$\frac{1}{|a|} \in R$$

for every cell a in K. In particular, this is the case under either of the two conditions below:
- the ring R of coefficients is a field, or
- the geometry of the complex K is *standard*: $|AB| = 1$.

The formula is simply the exterior derivative – with the extra coefficient equal to the reciprocal of the length of this edge. We recognize this coefficient from the definition of the Hodge star operator $\star : C^k(K) \to C^{1-k}(K^\star),\ k = 0, 1$. The operator's matrix is diagonal with

$$\star_{ii} = \frac{1}{|a_i|},$$

where a_i is the ith k-cell of K ($|A| = 1$ for 0-cells). Therefore, we have an alternative formula for the derivative.

Proposition 4.59.

$$g' = \star dg.$$

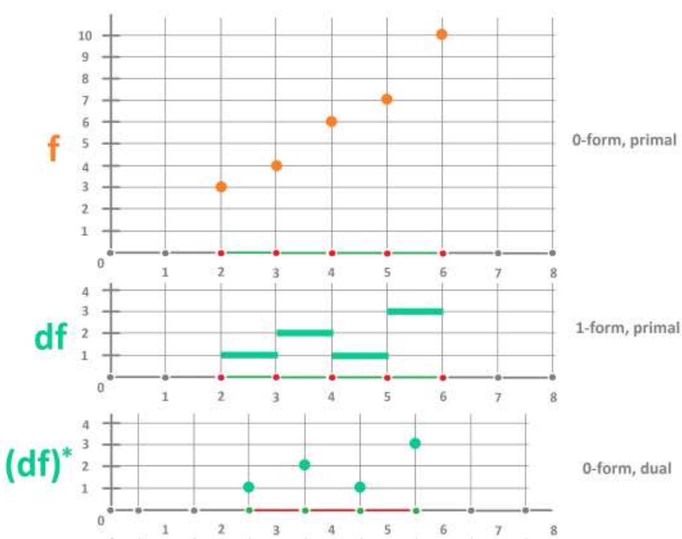
Thus, *differentiation is a linear operator*:

$$\tfrac{d}{dx} = \star d : C^0(K) \to C^0(K^\star).$$

It is seen as the diagonal of the following *Hodge star diagram*:

$$\begin{array}{ccccc} g \in & C^0(K) & \xrightarrow{\ \ d\ \ } & C^1(K) & \\ & & \tfrac{d}{dx} \searrow & \downarrow \star & \\ & & & C^0(K^\star) & \ni g' \end{array}$$

This is how the formula works in the case of edges with equal lengths:



Exercise 4.60. Prove: $f' = 0 \Longrightarrow f = const.$

Exercise 4.61. Create a spreadsheet for computing the first derivative as the "composition" of the spreadsheets for the exterior derivative and the Hodge duality.

Exercise 4.62. Define the first derivative of a 1-form h and find its explicit formula. Hint:

$$\begin{array}{cccc}
 & & C^1(K) & \ni h \\
 & \frac{d}{dx} \swarrow & \downarrow \star & \\
h' \in & C^1(K^\star) & \xleftarrow{\ d\ } C^0(K^\star) &
\end{array}$$

4.8 The second derivative

The concavity of a real-valued function is determined by the sign of its second derivative, which is *the rate of change of the rate of change*. In contrast to elementary calculus, we define it as a single limit:

$$g''(u) = \lim_{h \to 0} \frac{\frac{g(u+h)-g(u)}{h} - \frac{g(u)-g(u-h)}{h}}{h}.$$

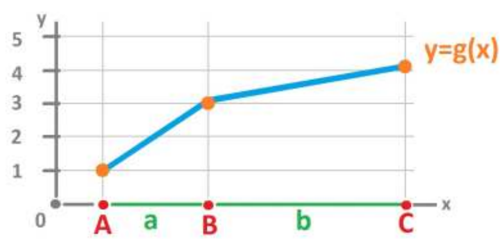
It is important to notice here that the definition remains equivalent when h is required to be *positive*.

Proposition 4.63. For a twice differentiable function g, we have

$$g''(u) = \lim_{v-u=u-w \to 0+} \frac{\frac{g(v)-g(u)}{v-u} - \frac{g(u)-g(w)}{u-w}}{(w-v)/2}.$$

Exercise 4.64. Prove the proposition.

This is the setting of the discrete case:



We would like to know the rate of change of the first derivative – from one edge (or vertex) to an adjacent one. Then the discrete analog of the second derivative above is

$$\frac{\frac{g(C)-g(B)}{|BC|} - \frac{g(B)-g(A)}{|AB|}}{|AC|/2}.$$

Let's take a closer look at this formula.

The numerators of the two fractions are easily recognized as the exterior derivative of the 0-form g evaluated at the 1-cells a and b respectively. Meanwhile, the denominators are the *lengths* of these cells.

The denominator is, in fact, the distance between the centers of the cells a and b. Therefore, we can think of it as the length of the *dual cell* of the vertex B. Then, we define the second derivative (not to be confused with dd) of a 0-form g as follows.

Definition 4.65. The *second derivative* of a primal 0-form g on a metric cell complex K is a primal 0-form given by its values at vertices of K:

$$g''(B) = \frac{\frac{dg(b)}{|b|} - \frac{dg(a)}{|a|}}{|B^\star|}, \ \forall B \in K$$

where a, b are the 1-cells adjacent to B.

We take this one step further. What is the meaning of the difference in the numerator? Combined with the denominator, it is the derivative of g' as a dual 0-form. The second derivative is indeed *the derivative of the derivative*:

$$g'' = (g')'.$$

Let's find this formula in the Hodge star diagram:

$$\begin{array}{cccccc}
g, g'' \in & C^0(K) & \xrightarrow{\quad d \quad} & C^1(K) & \\
& \uparrow \star & \neq & \downarrow \star & \\
& C^1(K^\star) & \xleftarrow{\quad d \quad} & C^0(K^\star) & \ni g'
\end{array}$$

If, starting with g in the left upper corner, we go around this, non-commutative, square, we get:

$$\star d \star dg.$$

Just as g itself, its second derivative

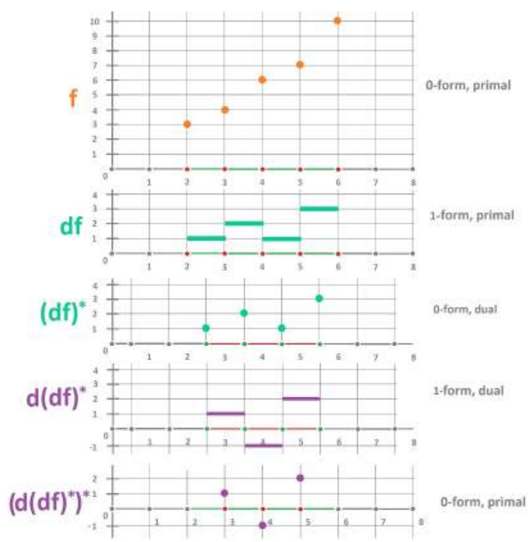
$$g'' := \star d \star dg = (\star d)^2 g$$

is another primal 0-form.

Definition 4.66. A 0-form g is called *twice differentiable* if it is differentiable and so is its derivative.

Proposition 4.67. All forms are twice differentiable when we have $\frac{1}{|a|} \in R$ for every cell a in K or $K^\star$.

This is how the formula works in the case of edges with equal lengths:



The second derivative is computed with a spreadsheet using the Hodge duality (dimension 1):

Primal complex:		0-cells, primal	A_i	A_1	A_2	A_3	A_4
		1-cells, primal	a_i	a_1	a_2	a_3	a_4
Dual complex:		0-cells, dual	B_i = dual of a_i	B_1	B_2	B_3	B_4
		1-cells, dual	b_i = dual of A_i		b_2	b_3	b_4
PROBLEM:							
Given	f	*1-cochain, primal*	numbers assigned to primal 1-cells = values f(a_i) of function f	*3*	*2*	*0*	*6*
find	f"	*1-cochain, primal*		?	?	?	?
SOLUTION:							
given	f	*1-cochain, primal*	*numbers f_i=f(a_i) assigned to primal 1-cells*	*3*	*2*	*0*	*6*
compute	g=f*	*0-cochain, dual*	*numbers g_i=g(B_i) assigned to dual 0-cells*	*3*	*2*	*0*	*6*
compute	h=dg	*1-cochain, dual*	*numbers h_i=h(b_i) assigned to dual 1-cells = differences of adjacent values of g*		*2*	*-6*	*5*
compute	k=h*	*0-cochain, primal*	*numbers k_i=k(A_i) assigned to primal 0-cells*		*2*	*-6*	*5*
compute	l=dk	*1-cochain, primal*	*numbers l_i=l(a_i) assigned to primal 1-cells = differences of adjacent values of k*		*8*	*-11*	*8*

Computing the second derivative over a metric complex with a non-trivial geometry is slightly more complex:

primal comple	K	0-cells, primal	A_i	A_1	A_2	A_3	A_4	A_5
		1-cells, primal	a_i	a_1	a_2	a_3	a_4	a_5
Dual complex:	K*	0-cells, dual	B_i = dual of a_i	B_1	B_2	B_3	B_4	B_5
		1-cells, dual	b_i = dual of A_i		b_2	b_3	b_4	b_5
Volumes		number	volume of primal 0-cell, |A_i|	1	2	3	1	9
		number	volume of primal 1-cell, |a_i|	2	3	2	3	4
		number	volume of dual 0-cell, |B_i|	1	1	1	1	1
		number	volume of dual 1-cell, |b_i|		2	2	2	1
Hodge star St	K->K*	number	coefficient of primal, 0->1, |b_i| / |A_i|		1.00	0.67	2.00	0.11
	K->K*	number	coefficient of primal, 1->0, |B_i| / |a_i|	0.50	0.33	0.50	0.33	0.25
	K*->K	number	coefficient of dual 0->1, |a_i| / |B_i|	2.00	3.00	2.00	3.00	4.00
	K*->K	number	coefficient of dual 1->0, |A_i| / |b_i|		0.67	1.00	0.67	0.25
PROBLEM:								
Given	f	*1-cochain, primal*	numbers assigned to primal 1-cells = values f(a_i) of function f	*3*	*2*	*0*	*6*	*1*
find	ϕээ	*1-cochain, primal*		?	?	?	?	?
SOLUTION:								
given	f	*1-cochain, primal*	*numbers f_i=f(a_i) assigned to primal 1-cells*	*3.00*	*2.00*	*0.00*	*6.00*	*1.00*
compute	g=f*	*0-cochain, dual*	*numbers g_i=g(B_i) assigned to dual 0-cells, g_i = g(B_i) = St_i * f_i*	*1.50*	*0.67*	*0.00*	*2.00*	*0.25*
compute	h=dg	*1-cochain, dual*	*numbers h_i=h(b_i) assigned to dual 1-cells = differences of adjacent values of g*		*0.83*	*0.67*	*-2.00*	*1.75*
compute	k=h*	*0-cochain, primal*	*numbers k_i=k(A_i) assigned to primal 0-cells, k_i = k(A_i) = St_i * h_i*		*0.56*	*0.67*	*-1.33*	*0.44*
compute	l=dk	*1-cochain, primal*	*numbers l_i=l(a_i) assigned to primal 1-cells = differences of adjacent values of k*		*-0.11*	*2.00*	*-1.77*	*26.69*

Note: See the files online.

Exercise 4.68. Instead of a direct computation as above, implement the second derivative via the spreadsheet for the first derivative.

Proposition 4.69. If the complex K is a directed graph with unit edges, the matrix D of the second derivative is given by its elements indexed by ordered pairs of vertices (u, v) in K,

$$D_{uv} = \begin{cases} -1 & \text{if } u, v \text{ are adjacent;} \\ \deg u, & \text{if } v = u; \\ 0 & \text{otherwise,} \end{cases}$$

with $\deg u$ the degree of node u.

Exercise 4.70. (a) Prove the formula. (b) Generalize the formula to the case of arbitrary lengths.

Exercise 4.71. Define the second derivative of a 1-form h and find its explicit formula.

4.9 Newtonian physics, the discrete version

The concepts introduced above allow us to state some elementary facts about the discrete Newtonian physics, in dimension 1.

First, the *time* is given by the standard metric complex $K := \mathbb{R}$, at the simplest. Note that, even though it is natural to assume that the complex that represents time has no curvature, it is still might be important to be able to handle time intervals of various lengths.

Second, the *space* is given by any ring R, in general. For all the derivatives to make sense at once, we can choose the standard geometry for time $K := \mathbb{R}$. Alternatively, we can choose R to be a field such as $R := \mathbf{R}$. (Note that even in the latter case, there is no topology and, consequently, there can be no assumptions about continuity.)

The main quantities we are to study are:

- the *location* r is a primal 0-form;
- the *displacement* dr is a primal 1-form;
- the *velocity* $v = r'$ is a dual 0-form;
- the *momentum* $p = mv$, where m is a constant mass, is also a dual 0-form;
- the *impulse* $J = dp$ is a dual 1-form;
- the *acceleration* $a = r''$ is a primal 0-form.

Some of these forms are seen in the following Hodge star diagram:

$$\begin{array}{ccc} a\ ..\ r & \xrightarrow{d} & dr \\ \uparrow\star & \neq & \downarrow\star \\ J/m & \xleftarrow{d} & v \end{array}$$

Newton's First Law: If the net force is zero, then the velocity v of the object is constant:

$$F = 0 \Longrightarrow v = const.$$

The law can be restated without invoking the geometry of time:

- If the net force is zero, then the displacement dr of the object is constant:

$$F = 0 \Longrightarrow dr = const.$$

The law shows that the only possible type of motion in this force-less and distance-less space-time is uniform; i.e., it is a repeated addition:

$$r(t+1) = r(t) + c.$$

Newton's Second Law: The net force on an object is equal to the derivative of its momentum p:

$$F = p'.$$

As we have seen, the second derivative is meaningless without specifying geometry, the *geometry of time.*

The Hodge star diagram above may be given this form:

$$\begin{array}{ccc} F\ ..\ r & \xrightarrow{d} & dr \\ \uparrow\star & \neq & \downarrow m\star \\ J & \longleftarrow\xleftarrow{d} & p \end{array}$$

Newton's Third Law: If one object exerts a force F_1 on another object, the latter simultaneously exerts a force F_2 on the former, and the two forces are exactly opposite:

$$F_1 = -F_2.$$

Law of Conservation of Momentum: In a system of objects that is "closed"; i.e.,

- there is no exchange of matter with its surroundings, and

- there are no external forces;

the total momentum is constant:

$$p = const.$$

In other words,

$$J = dp = 0.$$

To prove, consider two particles interacting with each other. By the third law, the forces between them are exactly opposite:

$$F_1 = -F_2.$$

Due to the second law, we conclude that

$$p_1' = -p_2',$$

or

$$(p_1 + p_2)' = 0.$$

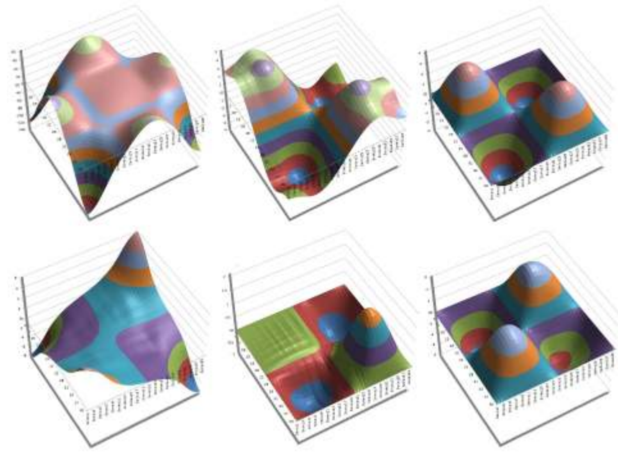
Exercise 4.72. State the equation of motion for a variable-mass system (such as a rocket). Hint: apply the second law to the entire, constant-mass system.

Exercise 4.73. Create a spreadsheet for all of these quantities.

Whatever ring we choose, $R = \mathbf{Z}$ or $R = \mathbf{R}$, these laws of physics are satisfied exactly not approximately.

Chapter VII

Flows



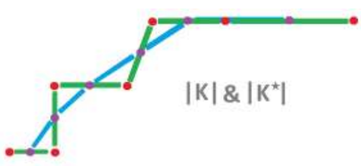
1 Metric complexes

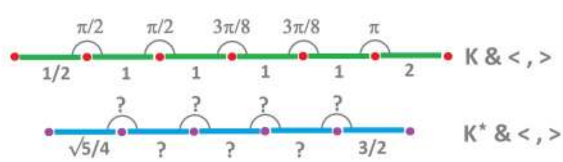
1.1 Hodge-dual complexes

Let's recall the mechanical interpretation of a realization $|K|$ of a metric complex K of dimension $n = 1$:

- the edges are *rods* and the vertices are *hinges*.

Furthermore, we connect the centers of the adjacent rods using an extra set of rods (and hinges) that form a new, Hodge-dual, complex $K^\star$.

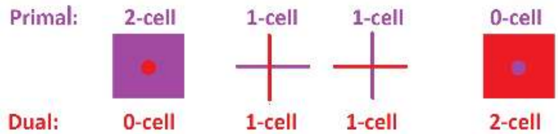




Given now a cell complex K of an *arbitrary* dimension, the first step in the construction its dual $K^\star$ is to choose the *dimension* n. The rule remains the same:

• each primal/dual k-cell α corresponds to a dual/primal $(n-k)$-cell $\alpha^\star$.

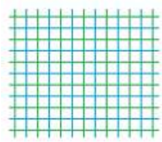
First we consider the case of $n = 2$:



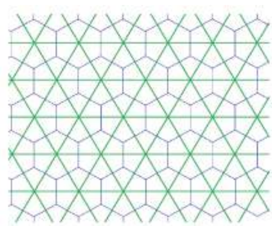
For the illustrations below, we will assume that each primal cell and its dual are directly "linked" in the realization, as follows:

• If α is a 2-cell then $\alpha^\star$ is a 0-cell located at the "center" of α.
• If a is a 1-cell then $a^\star$ is a 1-cell crossing at the "center" of a.
• If A is a 0-cell then $A^\star$ is a 2-cell centered at A.

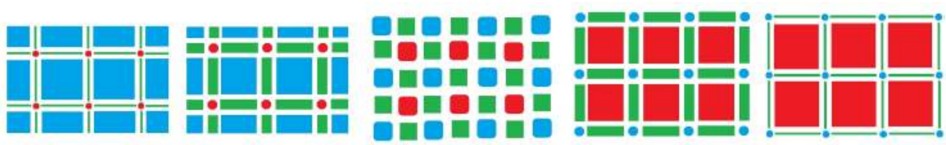
The result is a "staggered" grid of tile-like cells. It is simple for $\mathbb{R}^2$, as well as for any rectangular grid:



Other decompositions of the plane produce more complex patterns:

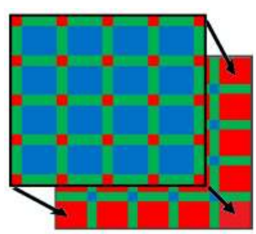


For the 2-cells, their duals *are* their centers, the 0-cells. We see below how these squares shrink to become the dots.

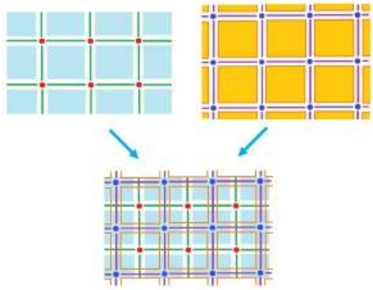


Meanwhile, the 0-cells are blown into the squares. Finally, the duality makes every 1-cell turn 90 degrees around its center.

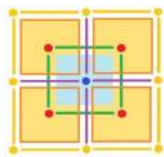
It is easy to illustrate the Hodge duality in a spreadsheet: one simply resizes the rows and columns in order to make the narrow cells wide and the wide cells narrow:



That's how you could build the dual of the whole-plane complex – following this one-to-one correspondence. Furthermore, you can rebuild the primal complex from the dual following the same procedure. Below, we show a primal grid, the dual grid, and then the two overlapping each other:



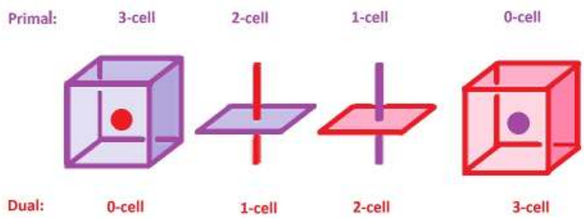
What happens if we have a *finite* cubical complex K in $\mathbf{R}^2$? Just as in dimension 1, some boundary cells are missing in the dual complex, as the example shows:



Then, in general,

$$K^{\star\star} \neq K.$$

There is more going on in $\mathbf{R}^3$:



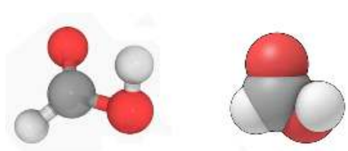
The k-cells and the $(n-k)$-cells are still paired up.

Exercise 1.1. Express the dual cell of a cube Q in $\mathbb{R}^N$ in terms of its primal via the orthogonal complement construction.

Exercise 1.2. What is the Hodge dual of $\mathbb{R}^2$ in the $\mathbb{R}^3$?

Exercise 1.3. In $\mathbb{R}^n$, each k-cube Q is represented as the product of k edges and $n-k$ vertices. Find such a representation for $Q^\star$.

Example 1.4 (molecule). To illustrate the idea of duality in dimension 3, one may consider the two ways a chemist would construct a model of a molecule:

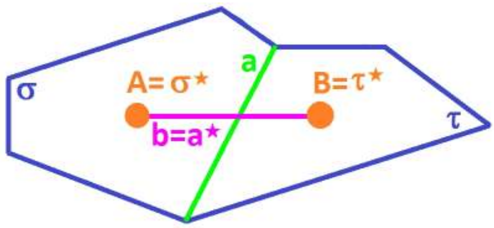


The two equally valid ways are:
- small balls connected by rods, or
- large balloons pressed against each other. □

Exercise 1.5. Describe the cells in the above example.

Exercise 1.6. Two polyhedra are called *dual* if the number of faces of one is equal to the number of vertices of the other (and vice versa). Find the dual polyhedra of Platonic solids and explain the relation to the Hodge duality:

This is what the Hodge duality looks like in the case of a general 2-dimensional cell complex. The dual cell is still placed – for the sake of illustration – at the center of mass of the primal cell:



Exercise 1.7. Suppose a simplicial complex K is realized in $\mathbf{R}^n$. Describe explicitly how the dual $K^\star$ can also be realized in $\mathbf{R}^n$ side by side with $|K|$. Hint: use barycentric coordinates.

We summarize the new concept as follows.

Definition 1.8. Suppose a *dimension* n is given. Two cell complexes K and $K^\star$ with $\dim K \leq n$, $\dim K^\star \leq n$, are called *Hodge-dual* of each other if there is a one-to-one correspondence,

$$K \ni a \longleftrightarrow a^\star \in K^\star,$$

between their cells of complementary dimensions:

$$\dim a^\star + \dim a = n.$$

1.2 Metric complexes

We have already seen metric complexes of dimension $n = 1$ as cell complexes equipped with a geometric structure: the metric tensor of K and $K^\star$. This data allows one to find the lengths and the curvatures of curves. The Hodge star operator is then built from this data.

In this subsection, we present the necessary geometric data for the case of arbitrary dimension.

In dimension $n = 2$, we consider lengths and areas; in dimension $n = 3$, lengths, areas, and volumes, etc. In general, we need to find the "sizes" of all cells of all dimensions in the complex. These sizes are called the *k-dimensional volumes*:

dimension k	k-cell a	k-volume $\lvert a \rvert$
0	vertex	1
1	edge	length
2	face	area
3	solid	volume
...		...
n	n-cell	n-volume

Instead of trying to *derive* these quantities from the metric tensor, we simply *list* them as a part of the definition of a metric complex. In the abstract, we simply have a pair of positive numbers $|b|, |b^\star|$ assigned to each cell $b \in K$.

Definition 1.9. Given a cell complex K, the *volume function* on K defines the k-volumes of the k-cells of K:

$$|\cdot| : K \to R \setminus \{0\},$$

normalized for dimension 0:

$$|A| = 1 \text{ for all vertices } A \in K.$$

For $R = \mathbf{R}$, the definition matches the one for the case of $n = 1$ considered previously if we limit ourselves to positive volumes.

Example 1.10. The simplest choice of the volumes for the 2-cell representation of the sphere is

$$A^0, a^2 \text{ with } |A^0| = |a^2| = 1. \qquad \square$$

In general, we don't assume that $|AB|, |BC|, |CA|$ determine $|ABC|$. Even though the three sides of a triangle determine the triangle and, therefore, its area, this only works under a particular choice of geometry, such as Euclidean. We take the opposite approach and let the lengths and areas determine the geometry.

Exercise 1.11. Give a metric complex representation of a triangulation of the sphere, realized as the unit sphere in $\mathbf{R}^3$.

Example 1.12 (measure). Suppose K is a cell complex with finitely many cells at each vertex and suppose $|K|$ is its realization. Then the set of all cells (open or closed) Σ of K as subsets of $|K|$ form a σ-algebra, i.e., a collection closed under the operations of complement, countable union, and countable intersection. If, furthermore, K has a volume function, the function $\mu : \Sigma \to \mathbf{R} \cup \{\infty\}$ given by

$$\mu(a) = \begin{cases} |a| & \text{if } \dim a = \dim K, \\ 0 & \text{if } \dim a < \dim K, \end{cases}$$

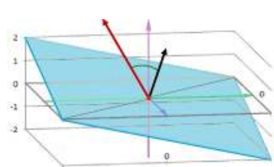
is a *measure* on $|K|$; i.e., it satisfies:

- non-negativity: for all $a \in \Sigma$ we have $\mu(a) \geq 0$;
- null empty set: $\mu(\emptyset) = 0$;
- σ-additivity: for all countable collections $\{a_i\}$ of pairwise disjoint sets in Σ we have:

$$\mu\Big(\bigcup_{k=1}^{\infty} a_k\Big) = \sum_{k=1}^{\infty} \mu(a_k). \qquad \square$$

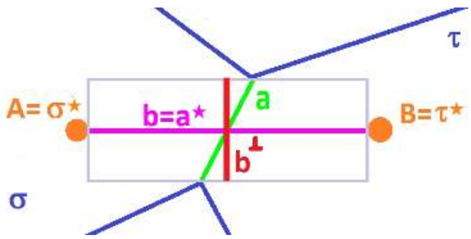
In addition to the volumes, the *angles* between every pair of dual cells will be taken into account in the new representation of the geometry.

We don't need to specify the meaning of the angle between *every choice* of a primal k-cell and its dual $(n-k)$-cell, for all k. For our purposes, it suffices to take care of the case $k = 1$. We are then in a familiar situation because the angle α between 1- and $(n-1)$-dimensional subspaces of an n-dimensional inner product space is always understood as the angle between the direction vector of the former and the normal vector of the latter.

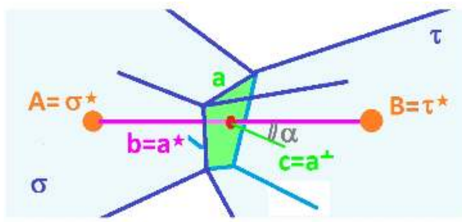


They are orthogonal when this angle is zero. It is also assumed that (for $k = 0$) a point and the whole space are orthogonal and $\cos\alpha = 1$. This assumption explains why these angles did not have to come into play in the case of a 1-dimensional complex.

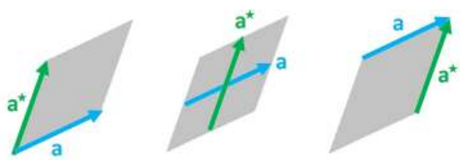
For cells in the higher dimensional case, we are looking at the angle between a and the normal $b^\perp$ of $b = a^\star$, illustrated for dimension $n = 2$:



Dually, this angle can also be understood as the angle between the normal $c = a^\perp$ of a and $a^\star$, illustrated for dimension $n = 3$:



Even though these angles don't affect the volumes of either primal or dual cells, they do affect the volumes of *new* n-cells, the cells determined by each dual pair:



The area of this parallelogram does not depend on how a and $a^\star$ are attached to each other (they aren't) but only on the angle between them.

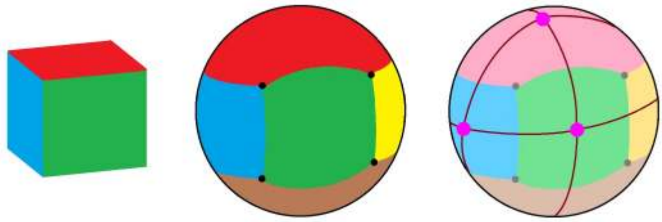
The way we handle angles is, of course, via the inner product: $\langle a, b\rangle \in R$ is defined for these three cases:

- adjacent $a, b \in K$,
- adjacent $a, b \in K^\star$, and now also
- $a \in K, a^\star \in K^\star$.

Definition 1.13. A *metric (cell) complex of dimension* n is a pair of Hodge-dual cell complexes K and $K^\star$ of dimension n equipped with a *joint metric tensor*, which is a combination of metric tensors for K and $K^\star$ as well as inner products of all dual pairs:

$$(a, a^\star) \mapsto \langle a, a^\star\rangle \in R.$$

Exercise 1.14. Represent a hollow cube as a metric complex. Hint:



Exercise 1.15. Suppose K is a metric complex. Show how the set of its vertices K^0 becomes a metric space.

It follows from the definition that, in addition to the two volume functions of these complexes, we can also derive the *angle function*:

$$(a, a^\star) \mapsto \cos\widehat{aa^\star} = \frac{\langle a, a^\star\rangle}{|a||a^\star|}.$$

The function is well-defined whenever the expression in the right-hand belongs to our ring R. This is always the case when R is a field or $|a| = |a^\star| = 1$.

Definition 1.16. We say that K has a *rectangular geometry* if

$$\langle a, a^\star\rangle = \pm|a||a^\star|.$$

In other words, all dual pairs are orthogonal:

$$\cos\widehat{aa^\star} = \pm 1.$$

The case of $n = 1$ previously considered is rectangular.

