Math 872 Algebraic Topology

Running lecture notes

First a word from our sponsor...

Algebraic topology is an umbrella term for that part or topology which uses algebraic tools to study and answer topological problems. The most basic problem in topology is, given two topological spaces X and Y, to determine whether or not they are homeomorphic. Other basic questions are to understand the different ways (if any) that one space can be embedded in another (i.e., the ways in which X can be homeomorphic to a subspace of Y), and to understand the continuous maps from X to Y. In all of these tools from alebraic topology have a role to play.

At its heart the idea to assign to each topological space (in some reasonable collection (read: category) of spaces that you are interested in) an algebraic object (group, ring, field, module...) in some intelligent way. Typically, to tie the construction to the topology on X, the object is built using continuous functions into or out of X. That is, after all, what a topology is really good for; it tells you what maps are continuous, i.e., aren't ripping your space apart. A construction which "really" only uses the topology on X (and not something more) will have the property that homeomorphic spaces have isomorphic objects assigned to them (you can almost take this as a definition of "really using the topology"). Then the algebraic objects can be used to distinguish spaces; if the objects aren't isomorphic, then the spaces they came from can't be homeomorphic. The basic idea is that distinguishing groups from one another is "easier" than distinguishing spaces; whether or not this is really true we will discuss at a later date! But, for example, finitely generated abelian groups are easy to distinguish (when given to you as a direct sum of cyclic groups, for example), which can be turned into a method for distinguishing spaces, when our method assigns such groups to spaces.

This kind of process can tell us that two spaces are different, if the *algebraic invariants* that we assign to the spaces are different, that is enough. But it doesn't run the other way; if the algebraic invariants are the same, we cannot conclude that the spaces are the same. (Dumb example: we assign to every topological space the field with two elements. Homeomorphic spaces have the same associated field, but...) But this doesn't stop us from continuing to try to find invariants that continue to do better at distinguishing spaces....

In this course we will explore two basic approaches to building algebraic invariants: homotopy theory and homology theory. Each builds a sequence of groups (all but one of them, the first homotopy group, or *fundamental group*, are abelian) which serve as algebraic invariants of the space. History has shown us that the homotopy groups, typically, are more powerful; they are better at distinguishing spaces. They pay for this, however by being more difficult to compute in practice. For example, there is no general formula for the homotopy groups of the 2-sphere S^2; all that is generally known is that all but two are finite, and all but one (?) are non-trivial. The first few hundred, probably, have actually been computed. We will focus mostly on the fundamental group $\pi_1(X)$; its computation, properties, and applications. The fundamental group has found its way into a wide variety of mathematical fields (essentially, anywhere that continuity has?).

Homology groups, on the other hand, are typically "easier" to compute, at least once you have gotten some theory out of the way! They pay for this by being less adept at distinguishing spaces. At least straight out of the box; a lot of effort has been invested in finding more subtle ways to use the techniques of homology theory to winkle ever more detailed information out of a topological space. Homology groups are designed to be abelian (for the higher homotopy groups this is more a matter of some delightful accident); literally, they are each the quotient of a free abelian group by a subgroup. So the main computational tool is some fairly straighforward linear algebra.

Ultimately our goal is to construct these two theories, explore their properties, and then use them to prove some topological results that, in the end, one might never have guessed that algebraic techniques would have played a role in proving. Some sample results:

There is a (continuous) map from the surface of genus 3, Σ_3 to the surface of genus 2, Σ_2, such that every point inverse is finite. There is no such map $\Sigma_2 \to \Sigma_3$ (or $\Sigma_4 \to \Sigma_3$ or...). One can in fact give a precise statement of when such a map $\Sigma_n \to \Sigma_m$ exists; it is that $m-1|n-1$.

The real projective plane $\mathbb{R}P^2$ cannot embed in $\mathbb{R}^3$.

Invariance of Domain: If $U \subseteq \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^n$ is continuous and injective, then $f(U) \subseteq \mathbb{R}^n$ is open. (I.e., being a *domain*, an open subset of $\mathbb{R}^n$, is invariant under continuous injections.)

There is one topological fact which we will use constantly which you might not have seen in your point-set topology class (although it may have come up in an analysis class?): the Lebesgue Number Theorem. If (X, d) is a compact metric space (in our applications, it is always a compact subset of Euclidean space), and $\{\mathcal{U}_i\}$ is an open cover of X, then there is an $\epsilon > 0$ so that for every $x \in X$, its ϵ-neighborhood $N_d(x, \epsilon)$ is contained in $\mathcal{U}_i$ for some i. For if not, then for every $n \in \mathbb{N}$ there is an $x_n \in X$ whose $1/n$-neighborhood is contained in no $\mathcal{U}_i$; that is, for every $i \in I$, there is an $x_{n,i}$ with $d(x_n, x_{n,i}) < 1/n$ and $x_{n,i} \notin \mathcal{U}_i$, so $x_{n,i} \in C_i = X \setminus \mathcal{U}_i$, a closed set. But since X is compact, there is a convergent subsequence of the x_n; $x_{n_k} \to y \in X$. [Proof: if not, then no point is the limit of a subsequence, so for every $x \in X$ there is an $\epsilon(x) > 0$ and an $N = N(x)$ so that $n \geq N$ implies $x_n \notin N_d(x, \epsilon(x))$. But these neighborhoods cover X, so a finite number of them do; for any n gretaer than the maximum of the associated $N(x)$'s x_n lies in none of the neighborhoods, a contradiction, since $x_n \in X =$ the union of theose neighborhoods.] But then since $d(x_{n_k}, y) \to 0$ and $d(x_{n_k}, x_{n_k,i}) \to 0$, for every i the $x_{n_k,i}$ also converge to y; since the $x_{n_k,i}$ all lie in the closed set C_i, so does y. So $y \in C_i$ for all i, so $y \notin \mathcal{U}_i$ for all i, a contradication, since the $\mathcal{U}_i$ cover X. So some $\epsilon > 0$, a *Lebesgue number* for the covering, must exist.

The fundamental group:

One of the main themes of topology is that a space X can be studied by looking at maps (= continuous functions) of "useful" spaces into X. Maps provide the means to explore a space, and literally map out the topology of the space. The most basic useful space to use is an interval $I = [0, 1]$; understanding how intervals map into a space provides information on the many ways to get from "here" to "there". This is the basis for the fundamental group.

The key to understanding the fundamental group is to understand how you could conceivably turn maps of intervals (i.e., *paths*) into a group, i.e., how to multiply them. That is, how do you take two maps $f, g : I \to X$ and produce a single map $f \cdot g : I \to X$? After fussing a bit, you would probably hit upon what Poincaré did; two intervals glued end to end form an interval; so two maps glued end to end build a map. *If* you can do the gluing; that requires $f(1) = g(0)$ in order to be well-defined. In order to multiply any two paths in any order (so as to form a group) we need to impose a compatibility condition, that is we need to assume that all paths have compatible endpoints, so we focus on loops, that is paths $f : [0, 1] \to X$ so that $f(0) = f(1) = x_0 \in X$ for some fixed basepoint $x_0 \in X$. These can be concatenated in any order;

$$f * g(t) = \begin{cases} f(2t), & \text{if } t \leq 1/2 \\ g(2t-1) & \text{if } t \geq 1/2 \end{cases}$$

runs across f first, and then g. This gives us a mutliplication on the loops at x_0. But this really can't be turned into a group; we could never find an appropriate identity element, since the product of f with anything contains a copy of f in it, so our identity would have to have a copy of every loop in it?

The solution is to define our group elements to be equivalence classes of loops, so that our identity element can "have" a copy of every map f in it! The other point is that by making elements of the group "big" (having lots of representatives), this will make the group 'smaller", and more manageable. Finally, by letting many loops really be the "same", we can focus on more important, global, features of a space rather than inessential local information. The equivalence relation we use is homotopy, that us, continuous deformation. In general, two maps $f, g : Y \to X$ are homotopic if there is a map $H : Y \times I \to X$ such that $H(y, 0) = f(y)$ and $H(y, 1) = g(y)$ for every $y \in Y$. That is, the map $f \coprod g : Y \times \{0, 1\} \to X$ extends to a map on $Y \times I$. There is also a notion of homotopy for a map of pairs $f : (Y, B) \to (x, A)$ (that is, $f(B) \subseteq A$), requiring that H is also a map of pairs, $H(B \times I) \subseteq A$. A loop at x_0 is really a map of pairs $f : (I, \partial I) \to (X, \{x_0\})$, and the elements of the fundamental group $\pi_1(X, x_0)$ will be equivalence classes of loops at x_0, under the equivalence relation of homotopy as maps of pairs. We of course need to show that homotopy is an equivalvence relation, that is, $f \simeq f$, if $f \simeq g$ then $g \simeq f$, and if $f \simeq g$ and $g \simeq h$ then $f \simeq h$. For each of these it is fairly straightforward to build the required homotopy ($H(t, s) = f(t)$, $K(t, s) = H((t, 1 - s)$, and $L(t, s) =$ the concatenation of two homotopies H and K, on the second variable; the Pasting Lemma assures its continuity).

