Fundamental groups

1. Motivation

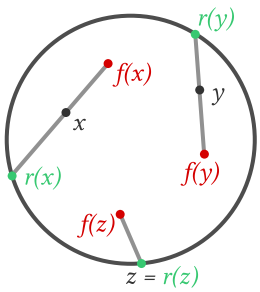
1.1. Brouwer's fixed point theorem

We cannot perfectly mix coffee!

Let $\mathbb{D}^2$ be the unit disk $\{x \in \mathbb{R}^2 \mid \|x\| \leq 1\}$.

Theorem 1.1.1 (Brouwer): For every continuous map $f : \mathbb{D}^2 \to \mathbb{D}^2$, there exists $x \in \mathbb{D}^2$ such that $f(x) = x$.

A natural attempt at the proof would be by contradiction. Suppose there is such a map f. Since x and $f(x)$ are distinct, we can draw a half line from x to $f(x)$. Denote by $r(x)$ the point where it intersects the circle $\mathbb{S}^1$ (the boundary of $\mathbb{D}^2$). Since f is a continuous map, it is an easy exercise that the map $x \mapsto r(x)$ is continuous as well. Moreover, it is an identity for all points on the boundary $\mathbb{S}^1$.



We have the inclusion $i : \mathbb{S}^1 \hookrightarrow \mathbb{D}^2$ of the boundary to the circle. So overall we have two maps

$$\mathbb{S}^1 \overset{i}{\hookrightarrow} \mathbb{D}^2 \overset{r}{\to} \mathbb{S}^1$$

whose composition is the identity on $\mathbb{S}^1$.

The argument could now proceed by saying such a map r can never be continuous, since it must somehow involve creating a "hole" in the middle of the disk. To make this argument formal, it is convenient to first develop a general theory of continuous maps. This theory is called *topology*.

Specifically, we will study a way to detect "holes." To each suitable geometric object S (like a disk or a sphere), we will assign a group $\pi_1(S)$ called the *fundamental group* of S. It will consist of loops which cannot be contracted in S to a point, suggesting a presence of some obstruction - a "hole". This will allow us to transform the statements about continuous maps between spaces to statements about functions between groups, which are often more tractable.

1.2. Topology

Definition 1.2.1: In this lecture, a *space* will be a subset of $\mathbb{R}^n$ for some $n \in \mathbb{N}$.

A map $f : S \to T$ between spaces $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$ is called *continuous* if for every $\varepsilon > 0$, there is $\delta > 0$ such that for all $x, y \in S$ with $\|x - y\| < \delta$, it holds that $\|f(x) - f(y)\| < \varepsilon$.

Intuitively, the definition of continuity means that f maps sufficiently close points to sufficiently close points.

1.3. Aside: topological spaces formally

For the present series, understanding the notions of this section is not substantial and it suffices to consider spaces as subsets of $\mathbb{R}^n$. It is only put here for interested readers.

You may have already seen a more general definition of a metric space. Inside of it, you have a *metric* funcion, which computes a distance between points.

Definition 1.3.1: A *metric space* is a set M and a function $d : M \times M \to \mathbb{R}_{\geq 0}$ to non-negative real numbers called a *metric*, satisfying for all $x, y, z \in M$
- $d(x, y) = 0$ if and only if $x = y$
- $d(x, y) = d(y, x)$ (symmetry)
- $d(x, y) + d(y, z) \geq d(x, z)$ (triangle inequality)

It is not hard to see that $\mathbb{R}^n$ and all of its subsets, along with the classic Euclidean distance, are metric spaces.

We can, however, consider even a much more general concept of a topological space. There, the notion of the distance is lost and we retain only the information about which sets are the neighbourhoods of which points. The advatage for our purposes is that it is often convenient to consider quotients of spaces, i.e. spaces factored by some equivalence. Those no longer have a natural metric, but still have a natural induced topological structure.