Definition 1.17. We will say that the complex K has the *standard geometry* if it is rectangular and also fully normalized:

$$|a| = 1 \text{ for all cells } a \in K \sqcup K^\star.$$

We always assume that $\mathbb{R}^n$ has the standard geometry.

1.3 The Hodge star operator

We continue to take the concepts we have considered previously for the case of dimension 1 and define them for the general case of dimension n. We define an operator from primal to dual chains that incorporate the geometry of the complex. Its formula is based on the analysis above.

Definition 1.18. The (chain) *Hodge star operator* of a metric complex K of dimension n is the pair of homomorphisms on chains of complementary dimensions:

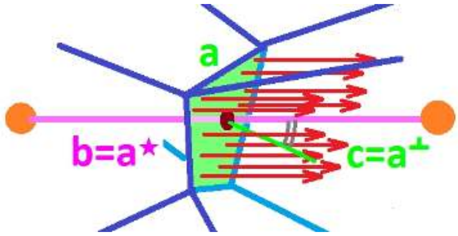
$$\star : C_k(K) \to C_{n-k}(K^\star), \quad k = 0, 1, ..., n,$$
$$\star : C_k(K^\star) \to C_{n-k}(K), \quad k = 0, 1, ..., n,$$

defined by

$$\star(a) := \pm\frac{\langle a, a^\star\rangle}{|a^\star|}a^\star.$$

for any cell a, dual or primal, with the sign to be defined later.

The expression can be understood as the total flux of a across a region of area 1 on $a^\star$:



The two operators then satisfy:

$$\star(a)|a^\star| = \langle a, a^\star \rangle = \star(a^\star)|a|$$

The following statements are easy to prove.

Proposition 1.19. The matrix of the Hodge star operator $\star$ is diagonal with:

$$\star_{ii} = \pm \tfrac{|a_i|}{|a_i^\star|} \cos \widehat{a_i a_i^\star},$$

where a_i is the ith cell of K.

Theorem 1.20. The two star operators are the inverses of each other, up to a sign:

$$\star\star = \pm \operatorname{Id}.$$

Theorem 1.21. For a rectangular geometry, the Hodge star operator $\star$ preserves volumes of cells:

$$|\star(a)| = |a|.$$

Proposition 1.22. The Hodge star operator is always well-defined when
- R is a field, or
- the geometry of K is standard.

Exercise 1.23. Provide the Hodge star operator for a metric complex representation of a triangulation of the sphere realized as the hollow unit cube in $\mathbf{R}^3$.

Exercise 1.24. Provide the Hodge star operators for the examples of metric tensors presented in this and the last sections.

Next, what happens to the *forms* under Hodge duality?

Definition 1.25. In a metric complex K, the *Hodge-dual form* $\psi^\star$ of a form ψ over K is given by

$$\psi^\star(\sigma) := \psi(\sigma^\star).$$

The operation is extended from cells to chains by linearity making these diagrams commutative:

$$\begin{array}{ccc} C_k(K) & \xrightarrow{\star} & C_{n-k}(K^\star) \\ \downarrow \varphi & & \downarrow \varphi^\star \\ R & = & R \end{array} \qquad \begin{array}{ccc} C_k(K) & \xleftarrow{\star} & C_{n-k}(K^\star) \\ \downarrow \psi^\star & & \downarrow \psi \\ R & = & R \end{array}$$

Then the formula is written for any chain:

$$(\star\psi)(a) := \psi(\star a),$$

where $\star$ in the right-hand side is the chain Hodge star operator.

Definition 1.26. The (cochain) *Hodge star operator* of a metric complex K is the dual of the chain Hodge star operator ($\star = \star^*$); i.e., it is the pair of homomorphisms,

$$\begin{aligned} \star : C^k(K) &\to C^{n-k}(K^\star), \quad k = 0, 1, ..., n, \\ \star : C^k(K^\star) &\to C^{n-k}(K), \quad k = 0, 1, ..., n, \end{aligned}$$

defined by

$$\star(\psi) := \psi^\star.$$

Proposition 1.27. The matrix of the Hodge star operator is diagonal with

$$\star_{ii} = \frac{|a_i|}{|a_i^\star|} \cos \widehat{aa^\star},$$

where a_i is the ith cell of K.

Exercise 1.28. Prove the proposition.

As before, we use in dimension $n = 1$:

$$\star^1_{ii} = \frac{|\text{point}|}{|\text{edge}|} = \frac{1}{\text{length}} = \frac{1}{\Delta x}.$$

Computing Hodge duality of forms over a complex with a non-trivial geometry is only slightly more complicated than for the standard case:

primal complex:	K	0-cells, primal	A_i	A_1	A_2	A_3	A_4	A_5
		1 cells, primal	a_i	a_1	a_2	a_3	a_4	a_5
Dual complex:	K*	0-cells, dual	B_i = dual of a_i	B_1	B_2	B_3	B_4	B_5
		1 cells, dual	b_i = dual of A_i		b_2	b_3	b_4	b_5
Volumes		number	volume of primal 0-cell, |A_i|	1	2	3	1	9
		number	volume of primal 1 cell, |a_i|	2	3	2	3	4
		number	volume of dual 0-cell, |B_i|	1	1	1	1	1
		number	volume of dual 1-cell, |b_i|		2	2	2	1
Hodge star St	K->K*	number	coefficient of primal, 0->1, |b_i| / |A_i|		1.00	0.67	2.00	0.11
	K->K*	number	coefficient of primal, 1->0, |B_i| / |a_i|	0.50	0.33	0.50	0.33	0.25
	K*->K	number	coefficient of dual 0->1, |a_i| / |B_i|	2.00	3.00	2.00	3.00	4.00
	K*->K	number	coefficient of dual 1->0, |A_i| / |b_i|		0.67	1.00	0.67	0.25
PROBLEM:								
Given	f	*0-cochain, primal*	numbers assigned to primal 0-cells, f_i = f(A_i)	1	2	3	6	1
compute	f*	*1-cochain, dual*	numbers assigned to dual 1-cells, f*_i = f*(b_i) = St_i * f_i		2.00	2.00	12.00	0.11
PROBLEM:								
Given	f	*1 cochain, primal*	numbers assigned to primal 1 cells, f_i = f(a_i)	3	2	0	6	1
compute	f*	*0-cochain, dual*	numbers assigned to dual 0-cells, f*_i = f*(b_i) = St_	1.50	0.67	0.00	2.00	0.25
PROBLEM:								
Given	f	*0 cochain, dual*	numbers assigned to primal 0 cells, f_i = f(A_i)	1	2	3	6	1
compute	f*	*1-cochain, primal*	numbers assigned to dual 1-cells, f*_i = f*(b_i) = St_i * f_i		6.00	6.00	18.00	4.00
PROBLEM:								
Given	f	*1-cochain, dual*	numbers assigned to primal 1 cells, f_i = f(a_i)		2	0	6	1
compute	f*	*0-cochain, primal*	numbers assigned to dual 0-cells, f*_i = f*(b_i) = St_i * f_i		2.33	0.00	4.00	0.25

Note: See the files online.

In dimension $n = 2$, we have the following diagonal entries for a rectangular grid:

$$\begin{aligned} \star^2_{ii} &= \frac{|\text{point}|}{|\text{rectangle}|} &= \frac{1}{\text{area}} &= \frac{1}{\Delta x \Delta y}, \\ \star^1_{ii} &= \frac{|\text{edge}|}{|\text{edge}|} &= \frac{\text{length}}{\text{length}} &= \frac{\Delta y}{\Delta x}. \end{aligned}$$

In a 2-dimensional cubical complex, combining two 1-cells may produce different orientations and, therefore, different signs. Since $a \times d = \tau$ but $d \times a = -\tau$, the duality operator is

$$\begin{aligned} \Big([0,1] \times \{0\}\Big)^\star &= \{0\} \times [0,1], \\ \Big(\{0\} \times [0,1]\Big)^\star &= -[0,1] \times \{0\}. \end{aligned}$$

In general, we are guided by the following rule.

Sign convention: The sign of the Hodge star is determined by:

$$\star\star\,\alpha = (-1)^{k(n-k)}\alpha,$$

for a k-form α.

Exercise 1.29. Confirm the sign convention for $\mathbb{R}^n$.

1.4 The Laplace operator

Let's review what we know about the derivatives.

Given a function f, we know that $ddf = 0$. Therefore, the "second exterior derivative" is always trivial. Instead, one can define the "Hodge-dualized" second derivative computed following these four steps:

- find the exterior derivative df of,
- "dualize" df,
- find the exterior derivative of the result, and then
- "dualize" the outcome back.

We can write the formula explicitly for $\mathbb{R}$:

$$f \stackrel{d}{\longmapsto} df = f'dx \stackrel{\star}{\longmapsto} f' \stackrel{d}{\longmapsto} f''dx \stackrel{\star}{\longmapsto} f''.$$

Then, for a 0-form, this is, again, a 0-form. This computation relies on a new, dual grid. In it, the new vertices are the centers of each 1-cell. We assign the values of df to these new vertices and then evaluate the exterior derivative of the result. Note that no derivative is defined at the endpoints of a 1-dimensional complex.

Thus, given two values assigned to two adjacent cells, one assigns their difference to the cell in-between, illustrated below with a spreadsheet:

Suppose now a metric complex K of an arbitrary dimension n is given. Consider its *Hodge duality diagram*:

$$\begin{array}{ccccccccccccc}
0 & \xrightarrow{d} & C^0(K) & \xrightarrow{d} & C^1(K) & \xrightarrow{d} & \dots & \xrightarrow{d} & C^{n-1}(K) & \xrightarrow{d} & C^n(K) & \xrightarrow{d} & 0\\
\downarrow{\scriptstyle 0} & \neq & \downarrow{\scriptstyle\star}\uparrow & \neq & \downarrow{\scriptstyle\star}\uparrow & & & & \downarrow{\scriptstyle\star}\uparrow & \neq & \downarrow{\scriptstyle\star}\uparrow & \neq & \downarrow{\scriptstyle 0}\\
0 & \xleftarrow{d} & C^n(K^\star) & \xleftarrow{d} & C^{n-1}(K^\star) & \xleftarrow{d} & \dots & \xleftarrow{d} & C^1(K^\star) & \xleftarrow{d} & C^0 & \xleftarrow{d} & 0
\end{array}$$

Then, the generalizations of both first and second derivatives of 0-forms are straight-forward and visible in the diagram.

Definition 1.30. The *second derivative of a* 0*-form* f is defined to be

$$f'' := \star d \star df.$$

We recognize f'' as the result of going around the leftmost square starting from the top left.

But what about higher order forms? What is the second derivative of $\varphi \in C^n$? We can still consider $\star d \star d\varphi$ but, as we follow the square on the right, the value is zero!

We notice that this time there is another square adjacent to C^n, on the left. Then we can, alternatively, first take the Hodge dual, then the exterior derivative, and continue around that square. The result is another "Hodge-dualized" second derivative.

Definition 1.31. The *second derivative of an* n*-form* φ is defined to be

$$\varphi'' := d \star d \star \varphi.$$

We will use the same notation when there is no confusion. Both are implicitly present in the following:

$$(\varphi'')^\star = \star(d \star d \star \varphi) = (\star d \star d \star)\varphi = (\varphi^\star)''.$$

As we move to the general k-forms, we discover that, for $0 < k < n$, there are *two* squares to follow and two, possibly non-trivial, second derivatives to be defined! Starting with a k-form, taking either of the two routes results in a new k-form:

$$\begin{array}{ccccc}
C^{k-1} & \xrightarrow{d} & C^{k} & \xrightarrow{d} & C^{k+1} \\
\uparrow \star & \neq & \downarrow \star \ \uparrow \star & \neq & \downarrow \star \\
C^{n-(k-1)} & \xleftarrow{d} & C^{n-k} & \xleftarrow{d} & C^{n-(k+1)}
\end{array}$$

Neither of the two but rather their *sum* is the true substitute for the second derivative.

Definition 1.32. The *Laplace operator* of a metric complex K is the operator

$$\Delta : C^k(K) \to C^k(K),\ k = 0, 1, ...,$$

defined to be

$$\Delta = \star d \star d + d \star d \star .$$

Proposition 1.33. The Laplace operator is a cochain map on K.

Exercise 1.34. (a) Prove the proposition. (b) Are $\star d$, $d\star$, $\star d \star d$, $d \star d\star$ also cochain maps? (c) Is the Laplace operator generated by a cell map?

Exercise 1.35. Compute the cohomology map of the Laplace operator for (a) $K = \mathbb{R}^1$, (b) $K = \mathbb{S}^1$, with the standard geometry.

---

2 ODEs

2.1 Motion: location from velocity

Let R be our ring of coefficients.

We start with a few simple *ordinary differential equations (ODEs)* with respect to the exterior derivative d that have explicit solutions.

Given a complex K, this is the most elementary ODE with respect to a 0-form f:

$$df = G,$$

with G some 1-form over K. This equation has a solution, i.e., a 0-form that satisfies the equation everywhere, if and only if G is exact (i.e., it is a coboundary). In that case,

$$f \in d^{-1}(G),$$

with d understood as a linear map $d : C^0(K) \to C^1(K)$. There are, in fact, multiple solutions here. We know that they differ by a constant, as antiderivatives should.

If we choose an *initial condition*:

$$f(A) = r \in R,\ A \in K^{(0)},$$

such a solution is also unique. The explicit formula is:

$$f(X) = r + G(AX),\ \ X \in K^{(0)},$$

where AX is any 1-chain from A to X in K, i.e., any 1-chain $AX = a \in K^{(1)}$ with $\partial a = X - A$.

The solution is well-defined for an exact G (which is always the case when K is acyclic). Indeed, if a' is another such chain, we have

$$[r + G(a)] - [r + G(a')] = G(a) - G(a') = G(a - a') = 0,$$

because $a - a'$ is a cycle and exact forms vanish on cycles. To verify that this is indeed a solution, let's suppose $b := BC \in K^{(1)}$. Then, by the Stokes Theorem, we have

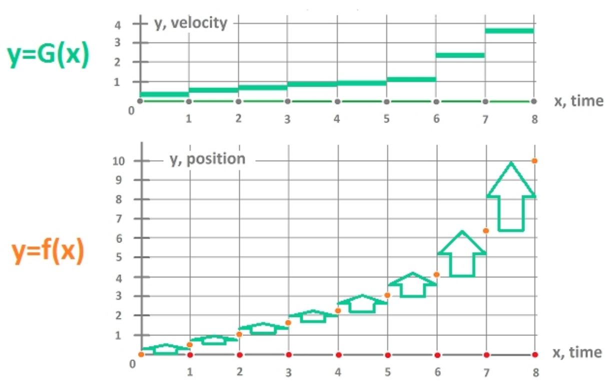
$$\begin{array}{ll} df(b) & = f(\partial b) = f(B - C) = f(B) - f(C) \\ & = [r + G(AB)] - [r + G(AC)] = G(AB - AC) = G(BC) \\ & = G(b), \end{array}$$

due to $G(AB + BC + CA) = 0$.

Exercise 2.1. Give an example of K and G so that the ODE has no solution.

We can use the above equation to model *motion*. Our domain is then the standard 1-dimensional cubical complex $K = \mathbb{R}$ and we study differential forms over ring $R = \mathbf{R}$, the reals. We have:

- $f = f(t)$ is the location of an object at time t, a 0-form defined on the integers $\mathbf{Z}$, and
- $G = G(t)$ is its displacement, a 1-form defined on the intervals $[0,1], [1,2], ...$.



We can give the above formula a more familiar, integral form:

$$f(X) = r + \int_A^X G(x)dx,$$

which, in our discrete setting, is simply a summation:

$$f(X) = r + \sum_{i=A}^{X-1} G\Big([i, i+1]\Big).$$

What about the question of the discrete vs. the continuous? The space is $R = \mathbf{R}$, which seems continuous, while the time is $K = \mathbb{R}$, which seems discrete. However, what if we acknowledge only the structures that we actually *use*? First, since ring R has no topology that we use,

- the space is algebraic.

Second, since complex K has no algebra that we use,

- the time is topological.

Exercise 2.2. Solve the above ODE for the case of (a) circular space $R = \mathbf{Z}_p$, and (b) circular time $K = \mathbb{S}^1$. Hint: make sure your solution is well-defined.

A more conventional way to represent motion is with an ODE with respect to:

- the first derivative f' instead of the exterior derivative, and
- the velocity v instead of the displacement.

For the derivative to be defined, the time has to be represented by a *metric* complex K. Then the ODE becomes:

$$f' = v.$$

Recall that both are dual 0-forms.

Exercise 2.3. Show that the solution f of this ODE is the same as above for the case of $K = \mathbb{R}$ with the standard geometry: $|a| = 1$ for all $a \in K, a \in K^{\star}$.

2.2 Population growth

In more complex equations, the exterior derivative of a form depends on the form. When $K = \mathbb{R}$, the simplest example is:

$$df = Gfq,$$

where $G : R \to R$ is some function and $q : C_1(\mathbb{R}) \to C_0(\mathbb{R})$ is given by

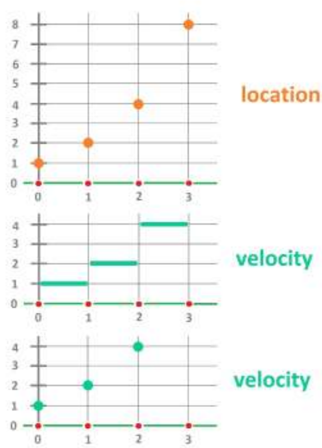
$$q\Big([n, n+1]\Big) = n.$$

The latter is used to make the degrees of the forms in the equation match.

In particular, f may represent the population with $G = k$ a constant growth rate. Then the equation becomes:

$$df = kfq.$$

The actual dynamics resembles that of a bank account: once a year the current amount is checked and then multiplied by a certain constant number $k > 1$. The case of $k = 2$ is illustrated below:



For the initial condition

$$f(0) = r \in R,$$

there is an explicit formula:

$$f(X) = r(k+1)^X, \ X \in K^{(0)} = \mathbf{Z}.$$

The growth is exponential (geometric), as expected. To verify, suppose $b := [B, B+1] \in K^{(1)}$. Then compute:

$$\begin{aligned} df(b) &= f(B+1) - f(B) \\ &= r(k+1)^{B+1} - r(k+1)^B = r(k+1)^B((k+1)-1) \\ &= r(k+1)^B k = f(B)k = fq(b)k. \end{aligned}$$

Exercise 2.4. Confirm that the dynamics is as expected for all values of $k > 0$ and $k < 0$.

Just as in the last subsection, we can represent the dynamics more conventionally with an ODE with respect to the first derivative f' provided the time is represented by a metric complex. This geometry allows us to consider variable time intervals.

Exercise 2.5. Set up an ODE for the derivative f' and then confirm that its solution is the same as above for the case of the standard geometry of the complex of time $K = \mathbb{R}$.

2.3 Motion: location from acceleration

If we are to study the motion that is produced by *forces* exerted on an object, we are compelled to specify the *geometry* of the space, in contrast to the previous examples.

Suppose K is a metric complex of dimension 1.

Recall that

- the location r is a primal 0-form;
- the velocity $v = r'$ is a dual 0-form;
- the acceleration $a = v'$ is a primal 0-form.

The ODE is:

$$r'' = a,$$

for a fixed a. Here $r'' = \star d \star dr$, with $\star$ the Hodge star operator of K. The motion is understood as if, at the preset moments of time, the acceleration steps in and instantly changes the velocity, which stays constant until the next time interval.

This is a second order ODE with respect to r. Its initial value problem (IVP) includes both an initial location and an initial velocity.

To find the explicit solution, let's suppose that $K = \mathbb{R}$ has the standard geometry.

Let's recall what we have learned about antidifferentiation. If we know the velocity, this is how we find the location:

$$r(t) = r_0 + \sum_{i=0}^{t-1} v\Big([i, i+1]\Big),$$

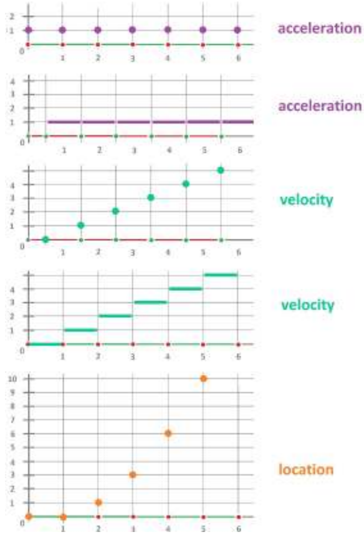
where r_0 is the initial location. And if we know the velocity, we can find the acceleration by the same formula.

$$v\Big([i, i+1]\Big) = v_0 + \sum_{j=0}^{t-1} a(j),$$

where v_0 is the initial velocity. Therefore,

$$r(t) = r_0 + \sum_{i=0}^{t-1} \Big(v_0 + \sum_{j=0}^{i-1} a(j)\Big) = r_0 + v_0 t + \sum_{i=0}^{t-1} \sum_{j=0}^{i-1} a(j).$$

The case of a constant acceleration is illustrated below:



The formula is, of course,

$$r(t) = r_0 + v_0 t + \frac{at(t-1)}{2}.$$

The dependence is quadratic, as expected. Note that the formula works even with $R = \mathbf{Z}$ as $t(t-1)$ is even.

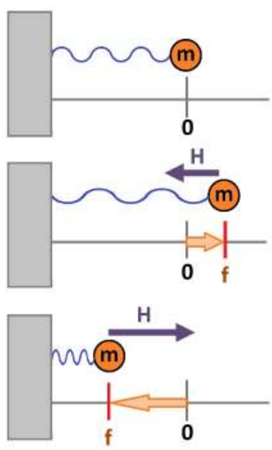
Exercise 2.6. Solve the ODE for $K = \mathbb{S}^1$ with the standard geometry.

Exercise 2.7. Solve the ODE for $K = \mathbb{R}$ with variable time intervals.

2.4 Oscillating spring

More elementary physics...

Imagine an object of mass $m \in R$ connected by (mass-less) spring to the wall.



The motion is, as before, within our ring of coefficients R. We let $f(t) \in R$ be the location of the object at time t and assume that the equilibrium of the spring is located at $0 \in R$. As before, we think of f as a form of degree 0 over a metric complex K.

The equation of the motion is derived from Hooke's law: the force exerted on the object by the spring is proportional to the displacement of the object from the equilibrium:

$$H = -kf,$$

where $k \in R$ is the spring constant.

Now, by the Second Newton's Law, the total force affecting the object is

$$F = ma,$$

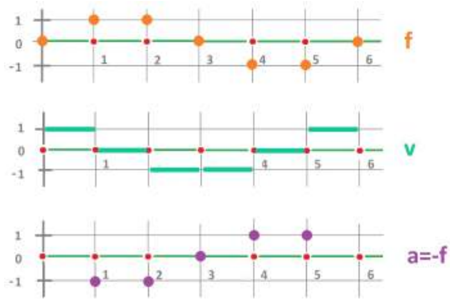
where a is the acceleration:

$$a = f'' = \star d \star df.$$

As there are no other forces, the two forces are equal and we have our *second order ODE*:

$$mf'' = -kf.$$

Let's assume that the geometry of $K = \mathbb{R}$ is standard and let $m := 1$ and $k := 1$. Then one of the solutions is the sequence $0, 1, 1, 0, -1, -1, 0, \ldots$. It is shown below along with its verification:



The dynamics is periodic, as expected.

Exercise 2.8. Find the explicit solutions of this ODE.

2.5 ODEs of forms

We will not be concerned with geometry in the rest of this section. We limit ourselves to "time-variable" ODEs: $K := \mathbb{R}$, the standard cubical complex.

Broadly, an ODE is a dependence of directions on locations in *space*, provided by our ring R. To make this point clearer, we can exhibit directions and locations in *time*, provided by the tangent bundle $T(\mathbb{R})$ of our complex $\mathbb{R}$. This is why we, initially, start by considering "local" 1-forms on $\mathbb{R}$, $\varphi \in T^1(\mathbb{R})$, i.e., R-valued functions on the tangent bundle of $\mathbb{R}$:

$$\varphi : T(\mathbb{R}) \to R,$$

linear on each tangent space $T_A(\mathbb{R}), A \in \mathbb{R}$. The exterior derivative of a 0-form f is a local 1-form given by the same formula as before:

$$df(A, AB) = f(B) - (A).$$

Definition 2.9. Given a function

$$P : C_0(\mathbb{R}) \times C_1(\mathbb{R}) \times R \to R,$$

linear on the first and second arguments, an *ordinary differential equation (ODE) of forms* of order 1 with right-hand side function P is:

$$df(A, AB) = P(A, AB, f(A)),$$

where

- $f \in C^0(\mathbb{R})$,
- $A \in C_0(\mathbb{R})$, and
- $AB \in C_1(\mathbb{R})$.

Note that the linearity requirements are justified by the linearity of the exterior derivative df in the left-hand side.

The abbreviated version of the equation is below.

$$\boxed{df = Pf}$$

Note that the variables of P encode: a time instant, a direction of time at this instant, and the value of the form at this instant, respectively.

Exercise 2.10. A function $h : \mathbf{R}^n \to \mathbf{R}$ is called a *potential* of a vector field V if $\operatorname{grad} h = V$. Interpret the problem of finding a potential of a given V as an ODE and solve it.

Example 2.11. Let's choose the following right-hand side functions. For a location-independent ODE, we let

$$P(t, v, x) := G(v);$$

then the equation describes the location in terms of the velocity:

$$df(A, AB) = G(AB).$$

For a time-independent ODE, we let

$$P(t, v, x) := kx;$$

then the equation describes population growth:

$$df(A, AB) = kf(A). \qquad \square$$

The nature of the process often dictates that the way a quantity changes depends only on its current value, and not on time. As a result, the right-hand side function P is often independent of the first argument. This will be our assumption below.

Let's simplify more. There are only two directions in time $\mathbb{R}$, so the ODE $df = Pf$ stands for the following combination of two (one for each direction) equations with respect to local forms:

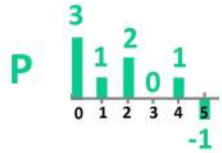
$$df\Big(A, [A, A+1]\Big) = P\Big([A, A+1], f(A)\Big),\ df\Big(A, [A, A-1]\Big) = P\Big([A, A-1], f(A)\Big).$$

They are independent of each other and it makes sense to study just one of them. We choose the former: *forward propagation.* In fact, the two equations are identical if we assume that

$$\Big(A, [A, A+1]\Big) = -\Big(A, [A, A-1]\Big).$$

In other words, if times reverses, so does the change.

Then the right-hand function can be seen as a function of one variable $P : R \to R$:



As a result, df is simply a cubical 1-cochain:

$$df : [A, A+1] \mapsto r \in R.$$

Definition 2.12. An *initial value problem (IVP)* is a combination of an ODE and an *initial condition (IC)*:

$$df = Pf, \ f(A_0) = x_0 \in R.$$

Then a 0-form f on the ray $\mathbb{R} \cap \{A \geq A_0\}$ is called a (forward) *solution* of the IVP if it satisfies:

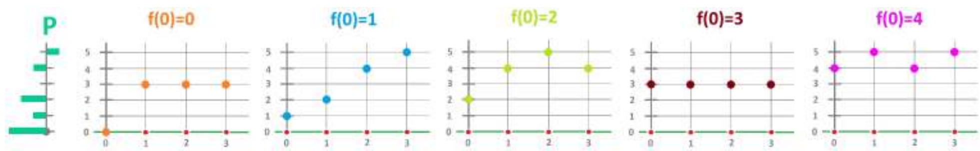
$$df\Big([A, A+1]\Big) = P(f(A)), \ \forall A \geq A_0.$$

Because the exterior derivative in this setting is simply the difference of values, a solution f is easy to construct iteratively.

Theorem 2.13 (Existence). The following is a solution to the IVP above:

$$f(A_0) := x_0, \ f(A+1) := f(A) + P(f(A)), \ \forall A \geq A_0.$$

A few solutions for the P shown above are illustrated below:



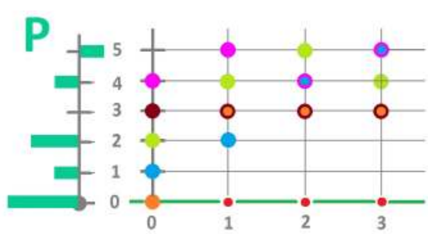
Exercise 2.14. Provide an analog of this iterative procedure for an IVP with domain a graph K and find out under what circumstances this procedure works or fails.

Exercise 2.15. Define (a) a backward solution and (b) a two-sided solution of the IVP and (c) devise iterative procedures to construct them. (d) Do the three match?

Because of the existence property, the solutions to all possible IVPs fill the space, i.e., R.

Theorem 2.16 (Uniqueness). The solution to the IVP given above is the only one.

When we plot the above solutions together, we see that they do overlap:



This reminds us that only the right-sided uniqueness is guaranteed: whenever two solutions meet, they stay together.

Exercise 2.17. Find necessary conditions for the space to be filled, in an *orderly* manner:

Next, is there anything we can we say about continuity?

Our ring R has no topology. On the other hand, if we choose $R = \mathbf{R}$, *every* 0-form f^0 on $\mathbb{R}$ can be extended to a continuous function on $f : \mathbf{R} \to \mathbf{R}$. In contrast to this trivial conclusion, below continuity appears in a meaningful way.

Definition 2.18. For the given ODE, the *forward propagation map* of depth $c \in \mathbf{Z}^+$ at A_0 is a map

$$Q_c : R \to R$$

defined by

$$Q_c(x_0) := f(A_0 + c),$$

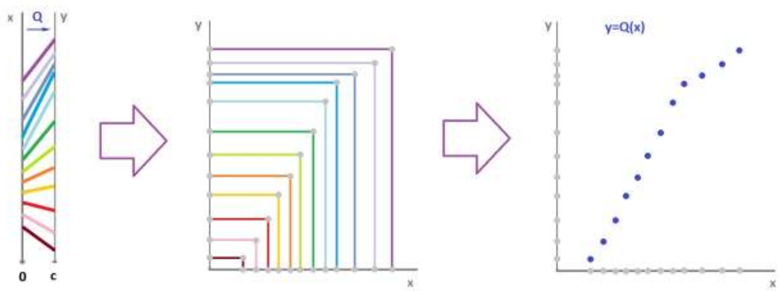
where f is the solution of the IC:

$$f(A_0) = x_0.$$

In other words, the forward propagation map is a self-map of the space of locations.

