METRIC SPACES FROM THE POINT OF VIEW OF KNOT THEORY

SCOTT A. TAYLOR

ABSTRACT. These notes are intended for an undergraduate course in knot theory, where the students know how to read, write, and construct proofs.

CONTENTS

1. METRIC SPACES

When we say that two locations on earth are a certain distance apart, we are assigning a non-negative real number to pairs of points on the earth. That is, "distance on earth" is a function from the Cartesian product of the earth with itself to the interval $[0, \infty)$. We expect any talk of distance to make use basic "facts" such that the distance between two points doesn't depend on which one comes first, distinct points can't have a distance of 0, and taking detours can only increase distance. We enshrine these principles in a collection of axioms.

Definition 1.1. A set X together with a function $d\colon X \times X \to [0, \infty)$ is a **metric space** if the following hold:

(1) For all $x \in X$, $d(x, x) = 0$
(2) (Definite) If $x, y \in X$ and $d(x, y) = 0$, then $x = y$.
(3) (Symmetry) For all $x, y \in X$, $d(x, y) = d(y, x)$.
(4) (Triangle Inequality) For all $x, y, z \in X$,
$$d(x, z) \leq d(x, y) + d(y, z).$$

Date: October 13, 2020.

The function d is a **metric**. The elements of X are **points**. If d is clear from the context, we just refer to X as a metric space. We also say that (X, d) is a metric space. If (X, d) satisfies all the conditions for being a metric space except, possibly, the Definite Property, we say that (X, d) is a **pseudo-metric space** and that d is a **pseudo-metric**.

Example 1.2. (Euclidean metric) Recall that if $a = (a_1, a_2, \dots, a_n)$ and $y = (b_1, b_2, \dots, b_n)$ are elements of $\mathbb{R}^n$, then the **dot product** is

$$a \cdot b = a_1 b_1 + a_2 b_2 + \cdots a_n b_n.$$

We define $||a|| = \sqrt{a \cdot a}$. Suppose that $X \subset \mathbb{R}^n$. The **euclidean metric** on X is defined by:

$$d(x, y) = ||x - y|| = \sqrt{(x - y) \cdot (x - y)}$$

for all $x, y \in X$. Proving that (X, d) is a metric space is a little tricky. A proof of a more general fact is proved at the end of the section. When $n = 1$, notice that

$$d(x, y) = \sqrt{(x - y)^2} = |x - y|,$$

which is the usual measure of distance in $\mathbb{R}$. It is a bit tedious, but hopefully straightforward, to prove that d is a metric in this case. In the general case, the proof would take us too far afield. The proof is easy to find online if you would like to see it.

Here are some more exotic examples of metrics on subsets of $\mathbb{R}^2$. Which of the metric space axioms are easy (if tedious) to prove? Which are difficult?

Example 1.3. Let $X \subset \mathbb{R}^2$. Define

$$d((x_1, x_2), (y_1, y_2)) = |x_1 - x_2| + |y_1 - y_2|$$

for all $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$. This is called the **taxicab metric**.

Example 1.4. Let $X \subset \mathbb{R}^2$. For $x, y \in X$ if there is a line in $\mathbb{R}^2$ passing through all of x and y and the origin, define $d(x, y) = ||x - y||$. Otherwise define $d(x, y) = ||x|| + ||y||$. The metric d is called the **Paris metric** on X, since unlesss two points are on the same line through the origin, the shortest path between them is to travel from x to the origin and then out to y. Do you see why? What's the connection to Paris?

Example 1.5. Let $X \subset \mathbb{R}^2$. Suppose that $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are elements of X. If $x_1 = y_1$, define $d(x, y) = |x_2 - y_2|$. Otherwise, define $d(x, y) = |x_2| + |x_1 - y_1| + |y_2|$. The metric d is called the **comb metric** on X. Sketch a picture of the shortest paths between different pairs of points to see why.

The following theorem shouldn't be surprising:

Theorem 1.6. *Suppose that (A, d) is a metric space and that $X \subset A$. Then (X, d) is a metric space, where we consider $d \colon A \times A \to [0, \infty)$ as the restriction of $d \colon X \times X \to [0, \infty)$. (That is, we only allow ourselves to measure the distance between elements of X, rather than between elements of A.)*

Despite the previous suggestions for trying to visualize the example metrics, we don't yet have a precise definition of "path" in a metric space. Don't worry, we'll get there! Here are two examples of metric spaces where it's not clear how to visualize paths.

Example 1.7. Suppose that X is any nonempty set. For $x, y \in X$, define $d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$.

The metric d is called the **discrete metric** on X. How does the discrete metric on $\mathbb{R}^2$ compare to the euclidean metric?

Example 1.8. Let $X \subset \mathbb{R}^n$ be any nonempty, closed and bounded subset of $\mathbb{R}^n$. Let $C^0(X) = \{f\colon X \to \mathbb{R}\}$ denote the set of continuous functions from X to $\mathbb{R}$. For $f, g \in C^0(X)$, define

$$d(f,g) = \max\{|f(x) - g(x)| : x \in X\}.$$

By the Extreme Value Theorem from (multivariable) calculus, $d(f,g)$ is well defined. Showing that d is a metric on $C^0(X)$ requires various facts from Calculus that we won't go into. You might like to try to prove that d is a metric.

We conclude with the product metric, something we'll make a lot of use of.

Example 1.9. Suppose that $(X, d_X)$and (Y, d_Y) are metric spaces. For $(a,b), (c,d) \in X \times Y$, define

$$d\big((a,b),(c,d)\big) = \sqrt{d_X(a,c)^2 + d_Y(b,d)^2}.$$

Then d is called the **product metric** on $X \times Y$. The next lemma shows it is a metric.