With this dealt with, elements are equivalence classes $[f]$ of loops at x_0, we define our multiplication to be $[f] \cdot [g] = [f * g]$. For this to be well-defined, we need to check that if $f \simeq f'$ and $g \simeq g'$, then $f * g \simeq f' * g'$, but the required homotopy can be built by concatenating the hypothesized homotopies, along the *first* variable. But now, since the elements are so big, we can construct a meaningful identity element, and verify that we under this multiplication we have a group. The identity element is the loop which does nothing, that is, the equivalence class containing the constant map c_0 at x_0; $e = [c_0]$. The inverse $[f]^{-1}$ is $[\overline{f}]$, where $\overline{f}(t) = f(1-t)$ is f run in the reverse direction. The verifications $f * c_0 \simeq f \simeq c_0 * f$ and $f * \overline{f} \simeq c_0 \simeq \overline{f} * f$ can be verified by building the appropriate homotopies:

$$H(t,s) = \begin{cases} f(2t/(s+1)), & \text{if } t \leq (s+1)/2 \\ x_0, & \text{if } t \geq (s+1)/2 \end{cases} \qquad H(t,s) = \begin{cases} x_0, & \text{if } t \leq s/2 \\ f((t-s/2)/(1-s/2)), & \text{if } t \geq s/2 \end{cases}$$

$$H(t,s) = \begin{cases} x_0, & \text{if } t \leq s/2 \\ f(2t-s), & \text{if } s/2 \leq t \leq 1/2 \\ f(2-s-2t), & \text{if } 1/2 \leq t \leq 1-s/2 \\ x_0, & \text{if } 1-s/2 \leq t \leq 1 \end{cases} \qquad H(t,s) = \begin{cases} x_0, & \text{if } t \leq (1-s)/2 \\ f(2-s-2t), & \text{if } (1-s)/2 \leq t \leq 1/2 \\ f(2t-s), & \text{if } 1/2 \leq t \leq (1+s)/2 \\ x_0, & \text{if } (1+s)/2 \leq t \leq 1 \end{cases}$$

These give us that $g \cdot e = g = e \cdot g$ and $g \cdot g^{-1} = e = g^{1} \cdot g$ for every $g \in \pi_1(X, x_0)$. Associativity, $g \cdot (h \cdot k) = (g \cdot h) \cdot k$, by another homotopy: if $g = [\alpha], h = [\beta]$ and $k = [\gamma]$, then $\alpha * (\beta * \gamma) \simeq (\alpha * \beta) * \gamma$ via the homotopy

$$H(t,s) = \begin{cases} \alpha(4t/(2-s)), & \text{if } t \leq (2-s)/4 \\ \beta(4t-2+s), & \text{if } (2-s)/4 \leq t \geq (3-s)/4 \\ \gamma((4t-3+s)/(s+1)), & \text{if } t \geq (3-s)/4 \end{cases}$$

The Pasting Lemma assures that all of these maps are continuous. All of these homotopies are probably best understood pictorially, looking at what is happening in each individual region of definition.

So with this (well-defined) multiplication, we have an associative product on $\pi_1(X, x_0)$ with an identity and two-sided inverse, making $\pi_1(X, x_0)$ a group, the *fundamental group of X based at x_0*. This group literally enumerates the number of different ways to walk around the space X and return to our starting point, where two ways are different if one cannot be continuously deformed to the other. The fact that this collection of different ways together form a group under concatenation provides extra structure, giving us a better chance to be able to compute this object when we need to.

But just jumping in and computing it turns out to be rather difficult, at least straight from the definition. How do you really decide when two loops are homotopic? What we need to do is to erect a theory around our basic definitions, to give us a way to work with them and to illuminate their importance and utility.

Several basic properties are important to the utility of the fundamental group. The first is that if $f : (X, x_0) \to (Y, y_0)$ is a map of pairs, then there is an induced homomorphism $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$ given by $f_*[\gamma] = [f \circ \gamma]$. Since $\gamma \simeq \beta$ implies $f \circ \gamma \simeq f \circ \beta$ via $K(t,s) = f(H(t,s))$, the map is well-defined, and since $f \circ (\gamma * \beta) = (f \circ \gamma) * (f \circ \beta)$, it is a homomorphism. Further, $(f \circ g)_* = f_* \circ g_*$ follows directly, as does $(I_X)_* = I_{\pi_1(X)}$. From which it follows that if $f : X \to Y$ is a homeomorphism, then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism. So homeomorphic spaces have isomorphic fundamental groups; the fundamental group is a homeomorphism invariant.

In many instances, we supress the basepoint x_0 in our work with the fundamental group. The basis for doing this is that if $x_0, x_1 \in X$ and $\gamma : I \to X$ is a path in X with $\gamma(0) = x_0$ and $\gamma(1) = x_1$, then given a loop f at x_0, $\overline{\gamma} * f * \gamma$ is a loop at x_1. The map $F = F_\gamma : \pi_1(X, x_0) \to \pi_1(X, x_1)$ given by $F([f]) = [\overline{\gamma} * f * \gamma]$ is well-defined and provides an isomorphism between the two groups, with inverse $F^{-1}([g]) = [\gamma * g * \overline{\gamma}]$. (The relevant homotopies are readily constructed.) So for a path-connected space, the fundamental groups based at any two points are isomorphic. So we can sensibly talk about <u>the</u> fundamental group of a path-connected space, without mentioning basepoints. (There needn't be a *canonical* isomorphism between the groups; a different path between two basepoints might yield a different isomorphism. This is occasionally a *very* important consideration!)

An important combination of these two facts allows us to understand the behavior of induced maps under homotopy. If $f, g : X \to Y$ are homotopic maps, via a homotopy H, and x_0 is a baspoint in X, let $\gamma(t) = H(x_0, t)$ be the path in Y traced out by x_0 under the homotopy, and set $y_0 = \gamma(0), y_1 = \gamma(1)$. Then $f : (X, x_0) \to (Y, y_0)$ and $g : (X, x_0) \to (Y, y_1)$ as maps of pairs, and we have an isomorphism $F_\gamma : \pi_1(Y, y_0) \to \pi_1(Y, y_1)$. Then we can see that $g_* = F_\gamma \circ f_* : \pi_1(X, x_0) \to \pi_1(Y, y_1)$, by properly "reparametrizing" the homotopy H: given a loop $\alpha : (I, \partial I) \to (X, x_0)$, $g \circ \alpha \simeq \overline{\gamma} * (f \circ \alpha) * \gamma$ via the homotopy $K(s, t) = H(\alpha(t), s)$ together with a map of the square $I \times I$ to itself which takes the top edge to itself and stretches the bottom edge around the remaining three edges (thereby taking the two vertical edges each to the endpoints of the top edge). Such a map is readily constructed, although a formula for it might be a bit ugly...

This has special meaning when one of f, g is a homeomorphism (think: the identity); then since the homeo induces an iso in π_1 and F_γ is an iso, the other map induces an iso on π_1, as well. Another special case is when the homotopy between f and g is basepoint-preserving; that is, γ is a constant map. Then $F_\gamma = I_{\pi_1}$, since we can smooth out the constant maps we pre- and post-append to a given loop to return us to the given loop, as we did above in verifying that the constant loop is the identity element in pi_1. So basepoint-preserving homotopic maps induce the same map on π_1.

This line of thought reaches its logical conclusion with the introduction of homotopy equivalences. Two spaces X, Y are homotopy equivalent if there are maps $f : X \to Y$ and $g : Y \to X$ so that $f \circ g : Y \to Y$ and $g \circ f : X \to X$ are both homotopic to the identity. It is straighforward (as in the discussion of homotopy above) that "are homotopy equivalent" is an equivalence relation (hence the name). Each of f and g are called homotopy equivalences. The discussions above combine to give the result: if $f : X \to Y$ is a homotopy equivalence, then $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism. This is because the compositions $f_* \circ g_* = (f \circ g)_*$ and $g_* \circ f_* = (g \circ f)_*$ are isomorphisms, so the first is surjective (hence f_* is surjective) and the second is injective (so f_* is injective). So homotopy equivalent spaces have isomorphic fundamental groups.