Exercise 2.19. Prove that Q_c is independent of A_0.

Theorem 2.20 (Continuous dependence on initial conditions). Suppose $R = \mathbf{R}$. If P is continuous on the second argument, the forward propagation map $Q_c : \mathbf{R} \to \mathbf{R}$ is continuous for any $c \in \mathbf{Z}^+$ and any $A_0 \in \mathbb{R}$.



Exercise 2.21. Prove the theorem.

Exercise 2.22. Define and analyze an IVP for an ODE of order 2.

2.6 Vector fields and systems of ODEs

The IVP we have studied is:

$$df = Pf, \ f(A_0) = x_0 \in R.$$

Here a 0-form f, which represents the variable quantity to be found, is defined on the ray $\mathbb{R} \cap \{A \geq 0\}$ and satisfies the following:

$$df\Big(A, [A, A+1]\Big) = P(f(A)), \ \forall A \in \mathbf{Z}^+.$$

A solution is given by

$$f(0) := x_0, \ f(A+1) := f(A) + P(f(A)), \ \forall A \geq 0.$$

What if we have *two* variable quantities?

Example 2.23. The simplest example is as follows:

- x is the horizontal location, and
- y is the vertical location.

Both are 0-forms on $\mathbb{R}$. Their respective displacements are dx and dy. If either depends only on its own component of the location, the motion is described by this pair of ODEs of forms:

$$dx = g(x),\ dy = h(y).$$

The solution consists of two solutions to the two, unrelated, ODEs either found with the formula above. □

The solution is easy because the two components are independent of each other.

Example 2.24 (predator-prey model). As an example of quantities that do interact, let x be the number of the prey and y be the number of the predators in the forest. Let $R := \mathbf{R}$.

The prey have an unlimited food supply and reproduces geometrically as described above – when there is no predator. Therefore, the gain of the prey population per unit of time is αx for some $\alpha \in \mathbf{R}^+$. The rate of predation upon the prey is assumed to be proportional to the rate at which the predators and the prey meet, which, in turn, is proportional to xy. Therefore, the loss per unit of time is βxy for some $\beta \in \mathbf{R}^+$. Then the prey's ODE is:

$$dx = \alpha x - \beta xy.$$

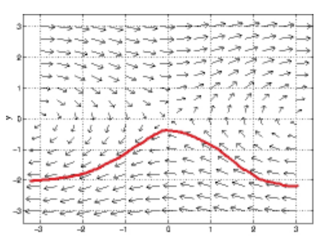
The predators have only a limited food supply, i.e., the prey. We use, again, the rate of predation upon the prey proportional to xy. Then the gain of predator population per unit of time is δxy for some $\delta \in \mathbf{R}^+$. Without food, the predators die out geometrically as described previously. Therefore, the loss per unit of time is γy for some $\gamma \in \mathbf{R}^+$. The predator's ODE is:

$$dy = \delta xy - \gamma y.$$

This system of these two, dependent, ODEs cannot be solved by reusing the formula for the 1-dimensional case. □

Exercise 2.25. (a) Set up and solve the IVP for this system. (b) Find its equilibria.

Example 2.26. We also take as a model a fluid flow. The "phase space" $\mathbf{R}^2$ is the space of all possible locations. Then the position of a given particle is a function $u : \mathbf{Z}^+ \to \mathbf{R}^2$ of time $t \geq 0$. Meanwhile, the dynamics of the particle is governed by the velocity of the flow, at each moment of time. Let $V : \mathbf{R}^2 \to \mathbf{R}^2$ be a function. Then the velocity of a particle if it happens to be at point u is $V(u)$ and the next position may be seen as $u + V(u)$.



□

A *vector field* is given when there is a vector attached to each point of the plane:

$$\text{point} \mapsto \text{vector}.$$

Then it is just a function

$$V : \mathbf{R}^2 \to \mathbf{R}^2.$$

Further, one can think of a vector field as a system of time-independent ODE:

$$\begin{cases} dx = g(x, y), \\ dy = h(x, y), \end{cases}$$

where $g, h : \mathbf{R}^2 \to \mathbf{R}$ are functions and $x, y : \mathbf{Z} \to \mathbf{R}$ are 0-forms. The corresponding IVP adds an initial condition:

$$\begin{cases} x(t_0) = x_0, \\ y(t_0) = y_0. \end{cases}$$

As before, a (forward) solution to this IVP is a 0-form on the ray $\mathbb{R} \cap \{A \geq 0\}$ that satisfies, $\forall A \geq 0$:

$$\begin{aligned} dx\Big(A, [A, A+1]\Big) &= g(x(A), y(A)), \\ dy\Big(A, [A, A+1]\Big) &= h(x(A), y(A)). \end{aligned}$$

Next, we combine the two variables to form a vector:

$$\begin{aligned} u &:= (x, y) \in \mathbf{R}^2, \\ du &:= (dx, dy) \in \mathbf{R}^2. \end{aligned}$$

Note that the new vector variables still form a ring!

Then the setup given at the top of the subsection:

$$df = Pf, \ f(A_0) = x_0 \in R,$$

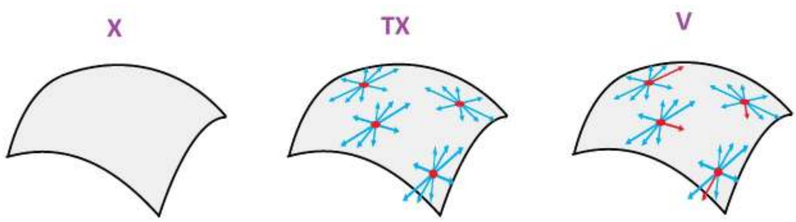
is now applicable to vector fields: in all the three examples, we set

$$R := \mathbf{R}^2.$$

More generally, a vector field supplies a *direction to every location*. In a cell complex X, it is a map from X to its tangent bundle:

$$T(X) := \bigsqcup_{A \in X} \Big(\{A\} \times T_A(X)\Big).$$

Thus, there is one vector at each point picked from a whole vector space:



We next make this idea more precise.

Definition 2.27. The *bundle projection* of complex X is the map

$$\pi = \pi_X : T(X) \to X$$

given by

$$\pi(A, v) := A.$$

Proposition 2.28. If $i_X : X \hookrightarrow T(X)$ is the *inclusion*, $i_X(A) = (A, 0)$, then

$$\pi_X i_X = \mathrm{Id}_X .$$

Definition 2.29. A *vector field* over complex X is any function

$$V : X \to T(X)$$

that is a section (i.e., a right inverse) of the bundle projection:

$$\pi_X V = \mathrm{Id}_X .$$

Exercise 2.30. What algebraic structure do the vector fields over X form?

Note that the inclusion serves as the trivial (zero) vector field.

A typical vector field is simply

$$V(A) = AB \in X,$$

where $B = B(A)$ is a vertex adjacent to A that depends on A. Then the motion produced by the vector field takes a particle from vertex to vertex along an edge, or simply:

$$A \mapsto B.$$

Its solution can be written as an iteration:

$$A_{n+1} := A_n + \partial_1 V(A_n),$$

where $\partial_1 : C_1(X) \to C_0(X)$ is the boundary operator of X.

Exercise 2.31. Verify the last statement. What kind of function is this?

2.7 Simulating a flow with a spreadsheet

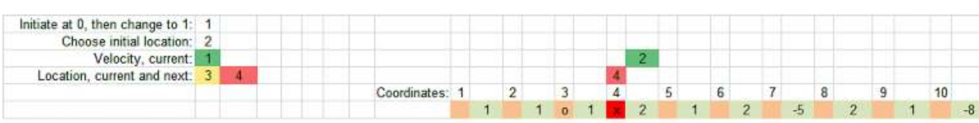
Spreadsheets, such as Excel, can be used to model various dynamics and some of the processes can also be visualized.

We represent a flow with a vector field starting with *dimension* 1; i.e., $R := \mathbf{Z}$. This vector field is the right-hand side function $P : \mathbf{Z} \to \mathbf{Z}$ of our ODE:

$$df = Pf.$$

The number in each cell tells us the velocity of a particle if it is located in this cell. The values of the velocities are provided in the spreadsheet. They were chosen in such a way that the particle can't leave this segment of the spreadsheet. They can be changed manually.

Then one can trace the motion of a single particle given by a 0-form $f : \mathbf{Z}^+ \to \mathbf{Z}$, shown as a red square:



Here, the velocities are seen in the elongated cells (edges), while the locations of the particles are the small square cells (vertices). Also,

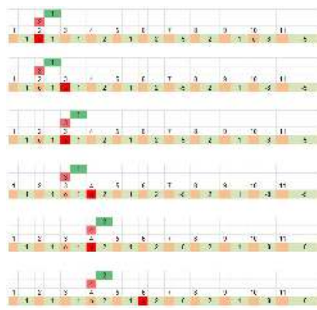
- "x" stands for the current position,
- "o" for the previous position.

The formula is as follows:

```
= IF( R[-1]C=R9C3,''x'', IF(R[-1]C=R9C2, ''o'','' '') )
```

The data, as it changes, is shown on the left.

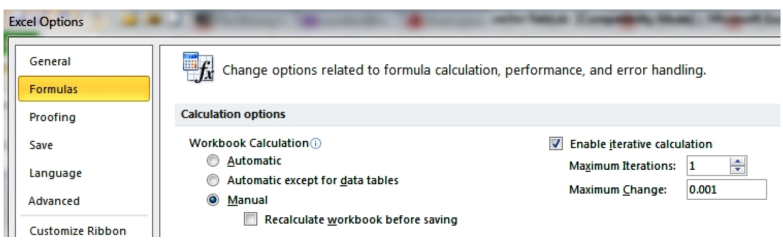
For motion simulation, we carry out one iteration for every two pushes of the button. This is an example of such dynamics with three iterations:



Note: See the files online.

For the file to work as intended, the settings should be set, under "Formulas":

- Workbook Calculation: Manual,
- Maximum Iterations: 1.



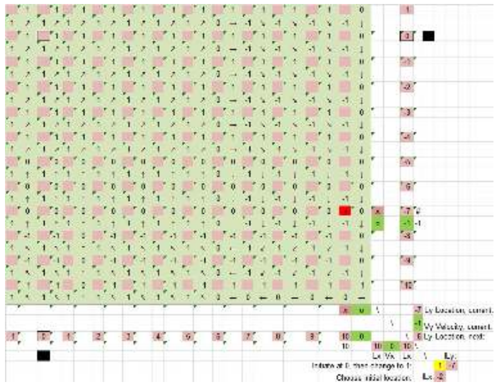
To see the motion of the particle, simply push "Calculate now" as many times as necessary.

Exercise 2.32. Modify the code so that the spreadsheet models motion on the circle, no matter what velocities are chosen. Hint: choose an appropriate ring R.

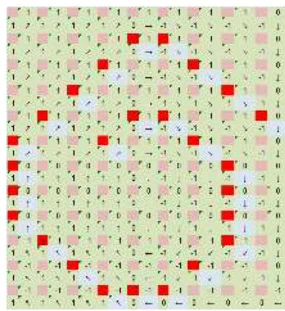
Exercise 2.33. Modify the code so that, no matter what velocities are chosen, the object stays within, say, interval $[0, ..., 7]$, and bounces off its ends.

Next, we simulate a flow with a vector field in *dimension* 2; i.e., $R := \mathbf{Z}^2$. This vector field is the right-hand side function $P : \mathbf{Z}^2 \to \mathbf{Z}^2$ of our ODE. Each pair (right and down) of elongated cells starting at every square cell contains a pair of numbers that represents the velocity of a particle located in this cell.

The vector field of dimension 2 is made of two vector fields of dimension 1 (on the same square) independently evaluated for the x and y variables. The components of the velocities are chosen to be only -1, 0, or 1 (illustrated with the little arrows) so that the jump is always to an adjacent location. They were also chosen in such a way that the particle can't leave this rectangle.



We trace the motion of a single particle given by a 0-form $f:\mathbf{Z}^+\to\mathbf{Z}^2$, shown as a red "x". Several iterations are shown in a single picture below:



Exercise 2.34. Modify the code to model motion on (a) the cylinder and (b) the torus.

Exercise 2.35. Modify the code so that, no matter what velocities are chosen, the object stays within, say, rectangle $[0,8]\times[0,7]$, and bounces off its walls.

2.8 The derivative of a cell map

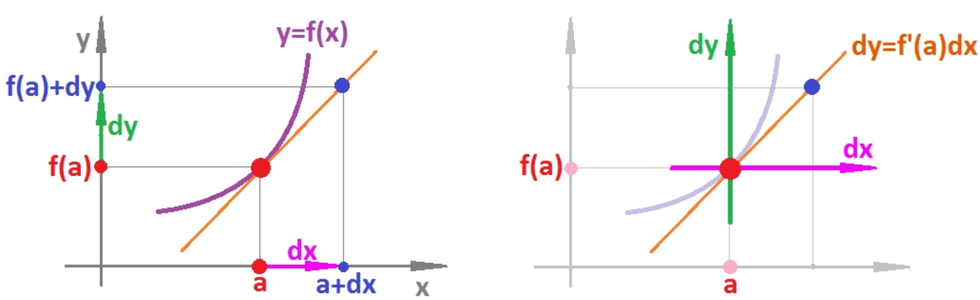
Consider the two standard ways to write the derivative of function f at $x=a$:

$$\tfrac{dy}{dx}=f'(a).$$

What we know from calculus is that the left-hand side is *not* a fraction but the equation can be rewritten as if it is:

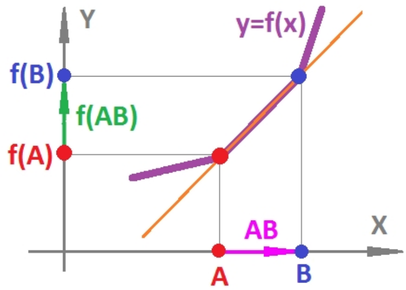
$$dy=f'(a)dx.$$

The equation represents the relation between the increment of x and that of y – *in the vicinity* of a. This information is written in terms of a new coordinate system, (dx,dy) and the best affine approximation (given by the tangent line) becomes a linear function in this system:

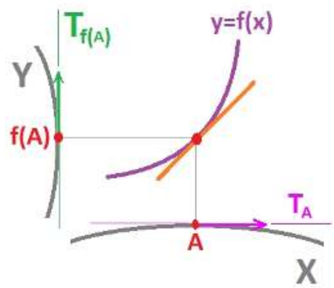


Things are much simpler in the discrete case.

Suppose X and Y are two cell complexes and $f : X \to Y$ is a cell map. Then "in the vicinity of point a" becomes "in the star of vertex A":



In fact, we can ignore the algebra of the x- and y-axis if we think of our equation as a relation between the elements of the *tangent spaces* of A and $f(A)$. If we zoom out on the last picture, this is what we see:

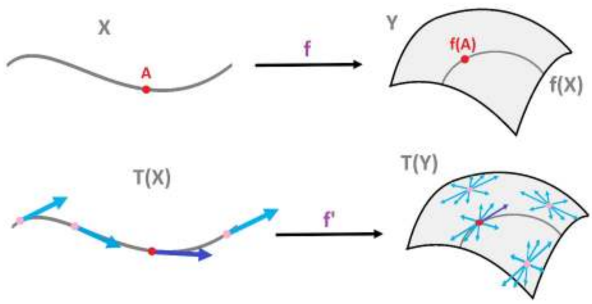


Then the above equation becomes:

$$e_Y = f(AB)e_X,$$

where e_X, e_Y are the basis vectors of $T_A(X), T_{f(A)}(Y)$, respectively.

The idea is: our function maps both locations *and* directions. The general case is illustrated below:



A cell map takes vertices to vertices and edges to edges and that's what makes the 0- and 1-chain maps possible. Then,

- the locations are taken care of by $f_0 : C_0(X) \to C_0(Y)$, and
- the directions are taken care of by $f_1 : C_1(X) \to C_1(Y)$.

Suppose also that $f(A) = P$, so that the location is fixed for now. Then the tangent spaces at these vertices are:

$$T_A(X) := < \{AB \in X\} > \subset C_1(X), \quad T_P(Y) := < \{PQ \in Y\} > \subset C_1(Y).$$

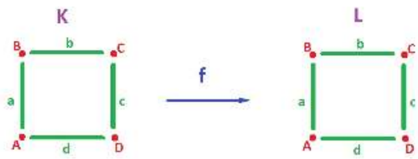
Definition 2.36. The *derivative of a cell map f at vertex A* is a linear map

$$f'(A) : T_A(X) \to T_P(Y)$$

given by

$$f'(A)(AB) := f_1(AB).$$

Example 2.37. Let's consider cell maps of the "cubical circle" (i.e., $\mathbf{S}^1$ represented by a 4-edge cubical complex) to itself, $f : X \to X$:



Given a vertex, we only need to look at what happens to the edges adjacent to it. We assume that the bases are ordered according to their letters, such as $\{AB, BC\}$.

The derivatives of these functions are found below.

Identity:

$$\begin{array}{llll} f_0(A) = A, & f_0(B) = B, & f_0(C) = C, & f_0(D) = D, \\ \Longrightarrow & f'(A)(AB) = AB, & f'(A)(AD) = AD. & \end{array}$$

It's the identity map.

Constant:

$$\begin{array}{llll} f_0(A) = A, & f_0(B) = A, & f_0(C) = A, & f_0(D) = A, \\ \Longrightarrow & f'(A)(AB) = AA = 0, & f'(A)(AD) = AA = 0. & \end{array}$$

It's the zero map.

Vertical flip:

$$\begin{array}{llll} f_0(A) = D, & f_0(B) = C, & f_0(C) = B, & f_0(D) = A, \\ \Longrightarrow & f'(A)(AB) = DC, & f'(A)(AD) = DA. & \end{array}$$

The matrix of the derivative is

$$f'(A) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

□

Exercise 2.38. Repeat these computations for (a) the rotations; (b) the horizontal flip; (c) the diagonal flip; (d) the diagonal fold. Hint: the value of the derivative varies from point to point.

As this construction is carried out for each vertex in X, we have defined a function on the whole tangent bundle.

Definition 2.39. The *derivative of cell map*

$$f : X \to Y$$

between two cell complexes is the map between their tangent bundles,

$$f' : T(X) \to T(Y),$$

given by

$$f'(A, AB) := (f_0(A), f_1(AB)).$$

Exercise 2.40. In this context, define the directional derivative and prove its main properties.

Theorem 2.41 (Properties of the derivative). For a given vertex and an adjacent edge, the derivative satisfies the following properties:

- The derivative of a constant is zero in the second component:

$$(C)' = (C, 0),\ C \in Y.$$

- The derivative of the identity is the identity:

$$(\mathrm{Id})' = \mathrm{Id}\,.$$

- The derivative of the composition is the composition of the derivatives:

$$(fg)' = f'g'.$$

- The derivative of the inverse is the inverse of the derivative:

$$(f^{-1})' = (f')^{-1}.$$

Exercise 2.42. Prove the theorem.

Exercise 2.43. Prove that if $|f|$ is a homeomorphism, then $f' = \mathrm{Id}$.

Notation: An alternative notation for the derivative is Df. It is also often understood as the *tangent map* of f denoted by $T(f)$.

Exercise 2.44. Show that T is a functor.

We have used the equivalence relation

$$(A, AB) \sim (B, -BA)$$

to glue together the tangent spaces of a cell complex:

$$T(K)/_{\sim} = C_1(K).$$

Theorem 2.45. Suppose X, Y are two cell complexes and $f : X \to Y$ is a cell map. Then the quotient map of the derivative

$$[f'] : T(X)/_{\sim} \to T(Y)/_{\sim}$$

is well-defined and coincides with the 1-chain map of f,

$$f_1 : C_1(X) \to C_1(Y).$$

Proof. Suppose $f_0(A) = P$ and $f_0(B) = Q$. Then we compute:

$$\begin{array}{ll} f'(A, AB) & = (f_0(A), f_1(AB)) \\ & = (P, PQ) \\ & \sim (Q, -QP) \\ & = (f_0(B), -f_1(BA)) \\ & = f'(B, -BA). \end{array}$$

We have proven the following:

$$(A, AB) \sim (B, -BA) \Longrightarrow f'(A, AB) \sim f'(B, -BA).$$

Therefore, $[f']$ is well-defined. ■

2.9 ODEs of cell maps

So far, we have represented motion by means of a 0-form, as a function

$$f : \mathbb{R}^+ \to R,$$

where R is our ring of coefficients. Now, is this motion *continuous*?

Here, R is our space of locations and it is purely algebraic. Even if we extend f from $\mathbb{R}^+$ to $\mathbf{R}^+$, without a topology on R, the continuity of such an f has no meaning! In contrast, motion is typically represented by a *parametric curve*, i.e., a continuous map

$$p : \mathbf{R}^+ \to U \subset \mathbf{R}^n,$$

with $\mathbf{R}^+$ serving as the time and U as the space.

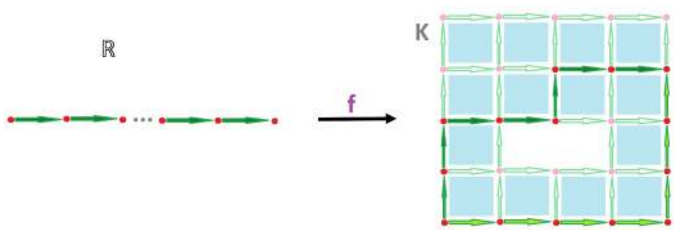
If we choose to recognize $R = \mathbf{R}^n$ as a topological space with the Euclidean topology, then, on the flip side, its subsets, such as $U = \mathbf{R}^n \setminus \{0\}$, may lack an algebraic structure. For example, these rings aren't *rings*:

Then we can't use forms anymore.

To match the standard approach, below we represent motion in cell complex K by a cell map,

$$f : \mathbb{R} \supset I \to K,$$

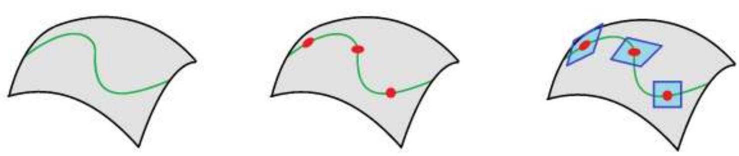
where I is a cell complex, typically an interval.



This motion is then continuous in the sense that this cell map's realization is a continuous map:

$$|f| : \mathbf{R} \supset |I| \to |K|.$$

We can still see forms if we look at the curve one point at a time:



Next, we will demonstrate how these curves are generated by vector fields. The analysis largely follows the study of ODEs of forms discussed previously, in spite of the differences. Compare their derivatives in the simplest case:

- for a 0-form $f : K \to R$, we have

$$df\Big([A, A+1]\Big) = f(A+1) - f(A) \in R;$$

- for a cell map $f:\mathbb{R}\to\mathbb{R}$, we have

$$f'\Big(A,[A,A+1]\Big)=\Big(f_0(A),f_1([A,A+1])\Big)\in C_0(\mathbb{R};R)\times C_1(\mathbb{R};R).$$

Definition 2.46. Suppose $P(A,\cdot)$ is a vector field on complex K parametrized by $A\in\mathbf{Z}$. Then an *ODE of cell maps* generated by P on complex K is:

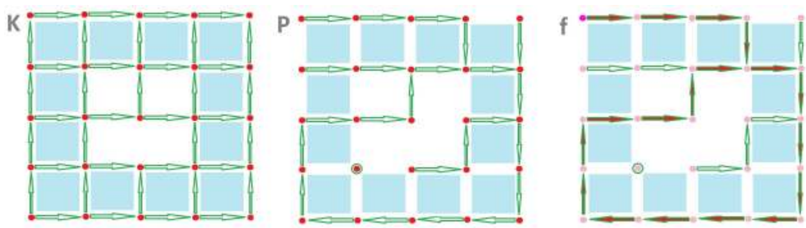
$$f'=Pf,$$

and its *solution* is a cell map $f:\mathbb{R}\supset I\to K$ that satisfies:

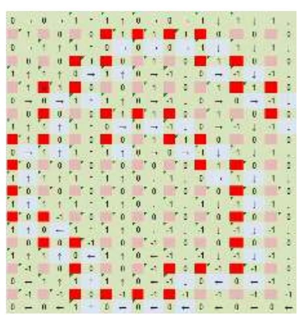
$$f'\Big(A,[A,A+1]\Big)=P(A,f(A)),\ \forall A\in I\cap\mathbf{Z}.$$

For $K=\mathbb{R}$, we can simply take the examples of 1st order ODEs of forms given in this section and use them to illustrate ODEs of maps. We just need to set $R:=\mathbf{Z}$. The difference is, this time the values of the right-hand side P can only be 1, 0, or -1.

For $K=\mathbb{R}^2$, the cubical case illustrated with a spreadsheet above doesn't apply anymore because a cell map has to follow the edges and can't jump diagonally. Then one of the components of the vector field has to be 0, as shown below:



Example 2.47. We change the values of the vector field in the spreadsheet above as described above in order to model an ODE of cubical maps. Several iterations can be shown in a single picture, below:



□

Exercise 2.48. Can the original spreadsheet be understood as a model of an ODE of cell maps?

Thus, if we represent motion with an ODE of cell maps, *both time and space are topological.*

Definition 2.49. An *initial value problem (IVP)* on complex K is a combination of an ODE and an *initial condition (IC)*:

$$f'=Pf,\ f(A_0)=x_0\in K,\ A_0\in\mathbf{Z},$$

and a solution $f:\mathbb{R}\cap\{t\geq A_0\}\to K$ of the ODE is called a *solution of the IVP* if it satisfies the initial condition.

Theorem 2.50 (Uniqueness). The solution to the IVP above, if exists, is given iteratively by

$$f(A_0) := x_0,\ f(A+1) := f(A) + \partial P(A, f(A)),\ \forall A \geq A_0.$$

Proof. Suppose f is a solution, $f' = Pf$. Then we compute:

$$\begin{aligned} f_0(A+1) - f_0(A) &= f_0((A+1) - (A)) \\ &= f_0\Big(\partial[A, A+1]\Big) \\ &= \partial f_1[A, A+1] \\ &= \partial P(A, f(A)). \end{aligned}$$

■

Just as before, only the right-sided uniqueness is guaranteed: when two solutions meet, they stay together.

What about existence? We have the formula, but is this a cell map? It can only be guaranteed under special restrictions.

Theorem 2.51 (Existence). If the values of the vector field P are *edges* of K, a solution to the IVP above always exists.

Proof. For f to be a cell map, $f(A)$ and $f(A+1)$ have to be the endpoints of an edge in K. Clearly, they are, because $f(A+1) - f(A) = \partial P(A, f(A))$. ■

Definition 2.52. For a given ODE, the *forward propagation map* of depth $c \in \mathbf{Z}^+$ at A_0 is a map

$$Q_c : K \to K$$

defined by

$$Q_c(x_0) := f(A_0 + c),$$

where f is the solution with the IC:

$$f(A_0) = x_0.$$

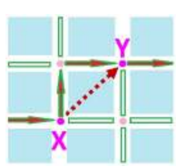
Exercise 2.53. Prove that if the vector field P is time-independent, Q_c is independent of A_0.

Exercise 2.54. Compute the forward propagation map for the vector field illustrated above.

Exercise 2.55. Suppose P is edge-valued. Is the resulting forward propagation map $Q_c : K \to K$ a cell map?

2.10 ODEs of chain maps

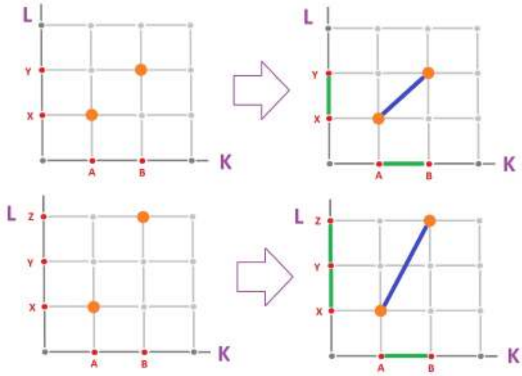
A cell map can't model jumping diagonally across a square.



That's why the existence theorem above requires the values of the vector field to be edges. Then, in the case of $\mathbb{R}^n$, all but one of the components must be 0.

This is a significant limitation of ODEs of cell maps, even in comparison to ODEs of forms.

The issue is related to one previously discussed: cell map extensions vs. chain map extensions (subsection V.3.10). Recall that in the former case, extensions may require subdivisions of the cell complex. The situation when the domain is 1-dimensional is transparent:



In the former case, we can create a cell map:

$$g(AB) := XY,$$

by extending its values from vertices to edges. In the latter case, an attempt of cell extension (without subdivisions) fails as there is no *single* edge connecting the two vertices. However, there is a *chain* of edges:

$$g(AB) := XY + YZ.$$

Even though the linearity cannot be assumed, the illustration alone suggests a certain continuity of this new "map". In fact, we know that chain maps are continuous in the algebraic sense: they preserve boundaries,

$$g_0 \partial = \partial g_1.$$

The idea is also justified by the meaning of the derivative of a cell map f:

$$f'\Big(A, [A, A+1]\Big) = \Big(f_0(A), f_1([A, A+1])\Big).$$

It is nothing but a combination of the 0- and the 1-chain maps of f...

Now, ODEs.

Once again, a vector field at vertex X only takes values in the tangent space $T_X(K)$ at X. We need to incorporate the possibility of moving in the direction $XY + YZ$ (while $YZ \notin T_X(K)$). From this point on, the development is purely algebraic.

Suppose we are given a (short) chain complex C to represent *space*:

$$\partial : C_1 \to C_0.$$

Here, of the two modules, C_0 is meant to represent the chains of locations and C_1 the chains of directions. Even though all the information is contained in the complex, we still define an analog of the tangent bundle to demonstrate the parallels with ODEs of forms and cell maps.

Definition 2.56. The *chain bundle* $T(C)$ of chain complex C is the module

$$T(C) := C_0 \oplus C_1;$$

while the *bundle projection*

$$\pi = \pi_C : T(C) \to C_0, \ \pi(X, v) = X,$$

is simply the projection of the product.