Lemma 1.10. *The product metric is a metric.*

Proof. Let d be the product metric on $X \times Y$ where (X, d_X) and (Y, d_Y) are metric spaces. We prove only that d satisfies the triangle inequality. The other aspects of being a metric are hopefully straightforward to prove.

Let $(a,b), (c,d), (e,f) \in X \times Y$. Set $A = d_X(a,c)$, $B = d_Y(b,d)$, $C = d_X(c,e)$, $D = d_Y(d,f)$. Then

$$\begin{aligned}
\Big(d\big((a,b),(c,d)\big) + d\big((c,d),(e,f)\big)\Big)^2 &= d\big((a,b),(c,d)\big)^2 + d\big((c,d),(e,f)\big)^2 + 2d\big((a,b),(c,d)\big)d\big((c,d),(e,f)\big) \\
&= A^2 + C^2 + B^2 + D^2 + 2d\big((a,b),(c,d)\big)d\big((c,d),(e,f)\big) \\
&= \big(A+C\big)^2 + \big(B+D\big)^2 + 2d\big((a,b),(c,d)\big)d\big((c,d),(e,f)\big) - 2AC - 2BD
\end{aligned}$$

Observe that

$$d\big((a,b),(c,d)\big)d\big((c,d),(e,f)\big)^2 = (A^2+B^2)(C^2+D^2) = A^2C^2 + B^2D^2 + B^2C^2 + A^2D^2$$

and that

$$(AC+BD)^2 = A^2C^2 + B^2D^2 + 2ACBD.$$

Hence,

$$\begin{aligned}
2d\big((a,b),(c,d)\big)d\big((c,d),(e,f)\big) - 2AC - 2BD &= 2(B^2C^2 + A^2D^2 - ACBD) \\
&= B^2C^2 + A^2D^2 + (B^2C^2 + A^2D^2 - 2ABCD) \\
&= B^2C^2 + A^2D^2 + (BC - AD)^2 \\
&\geq 0
\end{aligned}$$

Thus, applying the triangle inequality for d_X and d_Y, we see:

$$\begin{aligned}
\Big(d\big((a,b),(c,d)\big) + d\big((c,d),(e,f)\big)\Big)^2 &\geq \big(A+C\big)^2 + \big(B+D\big)^2 \\
&\geq d_X(a,e)^2 + d_Y(b,f)^2
\end{aligned}.$$

Consequently,

$$d\big((a,b),(c,d)\big) + d\big((c,d),(e,f)\big) \geq \sqrt{d_X(a,e)^2 + d_Y(b,f)^2} = d\big((a,b),(e,f)\big).$$

□

Corollary 1.11. *The euclidean metric is a metric on any subsets of* $\mathbb{R}^n$.

Proof. The case when $n = 1$ is straightforward, if tedious. When $n \geq 2$, the result can be proved by induction. □

2. Open Sets

Definition 2.1. Suppose that (X,d) is a metric space, that $a \in X$ and that $r > 0$. The **open ball centered at a of radius r** is

$$B_r(a) = \{x \in X : d(a,x) < r\}.$$

A subset $U \subset X$ is **open** (more precisely, an **open subset of** X) if for every $a \in U$, there exists $r > 0$ such that $B_r(a) \subset U$.

We think of an open ball centered at a as the set of points that are "near" to a. A subset is open if whenever a point is in the subset, nearby points are as well; but what "near" means gets to depend on the point.

Example 2.2. Let $X = \{(x,y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$ and let d be the euclidean metric on X. Let $a = (0,0)$ and let $b = (2,2)$. Set $r = 1$. Sketch pictures of $B_r(a)$ and $B_r(b)$.

Example 2.3. Let $X = \mathbb{R}^2$ and let d be the comb metric. Let $a = (0,0)$. Let $b = (2,2)$. Let $c = (1/2, 1/2)$. Sketch $B_r(a)$, $B_r(b)$, and $B_r(c)$.

The next lemma ensures that we have consistent uses of the word "open."

Lemma 2.4. *Suppose that (X,d) is a metric space.*

(1) Suppose that $a \in X$ and $r > 0$. Then $B_r(a)$ is an open subset of X.
(2) Suppose that $U \subset X$ is an open set. Consider (U,d) as a metric space (i.e. d is the subspace metric on U; the restriction of d to U). Suppose that $A \subset U$. Then A is an open subset of U if and only if A is an open subset of X.

Example 2.5. Let $X = \mathbb{R}^2$ with the Euclidean metric. Let $U = \{(x,y) \in X : x \geq 0, y \geq 0\}$. Explain why U is not open. Consider the set $V = \{(x,y) : x \geq 0, y \geq 0, x^2 + y^2 \leq 1\}$. (This is the open ball centered at $(0,0)$ in U.) Explain why V is open in U but not in X. Thus, the hypothesis that U be open in part (2) of Lemma 2.4 is necessary.

The next theorem lists some fundamental properties of open sets. These properties can be abstracted into the set of axioms for a "topology," but we will not need to do that in this course.

Theorem 2.6 (Open sets in metric spaces form a topology)**.** *Suppose that (X,d) is a metric space. Then the following hold.*

(1) The subsets $\varnothing \subset X$ and $X \subset X$ are open subsets of X.
(2) (Finite intersections of open sets are open) If $U, V \subset X$ are both open, then $U \cap V$ is open.
(3) (Arbitrary unions of open sets are open) If $U_\alpha \subset X$ is open for all $\alpha \in \Lambda$, then $\bigcup_{\alpha \in \Lambda} U_\alpha$ is open. (Λ is a nonempty index set.)