A subset $a \subseteq X$ is a *retract* of X if there is a map $r : X \to A$ so that $r(a) = a$ for all $a \in A$. That is, for $\iota : A \to X$ the inclusion map, $r \circ \iota = I_A$. A is a *deformation retract* of X if in addition $\iota \circ r : X \to X$ is homotopic to the identity. r is a *strong deformation retraction* if this homotopy leaves every element of A fixed; $H(z, t) = a$ for all $a \in A$. If r is a deformation retraction, then it is a homotopy equivalence (we take the identity homotopy for $r \circ \iota$). So the inclusion $\iota : A \to X$ induces an isomorphism on π_1. This idea simplifies many computations, by allowing us to compute $\pi_1(A)$ instead of $\pi_1(X)$. FOr example, $\mathbb{R}^n$ deformation retracts to the origin $x_0 = 0$; the homotopy $H(x,t) = tx$ interpolates between the identity and $\iota \circ r = c_{x_0}$. Since $\pi_1(x_0) = \{1\}$ (there is only one map $\gamma : I \to \{x_0\}$), we deduce that $\pi_1(\mathbb{R}^n) = 1$. The same is true for disks D^n. More generally, a space is called *contractible* if it is homotopy equivalent to a point; every contractible space has trivial fundamental group.

A loop is map $\gamma : I \to X$ with $\gamma(0) = \gamma(1) = x_0$; it therefore decends to a well-defined map of the quotient space $I/0, 1 \cong S^1$ to X, and so a loop can be thought of as a map from the unit circle $(S^1, 1) \to (X, x_0)$. Elements of $\pi_1(X, x_0)$ could have been defined as equivalence classes of such maps under homotopy of pairs (it is not difficult to see how to lift such a homotopy to a homotopy $(I \times I, \partial I \times I) \to (X, x_0)$). The multiplication is a little more convoluted to work out; the reader is invited to do so. From this perspective, though, understanding what the identity element looks like has a more geometric feel: $\gamma : S^1 \to X$ represents the identity in $\pi_1(X) \Leftrightarrow$ γ extends to a map $\Gamma : D^2 \to X$ (i.e., $\Gamma|_{\partial D^2} = \gamma$), where D^2 is the unit disk in $\mathbb{R}^2$. The basic idea is that if γ is trivial, there is a homotopy $S^1 \times I \to X$ which on $S^1 \times \{0\}$ is γ and which sends $Y = \{1\} \times I \cup S^1 \times \{1\}$ to x_0. The homotopy descends to a map from $S^1 \times I$, with Y crushed to a point, to X. But $S^1 \times I/Y$ is homeomorphic to D^2, with $S^1 \times \{0\}$ being sent to ∂D^2; the composition $D^2 \to S^1 \times I/Y \to X$ is the desired extension.
In a similar vein, two paths $\alpha, \beta : I \to X$ joining the same pair of points $x_0, x_1 \in X$ are homotopic rel endpoints (i.e., the maps $(I\partial I) \to (X, \{x_0, x_1\})$ are homotopic as maps of pairs) $\Leftrightarrow$ the loop $\alpha * \overline{\beta}$ is trivial in $\pi_1(X, x_0)$. (The extension to D^2 is built from the homotopy by crushing each of the vertical boundary segments to points.) So, for example, in a contractible space, any two paths between the same two points are homotopic rel endpoints.

The fundamental group of the circle:

Our first really non-trivial computation is to determine the fundamental group of the circle S^1. Since S^1 is path-connected, the answer is independent of basepoint, so we will choose $x_0 = (1, 0) \in S^! \subseteq \mathbb{R}^2$. First the answer: $\pi_1(S^1) \cong \mathbb{Z}$. The proof is a little involved, but it will introduce several basic ideas that will become central to the development of our more general theory.
The basic idea is, given an element $[\gamma] \in \pi_1(S^1)$, to find a (more or less) canonical representative $\alpha \in [\gamma]$, and show that these canonical representatives can be put into one-to-one correspondence with $\mathbb{Z}$. The idea is to cover S^1 by a pair of contractible open sets, which we can take to be $\mathcal{U}_- = \{(x, y) \in S^1 : y < \epsilon\}$ and $\mathcal{U}_+ = \{(x, y) \in S^1 : y > -\epsilon\}$ for some small $\epsilon > 0$. (These are slightly larger than the lower and upper semicircles in S^1.) Given a loop $\gamma : I \to S^1$, the sets $\mathcal{V}_\pm = \gamma^{-1}(\mathcal{U}_\pm)$ form an open cover of the compact metric space I, so, by the Lebesgue Number Theorem, there is a $\delta > 0$ so that every interval of length δ in I lies in either $\mathcal{V}_-$ or $\mathcal{V}_+$ Choose an $n > 1/\delta$ and cut I into n subintervals of equal length; then each (closed) subinterval (has length less than δ so) is mapped by γ into either $\mathcal{U}_-$ or $\mathcal{U}_+$ (or both); choose one, assigning a $+$ or $-$ to each subinterval. If two

adjacent subintervals have the same sign, amalgamate them into a single larger interval. Continuing in this fashion, we arrive at a collection of subintervals whose signs alternate. The endpoints of the intervals (since they really have both signs) map into $\mathcal{U}_- \cap \mathcal{U}_+$= two short intervals.

Now we start "straightening" γ. We have I cut into subintervals $I_1, \dots, I_k$, each mapping into $\mathcal{U}_\pm$. These subsets are contractible, so any pair of paths between the same two points are homotopic. Put differently, we can replace $\gamma|_{I_j}$ with any other path between the same two points, to obtain a new map homotopic, rel endpoints, to γ, yielding a new representative of $[\gamma]$. This is our basic simplification process; the new path we continually choose is the arc (in $\mathcal{U}_\pm$) between the points. So for each subinterval make this switch, yielding a new loop (because the endpoints are unaffected) homotopic to γ, which we still call γ. If any of the subintervals maps into $\mathcal{U}_- \cap \mathcal{U}_+$, then we can switch its sign and amalgamate it with its neighboring subintervals, producing fewer subintervals, and repeat the process. So eventually (read: by induction) we reach a stage where each subinterval "crosses" $\mathcal{U}_\pm$; each has its endpoints in different components of $\mathcal{U}_- \cap \mathcal{U}_+$. (There is a degenerate case where the entire interval I lies in $\mathcal{U}_- \cap \mathcal{U}_+$; this implies that our altered γ is the constant map, which is our canonical form for the identity element...) By inserting short paths from the endpoints of our subintervals to $(1,0)$ or $(-1,0)$ (whichever one is in the component of $\mathcal{U}_- \cap \mathcal{U}_+$ containing our endpoint) and back, and straightening, we may assume the our endpoints map to $(\pm 1, 0)$. Finally, a reparametrization of the interval makes all of the subintervals we now have of the same length $I_j = [j/m, (j+1)/m]$, making $\gamma|_{I_j}$ one of precisely four maps: reparametrizing I_j as I, for convenience, they are $\alpha_1 : t \mapsto (\cos(\pi t), \sin(\pi t))$, $\overline{\alpha_1} : t \mapsto (-\cos(\pi t), \sin(\pi t))$, $\alpha_2 : t \mapsto (-\cos(\pi t), -\sin(\pi t))$, $\overline{\alpha_2} : t \mapsto (\cos(\pi t), -\sin(\pi t))$. ($\alpha_1$ and α_2 map counterclockwise around the top and bottom of the circle, respectively; their reverses go clockwise.) Any occurances of $\alpha_1 * \overline{\alpha_1}$,$\overline{\alpha_1} * \alpha_1$, $\alpha_2 * \overline{\alpha_2}$,$\overline{\alpha_2} * \alpha_2$ may be replaced by the constant map (since these loops are null-homotopic) and amalgamated away, and the combintations $\alpha_1 * \overline{\alpha_2}$,$\overline{\alpha_2} * \alpha_1$, $\alpha_2 * \overline{\alpha_1}$,$\overline{\alpha_1} * \alpha_2$ cannot occur because in each the two paths do not share endpoints properly. So after this further amalgamation the only posibilities are c_{x_0} (the degenerate case), $(\alpha_1 * \alpha_2)^n$, or $(\overline{\alpha_1 * \alpha_2})^n = (\overline{\alpha_2} * \overline{\alpha_1})^n$ (where we must have an even number of factors by basepoint considerations; we have loops). These are our canonical forms. **If** these canonical forms are unique (no two of them are homotopic), then we can construct our isomorphism $\pi_1(S^1) \to \mathbb{R}$ by sending $[c_{x_0}] \mapsto 0$, $(\alpha_1 * \alpha_2)^n \mapsto n$, and $(\overline{\alpha_1 * \alpha_2})^n \mapsto -n$. That this is an isomorphism follows by inspection.

