APPLIED TOPOLOGY LECTURE NOTES

CHAD GIUSTI

UPDATED: NOVEMBER 13, 2017

Contents

Author's Note

This is a collection of course notes for an undergraduate course on applied algebraic topology. The goal is a rigorous survey of the fundamental notions in modern topology, providing both exposure to a collection of central ideas in classical algebraic topology and a foundation for the use of homological algebra and the philosophy of algebraic topology as tools in science and engineering. In its current form, the presentation assumes no exposure to topology, but expects that students have courses in linear algebra and multivariable calculus, basic exposure to sets and combinatorics, and some proficiency reading and writing proofs and computer code.

These notes are still in early development. Please excuse the glaring errors, general lack of organization, and the fact that most of the material and all of the images and citations are absent or underdeveloped. If something looks suspicious, it's probably wrong; if it doesn't look suspicious, it probably should. This state of affairs will eventually improve. Many of the deeper proofs are omitted as they were not presented in the course as taught, but all proofs that are not intended to be exercises will be added to a later version of the notes.

Questions and comments are always welcome.

Introduction: Thinking Topologically

WHAT IS TOPOLOGY? Here, we will take the slightly unorthodox (but useful) view that *topology* is the mathematics of qualitative notions of similarity. Possibly the simplest example of such a notion is a fixed *set* of objects with a binary notion of pairwise similarity: for any pair of objects in the set, the pair is either similar or dissimilar. This type of data is exceedingly common and is called a *graph*. However, before we dive into graphs, we need to pause and "recall" some fundamental notions about sets; we will use this review to highlight in a familiar context some ideas that will recurr throughout our investigation.

Sets

Sets are elemental objects in mathematics[1], with properties laid out by a system of axioms which we will not concern ourselves with here. Rather, we will settle for the standard, informal notion that sets are well-defined collections of unique elements[2].

Once we have the notion of sets, many natural questions arise. Foremost among them, and of strong interest to us moving forward, is: how do we know if two sets S and T are "the same"? Here, we are working in the absence of any further structure or semantics, so all we can do is think about what elements are in the two sets. Clearly, if the elements of S and T identical, we should consider them the same, and we write $S = T$. However, this notion is too rigid for most applications: simply renaming the elements is sufficient to force us to call two sets different if we rely on equality. We will require a more abstract notion of equivalence, under which we set up an abstract correspondence between elements of the two sets.

Before we move forward with this plan, it's good to perform a sanity check: under a reasonable notion of sameness, each set S should be the same as itself, so each element in S should uniquely correspond to some element in S. Without any further information, the only sensible choice is for each $s \in S$ to correspond to itself. Such a correspondence is recorded by the *identity function on* S, written $\mathrm{id}_S : S \to S$ and given by $\mathrm{id}_S(s) = s$.

To proceed by analogy to this example, if two sets S and T are to be considered the same, we require a function f which assigns to each $s \in S$ a unique corresponding element $f(s) \in T$; that is, an *injective* function $f : S \to T$. The existence of an injective function f tells us that there are sufficiently many elements in T to account for all elements of S, so we say that the *cardinality* of S is no greater than that of T, written $|S| \leq |T|$. However, on its own this is clearly in-

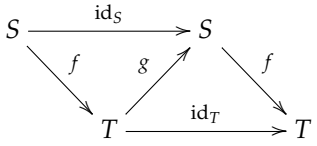
[1] Though there is an alternative modern formulation of the foundation of mathematics called *topos theory* which is fundamentally topological in nature.

[2] That is, a *set* is an unordered collections of *elements*, we can determine if a given element is included in the set or not, and that each element which is in the set appears precisely once.

sufficient as there may be elements of T not in the image $\mathrm{im}(f) \subset T$. One might attempt to remedy this issue by requiring a second injective function $g : T \to S$, thus ensuring that the opposite inequality $|T| \leq |S|$. This shows that the sets have the number of elements so implies that the sets the same abstract structure, but fails to explicitly capture a correspondence between the elements, For that, we require a *compatibility condition* between f and g: $g(f(s)) = t$ and $f(g(t)) = t$.

Definition. Let S, T be sets. An *set isomorphism between S and T* is a function $f : S \to T$ for which there exists a *two-sided inverse* function $g : T \to S$ with $(g \circ f) = \mathrm{id}_S$ and $(f \circ g) = \mathrm{id}_T$. We say S is *(set) isomorphic* to T and write $S \cong T$.

When we are attempting to understand relationships between several sets or functions, we will draw schematics called *diagrams*. The maps and sets in the definition of an isomorphism fit together as follows.



The requirement that $(g \circ f) = \mathrm{id}_S$ is equivalent to saying that the process of following either path in the diagram from S to itself, composing functions as you go, will result in the same function. If this holds for all possible paths in a diagram, we say the diagram *commutes*. Thus, we can say that f is an isomorphism between S and T if there is a function $g : T \to S$ so that the above diagram commutes.

You might have been expecting us to simply require that f also be surjective. It turns out that this is an equivalent notion.

Lemma 1. *Let S and T be sets. A function $f : S \to T$ is an isomorphism if and only if it is a bijection.*

Proof. Exercise. □

The existence of a set isomorphism between two sets is a *certificate* that the sets are the same; if such a map doesn't exist, then the sets aren't the same. In general, finding a certificate is easier than proving one doesn't exist, but as the objects of study grow in complexity, even the former task becomes formidable. One of our primary goals in developing the tools of algebraic topology is the construction of understandable and computable certificates of sameness for objects of interest.

It bears repeating that this notion of isomorphism is abstract: any data besides the fact that the sets contain matched pairs of elements

is lost. As we move forward, we will need to be carefully extend this notion of isomorphism to that the other structures we build on top of our sets are maintained.

We close our discussion of sets by fixing some notation that will get a lot of use.

Definition. Let S, T be a sets.

- The *cartesian product* of S and T is the set of ordered pairs of elements $S \times T = \{(s,t) \mid s \in S, t \in T\}$. The set of *ordered k-tuples of elements of S* is written $S^{\times k} = S \times S \times \cdots \times S$, and elements are written $(s_1, s_2, \ldots, s_k)$.
- The set of *unordered k-element subsets of S* is $\binom{S}{k}$. Denote by $s_1 s_2 \ldots s_k$ (or any permutation thereof) the element $\{s_1, s_2, \ldots s_k\}$. The *power set of S* is the collection of all subsets of S, written 2^S.[3]
- The *disjoint untion* of S and T is the set obtained by adding an index to the elements in each set indicating the set of its origin (written as subscripts), so elements in S and T are necessarily disjoint, and taking the standard union. Formally,

$$S \sqcup T = \{(S \times \{S\}) \cup (T \times \{T\})\}.$$

[3] This notation is fantastic because $2^{|S|} = \sum_{k=0}^{|S|} \binom{|S|}{k}$, and $2^S = \bigcup_{k=0}^{|S|} \binom{S}{k}$.

For example, if $S = \{0,1,2\}$ and $T = \{1,2,3\}$, then $S \cup T = \{0,1,2,3\}$, while

$$\begin{aligned} S \sqcup T &= \{(0,S),(1,S),(2,S),(1,T),(2,T),(3,T)\} \\ &= \{0_S, 1_S, 2_S, 1_T, 2_T, 3_T\}. \end{aligned}$$

Graphs

Now we can return our attention to the notion of a graph, which is a set of objects along with a binary notion of similarity or connection between any pair of those objects. We record the objects and the similarity measure as sets, where elements of the latter carry intrinsic meaning in terms of the former.

Definition. An *(abstract, simple) graph* Γ is a pair of sets $\Gamma = (V, E)$ with $E \subseteq \binom{V}{2}$. We say V is the set of *vertices*, and E is the set of *edges*.

Throughout these notes, unless explicitly stated otherwise, all graphs are simple and *finite*; that is, $|V| < \infty$.

Conceptually, the vertices $v \in V$ are the objects under consideration, and the presence of an edge $vv' \in E$ indicates that v and v' are similar. Without any further information, this is a very general notion of similarity: knowing that $v_1 v_2 \in E$ and $v_2 v_3 \in E$ tells us nothing about whether v_1 is similar to v_3.

Here are some real-world examples.

Examples.

- $V =$ people, $E =$ self-identified friends (social media)
- $V =$ locations, $E =$ connecting paths (map software)
- $V =$ products, $E =$ purchased together (online shopping)

These are too big to think about directly. In order to build intuition, we should focus on small examples. Of course, even trying to enumerate small examples will bring up an important question: have we got them all?

Examples.

- $V = \varnothing$, $E = \varnothing$ (the empty graph)
- $V = \{p\}$, $E = \varnothing$
  Is this the only graph with one vertex? Under what notion of "sameness" might it fail to be unique?
- a) $V_1 = \{p, q\}$, $E_1 = \varnothing$
  b) $V_2 = \{p, q\}$, $E_2 = \{pq\}$
  These are distinct graphs on the same vertex set $V_1 = V_2$ under any sensible notion of sameness. How can we tell them apart? The easiest way is to observe that there can be no isomorphism between E_1 and E_2. Is this a complete list of graphs on two vertices?
- a) $V_1 = \{p, q, r, s\}$, $E_1 = \{pq, ps, qr, rs\}$
  b) $V_2 = \{w, x, y, z\}$, $E_2 = \{wx, wz, wy, yz\}$
  c) $V_3 = \{i, j, k, \ell\}$, $E_3 = \{ij, ik, i\ell, k\ell\}$
  Are these the same? $|V_1| = |V_2| = |V_3|$ and $|E_1| = |E_2| = |E_3|$, so we can't rule out the possibility just by counting elements[4].

[4] Indeed, it is perfectly valid to take $|V_1| = |V_2|$ and $|E_1| = |E_2|$ to be a notion of sameness between graphs, in which case tese three graphs are "the same".

Let's think a bit about the structure we've built. Graphs consist of two sets, V and E, but the elements of the set E have a meaning endowed by V; each edge corresponds to a choice of two vertices. Thus, if the vertices of one graph can be assigned a correspondence to those in another, we automatically know what should happen to the edges.

Definition. Let $\Gamma_1 = (V_1, E_1)$ and $\Gamma_2 = (V_2, E_2)$ be graphs. A function $\varphi : V_1 \to V_2$ *induces* a relation[5] $\widetilde{\varphi} : E_1 \to E_2$ given by $\widetilde{\varphi}(vv') = \varphi(v)\varphi(v')$ whenever $vv' \in E_1$ and $\varphi(v)\varphi(v') \in E_2$.

[5] Recall that a *relation* R between sets S and T is just a collection of ordered pairs $R \subseteq S \times T$.

Note that this induced relation is not necessarily a function: if $vv' \in E_1$ but $\varphi(v)\varphi(v') \notin E_2$, the relation $\widetilde{\varphi}$ is undefined on vv'. However, if the relation *is* defined for every element of E_1, the induced relation is a function, and this pair is our notion of "structure preserving function" or *homomorphism* between graphs.

Definition. Let $\Gamma_1 = (V_1, E_1)$ and $\Gamma_2 = (V_2, E_2)$ be graphs, and $\varphi : V_1 \to V_2$ a function for which the induced relation $\widetilde{\varphi}$ is a function. The the pair of functions $(\varphi, \widetilde{\varphi})$ is a *(graph) homomorphism*, written $\Phi : \Gamma_1 \to \Gamma_2$.

So, when are two graphs "the same"? When the underlying vertices can be assigned a correspondence in such a way that the induced map on the edges is also a correspondence.

Definition. Let $\Gamma_1 = (V_1, E_1)$ and $\Gamma_2 = (V_2, E_2)$ be graphs. If there exists a homomorphism $\Phi : \Gamma_1 \to \Gamma_2$ so that $\varphi : V_1 \to V_2$ and $\widetilde{\varphi} : E_1 \to E_2$ are both bijections, then Γ_1 is *(graph) isomorphic* to Γ_2, written $\Gamma_1 \cong \Gamma_2$.

Graph isomorphisms preserve all of the structure inherent in graphs. The collection of all graphs isomorphic to a fixed graph Γ is called its *isomorphism class*, and except when considering specific graphs, it is common parlance to say "the graph Γ" meaning "the graph isomorphism class of Γ".
Back to our last example:

a) $V_1 = \{p, q, r, s\}$, $E_1 = \{pq, ps, qr, rs\}$
b) $V_2 = \{w, x, y, z\}$, $E_2 = \{wx, wz, xy, yz\}$
c) $V_3 = \{i, j, k, \ell\}$, $E_3 = \{ij, ik, i\ell, k\ell\}$

Suppose we want to try to construct an isomorphism from Γ_1 to Γ_2. There are $4! = 24$ possible bijections[6] $\varphi : V_1 \to V_2$. Some of these induce functions $\widetilde{\varphi} : E_1 \to E_2$ and others don't. If we choose a target for p, say $\varphi(p) = x$, then our further choices are constrained as p is implicated in two edges, pq and ps. That tells us that $\varphi(p)\varphi(q) = x\varphi(q)$ and $\varphi(p)\varphi(s) = x\varphi(s)$ must be elements of E_2. Since the edges in E_2 containing x are wx and wz, we must have either $\varphi(q) = x$ or z and $\varphi(s)$ assigned to the other. Either choice forces $\varphi(r) = y$, and the resulting function induces a graph isomorphism.

However, there are no graph isomorphisms between the first (or second) and third graphs. Why? Whatever vertex from V_1 gets sent to vertex i must be implicated in three edges in order for the induced map on edges to be surjective. However, there are no such vertices in the first graph, so there is no choice of bijection of vertices which induces a graph isomorphism.

[6] There are $n!$ possible bijections between two sets with n elements, so, if we have nothing else to go on, the problem of finding a graph isomorphism is extremely difficult. In computer science parlance, it is in *class NP*.

Definition. Let $\Gamma = (V, E)$ be a graph and $v \in V$. The *degree* of v is $||v|| = |\{e \in E \mid v \in e\}|$, the number of edges $e \in E$ which contain v. The *degree sequence* of a Γ is the list of the degrees of all vertices of Γ, sorted in decreasing order.

In order, the degree sequences of our graphs above are $(2, 2, 2, 2)$, $(2, 2, 2, 2)$, and $(3, 2, 2, 1)$.

Lemma 2. *Let Γ_1 and Γ_2 be graphs with degree sequences (d_i) and (d'_i) respectively. If $\Gamma_1 \cong \Gamma_2$, then $(d_i) = (d'_i)$.*

Proof. Exercise. □

The contrapositive of this statement is that the degree seqeunce of a graph is an *isomorphism invariant*: if two graphs have different degree sequences, they cannot be isomorphic. However, this is not a *complete* isomorphism invariant: there are graphs with the same degree sequence which are not isomorphic.

Example.

a) $V_1 = \{p, q, r, s, t, u\}$, $E_1 = \{pq, pr, qr, st, su, tu\}$

b) $V_2 = \{a, b, c, d, e, f\}$, $E_2 = \{ab, ac, bd, ce, df, ef\}$

All vertices in both graphs are of degree 2, but we claim that these graphs are not isomorphic. To check blindly, we'd have to exhaust all $6! = 720$ possible bijections between V_1 and V_2, which will be a little tedious (or require a computer). However, there are other interesting properties of graphs that are isomorphism invariants.

Definition. The *line graph of length n* is the graph

$$L_n = (\{v_1, \ldots v_n, v_{n+1}\}, \{v_1v_2, v_2v_3, \ldots, v_nv_{n+1}\}).$$

Here, "length" refers to the number of edges in the graph.

Definition. Let $\Gamma = (V, E)$ be a graph with $a, b \in V$. A *path of length n from a to b* is a homomorphism $P : L_n \to \Gamma$ so that $p(v_1) = a$ and $p(v_{n+1}) = b$. There is a *path from a to b in* Γ if there is a path of length n from a to b for some n. A path P is *simple* if it is injective on vertices. A *cycle* is a path P of length at least 3 which is injective on vertices except that $p(v_1) = p(v_{n+1})$.

Observe that this definition coincides with the "regular" definition of a path in a graph being a sequence of vertices with consecutive vertices joined by edges. However, it makes use of a *reference object*, the line graph, which is simple to understand. Exploring the existence of maps out of or into reference objects are one of the fundamental ways we will deconstruct complicated objects we encounter.

Definition. There is a unique partition of V into subsets $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ so that for every i and pair of vertices $v, v' \in C_i$, there is a path from v to v', but for any $v \in C_i$ and $v' \in C_j$, $i \neq j$, there is no path from v to v'. The sets $\mathcal{C}$ are called the *(path) components* of Γ, and a graph with exactly one path component is called *(path) connected*.

Definition. Let $\Gamma = (V, E)$ be a graph. Let $V' \subseteq V$ and $E' \subseteq E \bigcap \binom{V'}{2}$. The graph $\Gamma' = (V', E')$ is called a *subgraph* of Γ. If $E' = E \bigcap \binom{V'}{2}$, we say Γ' is the subgraph of Γ *induced* by V'.

Lemma 3. *Let Γ_1 and Γ_2 be graphs with path components $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$ and $\mathcal{D} = \{D_1, D_2, \ldots, D_m\}$ respectively. If $\Gamma_1 \cong \Gamma_2$, then $n = m$. Further, there is a bijection $\psi : \mathcal{C} \to \mathcal{D}$ so that, for each $i = 1, \ldots, n$, the subgraph of Γ_1 induced by C_i is isomorphic to the subgraph of Γ_2 induced by $\psi(C_i)$.*

Proof. Exercise. □

In light of this observation, it is common practice to restrict attention to the study of path connected graphs, as we can always decompose a graph into path-connected components, which essentially do not interact.

How many path components do the graphs in our example have? Observe that in E_1, there is no path from p to t: from p one can reach q or r, and each of these has an edge to the other, but there are no edges to any of the other nodes. Thus, $\{p, q, r\}$ is a path component, and similarly so is $\{s, t, u\}$. On the other hand, the second graph is path connected. Thus, by Lemma 3, the two graphs aren't isomorphic.

Before we move on, we'll add another reference object to our bag.

Definition. Let $n > 2$. The *cycle graph of length n* is

$$C_n = \{(v_1, \ldots v_n), \{v_1v_2, v_1v_n, v_2v_3, \ldots, v_{n-1}v_n\}).$$

A *cycle of length n in Γ* is a homomorphism $\sigma : C_n \to \Gamma$ which is injective on vertices.

Quick exercise: check that the two definitions of "cycle" are equivalent. Cycles and analogous structures are going to be fundamental tools as we move forward.

Embedded graphs

By now, those who are familiar with graphs are likely getting a little grumpy about the lack of pictures, and may have been serriptitiously drawing them while thinking about the examples. The reason we've been avoiding them is that drawing a picture of a graph isn't free: in doing so, we're assigning geometry to the purely combinatorial information in the graph, and that involves making choices. There will be a great many different pictures we can draw to represent a single graph, and each might influence our thinking about it in subtle ways. When dealing with data, we should always be acutely aware of what's actually given and what we've added, since any inferrences we make based on the latter are dangerous unless we understand exactly what we've chosen and how it affects the outcome.

Such geometry is often very useful, however. A "picture" of a graph, at least in the most familiar form, is an *embedding* (called a

realization) of the graph in Euclidean space: a collection of points representing the vertices, and continuous arcs connecting them representing the edges.

Definition. Let $\Gamma = (V, E)$ be a graph. A *(geometric) realization* ρ of Γ in $\mathbb{R}^d$ is a function $\rho_V : V \to \mathbb{R}^d$ along with a collection of continuous functions $\{\rho_{vv'} : I \to \mathbb{R}^d\}_{vv' \in E}$ so that

1. either $\rho_{vv'}(0) = \rho_V(v)$ and $\rho_{vv'}(1) = \rho_V(v')$ or vice versa, for all $vv' \in E$,
2. $\rho_{vv}(t_1) \neq \rho_{vv'}(t_2)$ for all $t_1 \neq t_2$ and $vv' \in E$, and
3. $\rho_{vv'}((0,1)) \cap \rho_{ww'}((0,1)) = \varnothing$ for all $vv' \neq ww' \in E$.