The complements of open sets are also useful.

Definition 2.7. Suppose that (X,d) is a metric space. A subset $F \subset X$ is **closed** or a **closed subset of** X if its complement F^C is open.

Warning: "Open" and "Closed" are **not** negations of each other.

Example 2.8. Let $X = \mathbb{R}^2$ and let d be the Euclidean metric. The set X itself is both open and closed. The set $\{(x,y) \in \mathbb{R}^2 : |x| < 1, y = 0\}$ is neither open nor closed. The open ball $B_1(0,0)$ is open, but not closed. The set $\{(x,y) \in \mathbb{R}^2 : |x| \leq 1, |y| \leq 1\}$ is closed but not open.

Definition 2.9. Suppose that (X, d) is a metric space. A sequence (x_n) in X **converges** to a point $a \in X$ if for every $\epsilon > 0$, the sequence (x_n) is eventually entirely contained in $B_\epsilon(a)$. More formally, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $x_n \in B_\epsilon(a)$. We say that a is the **limit** of (x_n).

Example 2.10. Let $X = \mathbb{R}^2$ and let d be the Euclidean metric. Let $x_n = (1/n, 0)$ for every $n \in \mathbb{N}$. The sequence (x_n) converges to $(0, 0)$. To see this, assume $\epsilon > 0$ is given. Let $N \in \mathbb{N}$ be any integer such that $N > 1/\epsilon$. If $n \geq N$, then $n > 1/\epsilon$ and so $1/n < \epsilon$. Consequently, $x_n \in B_\epsilon(0, 0)$ for all $n \geq N$.

Example 2.11. Let X be any nonempty set and let d be the discrete metric on X. Suppose that (x_n) is a sequence in X and that $a \in X$. Then (x_n) converges to a if and only if (x_n) is eventually constant at a. To see this, suppose first that (x_n) converges to a. Let $\epsilon = 1/2$. Then for large enough n, $d(x_n, a) < 1/2$. However the discrete metric only takes the values 0 or 1, so for large enough n, $d(x_n, a) = 0$. By the definition of metric, for large enough n, $x_n = a$.

Now suppose that (x_n) is eventually constant at a. That is, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $x_n = a$. Let $\epsilon > 0$ be given. Then for all $n \geq N$,

$$d(x_n, a) = d(a, a) = 0 < \epsilon.$$

Thus, for all $n \geq N$, $x_n \in B_\epsilon(a)$ and so (x_n) converges to a.

More generally:

Example 2.12. Suppose that (X, d) is a metric space and that (x_n) is a sequence which is eventually constant at $a \in X$. Then (x_n) converges to a.

Example 2.13. Give $\mathbb{R}^2$ and let $U = B_1(0, 0) \subset \mathbb{R}^2$ also have the euclidean metric. For all $n \in \mathbb{N}$, let $x_n = (1 - 1/n, 0)$. Then (x_n) converges to $(1, 0)$ in $\mathbb{R}^2$, but does not converge to any point in U.

Theorem 2.14 (Closed sets contain their limit points)**.** *Let (X, d) be a metric space and let $F \subset X$. Then F is closed if and only if whenever a sequence (x_n) in F converges to $a \in X$, then $a \in F$.*

Theorem 2.15. *Suppose that (X, d_X) and (Y, d_Y) are metric spaces and that d is the product metric on $X \times Y$. Prove that a subset $U \subset X \times Y$ is open in $X \times Y$ with the product metric if and only if for all $(a, b) \in U$, there exist open sets $U_X \subset X$ and $U_Y \subset Y$ such that $U_X \times U_Y \subset U$ and $a \in U_X$ and $b \in U_Y$.*

Definition 2.16. Suppose that X is a set and that d and d' are metrics on X. We say that d and d' are **topologically identical** if a subset $U \subset X$ is open in (X, d) if and only if it is open in (X, d').

Lemma 2.17. *Suppose that X is a set and that d and d' are metrics on X. Then d and d' are* ***topologically identical*** *if and only if both of the following conditions hold:*

(1) For every $a \in X$ and $r > 0$, the ball $B_r(a) = \{x \in X : d(x, a) < r\}$ is an open set in (X, d').
(2) For every $a \in X$ and $r > 0$, the ball $B'_r(a) = \{x \in X : d'(x, a) < r\}$ is an open set in (X, d).

Example 2.18. Let $X = \mathbb{R}^2$ and let d be the euclidean metric and d' be the taxicab metric. Then d and d' are topologically identical.

Example 2.19. Let $X = \mathbb{R}^2$ and let d be the euclidean metric and d' be the comb metric. Then d and d' are not topologically identical.

3. Continuity

In this section we explore different definitions of continuity, as well as some basic properties.

Definition 3.1. Suppose that X and Y are sets and that $f\colon X \to Y$ is a function. (Recall that this means that there is a unique $f(x) \in Y$ for every $x \in X$.) If $A \subset X$, we abuse notation and define

$$f(A) = \{y \in Y : \exists x \in A \,\text{s.t.}\, f(x) = y\}$$

and if $B \subset Y$

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

Notice that $f(A)$ is a subset of Y and $f^{-1}(B)$ is a subset of X.

Definition 3.2 (Metric definition of continuity)**.** Suppose that (X, d_X) and (Y, d_Y) are metric spaces and that $f\colon X \to Y$ is a function. Then f is (metrically) continuous if for every $a \in X$ and every $\epsilon > 0$, there exists $\delta > 0$ such that

$$f(B_\delta(a)) \subset B_\epsilon(f(a)).$$

Here is the definition of continuous function usually presented in calculus or analysis classes.