Furthermore, if $i_{C_0} : C_0 \hookrightarrow T(C)$ is the *inclusion*, $i_{C_0}(X) = (X, 0)$, then

$$\pi_C i_{C_0} = \mathrm{Id}_{C_0}.$$

Definition 2.57. A *chain field* over chain complex C is any linear map

$$W : C_0 \to T(C)$$

that is a section of the bundle projection:

$$\pi_C W = \mathrm{Id}_{C_0} .$$

The meaning is simple:

$$W(X) = (X, P(X)),$$

for some $P : C_0 \to C_1$.

Here is an illustration of possible values of W at vertex X (even though the chains don't have to start at X):

The boundary operator of C is extended to the *boundary operator of the chain bundle*:

$$\partial : T(C) \to C_0,$$

by setting

$$\partial(X, a) := \partial a.$$

The dynamics produced by the chain field is a sequence of 0-chains given by this iteration:

$$X_{n+1} := X_n + \partial W(X_n) = X_n + \partial P(X_n).$$

The chain bundle of a cell complex K is $T(C(K))$. The following is obvious.

Proposition 2.58. A vector field

$$V : K^{(0)} \to T(K)$$

over a cell complex generates – by linear extension – a chain field over its chain bundle:

$$W : C_0(K) \to T(C(K)).$$

With the new definition, the motion produced by such a chain field is as before. For example, if $W(X) = XY$, then

$$X_1 = X + \partial W(X) = X + (Y - X) = Y.$$

Definition 2.59. Suppose $W(A, \cdot)$ is a chain field on a chain complex C parametrized by $A \in \mathbf{Z}$ (for time). Then an *ODE of chain maps* generated by W on complex C is:

$$g_1 = W g_0,$$

and its *solution* is a chain map

$$g = \{g_0, g_1\} : C(I) \to C,$$

for some subcomplex $I \subset \mathbb{R}$, that satisfies the equation

$$g_1\Big([A, A+1]\Big) = W(g_0(A)), \ \forall A \in I \cap \mathbf{Z}.$$

Example 2.60. Under this definition, a step of the moving particle on a square grid can be diagonal, without explicitly following the vertical and horizontal edges. For example, to get from $X = (0, 0)$ to $Z = (1, 1)$ without stopping at $Y = (0, 1)$, set

$$W(X) := (X, XY) + (Y, YZ) \in T(C).$$

Indeed, we have

$$X \mapsto X + \partial W(X) = X + (Y - X) + (Z - Y) = Z. \qquad \square$$

Since every cell map generates its chain map, we have the following:

Proposition 2.61. If a map f is a solution of the ODE of cell maps generated by a vector field V, then $g := f_*$ is a solution of the ODE of chain maps generated by the chain field $W = V_*$.

Definition 2.62. An *initial value problem* is a combination of an ODE and an *initial condition*:

$$g_1 = W g_0, \;\; g_0(A_0) = X_0 \in C, \;\; A_0 \in \mathbf{Z},$$

and a solution $g : C(\mathbb{R} \cap \{t \geq A_0\}) \to C$ of the ODE is called a (forward) *solution of the IVP* if it satisfies the initial condition.

Theorem 2.63 (Existence and Uniqueness). The only (forward) solution to the IVP above is given iteratively by

$$\begin{array}{ll} g_0(A_0) := X_0, & \text{and for all } A \geq A_0, \\ g_0(A+1) := g_0(A) + \partial W(A, g_0(A)), & \\ g_1\Big([A, A+1]\Big) := W(A, g_0(A)). & \end{array}$$

Proof. The linearity of W on the second argument implies that g_0, g_1 so defined form a chain map. The proof that the ODE holds is identical to the proof of the uniqueness theorem in the last subsection. ■

The uniqueness holds only in terms of chains. When applied to cell complexes, the result may be that the initial location is a vertex X_n but the next "location" X_{n+1} is not.

Example 2.64. Where do we go from $X = (0, 0)$ if

$$W(X) := (X, XY) + (X, XZ),$$

with $Y = (0, 1), Z = (1, 1)$? This is the result:

$$X \mapsto X + \partial W(X) = X + (Y - X) + (Z - X) = -X + Y + Z. \qquad \square$$

This isn't a cell map because the output isn't a single vertex! The latter can be guaranteed though under certain constraints on the equation.

2.11 Cochains are chain maps

Since every cell map generates its chain map, the existence theorem from the last subsection is still applicable: if a vector field V is edge-valued, a solution to the IVP for maps always exists. Below is a stronger statement.

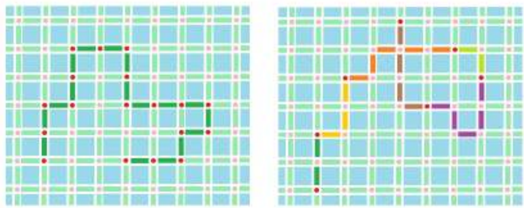
Theorem 2.65 (Cell maps as solutions). If a chain field over the chain complex $C(K)$ of a cell complex K satisfies the condition:

$$\partial W(A, X) = Y - X, \text{ for some } Y \in K^{(0)},$$

then the solution to the IVP for chain maps on K is a cell map.

The difference between a cell map as a solution of
- an ODE of cell maps and
- an ODE of chain maps

is illustrated below:



We have thus demonstrated that ODEs of *cell maps* are included in the new setting of ODEs of chain maps. What about ODEs of *forms*?

Suppose we are given f, a 0-form on $\mathbb{R}$. Then we would like to interpret the pair $g = \{f, df\}$ as some chain map defined on $C(\mathbb{R})$, the chain complex of time. What is the other chain complex C, the chain complex of space? Since these two forms take their values in ring R, we can choose C to be the trivial combination of two copies of R:

$$\partial = \mathrm{Id} : R \to R.$$

Below, we consider a more general setting of k-forms.

Theorem 2.66. Cochains are chain maps, in the following sense: for every k-cochain f on K, there is a chain map from $C(K)$ to the chain complex C with only one non-zero part, $\mathrm{Id} : C_{k+1} = R \to C_k = R$, as shown in the following commutative diagram:

$$\begin{array}{r|ccccccccc}
C(K): & ... & C_{k+2}(K) & \xrightarrow{\partial} & C_{k+1}(K) & \xrightarrow{\partial} & C_k(K) & \xrightarrow{\partial} & C_{k-1}(K) & ... \\
f: & & \downarrow{0} & & \downarrow{df} & & \downarrow{f} & & \downarrow{0} & \\
C: & ... & 0 & \xrightarrow{\partial=0} & R & \xrightarrow{\partial=\mathrm{Id}} & R & \xrightarrow{\partial=0} & 0 & ...
\end{array}$$

Proof. We need to prove the commutativity of each of these squares. We go diagonally in two ways and demonstrate that the result is the same. We use the duality $d = \partial^*$.

For the first square:

$$df\partial = (\mathrm{Id}^{-1} f\partial)\partial = \mathrm{Id}^{-1} f0 = 0.$$

For the second square:

$$f\partial = df = \mathrm{Id}\, df.$$

The third square (and the rest) is zero. ■

Now, how do ODEs of forms generate ODEs of chain maps?

Suppose $df = Pf$ is an ODE of forms with $P : R \to R$ some function. Then, for $K = \mathbb{R}$ and $k = 0$, the above diagram becomes:

$$\begin{array}{r|cccccc}
C(\mathbb{R}): & ... \xrightarrow{0} & C_1(\mathbb{R}) & \xrightarrow{\partial} & C_0(\mathbb{R}) & \xrightarrow{0} \\
f: & & \downarrow{df} & & \downarrow{f} & \\
C: & ... \xrightarrow{0} & R & \xrightarrow{\mathrm{Id}} & R & \xrightarrow{0}
\end{array}$$

The corresponding chain field,

$$W : C_0 = R \to T(C) = R \times R,$$

is chosen to be

$$W(A) := (A, P(A)).$$

Proposition 2.67. A solution of an ODE of forms is a solution of the corresponding ODE of chain maps.

Exercise 2.68. Prove the proposition.

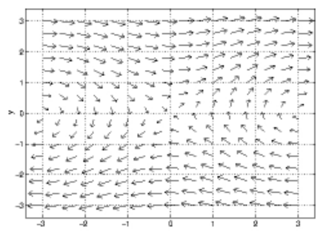
Thus, the examples at the beginning of this section can be understood as ODEs of chain maps.

Exercise 2.69. Provide chain fields for the examples of ODEs in this section.

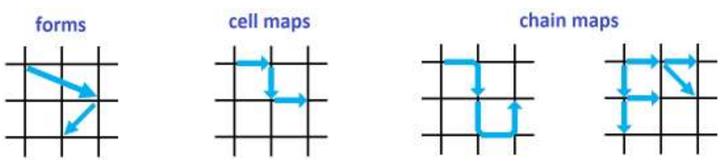
Exercise 2.70. (a) Define the forward propagation map for ODEs of chain maps. (b) Explain its continuity.

2.12 Simulating advection with a spreadsheet

We have shown that we can model a flow with a *discrete vector field* that provides the velocities of the flow:



We have also modeled motion with cell maps and chain maps:



The last example of an ODE of chain maps, however, gives us something new.

Example 2.71. Suppose our chain field on $\mathbb{R}^2$ is given by

$$P(X) := \tfrac{1}{2}(X, XY) + \tfrac{1}{2}(X, XZ).$$

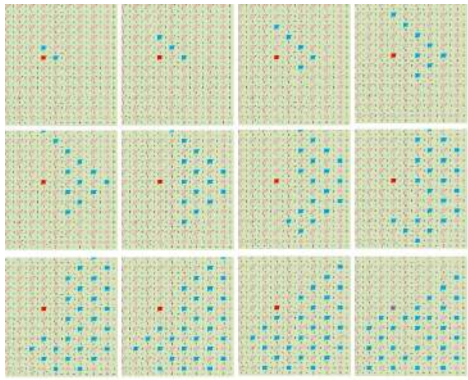
with $Y = (1, 0), Z = (0, 1)$. Where do we go from $X = (0, 0)$? This is the result:

$$X \mapsto X + \partial P(X) = X + \tfrac{1}{2}(Y - X) + \tfrac{1}{2}(Z - X) = \tfrac{1}{2}Y + \tfrac{1}{2}Z.$$

We can interpret this outcome as if some material initially located at X has been split between the two adjacent edges and these halves ended up in Y and Z. $\square$

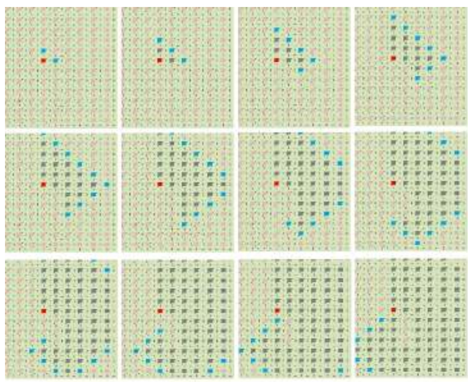
A process that transfers material along a vector field is called *advection*.

Example 2.72. To model advection, we can use the vector field on the spreadsheet given above. It is understood differently though. If the values at the two adjacent to X cells are 1 and 1, it used to mean that the particle follows vector $(1, 1)$. Now it means that the material is split and each half goes along one of the two directions. Several iterations of this process are shown with the new cells with non-zero values shown in blue:



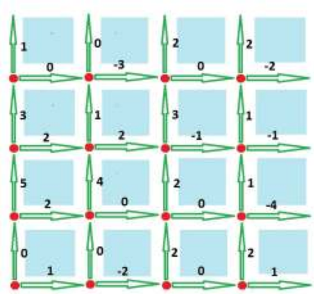
That's what a forward propagation map of a chain ODE looks like!

The coloring in the illustration does not reflect the fact that the material is spread thinner and thinner with time. That's why the simulation also resembles spreading of an *infection* – along preset directions – with blue squares indicating newly infected individuals, as below:



□

For a more direct interpretation of the spreadsheet, imagine that a liquid moves along the square grid which is thought of as a *system of pipes*:



Then, instead of the velocity vector attached to each cell, we think of our vector field as two numbers each of which represents the amount of liquid that follows that pipe. (Then this is just a cubical 1-form.)

In the picture above, the numbers represent "flows" through these "pipes", with the direction determined by the direction of the x, y-axes. In particular,

- "1" means "1 cubic foot per second from left to right".
- "2" means "2 cubic feet per second upward", etc.

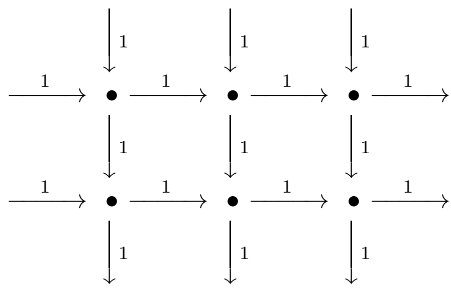
For the sake of *conservation of matter*, these numbers have to be normalized. Then 1/3 of the amount goes right and 2/3 goes up.

To confirm that this model makes sense, let's see what happens if a drop of dye be taken by such a flow. For computation, we follow the rule that

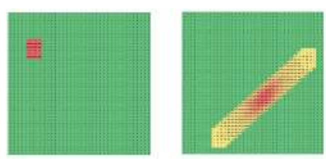
• the amount of dye in each cell is split and passed to the adjacent cells along the vectors of the vector field proportional to the values attached to the edges.

For instance, we can choose 1's on the horizontal edges and 0s on the vertical edges, i.e., dx. Then the flow will be purely horizontal. If we choose dy, it will be purely vertical.

Example 2.73. Let's choose the 1-form with 1's on both vertical and horizontal edges, i.e., $dx + dy$:



It is simple then: the amount of dye in each cell is split in half and passed to the two adjacent cells along the vectors: down and right. How well does it model the flow? Even though dispersal is inevitable, the predominantly diagonal direction of the spreading of the dye is evident in this spreadsheet simulation:



□

Exercise 2.74. Find the formula for the diagonal elements.

Exercise 2.75. Create such a spreadsheet and confirm the pattern. What happens if the flow takes horizontally twice as much material as vertically?

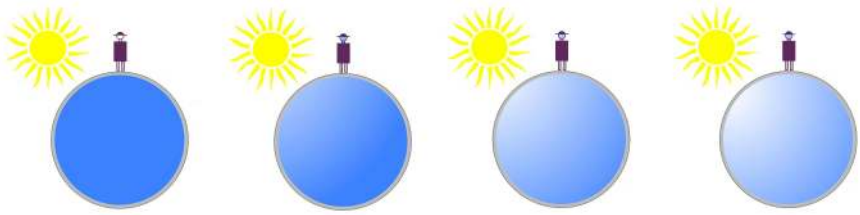
Exercise 2.76. Devise an advection simulation with a circular dispersal pattern.

Exercise 2.77. (a) Explain how an arbitrary directed graph gives rise to advection and present its ODE. (b) What if, in addition to carrying material around, the flow also deposits a proportion of it in every location?

3 PDEs

3.1 The PDE of diffusion

Just as advection, heat transfer exhibits a dispersal pattern. We can see, however, that without a flow to carry material around, the pattern doesn't have a particular direction:



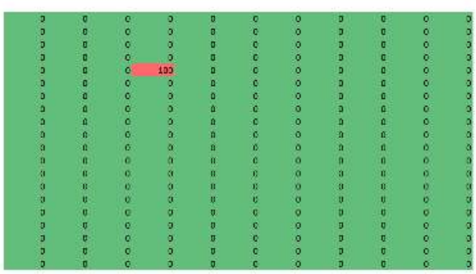
Instead of being carried around, the heat is *exchanged* – with adjacent locations. It's a circle.

The process is also known as *diffusion.*

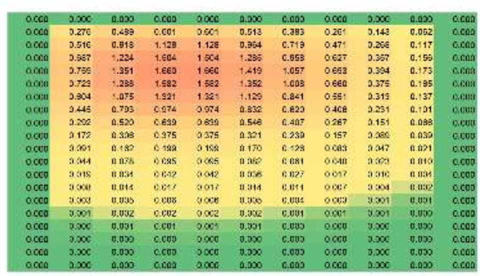
To model heat propagation, imagine a grid of square rooms and the temperature of each room changes by a proportion of the average of the temperature of the four adjacent rooms. Its spreadsheet simulation is given by the following short formula:

```
RC = RC - .25*h*( (RC-RC[-1]) + (RC-RC[1]) + (RC-R[-1]C) + (RC-R[1]C) )
```

Only a proportion h, dependent on the presumed length of the time interval, of the total amount is exchanged. The two images below are the initial state (a single initial hot spot) and the results (after 1,700 iterations) of such a simulation for $h = .01$ and the temperature at the border fixed at 0:

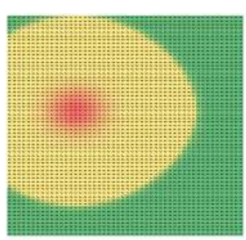


0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	100	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

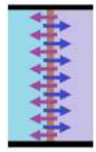


0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.276	0.499	0.601	0.601	0.513	0.383	0.261	0.143	0.052	0.000
0.000	0.516	0.918	1.128	1.128	0.964	0.719	0.471	0.266	0.117	0.000
0.000	0.667	1.224	1.604	1.604	1.296	0.958	0.627	0.367	0.159	0.000
0.000	0.765	1.351	1.660	1.660	1.419	1.057	0.663	0.394	0.173	0.000
0.000	0.723	1.288	1.582	1.582	1.352	1.008	0.660	0.375	0.195	0.000
0.000	0.604	1.075	1.321	1.321	1.129	0.841	0.551	0.319	0.137	0.000
0.000	0.445	0.793	0.974	0.974	0.832	0.620	0.408	0.231	0.101	0.000
0.000	0.292	0.520	0.639	0.639	0.546	0.407	0.267	0.151	0.066	0.000
0.000	0.172	0.306	0.375	0.375	0.321	0.239	0.157	0.089	0.039	0.000
0.000	0.091	0.162	0.199	0.199	0.170	0.128	0.083	0.047	0.021	0.000
0.000	0.044	0.078	0.065	0.065	0.082	0.081	0.040	0.023	0.010	0.000
0.000	0.019	0.034	0.042	0.042	0.036	0.027	0.017	0.010	0.004	0.000
0.000	0.008	0.014	0.017	0.017	0.014	0.011	0.007	0.004	0.002	0.000
0.000	0.003	0.005	0.008	0.008	0.005	0.004	0.003	0.001	0.001	0.000
0.000	0.001	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.000	0.000
0.000	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

On a larger scale, the simulation produces a realistic circular pattern:



A more careful look reveals that to model heat transfer, we need to separately record the exchange of heat with each of the adjacent rooms.



The process we are to study obeys the following law of physics.

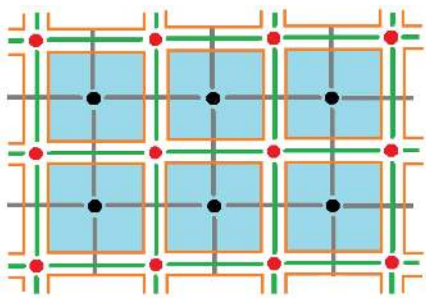
Newton's Law of Cooling: The rate of cooling of an object is proportional to the difference between its temperature and the ambient temperature.

This law is nothing but a version of the ODE of population growth and decay – with respect to the exterior derivative d_t over time. For each cell there are four adjacent cells and four temperature differences, d_x, to be taken into account. The result is a *partial differential equation* (PDE).

Below, we derive a PDE of forms for the diffusion process. The setup is as follows. A certain material is contained in a grid of rooms and each room (an n-cell) exchanges the material with its neighbors through its walls ($(n-1)$-cells). We will assume initially that the *space* is the standard cubical complex $\mathbb{R}^n$ and the *time* is the standard complex $\mathbb{R}$. For now, we ignore the geometry of time and space.

Dually, one can think of pipes (1-cells) connecting compartments or ponds (0-cells) in the centers of the rooms each of which exchanges the material with its neighbors. We will use both of the two approaches.

Below, the rooms are blue, the walls are green, and the pipes are gray:



We assume that the time increment is long enough
- for the material in each room to spread uniformly; or
- for the material to pass from one end of the pipe to the other.

What makes this different from ODEs is that the forms will have *two degrees* – with respect to location x and with respect to time t.

The *amount of material* $U = U(\alpha, t)$ is simply a number assigned to each room α which makes it an n-form. It also depends on time which makes it a 0-form. We call it an $(n, 0)$-form, an n-form with respect to location and a 0-form with respect to time.

Definition 3.1. In the product of two cubical complexes $K \times L$, an differential (k, m)-*form* is a real-valued linear operator defined on cells $a \times b$, where a is a k-cell in K and b is an m-cell in L.

The *outflow* gives the amount of flow across an $(n-1)$-face (from the room to its neighbor) per unit of time. It is also a number assigned to each "pipe" p that goes through each wall from one cell to the next. So, $F = F(p, t)$ is a dual $(n-1, 0)$-form.

Notation:
- d_t is the exterior derivative with respect to time (simply the difference in $\mathbb{R}$); and
- d_x is the exterior derivative with respect to location.

Below, we routinely suppress t for the second variable.

The "conservation of matter" in cell α gives us the following. The change of the amount of the material in room α over the increment of time is equal to

$$d_t U(\alpha) = -\Big(\text{sum of the outflow } F \text{ across the walls of } \alpha \Big).$$

Naturally, the walls form the boundary $\partial\alpha$ of α. Therefore, we have

$$d_t U(\alpha) = -F^\star(\partial\alpha)$$

or, by the Stokes Theorem,

$$d_t U(\alpha) = -(d_x F^\star)(\alpha).$$

Now, we need to express F in terms of U. The flow $F(a^\star) = F^\star(a)$ through wall a of room α is proportional to the difference of the amounts of material (or, more precisely, the density) in α and the other room adjacent to a. So,

$$F^\star(a) = -k(a) d_x (\star U)(a^\star).$$

Here, $k(a) \geq 0$ represents the *permeability* of the wall a at a given time. Since a is an $(n-1)$-cell, k is an $(n-1)$-form with respect to space. It is also a 0-form with respect to time.

The result of the substitution is a PDE of second degree called the *diffusion equation* of forms:

$$\boxed{d_t U = d_x \star k d_x \star U}$$

The right-hand side is seen in the Hodge duality diagram below. We start with U in the right upper corner and make the full circle:

$$\begin{array}{ccccc}
C^{n-1}(K) & \xrightarrow{d_x} & C^n(K) & \ni U \\
\downarrow \star \uparrow & \neq & \downarrow \star \uparrow & \\
C^1(K^\star) & \xleftarrow{d_x} & C^0(K^\star) &
\end{array}$$

The multiplication by k is implied.

In general, k cannot be factored out. Unless of course k happens to be constant with respect to space; then the right-hand side is kU''.

This was an outline. In the following, we develop both the mathematics and the simulations for progressively more complex settings.

3.2 Simulating diffusion with a spreadsheet

For the simplest spreadsheet simulation, let's "trivialize" the above analysis. We start with dimension 1.

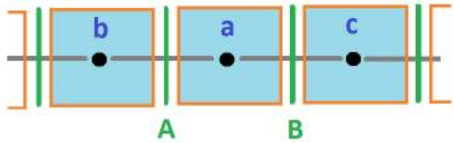
The material is contained in a *row of rooms* and each room exchanges the material with its *two neighbors* through its walls. The amount of material is a $(1,0)$-form.

The outflow, the amount of flow across a wall (from the room to its neighbor) per unit of time, $F = F(p,t)$ is a $(1,1)$-form over the dual complex.

Suppose

- $a = AB$ is one of the rooms;

- b, c are the two adjacent rooms, left and right;
- A, B are the walls of a, left and right;
- $p = A^\star, q = B^\star$ are the two pipes from a, left and right.



As the material is conserved, we know that $d_t U(a)$ is the negative of the sum of the outflow $F(\cdot)$ across the two walls of a. The flow is positive at A if it is from left to right and the opposite for B; then:

$$d_t U(a) = -\big(F^\star(A) - F^\star(B)\big) = F^\star(A) - F^\star(B),$$

which is the exterior derivative of a 0-form. The flow through a wall is proportional to the difference of amounts of material in the two adjacent rooms:

$$\begin{array}{ll} F^\star(A) & = -k(A)\big(U(a) - U(b)\big), \\ F^\star(B) & = -k(B)\big(U(c) - U(a)\big), \end{array}$$

where $k \geq 0$ is the permeability, a $(0,0)$-form. Then:

$$d_t U = \Big(-k(A)\big(U(a) - U(b)\big)\Big) - \Big(-k(B)\big(U(c) - U(a)\big)\Big).$$

The right-hand side becomes the increment in the recursive formula for the simulation:

$$U(a, t+1) := U(a,t) + \Big[-k(a)\big(U(a) - U(a-1)\big) + k(a+1)\big(U(a+1) - U(a)\big)\Big].$$

The initial state is shown below:

primal complex	K	0-cells, primal	A_i	A_1	A_2	A_3	A_4	A_5
		1-cells, primal	a_i	a_1	a_2	a_3	a_4	a_5
Dual complex:	K*	0-cells, dual	B_i = dual of a_i	B_1	B_2	B_3	B_4	B_5
		1-cells, dual	b_i = dual of A_i		b_2	b_3	b_4	b_5
PROBLEM:								
Given	f	1-cochain, primal	numbers assigned to primal 1-cells = values f(a_i) of function f	3.00	2.00	0.00	6.00	1.00
	k	0-cochain, primal	numbers assigned to primal 0-cells = values k(A_i) of function k	0.10	0.00	0.20	0.50	0.90
find	f-kΔf	1-cochain, primal		?	?	?	?	?
SOLUTION:								
given	f	1-cochain, primal	numbers f_i=f(a_i) assigned to primal 1-cells	3.00	2.00	0.00	6.00	1.00
compute	g=f*	0-cochain, dual	numbers g_i=g(B_i) assigned to dual 0-cells	3.00	2.00	0.00	6.00	1.00
compute	h=dg	1-cochain, dual	numbers h_i=h(b_i) assigned to dual 1-cells = differences of adjacent values of g	0.00	2.00	-6.00	5.00	-3.00
compute	l=h*	0-cochain, primal	numbers l_i=l(A_i) assigned to primal 0-cells	0.00	2.00	-6.00	5.00	-3.00
compute	kl	0-cochain, primal		0.00	0.00	-1.20	2.50	-2.70
compute	m=d(kl	1-cochain, primal	numbers m_i=m(a_i) assigned to primal 1-cells = differences of adjacent values of	0.00	-1.20	-3.70	5.20	-2.70
compute	f-m	1-cochain, primal		3.00	0.80	3.70	0.80	3.70

Note: When the domain isn't the whole space, the pipes at the border of the region have to be removed. In the spreadsheet, we use boundary conditions to substitute for the missing data.

The result after 1,500 iterations is shown next:

1	N=	1500										
primal complex:		0-cells, primal	A_i	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
		1-cells, primal	a_i	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
Dual complex:		0-cells, dual	B_i = dual of a_i	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9
		1-cells, dual	b_i = dual of A_i		b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
PROBLEM:												
Given	f	1-cochain, primal	numbers assigned to primal 1-cells = values f(a_i) of function f	3.00	2.00	0.00	6.00	2.00	4.00	0.00	3.00	5.00
	k	0-cochain, primal	numbers assigned to primal 0-cells = values k(A_i) of function k		0.02	0.02	0.05	0.02	0.01	0.01	0.02	0.03
find	f-kΔf	1-cochain, primal		?	?	?	?	?	?	?	?	?
SOLUTION:												
given	f	1-cochain, primal	numbers f_i=f(a_i) assigned to primal 1-cells	2.75	2.75	2.75	2.75	2.60	2.60	2.60	2.60	2.60
compute	g=f*	0-cochain, dual	numbers g_i=g(B_i) assigned to dual 0-cells	2.75	2.75	2.75	2.75	2.60	2.60	2.60	2.60	2.60
compute	h=dg	1-cochain, dual	numbers h_i=h(b_i) assigned to dual 1-cells = differences of adjacent values of g	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00
compute	l=h*	0-cochain, primal	numbers l_i=l(A_i) assigned to primal 0-cells	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00
compute	kl	0-cochain, primal		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
compute	m=d(kl	1-cochain, primal	numbers m_i=m(a_i) assigned to primal 1-cells = differences of adjacent values of g	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
compute	f+m	1-cochain, primal		2.750	2.750	2.750	2.750	2.593	2.600	2.601	2.600	2.600
Check	sum m		total change of f	0.00000000								
	sum |m|		total net change of f	0.00000000								

One can clearly see how the density becomes uniform eventually – but not across an impenetrable wall ($k(A_5) = 0$).

Note: See the files online.

Exercise 3.2. For an infinite sequence of rooms, what is the limit of U as $t \to \infty$?

Exercise 3.3. Create a simulation for a circular sequence of rooms. What is the limit state?

Exercise 3.4. Generalize the formula to the case when the permeability of walls depends on the direction.

Next, dimension 2.

The amount of material $U = U(\tau, t)$ is a 2-form with respect to location and a 0-form with respect to time.

The outflow gives the amount of flow across a 1-face (from the room to its neighbor) per unit of time. It is simply a number assigned to each "pipe" p that goes through each wall from one cell to the next. So, $F = F(p, t)$ is a $(1, 1)$-form, but over the dual grid.

The "conservation of matter" for cell τ implies that the change of the amount of the material over the increment of time $d_t U(\tau)$ is, as before, the negative sum of the outflow $F(\cdot)$ across the four walls of τ: a, b, c, d. Let's rewrite the latter:

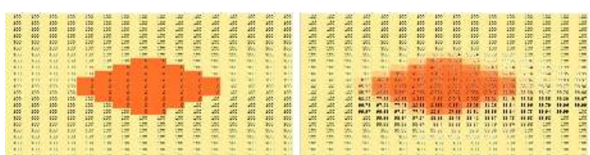
$$= -\Big(F^\star(a) + F^\star(b) + F^\star(c) + F^\star(d)\Big) = -F^\star(a + b + c + d).$$

Now, we need to express F in terms of U. The flow $F(a^\star) = F^\star(a)$ through face a of cell τ is proportional to the difference of the amounts of material in τ and the other 2-cell adjacent to a. So,

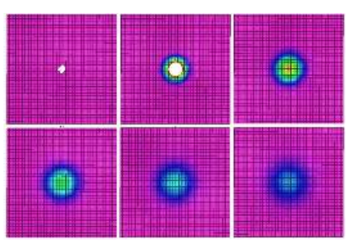
$$F^\star(a) = -k d_x(\star U)(a^\star).$$

Exercise 3.5. Demonstrate how the above analysis leads to the "naive" spreadsheet simulation presented in the beginning of the section.