Definition 1.3.2: A topological space is a set S, along with for each $x \in S$ a system $\mathcal{N}(S)$ of subsets of S called *neighbourhoods* of x, which satisfy:
- $x \in N$ for every $N \in \mathcal{N}(x)$
- if $M \supseteq N \in \mathcal{N}(x)$, then $M \in \mathcal{N}(x)$ (superset of a neighbourhood is a neighbourhood)
- for $M, N \in \mathcal{N}(x)$, also $M \cap N \in \mathcal{N}(x)$ (intersection of neighbourhoods is a neigbourhood)

A subset $O \subset S$ of a topological space S is called *open* if for every $x \in S$, there exists a neighbourhood N of x such that $N \subset O$.

- for every $N \in \mathcal{N}(x)$, there is $N \supseteq M \in \mathcal{N}(x)$ with M open

Every metric space is a topological space.

Definition 1.3.3: For a metric space M and a point $x \in M$, an open ball (of radius r) around x is the set $B(x, r) = \{y \in M \mid d(x, y) < r\}$.

A set $O \subset M$ is a neighbourhood of x if it contains an open ball around x $B(x, r)$ for some $r \in \mathbb{R}_{>0}$.

Then equivalently, an open set of a metric space is a set containing some open ball around each of its points.

Exercise: Show that the definition of the topological space we have given is equivalent to the following one, more often found in the literature, in terms of open sets: A topological space is a set T, along with a set τ of subsets of T called *open sets*, such that
- $\emptyset, T$ are open

- a union of any collection of open sets is open
- an intersection of any *finite* collection of open sets is open

(Hint: define neighbourhoods of X as the sets containing an open set containing X.)

Definition 1.3.4: A map $f : S \to T$ between topological spaces is continuous if for every $x \in S$ and $N \in \mathcal{N}(f(x))$, it holds that $f^{-1}(N) \in \mathcal{N}(x)$.

Exercise: Show that the definition is equivalent to the following: for every $O \subset T$ open, the preimage $f^{-1}(O)$ is open. Moreover, show that for a metric space, this definition is equivalent to the definition of a continuous map there.

1.4. Dirac belt trick

Feel free to try the following trick a home. Firmly fix the end of the belt opposite the buckle, so that it cannot move. Now twist the buckle 360 degrees around the axis of the belt while keeping it pointed in the same direction. In this state, the belt is twisted and no movement of the buckle which keeps it horizontal and pointing in the same direction can undo the twist.

Obviously, a rotation of the buckle by 360 degrees in the opposite direction would undo the twist. What is surprising is that after rotating the buckle by 360 degrees in the same direction as before, making the belt "doubly twisted", the twist can be undone by simply moving the buckle while keeping its orientation fixed.

To understand what's going on, consider the curve in the middle of the belt that goes from the opposite end to the buckle. In each point, we can draw a frame of 3 unit vectors - one in the direction of the curve, one normal to the belt and one orthogonal to the previous 2. So the state of the belt can be represented by the path in the space of triples of orthogonal vectors in $\mathbb{R}^3$. If we write these vectors in a matrix, they form precisely the rotation matrices, or the *special orthogonal group* $\mathrm{SO}(3)$.

Definition 1.4.1: A real 3×3 matrix A is in $\mathrm{SO}(3)$ if $A^T A$ is the identity matrix (equivalently, A has orthonormal columns) and its determinant is 1.

Although $\mathrm{SO}(3)$ is a subset of $\mathbb{R}^9$, its elements can actually be described by just 3 parameters (exercise), so it makes sense to think of it as of a 3-dimensional subspace.

Now, the belt represents a path $\varphi : [0,1] \to \mathrm{SO}(3)$. Having the buckle and the other end fixed amounts to fixing the values of $\varphi(0)$ and $\varphi(1)$, say to the identity matrix. Moving the belt while keeping the ends fixed corresponds to deforming these paths. The Dirac belt trick suggests that some paths can be deformed to the constant paths at the identity ($\varphi(t) = \mathrm{Id}\ \forall t \in [0,1]$) which is the straight belt, while others can't.

To describe it formally, we first have to define deformation of maps.