This definition immediately suggests a question: can every graph be realized in every dimension? It's quite easy to convince oneself that in $d = 0$ or $d = 1$, there are very few graphs with realizations – a good exercise is to characterize those that do.

On the other extreme, if $d = 3$ then every graph has a realization. To see this, map the vertices into $\mathbb{R}^3$ so that no three lie on a line, and no four lie in a plane, and use straight line segments for each edge. The former condition prevents edges from coinciding, while the latter keeps them from crossing at a point in their interiors. If you happen to have a collection of vertices for which these conditions don't hold, you can "repair" it by perturbing the collection of vertices slightly in almost any direction. This perturbation property tells us that every non-contrived choice of vertices will work to embed any graph in $\mathbb{R}^3$ with straight line segments as edges. Thus, we say that sets vertices which satisfy both the non-colinearity and non-planarity conditions are in *general position*. Above $d = 3$ the same game works.

That just leaves $d = 2$, the plane, where we can't play this trick: if we take the vertices of a convex quadrilateral, the diagonals cross and can't just be nudged out of doing so. There's hope, though, since we can take that diagonal move it elsewhere so the edges don't cross, or maybe move the vertices so they don't form a convex quadrilateral. However, it's definitely unclear that we can always find a good embedding. Since we're looking for counterexamples, notice that adding edges can only make it harder to draw a proper realization, so we can start by restricting our attention to graphs with all possible edges.

Definition. Let $V = \{v_1, v_2, \ldots v_n\}$. The *complete graph on n vertices* is $K_n = (V, \binom{V}{2})$.

It appears that we have the following:

Proposition 4. *K_5 does not admit a realization in $\mathbb{R}^2$.*

If it's true, there are graphs which cannot be embedded in $\mathbb{R}^2$. This being a math course, that calls for the creation of an adjective.

Definition. If there is a realization of a graph Γ in $\mathbb{R}^2$, we say Γ is *planar*.

How can we prove that K_5 isn't planar? Let's think a bit about the properties of such graphs to get a feel for them. Clearly we can restrict our attention to path connected graphs, since a graph with multiple components is planar only if all of its constituent pieces are.

The Euler characteristic

Now that we've made the decision to draw pictures, let's draw a bunch. We also got some mileage earlier out of counting things, so let's keep track of things we can count as we go.

A pattern pops out very quickly: if there are no "loops" in Γ, there is one more vertex than there are edges.

Definition. A graph Γ which is connected and contains no cycles is a *tree*.

Lemma 5. *Let Γ be a graph. If two vertices in Γ are connected by a path π, there is a simple path containing only edges from π which connects them.*

Proof. Exercise. □

Lemma 6. *Let Γ be a tree. Any two vertices in Γ are connected by a unique simple path.*

Proof. Exercise. □

Lemma 7. *If $\Gamma = (V, E)$ is a non-empty tree then $|V| - |E| = 1$.*

Proof. We proceed by strong induction on the number of vertices. Our inductive base will be the unique trees with one vertex and no edges, and two vertices and one edge, for which $|V| - |E| = 1$.

Now, assume that for $0 < k < n$, all trees with k vertices have $(k-1)$ edges. Suppose $\Gamma = (V, E)$ is a tree with n vertices. Select an edge $vv' \in E$. By Lemma 6, there is a unique simple path from v to v' in Γ which must be the edge vv'. Further, by Lemma 5, any other path connecting v to v' must include this edge, so the subgraph $\Gamma' = (V, E \setminus vv')$ has two path components C_1 and C_2, k_1 and k_2 vertices respectively, $k_1 + k_2 = n$. Because no subgraph of Γ contains a cycle, the induced subgraphs on C_1 and C_2 are each trees, so by our inductive hypothesis, C_1 has $(k_1 - 1)$ edges and C_2 has $(k_2 - 1)$ edges. Thus, Γ has $(k_1 - 1) + (k_2 - 1) + 1 = n - 1$ edges, as required. □

What about graphs that aren't trees? How do cycles interact with the plane? From the picture, it looks like they divide it up into pieces.

Theorem 8 (Jordan curve theorem)**.** *Let $\gamma : [0,1] \to \mathbb{R}^2$ be a continuous function which is injective on $[0,1)$ and with $\gamma(0) = \gamma(1)$, whose image is called a* Jordan curve*. The complement of the Jordan curve, $\mathbb{R}^2 \setminus \gamma([0,1])$, consists of two path components: a bounded region and an unbounded region which share as their boundary the Jordan curve.*

Perhaps surprisingly, this is not trivial to prove, even if we had all the terminology[7]. Right now, we'll take it at face value – it really *looks* true – but we'll come around and prove a stronger version of it later in the course. Here, we rephrase it to obtain the following useful fact.

[7] What's a path component of a subset of $\mathbb{R}^2$? Just the same as in a graph: a maximial collection of points which can be pairwise joined by continuous paths. We'll give formal definitions once we've got the right terminiology.

Corollary 9. *Let Γ be a planar graph with a realization ρ in $\mathbb{R}^2$. The complement of the image under ρ of a cycle in Γ consists of two path components.*

Right now, there are different kinds of path components: bounded, and unbounded. This sort of heterogeneity always complicates terminology and forces us to consider special cases, so we like to remove them whenever possible. Today we're going to use a bit of geometry to line things up just so. If we were doing the same thing two weeks from now, we would choose a much cleaner (and more topological) *one-point compactification* to do the same job.

Definition (Stereographic projection)**.** Map $\mathbb{R}^2$ into $\mathbb{R}^3$ by $(x,y) \mapsto (x,y,0)$, let $S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid \sqrt{x^2+y^2+z^2} = 1\}$, and take $P = (0,0,1)$. Define the *stereographic projection* of a point $p \in S^2$ to be the point of intersection between the line through P and p and the (x,y)-plane.

This map is a bijection from the sphere minus the north pole to points in the plane. As such, it provides a way to make any realization of a graph in $\mathbb{R}^2$ into a realization in S^2. Any realization in S^2 can be perturbed to miss the north pole, so any graph realization in S^2 can be turned into a realization in $\mathbb{R}^2$. Thus, *2-spherical* graphs are exactly the same as planar graphs[8].

[8] Of course, it's hard to draw things on the sphere, so we'll also use this fact to allow us to work in the plane when drawing pictures, keeping in mind that we have a bit of extra flexibility when moving things around on the sphere.

Why do we care? When we move to the sphere, there are no more unbounded components – the unbounded component just becomes the region containing the north pole.

Definition. Let Γ be a connected planar graph and ρ a realization of Γ in S^2. The connected components of the complement of the spherical realization are called the *faces* of the realization, and the collection of all such is denoted F.

Now the relationship is clearer: we can add an edge either by introducing a new vertex or not. If we add a vertex, we don't change the number of faces. If we don't, we introduce a new Jordan curve and up the number of faces by one. Thus, there is a correspondence

between (a choice of) a subset of edges in Γ and the faces in the realization.

Lemma 10. *Let $\Gamma = (V, E)$ be a connected graph. There is a tree $\tau(\Gamma) = (V, E' \subseteq E)$ which is a subgraph of Γ, called a* spanning tree for Γ.

Proof. Suppose Γ is a connected graph with $n > 0$ cycles. Select one, say $(vv_1, v_1v_2, \ldots, v_kv)$. Removing the edge vv_1 does not disconnect the graph, since the remaining edges of the cycle form an alternative path from v_1 to v, so any path that contained the edge can be altered to contain this sequence of edges instead. The subgraph $\Gamma' = (V, E \setminus vv_1)$ has fewer than n cycles, and so iterating this process a finite number of times must result in a tree, as required. □

This proof says that building a spanning tree precisely requires removing a set of edges that correspond to cycles. Cycles correspond to faces through Jordan curves, which cut existing faces in two pieces – and when we remove an edge that kills a cycle, we're "recombining" two faces. Let's put that together formally.

Definition. Given a planar graph Γ with realization ρ in S^2, choose a spanning tree $\tau(\Gamma) = (V, E')$ of Γ. Define the *spherical dual* graph $\tau^*(\Gamma, \rho) = (F, \widetilde{E})$ with vertices corresponding to the faces of the realization and an edge between two faces if they share a boundary edge that was removed in creating $\tau(\Gamma)$.

Lemma 11. *Let Γ be a planar graph with realization ρ in S^2 and spanning tree $\tau(\Gamma)$. The spherical dual $\tau^*_\rho(\Gamma)$ is a tree.*

Proof. Exercise. □

So, given a planar graph $\Gamma = (V, E)$, we can choose a realization ρ in S^2 and build a spanning tree $\tau(\Gamma) = (V, E')$ and its spherical dual tree $\tau^*(\Gamma, \rho) = (F, \widetilde{E} \cong E \setminus E')$. Lemma 7 then tells us that $|V| - |E'| = 1$ and that $|F| - |\widetilde{E}| = |F| - (|E| - |E'|) = 1$. Summing, we get

Theorem 12. *Let Γ be a connected planar graph with realization ρ in S^2. Then $|V| - |E| + |F| = 2$.*

This alternating sum is called the *Euler characteristic* of the realization. It will be a while before we encounter it again, but in the interim, it will provide us with the leverage we need to prove that K_5 is non-planar. However, if it isn't planar, we can't have a realization and thus can't possibly be counting faces, so we need to get rid of that F.

Corollary 13. *Let Γ be a connected planar graph with $|V| \geq 3$. Then $|E| \leq 3|V| - 6$.*

Proof. Exercise. □

Finally, we obtain

Corollary 14. *K_5 does not admit a realization in $\mathbb{R}^2$.*

Proof. K_5 has 5 vertices and $\binom{5}{2} = 10$ edges, and $10 \not\leq 3(5) - 6$. By Corollary 13, K_5 cannot be planar. □

Before we declare our mission accomplished, let's revisit our decision to map graphs to the sphere. What Theorem 12 appears to say is that any (nice) way we chop up the sphere into regions, we have some combinatorial relationship appearing. This association of the number 2 with the sphere is a very, very deep rabbit hole.

But we made a choice there – there are a lot of different shapes we can map these graphs onto. In particular, spherical polyhedra are the class of polyhedra that can be obtained by "flattening" the faces of a graph embedded on a sphere, so they all have this property. Another favorite of mine is the soccer ball – twelve pentagons and twenty hexagons. It's a weird coincidence that there are twelve pentagons on both of the objects with pentagons, isn't it?

Theorem 15. *Suppose $\mathcal{P}$ is a spherical polyhedron with only regular pentagonal and regular hexagonal faces. Then $\mathcal{P}$ has 12 pentagonal faces.*

Proof. Suppose there are P pentagons and H hexagons, so $|F| = P + H$. Each pentagon has five edges and vertices, and each hexagon has six. However, there are two faces sharing each edge, so we have $2|E| = 5P + 6H$. Finally, each vertex appears in three faces[9], so $3|V| = 5P + 6H$. Putting this all together, we get

$$2 = |V| - |E| + |F| = \frac{5}{3}P + 2H - \frac{5}{2}P - 3H + P + H = \frac{P}{6}.$$

So, $P = 12$. □

[9] Some geometry required. We don't *really* need the faces to be regular polyhedra, we just need three to meet at each vertex, which is guaranteed for the regular ones.

Topological spaces and continuous functions

Our goal is to build a general framework for understanding and comparing the intrinsic structure of objects. We will rely on intuition from Euclidean space as a roadmap.

Topological spaces

In $\mathbb{R}^d$, we (usually) measure distance using the *Euclidean metric*[10]. Points x and y are "close together" if the distance between them is small. How small? That depends on a notion of *scale*, which depends

[10] What is to follow will work in any metric space, though we don't have time to get into the details.

on context. As mathematicians, we like to use $\epsilon > 0$, so y is *ϵ-close* to x if $d(x, y) < \epsilon$ To conform with our plan to work with non-quantitative structures, though, what we really want is a check we can make without looking at distance directly.

Definition. Let $\vec{x} \in \mathbb{R}^d$ and $\epsilon > 0$. The *open ϵ-ball around x* is

$$B_\epsilon(x) = \{y \mid d(x, y) < \epsilon\}.$$

Now, if someone hands us the collection of ϵ-balls around points we can ask, "Is $y \in B_\epsilon(x)$?" as a proxy for closeness without directly computing distance.

Another useful notion from Euclidean space is that of an *open set*.

Definition. Let $U \subseteq \mathbb{R}^d$. If for every $x \in U$ there is some $\epsilon(x) > 0$ with $B_{\epsilon(x)}(x) \subseteq U$, we call U an *open set*.

In real-world systems, it is often sane to assume a notion of *genericity* - it is impossible to make perfectly precise measurements, and systems must operate in the presence of noise and other complications so cannot be too finely tuned. Fron a mathematical perspective, we can imagine that if a point x has a property, then those points nearby should as well. This allows us a measure of inductivity in reasoning: if x has a property, then there is some neighborood in which I expect that property to hold, possibly with slight modification. Wandering out into that neighboorhood, I find a new point x', and know that every point in some neighboorhood of that point is going to have a similar property, and so on. This is, for example, is a standard method in proving the existence of global solutions to differential equations: we show that such solutions exist and are unique on local patches, then "sew together" such solutions to obtain unique solutions everywhere.

Notice, we can flip this characterization around: an open set is one which is a union of open balls:

$$U = \bigcup_{x \in U} B_{\epsilon(x)}(x).$$

In particular, this is true *for open balls* – they are themselves open sets!

In our setting, we won't have all of these ϵs lying around, but we're going to mimic this notion.

Definition. Let X be a set. A *topology on X* is a collection $\tau \subseteq 2^X$ of *open sets* so that

1. $X \in \tau$ and $\varnothing \in \tau$,
2. If $U_i \in \tau$ for all i in some index set A, then $(\bigcup_{i \in A} U_i) \in \tau$, and

3. If $U_i \in \tau$ for all $i \in A$, $|A| < \infty$, then $(\bigcap_{i \in A} U_i) \in \tau$.[11]

A *topological space* is a pair (X, τ) where X is a set and τ is a topology on X. If $U \subseteq X$ has $U \in \tau$, we say U is *open in* X. Given a point $x \in X$, every open set containing x is an *open neighborhood of* x. A *closed set* is the complement of an open set; i.e. $C \subseteq X$ is closed if there is $U \in \tau$ so that $C = X \setminus U$.[12]

[12] Bad news! We have that $\varnothing$ is an open set, but X is open, so its complement, $\varnothing$, must be closed. Open and closed are *not opposites*!

Topologies are complex objects, so we usually build them by specifying simpler families of sets that *generate* them, like our ϵ-balls from $\mathbb{R}^d$.

Definition. Let X be a set. If $\mathcal{B}$ is a collection of subsets of X, containing $\varnothing$ and X, so that every finite intersection of elements of $\mathcal{B}$ can be written as a union of elements of $\mathcal{B}$, then $\mathcal{B}$ is a *base* and *generates* some topology τ on X.

Examples.

- $\mathbb{R}^d$
  a. $\{B_\epsilon(x) \mid x \in X, \epsilon > 0\}$ is a base which generates the *standard* or *Euclidean topology* on X.
  b. $\{B_\epsilon(x) \mid x \in X, \epsilon > 0, \epsilon \in \mathbb{Q}\}$ also generates the standard topology.
  c. So does $\{B_\epsilon(x) \mid x \in \mathbb{Q}^k, \epsilon > 0, \epsilon \in \mathbb{Q}\}$
- X a set, $\tau = \{\varnothing, X\}$, the *trivial topology*.
- X a set, $\tau = 2^X$, the *discrete topology*.
- $X = \{p, q\}$, $\tau = \{\varnothing, \{p\}, \{p, q\}\}$, called the *Sierpinski space*

While it is occasionally convenient to describe a topological space from the ground up, it is usually conceptually cleaner and easier to understand a description if we describe it in relation to a space we already know.

Definition. Let (X, τ) be a topological space and $Y \subseteq X$. The subspace topology on Y is given by $\{U \cap Y \mid U \in \tau\}$.

Examples.

- Let $I = [0, 1] \subset \mathbb{R}$, the *standard closed interval*. What are the open sets in the subspace topology?
- The *standard d-sphere* is

  $$S^d = \{x \in \mathbb{R}^{d+1} \mid ||x|| = 1\}$$

  Open sets in S^d under the subspace topology are given by intersections of Euclidean open sets with the sphere.

- The *standard d-disk* is

$$D^d = \{x \in \mathbb{R}^d \mid ||x|| \leq 1\}$$

  Observe that there is a canonical (identity) map $S^d \hookrightarrow D^{d+1}$. This is going to be one of our favorite maps.

- The *standard d-simplex* is the set

$$\Delta^d = \{x \in \mathbb{R}^{d+1} \mid \sum_{i=1}^{d+1} x_i = 1, x_i \geq 0, i = 1, \ldots, d+1\}.$$

  We should actually be thinking of the d-simplex as a subspace of the hyperplane $\sum_{i=1}^{d+1} x_i = 1$ which is itself a subspace of $\mathbb{R}^{d+1}$.

This last comment is important for the following reason: subspaces have structure relative to the larger space.

Definition. Let $Y \subseteq X$ be a subspace. The *boundary* of Y, ∂Y is the collection of all $x \in X$ so that every open neighborhood U_x of x has $U_x \cap Y \neq \varnothing$ and $U_x \cap (X \setminus Y) \neq \varnothing$. The *interior* of Y is $Y \setminus \partial Y$ and the *closure* of Y is $Y \cup \partial Y$.

Note that these terms depend on both X and Y. The d-simplex is its own boundary as a subspace of $\mathbb{R}^d$, but has a non-empty interior as a subspace of the hyperplane.

Continuous functions

Just like with graphs, we want a notion of how to compare topological spaces. The "correct" choice of maps should preserve the data of the topological space: open sets.

We might want to say that $f : X \to Y$ is continuous if it sends open sets to open sets. However, consider the constant map

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto 0.$$

Clearly this should be continuous, but it maps the open set $\mathbb{R}$ to the non-open set $\{0\}$.

Note that a function f is continuous at x if the image of a small neighborhood N_x of x, $f(N_x)$ gets arbitrarily small as N_x does. That is, if we consider a small open neighborhood of $f(x)$, it should be the image of a small open neighborhood of x.

Definition. Let X and Y be topological spaces. A function $f : X \to Y$ is *continuous* if for every open set $U \subseteq Y$, the set $f^{-1}(U) \subseteq X$ is open. We call continuous functions *maps*.

In our constant function example above, if $U \subseteq \mathbb{R}$ is an open neighborhood of zero, $f^{-1}(U) = f^{-1}(0) = \mathbb{R}$ which is open. If V is an open set in $\mathbb{R}$ which doesn't contain zero, then $f^{-1}(V) = f^{-1}(\varnothing) = \varnothing$, which is also open. Thus, the function is continuous.

Theorem 16. *For a function $f : \mathbb{R} \to \mathbb{R}$, the above definition of continuity agrees with the usual ϵ-δ definition from calculus.*

Proof. Exercise. □

Our definitions of the subspace topology is motivated by the desire to make the natural inclusion of sets into a map in the simplest possible way.

Lemma 17. *Let X be a topological space, $Y \subseteq X$ equipped with the subpace topology. The* inclusion map *$\iota : Y \hookrightarrow X$ given by $\iota(x) = x$ is continuous.*

Proof. Let $U \subseteq X$ be open. The preimage is $\iota^{-1}(U) = \{y \in U \mid y \in Y\} = U \cap Y$, which is open in Y by definition, so ι is continuous. □

One of the most powerful tools we'll have in our quest to build new topological spaces are *quotient spaces*. Quotients provide us with the power to "glue pieces together" or "forget information", depending on context. We build quotient spaces by defining an *equivalence relation*[13] on the space X and then building the simplest topology that makes the induced *projection* map continuous.