Note: One needs to use a buffer (an extra sheet); otherwise Excel's sequential – cell-by-cell in each row and then row-by-row – manner of evaluation will cause a skewed pattern:



A simulation of heat transfer from a single cell is shown below:



Note: See the files online.

Exercise 3.6. (a) What kind of medium would create an *oval* diffusion pattern? Demonstrate. (b) What about an oval not aligned with the axes?

Exercise 3.7. Modify the formula to create a simulation for diffusion on the torus.

3.3 Diffusion on metric complexes

We have used interchangeably "amount of material" and "density of material". The reason is that we ignore the possibility of cells of different sizes (and shapes) by assuming the simplest geometry.

We have modeled heat transfer with this formula:

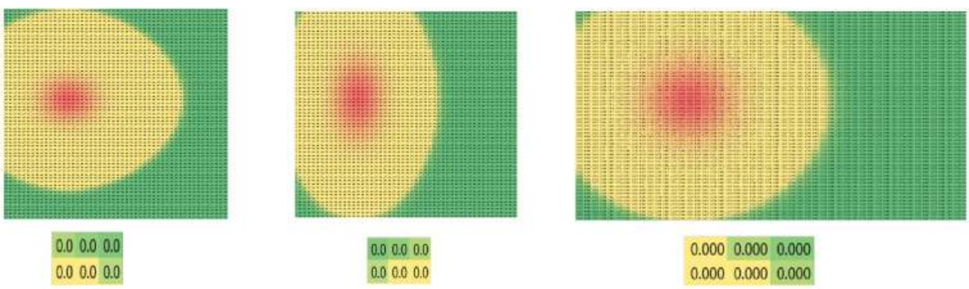
```
RC = RC - k*( (RC-RC[-1]) + (RC-RC[1]) + (RC-R[-1]C) + (RC-R[1]C) )
```

What if the horizontal walls are longer than the vertical ones? Then more heat should be exchanged across the vertical walls than the horizontal ones.

We can change the coefficients of "vertical" differences in the above formula to reflect that:

```
RC = RC - k*( .5*(RC-RC[-1]) + .5*(RC-RC[1]) + 2*(RC-R[-1]C) + 2*(RC-R[1]C) )
```

We can see how the material travel farther – as measured by the number of cells – in the vertical direction (second image) than normal (first image):



Yet, if we stretch – justifiably – the cells in the spreadsheet horizontally (third image), the pattern becomes circular again!

In summary, the amount of heat exchanged between two rooms is proportional to:
- the temperature difference,
- the permeability of the wall (dually: the conductance of the pipe),
- the area of the wall that separates them (dually: the cross section of the pipe), and
- inversely, to the distance between the centers of mass of the rooms (dually: the length of the pipe).

Let's split the data into three categories:
- the adjacency of rooms (and the pipes) is the *topology*,
- the areas of the walls (and the lengths of the pipes) is the *geometry*, and
- the properties of the material of the walls (and the pipes) is the *physics*.

They are given by, respectively:
- the cell complex K,
- the Hodge star operator $\star^m : C^m(K) \to C^{n-m}(K^\star)$, and
- the $(n-1)$-form k over K.

Suppose the *geometry of space* is supplied by means of the m-dimensional volume $|b|$ of each m-cell b – in both primal and dual complexes K and $K^\star$. Now, the mth Hodge star is a diagonal matrix whose entries are the ratios of dual and primal volumes (up to a sign) in each dimension $m = 0, 1, ..., n$:

$$\star^m_{ii} = \pm\frac{|a_i^\star|}{|a_i|}.$$

Similarly the *geometry of time* is supplied by means of the length $|t|$ of each time interval (1-cell) t – in both primal and dual complexes $\mathbb{R}$ and $\mathbb{R}^\star$. The Hodge star is a diagonal matrix whose entries are the reciprocals of the lengths of these intervals:

$$\star^1_{ii} = \pm\frac{1}{|t_i|}.$$

Recall that the right-hand side of our equation is the familiar second derivative (with respect to space) of an n-form with an extra factor k:

$$(kU_x)_x := (kU')' = d_x \star k d_x \star U.$$

As before, this expression is seen in the Hodge duality diagram, with the factors that come from the star operators also shown:

$$\begin{array}{ccccc}
C^{n-1}(K) & \xrightarrow{\quad d_x \quad} & C^n(K) & \ni U \\
\uparrow \times \frac{|a^{n-1}|k(a^{n-1})}{|b^1|} & \neq & \downarrow \times \frac{|b^0|}{|a^n|} & \\
C^1(K^\star) & \xleftarrow{\quad d_x \quad} & C^0(K^\star), &
\end{array}$$

where a^m is a primal m-cell and b^m is a dual m-cell.

If we are to use the derivative, instead of the exterior derivative, with respect to time, we need to consider two issues. First, let's recall that when studying ODEs we used the function $q : C_1(\mathbb{R}) \to C_0(\mathbb{R})$ given by

$$q\Big([i, i+1]\Big) = i,$$

to make the degrees of the forms, with respect to time, match. Second, in addition to the above list, the amount of material exchanged between two rooms is also proportional to the length of the current time interval $|t|$. Then our PDE takes the form:

$$d_t U q^{-1} = (kU_x)_x |t|,$$

with both sides $(n, 0)$-forms. Consider the first derivative with respect to time:

$$U_t := U' = \star d_t U = \tfrac{1}{|t|} d_t U.$$

Definition 3.8. Given a metric complex K of dimension n, the *diffusion equation* is:

$$U_t q^{-1} = (kU_x)_x,$$

where U is an $(n, 0)$-form over $K \times \mathbb{R}$.

The abbreviated version is below.

$$\boxed{U_t = (kU_x)_x}$$

Definition 3.9. An *initial value problem (IVP)* for diffusion is a combination of the diffusion equation and an *initial condition (IC)*:

$$U_t = (kU_x)_x, \; U(\cdot, A_0) = u_0 \in C^n(K).$$

Then a $(n, 0)$-form U on $K \times \mathbb{R} \cap \{A \geq A_0\}$ that satisfies both conditions is called a (forward) *solution* of the IVP.

Because the exterior derivative with respect to time is simply the difference of values, a solution is easy to construct iteratively.

Theorem 3.10 (Existence and Uniqueness). The solution to the IVP above exists and is given by

$$U(\cdot, A_0) := u_0, \quad U(\cdot, A+1) := U(\cdot, A) + (kU_x)_x(\cdot, A)|[A, A+1]|, \ \forall A \geq A_0,$$

provided $\frac{1}{|a|} \in R$ for every cell a in $K \sqcup K^\star$.

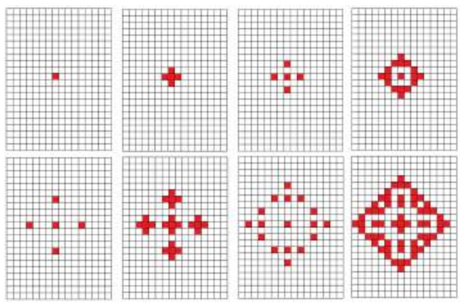
Exercise 3.11. Provide a weaker sufficient condition for existence.

Note: Boundary values problems lie outside the scope of this book.

Example 3.12. The choice of

$$K = \mathbb{R}^2, \quad R = \mathbf{Z}_2,$$

produces this simple "cellular automaton":



□

Exercise 3.13. Derive the dual diffusion equation: "Suppose the material is located in the nodes of a graph in $\mathbf{R}^n$ and the amount of material is represented by a 0-form V..."

Note: One can allow the physics to be absorbed into the geometry. As the transfer across a primal $(n-1)$-cell a is proportional to its permeability $k(a)$, we can simply replace the volume $|a|$ of a with $k(a)|a|$, provided $k \neq 0$. Then the right-hand side of our equation becomes the Laplacian (with respect to space) $U_{xx} = \Delta U$.

We test the performance of this equation next.

3.4 How diffusion patterns depend on the sizes of cells

Let's consider diffusion over a 1-dimensional metric cubical subcomplex of $\mathbb{R}$.

The diagonal elements of the Hodge star operator for dimension $n = 1$ are:

$$\star^1_{ii} = \frac{|\text{point}|}{|\text{edge}|} = \frac{1}{\text{length}} = \frac{1}{|a_i|}.$$

Then $\star^1$ and $\star^0$ give us our Hodge duality diagram:

$$\begin{array}{ccccc}
C^0(K) & \xrightarrow{\ d\ } & C^1(K) & & \ni U \\
\uparrow \times \frac{|b|}{1} & \neq & \downarrow \times \frac{1}{|a|} & & \\
C^1(K^\star) & \xleftarrow{\ d\ } & C^0(K^\star), & &
\end{array}$$

where a,b are primal and dual 1-cells respectively. The right-hand side of our equation will have extra coefficients: the lengths of the cells appear twice as we make a full circle.

First, it's the length of the cell itself, $\frac{1}{|a|}$. Then the equation will contain:

$$\frac{U(a)}{|a|}.$$

That's the *density* of the material inside a.

Second, it's the lengths of the 1-cells dual to the endpoints of $a = AB$:

$$\frac{1}{|A^\star|}, \frac{1}{|B^\star|}.$$

The denominators are the lengths of the pipes.

Conclusion: The amount of material exchanged by a primal 1-cell a with its neighbor a' is
- directly proportional to the difference of density in a and a', and
- inversely proportional to the length of the pipe that leaves a for a'.

To confirm these ideas, we run a spreadsheet simulation below:

1	N=	100						
primal complex:		0-cells, primal	A_i	A_1	A_2	A_3	A_4	A_5
		1-cells, primal	a_i	a_1	a_2	a_3	a_4	a_5
Dual complex:		0-cells, dual	B_i = dual of a_i	B_1	B_2	B_3	B_4	B_5
		1-cells, dual	b_i = dual of A_i		b_2	b_3	b_4	b_5
Volumes		number	volume of primal 0-cell, |A_i|	1	1	1	1	1
		number	volume of primal 1-cell, |a_i|	10	10	10	10	1
		number	volume of dual 0-cell, |B_i|	1	1	1	1	1
		number	volume of dual 1-cell, |b_i|		10	10	10	1
Hodge star St	K->K*	number	coefficient of primal, 0->1, |b_i| / |A_i|		10.00	10.00	10.00	1.00
	K->K*	number	coefficient of primal, 1->0, |B_i| / |a_i|	0.10	0.10	0.10	0.10	1.00
	K*->K	number	coefficient of dual 0->1, |a_i| / |B_i|	10.00	10.00	10.00	10.00	1.00
	K*->K	number	coefficient of dual 1->0, |A_i| / |b_i|		1.00	1.00	1.00	1.00
PROBLEM:								
Given	f	1-cochain, primal	numbers assigned to primal 1-cells = values f(a_i) of function f	0.00	0.00	0.00	0.00	50.00
	k	0-cochain, primal	numbers assigned to primal 0-cells = values k(A_i) of function k		0.05	0.05	0.05	0.05
find	f+kΔf	1-cochain, primal	after N iterations	?	?	?	?	?
SOLUTION:								
given	f	1-cochain, primal	numbers f_i=f(a_i) assigned to primal 1-cells	0.22	1.47	7.55	23.61	3.05
compute	g=f*	0-cochain, dual	numbers g_i=g(B_i) assigned to dual 0-cells	0.02	0.15	0.75	2.36	3.05
compute	h=dg	1-cochain, dual	numbers h_i=h(b_i) assigned to dual 1-cells = differences of adjacent values of g	0.00	-0.12	-0.61	-1.61	-0.69
compute	l=h*	0-cochain, primal	numbers l_i=l(A_i) assigned to primal 0-cells	0.00	-0.12	-0.61	-1.61	-0.69
compute	kl	0-cochain, primal		0.00	-0.01	-0.03	-0.08	-0.03
compute	m=d(kl	1-cochain, primal	numbers m_i=m(a_i) assigned to primal 1-cells = differences of adjacent values of	0.01	0.02	0.05	-0.05	-0.01
compute	f+m	1-cochain, primal		0.230	1.491	7.596	23.563	3.038

With a single initial spike in the middle, we see that the amounts of material in the smaller cells on the right:

$$|a_5| = |a_6| = |a_7| = |a_8| = |a_9| = 1,$$

quickly become uniform, while the larger ones on the left:

$$|a_1| = |a_2| = |a_3| = |a_4| = 10,$$

develop slower.

Note: See the files online.

Exercise 3.14. Find the speed of propagation of the material in a uniform grid as a function of the length of the cell.

Exercise 3.15. Suppose we have two rods made of two different kinds of metal soldered together side by side. The cells will expand at two different rates when heated. Model the change of its geometry and illustrate with a spreadsheet. Hint: Assign a thickness to both.

For dimension $n = 2$, we know the diagonal entries in the case of a rectangular grid $\mathbb{R}^2$:

$$\star^2_{ii} = \frac{1}{\text{area}}, \quad \star^1_{ii} = \frac{\text{length}}{\text{length}}.$$

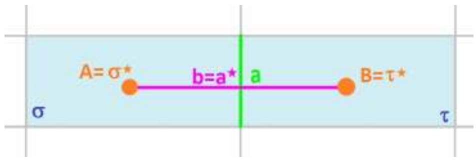
Then $\star^2$ and $\star^1$ give us our Hodge duality diagram:

$$\begin{array}{ccc} C^1(K) & \xrightarrow{\ d\ } & C^2(K) \quad \ni U \\ \uparrow \times \frac{|b|}{|a|} & \neq & \downarrow \times \frac{1}{|\sigma|} \\ C^1(K^\star) & \xleftarrow{\ d\ } & C^0(K^\star), \end{array}$$

where a, b are primal and dual 1-cells respectively and σ is a 2-cell.

Just as above, we notice that $U(\sigma)/|\sigma|$ is the density inside σ. The second coefficient $|a|/|a^\star|$ is new. We see here the 1-cell that represents the wall and its dual that represents the pipe:

$$\frac{\text{length of the wall}}{\text{length of the pipe}}.$$



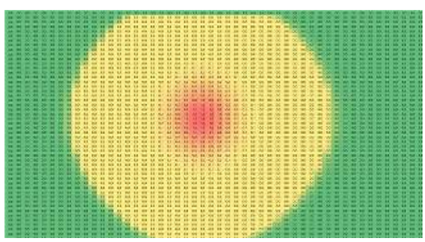
Conclusion: The amount of material exchanged by a primal 2-cell σ with its neighbor τ is
- directly proportional to the difference of density in σ and that of τ,
- inversely proportional to the length of the pipe from σ to τ, and
- directly proportional to the length of the wall (thickness of this pipe).

If we use the 2×1 grid, we have the following lengths of 1-cells:
- horizontal primal: $|a| = 2$, dual: $|a^\star| = 1$;
- vertical primal: $|a| = 1$, dual: $|a^\star| = 2$.

Then the Excel formula is the same as the one discussed above:

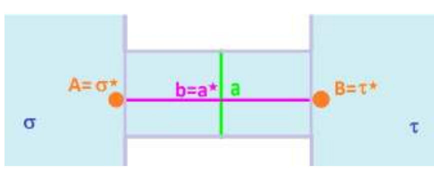
```
RC = RC - .0025*( .5*(RC-RC[-1]) + .5*(RC-RC[1]) + 2*(RC-R[-1]C) + 2*(RC-R[1]C) )
```

This formula, with a single source, produces a circular pattern on a spreadsheet with appropriately sized cells, as expected:



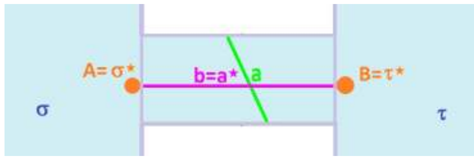
Exercise 3.16. Show that with the horizontal walls impenetrable, the transfer pattern will be identical to the 1-dimensional pattern.

To summarize, the material flows from room σ to room τ through what appears to be a pipe of length $|a^\star|$ and width $|a|$:

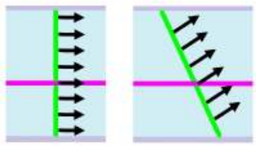


3.5 How diffusion patterns depend on the angles between cells

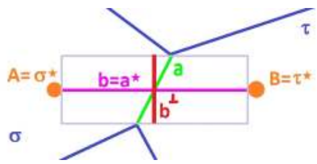
What if the wall between two rooms is slanted? Does it affect the amount of liquid that crosses to the other room?



If two walls are made of the same material (and of the same thickness), the slanted one will allow less liquid to flow through the pipe. In fact, what matters is the *normal component of the flow* across the wall.

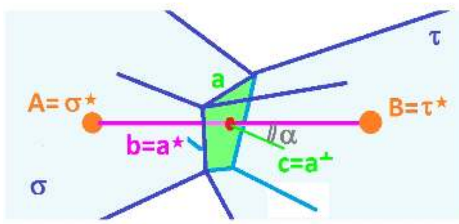


We arrive to the same idea if, dually, we expand our pipe, $a^\star$, to the largest possible width, $|a|$:



Then a is a (possibly non-orthogonal) cross-section of pipe $a^\star$.

Now, to make this more precise, consider the general case of an n-dimensional room, σ, with an $(n-1)$-dimensional wall, a, and still a 1-dimensional pipe, $a^\star$. It is illustrated below for $n = 3$:



As before, we are looking at the angle α between the normal $c = a^\perp$ of the wall a and the pipe $a^\star$. Then

• the normal component of the flow is acquired by multiplying by the cosine of the angle between a and $a^\star$.

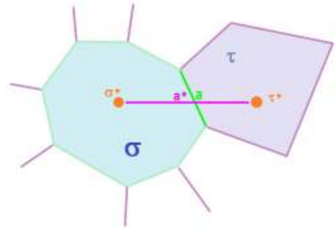
Conclusion: The amount of material exchanged by a primal n-cell σ with its neighbor τ is

- directly proportional to the difference of density in σ and that of τ; i.e., $U(\sigma)/|\sigma| - U(\tau)/|\tau|$,
- inversely proportional to the length $|a^\star|$ of the pipe $a^\star$ that leaves σ for τ,
- directly proportional to the area $|a|$ of the wall a, and
- directly proportional to the cosine of the angle between this pipe and this wall, $\cos\widehat{aa^\star}$.

Next, the iterative formula.

The boundary – as a chain – of a cell is the sum of its boundary cells taken with appropriate orientations:

$$\partial\sigma = \sum_{a \in \partial\sigma} \pm a.$$



Therefore, the net change of material in cell σ is the sum of the amounts exchanged through its boundary cells:

$$U(\sigma,t+1)-U(\sigma,t)=\sum_{\partial\sigma}\tfrac{|a|}{|a^\star|}\cos\widehat{aa^\star}\left[\tfrac{U(\sigma)}{|\sigma|}-\tfrac{U(\tau_a)}{|\tau_a|}\right],$$

where

$$\tau_a := \sigma + (\partial a^\star)^\star$$

is the n-cell that shares wall a with σ.

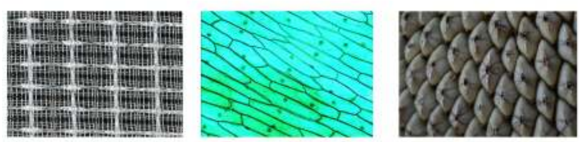
It is easy to confirm that the right-hand side is nothing but the same expression $\star d \star dU$ as before. What has changed is the coefficients of the matrix of the Hodge star operator of the new geometry of the complex:

$$\star^k_{ii}=\tfrac{|a_i|}{|a_i^\star|}\cos\widehat{a_i a_i^\star},$$

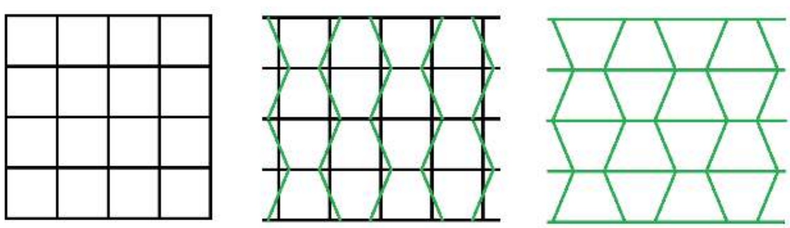
where a_i is the ith k-cell. Recall that the feature of a non-rectangular geometry is that the angles $\cos\widehat{aa^\star}\neq 1$.

Exercise 3.17. Create a spreadsheet for the above formula and confirm the results presented below.

With this formula, we are able to study diffusion or heat transfer through an anisotropic or even irregular pattern, such as fabric, plant cells, or sunflower seeds:



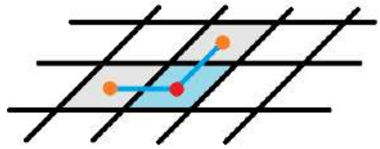
Example 3.18 (trapezoid grid). Consider this trapezoid grid:



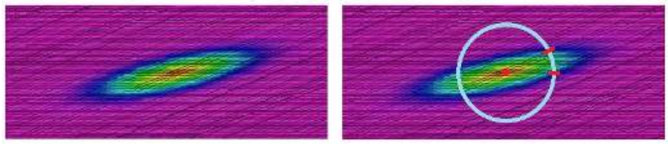
It is made from a square grid by turning the vertical edges by the same angle alternating left and right. According to our analysis, the horizontal flow will be slower than before, but what about vertical? □

Exercise 3.19. (a) Compare the speed of the horizontal flow and the vertical to those of the original grid. (b) Verify the conclusion with a simulation.

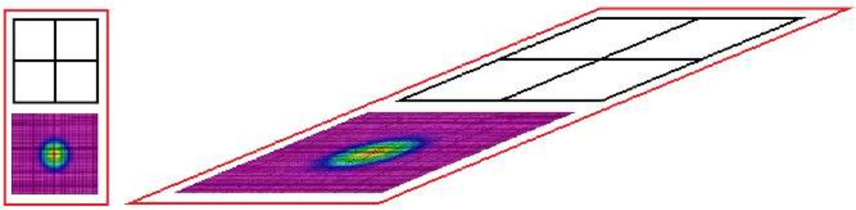
Example 3.20 (rhomboid grid). Consider now the rhomboid grid:



The pattern isn't circular anymore! And it not aligned with the directions of the walls:

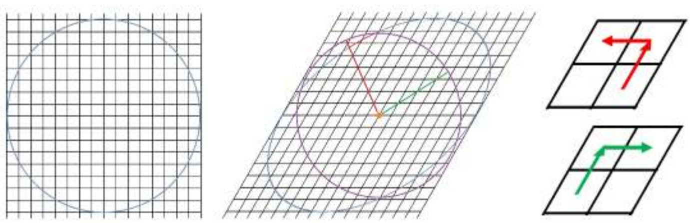


This is how this pattern is acquired. Even with slanted walls, the exchange with each of the four adjacent cells remains *identical*. That's why the result of the spreadsheet simulation is still circular (left):



The pattern is shown circular, however, only because the spreadsheet displays the shapes of the cells as squares. To see the true pattern, we have to skew the image (right) in such a way that the squares become rhombi.

The outcome may seem unexpected because the physics appears to be symmetric. It is true that the pattern of the flow is identical for the two directions along the grid, and therefore, the speed of propagation in either of the two directions is the same, just as before. Still, what about the other directions? Below, the red and green segments are equal in length, but to follow the red one, we would have to cross more walls than for the green:



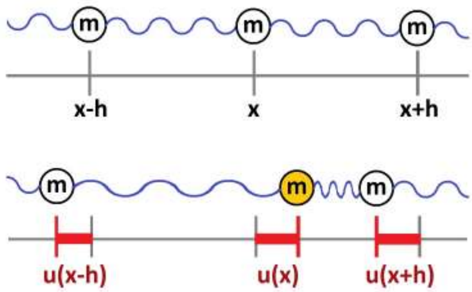
We conclude that the skewed pattern is caused by the "anisotropic" nature of the medium. □

Exercise 3.21. (a) Verify these conclusions with a simulation. (b) What are the directions of the fastest and the slowest propagation? Explain.

Exercise 3.22. To study a triangular grid, one needs a non-cubical cell complex. Use a spreadsheet to model diffusion on a triangular grid produced by diagonally cutting the squares.

3.6 The PDE of wave propagation

Previously, we studied the motion of an object attached to a wall by a (mass-less) spring. Imagine this time a *string of objects* connected by springs:



Just as above, we will provide the mathematics to describe the following three parts of the setup:

• the *topology* of the cell complex L of the objects and springs,

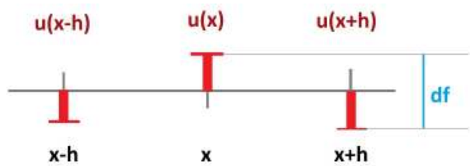
• the *geometry* given to that complex such as the lengths of the springs, and

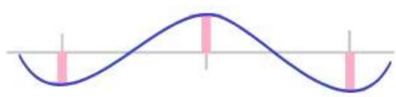
• the *physics* represented by the parameters of the system such as those of the objects and springs.

Let $u(t,x)$ be the function that measures the displacement from the equilibrium of the object associated with position x at time t (we will suppress t). It is an algebraic quantity:

$$u = u(t,x) \in R.$$

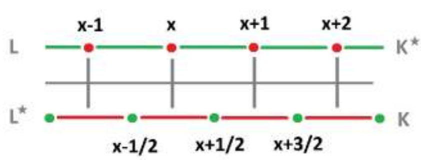
As such, it can represent quantities of any nature that may have nothing to do with a system of objects and springs; it could be an *oscillating string*:





Here, the particles of the string are vertically displaced while waves propagate horizontally (or we can see the pressure or stress vary in a solid medium producing sound).

First, we consider the *spatial variable*, $x \in \mathbf{Z}$. We think of the array – at rest – as the standard 1-dimensional cubical complex $L = \mathbb{R}_x$. The complex may be given a geometry: each object has a (possibly variable) distance $h = \Delta x$ to its neighbor and the distance between the centers of the springs has length $\Delta x^\star$. We think of u as a form of degree 0 – with respect to x.



According to Hooke's law, the force exerted by the spring is

$$F_{Hooke} = -k\,df,$$

where $df \in R$ is the displacement of the end of the spring from its equilibrium state and the constant, *stiffness*, $k \in R$ reflects the physical properties of the spring. If this is the spring that connects locations x and $x+1$, its displacement is the difference of the displacements of the two objects. In other words, we have:

$$df = u(x+1) - u(x).$$

Therefore, the force of this spring is

$$H_{x,x+1} = k\Big[u(x+1) - u(x)\Big].$$

Since k can be location-dependent, it is a 1-form over L.

Now, let H be the force that acted on the object located at x. There are two Hooke's forces acting on this object from the two adjacent springs: $H_{x,x-1}$ and $H_{x,x+1}$. Therefore, we have:

$$\begin{array}{llll}
H & = H_{x,x-1} & +H_{x,x+1} \\
 & = k\Big[u(x-1) - u(x)\Big] & +k\Big[u(x+1) - u(x)\Big] \\
 & = -(kd_x u)[x-1,x] & +(kd_x u)[x,x+1].
\end{array}$$

Next, we investigate what this means in terms of the Hodge duality. These are the duality relations of the cells involved:

$$\begin{array}{ll}
[x-1,x]^\star & = \{x-1/2\}, \\
[x,x+1]^\star & = \{x+1/2\}, \\
\{x\}^\star & = [x-1/2,x+1/2].
\end{array}$$

Then the computation is straight-forward:

$$\begin{array}{ll}
H & = (kd_x u)\Big([x+1,x] - [x,x-1]\Big) \\
 & = (kd_x u)\Big(\{x+1/2\}^\star - \{x-1/2\}^\star\Big) \\
 & = (\star kd_x u)\Big(\{x+1/2\} - \{x-1/2\}\Big) \\
 & = d_x(\star kd_x u)\Big([x-1/2,x+1/2]\Big) \\
 & = d_x(\star kd_x u)(x^\star) \\
 & = \star d_x \star kd_x u(x) \\
 & = \star d_x k^\star \star d_x u(x) \\
 & = (k^\star u_x)_x(x).
\end{array}$$

Note that for a constant k, we are dealing with the second derivative of the 0-form u with respect to space:

$$u'' = \star d_x \star d_x u = \Delta u.$$

Compare it to the second derivative of a 1-form U with respect to space:

$$U'' = d_x \star d_x \star U = \Delta U,$$

which we used to model 1-dimensional diffusion. The difference is seen in the two different starting points in the same Hodge duality diagram:

$$\begin{array}{ccccc}
u \in & C^0(\mathbb{R}_x) & \xrightarrow{\ d_x\ } & C^1(\mathbb{R}_x) & \ni U \\
 & \uparrow \star & \neq & \downarrow \star & \\
 & C^1(\mathbb{R}_x^\star) & \xleftarrow{\ d_x\ } & C^0(\mathbb{R}_x^\star) &
\end{array}$$

Then,

$$d_x u'' = (d_x u)''.$$

Second, we consider the *temporal variable*, $t \in \mathbf{Z}$. We think of time as the standard 1-dimensional cubical complex $\mathbb{R}_t$. The complex is also given a geometry. It is natural to assume that the geometry has no curvature, but each increment of time may have a different duration (and, possibly, $\Delta t \neq \Delta t^\star$). We think of u as a form of degree 0 with respect to t.

Now suppose that each object has mass m. Then, by the Second Newton's Law, the total force is

$$F = m \cdot a,$$

where a is the acceleration. It is the second derivative with respect to time, i.e., this 0-form:

$$a = u_{tt} := \star d_t \star d_t u.$$

The mass m is a 0-form too and so is F. Note that the stiffness k is also a 0-form with respect to time.

Now, with these two forces being equal, we have derived the *wave equation* of differential forms:

$$\boxed{m u_{tt} = (k^\star u_x)_x.}$$

If k and m are constant forms (and R is a field), the wave equation takes a familiar shape:

$$u_{tt} = \tfrac{k}{m} u_{xx}.$$

Exercise 3.23. Derive the dual (with respect to space) wave PDE.

3.7 Solutions of the wave equation

Now we will derive the recurrence relations.