Definition 3.3 (Analytic definition of continuity)**.** Suppose that (X, d_X) and (Y, d_Y) are metric spaces. A function $f\colon X \to Y$ is **(analytically) continuous** if for every $a \in X$ and every $\epsilon > 0$, there exists $\delta > 0$ such that if $d_X(a, x) < \delta$ then $d_Y(f(a), f(x)) < \epsilon$.

Here is the topological definition of continuity. We summarize it by saying that inverse images of open sets are open.

Definition 3.4 (Topological definition of continuity)**.** Suppose that (X, d_X) and (Y, d_Y) are metric spaces. A function $f\colon X \to Y$ is **(topologically) continuous** if for every open set $U \subset Y$, the subset $f^{-1}(U)$ is open in X.

We can also use sequences to define continuity.

Definition 3.5 (Sequential definition of continuity)**.** Suppose that (X, d_X) and (Y, d_Y) are metric spaces. A function $f\colon X \to Y$ is **(sequentially) continuous** if whenever a sequence (x_n) converges to a point $a \in X$, the sequence $(f(x_n))$ in Y converges to $f(a) \in Y$.

Theorem 3.6 (The different definitions of continuity are equivalent)**.** *Suppose that (X, d_X) and (Y, d_Y) are metric spaces and that $f\colon X \to Y$ is a function. Then the following are equivalent:*

(1) f is metrically continuous.
(2) f is analytically continuous.
(3) f is topologically continuous.
(4) f is sequentially continuous.

Proof. We show that (4) $\Rightarrow$ (3) by proving the contrapositive. Assume that f is not topologically continuous. Then there exists an open $U \subset Y$ such that $f^{-1}(U)$ is not open in X. By definition, there exists $a \in f^{-1}(U)$ such that for every $\epsilon > 0$, $B_\epsilon(a) \not\subset f^{-1}(U)$. In particular, for all $n \in \mathbb{N}$, there exists $x_n \in B_{1/n}(a) \setminus f^{-1}(U)$. Notice that (x_n) converges to a. Since $x_n \notin f^{-1}(U)$, by definition $f(x_n) \notin U$. However, $f(a) \in U$. Since U is open, there exists $r > 0$ such that $B_r(f(a)) \subset U$. Thus, for all $n \in \mathbb{N}$, $x_n \notin B_r(f(a))$. Thus, the sequence $(f(x_n))$ cannot converge to $f(a)$. Thus, f is not sequentially continuous.

Now we show that (3) $\Rightarrow$ (4). We prove this directly. Assume that f is topologically continuous and that (x_n) is a sequence in X converging to $a \in X$. We show that $(f(x_n))$ converges to $f(a)$. Let $\epsilon > 0$ and recall that $B_\epsilon(f(a))$ is open in Y. Since f is topologically continuous, $U = f^{-1}(B_\epsilon(f(a)))$ is open in X. Thus, there exists $\delta > 0$, such that $B_\delta(a) \subset U$.

$\langle$ Finish the proof that f is sequentially continuous. $\rangle$

We leave the other parts of the proof of the theorem as an exercise. $\square$

Henceforth, we just use "continuous" to mean any of these equivalent forms. Sometimes we refer to a continuous function as a **map**.

Exercise 3.7. Show that a function $f\colon X \to Y$ between metric spaces is continuous if and only if for every closed set $V \subset Y$, $f^{-1}(V)$ is closed in X.

Theorem 3.8. *Suppose that X, Y, and Z, are metric spaces and that $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous. Prove that $g \circ f\colon X \to Z$ is continuous.*

Definition 3.9. Suppose that X and Y are sets. The **projections** $\pi_X X \times Y \to X$ and $\pi_Y X \times Y \to Y$ are the functions such that $\pi_X((x, y)) = x$ and $\pi_Y((x, y)) = y$ for all $(x, y) \in X \times Y$.

Exercise 3.10. Suppose that X and Y are metric spaces and that $X \times Y$ has the product metric. Prove that the projections π_X and π_Y are continuous.

Exercise 3.11. Suppose that (X, d_X) and (Y, d_Y) are metric spaces and that $X \times Y$ is given the product metric d. Let (Z, d_Z) be any metric space and suppose that $f\colon Z \to X \times Y$ is a function. Prove that f is continuous if and only if the compositions $\pi_X \circ f$ and $\pi_Y \circ f$ are continuous.

Example 3.12. Define $\gamma\colon [0, 1] \to \mathbb{R}^2$ by $\gamma(t) = (\cos(t), \sin(t))$. Assume both $[0, 1]$ and $\mathbb{R}^2$ have the euclidean metrics, γ is continuous since the projections $\cos(t)$ and $\sin(t)$ are continuous.

Example 3.13 (Torus knots)**.**

The following lemma helps with proving that functions defined piecewise are continuous.

Lemma 3.14 (Gluing lemma)**.** *Suppose that (X, d_X) is a metric space that that $A, B \subset X$ with $X = A \cup B$. Assume that A, B are closed subsets of X. Suppose that (Z, d_Z) is a metric space that that $f\colon A \to Z$ and $g\colon B \to Z$ are continuous and have the property that $f(x) = g(x)$ for all $x \in A \cap B$. Define $f \cup g\colon X \to Z$ by*

$$f \cup g(x) = \begin{cases} f(x) & x \in A \\ g(x) & x \in B \end{cases}.$$

Then $f \cup g$ is continuous.