2. Groups

Here we just recall the basic definition of group theory.

Definition 2.1: A *group* is a set G, along with a binary operation $\cdot : G \times G \to G$ called the *product*, a unary operation $(-)^{-1} : G \to G$ called the *inverse* and a constant $e \in G$ called the *unit element*, satisfying for each $f, g, h \in G$ the following axioms:
- $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ (associativity)
- $e \cdot g = g = g \cdot e$ (e is the unit element)
- $g \cdot g^{-1} = e = g^{-1}$ (g^1 is the inverse to g)

Exercise: Show that for a set X, the set of bijections $X \to X$ forms a group with the product given by composition of funcions, inverse an inverse bijection and the identity bijection as the identity forms a group.

Originally, the term *group* was used on a subset of bijections of a set X closed under compositions and inverses. Later, it became apparent that the abstract axiomatic definition is often more convenient.

Definition 2.2: For a group G with the product $\cdot$ and a group H with the product $\star$, a function $f : G \to H$ is called a group homomorphism if it preserves products, i.e. for every $x, y \in G$:

$$f(x \cdot y) = f(x) \star f(y)$$

A homorphism f is called an *isomorphism* if there is a homomorphism $g : H \to G$ such that $f \circ g = \mathrm{id}_H$ and $g \circ f = \mathrm{id}_G$.

Isomorphisms are a good notion of "group equivalence" - in particular, the underlying sets are in bijection which preserves the products.

Exercise: Show that homomorphisms preserve units and inverses as well.

3. Homotopies

Definition 3.1: Let $f, g : S \to T$ be maps between spaces. They are called *homotopic* if there is a map $H : S \times [0, 1] \to T$ such that for all $x \in S$:
- $H(x, 0) = f(x)$
- $H(x, 1) = g(x)$

H is called *homotopy*. We will denote by $f \sim g$ that f and g are homotopic.

A map $f : S \to T$ is called a *homotopy equivalence* if there is $g : T \to S$ such that both composites fg and gf are **homotopic to** identity.

A space is called *contractible* if it is equivalent to the single point $*$ (a one element set).

Exercise: Show that being homotopic is an equivalence relation.

Example: $\mathbb{R}^n$ is contractible. Let $f : * \hookrightarrow \mathbb{R}^n$ be the inclusion of the origin and $g : \mathbb{R}^n \to *$ the unique map to the point. Then gf is the identity and fg is homotopic to the identity via the map $H(x, t) = tx$.

This example shows that while all homeomorphic spaces are homotopy equivalent, the converse is not true (there is not even a bijection between $\mathbb{R}^n$ and a single point). On the other hand, the spheres $\mathbb{S}^n$ are not contractible, but it is not completely trivial to prove.

3.1. Loop spaces

Definition 3.1.1: For a space S and a point $* \in S$, let the *loop space* $\Omega(S, *)$ be the set $\{\rho : [0,1] \to S \mid \rho(0) = \rho(1) = *\}$ of loops in S which begin and end in $*$.

For $\rho, \tau \in \Omega(S, *)$, define their product $\rho\tau : [0,1] \to S$ by:

$$\rho\tau(t) = \begin{cases} \rho(2t) \text{ for } t \in \left[0, \frac{1}{2}\right] \\ \tau\left(2\left(t - \frac{1}{2}\right)\right) \text{ for } t \in \left[\frac{1}{2}, 1\right] \end{cases}$$

We just go around the first loop and then around the second one.

This product almost gives the set $\Omega(S, *)$ the structure of a group - the group axioms in $\Omega(S, *)$ hold up to homotopy. Explicitly, for every $\rho, \sigma, \tau \in \Omega(S, *)$:

- $(\rho\sigma)\tau \sim \rho(\sigma\tau)$ (associativity)
- $e\rho \sim \rho \sim \rho e$ (unit)
- $\rho\rho^{-1} \sim e \sim \rho^{-1}\rho$ (inverses)

where $e(t) = *$ and $\rho^{-1}(t) = \rho(1-t)$.