First, we assume that the geometry of the time is "flat": $\Delta t = \Delta t^\star$. Then the left-hand side of our equation is

$$m d_{tt} u = m \frac{u(x,t+1) - 2u(x,t) + u(x,t-1)}{(\Delta t)^2}.$$

For the right-hand side, we can use the original expression:

$$\star d_x \star k d_x u = k\Big[u(x-1) - u(x)\Big] + k\Big[u(x+1) - u(x)\Big].$$

Second, we assume that k and m are constant. Then just solve for $u(x,t+1)$:

$$u(x,t+1) \quad = 2u(x,t) - u(x,t-1) + \alpha\Big(u(x+1,t) - 2u(x,t) + u(x-1,t)\Big),$$

where

$$\alpha := (\Delta t)^2 \frac{k}{m}.$$

To visualize the formula, we arrange the terms in a table:

	$x-1$	x	$x+1$
$t+1$		$u(x,t+1)$	
t	$= \alpha u(x-1,t)$	$+2(1-\alpha)u(t,x)$	$+\alpha u(x+1,t)$
$t-1$		$-u(x,t-1)$	

Even though the right-hand side is the same, the table is different from that of the (dual) diffusion equation. The presence of the second derivative with respect to time makes it necessary to look two steps back, not just one. That's why we have two initial conditions.

We suppose, for simplicity, that $\alpha = 1$.

Example 3.24. Choosing the simplified settings allows us to easily solve the following initial value problem:

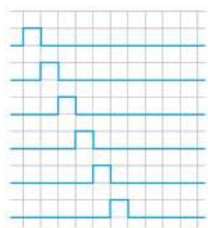
$$u_{tt} = u_{xx};$$
$$u(x,t) = \begin{cases} 1 & \text{if } t=0, x=1; \\ 0 & \text{if } t=0, x\neq 1; \\ 1 & \text{if } t=1, x=2; \\ 0 & \text{if } t=1, x\neq 2. \end{cases}$$

Initially, the wave has a single bump and then the bump moves one step from left to right. The negative values of x are ignored.

Now, setting $k = 1$ makes the middle term in the table disappear. Then every new term is computed by taking an alternating sum of the three terms above, as shown below:

$t\backslash x$	1	2	3	4	5	6	7	..
0	1	0	0	0	0	[0]	0	..
1	0	1	0	0	[0]	0	[0]	..
2	0	0	1	0	0	(0)	0	..
3	0	0	0	1	0	0	0	..
4	0	0	0	0	1	0	0	..
5	0	0	0	0	0	1	0	..
..	..	..	..	..	..	..	..	..

We can see that the wave is a single bump running from left to right at speed 1:



□

Exercise 3.25. (a) Solve the two-sided version of the above IVP. (b) Set up and solve an IVP with 2 bumps, n bumps.

Exercise 3.26. Implement a spreadsheet simulation for the case of non-constant m. Hint: you will need *two* buffers.

Next, we consider the case of *finite string* of springs and weights:

- the array of identical weights (m is constant) consists of N weights,
- the weights are distributed evenly over the length $L := (N-1)\Delta x$,
- the total mass is $M := Nm$,
- the springs are identical (k is constant),
- the total spring constant of the array is $K := k/N$,
- the geometry of the space is "flat": $\Delta x = \Delta x^\star$.

We can rewrite the above explicitly:

$$m\frac{u(x,t+1) - 2u(x,t) + u(x,t-1)}{(\Delta t)^2}$$
$$= k\Big[u(x+1,t) - 2u(x,t) + u(x-1,t)\Big]$$

Or,

$$\begin{aligned}\frac{u(x,t+1)-2u(x,t)+u(x,t-1)}{(\Delta t)^2} \\ &= \frac{k(\Delta x)^2}{m}\frac{u(x+1,t)-2u(x,t)+u(x-1,t)}{(\Delta x)^2} \\ &= \frac{k/N(N\Delta x)^2}{Nm}\frac{u(x+1,t)-2u(x,t)+u(x-1,t)}{(\Delta x)^2} \\ &= \frac{K(L+\Delta x)^2}{M}\frac{u(x+1,t)-2u(x,t)+u(x-1,t)}{(\Delta x)^2}.\end{aligned}$$

Solving for $u(x,t+1)$, we obtain the same formula:

$$u(x,t+1) \quad = 2u(x,t)-u(x,t-1)+\alpha\Big[u(x+1,t)-2u(x,t)+u(x-1,t)\Big],$$

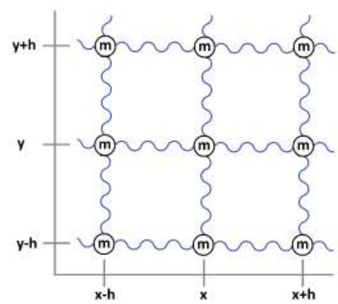
with a new coefficient:

$$\alpha := \left(\frac{\Delta t}{\Delta x}\right)^2 \frac{K(L+\Delta x)^2}{M}.$$

Exercise 3.27. Set up and solve the IVP for the finite string, for the simplified settings. Hint: mind the ends.

3.8 Waves in higher dimensions

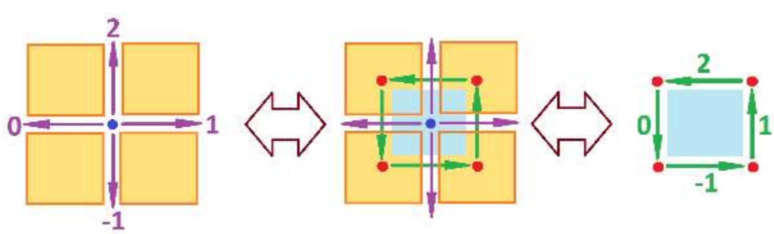
We next consider an *array* of objects connected by springs, both vertical and horizontal:



The forces exerted on the object at location x are the *four* forces of the four springs attached to it:

$$\begin{aligned}H &= H_{x,x-1}+H_{x,x+1}+H_{y,y-1}+H_{y,y+1} \\ &= k[u(x-1,y)-u(x,y)] \\ &+k[u(x+1,y)-u(x,y)] \\ &+k[u(x,y-1)-u(x,y)] \\ &+k[u(x,y+1)-u(x,y)].\end{aligned}$$

We still consider only 0- and 1-forms but on the standard 2-dimensional cubical complex $L=\mathbb{R}^2$. Hodge duality is slightly more complicated here. As an example, these are a 1-form φ and its dual 1-form φ^*:



Just as in the 1-dimensional case, each bracketed term in the expression for F is the exterior derivative: two with respect to x and two with respect to y:

$$\begin{array}{ll} k[u(x-1,y)-u(x,y)] & = kd_x u([x,x-1],y),\\ k[u(x+1,y)-u(x,y)] & = kd_x u([x,x+1],y),\\ k[u(x,y-1)-u(x,y)] & = kd_y u(x,[y,y-1]),\\ k[u(x,y+1)-u(x,y)] & = kd_y u(x,[y,y+1]). \end{array}$$

We can rearrange these terms as they all start from (x,y):

$$H = kdu \left\{ \begin{array}{ccc} & +\{x\}\times[y,y+1] & \\ +[x,x-1]\times\{y\} & & +[x,x+1]\times\{y\} \\ & +\{x\}\times[y,y-1] & \end{array} \right\},$$

where $d=(d_x,d_y)$. To get the duals of these edges, we just rotate them counterclockwise. Then,

$$H = kd^\star u \left\{ \begin{array}{ccc} & +[x^+,x^-]\times\{y^+\} & \\ +\{x^-\}\times[y^+,y^-] & & +\{x^+\}\times[y^-,y^+] \\ & +[x^-,x^+]\times\{y^-\} & \end{array} \right\},$$

where "$\pm$" stands for "$\pm 1/2$".

Observe that the dual edges are aligned counterclockwise along the boundary of the square neighborhood $[x^-,x^+]\times[y^-,y^+]$ of the point (x,y). That neighborhood is the Hodge dual of this point. We recognize this expression as a line integral:

$$H = \int_{\partial(\star(x,y))} \star kdu.$$

Now by the Stokes Theorem, the integral is equal to:

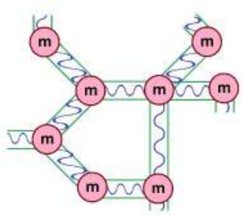
$$H = \star d_x \star kd_x u.$$

Since the left-hand side is the same as before, we have the same wave equation:

$$mu_{tt} = (k^\star u_x)_x.$$

In general, the medium may be non-uniform and anisotropic, such as wood:

We model the medium with a *graph* of springs and objects with a possibly non-trivial geometry:



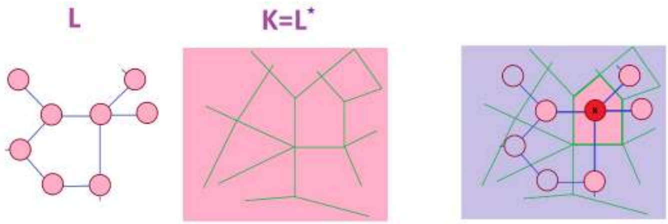
We are now in the same place as with the diffusion equation. We split the data into three categories:

- 1. the adjacency of the springs (and the objects) is the *topology*,
- 2. the lengths of the springs (and the angles and the areas of their cross-sections) is the *geometry*, and
- 3. the Hooke's constants of the springs is the *physics*.

They are given by, respectively:

- 1. the cell complex L,
- 2. the Hodge star operator $\star^m : C^m(L) \to C^{n-m}(L^\star)$, and
- 3. the primal 1-form $k \in C^1(L)$.

The two complexes dual to each other are shown below:



The total force that affects the object located at vertex x in L is

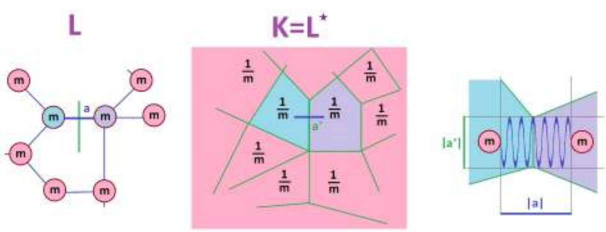
$$F = \star \int_{\partial(\star x)} \star k du.$$

Therefore, we end up with the same wave equation as above. Its right-hand side is seen as the full circle in the Hodge duality diagram:

$$\begin{array}{ccccc}
u \in & C^0(L) & \xrightarrow{\ d_x\ } & C^1(L) \\
& \uparrow \times \frac{|a^0|}{|b^n|} & \neq & \downarrow \times \frac{|b^{n-1}|}{|a^1|} \\
& C^n(L^\star) & \xleftarrow{\ d_x\ } & C^{n-1}(L^\star),
\end{array}$$

where a and b are dual.

The geometry of the primal complex L is given first by the lengths of the springs. What geometry should we give to the dual complex $K = L^\star$?



First, the meaning of the coefficients of the Hodge star operator are revealed to be represented by the stiffness $k(a)$ of spring a. In fact, it is known to be directly proportional to its cross-sectional area $|b|$ and inversely proportional to its length $|a|$:

$$k(a) := E\frac{|b|}{|a|},$$

where E is the "elastic modulus" of the spring. Plainly, when the angles are right, we have simply $b = a^\star$.

Second, we assign the reciprocals of the masses to be the n-volumes of the dual n-cells:

$$|x^\star| := \frac{1}{m(x)}.$$

Then we have the simplified – with the physics taken care of by the geometry – wave equation,

$$u_{tt} = (E^\star u_x)_x,$$

where E is the 1-form of elasticity. Its right-hand side is seen as the full circle in the Hodge duality diagram modified accordingly:

$$\begin{array}{cccc}
u \in & C^0(L) & \xrightarrow{d_x} & C^1(L) \\
& \Big\uparrow \times \frac{1}{m(b^n)} & \neq & \Big\downarrow \times k(a^1) \\
& C^n(L^\star) & \xleftarrow{d_x} & C^{n-1}(L^\star)
\end{array}$$

3.9 Simulating wave propagation with a spreadsheet

The recurrence formula for dimension 1 wave equation and constant k and m is:

$$u(x, t+1) \quad = 2u(x,t) - u(x, t-1) + \alpha\Big[u(x+1, t) - 2u(x,t) + u(x-1,t)\Big],$$

with

$$\alpha := (\Delta t)^2 \frac{k}{m}.$$

We put these terms in a table to be implemented as a spreadsheet:

	$x-1$	x	$x+1$
$t-1$		$-u(x, t-1)$	
t	$= \alpha u(x-1, t)$	$+2(1-\alpha)u(t, x)$	$+\alpha u(x+1, t)$
$t+1$		$u(x, t+1)$	

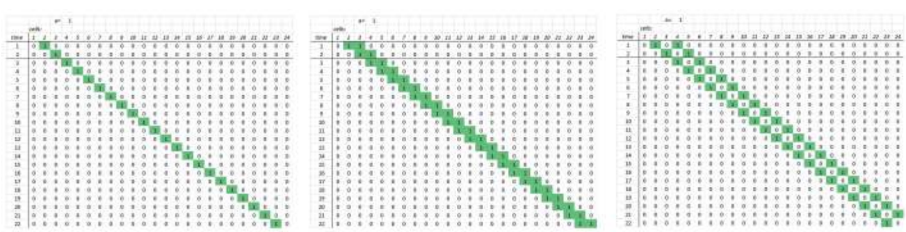
The simplest way to implement this dynamics with a spreadsheet is to use the first two rows for the initial conditions and then add one row for every moment of time. The Excel formula is:

```
= R1C5*R[-1]C[-1] + 2*(1-R1C5)*R[-1]C + R1C5*R[-1]C[1] - R[-2]C
```

Here cell `R1C5` contains the value of α.

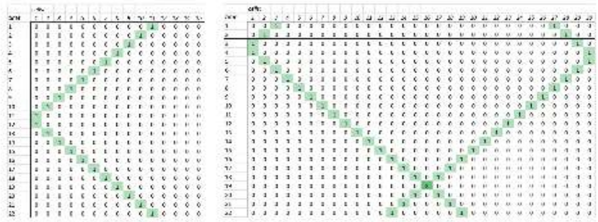
Note: See the files online.

Example 3.28. The simplest propagation pattern is given by $\alpha = 1$. Below we show the propagation of a single bump, a two-cell bump, and two bumps:

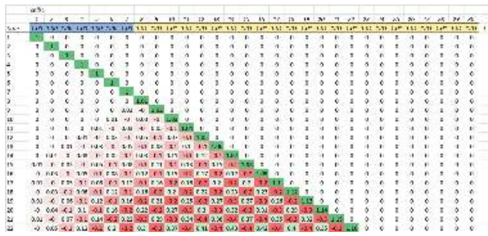


□

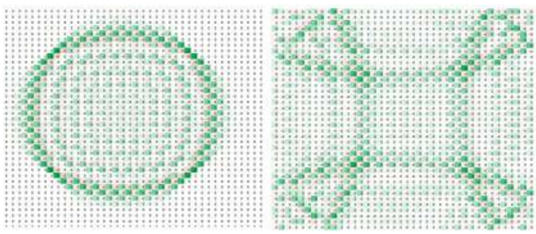
Exercise 3.29. Modify the spreadsheet to introduce walls into the picture:



Exercise 3.30. Modify the spreadsheet to accommodate non-constant data by adding the following, consecutively: (a) the stiffness k (as shown below), (b) the masses m, (c) the time intervals $|\Delta t|$.



The 2-dimensional case is treated with the formula given in the last subsection. This is the result of a circular wave with a single source bouncing off the walls of a square box.

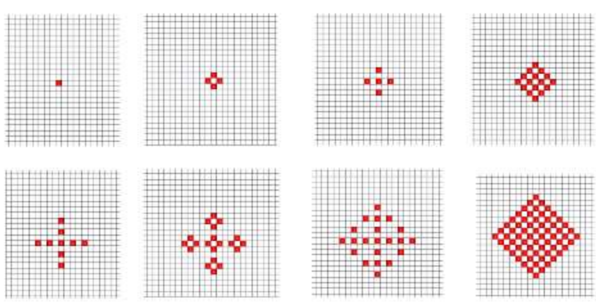


Exercise 3.31. Implement the wave illustrated on the first page of this chapter.

Example 3.32. The choice of

$$K = \mathbb{R}^2, \quad R = \mathbf{Z}_2,$$

produces this simple "cellular automaton":



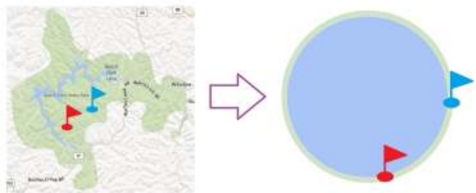
□

4 Social choice

4.1 The paradox of social choice

Let's review what we have learned about the problem of social choice.

Suppose we are to develop a procedure for finding a fair compromise on a choice of location (such as a camp in a forest) for two individuals:



We discovered that there is no such solution when the homology of the forest is non-trivial, such as one with a lake in the middle.

This is a more general setup. There are m voters, or agents, making their selections:
- the space of choices is a simplicial complex W;
- the choice made by the kth voter is a vertex A in W.

Then a solution to the *social choice problem* is a compromise-producing rule, i.e., a function that converts the m choices into one:

$$f : (A_1, ..., A_m) \mapsto B,$$

where $A_1, ..., A_m$, and B are vertices in K, that satisfies the following three conditions:
- **Continuity (Stability Axiom):** The function $f : W^m \to W$ generated by its values on the vertices is a simplicial map.
- **Symmetry (Anonymity Axiom):** The function is symmetric; i.e.,

$$fs = f, \ \forall s \in S_m.$$

- **Diagonality (Unanimity Axiom):** The function restricted to the diagonal of W^m is the identity; i.e.,

$$f\delta = \mathrm{Id}\,.$$

Here, and below, δ is the diagonal function:

$$\delta(x) = (x, ..., x).$$

The analysis relied on the concept of a mean of degree m on set X as a function $f : X^m \to X$ which is both symmetric and diagonal. Algebraic means are homomorphisms while topological means are continuous maps.

Theorem 4.1. There is an algebraic mean of degree m on an abelian group G if and only if G allows division by m. In that case, the mean is unique and is given by the standard formula.

Theorem 4.2. If G is a free abelian group and $f : G^m \to G$ is a symmetric homomorphism such that, for some integer q, we have:

$$f\delta = q\,\mathrm{Id}\,.$$

then, $m\big|q$. Moreover, f is a multiple of the sum,

$$\Sigma(x_1, ..., x_m) = x_1 + ... + x_m.$$

From these theorems, we concluded that the topological diagram below cannot be completed – because the algebraic one cannot be completed (over ring $R = \mathbf{Z}$):

$$\begin{array}{ccc} X & \xrightarrow{\delta} & X^m \\ & {\scriptstyle \mathrm{Id}} \searrow & \downarrow {\scriptstyle f=?} \\ & & X \end{array} \quad \xrightarrow{H} \quad \begin{array}{ccc} H_p(X) & \xrightarrow{\delta_*} & H_p(X^m) \\ & {\scriptstyle \mathrm{Id}} \searrow & \downarrow {\scriptstyle f_*=?} \\ & & H_p(X) \end{array}$$

The topological conclusion is the following.

Theorem 4.3. Suppose X is path-connected and has torsion-free integral homology. If there is a topological mean on X then X is acyclic.

As a corollary, we proved the following "impossibility theorem".

Theorem 4.4 (Impossibility). Suppose the space of choices W is path-connected and has torsion-free homology. Then the social choice problem on W has no solution unless W is acyclic.

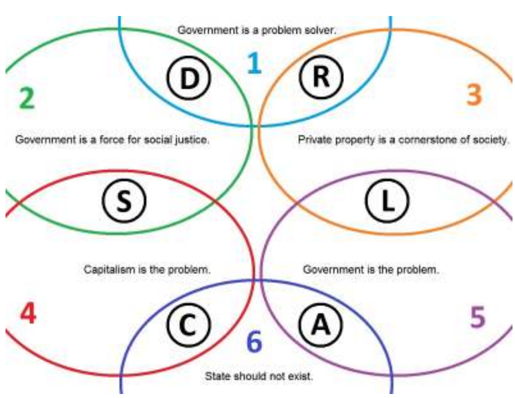
The conclusion, though surprising, isn't counter-intuitive. It is a common sense observation that excluding some options might make compromise impossible. Inevitably, a compromise-producing rule also becomes impossible.

Example 4.5 (political creeds). The six statements below are meant to represent an open cover of the political spectrum (in the US). In other words, we assume that *everyone supports at least one* of these statements:

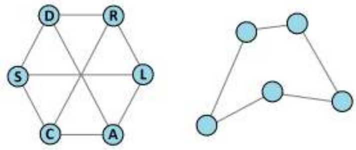
- 1. Government is a problem solver.
- 2. Government is a force for social justice.
- 3. Private property is a cornerstone of society.
- 4. Capitalism is the problem.
- 5. Government is the problem.
- 6. State should not exist.

We call the sets of individuals supporting these statements $U_1, ..., U_6$ respectively. We now assign letters to the intersections of these sets. This way, we classify all individuals based on which statements they support under the assumption that *no-one supports more than two*:

- 1 and 2: $D := U_1 \cap U_2$,
- 1 and 3: $R := U_1 \cap U_3$,
- 2 and 4: $S := U_2 \cap U_4$,
- 3 and 5: $L := U_3 \cap U_5$,
- 4 and 6: $C := U_4 \cap U_6$,
- 5 and 6: $A := U_5 \cap U_6$.



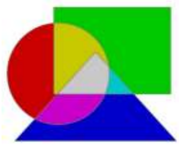
We now build the nerve of this cover. The sets become the vertices and the intersections become the edges. The result is the circle $\mathbf{S}^1$. Keep in mind that only the adjacency is represented in the complex and only the adjacency is illustrated; none of these concepts apply: "left-right", "close-far", "between", "opposite", or "extreme":



The theorem above then claims that, if these issues are put to the vote, there can be no electoral system that would always produce a fair compromise, whether it is an idea or an individual. □

Exercise 4.6. For the above example, use the possible sizes of the voter populations to create a filtration and evaluate its persistent homology.

A similar analysis shows that a compromise on *colors* may also be impossible – without the gray:

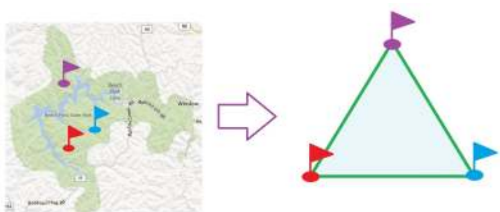


In an attempt to resolve this "paradox", we will allow choices that are more complex than just a single location or outcome. The topology of the space of choices will no longer be a problem.

4.2 Ratings, comparisons, ranking, and preferences

Instead of just pointing out an ideal location, either of the two hikers may attempt to express a more complex set of preferences.

Typically, the hiker can assign a number to each location in the forest reflecting his appreciation (i.e., the utility) of this choice. In the discrete interpretation, we suppose there are three camp sites, A, B, C, set up in the forest. Then hiker X assigns a number to each camp site.



Then X's vote is a triple $(a, b, c) \in R^3$, and so is Y's, where $R = \mathbf{R}$ or $\mathbf{Z}$ is our ring of coefficients. The goal is to find a triple that best approximates the desires of the two. A compromise-producing rule for the two hikers is then a function:

$$f : R^3 \times R^3 \to R^3.$$

Even though Continuity doesn't apply anymore, Symmetry and Diagonality are natural requirements. We will see that such a function still might not exist, depending on our choice of R.

The main examples of this, more sophisticated choice-making are:

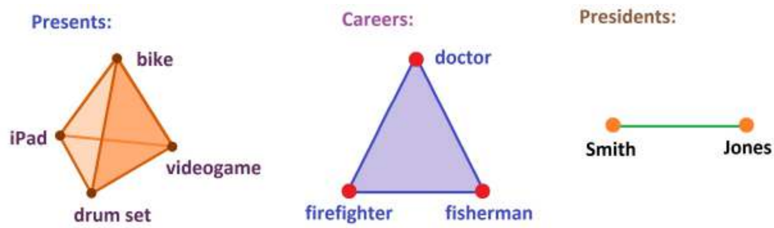
- *ratings*: $X(A) = 1,\ X(B) = 4$, etc.;
- *comparisons*: $x(AB) = -1,\ x(BC) = 3$, etc.;
- *rankings*: $X(A) = \#1,\ X(B) = \#2$, etc.; and
- *preferences*: $A <_x B,\ B <_x C$, etc.

Can any of these votes be combined into one compromise vote following the fairness principles presented in the last subsection?

More generally, there are n alternatives or candidates:

$$A := \{0, 1, 2, ..., n-1\} = \{A_0, A_1, A_2, ..., A_{n-1}\}.$$

They are ordered. We also suppose that these candidates are the vertices of a *path-connected* simplicial complex K. All candidates are subject to evaluation but the presence of edge AB in K reflects the fact that A and B are *comparable*.



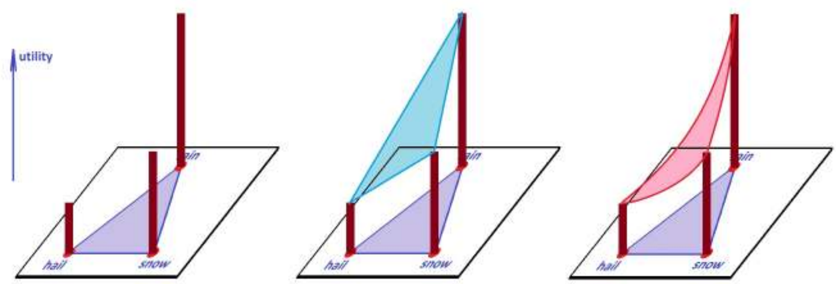
We will look into the first two options as they are subject to the algebra we have developed previously.

A vote is an element of $C^* = C^*(K)$. In particular, a rating vote is a combination of numbers assigned to each candidate. Therefore, this is a 0-cochain on K, $X \in C^0(K)$.

Example 4.7 (utility). Once such a vote is given, we may also produce a vote for every *linear combination* of candidates:

$$X\Big(\sum_i p_i A_i\Big) := \sum_i p_i X(A_i).$$

We recognize this as the expected utility on K:



□

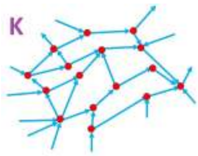
Furthermore, a comparison vote is a combination of numbers assigned to each pair of candidates. Therefore, this is a 1-cochain on K, $x \in C^1(K)$.

Note: One can also study rankings via the complex of orderings (subsection IV.4.4).

How these votes are tallied is discussed in the next subsection. Suppose the elections have produced a single vote, who wins?

If this is a *rating vote* X, the winner is the one (possibly tied) with the largest rating.

If this is a *comparison* vote x, the result is a weighted directed graph:



With many circular patterns present, it may be impossible to determine a winner...

4.3 The algebra of vote aggregation

There may be several *types* of elections, or evaluating procedures, to run – based on the types of votes allowed. For example, we may limit the votes to:

$$Q = C^k, Z^k, B^k, ...,$$

as subgroups of $C^*(K)$. For a given Q, an *election* is any combination of m votes from Q. Then the set of all possible elections is the product Q^m.

For any type of elections, one needs to convert an election, an element of Q^m, into a single vote *of the same type*, an element of Q.

Definition 4.8. A *tally* is any function:

$$f : Q^m \to Q,$$

written as

$$x = f(x_1, ..., x_m).$$

Exercise 4.9. Show that a tally doesn't have to be generated by a simplicial map.

The simplest and the most natural way to tally is to add the votes of the m voters:

$$x = \Sigma(x_1, ..., x_m) := x_1 + ... + x_m.$$

We will call Σ the *sum tally*. For $R = \mathbf{R}$, there is also the *mean tally*:

$$x = M(x_1, ..., x_m) := \frac{1}{m}(x_1 + ... + x_m).$$

Exercise 4.10. Show that the sum and the mean tallies are well-defined for both non-circular, $f^k : (Z^k)^m \to Z^k$, and rating, $f^k : (B^k)^m \to B^k$, comparison votes.

Both are homomorphisms, but does all tally have to be?

Suppose, hypothetically, that two elections of the same type are run two days in a row. Suppose every voter casts his vote twice with the possibility that he changes his mind by the next day or simply splits his vote arbitrarily. We want to have a single outcome. The first option is to take these two votes and add them together to create a combined vote of this voter, and after this is done for all voters, tally the votes according to the election rule to find the outcome vote. The second option is to tally at the end of the first day, then, separately, at the end of the second, and then add the two outcomes together. The results should match.

We introduce a new axiom.

Additivity: Tally f is a homomorphism:

$$f(x_1 + y_1, ..., x_m + y_m) = f(x_1, ..., x_m) + f(y_1, ..., y_m).$$

Exercise 4.11. What if we replace "add" with "average"?

Exercise 4.12. What should we require about scalar multiplication?

Now, what makes a good electoral system?

A *fair tally* has to be additive and but should also these two axioms:

- If two voters flip their votes, it won't affect the tallied vote (Symmetry).
- If all voters vote identically, then the tallied vote will coincide with this vote (Diagonality).

Therefore, a fair tally f is a mean on Q! What we know about means on groups implies the following.

Theorem 4.13.

- Over $\mathbf{Z}$, there is no fair tally.
- Over $\mathbf{R}$, the only fair tally is the mean tally.

4.4 Aggregating rating votes

So far, the votes have been treated as mere elements of an arbitrary subgroup $Q \subset C^*(K)$ and a tally for such an election has to satisfy some generic rules: Additivity and Symmetry.

This time, we consider ratings only; we choose our subgroup to be $Q := C^0(K)$. Below, we speak of *a rating tally*

$$f : (C^0)^m \to C^0.$$

We restate the latter axiom.

Symmetry I: The election result is independent of the permutation of the votes; i.e.,

$$f(sX) = f(X), \ \forall s \in \mathcal{S}_m.$$

Exercise 4.14. Here, s is meant to be an $n \times n$ matrix that permutes the components of the vector. Describe such a matrix.

In other words, as the voters, $1, ..., m$, interchange their votes, $X_1, ..., X_m$, the election outcome remains unchanged.

This requirement is meant to guarantee that there is no special voter, a dictator. But what about a special *candidate*?