Provided we show uniqueness! To do this, we use another technique which we will exploit much further later on. Essentially, we wish to show that for $n \neq m$, the maps $t \mapsto (\cos(2\pi mt), \sin(2\pi mt))$ and $t \mapsto (\cos(2\pi nt), \sin(2\pi nt))$, which is what our canonical forms really turn out to be, represent distinct elements of $\pi_1(S^1)$, i.e., are not homotopic to one another. To do this, we introduce the *winding number*; Given a loop $\gamma : I \to S^1$, we look at the covering $\mathcal{U}_{x+} = \{(x,y) \in S^1 : x > 0\}$,$\mathcal{U}_{x-} = \{(x,y) \in S^1 : x < 0\}$,$\mathcal{U}_{y+} = \{(x,y) \in S^1 : y > 0\}$,$\mathcal{U}_{y-} = \{(x,y) \in S^1 : y < 0\}$ and choose a partition $x_0 = 0, < x_1 < \cdots < x_k = 1$ of I so that $\gamma|_{[x_i, x_{i+1}]}$ maps into one of the $\mathcal{U}$'s (by Lebesgue number). Then for each i let θ_i = the angle (strictly between $-\pi$ and π) between the rays from the origin through $\gamma(x_i)$ and $\gamma(x_{i+1})$ (measured from the first to the second). Finally, let $w(\gamma) = \sum \theta_i / 2\pi$.

There are three things to prove:

(1) $w(\gamma)$ is independent of the partition used to compute it. This is a standard trick: given two partitions, show that both compute the same number as the union of the two partitions. This follows by showing the the number doesn't change by adding a single extra point to a partition (which is immediate: don't change the $\mathcal{U}$'s you assign (noting that it makes no difference to the angle computation if you switch between two such that can be assigned), and note that angles add).

(2) If $\gamma \simeq \beta$, then $w(\gamma) = w(\beta)$. This is also a standard sort of argument; the basic idea is that a homotopy can be thought of as a long sequence of "small" homotopies. Given the homotopy $H : I \times I \to S^1$, a Lebesgue number argument, applied to the same cover above, implies that there is an $\epsilon > 0$ so that every square with side $< \epsilon$ maps into one of the four sets. Then choose an $n > 1/\epsilon$ and partition $I \times I$ into n rows of n squares, each of which map into one of the four sets. Let the corners of these squares be denoted (x_i, x_j). If we let $\theta_{i,j}$ denote the angle between (x_i, x_j) and (x_{i+1}, x_j) and $\varphi_{i,j}$ the angle between (x_i, x_j) and (x_i, x_{j+1}), then since the four corners of a square are all in the same set and angles add, we find that $\theta_{i,j} + \varphi_{i+1,j} = \varphi_{i,j} + \theta_{i,j+1}$, so $\theta_{i,j} + \varphi_{i+1,j} - \varphi_{i,j} = \theta_{i,j+1}$. Summing both sides over i, most of the left terms telescope, and since $\varphi_{0,j} = \varphi_{n,j} = 0$ (since these lie on the vertical sides, where the homotopy is constant), we find that $w(H|_{I\times\{x_j\}}) = \sum \theta_{i,j} = \sum \theta_{i,j+1} = w(H|_{I\times\{x_{j+1}\}})$ So, by induction, $w(\gamma) = w(H|_{I\times\{x_0\}}) = w(H|_{I\times\{x_{n+1}\}}) = w(\beta)$.

(3) Each of our canonical forms have different winding numbers. This is immediate; they can be computed to be 0, n, and $-n$, respectively.

Together these facts imply that $\pi_1(S^1) \cong \mathbb{Z}$. Note that, in fact, The map $[\gamma] \mapsto w(\gamma)$ is our isomorphism, but we couldn't know that without both parts of the argument. The second part could be extended to show that this map is a homomrphism, and onto, but the first part is needed to show that it is injective, i.e.., loops with the same winding number are both homotopic to the same canonical form.

This is a very fundamental (pardon the pun) computation in homotopy theory, and a great deal can be proved just this one fact. It is also the basis for nearly every other fundamental group calculation that we will do. Spheres and disk make up the basic building blocks for the topological spaces which we will study in this course, and as we shall see the circle is the only one of these whose fundamental group is non-trivial, and so the fundamental group of every space is founded upon the circles that are built into its initial construction. To formalize this, we need to understand how that fundamental group of a space can be assembled out of the fundamental groups of the pieces used to build it. The basic idea, formalized in the Seifert-van Kampen Theorem, is that if $X = A \cup B$, and we understand the fundamental groups of A, B, and $A \cap B$, then we can compute $\pi_1(X)$ from this. But before embarking on this line of thought, let us first put our computation $\pi_1(S^1) \cong \mathbb{Z}$ to work.

One of the standard results of calculus is that the intermediate value theorem implies that every map $f : I \to I$ has a fixed point: $f(x_0) = x_0$ for some $x_0 \in I$. This has a higher-dimensional analogue:

Brouwer Fixed Point Theorem: Every map $f : D^2 \to D^2$ has a fixed point. Proof: If not, then we can construct a retraction $r : D^2 \to \partial D^2$ by sending $x \in D^2$ to the point on ∂D^2 lying on the ray from $f(x)$ to x (this uses the hypothesis that $f(x) \neq x$); such a ray intersects the boundary in exactly one point. A little analytic geometry will allow you to write down a formula for this map, which uses only elementary operations and the function f, so r is continuous. But a retraction induces a surjective homomorphism on π_1, so r_* is a surjection from $\pi_1(D^2) = 1$ to $\pi_1(\partial D^2) = \pi_1(S^1) = \mathbb{Z}$, a contradiction. So f must have a fixed point.

Another quick result using $\pi_1(S^1) \cong \mathbb{Z}$ is the **Fundamental Theorem of Algebra:** Every non-constant polynomial (with complex coefficients) has a complex root: for every $f(z) = a_n z^n + \cdots + a_0$ with $n \geq 1$ and $a_n \neq 0$, $a_i \in \mathbb{C}$, there is a $z_0 \in \mathbb{C}$ with $f(z_0) = 0$.For, thinking of $\mathbb{C} = \mathbb{R}^2$, if not, then f is a map $\mathbb{R}^2 \to \mathbb{R}^2 \setminus \{0\}$. We can divide through by a_n without affeting this, and assume that f is monic. Setting $\gamma_m(t) = f(m\cos(2\pi t), m\sin(2\pi t))$, then thought of as a map of the circle into $\mathbb{R}^2 \setminus \{0\}$, it manifestly extends to a map of the disk D^2, as $\Gamma_n(x) = f(mx)$, so γ_m is null-homotopic for all m. But $\mathbb{R}^2 \setminus \{0\}$ deformation retracts to the unit circle (the retraction is $r(z) = z/|z|$), so $\pi_1(\mathbb{R}^2 \setminus \{0\}) \cong \mathbb{Z}$, and by the above all of the $[\gamma_m]$ represent 0 in $\mathbb{Z}$, and so $r_*[\gamma_m] = [r \circ \gamma_m] = 0$, as well. But for large m, it turns out, we can compute $w(r \circ \gamma_m) = n$, which, since $n \geq 1$, is a contradiction. To see this, we write $\gamma_m(t) = f(me^{2\pi it}) = m^n(e^{2\pi nit} + (a_{n-1}/m)e^{2\pi(n-1)it} + \cdots + (a_0/m^n)) = m^n(e^{2\pi nit} + R(m,t))$, so $r \circ \gamma_m(t) = (e^{2\pi nit} + R(m,t))/|e^{2\pi nit} + R(m,t)|$. But as $m \to \infty$, $R(m,t) \to 0$ uniformly in t, since every term in $R(m,t)$ is a number with constant norm divided by a positive power of m. So for large enough m $|R(m,t)| < 1/2$ for all t, and then for every $s \in I$, $|e^{2\pi nit} + sR(m,t)| \neq 0$, since this could be 0 only if $e^{2\pi nit} = -sR(m,t)$, which is impossible since the left-hand side has norm 1 and the right has norm at most 1/2. Then the homotopy $H(t,s) = (e^{2\pi nit} + sR(m,t))/|e^{2\pi nit} + sR(m,t)|$ is well-defined and continuous, $H : I|timesI \to S^1$, and defines a homotopy from $\alpha : t \mapsto e^{2\pi nit}$ (at $s = 0$) to $r \circ \gamma_m$ (at $s = 1$). Since $w(\alpha) = n$, we have shown that, for large enough m, $w(r \circ \gamma_m) = n$. This contradiction implies that f must have a root, as desired.