Definition. Let (X, τ) be a topological space and $\sim$ an equivalence relation on X. The quotient map $\pi : X \to (X/\sim)$ is given by $\pi(x) = [x]$, and the quotient topology on $(X/\sim)$ is the smallest topology that makes the quotient map continuous; i.e. $U \in (X/\sim)$ is open if and only if the preimage $\pi^{-1}(U)$ is open[14].

Quotient spaces are one of the most powerful tools we have for building spaces.

Examples.

- Let $\Gamma = (V, E)$ be a graph. The *topological realization* of Γ is the topological space given by[15]

$$|\Gamma| = (\coprod_{v \in V} \Delta^0 \times \coprod_{vv' \in E} \Delta^1)/((1,0)_{vv'} \sim v, (0,1)_{vv'} \sim v').$$

  We call any space obtained in this fashion a *topologicalgraph*. Open sets in this space inside of edges are just the same as usual. If they cross a vertex, they've got to bleed into *all* nearby edges.

With this notion of the topological realization of a graph, we can re-define the earlier notion of a geometric realization in a cleaner, more understandable way.

[13] Recall that an equivalence relation on a set X is a relation $\sim$ which is reflexive, symmetric and transitive. Write $[x]$ for the *equivalence class* of x. The equivlance classes under $\sim$ partition X, so we can think of the collection of classes as a new set $(X/\sim) = \{[x] \mid x \in X\}$, the *quotient of X by $\sim$*.

[14] More explicitly, $U \subseteq (X/\sim)$ is open when $\bigcup_{[x] \in U} \bigcup_{x \in X} x$ is open in X.

[15] This is a lie. In order to do this correctly, what we really need to do is make a choice for each edge of a vertex to attach at 0 and the other to attach at 1. That requires notation that obscures a simple idea, so I didn't write it down.

Definition. Let Γ be a graph. A geometric realization ρ of Γ is an injective, continuous function $\rho : \Gamma \to \mathbb{R}^d$. We call ρ an *embedding* of the topological graph into $\mathbb{R}^d$.

Reeb spaces

When we were discussing Jordan Curves, we mentioned path components in a context more general than graphs, but left the notion intuitive. We now have the technology to make a formal definition.

Definition. Let X be a topoogical space. A *path* in X is a continuous function $\gamma : I \to X$. The *endpoints* of the path are $a = \gamma(0)$ and $b = \gamma(1)$, and we say γ is a *path from a to b*. The existence of a path from a to b induces an equivalence relation $\sim$ on X[16]. Write $\pi_0(X) = X/\sim$, and call $\pi_0(X)$ the *path components* of X. Say X is *path connected* if $|\pi_0(X)| = 1$.

[16] Exercise.

Path components are preserved by continuous functions.

Lemma 18. *Let X, Y be topological spaces and $f : X \to Y$ a continuous function. There is an induced function on the path components $f_* : \pi_0(X) \to \pi_0(Y)$ given by $f_*([x]) = [f(x)]$.*

Proof. Because we are defining a function on a equivalence classes, what we need to show is that it is well-defined: if $[x] = [y] \in \pi_0(X)$ then $f_*([x]) = f_*([y])$. By definition, if [x] = [y] then there is a path $\gamma_{xy} : I \to X$. If we post-compose γ by f, we obtain a new continuous function[17] $(f \circ \gamma) : I \to X \to Y$ with $f \circ \gamma(0) = f(x)$ and $f \circ \gamma(1) = f(y)$. Thus, $[f(x)] = [f(y)]$ in $\pi_0(Y)$. □

[17] Exercise.

Another useful tool that comes up over and over in the study of topological spaces are real-valued functions on a topological space; we can use the ordering of points on the real line to "decompose" the space through the lens of the function.

Definition. Let X be a topological space, and $f : X \to \mathbb{R}$ a continuous function. For $c \in \mathbb{R}$, the *level set* for f at c is $X_c = f^{-1}(c)$[18]. The *sublevel set* for f at c is $X_{\leq c} = f^{-1}((-\infty, c])$, and the *superlevel set* is $X_{\geq c} = f^{-1}([c, \infty))$.

[18] This notation X_c isn't optimal: the level set clearly depends on the choice of f. However, it is standard in the literature, so we will establish it here and use it when f is clear from context, falling back on $f^{-1}(c)$ when there might be confusion.

Studying how these subspaces change as c varies is an example of *persistence*, which will come back as a major topic later in the course. For now, we'll combine level sets with components to get a tool that sees use in computational geometry and data visualization.

Observe that for $c < c'$, the sublevel set $X_{\leq c}$ is a subspace of the sublevel set $X_{\leq c'}$, and the inclusion map thus induces a map $(i_{cc'})_* : \pi_0(X_{\leq c}) \to \pi_0(X_{\leq c'})$. Similarly, we have $(i_{c'c})_* : \pi_0(X_{\geq c'}) \to \pi_0(X_{\geq c})$.

Definition. Let X be a topological space, and $f : X \to \mathbb{R}$ a continuous function. We construct an equivalence relation on X as follows: for each $c \in \mathbb{R}$ and $x, y \in f_c$, $x \sim y$ if they are in the same path component of f_c. The *Reeb space* of f is $X/\sim$.

Under some mild hypotheses, the Reeb space is a topological graph.

Lemma 19. *Suppose X is a topological space and $f : X \to \mathbb{R}$ a continuous function. If*

1. *there are $a, b \in \mathbb{R}$ so that $f^{-1}((-\infty, a]) = f^{-1}([b, \infty)) = \varnothing$;*
2. *there are finitely many points $a = c_1 < c_2 < \cdots < c_k = b \in \mathbb{R}$ so that for $i = 1, \ldots, k-1$*
   - *if $c < c' \in [c_i, c_i + 1)$, the map $(i_{cc'})_* : \pi_0(X_{\leq c}) \to \pi_0(X_{\leq c'})$ is a bijection, and*
   - *if $c < c' \in (c_i, c_i + 1]$, the map $(i_{c'c})_* : \pi_0(X_{\geq c'}) \to \pi_0(X_{\geq c})$ is a bijection; and*
3. *$|\pi_0(X_c)| < \infty$ for all $c \in \mathbb{R}$.*

Then the Reeb graph of f is a topological graph. That is, there is a graph Γ for which $|\Gamma|$ "is" the Reeb space[19].

[19] We're going to have to hold our breath on the meaning of the word "is" here for just a little bit.

The upshot of this definition is that if your space and function are nice enough, then you get a graph back and use your favorite graph tools to study it. "Nice enough" includes things like finitely triangulatable surfaces, which get a lot of use in computer graphics and visualization.

Homeomorphisms and homotopy equivalence

Now that we know what our fundamental objects of study are, we need a way of deciding which topological space we're looking at when we find one in the wild – that is, we need a notion of "sameness." There are several to choose from, and here we will discuss two. The first, *homeomorphism*, has a very natural definition in terms of continuous functions and will be useful for describing *topological manifolds*, a fundamental class of topological spaces. The second, *homotopy*, is weaker and somewhat harder to understand conceptually, but is compatible with a powerful collection of algebraic tools, so we will tend to use it moving forward.

The natural notion of sameness, as we discussed in the case of graphs, is a homomorphism – a map which respects the instrinsic structure – and which further presercves all of that structure. In the case of topological spaces, homomorphisms are open sets; it is

tempting to suggest that an isomorphism is a bijective continuous function. However, $f : [0, 2\pi) \to S^1$ by $f(t) = (\cos(t), \sin(t))$ is clearly a continuous bijection, and it is reasonable to hope that a circle and an interval are not topologically equivalent. The issue is that the inverse is not continuous: open neighborhoods of $[0, \epsilon)$ of $0 \in [0, 2\pi)$ pull back to sets that aren't open in S^1.

Definition. Let X and Y be topological spaces. A *homeomorphism* $f : X \to Y$ is a continuous bijection with a continuous inverse. If such a map exists, X and Y are *homeomorphic* and write $X \cong Y$.

Homeomorphism is difficult to deal with because of its strength: many features can serve as obstructions to the existence of a homeomorphism. While we will not spend much time developing them here, but two features of homeomorphisms will be important.

Examples.

- $\mathbb{R}^n \ncong \mathbb{R}^m$ if $n \neq m$ – homeomorphism preserves dimension.
- $B_\epsilon(0) \ncong D^n$ – homeomorphism preserves "special" points like those on the boundary

This dimension invariance property is vital for the following ubiquitous class of topological spaces.

Definition. A *topologcal manifold* is a topological space X so that

- there is some $n \in \mathbb{N}$ so that for every $x \in X$ has an open neighborhood U_x with $U_x \cong \mathbb{R}^n$,
- for every pair of points $x, y \in X$, there are open neighborhoods U_x and U_y respectively with $U_x \cap U_y = \varnothing$, and
- the topology of X has a countable base.

Topological manifolds are topological spaces which are *locally Euclidean*, including arbitrary open subsets of Euclidean space and spheres. Non-examples include topological realizations of graphs, unless each vertex has degree exactly two.

Definition. Let X and Y be topological spaces. The *product* space $X \times Y$ is the topological space on the set $X \times Y$ with topology generated by the collection $U \times V$, where $U \in \tau_X$ and $V \in \tau_Y$.[20]

[20] This topology works for products of finitely many spaces. If you want infinite products, you need to refine the definition a bit.

The open sets on $X \times Y$ are precisely the minimal ones needed to make the *projection* maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ continuous.

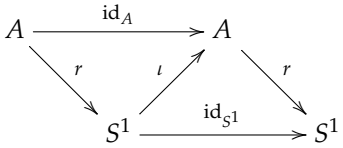
Definition. Let X and Y be topological spaces and $f, g : X \to Y$ be continuous functions. We say f is *homotopic* to g, written $f \simeq g$, if there is a continuous function $H : I \times X \to Y$ with $H(0, x) = f(x)$ and $H(1, x) = g(x)$.

The right intuition for homotopy is that there is a "continuous path through continuous functions" from f to g.

Examples.

- Let $\iota : S^2 \hookrightarrow \mathbb{R}^3$ be the standard inclusion $\iota((x, y, z)) = (x, y, z)$ and $z : S^2 \to \mathbb{R}^3$ be the constant map $z(x, y, z) = (0, 0, 0)$. Take $H : I \times S^2 \to \mathbb{R}^3$ to be $H(t, (x, y, z)) = t(x, y, z)$, so $H(0, (x, y, z)) = z(x, y, z)$ and $H(1, (x, y, z)) = \iota$. If a function if homotopic to a constant map, we call it *null-homotopic* and write $\iota \simeq *$.
- Let $e : S^1 \hookrightarrow S^2$ by $e(x, y) = (x, y, 0)$ and $p(x, y) = (0, 0, 1)$. Let $H : I \times S^1 \to S^2$ by $H(t, (x, y)) = (x\sqrt{1 - t^2}, y\sqrt{1 - t^2}, , t)$. This is a null-homotopy of e.
- Let $A = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq ||(x, y)|| \leq 2\}$, take $\mathrm{id}_A, r : A \to A$ by $\mathrm{id}_A(x, y) = (x, y)$ and $r(x, y) = \frac{(x,y)}{||(x,y)||}$. We have $\mathrm{id}_A \simeq r$ via $H(t, (x, y)) = \frac{(x,y)}{(1-t)+t||(x,y)||}$.

In the last example, we deformed the identify function on an annulus to a projection onto the inner boundary circle. If we "run the homotopy backwards", we get an expansion of the circle into the whole annulus, so it seems like the annulus and the circle are "the same" in terms of continuous deformation. Let's break that up and look at the component pieces.



If we take the entire annulus and pass it through r, we are collapsing it onto the circle. Then, we can send it back through the canonical inclusion to get the inner boundary of the annulus as a subspace of the annulus. Finally, we use the homotopy H to re-expand to the identiy map on the annulus. Thus, $\iota \circ r \simeq \mathrm{id}_A$. This says, "Passing A *through* S^1 via $\iota \circ r$ is the same as leaving it be, up to continuous deformation." So, S^1 is "big enough" to hold all of the information in A, up to homotopy. Observe that the same is true if we invert the roles of A and S^1: $r \circ \iota \simeq \mathrm{id}_{S^1}$ (indeed, $r \circ \iota = \mathrm{id}_{S^1}$.) Thus, each space is "at least as big" as the other, and so we should think of them as equivalent.

Definition. Let X, Y be topological spaces, $f : X \to Y$ and $g : Y \to X$ be continuous functions. If $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$, we say X and Y are *homotopy equivalent* and write $X \simeq Y$. If a space is homotopy equivalent to a point, it is called *contractible*.

Examples.

- $\mathbb{R}^d \simeq *$ via $\iota : \{0\} \to \mathbb{R}^d$ by $\iota(0) = 0$ and $p : \mathbb{R}^d \to \{0\}$ the constant map.
- $\mathbb{R}^d \setminus \{0\} \simeq S^{d-1}$ via $\iota : S^{d-1} \hookrightarrow \mathbb{R}^d \setminus \{0\}$ the standard inclusion and $\nu : \mathbb{R}^d \setminus \{0\} \to S^{d-1}$ by $\nu(x) = x/||x||$.

In general, to show two things are homotopy equivalent, we must demonstrate the appropriate maps and homotopies. To show two things are *not* homotopy equivalent, we find a property of the spaces that is a *homotopy invariant* and show that that property is different between the two spaces.

Simplicial complexes

Topological spaces are extremely general, and so working with them without constraints is difficult. Further, the application-focused reader might wonder how we can perform any computations with them – how do we put all of this open set data into a computer?

One of the foundational results in algebraic topology is the fact that we can approximate any sufficiently nice space with a combinatorial object called a simplicial complex, which is a dramatic generalization of a graph which encodes relations that inherently involve multiple objects.

Definition. An *(abstract) simplical complex* Σ is a pair of sets $\Sigma = (V, S)$ of *vertices* and *simplices*, with $S \subseteq 2^V$ so that i. if $\sigma \in S$, $|\sigma| < \infty$; and, ii. if $\sigma \in S$ and $\tau \subset \sigma$ then $\tau \in S$.

This last property is called *subset closure*[21], and implies that we only need to specify faces which are maximal under inclusion; e.g., if $\tau \subsetneq \sigma$ and $\sigma \in S$, we can omit mention of τ when describing Σ. This dramatically reduces the amount of information needed to specify a simplicial complex. We call such maximal simplices the *facets* of the complex. As shorthand, we will usually write the set of facets, $F(S)$ in place of S when specifying Σ.

[21] This closure condition is essentially true for graphs, too: if an edge is contained in the graph, it is necessarily the case that the vertices of that edge are in the graph as well. We just don't add the vertices (or the empty set) to the list of edges.

Example.

$V =$ students, $S =$"have shared a class"

Facets of this a complex are maximal collections of students who have shared a clsss – that is, the entire class roster. Note that if we

encode this information in a graph using the given relation, it is impossible to tell if three students shared a single class or each of the three pairs shared distinct classes[22].

Simplices $\sigma \in S$ with $|\sigma| = (k+1)$ are called k-simplices, written $S_k \subseteq S$. This unintuitive naming convention comes from fact that $(k+1)$ points generically span a k-dimensional space, and we are planning to make a leap to geometry shortly[23]. The k-skeleton of Σ is the subcomplex consisting of all ℓ-simplices for $\ell \leq k$.

Definition. Let $\Sigma_1 = (V_1, S_1), \Sigma_2 = (V_2, S_2)$ be simplicial complexes. A *(simplicial complex) homomorphism* $\Phi : \Sigma_1 \to \Sigma_2$ is a function $\phi : V_1 \to V_2$ which induces a function on k-simplices $\widetilde{\phi}_k : (S_1)_k \to (S_2)_k$ for each k.

As with graphs, unless otherwise noted we will assume all simplicai complexes have $|V| = n < \infty$. In order to be explicit about constructions, it will be convenient to choose an ordering on them[24]. Thus, we will assume from here forward that $V = (1, 2, \ldots, n) = [n]$.

Using this ordering, $\sigma \in S_k$ can be written uniquely as an ordered $(k+1)$-tuple $\sigma = (i(0), i(1), \ldots i(k))$ with $i(0) < i(1) < \cdots < i(k)$. This is unwieldy, so we will denote this simplex by $\sigma_{i(0)i(1)\cdots i(k)}$ or simply $i(0)i(1)\cdots i(k)$. For example, the 3-simplex $(2, 3, 5, 8)$ is denoted σ_{2358} or 2358.[25]

Recall from our discussion of the standard k-simplex Δ^k that the boundary of Δ^k is made up of standard $(k-1)$-simplices. In an abstract simplicial complex, the subset closure property ensures that all subsets, called *faces*, of each simplex are contained in the complex. In particular, those faces obtained by omitting a single vertex are present.

Definition. Let $\Sigma = (V, S)$ a simplicial complex and $\sigma = i(0) \cdots i(k) \in S_k$. The *boundary* of σ is

$$\partial(\sigma) = \{i(0) \cdots \widehat{i(\ell)} \cdots i(k) \mid \ell = 0, \ldots, k\} \subseteq S_{k-1},$$

where $\widehat{i(\ell)}$ denotes omitting an index.

For example, $\partial(2358) = \{358, 258, 238, 235\}$.

With our orientation and the standard[26] ordered basis $e_0, \ldots e_k$ on $\mathbb{R}^{k+1}$, there is a canonical way to associate a k-simplex $\sigma = i(0)i(1)\cdots i(k)$ with the standard k-simplex Δ^k: send the vertex $i(\ell)$ to the unit basis vector $e_\ell \in \mathbb{R}^{k+1}$. Write Δ^k_σ for this associated space.

Now, if $\tau \subset \sigma \in S$, τ consists of some subset $\{i(\ell_1), i(\ell_2), \ldots i(\ell_m)\}$ of the vertices of σ. Thus, the standard simplex Δ^m_τ is canonically a subspace of Δ^k_σ given by the non-negative span of $\{e_{\ell_1}, \ldots e_{\ell_m}\}$, where the inclusion $\iota_{\tau,\sigma}$ is the linear map which takes each unit vector in Δ^m_τ to its associated unit vector in Δ^k_σ.

[22] While we can work around this limitiation by enriching the graph structure, it is reaonable instead to work with data structures that naturally support the desired information (and many of the solutions people find turn out to be equivalent to moving to simplicial complexes and reinventing the wheel!)

[23] Thus, by convention, the empty simplex is a (-1)-simplex.

[24] Selecting this order is called choosing a *global orientation* and is analogous to choosing a left- or right-handed coordinate system when doing calculus in $\mathbb{R}^3$: potentially, it changes some quantities up to a sign, but these signs fall out in the end so long as everything is consistent. Since we're planning to work in $\mathbb{F}_2$ where signs don't exist, we can just ignore the issue entirely.

[25] If this notation is ambiguous – say 23 and 58 are vertices – add delimiters as necessary.

[26] Standard, but indexed in this non-standard way to make life simpler.

For example, Δ^2_{258} includes as the subspace of Δ^3_{2358} spanned by e_0, e_2, and e_3, via the linear map

$$\iota_{258,2358} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Definition. Let $\Sigma = (V, S)$ be a simplicial complex. The *(topological) realization* of Σ is the topological space given by

$$\left(\coprod_k \coprod_{\sigma \in S_k} \Delta^k_\sigma \right) / \sim$$

where if $\tau \in \partial(\sigma)$ and $x \in \Delta^{k-1}_\tau$, then $\iota_{\tau,\sigma}(x) \sim x$.

Simplicial approximations of topological spaces

Simplicial complexes are powerful both because they can be used to encode very general types of data and because they can be used to approximate an extensive range of topological spaces. Let's focus on making that last statement precise by constructing such an approximation.