For example, the two curves in the associativity relation (the two ways to bracket the product of three curves) have the same image, but different parametrisation. The homotopy $H : [0,1] \times [0,1] \to S$ between them is given by linearly interpolating between the two parametrizations. For the sake of explicitness, this is:

$$H(s,t) = \begin{cases} \rho((4-2s)t) \text{ for } t \in \left[0, \frac{1}{4-2s}\right] \\ \sigma(4t - (1+s)) \text{ for } t \in \left[\frac{1}{4-2s}, \frac{1}{4} + \frac{1}{4-2s}\right] \\ \tau\left((1+s)\left(2t - \frac{4-s}{4-2s}\right)\right) \text{ for } t \in \left[\frac{1}{4} + \frac{1}{4-2s}, 1\right] \end{cases}$$

Exercise: Construct the homotopies witnessing the unit and inverse axiom.

3.2. Fundamental groups

To get a proper group, we factor the loop space by the equivalence relation of being homotopic.

Definition 3.2.1: For a space S and $* \in S$, let the *fundamental group* $\pi_1(S, *)$ be the set $\Omega(S, *)/\sim$, along with the induced group structure. Its elements are the equivalence classes of paths starting and ending at $*$.

For φ in $\Omega(S, *)$, we will denote its equivalence class in $\pi_1(S, *)$ by $[\varphi]$.

To avoid having to pick the base point $* \in S$, we can without loosing much generality restrict ourselves to the spaces with a single connected component and show that the group does not depend on the choice of the base point there.

Definition 3.2.2: A space S is called *path connected* if for all pairs of points $x, y \in S$, there exists a path $\varphi : [0,1] \to S$ connecting them, i.e $\varphi(0) = x$ and $\varphi(1) = 1$.

Proposition 3.2.1: For a path connected space S and two points $x, y \in S$, the groups $\pi_1(S, x)$ and $\pi_1(S, y)$ are isomorphic.

Proof: Pick a path φ connecting x and y. Construct a pair of maps

$$\Omega(S,x) \to \Omega(S,y), \psi \mapsto \varphi\psi\varphi^{-1}$$
$$\Omega(S,y) \to \Omega(S,x), \tau \mapsto \varphi^{-1}\tau\varphi$$

where $\varphi^{-1}(t) = \varphi(1-t)$ and the composition $\varphi\psi\varphi^1(t)$ is defined by

$$\varphi\psi\varphi^{-1}(t) = \begin{cases} \varphi(3t) \text{ for } t \in \left[0, \frac{1}{3}\right] \\ \psi\left(3\left(t - \frac{1}{3}\right)\right) \text{ for } t \in \left[\frac{1}{3}, \frac{2}{3}\right] \\ \psi\left(3\left(t - \frac{2}{3}\right)\right) \text{ for } t \in \left[\frac{2}{3}, 1\right] \end{cases}$$

These maps map homotopic paths to homotopic paths and preserve the composition of paths in the loop spaces up to homotopy. Moreover, they are inverse up to homotopy. So they descend to mutually inverse group homomorphisms between $\pi_1(S, x)$ and $\pi_1(S, y)$, i.e. isomorphisms. □

Definition 3.2.3: For a path connected space S, $\pi_1(S)$ will denote the fundamental group of S with respect to any choice of the base point in S.

Definition 3.2.4: A map between spaces $f : S \to T$ induces a homomorphism of groups $\pi_1(f) : \pi_1(S, *) \to \pi_1(T, f(*))$ by mapping $[\varphi]$ for $\varphi : [0,1] \to S$ to $[f \circ \varphi]$.

Remark: A homotopy $H : [0,1] \times [0,1] \to S$ is mapped to the homotopy $f \circ H$ in T, so this map is well defined (does not depend on the choince of representative for the equivalence class of φ). It is a homomorphism since a map of a composition of paths $f \circ (\varphi\psi)$ is the same curve as $(f \circ \varphi)(f \circ \psi)$.