Note that in the lemma, we are assuming that $f\colon A \to Z$ and $g\colon B \to Z$ are continuous when we consider A and B with the subspace metrics.

Proof. Note that $f \cup g$ is well-defined since f and g agree on $A \cap B$. Since A and B are closed so is $A \cap B$. Let $F \subset Z$ be a closed set. Observe that $(f \cup g)^{-1}(F) = f^{-1}(F) \cup g^{-1}(F)$. Since f is continuous $f^{-1}(F) \subset A$ is a closed subset of A. Since A is a closed subset of X, $V_A = f^{-1}(F)$ is also a closed subset of X. (Do you see why?) Similarly, $V_B = g^{-1}(F)$ is also a closed subset of X. By DeMorgan's Laws, $(V_A \cup V_B)^C = V_A^C \cap V_B^C$. Since V_A^C and V_B^C are open and since the finite intersection of open sets is open, $V_A^C \cap V_B^C$ is open. Thus, $V_A \cup V_B = (f \cup g)^{-1}(F)$ is closed. Thus the inverse image of closed sets is closed, and so $(f \cup g)$ is continuous. $\square$

Definition 4.1. Suppose that $f\colon X \times Y$ is a function between metric spaces. It is a **homeomorphism** if it is a bijection and both f and its inverse f^{-1} are continuous. (That is, a homeomorphism is a bicontinuous bijection.)

Notice that the inverse of a homeomorphism is also a homeomorphism. If there is a homeomorphism from X to Y we say that X and Y are **homeomorphic**. Being homeomorphic is an equivalence relation on metric spaces.

Example 4.2. Suppose that $X = [0,1]$ and $Y = [3,8]$, both with the euclidean metric. Let $f\colon X \to Y$ be defined by $f(t) = 5t+3$. Then f is a continuous bijection and its inverse $f^{-1}(s) = (s-3)/5$ is also continuous. Thus f is a homeomorphism.

Example 4.3. Let $A = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2+y^2 \leq 4\}$. Define $h\colon A \to A$ by $h(r,\theta) = (r, \theta + 2\pi r)$ where (r,θ) are the polar coordinates of a point in A. Then h is a homeomorphism of A to itself, called a **Dehn twist**. See Figure 1 for a depiction of the effect of the Dehn twist h. Let $\widehat{h}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$\widehat{h}(x,y) = \begin{cases} h(x,y) & (x,y) \in A \\ (x,y) & (x,y) \notin A \end{cases}.$$

Then $\widehat{h}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is also homeomorphism by the gluing lemma.

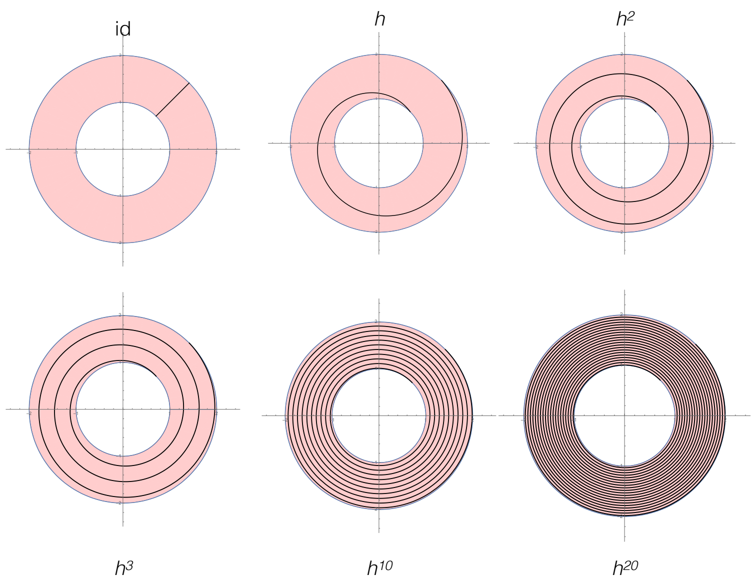

FIGURE 1. The Dehn twist h as well as some examples of its powers.

Example 4.4. Let S^1 be the unit circle $\{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 1\}$ and let $Q = \{(x,y) \in \mathbb{R}^2 : \max\{|x|,|y|\} \leq 1\}$ be the unit (hollow) square. For each point $q \in Q$, let $\rho(q)$ be the point on S^1 on the ray emanating from the origin and passing through q. Then $\rho\colon Q \to S^1$ is a homeomorphism.

Example 4.5 (Stereographic Projection)**.** Let S^1 be the unit circle as in the previous example. Let $N = (0, 1)$ be its north pole. Let L be the line with equation $y = -1$ in $\mathbb{R}^2$. Note that L with the euclidean metric is homeomorphic to $\mathbb{R}$. Define $\rho\colon\ S^1 \setminus \{N\} \to L$ as follows. Let $x \in S^1 \setminus \{N\}$. Let $\rho(x) \in L$ be the point on L that lies on the ray emanating from N and passing through x. Then ρ is a homeomorphism. In particular, $\mathbb{R}$ is homeomorphic to the result of removing a point from a circle.

More generally, let $S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$ and let N be its north pole. Then $S^n \setminus \{N\}$ is homeomorphic to $\mathbb{R}^n$. The case when $n = 2$ can be obtained from the case $n = 1$ by rotating the picture from the previous paragraph around the y-axis. Similarly, the statement for arbitrary $n \geq 2$ can be obtained from the statement for n by rotating around an axis. Of special interest to us is the fact that $S^3 \setminus \{\text{point}\}$ is homeomorphic to $\mathbb{R}^3$. We can alternatively phrase this by saying that the 3-sphere S^3 is obtained by adding a point "at infinity" to $\mathbb{R}^3$.