Written coordinate-wise, a tally is a function of matrices (over R). This matrix represents an election and its (i, j)-entry is the rating assigned by the jth voter to the ith candidate:

$$
\begin{array}{c|cccc}
 & & \text{voters} & & \\
\textit{Ballot} & & X_1 \quad ... \quad X_m & & \text{outcome} \\
\hline
\begin{array}{c}\text{candidate } A_1\\ ...\\ \text{candidate } A_n\end{array} & f & \left(\begin{array}{ccc} a_{11} & ... & a_{1m}\\ ... & & ...\\ a_{n1} & ... & a_{nm}\end{array}\right) & = & \left[\begin{array}{c} c_1\\ ...\\ c_n\end{array}\right]
\end{array}
$$

According to Symmetry I,

- interchanging the *columns* in the matrix doesn't affect the outcome.

In addition, we will require that

- interchanging the *rows* in the matrix interchanges the entries in the outcome accordingly.

This requirement is rephrased as follows:

Symmetry II: The election result is independent of the permutation of the candidates; i.e.,

$$f(X)(sa) = s^{-1}f(X)(a), \ \forall X \in (C^0)^m, \ \forall s \in \mathcal{S}_n, \ \forall a \in C_0.$$

So, renumbering the candidates, $A_1, ..., A_n$, in the ballots produces the same election outcome once we renumber them back.

Exercise 4.15. Show that a "constant" tally, such as $f(E) = A_1^*$ for every $E \in (C^0)^m$, satisfies Symmetry I but doesn't satisfy Symmetry II.

Proposition 4.16. The sum tally Σ satisfies Symmetry II.

We no longer require the outcome of a unanimous election to be exactly the same quantitatively as that vote but only that the order of the candidates is to be preserved. Let's suppose that the voters voted unanimously for candidate A; i.e., for all $i = 1, .., m$, we have

$$X_i(A) = 1, \; X_i(B) = 0,$$

with B any other candidate. In other words, $X_i = A^*$. Then the outcome of the elections should also be a zero vote for each candidate but A, which should have a positive vote. Therefore, there is a positive coefficient $k_A \in R$ such that:

$$f(A^*, ..., A^*) = k_A A^*.$$

We rephrase this condition as follows:

$$f\delta \text{ is a positive diagonal matrix.}$$

Furthermore, Symmetry II implies that k_A is independent of A; we have the final version of our new axiom.

Diagonality: The election result respects unanimous votes; i.e.,

$$f\delta = k \,\mathrm{Id}, \; k \in R, k > 0.$$

Then, what we know about means proves the following.

Theorem 4.17. A rating tally that satisfies Additivity, Symmetry I, Symmetry II, and Diagonality is a multiple of the sum tally; i.e.,

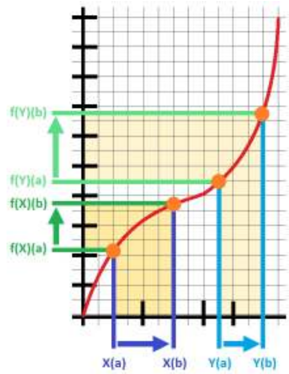
$$f = k\Sigma, \; k \in R, k > 0.$$

The next axiom is meant to prevent a voter from manipulating the outcome by voting strategically. It is commonly called *independence of irrelevant alternatives.*

Monotonicity: If no voter has changed his preference for a given pair of candidates from the election to the next, then the preference of the election result for these candidates hasn't changed either; i.e.,

$$\begin{aligned}
&\forall a, b \in C_0, \; X, Y \in (C^0)^m, \\
&\operatorname{sign}\Big(X_i(a) - X_i(b)\Big) = \operatorname{sign}\Big(Y_i(a) - Y_i(b)\Big) \quad \Longrightarrow \\
&\operatorname{sign}\Big(f(X)(a) - f(X)(b)\Big) = \operatorname{sign}\Big(f(Y)(a) - f(Y)(b)\Big).
\end{aligned}$$

Note: Tally f is a function of $X = (X_1, ..., X_m) \in (C^0)^m$. But with a, b fixed, X_i becomes an element of R and so does $f(X)$, while X is an element of R^m. Then we can restate the condition as follows: at two locations, if the input of the function increases (or decreases), then the output simultaneously increases or decreases in both locations. Such a function can only be an increasing or decreasing one, i.e., monotonic.



Exercise 4.18. What is the order relation on R^m being used here?

Exercise 4.19. Explain the meaning of s in the axiom as an operator and describe its matrix.

Example 4.20 (election manipulation). Let $n := 4$ and $m := 3$. Consider this election and its outcome:

	A	B	C	D
X	4	3	2	1
Y	3	4	2	1
Z	3	4	2	1
U	10	11	6	3

Here $U = \Sigma(X, Y, Z)$ is the resulting vote.

We can also express these ratings as rankings:

$$\begin{array}{rl} X: & A > B > C > D \\ Y: & B > A > C > D \\ Z: & B > A > C > D \\ \hline U: & B > A > C > D \end{array}$$

Here, each candidate receive four points for coming first, three for coming second, etc.

We have the election result:

$$U: \ B > A.$$

Now suppose that in the next election, voter X, now X', moves B from second place to last on his list:

	A	B	C	D
X'	4	1	3	2
Y	3	4	2	1
Z	3	4	2	1
V	10	9	7	4

Here $V = \Sigma(X', Y, Z)$ is the resulting vote.

We can express these ratings as rankings:

$$\begin{array}{rl} X': & A > C > D > B \\ Y: & B > A > C > D \\ Z: & B > A > C > D \\ \hline V: & A > B > C > D \end{array}$$

We have the election result:

$$V: \ A > B.$$

As we can see, while each voter's ranking of A with respect to B is the same in the two elections, the final rankings are different.

What happened is that voter X – wishing victory for A – correctly predicted that B was the most dangerous competition. He then pushed B to the bottom of his list to reduce B's rating as much as possible. This manipulative step alone lead to A's victory. □

Exercise 4.21. What is the minimal number of candidates that allows such a manipulation?

Exercise 4.22. If we normalize the tallied votes, the result is a "lottery" on K. Running an actual lottery can then be used to determine the winner. What is the effect of the manipulation described above on this lottery?

We have thus proven the following.

Proposition 4.23. The sum tally (and its multiples) does not satisfy Monotonicity.

Combined with the last theorem the proposition gives us the following.

Theorem 4.24 (Impossibility). There is no rating tally $f : (C^0)^m \to C^0$ that satisfies Additivity, Symmetry I, Symmetry II, Diagonality, and Monotonicity.

The theorem is an algebraic version of *Arrow's Impossibility Theorem.*

It appears that ratings alone aren't versatile enough to be a basis of a fair electoral system. This suggests that we should consider allowing comparison votes – even at the risk of facing "inconsistent" votes: $A > B > C > A$.

4.5 Aggregating comparison votes

This time, we consider comparison votes and choose our group to be $Q := C^1(K)$. Below, we speak of *a comparison tally*

$$f : (C^1)^m \to C^1.$$

We will restate some of the axioms. The first is without change.

Symmetry I: The election result is independent of the permutation of the votes; i.e.,

$$f(sX) = f(X),\ \forall s \in \mathcal{S}_m.$$

Here, the voters, $1, ..., m$, interchange their votes, $X_1, ..., X_m$, each of which is pairwise comparison, such as: $X_1(AB) = 3$.

This time, a tally is a function of 3-dimensional ($m \times m \times n$) matrices. Each such matrix represents a comparison election and its (i, j, k)-entry is the score assigned by the kth voter to the pair of the ith and jth candidates. This is what a single ballot looks like:

	candidates A_1	...	A_n
candidate A_1	a_{11}	...	a_{1n}
...	...		...
candidate A_n	a_{n1}	...	a_{nn}

A tallied election outcome is also meant to be such a table.

According to Symmetry I, reshuffling the ballots doesn't affect the outcome. In addition, we require that interchanging the rows and columns – identically – in all of these matrices interchanges the rows and columns in the outcome table in the exact same way.

Symmetry II: The election result is independent of the permutation of the candidates; i.e.,

$$f(X)(sa) = s^{-1}f(X)(a),\ \forall X \in (C^1)^m,\ \forall s \in \mathcal{S}_n,\ \forall a \in C_1.$$

Just as before, renumbering the candidates, $A_1, ..., A_n$, in the ballots produces the same election outcome once we renumber them back.

Proposition 4.25. The sum tally Σ satisfies Symmetry II.

We still require that the outcome of a unanimous election preserves the order of the candidates. We then follow the same logic as with a rating tally to make this precise. Suppose, the voters vote unanimously for candidate A over B; i.e., for all $i = 1, .., m$, we have

$$X_i(AB) = 1,\ X_i(CD) = 0,$$

with CD any other pair of candidates (including AC and BC for $C \neq A, B$). In other words, $X_i = AB^*$. Then the outcome of the elections should also produce a zero vote for each pair but AB, which should have a positive vote. Therefore, there is a positive coefficient $k_{AB} \in R$ such that:

$$f(AB^*, ..., AB^*) = k_{AB}AB^*.$$

This condition takes the same form as with a rating tally.

Diagonality: The election result respects unanimous votes; i.e.,

$$f\delta = k\,\mathrm{Id},\ k \in R, k > 0.$$

Once again, we apply the theorem about means to conclude that there is, essentially, just one such tally.

Theorem 4.26. A comparison tally that satisfies Additivity, Symmetry I, Symmetry II, and Diagonality is a (positive) multiple of the sum tally; i.e.,

$$f = k\Sigma,\ k \in R, k > 0.$$

The last axiom is the independence of irrelevant alternatives. For ratings, it takes the form of Monotonicity. For comparisons, we restate it by following the familiar link between 0- and 1-cochains: the difference of the rating votes for candidates A, B is understood as a comparison vote for AB. In other words, we use the following:

$$d(X_1)(AB) = X_1(B) - X_1(A).$$

Positivity: If no voter has changed his preference for a given pair of candidates from the election to the next, then the preference of the election result for these candidates hasn't changed either; i.e.,

$$\begin{array}{l}\forall a \in C_1,\ X, Y \in (C^1)^m,\\ \mathrm{sign}\, X_i(a) = \mathrm{sign}\, Y_i(a) \quad \Longrightarrow\\ \mathrm{sign}\, f(X)(a) = \mathrm{sign}\, f(Y)(a).\end{array}$$

After all, the derivative of a monotonic function is either all positive or all negative.

Example 4.27 (election manipulation). Let $n := 4$ and $m := 3$. Consider this election (and its outcome) from the last subsection. Each voter assigns 1 to each pairwise preference:

	AB	AC	AD	BC	BD	CD
X	1	1	1	1	1	1
Y	-1	1	1	1	1	1
Z	-1	1	1	1	1	1
U	-1	3	3	3	3	3

Here $U = \Sigma(X, Y, Z)$ is the resulting vote.

In the second election, voter X, now X', moves B from the second place to the last on his list:

	AB	AC	AD	BC	BD	CD
X'	1	1	1	-1	-1	1
Y	-1	1	1	1	1	1
Z	-1	1	1	1	1	1
V	-1	3	3	1	1	3

Here $V = \Sigma(X', Y, Z)$ is the resulting vote.

The election results match:

$$U, V: \ B > A.$$

Thus, even though X pushed B to the bottom of his list, B is still ahead of A. The manipulation failed. □

Unlike ratings,

- the comparison sum tally doesn't violate independence of irrelevant alternatives.

That's why there is no impossibility theorem. The following sums up the results.

Theorem 4.28 (Possibility). The sum tally $\Sigma : (C^1)^m \to C^1$ is a comparison tally that satisfies Additivity, Symmetry I, Symmetry II, Diagonality, and Positivity.

Exercise 4.29. Provide a similar analysis for votes of degree 2; i.e., $x \in C^2(K)$.

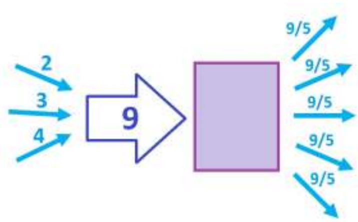
What remains to be decided, with the comparison votes successfully tallied, who is the winner?

4.6 Google's PageRank

PageRank is a popular method of evaluating the relative importance of web-pages based on their incoming and outgoing links. It is a mathematical idea that was at the core of Google's web search algorithm. According to Google, "PageRank thinks of links as 'votes', where a page linking to another page is essentially casting a vote for that page... PageRank also considers the importance of each page that casts a vote, as votes from some pages are considered to have greater value, thus giving the linked page greater value." This *is* how such evaluation is supposed to work; however, the actual formula for PageRank shows that it's *not* an electoral system in the sense we have put forward.

Let's review the basics of the algorithm of PageRank and subject them to mathematical scrutiny...

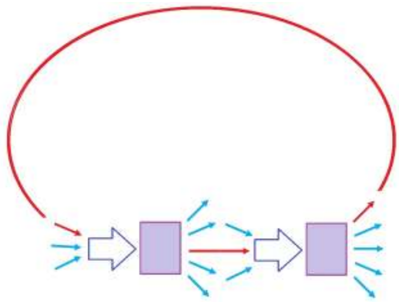
The idea of the definition is recursive.



Suppose a given page has three (or any other number) incoming links and five outgoing links. Let's assume that at the last stage the page received 9 points of "PageRank" from the incoming links. Then, at the next stage, these 9 points are distributed equally to the five outgoing links that carry these points to those pages. Consequently, each carries 9/5 points.

Starting with an equal number of points for every page, we allow the flow go around the web following the links and we recompute the PageRank for every page at every step. Once everything settles, the pages have accumulated their PageRanks as ratings and they are then ranked accordingly.

The simple formula fails when there are loops in the graph of links:



The reason is that a proportion of the initial PageRank of one of these pages will travel the full loop.

Example 4.30. Consider the propagation of PageRank through this simple graph:

$$\begin{array}{ccc} 1 & \longrightarrow & 1 \\ & \nwarrow & \downarrow \\ 1 & \longrightarrow & 1 \end{array} \rightsquigarrow \begin{array}{ccc} 1 & \longrightarrow & 1 \\ & \nwarrow & \downarrow \\ 0 & \longrightarrow & 2 \end{array} \rightsquigarrow \begin{array}{ccc} 2 & \longrightarrow & 1 \\ & \nwarrow & \downarrow \\ 0 & \longrightarrow & 1 \end{array} \rightsquigarrow \begin{array}{ccc} 1 & \longrightarrow & 2 \\ & \nwarrow & \downarrow \\ 0 & \longrightarrow & 1 \end{array} \rightsquigarrow \ ... \xrightarrow{\qquad\qquad} ?$$

There is no convergence! □

In order to ensure its convergence, the algorithm is modified. This time, only a *proportion*, r, of the current PageRank is passed to the target pages. This number, called the *decay coefficient*, is commonly chosen to be $r = .85$. As there is no justification for choosing this over another value, r is a non-mathematical, made-up parameter of this model.

Even after such a modification, there are examples of undesirable behavior, such as accumulation of PageRank at the dead-ends.

Example 4.31. Suppose pages are linked to each other consecutively:

$$A \to B \to C \to ... \to Y \to Z.$$

Then all pages will eventually pass their entire PageRanks to Z. As a result, pages A-Y are tied at 0. Then the seemingly obvious ranking,

$$A < B < C < ... < Y < Z,$$

is lost! □

In order to ensure that every page will get some of the PageRank, the algorithm is modified, again. This time, it is assumed that pages with no outbound links are linked to *all* other pages.

Further examples may cause (and perhaps may have caused) further modifications of the algorithm. Instead, we simply present the most common way to compute the PageRank.

Definition 4.32. Suppose the web is represented as a directed graph and let E be the set of edges and N the set of nodes in the graph. Then P_i, the PageRank of the ith node i, is defined by the recursive formula:

$$P_i = (1 - r) \sum_{ij \in E} \frac{P_j}{\deg j} + \frac{r}{\#N}.$$

The totality of the PageRanks is a rating (not a ranking).

There are still a number of surprising features. One is this:

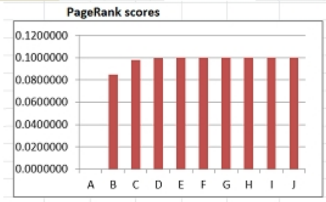
- adding *outbound* links to your page may improve its PageRank.

Example 4.33 (adding outbound links). Suppose we have ten "web-pages": $A, B, ..., J$. They are listed in the table as both rows and columns. The links are the entries in the table.

First we suppose that they link to each other consecutively:

$$A \to B \to C \to ... \to J.$$

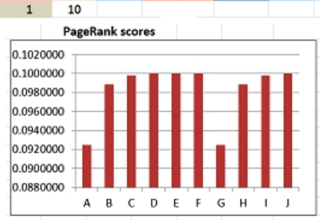
Links between pages		To:	To:	To:	To:	To:	To:	To:	To:	To:	To:	degree		PAGERANK		
	Pages:	A	B	C	D	E	F	G	H	I	J	out	Pages:	Score	Position	Pages:
From:	A	0	1	0	0	0	0	0	0	0	0	1	A	0.0000000	10	A
From:	B	0	0	1	0	0	0	0	0	0	0	1	B	0.0850000	9	B
From:	C	0	0	0	1	0	0	0	0	0	0	1	C	0.0977500	8	C
From:	D	0	0	0	0	1	0	0	0	0	0	1	D	0.0996625	7	D
From:	E	0	0	0	0	0	1	0	0	0	0	1	E	0.0999494	6	E
From:	F	0	0	0	0	0	0	1	0	0	0	1	F	0.0999924	5	F
From:	G	0	0	0	0	0	0	0	1	0	0	1	G	0.0999989	4	G
From:	H	0	0	0	0	0	0	0	0	1	0	1	H	0.0999998	3	H
From:	I	0	0	0	0	0	0	0	0	0	1	1	I	0.1000000	2	I
From:	J	0	0	0	0	0	0	0	0	0	0	0	J	0.1000000	1	J
degree	in	0	1	1	1	1	1	1	1	1	1	9				



The PageRank points, as well as the rankings, of the pages are computed and displayed on the right. The scores for $A, ..., J$ are increasing in the obvious way: from 0 to .1. In particular, F is #5.

Next, suppose F adds a link to A, $F \to A$, which completes the loop. You can see that 1 has appeared in the first column of the table:

Links between pages		To:	To:	To:	To:	To:	To:	To:	To:	To:	To:	degree		PAGERANK		
	Pages:	A	B	C	D	E	F	G	H	I	J	out	Pages:	Score	Position	Pages:
From:	A	0	1	0	0	0	0	0	0	0	0	1	A	0.0925000	10	A
From:	B	0	0	1	0	0	0	0	0	0	0	1	B	0.0988750	8	B
From:	C	0	0	0	1	0	0	0	0	0	0	1	C	0.0998312	6	C
From:	D	0	0	0	0	1	0	0	0	0	0	1	D	0.0999747	4	D
From:	E	0	0	0	0	0	1	0	0	0	0	1	E	0.0999962	2	E
From:	F	1	0	0	0	0	0	1	0	0	0	2	F	0.0999994	1	F
From:	G	0	0	0	0	0	0	0	1	0	0	1	G	0.0925000	10	G
From:	H	0	0	0	0	0	0	0	0	1	0	1	H	0.0988750	8	H
From:	I	0	0	0	0	0	0	0	0	0	1	1	I	0.0998312	6	I
From:	J	0	0	0	0	0	0	0	0	0	0	0	J	0.0999747	4	J
degree	in	1	1	1	1	1	1	1	1	1	1	10				



Now, suddenly F ranks first! Thus, adding an *outbound* link has brought this page from #5 to #1. □

Note: See the files online.

Another undesirable but not unexpected feature is:

- changing the decay coefficient may change your rankings.

PageRank's paradigm is about passing something valuable (popularity, significance, authority, etc., but not a vote) from a page to the next along its links. Then, the PageRank points show who, after all this "redistribution of wealth", ends up on top. This process is called *advection*: a fluid flow carries some substance, such as sand, around and gradually deposits it in different locations. Certainly, the amount of sand accumulated "at the end of the day" at a given location should be expected to depend on the percentage of sand deposited per unit of time.

Exercise 4.34. Give examples of how the rankings may change depending on the value of the decay coefficient.

The approach to voting explained in this section is applied to web search as follows:

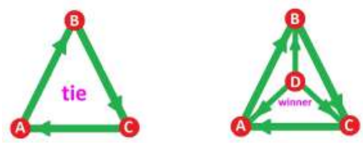
- the presence of a link from A to B is a comparison vote of -1 on AB.

4.7 Combining ratings with comparisons

The outcome of an election is a single vote. It may be a comparison vote and it can be "inconsistent" (even when each vote isn't), such as

$$A > B > C > A.$$

When the differences are identical, the outcome should be treated as a *tie* (left):



Definition 4.35. A *perfectly circular vote* is a comparison vote $x \in C^1$ with

$$x(A_iA_i) = p \in R,$$

for some circular sequence of adjacent vertices $A_0, A_1, ..., A_n = A_0$, and 0 for the rest.

Now, more complex votes may contain non-circular components (above right):

$$A > B > C > A, \quad D > A, D > B, D > C.$$

Here, D is the winner.

We need to learn how to extract the circular component (left) from a given vote so that we can discard it, as follows:

$$A = B = C, \quad D > A, D > B, D > C.$$

A related idea appears, for example, when we study the motion on an inclined surface. Then only the component of the force parallel to the surface matters and the normal component is to be discarded. Then we will need the concept of the *orthogonal complement* of a subset P of an inner product space V:

$$P^\perp := \{v \in V : v \perp u, \ \forall u \in P\}.$$

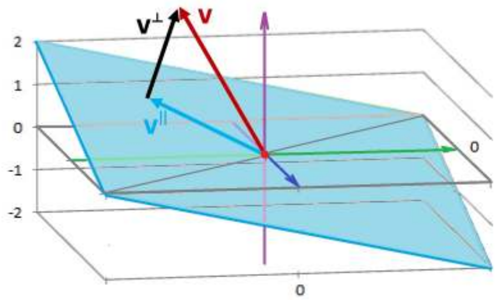
It is a submodule.

Proposition 4.36. Suppose P is a subset of an inner product space V. Then its orthogonal complement is a summand:

$$V =< P > \oplus P^\perp.$$

Moreover, every element of V is uniquely represented as the sum of its *parallel and orthogonal components*:

$$v = v^{||} + v^\perp, \quad v^{||} \in < P >, v^\perp \in P^\perp.$$



We next consider progressively more complex voting situations, which will eventually bring us to this concept.

In a basic *electoral system*, we have n candidates located at the vertices of a simplicial complex K. We assume that the ring of coefficients R is either $\mathbf{Z}$ or $\mathbf{R}$. A vote x is an element of $C^* = C^*(K)$:

$$a = a^0 + a^1 + a^2 + ..., \ a^i \in C^i(K).$$

The vote may be cast by a single voter or is the result of an election. In either case, we don't assume any consistency or strategy on behalf of the voter and, therefore, no interdependence among $a^0, a^1, a^2, ...$

Who is the winner?

Definition 4.37. For a given vote a, the *winner of degree* 0 is the candidate(s) with the largest value of a^0.

Thus,

$$winner := \arg \max_{i \in K^{(0)}} a^0(i).$$

Above, a^0 is a rating vote and a^1 is a comparison vote. We discarded the comparison vote a^1 in the definition. However, it is unwise to throw out meaningful information and it is in fact unacceptable when there is a tie.

Example 4.38. We have already seen that for the vote

$$a^0 = 1, \quad a^1(AB) = 1, \ a^1(BC) = -1, \ a^1(CA) = 0,$$

the tie in dimension 0 is broken by using the fact that a^1 is a coboundary:

$$b^0(A) = 0, \ b^0(B) = 1, \ b^0(C) = 0 \Longrightarrow \partial^0 b^0 = a^1.$$

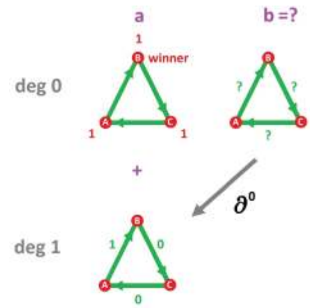
□

Definition 4.39. Suppose we are given a vote a with the comparison component a^1 that is a coboundary. Then the *winner of degree* 1 is the candidate(s) with the largest value of $a^0 + b^0$, where b^0 is any 0-chain that satisfies $\partial^0 b^0 = a^1$.

However, being a coboundary isn't necessary for the comparison component to be useful.

Example 4.40. We consider the simple case when the ratings are tied but the comparisons aren't:

- $a^0(A) = a^0(B)$,
- $a^1(A) < a^1(B)$.

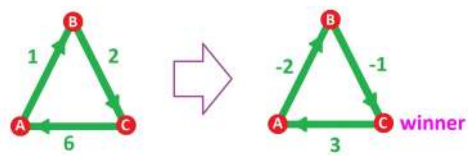


Then, plainly, B is the winner! □

Exercise 4.41. Show that a^1 doesn't have to be a coboundary to be useful.

Exercise 4.42. Show that using the comparison component of the vote may change the outcome even when there is no tie.

Example 4.43. In the next example, we have a circular comparison vote. Then subtracting the average of its values reveals a rating vote:



□

Exercise 4.44. State and prove a theorem that generalizes the above example.

Now, what if we can't see the winner by just an examination of the vote as in these examples? We do have an algebraic procedure for the case when a^1 happens to be a coboundary and now we need one for the general case.

Example 4.45. We consider the simple case above:

$$a^0 = (1, 1, 1),\ a^1 = (1, 0, 0).$$

Then a^1 isn't a coboundary: there is no b^0 with $\partial^0 b^0 = a^1$ to be used to determine the winner. The idea is to find the best *substitute* coboundary of a^1 as an approximation over the reals.

First, let's find the space of 1-coboundaries (i.e., the rating comparison votes),

$$B^1 = \partial^0 C^0 \subset C^1.$$

If x, y, z are the values of a coboundary on the three edges, then there are some values p, q, r of a 0-cochain on the vertices such that

$$x = q - p,\ y = r - q,\ z = p - r.$$

This is equivalent to the following:

$$x + y + z = 0.$$

This plane – in the 3-dimensional Euclidean space C^1 – is the coboundary group B^1. Then, in order to find the best coboundary approximation of $a^1 = (1, 0, 0)$, we minimize the distance to this plane:

$$b^1 := \arg\min\{(x-1)^2 + y^2 + z^2 : x + y + z = 0\}.$$

The meaning of b^1 is, clearly, the projection of a^1 on B^1. It can be found by simple linear algebra, as follows. A vector normal to this plane is chosen, $N = (1, 1, 1)$. Then,

$$b^1 = a^1 + tN = (1, 0, 0) + t(1, 1, 1) = (1 + t, t, t)$$

for some t. Because $b^1 \in B^1$, we have

$$(1 + t) + t + t = 0 \Longrightarrow t = -\tfrac{1}{3}.$$

Then the projection of $(1, 0, 0)$ is

$$b^1 = \left(\tfrac{2}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}\right).$$

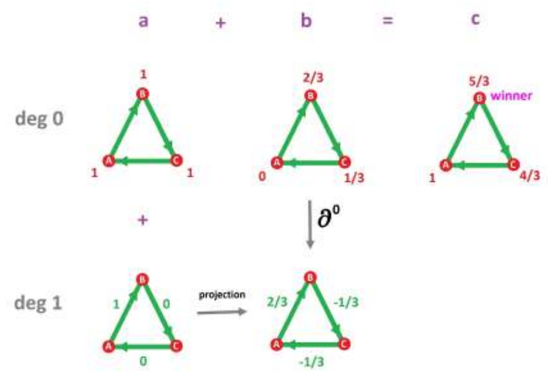
It is the coboundary of:

$$b^0 = \left(0, \tfrac{2}{3}, \tfrac{1}{3}\right).$$

Then,

$$a^0 + b^0 = (1, 1, 1) + \left(0, \tfrac{2}{3}, \tfrac{1}{3}\right) = \left(1, \tfrac{5}{3}, \tfrac{4}{3}\right),$$

and B is, again, the winner!



□

Exercise 4.46. Provide a similar analysis for the following vote:

$$a^0 = (1, 1, 0),\ a^1 = (1, 0, 1).$$

Exercise 4.47. What happens if in the last example we choose the ring of coefficients to be $R = \mathbf{Z}$ instead of $\mathbf{R}$?

4.8 Decycling: how to extract ratings from comparisons

To be able to speak of projections, we need the cochain group $C^1(K)$ to be an inner product space.

It is uncomplicated. In fact, we will supply such a structure to the whole cochain group $C^*(K)$, by specifying an *orthonormal basis.*

A basic rule we have been enforcing is the equality between any two candidates, i.e., the vertices of our complex K. The set of vertices has been the standard basis of $C_0(K)$; this time, in addition, we declare that

- the vertices $\{A\}$ of K form an orthonormal basis of $C_0(K)$, and
- the duals of vertices $\{A^*\}$ form an orthonormal basis of $C^0(K)$.

Similarly, the equality between any two *pairs* of candidates is also required. Then the set of edges of K has been the standard basis of $C_1(K)$; this time, in addition, we declare that

- the edges $\{a\}$ of K form an orthonormal basis of $C_1(K)$, and
- the duals of edges $\{a^*\}$ form an orthonormal basis of $C^1(K)$.

These assumptions provide us with *inner products* for all four spaces,

$$\langle x, y\rangle := \sum_i x_i y_i,$$

where x_i, y_i are the coordinates of x, y with respect to the basis, as well as corresponding norms.

Exercise 4.48. Are there other inner products that would guarantee equal treatment of the candidates?

Our interest is the 1-cochains, $C^1(K)$. Suppose $\{e_i\}$ is the set of edges of K, the chosen basis of $C_1(K)$. Then the coordinates of any 1-cochain x are given by

$$x_i = x(e_i).$$

Then,

$$\langle x, y\rangle = \sum_i x(e_i)y(e_i) = \sum_i (xy)(e_i) = (xy)\Big(\sum_i e_i\Big).$$

In the language of differential forms, the formula is seen as a line integral over the 1-chain of the whole complex:

$$\langle x, y\rangle = \int_K xy.$$

At this point, we choose our ring of coefficients to be the *reals* $R = \mathbf{R}$. Then we are in the realm of linear algebra: C^1 is a finite dimensional vector space and so is its subspace B^1. The projection of the former on the latter has a clear meaning which makes the following well-defined.