Definition. Let X be a topological space, A an indexing set. If $\mathcal{U} = \{U_a\}_{a \in A}$ is a family of open (closed) sets in X so that $\bigcup_{a \in A} U_a = X$, we say $\mathcal{U}$ is an *open (closed) cover* for X. If, in addition, for each $B \subseteq A$, $\bigcap_{b \in B} U_b$ is either contractible or empty[27], $\mathcal{U}$ is a *good open (closed) cover*. If X is a subspace of a Euclidean space, then if each U_a is a convex set[28] we call $\mathcal{U}$ a *convex open (closed) cover*.

[27] In particular this includes the case where B is a singleton, so each set $U_a, a \in A$ has $U_a \simeq *$ or $U_a = \emptyset$.

[28] A *convex set* A is one for which every pair of poinst $a, b \in A$, the line segment connecting a to b lies in A.

Note that we want the sets to be either all open or all closed. Open is more common in general topology, but it will be convenient for us to have the closed version for use in applications.

Examples.

- Let $X = S^1$, U_1 a contractible open set that covers slightly more than the top semicircle, U_2 similarly for the bottom. $\mathcal{U}$ is an open cover, but not a good open cover because the intersection $U_1 \cap U_2$ is not path connected, thus $U_1 \cap U_2 \not\simeq *$.
- Let $X = S^1$, U_1, U_2 and U_3 contractible open sets that each cover slightly more than a third of the circle. This is a good cover, since each $U_a cap U_b \simeq *$ for each pair a, b and $U_1 \cap U_2 \cap U_3 = \varnothing$.
- Let $X = \mathbb{R}^2$ and $U_a = \{x \in \mathbb{R}^2 \mid a - 1 < ||x|| < a + 1\}, i = 0, 1, 2, \ldots$. This is an open cover for which almost all individual U_1 are not contractible.

- Let $X = \mathbb{R}^2$ and $U_a = B_a(0), a = 0, 1, 2, \ldots$. This, on the other hand is a good cover, and a convex cover.

Covers are a way of breaking a topological space up into simpler pieces. Often, we can perform a computation on each of the simpler pieces and then glue the answers back together on the overlaps (using a quotient construction) to get the answer on the whole space, thus avoiding a more complex computation.

There are (usually) many different ways to build a cover for a space. Many spaces of interest admit *finite covers* where $|A| < \infty$, and the existence of a finite cover can be thought of as an analog of a set being finite: your intuition will generally hold because you won't have to take limits while gluing. Many useful topological spaces have the property that we can always build a finite cover.

Definition. Let X be a topological space. Let $\mathcal{U}$ be an open cover of X. A *subcover* of $\mathcal{U}$ is a subset $\mathcal{V} \subset \mathcal{U}$ so that $\mathcal{V}$ is also a cover of X. If every open cover of X admits a finite subcover, then X is called *compact*.

We can construct a simplicial complex from a finite cover by encoding the intersection patterns of the constituent sets.

Definition. Let X be a topological space and $\mathcal{U} = \{U_a\}_{a \in A}$ a finite open (closed) cover of X. The *nerve* of $|(U)$ is the simplicial complex $N(\mathcal{U}) = (A, S)$, where

$$\sigma \in S \iff \bigcap_{a \in \sigma} U_a \neq \varnothing.$$

Let's revisit our finite cover examples above.

Examples.

- In the two-set cover of S^1, we have $N(\mathcal{U})$ given by $A = \{1, 2\}$ and $F(S) = \{12\}$.
- In the three-set cover of S^1, $N(\mathcal{U})$ has $A = \{1, 2, 3\}$ and $F(S) = \{12, 13, 23\}\}$.

The second is built from a finite good cover of S^2. Further, the nerve of the cover is isomorphic to C_3, the three-node cycle graph, and $|N((U))| \simeq S^1$.[29]

Indeed, this is true in general.

[29] Contrast with the realization of the two-set cover of S^1 which is not a good cover.

Theorem 20 (Nerve theorem)**.** *Let X be a topological space and either i. $\mathcal{U}$ a finite, good open cover of X, or ii. $\mathcal{U}$ a finite, convex closed cover of X. Then $|N(\mathcal{U})| \simeq X$.*

The nerve theorem says that in topological spaces which admit such "nice" covers, the homotopy type of the space comes down to the combinatorics of a good/convex cover. If we *triangulate* a space, we are essentially building a closed, convex cover by simplices[30], we don't lose any homotopy-type information. Of course, we also need to make sure we get all of the continuous maps we expect.

Definition. Let Σ_1 and Σ_2 be simplicial complexes, and let $\phi : \Sigma_1 \to \Sigma_2$ a homomorphism. The *realization* of ϕ is the continuous function $|\phi| : |\Sigma_1| \to |\Sigma_2|$ given by linearly extending the map on the vertices to the interior of each simplex.

Theorem 21 (Simplicial approximation theorem)**.** *Let Σ_1 and Σ_2 be simplicial complexes, and let $f : |\Sigma_1| \to |\Sigma_2|$ be a continuous function. There is a subdivision[31] $D\Sigma_1$ of Σ_1 and a simplicial map $\widehat{f} : D\Sigma_1 \to \Sigma_2$ so that $|\widehat{f}| \simeq f$.*

Together, these results say that it is entirely reasonable (in most circumstances) to work with simplicial complexes rather than general topological spaces.

[31] A *subdivision* of a simplicial complex involves breaking some of its simplices into collections of simplices. For example, adding a vertex in the middle of a triangle, along with all edges from the new vertex to the vertices of that triangle. The result is a "more granular" complex with the same realization – $|D\Sigma| \simeq |\Sigma|$.

Homological Algebra

Now that we have a firm grasp on the objects we're going to be studying, we need a quantitative framework we can use to do the computations we will use to explore properties of those objects. As is nearly always the case in mathematics, the best tool for the job is algebra, in this case a sub-field called Homological Algebra. Here, we will focus on a version built out of vector spaces so we can rely on our understanding of linear algebra. Because computers are very good at linear algebra, this is the principal approach used in applications.

(Review of) Linear Algebra

What follows is a "review" in the mathematical sense: I'm going to pretend that you have a passing familiarity with these ideas whether you are or not, and it's up to you to choose where you need to spend time. You have probably seen much of this material at some point, but most of it is unlikely to have been the focus of your experience with linear algebra; some of it might not even have appeared in your linear algebra courses despite being a selection of fundamental tools in the subject, such are the vagaries of time limitations on course design. If you have not seen or do not feel comfortable with these ideas, clearly the review is doubly necessary.

We'll start from the beginning.

Definition. • A *field* $(\mathbb{F}, +, \cdot)$ is a set $\mathbb{F}$ equipped with binary operations $+$, called addition, and $\cdot$, called multiplication[32]. Both addition and multiplication are commutative, associative operations with identity elements 0 and 1 respectively. All elements of $\mathbb{F}$ have additive inverses and all elements besides 0 have multiplicative inverses, and multiplication distributes over addition.

[32] As usual, we will (almost always) drop the "dot" for multiplication from our notation.

- A *vector space* $(V, +)$ over $\mathbb{F}$ is a set V along with a binary operation $+$, called addition. and a map $- \cdot - : \mathbb{F} \times V \to V$. Addition is associative and commutative, has an identity element 0, and additive inverses exist for every element. The identity element of $\mathbb{F}$ acts as an identity on V under scalar multiplication, scalar multiplication associates with multiplication in $\mathbb{F}$, scalar multiplication distributes over $+$, and addition in $\mathbb{F}$ distributes over scalar multiplication.

- A *subspace* $W \subseteq V$ is a subset of V which is a vector space with the structure induced by that on V.

- Let $\mathbb{F}$ be a field and S a finite set. The $\mathbb{F}$-*vector space with basis* S is

$$\mathbb{F}\langle S\rangle = \{\sum_{s\in S} c_s e_s \mid c_s \in \mathbb{F}\}.$$

  The cardinality of S is the *dimension* of $\mathbb{F}\langle S\rangle$[33].

[33] The fact that dimension is well-defined is vital to us, but we do not have time to dive deeply into it here. Recall from linear algebra that every finite-dimensional vector space has a basis, and that the cardinality of that basis is well-defined. To skip ahead just slightly, if V is an $\mathbb{F}$-vector space then $V \cong \mathbb{F}\langle S\rangle$ for some finite set S. Further, the cardinality of S is a complete isomorphism invariant: *every* $\mathbb{F}$-vector space with a basis with the same cardinality is isomorphic to V.

That is, $\mathbb{F}\langle S\rangle$ is the set of *formal* $\mathbb{F}$-*linear combinations* of elements of S. If the elements of S are indexed by elements of an indexing set A in some way, we will often write $\sum_{a\in A} c_a e_a$. If the elements of A have an order $(a(1), a(2), \ldots a(N))$, then we can write a column vector $[c_{a(1)}\ c_{a(2)} \cdots\ c_{a(N)}]^T$.

Examples.

- $\mathbb{R}$ is a field, $\mathbb{R}^n$ is the vector space

$$\mathbb{R}\langle\{1, 2, \ldots n\}\rangle = \text{span}_{\mathbb{R}}(\{e_1, \ldots, e_n\}).$$

  The collection of vectors $\{(x, 2x) \mid x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

- $\mathbb{F}_2 = \{0, 1\}$ is a field, with additive identity 0, multiplicative identity 1 and $1 + 1 = 0$. Let $S = \{\sigma_1, \sigma_2, \sigma_3\}$. The vector space $\mathbb{F}_2\langle S\rangle$ consists of elements $c_1 e_{\sigma_1} + c_2 e_{\sigma_2} + c_3 e_{\sigma_3} = c_1 e_1 + c_2 e_2 + c_3 e_3$, where the c_i are either 0 or 1.

Maps between vector spaces should, as always, preserve the structure of the vector space.

Definition. Let V, W be $\mathbb{F}$-vector spaces and $T : V \to W$ be a function. If $T(a_1 v_1 + a_2 v_2) = a_1 T(v_1) + a_2 T(v_2)$, T is a *linear transformation*. If

T is a bijection, it is a *(linear) isomorphism*. The *kernel* of T, sometimes called the *null space*, is

$$\ker(T) = \{v \in V \mid T(v) = 0\},$$

and the *image* of T is

$$\text{im}(T) = \{w \in W \mid w = T(v) \text{ for some } v \in V\}.$$

Lemma 22. *Let $T : V \to W$ be a linear transformation. Then $ker(T)$ is a subspace of V and $im(T)$ is a subspace of W.*

Proof. Exercise. □

Recall that choosing a matrix A_T for T is the performed by choosing ordered bases for V and W. In the case where those bases are given to us, we obtain the matrix just by checking where the basis vectors go.

Given a matrix A_T for a linear transformation T, there are straightforward algorithms for computing the image and kernel of T using *Gaussian elimination*. Recall that in Gaussian row elimination, we iteratively add a scalar multiple of one row to another, or multiply a row by a scalar, to put the matrix in *row echelon form*, where each row has as its first non-zero entry a 1 that appears to the right of the first non-zero entry above it. These initial 1s are called *pivots*.

Lemma 23. *Let A_T be the matrix of a linear transformation $T : V \to W$ and $REF(A_T)$ the row echelon form of T. The columns of A corresponding to those of $REF(A_T)$ in which the pivots appear are a basis for $im(T)$.*

Example.

Consider the following matrix of a linear transformation between $\mathbb{F}_2$ vector spaces.

$$A_T = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \quad REF(A_T) = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and the image of T has basis

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

Gaussian column elimination works similarly, resulting in *column echelon form*.

Lemma 24. *Let A_T be the matrix of a linear transformation $T : V \to W$. Define a block matrix*

$$A'_T = \left(\frac{A_T}{id}\right).$$

If we perform Gaussian column elimination on A'_T to put only the top A_T block into column echelon form, $TCEF(A'_T)$ those columns in the bottom block of $TCEF(A'_T)$ for which the corresponding columns in the top block of $TCEF(A'_T)$ are zero form a basis for $ker(T)$.

Example.

Back to our matrix A_T from above, we have

$$A'_T = \left(\begin{array}{cccccc} 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \quad TCEF(A'_T) = \left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

and the kernel of T has basis

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

It is no coincidence that there are three vectors in each of these bases.

Theorem 25 (Rank-Nullity theorem)**.** *Let $T : V \to W$ be a linear map. Then*

$$dim(im(T)) + dim(ker(T)) = dim(V).$$

Proof. Let $\dim(V) = N$, $\dim(\ker(T)) = k$. Fix $\{b_1, \dots b_k\}$ a basis for $\ker(T)$, and extend this to a basis for V by adding $\beta_{k+1}, \dots \beta_N$.

Let $v \in V$, so v can be uniquely written as

$$v = c_1 b_1 + \cdots + c_k b_k + c_{k+1}\beta_{k+1} + \cdots + c_N \beta_N.$$

Now,

$$\begin{aligned} T(v) &= T(c_1 b_1 + \cdots + c_k b_k + c_{k+1}\beta_{k+1} + \cdots + c_N \beta_N) \\ &= c_1 T(b_1) + \cdots + c_k T(b_k) + c_{k+1} T(\beta_{k+1}) + \cdots + c_N T(\beta_N) \\ &= c_{k+1} T(\beta_{k+1}) + \cdots + c_N T(\beta_N). \end{aligned}$$

Thus, $\{T(B_{k+1}), \ldots T(B_N)\}$ spans $\mathrm{im}(T)$.

Now, suppose there are coefficients $c_{k+1}, \ldots c_N$ so that

$$c_{k+1} T(\beta_{k+1}) + \cdots + c_N T(\beta_N) = 0.$$

But then

$$T(c_{k+1}\beta_{k+1} + \cdots + c_N \beta_N) = 0,$$

so $c_{k+1}\beta_{k+1} + \cdots + c_N \beta_N \in \ker(T)$. But each $\beta_i \in \ker(T)^\perp$, so the only element of this form in $\ker(T)$ is the origin. Thus, the c_i are all zero, so the $T(\beta_i)$ are linearly independent and thus form a basis for $\mathrm{im}(T)$. Thus, $\dim(\mathrm{im}(T)) = N - k$. □

Chain complexes

Using the notation above, it is straightforward to encode simplicial complexes in the world of vector spaces: we build vector spaces whose bases are the simplices.

Definition. Let $\Sigma = (V, S = \coprod_{k=0}^N S_k)$ be a simplicial complex. The *k-th (simplicial) chain group (over $\mathbb{F}_2$)* of Σ is $C_k(\Sigma; \mathbb{F}_2) = \mathbb{F}_2\langle S_k \rangle$. To make indexing easier, we set $C_{-1}(\Sigma) = 0$.

Since we will only be working over $\mathbb{F}_2$, we will omit it from notation, writing $C_k(\Sigma) = C_k(\Sigma; \mathbb{F}_2)$. And, to avoid any possible confusion, we write $e_\sigma \in C_k(\Sigma)$ for the basis vector corresponding to $\sigma \in S_k$. Finally, observe that the order on the vertices induces a lexicographic ordering on the simplices, so these vector spaces come with *canonical* ordered bases.

The boundary maps on the simplices give us induced linear transformations between the chain groups.

Definition. Let $\Sigma = (V, S = \coprod_{k=0}^N S_k)$ be a simplicial complex, $0 < k \leq N$. The *k-th differential* is the linear map $d_k : C_k(\Sigma) \to C_{k-1}(\Sigma)$ given on basis vectors by

$$d_k(e_\sigma) = \sum_{\tau \in \partial(\sigma)} e_\tau.$$

More explicitly,

$$d_k(e_{i(0)\cdots i(k)}) = \sum_{\ell=0}^{k} e_{i(0)\cdots \widehat{i(\ell)} \cdots i(k)},$$

where as usual the hat indicates which index to drop. By necessity, $d_0 : C_0(\Sigma) \to 0$ is the zero map.

Example.

Let Σ have $V = \{1,2,3,4\}$ and $F(S) = \{123, 234\}$. The chain groups of Σ are

$$\begin{aligned} C_2(\Sigma) &= \mathbb{F}_2\langle\{123, 234\}\rangle \\ C_1(\Sigma) &= \mathbb{F}_2\langle\{12, 13, 23, 24, 34\}\rangle \\ C_0(\Sigma) &= \mathbb{F}_2\langle\{1, 2, 3, 4\}\rangle \end{aligned}$$

The differentials are the linear maps

$$d_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad d_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

These vector spaces and linear maps fit together in a sequence called a *chain complex*

$$C_2(\Sigma) \xrightarrow{d_2} C_1(\Sigma) \xrightarrow{d_1} C_0(\Sigma) \xrightarrow{d_9} 0\,,$$

which together we denote $C_\bullet(\Sigma)$. Consider the 2-simplex 123. The boundary of the 2-simplex is

$$\partial(123) = \{12, 13, 23\},$$

and the boundaries of each of these boundary simplices are, respectively

$$\begin{aligned} \partial(12) &= \{\ \ 1,\ \ 2\ \ \ \ \ \} \\ \partial(13) &= \{\ \ 1,\ \ \ \ \ 3\ \ \} \\ \partial(23) &= \{\ \ \ \ \ 2,\ \ 3\ \ \} \end{aligned}$$

That is, each simplex in the *boundary of the boundary* appears twice. On the level of the differentials over $\mathbb{F}_2$, we have

$$d_2(e_{123}) = e_{23} + e_{13} + e_{12},$$

and

$$d_1(e_{12} + e_{13} + e_{23}) = (e_2 + e_1) + (e_3 + e_1) + (e_3 + e_2) = 0$$

In fact, if we compose the linear maps by multiplying the matrices we wrote down (in the proper order), we get

$$d_1 d_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

To unpack that a little, let $c \in C_2(\Sigma)$. If $c' = d_2(c)$, then $d_1(c') = 0$. That is, for every $c' \in \mathrm{im}(d_2)$, $c' \in \ker(d_1)$, so $\mathrm{im}(d_2) \subseteq \ker(d_1)$.

In fact, this is true in general.

Lemma 26. *Let Σ be a simplicial complex, $C_k(\Sigma)$, $C_{k-1}(\Sigma)$ and $C_{k-2}(\Sigma)$ chain groups of Σ with differentials d_k and d_{k-1}. Then $im(d_k)$ is a subspace of $ker(d_{k-1})$; i.e. $d_{k-1} \circ d_k = 0$.*

Proof. Exercise. □

Definition. An $\mathbb{F}_2$ *chain complex,* $C_\bullet$, is a sequence of $\mathbb{F}_2$-vector spaces called *chain groups* $\{C_k\}_{k=-1}^N$ and linear maps $\{d_k : C_k \to C_{k-1}\}_{k=0}^N$ called *differentials* so that $d_{k-1} \circ d_k = 0$. By convention $C_{-1} = 0$ and d_0 is thus the zero map.

Technically, we also need to check that $\text{im}(d_1) \subseteq \ker(d_0)$ in our example above to call $C_\bullet(\Sigma)$ a chain complex, but this is trivial: it is always the case that $\text{im}(d_1) \subseteq C_0(\Sigma) = \ker(d_0 = 0)$.

The fact that $\text{im}(d_2)$ is a subspace of $\ker(d_1)$ raises a new question: is it the *entire* kernel? To find out, we use our algorithm from the previous section to compute a basis for $\ker(d_1)$.

$$\left(\begin{array}{ccccc} 1&1&0&0&0\\ 1&0&1&1&0\\ 0&1&1&0&1\\ 0&0&0&1&1\\ \hline 1&0&0&0&0\\ 0&1&0&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&0&1 \end{array}\right) \to \left(\begin{array}{ccccc} 1&0&0&0&0\\ 1&1&0&0&0\\ 0&1&0&1&0\\ 0&0&0&1&0\\ \hline 1&1&1&0&1\\ 0&1&1&0&1\\ 0&0&1&1&0\\ 0&0&0&1&1\\ 0&0&0&0&1 \end{array}\right)$$

It's clear from the matrix of d_2 that the image has dimension 2, and since we have a two-element basis for the kernel, the two spaces must be isomorphic. However, since one is a subspace of the other, $\text{im}(d_2) = \ker(d_1)$. Even better, we can see that the images of the basis vectors of $C_2(\Sigma)$ under d_2 are the same vectors given by the algorithm above – this sort of observation will be useful later. (*A priori,* we would also need to compute a basis for the image to work out its dimension.)