It is also easy to see that $\pi_1(g \circ f) = \pi_1(g) \circ \pi_1(f)$. In the language of the category theory, this means that π_1 is a *functor*.

Proposition 3.2.2: For homotopy equivalent path-connected spaces S and T, their fundamental groups $\pi_1(S)$ and $\pi_1(T)$ are isomorphic.

Proof: Pick maps $f : S \to T$ and $g : T \to S$ witnessing the homotopy equivalence, i.e. there is a homotopy H between fg and id_S, also there is a homotopy G between gf and id_T. A loop $\varphi : [0,1] \to S$ is mapped to a loop $f \circ \varphi$ in T and a loop $\psi : [0,1] \to T$ is mapped to a loop $g \circ \psi$ in S. The loop $g \circ f \circ \varphi$ is homotopic to φ via the homotopy H; analogously in the other direction. These maps obviously preserve the composition of loops, yielding the sought group isomorphisms. □

The fundamental group of a single point is a trivial group (with only one element - the identity) since there is only one map from any set to a one elemnt set. Thanks to the previous proposition, the fundamental group of any *contractible* space, i.e space homotopy equivalent to a point (such as $\mathbb{R}^n$ or $\mathbb{D}^n$) is trivial as well.

Proposition 3.2.3: The fundamental group $\pi_1(\mathbb{S}^n)$ of the sphere $\mathbb{S}^n$ for $n \geq 2$ is trivial.

Proof: If a loop $\varphi : [0,1] \to \mathbb{S}^n$ misses a point $x \in \mathbb{S}^n$, stereographically project it from x to $\mathbb{R}^n$, contract it there to the trivial loop, and stereographically project the homotopy back to $\mathbb{S}^n$.

There actually exist loops whose image is the whole $\mathbb{S}^n$. For them, we need more work. □

3.3. Coverings

To compute the fundamental group in the spaces where it is non-trivial, one needs to develop more theory. One approach is to construct a bigger space, which "covers" the original space and with that keeps track of the loops inside it.

Definition 3.3.1: A map $p : E \to B$ is called a *covering* if for every $b \in B$, there is a neighbourhood $b \in U \subseteq B$ such that $p^{-1}(B)$ is homeomorphic to a disjoint union of a copies of U.

For a point $b \in B$, the preimage $p^{-1}(b)$ is called a *fiber* of b.

A covering is called *universal* if E is path-connected and $\pi_1(E)$ is trivial.

Example: A (universal) covering $\mathbb{R}^1 \to \mathbb{S}^1$ given by $t \mapsto (\cos t, \sin t)$. This covering can be imagined as an infinite spiral staircase above a circle. Furthermore, one can imagine two people one on the staircase and one below, always moving in a way that they remain one above the other.

Proposition 3.3.1 (Unique path lifting): For a covering $E \to B$ along with:

- $b \in B$
- a path φ in B with $\varphi(0) = b$
- $e \in E$ with $p(e) = b$,

there is a unique path $\tilde{\varphi}$ in E with $\tilde{\varphi}(0) = e$ and $p \circ \tilde{\varphi} = \varphi$.

Proof: By the definition of a covering, for every $t \in [0,1]$, the point $\varphi(t)$ has a neighbourhood U_t whose p-preimage is homeomorphic to a disjoint union of copies of U_t. The sets $\varphi^{-1}(U_t)$ cover the unit interval and by the Heine-Borel theorem (TODO), it is covered by a finite number of them for some $t_0, ..., t_n$.

First compose φ the neighbourhood of 0 with the homeomorphism mapping U_0 to the connected component of $p^{-1}(U_0)$ containing the point e. From there, extend the curve in the finite number of steps using the neigbourhoods $U_{t_0}, ..., U_{t_n}$. □

Proposition 3.3.2 (Unique homotopy lifting): For a covering $E \to B$ along with:

- $b \in B$
- paths φ, ψ in B with $\varphi(0) = \psi(0) = b$
- a homotopy H between φ and ψ constant on the point b
- $e \in E$ with $p(e) = b$,

there is a unique homotopy $\tilde{H}$ between $\tilde{\varphi}$ and $\tilde{\psi}$ constant on endpoints such that $p \circ \tilde{H} = H$.