Definition 4.49. The *best rating approximation* of a comparison vote $a^1 \in C^1$ is defined to be

$$b^1 := \arg\min_{x\in B^1} ||x - a^1||.$$

Next, $(\partial^0)^{-1}(b^1)$ is an affine subspace of $C^0(K)$. The elements are linear maps over $C_0(K)$ and, since K is connected, they differ by a constant. We can say the same about $a^0 + (\partial^0)^{-1}(b^1)$ for any $a^0 \in C^0$. It follows each of them attains its maximum at the same location(s). Therefore, the following is also well-defined.

Definition 4.50. For a given vote a, a *winner of degree* 1 is a candidate with the largest value of $a^0 + b^0$, where b^0 is any 0-chain that satisfies $\partial^0 b^0 = b^1$ and b^1 is the best rating approximation of a^1.

Exercise 4.51. Define a winner of degree 2 and show how it can be used.

The construction is based on the following decomposition:

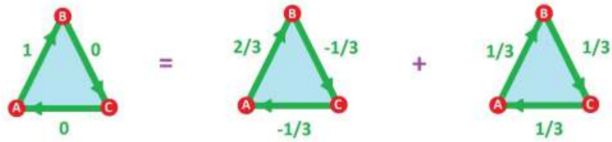
$$C^1 = B^1 \oplus (B^1)^\perp.$$

What is the meaning of $(B^1)^\perp$ in the voting context?

Example 4.52. Observe that in the above example, the difference between $a^1 = (1, 0, 0)$ and its best rating approximation is

$$a^1 - b^1 = (1, 0, 0) - \left(\tfrac{2}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}\right) = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right) = \tfrac{1}{3}(1, 1, 1).$$

Plainly, this is a vector perpendicular to the space of coboundaries and it is also a prefectly circular vote.



Thus, each comparison vote is (uniquely) decomposed into the sum of a non-circular vote and a perfectly circular vote. □

Definition 4.53. The projection $D : C^1 \to B^1$ will be called the *decycling operator* (technically, "de-co-cycling").

Exercise 4.54. Generalize the conclusion of the last example to the case of an n-simplex.

Example 4.55. Suppose we have a perfectly circular vote x, i.e., a special 1-cochain:

$$x(AB) = x(BC) = x(AC) = p \in R.$$

We recognize it as the dual of a 1-boundary:

$$x = p(\partial_2 ABC)^*.$$

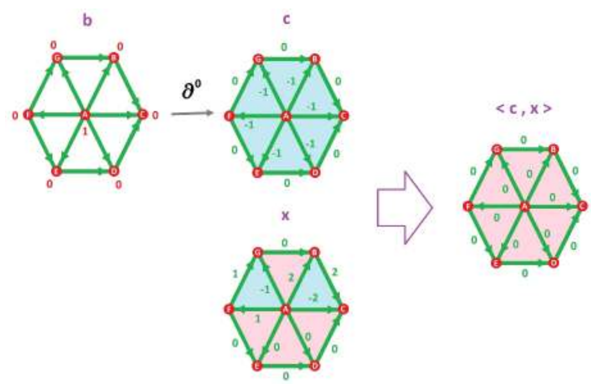
Next, let's consider the coboundary of the dual of vertex A:

$$c := \partial^0 A^* \in B^1.$$

Then, for every vertex B_i adjacent to A, we have

$$c(AB_i) = \partial^0 A^*(AB_i) = A^*(\partial_1(AB_i)) = A^*(B_i - A) = A^*(B_i) - A^*(A) = 0 - 1 = -1.$$

The rest are 0s. The inner product of such c and x is illustrated below:



It is zero! □

Theorem 4.56.

$$(B^1)^\perp = (B_1)^*.$$

Exercise 4.57. Finish the proof of the theorem.

Exercise 4.58. Find explicit formulas for the decycling operator for $n = 3$.

Corollary 4.59 (Hodge decomposition).

$$C^1 = B^1 \oplus (B_1)^*.$$

Accordingly, decycling doesn't just remove from each comparison vote something nonsensical, such as

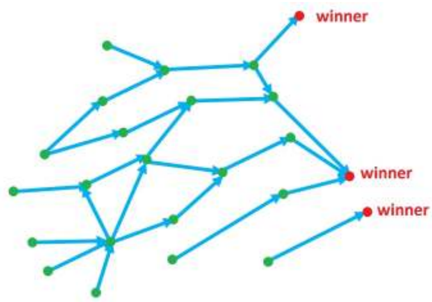
$$... < \text{ rock } < \text{ paper } < \text{ scissors } < \text{ rock } < ...,$$

it removes exactly the component that prevents this comparison from being a rating.

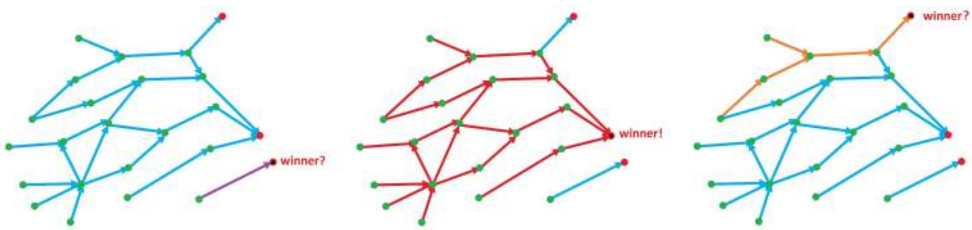
With the decycled comparison vote b^1 at hand, we proceed to maximize $(\partial^0)^{-1}(b^1)$ as explained before. Alternatively, we look at a comparison vote as a weighted directed graph, a flow. With the circular patterns removed by decycling, this is an irrotational flow. Then the solution is:

- follow this flow towards better and better candidates.

We find the best ones at its dead ends:



There may be many dead ends. If, as shown above, we ignore the edges with zero comparison votes, the graph may be disconnected. Then there will be at least one winner in each component of the graph. To choose the #1 winner, we may compare the number of voters *upstream* from each of them:



Exercise 4.60. Does this procedure produce the same winners as maximizing $(\partial^0)^{-1}(b^1)$ described before?

This is how we decycle data with a spreadsheet.

We start with a comparison vote given as an antisymmetric $n \times n$ matrix X. We need to construct a rating vote, as an n-vector R, so that $\partial^0 R$ is the closest to X.

As we have seen, there is a direct way, via linear algebra, to compute R from X. We treat the task as an optimization problem instead:

$$R := \arg\min_R ||X - \partial^0 R||.$$

To find the function to be minimized, we replace the norm with its square and then expand:

$$\Phi(R_1, ..., R_n) := \sum_{ij} \Big(X_{ij} - (R_i - R_j) \Big)^2 .$$

We will "chase" (the negative of) its gradient. The increment of R_j is then proportional to:

$$\begin{aligned} \Delta R_j \quad &:= \sum_{i \neq j} \Big(X_{ij} - (R_i - R_j) \Big) \\ &= \sum_{i \neq j} X_{ij} - \sum_{i \neq j} R_i + \sum_{i \neq j} R_j \\ &= \sum_i X_{ij} - \sum_i R_i + nR_j . \end{aligned}$$

These three terms are visible in the spreadsheet:

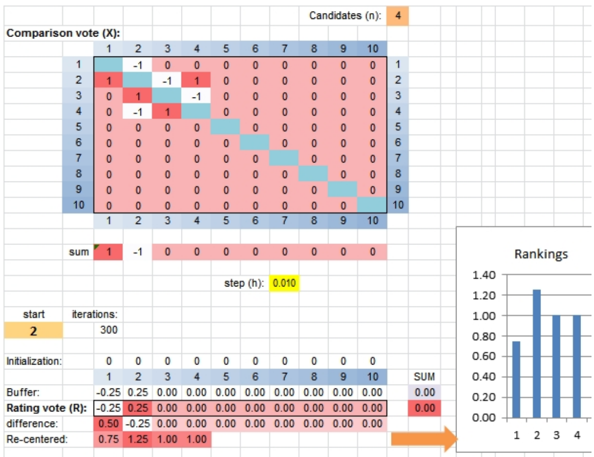
- the sum of the terms in the jth column of the matrix of the comparison vote,
- the sum of the terms of the current vector of the rating vote, and
- the jth entry of that vector times n.

We compute the iteration:

$$R_j' := R_j + h\Delta R_j,$$

where h is the parameter of the process that controls the length of its step.

The spreadsheet shown below finds the ranking of the webpages for the web discussed previously:



Note: See the files online.

Exercise 4.61. (a) Modify the formula to account for possibility of "incomparable" candidates. (b) Modify the spreadsheet accordingly.

4.9 Is there a fair electoral system?

In the face of both an impossibility and a possibility theorems, what is our conclusion? Is there a fair way to run elections?

First of all, the outcome must be a rating! Then the possible avenues presented in the diagram below have to end at the bottom right corner:

$$\begin{array}{ccccc}
\text{Election:} & & & & \text{Outcome:} \\
\text{comparison votes} & \xrightarrow{\ tally\ } & \text{a comparison vote} & \xrightarrow{\ decycling\ } & \text{a rating comparison vote} \\
\uparrow (\partial^0)^m & & & & \uparrow \partial^0 \\
\text{rating votes} & \cdots\cdots & \textit{impossible} & \cdots\cdots > & \text{a rating vote}
\end{array}$$

Let's make a few observations.

First, we cannot run a rating election, because we cannot tally it fairly according to the impossibility theorem in this section. (Moreover, if we replace ratings with rankings, there is still no fair tally according to the original Arrow's Impossibility Theorem.)

Second, we *can* run a comparison election.

Third, we can bypass the impossible direct route from rating election to the total rating: we first convert each voter's rating vote into a comparison vote, then tally the votes into a total comparison vote, and then finally convert it to a rating vote via decycling. Unfortunately, since the start and the endpoints of the route remain the same as in the first case, this scheme fails due to the commutative nature of the diagram.

Fourth, we can truly circumvent the impossibility theorem by choosing a different *starting point*: run a comparison election and then proceed to the ratings as described above. In other words, a *decycled comparison election* is implemented by the following composition of homomorphisms:

$$\begin{array}{ccccc} \text{Election:} & & & & \text{Outcome:} \\ (C^1)^m & \xrightarrow{\quad\Sigma\quad} & C^1 & \xrightarrow{\quad D\quad} & B^0 \\ & & & & \Big\downarrow (\partial^0)^{-1} \\ & & & & C^0 \end{array}$$

Is decycling the answer? Let's examine $D\Sigma$.

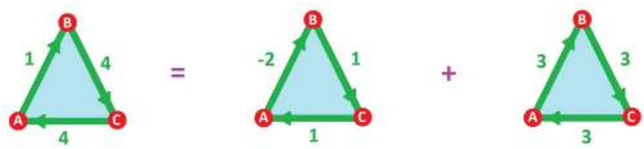
It is a comparison tally and it satisfies Additivity, Symmetry I, and Symmetry II.

What about Diagonality? Do we have $D\Sigma\delta = k\,\mathrm{Id}$? Not unless $B^1 = C^1$.

Exercise 4.62. Prove this statement.

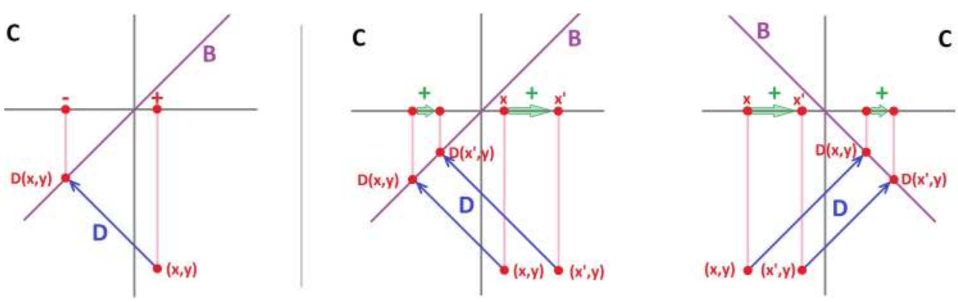
What about Positivity? Once again, not unless $B^1 = C^1$.

Example 4.63. This is how decycling replaces the vote of $A > B$ with $A < B$:



□

Algebraically, will a point with a positive first coordinate always have a projection with a positive first coordinate? Of course not: just consider the projection of the plane onto an inclined line (left):



There is, however, a *monotonic* relation (right) between these two numbers. This idea is put in the form of this simple algebraic lemma.

Lemma 4.64. Under an orthogonal projection, the dependence of any vector on itself is non-decreasing. In particular, the dependence on any coordinate on itself is non-decreasing.

Proof. Suppose we have an orthogonal projection $D : C \to C$ of a vector space C onto some subspace. The range of the orthogonal projection and its kernel are orthogonal. In other words, for every $x, y \in C$, we have

$$\langle Dx, y - Dy \rangle = 0, \text{ or } \langle Dx, y \rangle = \langle Dx, Dy \rangle.$$

If $x = y$, the equation becomes:

$$\langle Dx, x\rangle = \langle Dx, Dx\rangle = ||Dx||^2 \geq 0.$$

Now, if we apply the formula to $x = e^i$, the ith basis element of C, the result is D_{ii}, the (i, i)-entry of the matrix of D. □

In other words, an increase of a comparison vote on A vs. B will not cause a decrease of this vote after the decycling operator $D : C^1 \to C^1$. Combined with what we know about the sum tally Σ, this implies the following useful property.

Theorem 4.65 (Monotonicity). If a voter increases his comparison vote on A over B, in the outcome of the decycled comparison election, $D\Sigma$, candidate B will not improve his position relative to A.

Such a step, however, may change relative positions of *other* candidates. Let's take another look at the independence of irrelevant alternatives.

Example 4.66 (election manipulation). We consider a simple comparison vote $P \in C^1$:

$$P(AB) = 1, \ P(BC) = -1, \ P(AC) = 0.$$

It is a rating vote:

$$B > A = C.$$

Suppose this vote came – via the sum tally – from the following comparison election:

	AB	BC	CA
X	1	0	0
Y	0	-1	0
rest	0	0	0
P	1	-1	0

The winner is B...

Suppose the voters have changed their votes – but not about A vs. B:

	AB	BC	CA
X	1	0	0
Y	0	-1	0
rest	0	5	4
Q	1	4	4

Now, Q is not a rating comparison vote. As we know, it is decomposed as follows:

$$Q = (1, 4, 4) = (-2, 1, 1) + (3, 3, 3).$$

The outcome of the second election is then:

$$D(Q) = (-2, 1, 1),$$

and the winner is A!

Predictably, when enough voters change their minds about A vs. C or B vs. C, the outcome of A vs. B may change too. □

It appears that $D\Sigma$ does not produce a fair election method...

Let's, nonetheless, consider how this method applies to rating of web-pages. First, we set up the votes:

- if there is a link from page A to page B, edge AB is given a value of -1;
- if there isn't, it's 0.

The totality of these values is a comparison vote.

Example 4.67. Let's decycle this simple web presented as a comparison vote:

$$\begin{array}{ccccccccccc}
C & \xrightarrow{-1} & D & & C & \xrightarrow{0} & D & & C & \xrightarrow{0} & D \\
 & \nwarrow^{-1} & \downarrow_{-1} & \xrightarrow{D} & & \nwarrow^{0} & \downarrow_{0} & \text{OR} & & \nwarrow^{-1/4} & \downarrow_{1/4} \quad ? \\
A & \xrightarrow{-1} & B & & A & \xrightarrow{-1} & B & & A & \xrightarrow{1/2} & B
\end{array}$$

Even though removing the obvious circular vote gives us the former, the latter is the correct answer. The resulting ranking is:

$$A < C = D < B.$$

□

Exercise 4.68. Justify the answer. Hint: the diagrams are incomplete.

Exercise 4.69. What if we decycle each comparison vote before we tally?

What makes the setting of interlinked websites different is the source and the nature of the votes. In a standard electoral system, there are voters and there are candidates and the former vote for the latter. On the web, there are only pages and they vote for each other. What's crucial, each page can only vote *against itself*!

This inability to increase your own standing is preserved under decycling due to the Monotonicity Theorem. An immediate consequence is that under the decycling scheme, unlike under PageRank,

adding outbound links to your site won't increase your ranking.

This approach can be used as a foundation of an "against-self electoral system". In such a system, every voter is also a candidate but as a voter he has limited voting powers. Compare:

- the standard system:
  - ◇ voter X: candidate $A >$ candidate B;
- the against-self system:
  - ◇ voter X: candidate $A >$ candidate X.

By voting this way, voter X says: "voter A is a better choice than me to decide on this issue". There are only two comparison votes for each pair of candidates – by the candidates themselves.

Just as explained previously, with the circular patterns removed by decycling, this mode of voting creates

- a flow towards better and better voters, or, which is the same thing,
- a flow towards better and better candidates.

Finally, one can follow this flow to its end(s) to find the best candidate(s).

Exercise 4.70. Prove or disprove the "majority criterion": if a majority of voters vote $A > B$ for all B, then A wins.

Exercise 4.71. Prove or disprove the "monotonicity criterion": if A wins, he still wins even after some of the voters change their votes from $A < B$ to $A > B$.

Exercise 4.72. Prove or disprove the "participation criterion": adding a voter who votes $A > B$ to an existing election will not change the winner from A to B.

Exercise 4.73. Prove or disprove the "consistency criterion": if the voters are divided into parts and A wins in each of the separate elections, he also wins in an election of all voters.

Exercise 4.74. Prove or disprove the "independence of clones criterion": if A is a winner and $B \neq A$, then even after adding a new candidate C candidate A still wins provided there is no voter with a vote: $B \leq D \leq C$ or $C \leq D \leq B$ for any candidate $D \neq B, C$.

1 Appendix: Student's guide to proof writing

These are ten rules (or tests, or clues) that you can use to check whether there may be problems with what you are about to submit...

Once upon a time a lecture was being given:

> **Pythagorean Theorem:** $a^2 + b^2 = c^2$ for the sides of a right triangle...
> **Homework assignment:** Prove the theorem.
> .
> .
> **Law of Cosines:** $a^2 + b^2 - 2ab\cos\alpha = c^2$ for the sides of a triangle with angle α...

A week later this was submitted:

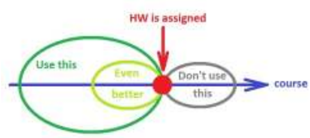
> **Homework solution:** To prove the Pythagorean Theorem, take the Law of Cosines and set $\alpha = 90$ degrees, done.

What is wrong with this picture?

The argument is circular as the proof of the Law of Cosines is (usually) based on the Pythagorean Theorem. That's why the lectures that follow the assignment should be off-limits.

Rule 1: *If you use in your proof results presented in lectures* after *the homework was assigned, this is very likely not what's expected from you.*

It may be impossible to check all the lectures and the purpose of the rule is to eliminate any chance of circular reasoning.



The version of this rule for the professor to follow: Don't assign homework until all the necessary background material has been covered.

The example provides a rationale for the following:

Rule 2: *If your proof is "The statement is just a particular case of this theorem", that's probably not what's expected from you.*

Be especially careful if this theorem comes from outside the course. Simply put, a very short proof is often a bad sign. This rule is meant to prevent you from being tempted by an easy way out. A proof from scratch – based on the concepts being discussed in the class you are taking – is expected just about every single time!

To summarize the two rules: the proof should be as *localized* as possible, but only on one side (think $(a - \varepsilon, a]$).

Rule 3: *If your proof doesn't provide the definition or quote a theorem for* each *concept used, it is probably flawed.*

The danger is that the proof is superficial and hand-wavy.

A proof that looks like an essay can probably use a lot more structure.

Rule 4: *If your proof is long but has no lemmas, it is likely to be flawed.*

It is certainly not very readable. It is not readable for the professor but for you too – that's why there is a chance it may have problems that you've overlooked.

Note: In just about any undergraduate or graduate course a one-page proof in TeX (with no illustrations) is long.

The rule isn't meant to proclaim that long is bad. Just the opposite is true: one shouldn't keep any part of the proof for himself.

Rule 5: *If you have an "easy" part of the proof in your head but don't put it on paper, this part is likely to be challenged.*

Rule 6: *If you don't introduce all or most of the objects in the proof by giving them letter names, your proof may be flawed.*

The danger is once again that the proof is superficial and hand-wavy. Always start with "Let x be..."

Note: How to choose good notation is a subject of a separate discussion.

Rule 7: *If you introduce an object in your proof and then never use it, this is a problem and might be a sign of other problems.*

At least it will leave a very bad impression...

Rule 8: *If you don't use* all *of the conditions of the theorem you are to prove, your proof is very likely to be flawed.*

You simply don't see a lot of theorems with redundant conditions.

Your drawings may fool you. In fact, draw a triangle...

...Now, take a closer look at it; you are likely to discover that your triangle is either a right triangle or an isosceles. A proof based on this picture could easily be flawed.

Rule 9: *If removing all pictures from your proof makes it incomplete or just hard to follow, it is probably not rigorous enough.*

Let illustrations illustrate...

And finally, just because your professor or your textbook violate, as they often do, some or all of these rules, don't assume that you are off the hook.

Rule 10: *If you write your proof as just a variation of a proof taken from a lecture or a book, it is likely that higher standards will be applied to yours.*

2 Appendix: Notation

Generalities

$\Longrightarrow$		"therefore"
$\Longleftrightarrow$		"if and only if"
$\forall$		"for any" or "for all"
$\exists$		"there exists" or "for some"
$A :=$ or $=: A$		"A is defined as"
$A \setminus B$	$:= \{x \in A : x \notin B\}$	the complement of B in A
A^n	$:= \{(x_1, ..., x_n) : x_i \in A\}$	the nth power of set A
$f : X \to Y$		a function from set X to set Y
$i_X : A \hookrightarrow X$	$i_X(x) := x$	the inclusion function of subset $A \subset X$ of set X into X
$\mathrm{Id}_X : X \to X$	$\mathrm{Id}_X(x) := x$	the identity function on set X
$f : x \mapsto y$		function f takes $x \in X$ to $y \in Y$; i.e., $f(x) = y$
$\#A$		the cardinality of set A
$f(A)$	$:= \{y \in Y : y = f(x), x \in A\}$	the image of subset $A \subset X$ under $f : X \to Y$
$\mathrm{Im}\, f$	$:= f(X)$	the (complete) image of $f : X \to Y$
$\mathrm{Graph}\, f$	$:= \{(x, y) : y = f(x)\} \subset X \times Y$	the graph of function $f : X \to Y$
2^X	$:= \{A \subset X\}$	the power set of set X

Basic topology

$\mathrm{Cl}(A)$		the closure of subset $A \subset X$ in X
$\mathrm{Int}(A)$		the interior of subset $A \subset X$ in X
$\mathrm{Fr}(A)$		the frontier of subset $A \subset X$ in X
$B(a, \delta)$	$:= \{u \in \mathbf{R}^n : \|\|u - a\|\| < \delta\}$	the open ball centered at $a \in \mathbf{R}^n$ of radius δ
$\bar{B}(a, \delta)$	$:= \{u \in \mathbf{R}^n : \|\|u - a\|\| \leq \delta\}$	the closed ball centered at $a \in \mathbf{R}^n$ of radius δ
$\dot{\sigma}$		the interior of cell σ
$F(X, Y)$		the set of all functions $f : X \to Y$
$C(X, Y)$		the set of all maps $f : X \to Y$
$C(X)$	$:= C(X, R)$	the set of all maps $f : X \to Y$ for a fixed R

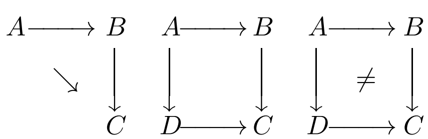
Sets

$\mathbf{R}$		the real numbers
$\mathbf{C}$		the complex numbers
$\mathbf{Q}$		the rational numbers
$\mathbf{Z}$		the integers
$\mathbf{Z}_n$	$:= \{0, 1, ..., n-1\}$	the integers modulo n
$\mathbf{R}^n$	$:= \{(x_1, ..., x_n) : x_i \in \mathbf{R}\}$	the n-dimensional Euclidean space
$\mathbb{R}^n$		the standard cubical complex representation (with unit cubes) of $\mathbf{R}^n$
$\mathbf{R}^n_+$	$:= \{(x_1, ..., x_n) : x_i \in \mathbf{R}, x_1 \geq 0\}$	the positive half-space of $\mathbf{R}^n$ (not $(\mathbf{R}_+)^n$)
$\mathbf{B}^n$	$:= \{u \in \mathbf{R}^n : \|\|u\|\| \leq 1\}$	the closed unit ball in $\mathbf{R}^n$
$\mathbf{S}^1$		the circle
$\mathbf{S}^n$		the sphere
$\mathbf{I}$	$:= [0, 1]$	the closed unit interval

$\mathbf{I}^n$		the n-cube
$\mathbf{T} = \mathbf{T}^2$		the torus
$\mathbf{T}^n$	$:= \mathbf{S}^1 \times \mathbf{S}^1 ... \times \mathbf{S}^1$	the n-torus
$\mathbf{M} = \mathbf{M}^2$		the Möbius band
$\mathbf{P} = \mathbf{P}^2$		the projective plane
$\mathbf{K} = \mathbf{K}^2$		the Klein bottle

Algebra

$\langle x, y \rangle$		an inner product
A^T		matrix A transposed
$< A \vert B >$	$\subset G$	the subgroup of group G generated by the subset $A \subset G$ with condition B
$< a >$	$:= \{na : n \in \mathbf{Z}\} \subset G$	the cyclic subgroup of group G generated by element $a \in G$
$x \sim y$		equivalence of elements with respect to equivalence relation $\sim$
$[a]$	$:= \{x \in A : x \sim a\}$	the equivalence class of $a \in A$ with respect to equivalence relation $\sim$
$A/_\sim$	$:= \{[a] : a \in A\}$	the quotient of set A with respect to equivalence relation $\sim$
$[f]$	$: A/_\sim \to B/_\sim$	the quotient function of function $f : A \to B$ with respect to these two equivalence relations; i.e., $[f]([a]) := [f(a)]$
$A \times B$	$:= \{(a, b) : a \in A, b \in B\}$	the product of sets A, B
$G \oplus H$	$:= \{(a, b) : a \in G, b \in H\}$	the direct sum of abelian groups (or vector spaces) G, H
$g \oplus h$	$: G \oplus H \to G' \oplus H'$	the direct sum of homomorphisms (or linear operators) $g : G \to G', h : H \to H'$; i.e., $(g \oplus h)(a, b) := (g(a), h(b))$
$\mathcal{S}_n$		the group of permutations of n elements
$\mathcal{A}_n$		the group of even permutations of n elements



two commutative diagrams and a non-commutative diagram

Homology

X, Y cell complexes, M, N chain complexes

$C_n = C_n(X)$		the group of n-chains of cell complex X
∂_n	$: C_n(X) \to C_{n-1}(X)$	the nth boundary operator of X
$C(X)$	$:= \{C_n(X) : n = 0, 1, 2, ...\}$,	the (total) chain group of X
$C(X)$	$:= \oplus_n C_n(X)$	
∂	$:= \{\partial_n : C_n(X) \to C_{n-1}(X)\}$	the (total) boundary operator of X
$C(X)$	$:= \{C(X), \partial\}$	the chain complex of X (i.e., $\partial : C(X) \to C(X)$)
f_n	$: C_n(X) \to C_n(Y)$	the nth chain map of map $f : X \to Y$
f_Δ	$:= \{f_n : n = 0, 1, 2, ...\}$	the (total) chain map $: C(X) \to C(Y)$

f_Δ	$:= \oplus_n f_n$	of map $f : X \to Y$
$Z_n = Z_n(X)$	$:= \ker \partial_n$	the group of n-cycles of X, or
$Z_n(M)$	$:= \ker \partial_n$	the group of n-cycles of M
$B_n = B_n(X)$	$:= \operatorname{Im} \partial_{n+1}$	the group of n-boundaries of X, or
$B_n(M)$	$:= \operatorname{Im} \partial_{n+1}$	the group of n-boundaries of M
$H_n = H_n(X)$	$:= Z_n(X)/B_n(X)$	the nth homology group of X, or
$H_n(M)$	$:= Z_n(M)/B_n(M)$	the nth homology group of M
$H(X)$	$:= \{H_n(X) : n = 0, 1, 2, ...\}$,	the (total) homology group of X
$H(X)$	$:= \oplus_n H_n(X)$	
$H(M)$	$:= \oplus_n H_n(M)$	the (total) homology group of M
$\beta_n(X)$	$:= \dim H_n(X)$	the nth Betti number of X, or
$\beta_n(M)$	$:= \dim H_n(X)$	the nth Betti number of M
$[f_n]$	$:= [f_\Delta] : H_n(X) \to H_n(Y)$	the nth homology map of map $f : X \to Y$, or
$[g_n]$	$: H_n(M) \to H_n(N)$	the nth homology map of the chain map $g = \{g_n : M_n \to N_n,\ n = 0, 1, 2, ...\}$
f_*	$:= \{[f_n] : n = 0, 1, 2, ...\}$	the (total) homology map $: H(X) \to H(Y)$
f_*	$:= \oplus_n [f_n]$	of map $f : X \to Y$
g_*	$:= \{[g_n] : n = 0, 1, 2, ...\}$	the (total) homology map $: H(M) \to H(N)$ of the chain map $g : M \to N$
	Other algebraic topology	
$X \sqcup Y$		the disjoint union of X, Y
$X \vee Y$	$:= \Big(X \sqcup Y\Big)/\{a \sim b\}$	the one-point union of spaces X, Y ($a \in X, b \in Y$)
ΣX	$:= [0,1] \times X/_{\{(0,x)\sim(0,y),(1,x)\sim(1,y)\}}$	the suspension of X
$\chi(X)$		the Euler characteristic of X
$[X, Y]$		the set of all homotopy classes of maps $X \to Y$
$\pi_1(X)$		the fundamental group of X
$\|K\|$		the realization of complex K
$\mathrm{St}_A(K)$	$:= \{\sigma \in K : A \in \sigma\} \subset K$	the star of vertex A in complex K
$T_A(K)$		the tangent space of vertex A in complex K
$T(K)$		the tangent bundle of complex K
$K^{(n)}$	$:= \{s \in K : \dim a \leq n\}$	the nth skeleton of complex K
$A_0A_1...A_n$		the n-simplex with vertices $A_0, A_1, ..., A_n$
$G \cong H$		isomorphism
$X \approx Y$		homeomorphism
$X \simeq Y$		homotopy equivalence
$f \simeq g$	$: X \to Y$	homotopy of maps

Index