Definition. Let $C_\bullet$ be a chain complex. We say $C_\bullet$ is *exact* at C_k if $\text{im}(d_{k+1}) = \ker(d_k)$. If $C_\bullet$ is exact at every k, we say it is *exact*.

A tremendous amount of modern mathematics is encoded by measuring the failure of chain complexes to be exact. We can apply the Rank-Nullity theorem in the case of exact chain complexes to obtain an interesting relationship between the dimensions of the chain groups.

Corollary 27. *Let $C_\bullet$ be an exact chain complex. Then*

$$\sum_{k=0}^{N} (-1)^k dim(C_k) = 0.$$

Homology

Now that we have our chain-complex bridge from the world of combinatorial topological spaces to algebra, we should give it a closer look.

Recall the example $\Sigma = (\{1,2,3,4\}, \{123, 234\})$, for which $C_\bullet(\Sigma)$ is

$$0 \xrightarrow{0} \mathbb{F}_2\langle e_{123}, e_{234}\rangle \xrightarrow{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}} \mathbb{F}_2\langle e_{12}, e_{13}, e_{23}, e_{24}, e_{34}\rangle \xrightarrow{\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}} \mathbb{F}_2\langle e_1, e_2, e_3, e_4\rangle \xrightarrow{0} 0$$

We showed that this chain complex is exact at C_1. It's clear that the kernel of d_2 is trivial (the zero vector space), and since it's a chain complex, the image of d_3 must also be zero – and it is, thus the complex is also exact at C_2. What about C_0? Since the complex is exact at C_1, $\ker(d_1) = \text{im}(d_2)$, which is clearly 2-dimensional. By the rank-nullity theorem,

$$\dim(\text{im}(d_1)) = \dim(C_1) - \dim(\ker(d_1)) = 5 - 2 = 3.$$

However, $\ker(d_0 = 0) = C_0$, so is dimension 4 – there is a gap of one dimension, so the complex is not exact at C_0. What is the meaning of this difference?

We can (and should) think about the problem from the perspective of the topological spaces, and since only 1- and 0- simplices are directly involved in constructing C_1, C_0, d_1, and d_0, we're really thinking about a topological graph.

As we just noted, the kernel of d_0 is spanned by all of the vectors affilited with the vertices, and since we're working over $\mathbb{F}_2$, elements in this vector space are just subsets of the vertices of the graph. What does it mean to be in the image of d_1? Let's look at some examples.

$$\begin{aligned} d_1(e_{12}) &= e_2 + e_1 \\ d_1(e_{13}) &= e_3 + e_1 \\ d_1(e_{12} + e_{13}) &= (e_2 + e_1) + (e_3 + e_1) = e_2 + e_3 \\ d_1(e_{12} + e_{34}) &= (e_1 + e_2) + (e_3 + e_4) \end{aligned}$$

It looks a lot like sums of pairs of basis vectors are in the image precisely if they are endpoints on a path. The final example takes two disjoint paths and gives us the endpoints of both, but that's the natural result of linearity: if we can build two things, we can also build their sum.

This all sounds strikingly familiar. We have a bunch of vertices, and we have a list of pairs of them that are path connected in the graph. Recall our definition of $\pi_0(X)$, the set of all path components of a topological space. If we knew how to take a quotient of vector spaces, this would give us exactly the same kind of information.

Definition. Let V be a vector space and W a subspace of V. The *quotient vector space* V/W is obtained by identifying all elements of W to the 0 element. Elements of V/W are *W-cosets*, written $[v] = v + W$.

Thus, in a quotient vector space we can introduce or remove linear combinations of elements of W without changing the element: $[v] = v + W = v + w + W = [v + w]$ for every $v \in V$ and $w \in W$.

Example.

Let $V = \mathbb{R}^2$ and $W = \text{span}([1\ 2]^T)$. In V/W, we have

$$\left[\begin{bmatrix}3\\5\end{bmatrix}\right] = \begin{bmatrix}3\\5\end{bmatrix} + \left\{ \begin{bmatrix}a\\2a\end{bmatrix} \;\middle|\; a \in \mathbb{R} \right\} = \left[\begin{bmatrix}7\\13\end{bmatrix}\right] = \left[\begin{bmatrix}0\\-1\end{bmatrix}\right]$$

Observe that $V/W \cong \mathbb{R}^1$. *However*, it is not $\mathbb{R}^1$ as we usually define it – here, elements are cosets of vectors. Just as with sets, whenever we work with quotient vector spaces, we need to be careful to do things like making sure functions are well-defined.

Since, in a chain complex, $\text{im}(d_k + 1) \subseteq \ker(d_k)$, we just do what comes naturally.

Definition. Let $C_\bullet$ be a chain complex, $k \geq 0$. The *k-th homology group (with $\mathbb{F}_2$ coefficients) of* $C_\bullet$ is

$$H_k(C_\bullet) = \frac{\ker(d_k)}{\text{im}(d_{k+1})}.$$

We abbreviate $H_k(\Sigma) = H_k(C_\bullet(\Sigma))$ for the /simplicial homology groups.

In our running example, what have we done? If the sum of two basis vectors corresponding to vertices in Σ are the endpoints of a path, say $e_1 + e_2 \in \text{im}(d_1)$, then in $H_0(\Sigma)$,

$$[e_1 + e_2] = e_1 + e_2 + \text{im}(d_1) = d_1(e_{12}) + \text{im}(d_1) = 0 + \text{im}(d_1) = [0].$$

The vector $[e_1]$, on the other hand, has for example

$$[e_1] = e_1 + \text{im}(d_1) = e_1 + d_1(e_{13}) + \text{im}(d_1) = e_3 + \text{im}(d_1) = [e_3].$$

Thus, in this vector space, every vertex is equivalent to every other vertex with which it is path connected. The one-dimensional vector space $H_0(\Sigma)$ corresponds to the single path component of the simplicial complex Σ.

It should be straightforward to extend this argument to convince yourself of the following.

Lemma 28. *Let Σ be a simplicial complex with ℓ path components. Then $H_0(\Sigma)$ is an ℓ-dimensional $\mathbb{F}_2$ vector space with one basis vector corresponding to each path component of Σ (resp. X).*

Swinging back around to our example one last time, we already know the complex is exact at C_1 and C_2, thus when we compute homology we will set every vector in the kernels equivalent to zero – $H_1(\Sigma) \cong H_2(\Sigma) \cong 0$. In summary, we have

$$H_*(\Sigma) = \begin{cases} \mathbb{F}_2 & * = 0 \\ 0 & * > 0 \end{cases}$$

What would these higher homology groups capture if they weren't zero?

In Σ, there were plenty of elements in the kernel of d_1, but too many in the image of d_2, which ended up "killing off" $H_1(\Sigma)$. Let's consider a modification of the complex Σ wherein we remove the facet 234 but leave its boundary edges. This new simplicial complex is $\Sigma' = (\{1,2,3,4\}, \{123, 24, 34\})$. Now we have $C_\bullet(\Sigma')$ as follows

$$0 \xrightarrow{0} \mathbb{F}_2\langle e_{123}\rangle \xrightarrow{\begin{pmatrix}1\\1\\1\\0\\0\end{pmatrix}} \mathbb{F}_2\langle e_{12}, e_{13}, e_{23}, e_{24}, e_{34}\rangle \xrightarrow{\begin{pmatrix}1&1&0&0&0\\1&0&1&1&0\\0&1&1&0&1\\0&0&0&1&1\end{pmatrix}} \mathbb{F}_2\langle e_1, e_2, e_3, e_4\rangle \xrightarrow{0} 0$$

Observe that nothing has changed on the right side, so we can safely conclude that $H_0(\Sigma') \cong \mathbb{F}_2$. Similarly, $\ker(d_2) = 0$ just as before, so $H_2(\Sigma') = 0$. However, the structure has changed at C_1. Using our computation from last time, we have that

$$\ker(d_1) = \mathbb{F}_2 \left\langle \begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\\1\\1\end{bmatrix} \right\rangle \qquad \operatorname{im}(d_2) = \mathbb{F}_2 \left\langle \begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix} \right\rangle$$

Thus, $H_1(\Sigma') \cong \mathbb{F}_2$, and so

$$H_*(\Sigma') = \begin{cases} \mathbb{F}_2 & * = 0, 1 \\ 0 & * > 1 \end{cases}$$

Observe that the generator of $H_1(\Sigma')$ is $[[1\ 1\ 0\ 1\ 1]^T]$. This equiva-

lence class consists of exactly two vectors:

$$\begin{bmatrix}1\\1\\0\\1\\1\end{bmatrix} = e_{12} + e_{13} + e_{24} + e_{34}$$

$$\begin{bmatrix}1\\1\\0\\1\\1\end{bmatrix} + \begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix} = e_{23} + e_{24} + e_{34}$$

The second is the sum of the vectors corresponding to the boundary of the missing simplex, 234 – this vector "picks out" the cycle that should bound the missing simplex. The first is another cycle, also surrounding the hole, though now running around the perimeter. Adding the vector from the image of d_2 had the effect of "pushing" the cycle through the simplex 123, replacing the edge 23 with the edges 12 and 13.

This generalizes to all dimensions: elements of the kernel of the differential are "closed cycles", collections of simplices whose boundaries cancel out[34], and two such elements are the same if they differ by the boundary of a simplex of higher dimension.

[34] Technically, the basis elements in the chain group corresponding to these simplices, but that's a mouthful.

Definition. Let $C_\bullet$ be a chain complex. The elements of $\ker(d_k)$ are called *k-cycles* and those of $\text{im}(d_{k+1})$ are called *k-boundaries*. Two cycles that differ by a boundary are called *homologous*, and they belong to the same *homology class*. The dimension of $H_k(C_\bullet)$ is the *kth Betti number*, $\beta_k(C_\bullet)$.

Informally, the *k-th Betti number* "counts the number of non-trivial k-cycles", though this is at best an approximation since we're dealing with equivalence classes, as we will see moving forward. In our examples, we have

$$\beta_k(\Sigma) = \begin{cases}1 & k = 0\\ 0 & k > 0\end{cases} \qquad \beta_k(\Sigma') = \begin{cases}1 & k = 0, 1\\ 0 & k > 1\end{cases}$$

Maps on Homology

As usual, the object of study is only useful to us in the context of how objects interact. We've encoded our simplicial complexes as chain complexes, then used the differentials induced by the maps to compute a new sequence of quotient vector spaces called homology

groups. What we haven't covered is what happens to homomorphisms between simplicial complexes, or more generally to continuous functions between topological spaces.

Definition. Let $C_\bullet$ and $C'_\bullet$ be chain complexes. A *(chain complex) homomorphism* $f : C_\bullet \to C'_\bullet$ is a collection of linear transformations $f_k : C_k \to C'_k$ so that $d'_k \circ f_k = f_{k-1} \circ d_k$.

That is, in the following diagram, the squares commute.

$$\begin{array}{ccccccccc}
\cdots & \xrightarrow{d_{k+2}} & C_{k+1} & \xrightarrow{d_{k+1}} & C_k & \xrightarrow{d_k} & C_{k-1} & \xrightarrow{d_{k-1}} & \cdots \\
 & & \downarrow{\scriptstyle f_{k+1}} & & \downarrow{\scriptstyle f_k} & & \downarrow{\scriptstyle f_{k-1}} & & \\
\cdots & \xrightarrow{d'_{k+2}} & C'_{k+1} & \xrightarrow{d'_{k+1}} & C'_k & \xrightarrow{d'_k} & C'_{k-1} & \xrightarrow{d'_{k-1}} & \cdots
\end{array}$$

Where can get chain complex homomorphisms? From simplicial complex homomorphisms, of course.

Lemma 29. *Let Σ, Σ' be simplicial complexes, $\Phi : \Sigma' \to \Sigma'$ a homomorphism. There is an induced homomorphism $\Phi_\# : C_\bullet(\Sigma) \to C_\bullet(\Sigma')$.*

Proof. We need to write down linear transformations $(\phi_\#)_k : C_k(\Sigma) \to C_k(\Sigma')$. It suffices to define the transformation on basis vectors, and there is only one sensible candidate:

$$(\phi_\#)_k(e_{i(0)i(1)\cdots i(k)}) = e_{\widetilde{\phi}_k(i(0)i(1)\cdots i(k))} = e_{\phi(i(0))\phi(i(1))\cdots\phi(i(k))}.$$

We need to check that this map commutes with the differentials. That is, that $(\phi_\#)_{k-1} \circ d_k = d_k \circ (\phi_\#)_k$. Again, it suffices to check on basis vectors, so we compute

$$\begin{aligned}
(\phi_\#)_{k-1}(d_k(e_{i(0)i(1)\cdots i(k)})) &= (\phi_\#)_{k-1}(\sum_{\ell=0}^{k} e_{i(0)\cdots\widehat{i(\ell)}\cdots i(k)}) \\
&= \sum_{\ell=0}^{k} (\phi_\#)_{k-1}(e_{i(0)\cdots\widehat{i(\ell)}\cdots i(k)}) \\
&= \sum_{\ell=0}^{k} e_{\phi(i(0))\cdots\widehat{\phi(i(\ell))}\cdots\phi(i(k))} \\
&= d_k(e_{\phi(i(0))\cdots\phi(i(k))}) \\
&= d_k(e_{\widetilde{\phi}_k(i(0)\cdots i(k))}) \\
&= d_k((\phi_\#)_k(e_{i(0)\cdots i(k)})).
\end{aligned}$$

□

Of course, all the work we put into defining homology would be less compelling if this wasn't a compatable story.

Lemma 30. *Let $C_\bullet$ and $C'_\bullet$ be chain complexes, and $f : C_\bullet \to C'_\bullet$ a homomorphism. For each k, there is an induced linear transformation $(f_k)_* : H_k(C_\bullet) \to H_k(C'_\bullet)$.*

Collectively, we write $f_* : H_*(C_\bullet) \to H_*(C'_\bullet)$.

Proof. There is only one sensible candidate for such a map:

$$(f_k)_*([\sigma]) = [f_k(\sigma)].$$

Clearly this map can be applied to any element in $H_k(C_\bullet)$. We need to show that this map has the right codomain, and that it is well-defined.

First, if $[\sigma] \in H_k(C_\bullet)$, $\sigma \in \ker(d_k)$. That is, $d_k(\sigma) = 0$, and since f_{k-1} is a linear transformation, $(f_{k-1} \circ d_k)(\sigma) = f_{k-1}(0) = 0$. Since f is a chain complex homomorphism, $d'_k(f_k(\sigma)) = f_{k-1}(d_k(\sigma)) = 0$, so $f_k(\sigma) \in \ker(d'_k)$ and $[f_k(\sigma)] \in H_k(C'_\bullet)$.

Now, suppose $[\sigma] = [\tau] \in H_k(C_\bullet)$. That is, $\tau = \sigma + d_{k+1}(\gamma)$ for some $\gamma \in C_{k+1}$. Thus,

$$\begin{aligned}
(f_k)_*([\tau]) &= [f_k(\tau)] \\
&= f_k(\tau) + \operatorname{im}(d'_{k+1}) \\
&= f_k(\sigma + d_{k+1}(\gamma)) + \operatorname{im}(d'_{k+1}) \\
&= f_k(\sigma) + f_k(d_{k+1}(\gamma)) + \operatorname{im}(d'_{k+1}) \\
&= f_k(\sigma) + d'_{k+1}(f_{k+1}(\gamma)) + \operatorname{im}(d'_{k+1}) \\
&= f_k(\sigma) + \operatorname{im}(d'_{k+1}) = [f_k(\sigma)] = (f_k)_*([\sigma]).
\end{aligned}$$

□

Putting these two lemmas together, any homomorphisms of simplicial complexes induces a map on homology.

Example.

Consider

$$\Sigma = (\{1,2,3,4\}, \{12,13,14,24,34\})$$

and

$$\Sigma' = (\{1',2',3',4'\}, \{1'2'3', 2'3'4', 1'4'\})$$

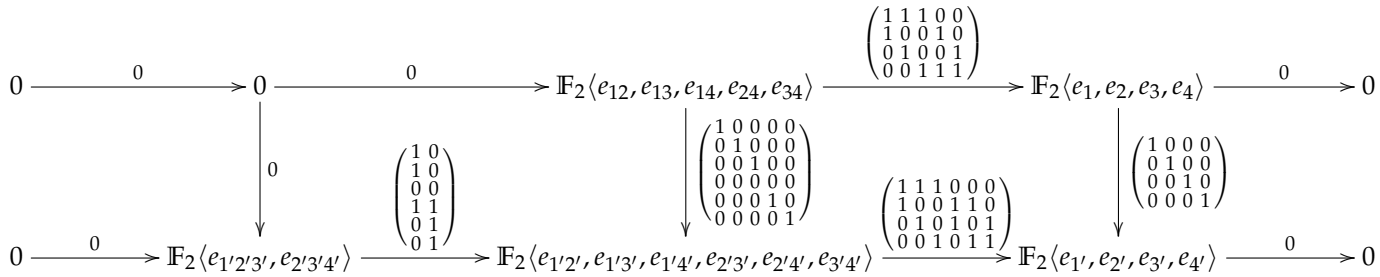
with homomorphism given on vertices by $\phi(i) = i'$. The chain complex for Σ is

$$0 \xrightarrow{0} \mathbb{F}_2\langle e_{12}, e_{13}, e_{14}, e_{24}, e_{34}\rangle \xrightarrow{\begin{pmatrix} 1&1&1&0&0 \\ 1&0&0&1&0 \\ 0&1&0&0&1 \\ 0&0&1&1&1 \end{pmatrix}} \mathbb{F}_2\langle e_1, e_2, e_3, e_4\rangle \xrightarrow{0} 0$$

and for Σ' we have

$$0 \xrightarrow{0} \mathbb{F}_2\langle e_{1'2'3'}, e_{2'3'4'}\rangle \xrightarrow{\begin{pmatrix} 1&0 \\ 1&0 \\ 0&0 \\ 1&1 \\ 0&1 \\ 0&1 \end{pmatrix}} \mathbb{F}_2\langle e_{1'2'}, e_{1'3'}, e_{1'4'}, e_{2'3'}, e_{2'4'}, e_{3'4'}\rangle \xrightarrow{\begin{pmatrix} 1&1&1&0&0&0 \\ 1&0&0&1&1&0 \\ 0&1&0&1&0&1 \\ 0&0&1&0&1&1 \end{pmatrix}} \mathbb{F}_2\langle e_{1'}, e_{2'}, e_{3'}, e_{4'}\rangle \xrightarrow{0} 0$$

Aligning these, we can add in the induced chain maps.



In the top complex,

$$\ker(d_1) = \mathbb{F}_2 \left\langle \begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\\1\end{bmatrix} \right\rangle \qquad \operatorname{im}(d_2) = 0$$

so

$$H_1(\Sigma) = \mathbb{F}_2 \left\langle \begin{bmatrix}\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}\end{bmatrix}, \begin{bmatrix}\begin{bmatrix}0\\1\\1\\0\\1\end{bmatrix}\end{bmatrix} \right\rangle$$

In the bottom complex,

$$\ker(d_1') = \mathbb{F}_2 \left\langle \begin{bmatrix}1\\1\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\1\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\\0\\1\end{bmatrix} \right\rangle \qquad \operatorname{im}(d_2') = \mathbb{F}_2 \left\langle \begin{bmatrix}1\\1\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\\1\\1\end{bmatrix} \right\rangle$$

In the quotient, the first vector in the basis of the kernel is in the image, so is set to zero. Further, the sum of the three basis vectors for the kernel is the other basis element for the image, so their sum is equal to zero in the quotient space. Since the first vector is already zero, that means the sum of the other two are – or, over $\mathbb{F}_2$, they are equal. Thusm

$$H_1(\Sigma') = \mathbb{F}_2 \left\langle \begin{bmatrix}\begin{bmatrix}1\\0\\1\\0\\1\\0\end{bmatrix}\end{bmatrix} \right\rangle$$

The map on homology is given by

$$(\phi_*)_1\left(\begin{bmatrix}\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}\end{bmatrix}\right) = \begin{bmatrix}(\phi_\#)_1\left(\begin{bmatrix}1\\0\\1\\1\\0\end{bmatrix}\right)\end{bmatrix}$$

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

$$(\phi_*)_1\left(\begin{bmatrix}\begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}\end{bmatrix}\right) = \begin{bmatrix}\begin{pmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix}\begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}\end{bmatrix}$$

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

In matrix form, using the ordered bases for homology given above, we have $(\phi_*)_1 = (1\ 1)$.