Definition 4.6. Suppose that X and Y are metric spaces and that $f\colon\ X \to Y$ is an injective continuous function. Note that we can give $f(X) \subset Y$ the subspace metric from Y. We say that f is an **embedding** of X in Y if restricting the codomain of f to its range, makes $f\colon\ X \to f(X)$ a homeomorphism. (That is an embedding is a homeomorphism onto its image.)

Example 4.7. Define $f\colon\ [0, 1] \to \mathbb{R}^2$ by $f(t) = (t, t^2)$ for all $t \in [0, 1]$. Then f is an embedding of the unit interval into $\mathbb{R}^2$. The continuous function $f\colon\ [0, 1] \to \mathbb{R}^2$ defined by $f(t) = (\cos 2\pi t, \sin 4\pi t)$ is not an embedding because it is not injective. Finally, consider the image of the continuous function shown in Figure 2. It is the image of $f\colon\ (0, 5] \to \mathbb{R}^2$ defined by

$$f(t) = \begin{cases} (t, \sin(1/t)) & t \in (0, 1] \\ (1, (2-t)\sin(1) + 2(t-1)) & t \in (1, 2] \\ ((3-t) + (-1)(t-2), 2) & t \in (2, 3] \\ (-1, 2(4-t)) & t \in (3, 4] \\ (-(5-t), 0) & t \in (4, 5] \end{cases}$$

This is injective and continuous, but not an embedding. To see that it is an embedding let $p = (0, 0) = f(5)$. In the interval $(0, 5]$, an open ball of radius ϵ is the interval $(5 - \epsilon, 5]$, which is a connected interval. On the other hand, in $\mathbb{R}^2$, no matter how small ϵ is the ball $B_\epsilon(p) \cap \operatorname{im}(f)$ is the union of infinitely many intervals. So we can believe (even if not rigorously prove right now) that f is not a homeomorphism onto its image.

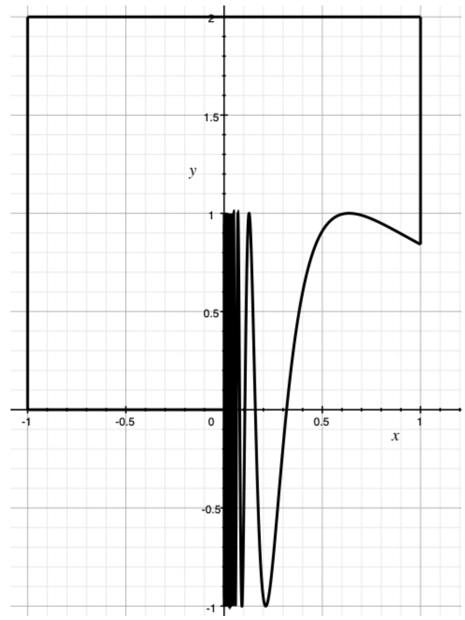

Figure 2. A version of the topologists' sine curve. Notice that as the curve comes into the y-axis from the right, there are infinitely many oscillations.

Definition 4.8. A **knot** is either an embedding of S^1 into S^3 or into $\mathbb{R}^3$. We also sometimes refer to the image of the embedding as a knot.

Although we do not explore them in this course, we note that we can discuss knots in all dimensions by considering embeddings of S^k into S^n for $n > k$. The study of knotted S^2s in S^4 is a particularly rich and active area of contemporary research.

We have defined a knot as an embedding of S^1 in S^3. This allows for so-called **wild knots**. Here is an example, which we explain rather informally.

Example 4.9. Consider the box $B = [-1,1] \times [-1,1] \times [-1,1]$. Let $\gamma_B \colon [0,1] \to B$ be a path from $(0,0,-1)$ to $(0,0,+1)$ and that forms a trefoil, as in the picture. Let B_n be the result of scaling B by a factor of $1/2^n$ and translating it a distance of $2(n-1)$ along the z-axis. Let $\gamma_n(t) = \frac{1}{2^n}\gamma(2^n t + 1/2 - 1/2^n)$ be the same curve as γ_B except traversing it over the interval $[(2^{n-1}-1)/2^{n-1}, (2^n-1)/2^n]$ and shifted to lie in B_n. Let $\phi = \bigcup_{n\in\mathbb{N}} \gamma_n$ and then construct a knot γ in $\mathbb{R}^3$ by taking the union of $\gamma(2t-1)$ for $t \in [0,1/2]$ with an embeddeding of $[1/2,1]$ from $(0,0,1)$ to $(0,0,0)$ having range disjoint from $\bigcup B_n$. See Figure 3.

5. Homotopy, Isotopy, and Knot Equivalence

For the remainder of the course we let $I = [0,1] \subset \mathbb{R}$ with the Euclidean metric.

Definition 5.1. Suppose (X,d) is a metric pace. A **path** in X is a continuous function $\gamma \colon I \to X$. If $p = \gamma(0)$ and $q = \gamma(1)$, we say that γ is a path from p to q. If $p = q$, then γ is a **loop** or a **loop** based at p. The metric space X is **path-connected** if there is a path between each pair of points.

Exercise 5.2. Let X be a metric space. For $x, y \in X$, define $x \sim y$ if and only if there is a path in X from x to y. Prove that $\sim$ is an equivalence relation. The equivalence classes are called the **path components** of X.