Group theory "done right": presentations

Our next task is to build up machinery for computing the fundamental group of still more spaces. The basic idea is that if we understand how a space it built up out of simpler pieces, then its fundamental group is similarly built up out of "simpler" groups. This building up of groups is best understood in the language of combinatorial group theory, using presentations of groups by generators and relations.

Free groups: Σ = a set; a *reduced word* on Σ is a (formal) product $a_1^{\epsilon_1} \cdots a_n^{\epsilon_n}$ with $a_i \in \Sigma$ and $\epsilon_i = \pm 1$, and either $a_i \neq a_{i+1}$ or $\epsilon_i \neq -\epsilon_{i+1}$ for every i. (I.e., no $aa^{-1}, a^{-1}a$ in the product.)

The free group $F(\Sigma)$ = the set of reduced words, with multiplication = concatenation followed by reduction; remove all possible $aa^{-1}, a^{-1}a$ from the site of concatenation.

identity element = the empty word, $(a_1^{\epsilon_1} \cdots a_n^{\epsilon_n})^{-1} = a_n^{-\epsilon_n} \cdots a_1^{-\epsilon_1}$. $F(\Sigma)$ is generated by Σ, with no relations among the generators other than the "obvious" ones.

Important property of free groups: any function $f : \Sigma \to G$, G a group, extends uniquely to a homomorphism $\phi : F(\Sigma) \to G$.

If $R \subseteq F(\Sigma)$, then $< R >^N$ = normal subgroup generated by R
$= \{\prod_{i=1}^{n} g_i r_i g_i^{-1} : n \in \mathbb{N}_0, g_i \in F(\Sigma), r_i \in R\}$
=smallest normal subgroup containing R.
$F(\Sigma)/ < R >^N$ = the group with *presentation* $< \Sigma | R >$; it is the largest quotient of $F(\Sigma)$ in which the elements of R are the identity. Every group has a presentation:

$$G = F(G)/ < gh(gh)^{-1} : g, h \in G >^N$$

where (gh) is interpreted as a single letter in G.

If $G_1 =< \Sigma_1 | R_1 >$ and $G_2 =< \Sigma_2 | R_2 >$, then their *free product* $G_1 * G_2 =< \Sigma_1 \coprod \Sigma_2 | R_1 \cup R_2 >$ (Σ_1, Σ_2 must be treated as (formally) disjoint). Each element has a unique reduced form as $g_1 \cdots g_n$ where the g_i alternate from G_1, G_2. G_1, G_2 can be thought of as subgroups for $G_1 * G_2$, in the obivous way. Important property of free products: any pair of homoms $\phi_i : G_i \to G$ extends uniquely to a homom $\phi : G_1 * G_2 \to G$ (exactly the way you think it does).

Gluing groups: given groups G_1, G_2, with subgroups H_1, H_2 that are isomorphic $H_1 \cong H_2$, how can we "glue" G_1 and G_2 together along their "common" subgroup? More generally (and with our eye on van Kampen's Theorem) given a group H and homomorphisms $\phi_1 : H \to G_i$, we wish to build the largest group "generated" by G_1 and G_2, in which $\phi_1(h) = \phi_2(h)$ for all $h \in H$.

We can do this by starting with $G_1 * G_2$ (to get the first part), and then take a quotient to insure that $\phi_1(h)(\phi_2(h))^{-1} = 1$ for every h. Using presentations $G_1 =< \Sigma_1 | R_1 >$, $G_2 =< \Sigma_2 | R_2 >$, if we insist on quotienting out by as little as possible to get our desired result, we can do this very succinctly as

$$G = (G_1 * G_2)/ < \phi_1(h)(\phi_2(h))^{-1} : h \in H >^N =< \Sigma_1 \coprod \Sigma_2 | R_1 \cup R_2 \cup \{\phi_1(h)(\phi_2(h))^{-1} : h \in H\} >$$

This group $G == G_1 *_H G_2$ is the *largest* group generated by G_1 and G_2 in which $\phi_1(h) = \phi_2(h)$ for all $h \in H$, and is called the *amalgamated free product* or *free product with amalgamation (over H)* . [**Warning!** Group theorists will generally use this term only if both homoms ϕ_1, ϕ_2 are injective. (This insures that the natural maps of G_1, G_2 into $G_1 *_H G_2$ are injective.) But we will use this term for all ϕ_1, ϕ_2. (Some people use the term *pushout* in this more general case.)]

Important special cases : $G *_H \{1\} = G/ < \phi(H) >^N =< \Sigma | R \cup \phi(H) >$, and $G_1 *_{\{} 1\} G_2 \cong G_1 * G_2$

The relevance to π_1: the Seifert-van Kampen Theorem.

If we express a topological space as the union $X = X_1 \cup X_2$, then we have inclusion-induced homomorphisms

$$j_{1*} : \pi_1(X_1) \to \pi_1(X) \ , \ j_{2*} : \pi_1(X_2) \to \pi_1(X)$$

- to be precise, we should choose a common basepoint in $A = X_1 \cap X_2$. This in turn gives a homomorphism $\phi : \pi(X_1) * \pi_1(X_2) \to \pi_1(X)$. Under the hypotheses

$$X_1, X_2 \text{ are open, and } X_1, X_2, X_1 \cap X_2 \text{ are path-connected}$$

we can see that this homom is onto:

Given $x_0 \in X_1 \cap X_2$ and a loop $\gamma : (I, \partial I) \to (X, x_0)$, we wish to show that it is homotopic rel endpoints to a product of loops which lie alternately in X_1 and X_2. But $\{\gamma^{-1}(X_1), \gamma^{-1}(X_2)\}$ is an open cover of the compact metric space I, and so there is an $\epsilon > 0$ (a *Lebesgue number*) so that every interval of length ϵ in I lies in one of these two sets, i.e., maps, under γ, into either X_1 or X_2. If we set $N = \lceil 1/\epsilon \rceil$, then setting $a_i = i/N$, then we get a sequence of intervals $J_i = [a_i, a_{i+1}], i = 0, \dots N-1$, each mapping into X_1 or X_2. If J_i and J_{i+1} both map into the same subpace, replace them in the sequence with their union. Continuing in this fashion, reducing the number of subintervals by one each time, we will eventually find a collection I_k, $k = 1, \dots m$, of intervals covering I, overlapping only on their endpoints, which alternately map into X_1 and X_2. Their endpoints, therefore, all map into $X_1 \cap X_2$. Setting $y_k = \gamma(I_k \cap I_{k+1})$, we can, since $X_1 \cap X_2$ is path-connected, find a path $\delta_k : I \to X_1 \cap X_2$ with $\delta_k(0) = y_k$ and $\delta_k(1) = x_0$. Choosing our favorite homeomorphisms $h_k : I \to I_k$ and defining $\eta_k = \gamma|_{I_k} \circ h_k$, we have that, in $\pi_1(X, x_0)$,

$$[\gamma] = [\eta_1 * \cdots * \eta_m] = [\eta_1 * (\delta_1 * \overline{\delta_1}) * \eta_2 * \cdots * \eta_{m-1} * (\delta_{m-1} * \overline{\delta_{m-1}}) * \eta_m]$$
$$= [\eta_1 * \delta_1][\overline{\delta_1} * \eta_2 * \delta_2] \cdots [\overline{\delta_{m-2}} * \eta_{m-1} * \delta_{m-1}][\overline{\delta_{m-1}} * \eta_m]$$

We can insert the $\delta_k * \overline{\delta_k}$ into these products because each is homotopic to the constant map, and $\eta_k *$(constant) is homotopic to η_k by the same sort of homotopy that established that the constant map represents the identity in the fundamantal group.

This results in a product of loops (based at x_0) which alternately lie in X_1 and X_2. This product can therefore be interpreted as lying in $\pi(X_1) * \pi_1(X_2)$, and maps, under ϕ, to $[\gamma]$. ϕ is therefore onto, and $\pi_1(X)$ is isomorphic to the free product modulo the kernel of this map ϕ.

Loops $\gamma : (I, \partial I) \to (A, x_0)$, can, using the inclusion-induced maps $i_{1*} : \pi_1(A) \to \pi_1(X_1)$, $i_{2*} : \pi_1(A) \to \pi_1(X_2)$, be thought as either in $\pi_1(X_1)$ or $\pi_1(X_2)$. So the word $i_{1*}(\gamma)(i_{2*}(\gamma))^{-1}$, in $\pi(X_1) * \pi_1(X_2)$, is set to the identity in $\pi_1(X)$ under ϕ. So these elements lie in the kernel of ϕ.