Even better, however, is the fact that moving from the world of simplicial complexes to the world of homology is compatible with our notion of sameness of maps.

Theorem 31. *Let Σ, Σ' be simplicial complexes, $f, g : \Sigma \to \Sigma'$ homomorphisms so that $|f| \simeq |g|$. Then $f_* = g_* : H_*(\Sigma) \to H_*(\Sigma')$.*

The same is true for general topological spaces with the appropriate notion of homology. We can apply this to conclude

Corollary 32. *Let Σ, Σ' be simplicial complexes. If $|\Sigma| \simeq |\Sigma'|$, then $H_*(\Sigma) \cong H_*(\Sigma')$.*

Proof. Exercise. □

The converse says that homology is a *homotopy invariant* – if two spaces have different homology, they cannot be homotopy equivalent.

The Snake Lemma

Computing homology directly is a lot of work – the vector spaces get quite large, quite quickly. Fortunately, homology is built in a way that facilitates building complicated computations up from simpler ones. One of the most useful of these is a kind of quotient space construction: given a simplicial complex Σ, select a subcomplex Σ' and define an equivalence relation on $|\Sigma|$ so that $x \sim y$ if $x, y \in |\Sigma'| \subseteq |\Sigma|$.

Definition. Let Σ be a simplicial complex, $\Sigma' \subseteq \Sigma$ a subcomplex with inclusion map ι. The *relative chain complex* $C_\bullet(\Sigma, \Sigma')$ is given by

$$C_k(\Sigma, \Sigma') = C_k(\Sigma)/(\iota_\#)_k(C_k(\Sigma'))$$

with differentials d_k^{rel} induced from d_k by $d_k^{\text{rel}}([e_\sigma]) = [d_k(e_\sigma)]$. The *relative homology* $H_*(\Sigma, \Sigma')$ is the homology of the relative chain complex.

The map $(\iota_\#)_k : C_k(\Sigma') \to C_k(\Sigma)$ simply maps basis vectors for simplices in Σ' to the corresponding basis vectors for simplices in Σ, and is clearly an injective map. There is also a surjective map $\pi_k : C_k(\Sigma) \to C_k(\Sigma, \Sigma')$ which quotients out elements in the image of $(\iota_\#)_k$ – that is, they are zero in the image. Indeed, they are precisely the elements which go to zero, so

$$0 \to C_k(\Sigma') \xrightarrow{(\iota_\#)_k} C_k(\Sigma) \xrightarrow{\pi_k} C_k(\Sigma, \Sigma') \to 0$$

is exact at $C_k(\Sigma)$. It is straightforward to check that this sequence is also exact at $C_k(\Sigma')$ and $C_k(\Sigma, \Sigma')$. A sequence of exactly three vector spaces which is exact at each position is called a *short exact sequence*. Since this is true for all k, and the maps commute with the differentials, we get a short exact sequence of chain complexes

$$0 \to C_\bullet(\Sigma') \xrightarrow{\iota_\#} C_\bullet(\Sigma) \xrightarrow{\pi} C_\bullet(\Sigma, \Sigma') \to 0.$$

Any longer sequence which is exact at every term is called a *long exact sequence*.

Lemma 33 (Snake lemma)**.** *Let*

$$0 \to C_\bullet \xrightarrow{f} C'_\bullet \xrightarrow{g} C''_\bullet \to 0$$

be a short exact sequence of chain complexes. This induces a long exact

sequence in homology

$$\begin{array}{ccccccc}
 & & \cdots & \xrightarrow{(g_*)_2} & H_2(C''_\bullet) & & \\
 & \overset{\delta_2}{\swarrow} & & & & & \\
H_1(C_\bullet) & \xrightarrow{(f_*)_1} & H_1(C'_\bullet) & \xrightarrow{(g_*)_1} & H_1(C''_\bullet) & & \\
 & \overset{\delta_1}{\swarrow} & & & & & \\
H_0(C_\bullet) & \xrightarrow{(f_*)_0} & H_0(C'_\bullet) & \xrightarrow{(g_*)_0} & H_0(C''_\bullet) & \longrightarrow & 0
\end{array}$$

Proof. The maps f_* and g_* are just the usual maps induced by a map on chain complexes. We need to work out what these new δ maps – called "connecting homomorphsims" – are, and show exactness.

Unlike our previous definitions, it's not immediately clear what the connecting map should be. Let's dig in to the object at hand.

$$\begin{array}{ccccccccc}
0 & \longrightarrow & C_{k+1} & \xrightarrow{f_{k+1}} & C'_{k+1} & \xrightarrow{g_{k+1}} & C''_{k+1} & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle d_{k+1}} & & \downarrow{\scriptstyle d'_{k+1}} & & \downarrow{\scriptstyle d''_{k+1}} & & \\
0 & \longrightarrow & C_k & \xrightarrow{f_k} & C'_k & \xrightarrow{g_k} & C''_k & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle d_k} & & \downarrow{\scriptstyle d'_k} & & \downarrow{\scriptstyle d''_k} & & \\
0 & \longrightarrow & C_{k-1} & \xrightarrow{f_{k-1}} & C'_{k-1} & \xrightarrow{g_{k-1}} & C''_{k-1} & \longrightarrow & 0
\end{array}$$

Suppose $\alpha \in \ker(d''_k) \subseteq C''_k$. Since g_k is surjective, there is $\beta \in C'_k$ with $g_k(\beta) = \alpha$. We can push β down through d'_k to get an element $d'_k(\beta) \in C'_{k-1}$. Now, use the fact that g is a map of chain complexes to obtain $g_{k-1}(d'_k(\beta)) = d''_k(g_k(\beta)) = d''_k(\alpha) = 0$. Since this is a short exact sequence, $\ker(g_{k-1}) = \operatorname{im}(f_{k-1})$, so there is $\gamma \in C_{k-1}$ with $f_{k-1}(\gamma) = d'_k(\beta)$. Define $\delta_k([\alpha]) = [\gamma]$.

Now we need to check that this is well-defined. Observe that the only choice we made was the lift under g_k, so suppose there are $\beta \neq \beta' \in C'_k$ with $g_k(\beta) = g_k(\beta') = \alpha$. Follow the process above to get $\gamma, \gamma' \in C_{k-1}$ respectively. Now, $\beta' - \beta \in \ker(g_k)$, so by exactness there is $\sigma \in C_k$ with $f_k(\sigma) = \beta' - \beta$. Apply the injective map f_{k-1} to $\gamma + d_k(\sigma)$, obtaining

$$\begin{aligned}
f_{k-1}(\gamma + d_k(\sigma)) &= f_{k-1}(\gamma) + f_{k-1}(d_k(\sigma)) \\
&= d'_k(\beta) + d'_k(f_k(\sigma)) \\
&= d'_k(\beta) + d'_k(\beta' - \beta) \\
&= d'_k(\beta') = f_{k-1}(\gamma').
\end{aligned}$$

But since f_{k+1} is injective, $\gamma' = \gamma + d_k(\sigma)$, so $[\gamma] = [\gamma'] \in H_k(C_\bullet)$.

Checking exactness is left as an exercise. □

Corollary 34 (Long exact sequence of a pair)**.** *Let Σ be a simplicial complex, Σ' a subcomplexes with inclusion map ι. There is a long exact sequence on homology*

$$\begin{array}{ccccccc} & & \cdots & \xrightarrow{(\pi_*)_2} & H_2(\Sigma, \Sigma') & & \\ & \overset{\delta_2}{\swarrow} & & & & & \\ H_1(\Sigma') & \xrightarrow{(\iota_*)_1} & H_1(\Sigma) & \xrightarrow{(\pi_*)_1} & H_1(\Sigma, \Sigma') & & \\ & \overset{\delta_1}{\swarrow} & & & & & \\ H_0(\Sigma') & \xrightarrow{(\iota_*)_0} & H_0(\Sigma') & \xrightarrow{(\pi_*)_0} & H_0(\Sigma, \Sigma') & \longrightarrow & 0 \end{array}$$

Definition. Let V, W be vector spaces. The *direct sum* $V \oplus W$ is the vector space whose elements are ordered pairs of vectors $(v, w), v \in V, w \in W$, with induced sum $(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_2 + w_2)$.

Corollary 35 (Mayer-Vietoris sequence)**.** *Let Σ be a simplicial complex, A, B subcomplexes. let $\iota_{X,Y}$ be the inclusion map from a subcomplex X into the larger complex Y. There is a long exact sequence on homology*

$$\begin{array}{ccccccc} & & \cdots & \longrightarrow & H_2(\Sigma) & & \\ & \overset{\delta_2}{\swarrow} & & & & & \\ H_1(A \cap B) & \xrightarrow{((\iota_{A\cap B,A}, \iota_{A\cap B,B})_*)_1} & H_1(A) \oplus H_1(B) & \xrightarrow{((\iota_{A,\Sigma} + \iota_{B,\Sigma})_*)_1} & H_1(\Sigma) & & \\ & \overset{\delta_1}{\swarrow} & & & & & \\ H_0(A \cap B) & \xrightarrow{((\iota_{A\cap B,A}, \iota_{A\cap B,B})_*)_0} & H_0(A) \oplus H_1(B) & \xrightarrow{((\iota_{A,\Sigma} + \iota_{B,\Sigma})_*)_0} & H_0(\Sigma) & \longrightarrow & 0 \end{array}$$

The Homology of Spheres

Let's write down some basic facts that will be useful when working with homology.

Lemma 36. *Let $V_1 \xrightarrow{f} V_2 \xrightarrow{g} V_3$ be a sequence of vector spaces which is exact at V_2. Then $dim(V_2) = dim(im(f)) + dim(im(g))$.*

Proof. By the Rank-Nullity theorem, $\dim(V_2) = \dim(\text{im}(g)) + \dim(\ker(g))$, and by exactness, $\dim(\text{im}(f)) = \dim(\ker(g))$. □

Lemma 37. *Let $0 \to V_1 \xrightarrow{f} V_2$ and $V_3 \xrightarrow{g} V_4 \to 0$ be sequences of vector spaces exact at V_1 and V_4 respectively. The f is injective and g is surjective.*

Proof. We have that $\ker(f) = \text{im}(0) = 0$ and $\text{im}(g) = \ker(0) = V_4$. □

Corollary 38. *Let W_1, W_2, W_3 be vector spaces.*

- *If* $0 \to W_1 \to 0$ *is an exact sequence,* $W_1 \cong 0$.
- *If* $0 \to W_1 \to W_2 \to 0$ *is an exact sequence,* $W_1 \cong W_2$.
- *If* $0 \to W_1 \xrightarrow{f} W_2 \xrightarrow{g} W_3 \to 0$ *is a short exact sequence,* f *is injective and* g *is surjective, and*

$$dim(W_2) = dim(W_1) + dim(W_3).$$

Proof. Exercise. □

Lemma 39. *Let* Σ, Σ' *be simplicial complexes, then for all* k*,*

$$H_k(\Sigma \sqcup \Sigma') \cong H_k(\Sigma) \oplus H_k(\Sigma).$$

Proof. Observe that $C_k(\Sigma \sqcup \Sigma) = \mathbb{F}_2\langle S_k \sqcup S'_k\rangle$, and if $\sigma \in S_k$, $d_k^{\sqcup}(e_\sigma) \in \mathbb{F}_2\langle S_{k-1}\rangle \subseteq C_{k-1}(\Sigma \sqcup \Sigma')$ (similarly for $\Sigma \in S'_k$). Thus, $\mathrm{im}(d_k^{\sqcup}) \cong \mathrm{im}(d_k^{\sqcup}|_{S_k}) \oplus \mathrm{im}(d_k^{\sqcup}|_{S'_k}) \cong \mathrm{im}(d_k) \oplus \mathrm{im}(d'_k)$, the latter being a decomposition into the individual chain complexes for Σ and Σ'. This works similarly for kernels, and thus also for homology. □

Let $Y \subseteq X$ be a subspace and denote by X/Y the quotient space obtained by identifying all points in Y to a single point. (That is, y $\sim$ y'\) for each $y, y' \in Y$.

Lemma 40. *Let* Σ *be a simplicial complex and* Σ' *subcomplex,* N *the number of path components of* Σ*, and* M *the number of path components of* Σ *containing a vertex from* Σ'*. Then* $H_0(\Sigma, \Sigma') \cong \mathbb{F}_2^{N-M}$.

In particular, if Σ' is a nonempty, connected subcomplex, $H_0(\Sigma, \Sigma') = \mathbb{F}_2^{\beta_0(\Sigma)-1}$. Note that this means the relative homology isn't quite the same at dimension zero as the usual way of computing homology – we're off by one dimension most of the time.

Proof. By definition, $C_0(\Sigma, \Sigma') \cong C_0(\Sigma)/C_0(\Sigma')$.. Each path component C_i of Σ containing simplices from Σ' has at least one vertex $v_i \in \Sigma'$. In the quotient, the equivlance class of each $[e_{v_i}]$, which correspond to those connected components, is equal to $[0]$. □

Now, let's do some homology computations using our lemmas. Let $\mathrm{pt} = (\{v\}, \{v\})$ be the unique one-vertex simplcial complex – that is, a point. The chain complex for the point is

$$0 \to C_0(\mathrm{pt}) \to 0$$

so the homology is

$$H_*(\mathrm{pt}) \cong \begin{cases} \mathbb{F}_2 & * = 0 \\ 0 & \text{else} \end{cases}$$

Building on that, we have

$$H_*(S^0) \cong H_*(\{-1,1\}) \cong H_*(\text{pt}) \oplus H_*(\text{pt}) \cong \begin{cases} \mathbb{F}_2^2 & * = 0 \\ 0 & \text{else} \end{cases}$$

Recall (from lecture and homework) that $D^k/S^{k-1} \simeq S^k$ and, since $S^{k-1} \simeq \partial(\Delta^k)$ and $\Delta^k \simeq D^k \simeq *$, $H_*(S^k) \cong H_*(\Delta^k, \partial(\Delta^k))$ so we can use the LES of a pair. Let's start with S^1.

Write

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & H_2(\Delta^1) & \xrightarrow{(\pi_*)_2} & H_2(\Delta^1, \partial(\Delta^1)) & & & & \\
 & & & \overset{\delta_2}{\searrow} & & & & & \\
H_1(S^0) & \xrightarrow{(\iota_*)_1} & H_1(\Delta^1) & \xrightarrow{(\pi_*)_1} & H_1(\Delta^1, \partial(\Delta^1)) & & & & \\
 & & & \overset{\delta_1}{\searrow} & & & & & \\
H_0(S^0) & \xrightarrow{(\iota_*)_0} & H_0(\Delta^1) & \xrightarrow{(\pi_*)_0} & H_0(\Delta^1, \partial(\Delta^1)) & \longrightarrow & 0 & &
\end{array}$$

and fill in the values we know

$$\begin{array}{ccccccc}
 & & 0 & \xrightarrow{(\pi_*)_2} & H_2(S^1) & & \\
 & & & \overset{\delta_2}{\searrow} & & & \\
0 & \xrightarrow{(\iota_*)_1} & 0 & \xrightarrow{(\pi_*)_1} & H_1(S^1) & & \\
 & & & \overset{\delta_1}{\searrow} & & & \\
\mathbb{F}_2^2 & \xrightarrow{(\iota_*)_0} & \mathbb{F}_2 & \xrightarrow{(\pi_*)_0} & 0 & \longrightarrow & 0
\end{array}$$

Since the first to vector spaces in each row will be zero all the way up after H_1, we have $0 \to H_k(S^1) \to 0$ for $k > 1$ and conclude that $H_k(S^1) \cong 0$ for $k > 1$. However, we are left with an interesting exact sequence:

$$0 \to H_1(S^1) \xrightarrow{\delta_1} \mathbb{F}_2^2 \xrightarrow{(\iota_*)_0} \mathbb{F}_2 \to 0$$

We can appeal to our lemma about short exact sequences above to get the answer immediately, but let's talk through the solution for practice. The rightmost map is zero, so the kernel is all of $\mathbb{F}_2$, thus by exactness, the image of $(\iota_*)_0$ is as well. By the Rank-Nullity theorem, $\dim(\ker((\iota_*)_0)) = \dim(\mathbb{F}_2^2) - \dim(\text{im}((\iota_*)_0)) = 2 - 1 = 1$. But by exactness, $\ker((\iota_*)_0) = \text{im}(\delta_1)$. Further, $\ker(\delta_1) = \text{im}(0) = 0$, so the map is injective. Thus, we conclude that

$$H_*(S^1) \cong \begin{cases} \mathbb{F}_2 & * = 0, 1 \\ 0 & \text{else} \end{cases}$$

Now on to S^2.

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & H_3(\Delta^2) & \xrightarrow{(\pi_*)_3} & H_3(\Delta^2, \partial(\Delta^2)) & & & & \\
& & & \overset{\delta_3}{\searrow} & & & & & \\
H_2(S^1) & \xrightarrow{(\iota_*)_2} & H_2(\Delta^2) & \xrightarrow{(\pi_*)_2} & H_2(\Delta^2, \partial(\Delta^2)) & & & & \\
& & & \overset{\delta_2}{\searrow} & & & & & \\
H_1(S^1) & \xrightarrow{(\iota_*)_1} & H_1(\Delta^2) & \xrightarrow{(\pi_*)_1} & H_1(\Delta^2, \partial(\Delta^2)) & & & & \\
& & & \overset{\delta_1}{\searrow} & & & & & \\
H_0(S^1) & \xrightarrow{(\iota_*)_0} & H_0(\Delta^2) & \xrightarrow{(\pi_*)_0} & H_0(\Delta^2, \partial(\Delta^2)) & \longrightarrow & 0 & &
\end{array}$$

and filling in, we have

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & 0 & \xrightarrow{(\pi_*)_3} & H_3(S^2) & & & & \\
& & & \overset{\delta_3}{\searrow} & & & & & \\
0 & \xrightarrow{(\iota_*)_2} & 0 & \xrightarrow{(\pi_*)_2} & H_2(S^2) & & & & \\
& & & \overset{\delta_2}{\searrow} & & & & & \\
\mathbb{F}_2 & \xrightarrow{(\iota_*)_1} & 0 & \xrightarrow{(\pi_*)_1} & H_1(S^2) & & & & \\
& & & \overset{\delta_1}{\searrow} & & & & & \\
\mathbb{F}_2 & \xrightarrow{(\iota_*)_0} & \mathbb{F}_2 & \xrightarrow{(\pi_*)_0} & 0 & \longrightarrow & 0 & &
\end{array}$$

Again, we conclude that $H_k(S^2) \cong 0$ for $k > 2$. Now we have two interesting pieces

$$0 \to H_2(S^2) \xrightarrow{\delta_2} \mathbb{F}_2 \to 0$$
$$0 \to H_1(S^2) \xrightarrow{\delta_1} \mathbb{F}_2 \xrightarrow{(\iota_*)_0} \mathbb{F}_2 \to 0$$

Now, we'll apply our corollary to both to obtain

$$H_*(S^2) \cong \begin{cases} \mathbb{F}_2 & * = 0, 2 \\ 0 & \text{else} \end{cases}$$

Note that this decomposition is going to happen again for S^3, this time with the top sequence in dimension 3. This continues inductively forever, giving us

Theorem 41. *Let $k > 0$, then*

$$H_*(S^k) \cong \begin{cases} \mathbb{F}_2 & * = 0, k \\ 0 & \textit{else} \end{cases}$$

Nash equilibria

Let's close this discussion with a pair of applications of the computation we just completed.