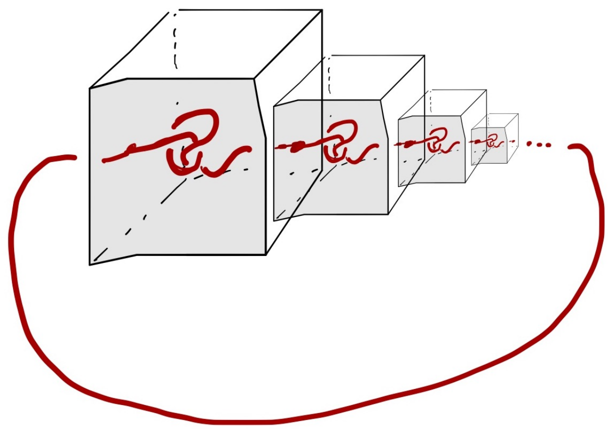

FIGURE 3. An example of a wild knot. The boxes just help us describe the scaling and translation.

There are examples of surjective paths in $I \times I$, so we will often need to put more conditions on our paths. Note that if $\gamma(t) = (x_1(t), x_2(t), \ldots, x_n(t))$ is a path in $\mathbb{R}^n$ (or some subset of $\mathbb{R}^n$), its derivative can be obtained by just taking the derivative of each of the coordinates.

Definition 5.3. If $X \subset \mathbb{R}^n$, a path γ in X is:

- **smooth** if $\gamma'(t)$ exists for all t and if $||\gamma'(t)|| \neq 0$ for any t. This means that the one-sided derivatives at 0 and 1 exist and are non-zero.
- **piecewise smooth** if there exist
$$t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$$
such for each $i \in \{0, \ldots, n-1\}$, on the interval $[t_i, t_{i+1}]$ the derivative $\gamma'(t)$ exists and is non-zero; at t_i and t_{i+1} we use one-sided derivatives.
- **piecewise linear** if there exist
$$t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$$
such for each $i \in \{0, \ldots, n-1\}$, on the interval $[t_i, t_{i+1}]$, γ is linear. That is, the derivative $\gamma'(t)$ exists and is a non-zero constant vector; at t_i and t_{i+1} we use one-sided derivatives.

Henceforth, we restrict our attention to tame knots.

Definition 5.4. A knot in $\mathbb{R}^3$ is **tame** if it is piecewise smooth or piecewise linear. A knot in S^3 is tame if for any point $p \in S^3$ not contained in the range of the knot, the result of stereographic projection from p is a tame knot in $\mathbb{R}^3$.

As we've discussed, two different knots can represent the same knot type. How can we formalize the notion of knot equivalence?

One natural method of formalizing it is by declaring two knots $K \subset S^3$ and $K' \subset S^3$ (thinking of a knot as a subset of S^3, rather than as a function) if and only if there exists a homeomorphism $h\colon S^3 \to S^3$ such that $h(K) = K'$. (We say that such a homeomorphism h is a **homeomorphism of pairs** and write $h\colon (S^3, K) \to (S^3, K')$.) The problem with this definition is that it makes a knot always equivalent to its mirror image. To correct this, we consider orientations, unfortunately a purely topological definition of orientation is beyond the scope of this course, so we allow ourselves to use some calculus.

Definition 5.5. Suppose that $h\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a homeomorphism. It is a **diffeomorphism** if all partial derivatives of all orders exist for both h and h^{-1}. A homeomorphism $h\colon S^3 \to S^3$ is a **diffeomorphism** if for all $p \in S^3$, the composition of the restricted map $f\colon S^3 \setminus \{p\} \to S^3 \setminus \{f(p)\}$ with stereographic projection from p is a diffeomorphism of $\mathbb{R}^3$.

Definition 5.6. Suppose that $h\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a diffeomorphism. Notice that for each $p \in \mathbb{R}^3$ we can write $h(p) = (h_1(p), h_2(p), h_3(p))$. Its **derivative** at a point $p \in \mathbb{R}^3$ is the matrix of partial derivatives:

$$Dh(p) = \begin{pmatrix} \frac{\partial h_1}{\partial x}(p) & \frac{\partial h_1}{\partial y}(p) & \frac{\partial h_1}{\partial z}(p) \\ \frac{\partial h_2}{\partial x}(p) & \frac{\partial h_2}{\partial y}(p) & \frac{\partial h_2}{\partial z}(p) \\ \frac{\partial h_3}{\partial x}(p) & \frac{\partial h_3}{\partial y}(p) & \frac{\partial h_3}{\partial z}(p) \end{pmatrix}$$

Suppose that $h\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a diffeomorphism and that $p \in \mathbb{R}^3$. At p, we choose coordinate vectors e_x, e_y, e_z pointing along the positive x, y, and z axes, respectively. Note that they form a right-handed coordinate system. The vectors $Dh(p)e_x$, $Dh(p)e_y$, and $Dh(p)e_z$ are three linearly independent vectors based at $h(p)$. They form either a right-handed or left-handed coordinate system. If they form a right handed system then h is **orientation-preserving**. If they form a left-handed system, then $Dh(p)$ is **orientation-reversing**. It is a fact that h is orientation preserving if and only if $\det Dh(p) > 0$ and orientation-reversing if and only if $\det Dh(p) < 0$. As we did with the definition of diffeomorphism we can extend the definitions of orientation-preserving and orientation-reversing to S^3 by using stereographic projection. We could then define two knots $K, K' \subset S^3$ to be equivalent if there is an *orientation-preserving* homeomorphism of pairs $(S^3, K) \to (S^3, K')$. This turns out to be a perfectly valid definition, although it does require that we consider diffeomorphisms rather than just homeomorphisms. Unfortunately, it does not accord with our intuition that two equivalent knots should be able to be moved within 3-space to coincide, e.g. via moves that project to Reidemeister moves on a projection sphere. To get a definition along those lines, we introduce homotopies and isotopies.