Seifert - van Kampen Theorem: $\ker(\phi) = < i_{1*}(\gamma)(i_{2*}(\gamma))^{-1} : \gamma \in \pi_1(A) >^N$, so $\pi_1(X) \cong \pi_1(X_1) *_{\pi_1(A)} \pi_1(X_2)$.

Before we explore the proof of this, let's see what we can do with it!

Fundamental groups of graphs: Every finite connected graph Γ has a *maximal tree* T, a connected subgraph with no simple circuits. Since any tree is the union of smaller trees joined at a vertex, we can, by induction, show that $\pi_1(T) = \{1\}$. In fact, if e is an outermost edge of T, then T deformation retracts to $T \setminus e$, so, by induction, T is contractible. Consequently (*Hatcher, Proposition 0.17*), Γ and the quotient space Γ/T are homotopy equivalent, and so have the same π_1. But $\Gamma/T = \Gamma_n$ is a bouquet of n circles for some n. If we let $\mathcal{U}$ = a neighborhood of the vertex in Γ_n, which is contractible, then, by singling out one petal of the bouquet, we have

$$\Gamma_n = (\Gamma_{n-1} \cup \mathcal{U}) \cup (\Gamma_1 \cup \mathcal{U}) = X_1 \cup X_2$$

with $\Gamma_k \cup \mathcal{U} \sim (\Gamma_k \cup \mathcal{U})/\mathcal{U} \cong \Gamma_k$. And since $X_1 \cap X_2 = \mathcal{U} \sim *$, we have that

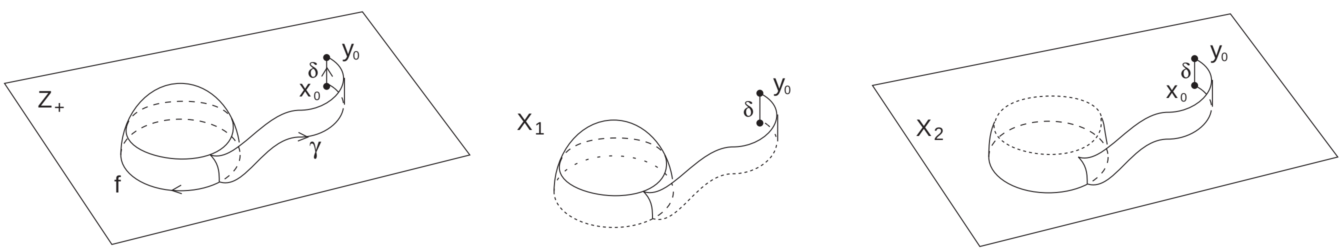
$$\pi_1(\Gamma_n) \cong \pi_1(\Gamma_{n-1}) *_1 \pi_1(\Gamma_1) = \pi_1(\Gamma_{n-1}) * \mathbb{Z}$$

So, by induction, $\pi_1(\Gamma) \cong \pi_1(\Gamma_n) \cong \mathbb{Z} * \cdots * \mathbb{Z} = F(n)$, the free group on n letters, where n = the number of edges not in a maximal tree for Γ. The generators for the group consist of the edges not in the tree, prepended and appended by paths to the basepoint.

Gluing on a 2-disk: If X is a topological space and $f : \partial\mathbb{D}^2 \to X$ is continuous, then we can construct the quotient space $Z = (X \coprod \mathbb{D}^2)/\{x \simeq f(x) : x \in \partial\mathbb{D}^2\}$, the result of gluing $\mathbb{D}^2$ to X along f. We can use Seifert - van Kampen to compute π_1 of the resulting space, although if we wish to be careful with basepoints x_0 (e.g., the image of f might not contain x_0, and/or we may wish to glue several disks on, in remote parts of X), we should also include a rectangle R, the mapping cylinder of a path γ running from $f(1,0)$ to x_0, glued to $\mathbb{D}^2$ along the arc from $(1/2,0)$ to $(1,0)$ (see figure). This space Z_+ deformation retracts to Z, but it is technically simpler to do our calculations with the basepoint y_0 lying above x_0. If we write $D_1 = \{x \in \mathbb{D}^2 : ||x|| < 1\} \cup (R \setminus X)$ and $D_2 = \{x \in \mathbb{D}^2 : ||x|| > 1/3\} \cup R$, then we can write $Z_+ = D_+ \cup (X \cup D_2) = X_1 \cup X_2$. But since $X_1 \sim *$, $X_2 \sim X$ (it is essentially the mapping cylinder of the maps f and γ) and $X_1 \cap X_2 = \{x \in \mathbb{D}^2 : 1/3 < ||x|| < 1\} \cap (R \setminus X) \simeq S^1$, we find that

$$\pi_1(Z, y_0) \cong \pi_1(X_2, y_0) *_{\mathbb{Z}} \{1\} = \pi_1(X_2)/ < \mathbb{Z} >^N \cong \pi_1(X_2)/ < [\overline{\delta} * \overline{\gamma} * f * \gamma * \delta] >^N$$

If we then use δ as a path for a change of basepoint isomorphism, and then a homotopy equivalence from X_2 to X (fixing x_0), we have, in terms of group presentations, if $\pi_1(X, x_0) =< \Sigma | R >$, then $\pi_1(Z) =< \Sigma | R \cup \{[\overline{\gamma} * f * \gamma]\} >$. So the effect of gluing on a 2-disk on the fundamental group is to add a new relator, namely the word represented by the attaching map (adjusting for basepoint).



This in turn opens up huge possibilities for the computation of $\pi_1(X)$. For example, for cell complexes, we can inductively compute π_1 by starting with the 1-skeleton, with free fundamental group, and attaching the 2-cells one by one, which each add a relator to the presentation of $\pi_1(X)$. [**Exercise:** (Hatcher, p.53, # 6) Attaching n-cells, for $n \geq 3$, has no effect on π_1.] As a specific example, we can compute the fundamental group of any compact surface.

CW complexes: The "right" spaces to do algebraic topology on.

The basic idea: CW complexes are built inductively, by gluing disks onto lower-dimensional strata. $X = \bigcup X^{(n)}$, where

$X^{(0)}$ = a disjoint union of points, and, inductively,

$X^{(n)}$ is built from $X^{(n-1)}$ by gluing n-disks D_i^n along their boundaries. That is we have $f_i : \partial D_i^n \to X^{(n-1)}$ and $X^{(n)} = X^{(n-1)} \cup (\coprod D_i^n)/\sim$ where $x \sim f_i(x)$ for all $x \in \partial D_i^n$. We have (natural) inclusions $X^{(n-1)} \subseteq X^{(n)}$, and $X = \bigcup X^{(n)}$ is given the *weak topology*; that is, $C \subseteq X$ is closed $\Leftrightarrow$ $C \cap X^{(n)}$ is closed for all n.
(Note: this is reasonable; $X^{(n-1)}$ is closed in $X^{(n)}$ for all n.)

Each disk D_i^n has a *characteristic map* $\phi_i : D_i^n \to X$ given by
$D_i^n \to X^{(n-1)} \cup (\coprod D_i^n) \to X^{(n)} \subseteq X$.
$f : X \to Y$ is cts $\Leftrightarrow$ $f \circ \phi_i : D_i^n \to X \to Y$ is cts for all D_i^n.
(This is a consequence of using the weak topology.)

A *CW pair* (X, A) is a CW complex X and a *subcomplex* A, which is a subset which is a union of images of cells, so it is a CW complex in its own right. We can induce CW structures under many standard constructions; e.g., if (X, A) is a CW pair, then X/A admits a CW structure whose cells are $[A]$ and the cells of X not in A. We can glue two CW complexes X, Y along isomorphic subcomplexes $A \subseteq X, Y$, yielding $X \cup_A Y$.

"CW"=closure finiteness, weak topology

Perhaps the most important property of CW complexes (for algebraic topology, anyway) is the *homotopy extension property*; given a CW pair (X, A), a map $f : X \to Y$, and a homotopy $H : A \times I \to Y$ such that $H|_{A\times 0} = f|_A$, there is a homotopy (extension) $K : X \times I \to Y$ with $K|_{A\times I} = H$. This is because $B = X \times \{0\} \cup A \times I$ is a retract of $X \times I$; K is the composition of this retraction and the "obvious" map from B to Y.

To build the retraction, we do it one cell of X at a time. The idea is that the retraction is defined on the cells of A (it's the identity), so look at cells of X not in A. Working our way up in dimension, we can assume the the retraction r_{n-1} is defined on (the image of) $\partial D^n \times I$, i.e., on $X^{(n-1)} \times I$. But $D^n \times I$ (strong deformation) retracts onto $D^n \times 0 \cup \partial D^n \times I$; composition of r_{n-1} with this retraction extends the retraction over $\phi(D^n) \times I$, and so over $X^{(n)} \times I$..