Theorem 42 (Brouwer fixed point theorem)**.** *Let $f : D^n \to D^n$ be a continuous function. Then f has a fixed point; e.g. there is $x \in D^n$ with $f(x) = x$.*

That is, when you stir your coffee, there's always a point on the surface of the liquid that's doesn't move between any pair of observations (which point, exactly, varies continuously with the Δt between measurements).

Proof. Suppose the converse, that f has no fixed point. Then we can use the function f to build another function $r : D^n \to S^{n-1}$ by taking $r(x)$ to be the point on the ray from $f(x)$ to x which intersects $S^{n-1} \subseteq D^n$. Because there are no fixed points, this ray is always well defined, and varying x (and thus $f(x)$ a small amount similarly varys the ray only slightly, so the map is continuous.

Further, if $x \in S^{n-1}$, then $r(x) = x - r(x)$ restricts to $\mathrm{id}_{S^{n-1}}$ on the boundary. In particular, writing ι for the subspace inclusion, $r \circ \iota = \mathrm{id}_{S^{n-1}} : S^{n-1} \to S^{n-1}$ is the identity map. Thus, the induced map $((r \circ \iota)_*)_{n-1} : H_{n-1}(S^{n-1}) \to H_{n-1}(D^n) \to H_{n-1}(S^{n-1})$ must equal the identity map $((\mathrm{id}_{S^{n-1}})_*)_{n-1} : H_{n-1}(S^{n-1}) = \mathbb{F}_2 \to H_{n-1}(S^{n-1}) = \mathbb{F}_2$. But $(\iota_{ast})_{n-1} : H_{n-1}(S^{n-1}) = \mathbb{F}_2 \to H_{n-1}(D^n) = 0$ must be the zero map, and thus so must the composite $((r \circ \iota)_*)_{n-1}$. Thus, no such map f can exist. □

We should finally stop and define a couple of things. Let X be a topological space and $Y \subseteq X$ a subspace. The *boundary* of Y in X, written $\partial_X Y$ (or just ∂Y when X is clear) is the set of all points $y \in X$ so that every open set U with $y \in U$ has both $U \cap Y \neq \varnothing$ and $U \setminus Y \neq \varnothing$ – that is, the boundary is all points so that every open neighborhood of the point contains points outside the set. The *interior* of Y is $\mathrm{int}(Y) = Y \setminus \partial Y$, and the *closure* of Y is $\mathrm{cl}(Y) = Y \cup \partial Y$. A set Y is closed if $Y = \mathrm{cl}(Y)$.

We want a slightly more general version of the Brouwer fixed point theorem.

Corollary 43. *Suppose $U \cong D^n$, then any map $f : U \to U$ has a fixed point.*

In particular, any closed, bounded, convex region $U \subseteq \mathbb{R}^n$ will work. To prove this corrolary, let $\varphi : U \to D^n$ be the homeomorphism and apply the Brouwer fixed point theorem to the continuous map $(\varphi \circ f \circ \varphi^{-1}) : D^n \to D^n$.

The most famous application of the Brower fixed point theorem comes from economics, specifically from game theory, where it led to a Nobel prize.

Definition. An *(n-player finite) game* is a triple $([n], A, u)$ where $[n]$ is the set of *players*, $A = \prod_{k=1}^{n} A_k$, $|A_k| < \infty$ is the set of *pure strategies*, and $u : A \to \mathbb{R}^n$ is the *utility function*.

Thus, for each $\vec{a} \in A$, a choice of pure strategies for each player, we assign to each player k a utility $u_k(\vec{a})$ based on what everyone chose to do. It is usually more useful to think in terms of probabilities – what is the likelihood that player k employs strategy a_k?

Definition. Let X be a finite set. Write $\mathcal{P}(X)$ for the *set of probability distributions* on X,

$$\mathcal{P}(X) = \{p : X \to \mathbb{R} \mid p(x) \geq 0 \text{ for all } x \in X, \sum_{x \in X} p_i(x) = 1\}.$$

This looks awfully familiar: each p is just a list of non-negative real numbers which sum to 1. In fact, $\mathcal{P}(X) = \Delta^{|X|-1}$.

Definition. The set of *mixed strategies* for a game ([n], A, u) is the set $M = \prod_{k=1}^{n} M_k$ where $M_k = \mathcal{P}(A_k)$ are the mixed strategies for player k.

Thus, $M = \prod_{k=1}^{n} \Delta_k^{|A_k|-1}$. Observe that this is a convex set of dimension $\sum_{k=1}^{n}(|A_k| - 1)$.

We need a way to measure how these mixed strategies affect utility. To do so, we'll ask what the average outcome for each player is, assuming all players make their strategy choices indepdendently.

Definition. The *expected utility for player k under a mixed strategy* $\vec{p} \in M$ is

$$\bar{u}_k(\vec{p}) = \sum_{\vec{a} \in A} \left(u_k(a) \prod_{k=1}^{n} p_k(a_k) \right).$$

This above expression can be rewritten in terms of the pure strategy choices for player k under uncertainty about the other player's decisions, producing (by slight abuse of notation)

$$\bar{u}_k(\vec{p}) = \sum_{a_k \in A_k} \left(\bar{u}_k(p_1, \ldots, p_{k-1}, a_k, p_{k+1}, \ldots p_n) p_k(a_k) \right).$$

Definition. Write $\vec{p}_{-k} = (p_1, \ldots, \widehat{p_k}, \ldots p_n)$. A *best response* for player k to $\vec{p}_{-k}$ is $p_k^* \in M_k$ so that $\bar{u}_k(p_1, \ldots, p_k^*, \ldots p_n) \geq \bar{u}_k(\vec{p})$ for all $\vec{p} \in M$. A *Nash equilibrium* for a game is a mixed strategy $\vec{p} \in M$ which is a best response to $\vec{p}_{-k}$ for each player $k \in [n]$.

Theorem 44 (Nash, 1951)**.** *Every finite, n-player game has a Nash equilibrium.*

Proof. We begin by defining functions

$$\begin{aligned}
\text{Gain}(\vec{p}, k, a_k) &= \max\left(0, \bar{u}_k(p_1, \ldots, p_{k-1}, a_k, p_{k+1}, \ldots, p_n) - \bar{u}_k(\vec{p})\right) \\
g(\vec{p}, k, a_k) &= p_k(a_k) + \text{Gain}(\vec{p}, k, a_k) \\
C(\vec{p}, k) &= \sum_{a_k \in A_k} g(\vec{p}, k, a_k) = 1 + \sum_{a_k \in A_k} \text{Gain}(\vec{p}, k, a_k) \\
f(\vec{p}, k, a_k) &= \frac{g(\vec{p}, k, a_k)}{C(\vec{p}, k)}
\end{aligned}$$

which respectively approximate the directional derivative along strategy changes, that derivative translated to the point in question, the gradient and gradient ascent in this context. We think of each of these as a function of $\vec{p}$, so, for example, $f(\vec{p}, -, -) : M \to M$.

Observe that $C \geq 1$

Now, under the discrete dynamics induced by iterating f on M, the existence of fixed points is immediate from Brouwer, since we have already observed that M is a closed, convex, bounded subset of Euclidean space. We need to check that fixed point are Nash Equilibria.

Suppose not: $f(\vec{p}^*, -, -) = \vec{p}^*$, but this is not a best response for some player. Then there are k and a_k so that $\text{Gain}(\vec{p}^*, k, a_k) > 0$, and thus $C(\vec{p}^*, k) > 1$. Now, unpacking f, and solving for the gain at the fixed point, we have

$$\begin{aligned}
\vec{p}^* &= f(\vec{p}^*, -, -) \\
&= \frac{g(\vec{p}^*, -, -)}{C(\vec{p}^*, -)} \\
&= \frac{\vec{p}^* + \text{Gain}(\vec{p}^*, -, -)}{C(\vec{p}^*, -)} \\
\text{Gain}(\vec{p}^*, -, -) &= (C(\vec{p}^*, -) - 1)\vec{p}^*
\end{aligned}$$

Thus, for any k and a_k with non-zero gain at $\vec{p}^*$, the gain is a positive multiple of $\vec{p}^*$.

Finally, we compute how this affects utility

$$\begin{aligned}
0 &= \bar{u}_k(\vec{p}^*) - \bar{u}_k(\vec{p}^*) \\
&= \sum_{a_k \in A_k} \bar{u}_k(p_1^*, \ldots, p_{k-1}^*, a_k, p_{k+1}^*, \ldots p_n^*) p_k^*(a_k) \\
&\quad - \Big(\sum_{a_k \in A_k} \bar{u}_k(\vec{p}^*) p_k^*(a_k) \\
&= \sum_{a_k \in A_k} \left(\bar{u}_k(p_1^*, \ldots, p_{k-1}^*, a_k, p_{k+1}^*, \ldots p_n^*) - \bar{u}_k(\vec{p}^*) \right) p_k^*(a_k) \\
&= \sum_{a_k \in A_k} \text{Gain}(\vec{p}^*, k, a_k) p_k^*(a_k) \\
&= \sum_{a_k \in A_k} (C(\vec{p}^*, k) - 1)(p_k^*(a_k))^2 > 0
\end{aligned}$$

Thus, every fixed point must be a Nash Equilibrium. □

Topology of Data

When we move to the realm of science and engineering one of the fundamental changes from that of mathematics is that we only get to measure things, rather than assuming we know what we're studying – to put it another way, we move from topological spaces to a finite sample of points from those spaces. However, with a bit of cleverness, we can still use our topological tools to recover information about the underlying spaces in a sensible way.

Data, clustering, and dendrograms

The concept of *data* is much subtler than one might assume from first impressions. This course is, at its core, about uncovering some of those subtleties, and so we'd better start by centering ourselves a bit. Since we don't have time to dive deep into topics from probability and statistics, we'll have to skirt certain important details. However, we won't let that scare us away from thinking big.

Definition. Let X be a topological space. A *point cloud* in $X = (P, c)$ is a finite set $P = \{p_1, \dots p_n\}$ and a map $c : P \to X$; that is, a finite collection of (possibly non-distinct) points in X. A *proximity measure* on X is a real-valued function $\mu : X \times X \to \mathbb{R}_{\geq 0}$, and the *$\mu$-proximity matrix* for P is $(M(P)_{i,j}) = \mu(x_i, x_j)$.

Proximity measures usually come in one of two types: *dissimilarity measures*, which take large values when two objects are dissimilar, as with distances, and *similarity* measures, which behave inversely, as with correlations.

The most common context in applications is for X to be some Euclidean space $\mathbb{R}^d$, or some subspace thereof, μ to be the Euclidean metric, a dissimilarity measure, on that space. This usually corresponds to some collection of d measurements taken independently or as a whole for several objects – spatial coordinates, physical properties, sale prices, time stamps, etc. It is important to note that the choice of μ as the Euclidean distance is not always sensible, particularly when we have several different types of unrelated data being by the coordinates. For example, we can think of vectors in $\mathbb{R}^d$ as being time series, in which case pointwise dissimilarity may not indicate overall differences – consider sampling $\sin(x)$ and $\cos(x)$ at a range of x values, for example. In this case, it is often saner to consider a similarity measure like correlation or coherence.

One of the most common problems in scientific and engineering applications is as follows.

Problem. Let X be a topological space representing the "ground truth" *object of study*, and L a topological space of *labels or features*. Suppose there is a continuous map $\ell : X \to L$, providing a labeling $\ell(x)$ to each point in X by some feature; this feature is what we want to recover, predict or understand. However, we do not have access to X or ℓ; rather, we can *select* a point cloud (P, c) in X, and, via a measurement function $r : X \to Y$, push that point cloud into a topological space Y of measurements, equipped with a proximity measure μ. Given some observations of the point cloud $(P, (\ell \circ c))$ and the point cloud $(P, r \circ c)$ (or the proximity matrix $M(r \circ c(P))$, how much of ℓ can we recover?

We tend to believe that r should be a continuous function, but we also usually assume that measurements are noisy – that we add some small, random amount of error to the measurement. Thus, in reality r isn't even a function, but rather a sample from a probability distribution. However, in well-behaved cases it is "nearly a continuous function", so we can try to make reasonable inferences in spite of things going a little wrong.

Let $A = \coprod_{i=1}^{k} A_i$ for some path-connected topological spaces A_i and $L = \pi_0(A)$ with the natural map $\ell : A \to L$ given by $\ell(a) = [a]$. The problem of recovering $\hat{\ell} \approx \ell$ from $r(S)$ is called *clustering*.

Suppose $r : A \to \mathbb{R}^d$ is also constant on each component, so $r(a) = r_i \in \mathbb{R}^d$ for all $a \in A_i$, but that we add noise to each evaluation. If $\mu(r_i, r_j)$ is much larger/smaller than the magnitude of the noise, it is usually possible to recover the distinction between ℓ_i and ℓ_j using standard techninques like k-means or linkage clustering, support vector machines, etc. For concreteness, let's work with (single-)linkage clustering under a dissimilarity measure, where $\hat{\ell}(r(s_i)) = \hat{\ell}(r(s_j))$ if $\mu(r(s_i), r(s_j)) \leq t$ for some $t \in \mathbb{R}_{\geq 0}$, extended by transitivity. These techniques extend nicely to $r(a)$ being non-constant, so long as the images of the components are well-separated by μ; in this case, kernel-based or deep-learning methods are often employed to overcome the potential geometric complexity introduced by r.

However, as separation decreases or happens across multiple scales, it often becomes less clear when to give two points distinct labels. One common way to deal with this is *hierarchical clustering*, where we consider the outcome of our chosen clustering algorithm as we vary the criterion for when to assign distinct labels to different subsets. Thus, we obtain a continuous sequence of imputed labelings of points, $\hat{\ell}_t$. If we draw the codomain of $\hat{\ell}_t$ as a function of t, we obtain a tree called a *dendrogram* or *merge tree*, which describes the proximity (or other parameter) at which the number of connected components changes. Each component of

the level set of the tree at a given proximity corresponds to a cluster – this is the point cloud version of a Reeb graph for a space Q with a function $f : Q \to \mathbb{R}_{\geq 0}$ for which the level sets are $f^{-1}(t) = \bigcup_{i=1}^{m} \{x \in X \mid \min(\mu(r(s_i), x), \mu(x, r(s_i))) < t\}$.

Restricting ourselves to a discrete label space provides a lot of structure, and thus provides powerful tools, but is extremely limiting. If we replace our label space by $L = \mathbb{R}^d$ – suppose that each object is labeled by its weight or position, for example – a range of techniques like principal component analysis (for maps where a linear $\hat{\ell}$ make sense) and multidimensional scaling (where $\hat{\ell}$ can be non-linear) are employed. In exchange, we have to assume certain things about X, usually that it is $\mathbb{R}^D$, $D >> d$, with μ the standard Euclidean distance, but we can get a very good recovery of the lower-dimensional labels if this is all true, however these come at the cost of distortion of distances and sometimes other feature loss.

Mapper

We want to understand the organizational structure underlying the point cloud $r(S)$. Clustering algorithms tend to collapse a lot of the detail in $r(S)$ while they're looking for component labels – this detail may tell us a great deal of interesting information about A and ℓ. Thus, we want to develop methods that retain that information.

One powerful tool is the *Mapper* algorithm. Mapper disassembles the point cloud into smaller pieces, and uses clustering in those individual pieces as building blocks for a low-dimensional visualization of the geometry structure of the point cloud.

Definition. Let $r(S) \subset X$ be a point cloud and fix a function $f : X \to \mathbb{R}$. Let $\mathcal{U} = \{U_i = (a_i, b_i)\}_{i=1}^{m}$ be a cover of $\mathbb{R}$ by overlapping open intervals $U_i = (a_i, b_i)$ with $a_1 = -\infty, b_m = \infty$ and $U_i \cap U_j \neq \varnothing$ if and only if $|i - j| \leq 1$. Write $r(S)_i = f^{-1}(U_i) \cap r(S)$. Select a clustering algorithm, and for each $r(S)_i$, perform the algorithm to obtain clusters $C_i^1, \ldots C_i^{p_i}$. The *mapper graph for f and $\mathcal{U}$* is the graph with vertices given by the collection of clusters for each U_i, $V = \{C_i^p\}$, and an edge $C_i^p C_j^q$ when $C_i^p \cap C_j^q \neq \varnothing$.

More generally, we can take $f : X \to Y$ for any topological space Y and use an open cover of Y to build a combinatorial structure. In this case, we build a *mapper complex* with simplices for each non-empty intersection of clusters.

Mapper is implemented commercially by its inventors through a company called Ayasdi, in a package called Ayasdi Core. The software offers a range of powerful visualization tools for the mapper

graph, as well as statistical tools for understanding the structure of the data set via manipulation of cluster

Simplicial complexes from data

Another option, the most common one in topological data analysis, is to use the points in $r(S)$ and the proximity function μ to build a simplicial complex directly. There are several possibilities.

Definition. Let $\epsilon > 0$ and P be a point cloud. The Čech complex for P at scale ϵ is the simplicial complex $\check{C}_\epsilon(P)$ with vertices P and simplices $\sigma \subseteq P$ whenever $\bigcap_{x\in\sigma} B_\epsilon(x) \neq \varnothing$.

The Čech complex is precisely the nerve of the canonical open cover by epsilon balls $B_\epsilon(x)$ of the space $\bigcup_{x\in P} B_\epsilon(x)$. Therefore, by the nerve theorem the simplicial complex built in this way recovers the homotopy type and, thus, has the proper homology for this union of local neighborhoods. If we were lucky enough to choose ϵ so that the structure of this union reflects the structure of interest, this is fantastic for us.

Unfortunately, the Čech complex is hard to compute – it involves looking at intersections of sets, which is an expensive task computationally. If we are fortunate enough to have μ symmetric, we can approximate the Čech complex using only the matrix $M(r(S))$.

Definition. Let μ a symmetric proximity function on X, $\epsilon > 0$ and P be a point cloud. The Vietoris-Rips complex for P at scale ϵ is the simplicial complex $VR_\epsilon(P)$ with vertices P and simplices $\sigma \subseteq P$ whenever $M(r(S))_{i,j} < \epsilon$ for all $i \neq j \in \sigma$.

The Vietoris-Rips complex is the clique complex of the graph $\Gamma_\epsilon(P)$ with vertices P and edges whenever the proximity of two points is less than ϵ. This saves the effort of checking for intersections in the ambient space, reducing the problem to finding cliques in a graph – which is still a hard computational problem, but more approachable in general.

In the special case when μ is a metric[35] (so X is a metric space), we have the following guarantee that there is a relationship between the Vietoris-Rips and Čech complexes.

[35] A *metric* is a function $\mu : X \times X \to \mathbb{R}_{\geq 0}$ for which (i) $\mu(x,y) = \mu(y,x)$, (ii) $\mu(x,y) = 0$ only if $x = y$ and (iii) $\mu(x,y) \leq \mu(x,z) + \mu(y,x)$ for every z.

Lemma 45. *Let $\epsilon > 0$ and P a point cloud in a metric space X, then*

$$\check{C}_\epsilon(P) \subseteq VR_\epsilon(P) \subseteq \check{C}_{2\epsilon}(P).$$

Proof. Exercise. □

Thus, if we want to approximate one of these complexes by another, we can "bound" the structure on either end by shifting ϵ.

Finally, suppose we have a collection of *witnesses* $\{w_1, \dots w_m\} \subseteq X$, and rather than measuring distance between the points in P, we have measurements of distance to these witnesses. Witnesses model sensors with limited range in a space.