Proof: The proof is completely analogous to the previous one, using the fact that every cover by open sets of the square $[0,1] \times [0,1]$ admits a finite subcover. □

In particular, the universal covering sees the group $\pi_1(B)$ geometrically. We will now exhibit it as certain symmetries of the *universal* covering.

Definition 3.3.2: For a covering $p : E \to B$, a map $f : E \to E$ is called a *deck transformation* if it preserves the covering, meaning $p \circ f = p$. All deck tranformations of a covering form a group, which will be denoted $G(E)$.

Proposition 3.3.3: For a universal covering $p : E \to B$ and $b \in B$, the deck transformations $G(E)$ correspond to points of the fiber $p^{-1}(b)$ (by where they send a single point $e \in p^{-1}(b)$.

Proof: Suppose the deck transformation maps e to the point e'. For a point $x \in B$, pick a path φ connecting e and x and let $\varphi_{e'}$ be the unique lift of $p \circ \varphi$ starting at e'. By the uniqueness of the path lifting, x must be mappend to $\varphi_{e'}(1)$, so the deck transformation is determined by e'.

On the other hand, since $\pi_1(E)$ is trivial, all the choices of φ are homotopic, yielding a homotopy between the corresponding $\varphi_{e'}$s (by lifting of homotopies). In particular, their endpoints are the same, so the image of every point is uniquely determined.

It remains to show that the constructed map $d : E \to E$ is continuous. A sufficiently small open neighbourhood V around $d(x)$ is homeomorphic to an open neighbourhood of $p(x) = p(d(x))$, whose p-preimage is an open neighbourhood $d^{-1}(V)$ of x. □

Finally, given the universal covering, we can compute the fundamental group of the base space.

Theorem 3.3.1: For a *universal* covering $p : E \to B$, the group of its deck tranformations $G(E)$ is isomorphic to $\pi_1(B)$.

Proof: Pick a point $b \in B$ and $e \in p^{-1}(b)$.

On one hand, assign to $[\varphi] \in \pi_1(B)$ (with φ a loop based at b) the unique deck transformation $d_{\tilde{\varphi}(e)}$ mapping e to $\tilde{\varphi}(e)$ where $\tilde{\varphi}$ is the unique lift of e starting at p. By lifting of homotopies, this does not depend on the choice of representative. If σ is another loop (based at b), then $d_{\widetilde{\sigma\rho}(e)} = d_{\tilde{\sigma}(e)} \cdot d_{\tilde{\varphi}(e)}$ by the construction from Proposition 3.3.3, so it is a homomorphism.

On the other hand, for a deck transformation $d : E \to E$, pick a path ρ_d connecting e and $d(e)$. Since it is a deck transformation, $d(e) \in p^{-1}(b)$, so we can define the loop

$$\psi_d \coloneqq p \circ \rho_d$$

Because $\pi_1(E)$ is trivial, all paths from e to $d(e)$ are homotopic (otherwise we would have a nontrivial loop going from e to $d(e)$ and back), so the homotopy class $[\psi_d]$ does not depend on our choice of ρ. If f is another deck transformation, then

$$\psi_{f \cdot d} = p \circ (\rho_f \rho_d) = (p \circ \rho_f)(p \circ \rho_d) = \psi_f \psi_d$$

so the assignment $d \mapsto [\psi_d]$ is really a group homomorphism.

Now, composing the homomorphisms in one direction, we have for a deck transformation d (by Proposition 3.3.3)

$$d_{\widetilde{\psi_d}(e)} = d_{d(e)} = d$$

In another direction, for a loop φ in B

$$\psi_{d_\varphi} = p \circ \rho_{d_{\tilde{\varphi}(e)}} \sim p \circ (\tilde{\varphi}) = \varphi$$

since all paths between points in E are homotopic. Thus we have inverse isomorphisms. □