This, for example, lets us show that if (X, A) is a CW pair and A is contractible, then $X/A \simeq X$. This is because the composition $A \to * \to A$ is homotopic to the identity I_A, via some map $H : A \times I \to A$, with $H|_{A\times 0} = I_A$. Thinking of H as mapping into X, then together with the map $I_X : X \to X$ the HEP provides a map $K : X \times I \to X$ with $K_0 = I_X$ and $K_1(A) = *$. Setting $K_1 = g : X \to X$, it induces a map $h : X/A \to X$. This is a homotopy inverse of the projection $p : X \to X/A$:

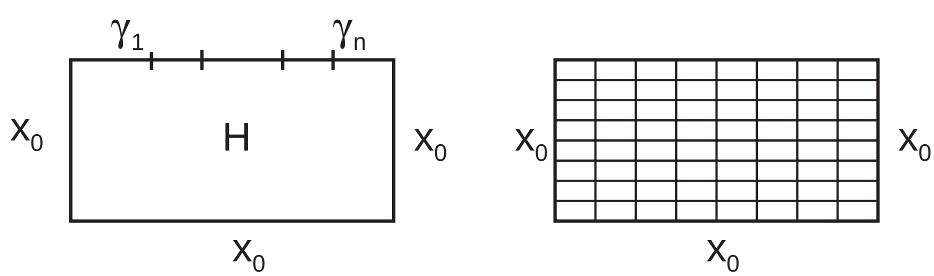
$h \circ p = g \simeq I_X$ via K, and $p \circ h : X/A \to X/A$ is homotopic to $I_{X/A}$ since $K_t(A) \subseteq A$ for every t, so induces a map $\overline{K}_t : X/A \to X/A$, giving a homotopy between
$\overline{K}_0 = \overline{I_X} = I_{X/A}$ and $\overline{K}_1 = \overline{g} = p \circ h$.

Proving Seifert - van Kampen:

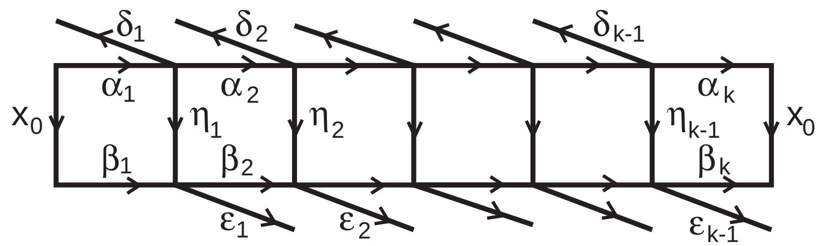
We now turn our attention to proving Seifert - van Kampen; understanding the kernel of the map $\phi : \pi_1(X_1) * \pi_1(X_2) \to \pi_1(X)$, under the hypotheses that X_1, X_2 are open, $A = X_1 \cap X_2$ is path-connected, and the basepoint $x_0 \in A$. So we start with a product

$g = g_1 \cdots g_n$ of loops alternately in X_1 and X_2, which when thought of in X is null-homotopic. We wish to show that g can be expressed as a product of conjugates of elements of the form $i_{1*}(a)(i_{2*}(a))^{-1}$ (and their inverses). The basic idea is that a "big" homotopy can be viewed as a large number of "little" homotopies, which we essentially deal with one at a time, and we find out how little "little" is by using the same Lebesgue number agument that we used before.

Specifically, if H is the homotopy, rel basepoint, from $\gamma_1 * \cdots * \gamma_n$, where γ_i is a based loop representing g_i, and the constant loop, then, as before, $\{H^{-1}(X_1), H^{-1}(X_2)\}$ is an open cover of $I \times I$, and so has a Lebesgue number ϵ. If we cut $I \times I$ into subsquares, with length $1/N$ on a side, where $1/N < \epsilon$, then each subsquare maps into either X_1 or X_2. The idea is to think of this as a collection of horizontal strips, each cut into squares. Arguing by induction, starting from the bottom (where our conclusion will be obvious), we will argue that if the bottom of the strip can be expressed as an element of the group
$N =< i_{1*}(\gamma)(i_{2*}(\gamma))^{-1} : \gamma \in \pi_1(A) >^N \subseteq \pi_1(X_1) * \pi_1(X_2)$
(i.e., as a product of conjugates of such loops), then so can the top of the strip.



And to do this, we work as before. We have a strip of squares, each mapping into either X_1 or X_2. If adjacent squares map into the same subpace, amalgamate them into a single larger rectangle. Continuing in this way, we can break the strip into subrectangles which alternately map into X_1 or X_2. This means that the vertical arcs in between map into $X_1 \cap X_2 = A$, and represent paths η_i in A. Their endpoints also map into A, and so can be joined by paths (δ_i on the top, ϵ_i on the bottom) in A to the basepoint. The top of the strip is homotopic, rel basepoint, to
$(\alpha_1 * \delta_1) * (\overline{\delta_1} * \alpha_2 * \delta_2) * \cdots * (\overline{\delta_{k-1}} * \alpha_k)$
each grouping mapping into either X_1 or X_2. The rectangles demonstrate that each grouping is homotopic, rel basepoint, to the product of loops
$(\overline{\delta_i} * \eta_i * \epsilon_i) * (\overline{\epsilon_i} * \beta_i * \epsilon_{i+1}) * (\overline{\epsilon_{i+1}} * \overline{\eta_{i+1}} * \delta_{i+1}) = a_i b_i a_{i+1}^{-1}$
where this is thought of as a product in either $\pi_1(X_1)$ or $\pi_1(X_2)$. The point is that when strung together, this appears to give $(b_1 a_2^{-1})(a_2 b_2 a_3^{-1}) \cdots (a_k b_k)$, with lots of cancellation, but in reality, the terms $a_i^{-1} a_i$ represent elements of N, since the two "cancelling" factors are thought of as living in the different groups $\pi_1(X_1), \pi_1(X_2)$. The remaining terms, if we delete these "cancelling" pairs, is $b_1 \cdots b_k = \beta_1 * \epsilon_1 * \cdots * \overline{\epsilon_i} * \beta_i * \epsilon_{i+1} * \cdots * \overline{\epsilon_k} * \beta_k$, which is homotopic rel endpoints to $\beta_1 * \cdots * \beta_k$, which, by induction, can be represented as a product which lies in N.



So, we can obtain the element represented by the top of the strip by inserting elements of N into the bottom, which is a word having a representation as an element of N. The final problem to overcome is that the insertions represented by the vertical arcs might not be occuring where we want them to be! But this doesn't matter; inserting a word w in the middle of another uv (to get uwv) is the same as multiplying uv by a conjugate of w; $uwv = (uv)(v^{-1}wv)$, so since the bottom of the strip is in N, and we obtain the top of the strip by inserting elements of N into the bottom, the top is represented by a product of conjugates of elements of N, so (since N is normal) is in N. And a final final point; the subrectangles may not have cut the bottom of the strip up into the same pieces that the inductive hypothesis used to express the bottom as an element of N. It didn't even cut it into loops; we added paths at the break points to make that happen. The inductive hypothesis would have, in fact, added its own extra paths, at possibly different points! But if we add both sets of paths, and cut the loop up into even more pieces, then we end up with a loop, which we have expressed as a product in $\pi_1(X_1) * \pi_1(X_2)$ in two (possibly different) ways, since the two points of view will have interpreted pieces as living in different subspaces. But when this happens, it must be because the subloop really lives in $X_1 \cap X_2 = A$. Moving from one to the other amounts to repeatedly changing ownership between the two sets, which in $\pi_1(X_1) * \pi_1(X_2)$ means inserting an element of N into the product (that is literally what elements of N do). But as before, these insertions can be collected at one end as products of conjugates. So if one of the elements is in N, the other one is, too.

Which completes the proof!

The inherent complications above derived from needing open sets can be legislated away, by introducing additional hypotheses:

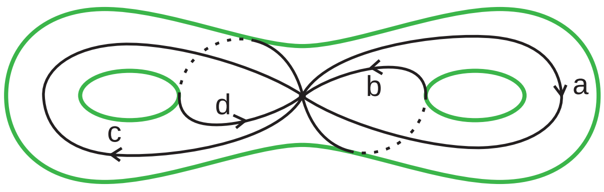
Theorem: If $X = X_1 \cup X_2$ is a union of closed sets X_1, X_2, with $A = X_1 \cap X_2$ path-connected, and if X_1, X_2 have open neighborhood $\mathcal{U}_1, \mathcal{U}_2$ so that $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_1 \cap \mathcal{U}_2$ deformation retract onto X_1, X_2, A respectively, then $\pi_1(X) \cong \pi_1(X_1) *_{\pi_1(A)} \pi_1(X_2)$ as before.