Definition. Let P be a point cloud in X and $\{w_1, \dots w_m\} \subset X$ a collection of witnesses, $\epsilon > 0$. The *witness complex* $W_\epsilon(P)$ at scale ϵ is the simplicial complex with vertices P and simplices $\sigma \in W_\epsilon(P)$ if $\mu(x_i, w_j) < \epsilon$.

The witness complex is a special case of a general construction of a simplicial complex for rectangular matrices called a *Dowker complex*.

Definition. Let M be an $n \times m$ matrix, $\epsilon > 0$. The *Dowker complex* of M at scale ϵ, $\mathrm{Dow}_\epsilon(M)$ is the simplicial complex with vertices $[n]$ and a face $\sigma_j, j = 1 \dots m$ for each column given by $i \in \sigma_j \Leftrightarrow M_{ij} < \epsilon$.

The process for constructing a Dowker complex may result in repeated addition of faces to the complex; this is not a problem because faces form a set. The witness complex is just the Dowker complex at scale ϵ for the matrix with rows corresponding to points in the cloud, columns corresponding to witnesses, and entries corresponding to their proximity.

The choice to make rows into vertices and columns into faces seems arbitrary, and it is. We could consider instead the alternative complex with roles reversed, which would just be $\mathrm{Dow}(M^T)$. These complexes are called Dowker complexes because of his remarkable observation about how this choice doesn't matter.

Theorem 46 (Dowker's theorem)**.** *Let M be a matrix, $\epsilon > 0$. Then $H_*(Dow_\epsilon(M)) \cong H_*(Dow_\epsilon(M^T))$.*

Persistent homology

Of course, all of these constructions depend on a choice of scale parameter ϵ, and we definitely don't want to have to make such a choice about data we don't yet understand. Fortunately, we can simply decide not to make a choice.

Write S_ϵ for any choice of $\check{C}_\epsilon(P), VR_\epsilon(P)$, or $\mathrm{Dow}_\epsilon(M)$. Choose a new scale paramter $\epsilon' > \epsilon$, and observe that $S_\epsilon \subseteq S'_\epsilon$ in each case: the vertices remain the same, but we might introduce some new higher simplices. Thus, there is a canoncial inclusion map

$$\iota_{\epsilon,\epsilon'} : S_\epsilon \hookrightarrow S_{\epsilon'}$$

which is the identity on the vertices. If we follow this down to the level of homology, we get

$$(\iota_{\epsilon,\epsilon'})_* : H_*(S_\epsilon) \to H_*(S_{\epsilon'}).$$

Further, observe that since there are finitely many points in P or elements in M, ϵ-balls will form new intersections or new vertices will be added to simplices in $\mathrm{Dow}_\epsilon(M)$ only finitely many times. Thus, the structure of S_ϵ can only change finitely many times as ϵ increases. We record the parameters at which this structure changes as a list $-\infty = t(0) < t(1) < \cdots < t(k) < t(k+1) = \infty$, and observe that $S_\epsilon = S_{\epsilon'}$ if $\epsilon, \epsilon' \in (t(i), t(i+1)]$, and thus $(\iota_{\epsilon,\epsilon'})_*$ is an isomorphism, so the only interesting information is contained in the maps $(\iota_{t(i),t(i+1)})_* : H_*(S_{t(i)}) \to H_*(S_{t(i+1)})$.

Definition. Let $S_{t(0)}, S_{t(1)}, \ldots, S_{t(k+1)}$ be simplicial complexes so that $\iota_{t(i),t(i+1)} : S_{t(i)} \hookrightarrow S_{t(i+1)}$ is an injective homomorphism. Such a sequence S of simplicial complexes is called a *filtration*. For $j > i$, the $(t(i), t(j))$*-persistent homology in degree* p of the filtration is

$$H_p^{t(i)\to t(j)}(S) = \mathrm{im}((\iota_{t(i),t(j)})_p).$$

If $[x] \in H_p(S_{t(i)}), [x] \neq [0]$ and $[x] \notin H_p^{t(i-1)\to t(i)}(S)$, we say $[x]$ is *born* at $t(i)$ and write $b_{[x]} = t(i)$. If $[x] \in H_p(S_{t(i)}), [x] \neq [0]$ is born at $t(i)$, $[\iota_{t(i),t(j)}(x)] \neq [0] \in H_p^{t(i)\to t(j)}(S)$ but $[\iota_{t(i),t(j+1)}(x)] = [0] \in H_p^{t(i)\to t(j+1)}(S)$ then we say $[x]$ *dies* at $t(j)$ and write $d_{[x]} = t(j)$. The *lifetime* of $[x]$ is $\ell_{[x]} = d_{[x]} - b_{[x]}$.

Barcodes and persistence diagrams

Let's investigate this last collection of terminology a bit. Cycles are elements of a vector space, which are at some point *born* and then *die*. That these are discrete events that apply to each cycle suggests that we should be able to understand the persistent homology of a filtered complex through such a lens.

Select $[x] \in H_p(S_{t(i)}), [x] \neq [0]$ born at $t(i) = b_{[x]}$ and dying at $t(j) = d_{[x]}$[36]. Write $[x_{t(k)}] = (\iota_{t(i),t(k)})_*([x]) \in H_p(S_{t(k)})$ and consider the maximal sequence $([x_{t(i)}], [x_{t(i+1)}], \ldots [x_{t(j)}])$ of non-zero cycles in the image of $[x_{t(i)}]$ under the induced maps in homology. Without loss of generality, assume that $b_{[x]}$ is minimal among all possible choices of $[x]$.

Since each of these is a non-zero element, we can consider their spans as a sequence of subspaces of the appropriate homology groups

$$\langle [x_{t(k)}] \rangle = \mathbb{F}_2 \langle [x_{t(k)}] \rangle \subseteq H_p(S_{t(k)}).$$

By definition, the restriction

$$(\iota_{t(k),t(k+1)})_p|_{\langle [x_{t(k)}] \rangle} : \langle [x_{t(k)}] \rangle \to \langle [x_{t(k+1)}] \rangle$$

has matrix (1). Without loss of generality, we can assume that all other basis vectors for each homology group are contained in the

[36] In doing this, we've made a choice and added information to the system. A different choice of $[x]$ may result in different induced bases in the homology groups of each complex in the filtration, which complicates the notion of following cycle representatives through the homology at the level of complexes.

orthogonal subspaces $\langle [x_{t(k)}] \rangle^{\perp}$, so the persistent homology decomposes as

$$\begin{array}{ccccccccccccc}
\cdots & \xrightarrow{0} & 0 & \xrightarrow{0} & \langle [x_{t(i)}] \rangle & \xrightarrow{1} & \langle [x_{t(i+1)}] \rangle & \xrightarrow{1} & \cdots & \xrightarrow{1} & \langle [x_{t(j)}] \rangle & \xrightarrow{0} & 0 \xrightarrow{0} \cdots \\
 & & \oplus & & \oplus & & \oplus & & & & \oplus & & \oplus \\
\cdots & \longrightarrow & H_p(S_{t(i-1)}) & \longrightarrow & \langle [x_{t(i)}] \rangle^{\perp} & \longrightarrow & \langle [x_{t(i+1)}] \rangle^{\perp} & \longrightarrow & \cdots & \longrightarrow & \langle [x_{t(j)}] \rangle^{\perp} & \longrightarrow & H_p(S_{t(j+1)}) \longrightarrow \cdots
\end{array}$$

Write $\mathrm{Int}_{b_{[x]}, d_{[x]}}$ for the *($(b_{[x]}, d_{[x]})$-interval module* given by the sequence of 1-dimensional $\mathbb{F}_2$ vector spaces indexed by $t(i), t(i+1), \ldots, t(j)$ with identity maps between consecutive spaces, as in the first row of the diagram above, and 0-dimensional vector spaces outside that range.

We can inductively apply this argument to the second row of the diagram to split off a sequence of interval modules[37]. Because each homology group is finite dimensional, this process will eventually terminate, resulting in a decomposition of the form

$$H_p(S) \cong \bigoplus_{a=1}^{M} \mathrm{Int}_{b_{[x_a]}, d_{[x_a]}}.$$

[37] This is directly analogous to the Gram-Schmidt process for constructing an orthonormal basis for a vector space, and is a special case of the classification of finitely generated modules over a PID.

Thus, the degree p persistent homology of a filtered simplicial complex can be summarized by a collection of (possibly repeated) pairs, $((b_{[x_a]}, d_{[x_a]}))_{a=1}^{M}$, of birth and death indices for a collection of *generating cycles* $GC(S) = \{[x_a]\}_{a=1}^{M}$ in $H_p(S)$.

There are two common ways to visualize this data, *barcodes* and *peristence diagrams*. Both are useful for understanding the structure of persistent homology, so we will present and use both.

Definition. Let S be a filtered simplicial complex with $H_p(S) \cong \bigoplus_{a=1}^{M} \mathrm{Int}_{b_{[x_a]}, d_{[x_a]}}$. The *degree p persistence diagram for S* is $D(S) = (\mathrm{im}(P), \mu)$, where the function $P : GC(S) \to \mathbb{R}^2$ is given by $P([x_a]) = (b_{[x_a]}, d_{[x_a]})$, and the multiplicity function $\mu : \mathrm{im(P)} \to \mathbb{N}_{>0}$ is given by $\mu((b,d)) = |P^{-1}((b,d))|$.

We draw a persistence diagram by plotting the image and labeling points with their multiplicity. Generically, each pair $(b_{[x_a]}, d_{[x_a]})$ is distinct, so we suppress any label of 1. Also, it is necessarily the case that $0 < b_{[x_a]} < d_{[x_a]} < \infty$, so these points are contained in the upper half of the first quadrant.

Definition. Let S be a filtered simplicial complex with $H_p(S) \cong \bigoplus_{a=1}^{M} \mathrm{Int}_{b_{[x_a]}, d_{[x_a]}}$. The *degree p barcode for S* is the collection of line segments joining the points $(b_{[x_a]}, \frac{1}{a})$ and $(d_{[x_a]}, \frac{1}{a})$, $a = 1, \ldots, M$.

It is common to sort the intervals to make the barcode visually clearer, either in terms of increasing birth or death times.

Distance and stability for persistence diagrams

Suppose we compute persistent homology for a pair of filtrations. How do we know if they're "similar"? Given our ability to summarize the structure using diagrams or barcodes, the popular answer is as one might expect: assign a notion of distance to the persistence diagrams.

Naively, we would like to compute this distance by building on some notion of distance between paired points, for which we would need a bijection (with multiplicity) of the points in the two diagrams. However, there is no reason to assume that there are the same number of points. To fix this, we note that in a persistence diagram, points (b, d) for which $\ell = d - b$ is small are "close to vanishing" – they could reasonably be deleted, because the corresponding cycle only appears for a short while in the persistent homology. Thus, we can reasonably pair points with the nearest point on the line $b = d$.

Definition. Let $D_i = (\mathrm{im}(P_i), \mu_i)$, $i = 1, 2$, with $P_i : GC_i \to \mathbb{R}^2$ be persistence diagrams. A *partial matching* for D_1 and D_2 is a choice of subsets $M_i \subseteq GC_i$, $i = 1, 2$ along with a bijection $\phi : M_1 \to M_2$.

A partial matching tells us which points to pair with points from the other diagram; those unmatched in either diagram will be matched to the nearest points on the $(b = d)$ line.

We will use the so-called ∞-norm to measure distance: the distance between two points is the largest amount they differ in a single coordinate. That is, $||(b, d)||_\infty = \max(|b|, |d|)$. In this context, the ∞-norm of a difference of points measures larger of the discrepancies between the birth and death time of the two cycles, which is perhaps more natural than a Euclidean norm-style combination of the two.

Definition. Let D_1 and D_2 be persistence diagrams. Let $\phi : M_1 \to M_2$ be a partial matching for D_1 and D_2, and define

$$d_\phi(D_1, D_2) = \max\{||P_1([x_a]) - P_2(\phi([x_a]))||_\infty \mid [x_a] \in M_1\}$$

and

$$d_{\phi^C}(D_1, D_2) = \max\{\ell_{[x_a]} \mid [x_a] \in (GC_1 \setminus M_1) \cup (GC_2 \setminus M_2)\}.$$

The *bottleneck distance* between D_1 and D_2 is

$$d_B(D_1, D_2) = \min_\phi \max\{d_\phi(D_1, D_2), d_{\phi^C}(D_1, D_2)\}.$$

Here, d_ϕ is the largest difference in the birth or death time of a matched generating cycle, and d_{ϕ^C} is the largest lifetime of an unmatched cycle (or, equivalently, the difference in birth or death to

the nearest point on the $b = d$ line). The bottleneck distance arises from selecting the matching with the smallest such discrepancy.

The bottleneck distance has one fundamentally nice property that makes it useful in the context of data analysis: perturbing the input data slightly only perturbs the diagram slightly in terms of d_B. First, we need the "right" notion of distance between point clouds, so we know what a small perturbation is.

Definition. Let $P = \{p_i\}_{i=1}^N, Q = \{q_j\}_{j=1}^M$ be point clouds in $\mathbb{R}^d$. The *Hausdorff distance* from P to Q is

$$d_H = \max\{\max_{p_i \in P} \min_{q_j \in Q} d(p_i, q_j), \max_{q_j \in Q} \min_{p_i \in P} d(p_i, q_j)\}.$$

That is, for each point in P (resp Q), we find the closest point in Q (resp P), and we take the largest such distance for all points in P (resp Q). The larger of these is the Hausdorff distance between the point clouds.

Theorem 47 (Stability theorem)**.** *Let P, Q be point clouds in $\mathbb{R}^d$, $\check{C}(P), \check{C}(Q)$ the Čech complexes, and $D(P)$ and $D(Q)$ the corresponding degree p persistence diagrams. Then*

$$d_B(D(P), D(Q)) \leq d_H(P, Q)$$

Thus, if we "wiggle" the points in a point cloud by ϵ, the diagrams will only change by at most ϵ.

Summary statistics for persistent homology

The bottleneck distance is provides a solid theoretical foundation for understanding persistence, as a notion of stability is absolutely necessary for any serious application: if small perturbations to the data created large changes in the persistence diagram, it would be impossible to interpret answers. However, the resulting geometry on the space of persistence diagrams is not as straightforward as we might like it to be.

Example.

Let $D_1 = \{(2,4), (2,6)\}$ and $D_2 = \{(1,5), (3,5)\}$. There are two partial matchings between D_1 and D_2 which induce the bottleneck distance: matching the cycle at $(2,4)$ to that at $(1,5)$, and $(2,6)$ to $(3,5)$ gives the correct bottleneck distance of 1, as does matching $(2,4)$ to $(3,5)$ and $(2,6)$ to $(1,5)$. We could use either matching to give a sensible notion of the "mean" of these two diagrams[38], and those two choices disagree.

[38] In particular, the Fr'echet mean, which is a choice of a centroid point for each matched pair which minimizes the total variance.

If we can't even take means of diagrams, we are in trouble with regard to doing statistical analyses using persistent homology. One of the most promising ways to fix this problem is to embed the diagrams in a function space where such problems can be easily overcome. The down-side is that we move even further away from the persistence module, often losing most or all of the underlying topological structure. Nonetheless, the trade-off often provides discriminatory power that is useful in applications. We'll consider three such approaches.

Definition. Let $H_p(S)$ be the pth persistent homology of a filtered simplicial complex. The *pth Betti curve* for S is the piecewise constant function $\beta_p(t) : \mathbb{R} \to \mathbb{R}_{\geq 0}$ given by $\beta_p(t) = \beta_p(S_{t(i)})$ for $t \in (t(i), t(i+1)]$.

The Betti curves of a complex provide a portrait of the number of classes that are present at any given filtration, but throw away information on which cycles are persisting from filtration to filtration – for example, imagine a structure with a few very long lived cycles having the same Betti curves as one with very many short-lived cycles that don't overlap much. However, as an aggregate measure of "how much homology" is present, they provide a useful summary. Further, it is straightforward to compute means (pointwise on the t-axis), norms (standard integral norms), variance/standard deviation and other statistics of curves. If we don't want to throw away lifetime information for individual cycles, we can use a richer embedding.

Definition. Let $H_p(S)$ be the pth persistent homology of a filtered simplicial complex, $GC(S)$ a collection of generating cycles for S. Transform the birth-death coordinates $(b_{[x_a]}, d_{[x_a]})$ for each $[x_a] \in GC(S)$ into $(m_{[x_a]}, h_{[x_a]}) = (\frac{1}{2}(d_{[x_a]} + b_{[x_a]}), \frac{1}{2}(d_{[x_a]} - b_{[x_a]}))$ coordinates, and define $R_{[x_a]} \subseteq \mathbb{R}^2$ to be the triangular region bounded by the m-axis, the slope 1 line segment $(b_{[x_a]}, 0)$ to $(m_{[x_a]}, h_{[x_a]})$, and the slope -1 line segment $(d_{[x_a]}, 0)$ to $(m_{[x_a]}, h_{[x_a]})$. Define a function $r(m,h) = |\{[x_a] \mid (m,h) \in R_{[x_a]}\}|$ which counts how many regions R a point lies inside, and define a sequence of functions $\lambda_k(m) = \sup_h\{r(m,h) \leq k\}$, which record the "top boundary" of the region of the plane with $r(m,h) \leq k$. The *pth persistence landscape* for S is $\Lambda(S) = \{\lambda_k(m) \mid k = 1, 2, \dots\}$.

Since only finitely many regions exist for any choice of S, the λ_k are uniformly zero for $k >> 0$. Thus, we can assume that for any finite family of diagrams there is a maximum k and truncate the sequence of functions at that value. Now, as with the Betti curves, we can compute statistics on each of the individual functions λ_k pointwise as a function of m, or using statistical measures build specifically for functions.

Landscapes retain a great deal more information about the lifetimes of individual cycles at the price of being much larger and potentially computationally cumbersome. They also fail to have a fixed size: the maximum k might vary from data set to data set, which may in turn complicate input into machine learning tools, etc. The following is a potential remedy.

Definition. Let $H_p(S)$ be the pth persistent homology of a filtered simplicial complex, $GC(S)$ its persistence diagram. Transform the birth-death coordinates $(b_{[x_a]}, d_{[x_a]})$ for each $[x_a] \in GC(S)$ into $(b_{[x_a]}, \ell_{[x_a]}) = (b_{[x_a]}, (d_{[x_a]} - b_{[x_a]}))$ coordinates. Let $g(b, \ell)$ be a differentiable *kernel* function with $\int_{\mathbb{R}^2} g = 1$ and mean $(0, 0)$, and $f(b, \ell)$ a differentiable weighting function which is zero when $\ell \leq 0$. The *persistence surface* for D and g is $\rho_D(b, \ell) = \frac{f(b,\ell)}{|GC(S)|} \sum_{[x_a] \in GC(S)} g(b - b_{[x_a]}, \ell - \ell_{[x_a]})$; with mild abuse of notation, this is just the convolution $D * g$. Given discretization parameters Δb and $\Delta \ell$, and a rectangular region $R = (B, B + k\Delta b) \times (L, L + m\Delta \ell)$ of the (b, ℓ)-plane, the *persistence image* of D and g on $(R, \Delta x)$ is the $(k \times m)$ matrix $PI(S)$ with

$$PI(S)_{i,j} = \int_{B+(i-1)\Delta b}^{B+i\Delta b} \int_{L+(j-1)\Delta \ell}^{L+j\Delta \ell} \rho_D d\ell db.$$

Thus, we divide the region of interest into rectangular 'pixels' and take the total value of the persistence surface on each pixel as a proxy for the number of "nearby" points in the persistence diagram. The weighting function ensures that adding points with short lifetimes does not cause arbitrary jumps in the value of the surface, preserving stability. For a fixed choice of region and discretization parameters, this approach allows the comparison of diagrams with arbitrary numbers of points as a vector of fixed size, which is then appropriate input for many machine learning algorithms.