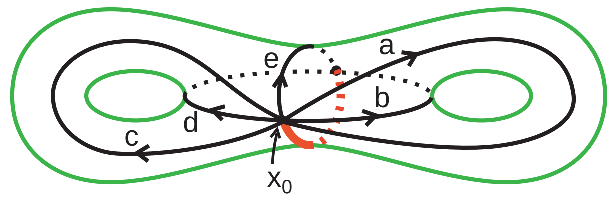
The hypotheses are satisfied, for example, if $X_1.X_2$ are subcomplexes of the cell complex X.

Some more computations:

The *real projective plane* $\mathbb{R}P^2$ is the quotient of the 2-sphere S^2 by the antipodal map $x \mapsto -x$; it can also be thought of as the upper hemisphere, with identification only along the boundary. This in turn can be interpreted as a 2-disk glued to a circle, whose boundary wraps around the circle twice. So $\pi_1(\mathbb{R}P^2) \cong < a | a^2 > \cong \mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$. A surface F of genus

2 can be given a cell structure with 1 0-cell, 4 1-cells, and 1 2-cell, as in the figure, as in the first of the figures below. The fundamental group of the 1-skeleton is therefore free of rank 4, and $\pi_1(F)$ has a presentation with 4 generators and 1 relator. Reading the attaching map from the figure, the presentation is $< a, b, c, d \mid [a, b][c, d] >$.



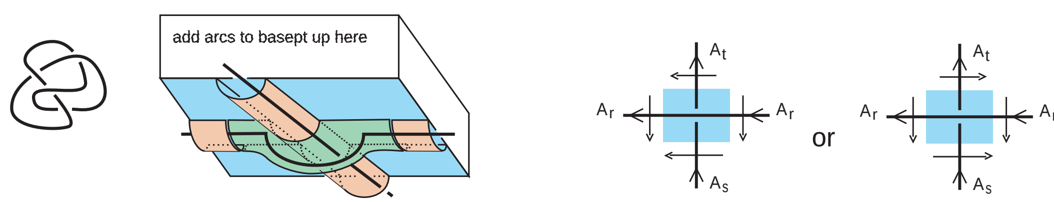


Giving it a different cell structure, as in the second figure, with 2 0-cells, 6 1-cells, and 2 2-cells, after choosing a maximal tree, we can read off the two relators from the 2-cells to arrive at a different presentation $\pi_1(F) =< a, b, c, d, e \mid aba^{-1}eb^{-1}, cde^{-1}c^{-1}d^{-1} >$. A posteriori, these two presentations describe isomorphic groups.

Using the same technology, we can also see that, in general, any group is the fundamental group of some 2-complex X; starting with a presentation $G =< \Sigma | R >$, build X by starting with a bouquet of $|\Sigma|$ circles, and attach $|R|$ 2-disks along loops which represent each of the generators of R. (This works just as well for infinite sets Σ and/or R; essentially the same proofs as above apply.)

Wirtinger presentations for knot complements:

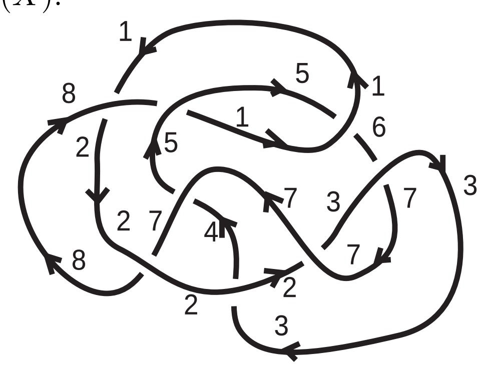
A *knot* K is (the image of) an embedding $h : S^1 \hookrightarrow \mathbb{R}^3$. Wirtinger gave a prescription for taking a planar projection of K and producing a presentation of $\pi_1(\mathbb{R}^3 \setminus K) = \pi_1(X)$. The idea: think of K as lying on the projection plane, except near the crossings, where it arches under itself. We build a CW-complex $Y \subseteq X$ that X deformation retracts to. A presentation for $\pi_1(Y)$ gives us $\pi_1(X)$.



To build Y, glue rectangles arching under the strands of K to a horzontal plane lying just above the projection plane of K. At the crossing, the rectangle is glued to the rectngle arching under the over-strand. X deformation retracts to Y; the top half of $\mathbb{R}^3$ deformation retracts to the top plane, the parts of X inside the tubes formed by the rectangles radially retract to the boundaries of the tubes, and the bottom part of X vertically retracts onto Y. Formally, we should really keep a "slab" above the plane, to give us a place to run arcs to a fixed basepoint in the interior of the slab.

We think of Y as being built up from the slab C, by gluing on annuli $A_i \cong S^1 \times I$, one for each rectangle R_i glued on; the rectangle S_i lying above R_i in the bottom of the slab C is the other half of the annulus. Then we glue on the 2-disks D_j, one for each crossing of the knot projection. A little thought shows that there are as many annuli as disks; the annuli correspond to the unbroken strands of the knot projetion, which each have two ends, and each crossing is where two ends terminate (so there are two ends for every A_i and two ends for every D_j, so there are half as many of each as there are total number of ends). To make sure that all of our interections are path connected, and to formally use a single basepoint in all of our computations, we join every one of the annuli and disks to a basepoint lying in the slab by a collection of (disjoint) paths.

Now starting with the slab (which is simply-connected), we begin to add the A_i one at a time; each has fundamental group $\mathbb{Z}$, generated by a loop which travels once around the S^1-direction, and its intersection with $C\cup$ the previously glued on annuli is the rectangle S_i, which is simply connected. So, inductively, $\pi_1(C \cup A_1 \cup \cdots \cup A_i) \cong \pi_1(C \cup A_1 \cup \cdots \cup A_{i-1}) * \pi_1(A_i) \cong F(i-1) * \mathbb{Z} \cong F(i)$ is the free group on i letters, so, adding all n (say) of the annuli yields $F(n)$. To finish, we glue on the n 2-disks D_j; these amount to adding n relators to the presentation $\langle x_1, \dots, x_n | \rangle$. To determine these relators, we need to choose specific generators for our $\pi_1(A_i)$; a standard choice is made by *orienting* the knot (choosing a direction to travel around it) and choosing the loop which goes counter-clockwise around the annulus (when you stand it vertically using the orientation of the strand it is going around. Then going around the boundary of the 2-disk D_j spells out the word $x_r x_s x_r^{-1} x_t^{-1}$, if the overstrand at the crossing corresponds to A_r and the understrand runs from A_s on the left to A_t on the right. [There is the possibility of the mirror image, when the orientation of the strands goes from right to left under the overstrand; then the proper relator is $x_r x_s^{-1} x_r^{-1} x_t$.] Carrying this out for every 2-disk completes the presentation of $\pi_1(Y) \cong \pi_1(X)$.



With practice, it becomes completely routine to read off a presentation for the fundamental group of $\mathbb{R}^3 \setminus K$ from a projection of K. For example, from the projection above, we have $\pi_1(\mathbb{R}^3 \setminus K) \cong \langle x_1, \dots, x_8 | x_8 x_1 = x_2 x_8, x_2 x_7 = x_8 x_2, x_5 x_8 = x_1 x_5, x_1 x_5 = x_6 x_1, x_3 x_6 = x_7 x_3, x_7 x_2 = x_3 x_7, x_3 x_2 = x_2 x_4, x_7 x_4 = x_5 x_7 \rangle$

Postscript: why should we care? The role of the fundamental group in distinguishing spaces has already been touched upon; if two (path-connected) spaces have non-isomorphic fundamental groups, then the spaces are not homeomorphic, and even not homotopy equivalent. It is one of the most basic, and in many cases the best such invariant we will have in our arsenal (hence the name “fundamental”). As we have seen with the circle, it captures the notion of how many times a loop “winds around” in a space. And the idea of using paths to understand a space is very basic; we explore a space by mapping familiar objects into it. (This is a theme we keep returning to in this course.) The concepts we have introduced play a role in analysis, for instance with the notion of a path integral; the invariance of the integral under homotopies rel endpoints is an important property, related to Green’s Theorem and (locally) conservative vector fields. And the space of all paths in X plays an important (theoretical, although pprobably not practical) role in what we will do next.