Definition 5.7. Suppose that X and Y are metrics spaces and that $f, g\colon X \to Y$ are continuous. A **homotopy** from f to g is a continuous function

$$H\colon X \times I \to Y$$

such that for all $x \in X$, $H(x,0) = f(x)$ and $H(x,1) = g(x)$. If $A \subset X$, the homotopy is **relative to** A if $H(a,t) = f(a)$ for all $a \in A$ and all $t \in I$. We often write $H_t(x)$ in place of $H(x,t)$. Note that $H_t\colon X \to Y$ is a continuous function for all $t \in I$. If each H_t is an embedding then H is an **isotopy** and we say that f is **isotopic** to g.

Exercise 5.8. Prove that the notion of being homotopic (or isotopic!) is an equivalence relation on the set of continuous functions from X to Y.

Example 5.9. Let $i\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the identity map and let $c\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the constant map such that $c(x) = (0,0,0)$ for all $x \in \mathbb{R}^3$. Then i and c are homotopic. One such homotopy is given by $H(x,t) = (1-t)x$ for all $x \in \mathbb{R}^3$ and $t \in I$.

Definition 5.10. A metric space is **contractible** if there is a homotopy from the identity map to a constant map. A map $S^1 \to X$ is **contractible** if it is homotopic to a constant map. A path-connected metric space is **simply connected** if every map $S^1 \to X$ is contractible.

Notice that $\mathbb{R}^3$ is contractible.

Exercise 5.11. A contractible space is both path-connected and simply-connected.

Exercise 5.12. The 0-sphere $S^0 = \{-1, 1\} = \{x \in \mathbb{R} : x^2 = 1\}$ is not path-connected. (Hint: use the Intermediate Value Theorem)

Theorem 5.13. *The circle $S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is path-connected but not simply connected.*

Proof Sketch. We leave the fact that S^1 is path-connected as an exercise (Hint: Use polar coordinates). The proof that S^1 is not simply connected will be addressed later, or possibly not at all. □

Theorem 5.14. *Every sphere $S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$ is path connected and simply connected for $n \geq 2$.*

Proof. We leave the fact that S^n is path-connected as an exercise. (Hint: Use stereographic projection to $\mathbb{R}^n$.)

Suppose that $\gamma\colon S^1 \to \S^n$ is a loop. Since $n \geq 2$, it turns out that it is possible to homotope γ so that there exists $p \in S^n$ such that $p \notin \gamma(S^1)$. Let $\pi\colon S^n \setminus \{p\} \to \mathbb{R}^n$ be stereographic projection. Let $H\colon \mathbb{R}^n \times I \to \mathbb{R}^n$ be a contraction of $\mathbb{R}^n$. Define $\widehat{H}(x, t) = \pi^{-1}(H(\pi(\gamma(x)), t))$. Then $\widehat{H}$ is a homotopy of γ to a constant map. Thus, γ is contractible. □

Returning to knots, it would now be tempting to define two knots (thought of as embeddings $S^1 \to S^3$ to be equivalent if and only there is an isotopy of one to the other. Unfortunately, this would make every tame knot in S^3 equivalent to every other tame knot.

Example 5.15. Suppose that $K \subset S^3$ is a tame knot. Because it is tame, there is a 3-ball $B \subset S^3$ intersecting K in an arc and with the property that if we crush B down to a point the resulting knot is unknotted. The crushing of B defines an isotopy of K to an unknot.

The key to definining knot equivalence is to insist that the isotopy be defined knot just on the knot, but also near the knot. We can do this by insisting that the isotopy itself be smooth or piecewise linear, but we can also do it as follows:

Definition 5.16. Two links K, K' in S^3 are **equivalent** if there is an isotopy $H\colon S^3 \times I \to S^3$ such that for all $k \in K$, $H(k, 0) = k$ and $H(k, 1) \in K'$.

Allan Hatcher proved the Smale Conjecture in 1983. One consequence is that our definition of knot equivalence is logically equivalent to the definition involving orientation-preserving homeomorphisms of pairs. In 1989, Cameron Gordon and John Luecke proved that if $K, K' \subset S^3$ are knots and if there is an orientation-preserving homeomorphism $h\colon S^3 \setminus K \to S^3 \setminus K'$ then K and K' are equivalent. That is, knots are determined by their complements. That, however, is not true of links.

Example 5.17. Here are two links with homeomorphic complements.

6. Based loops and Fundamental Groups

(See the introduction to Chapter 1 and Section 1.1 of Hatcher's *Algebraic Topology.* Available online.)

Throughout this section, assume that (X, d) is a path-connected metric space.

Definition 6.1. Suppose that $\gamma\colon I \to X$ is a path such that $\gamma(0) = \gamma(1) = p \in X$. We say that γ is a **loop based at** p. A **based homotopy** from γ to ψ is a map $H\colon I \times I \to X$ such that:

(1) $H(t, 0) = \gamma(t)$ for all $t \in I$.
(2) $H(t, 1) = \psi(t)$ for all $t \in I$.
(3) $H(0, s) = H(1, s) = p$ for all $s \in I$.

Exercise 6.2. If γ and ψ are both loops based at p, define $\gamma \sim \psi$ if and only if there is a based homotopy from γ to ψ. Then $\sim$ is an equivalence relation on the set of based loops at p. We let $[\gamma]$ denote the equivalence class of γ. We let $\pi(X, p)$ denote the set of equivalence classes of loops based at p.

Definition 6.3. Suppose that γ and ψ are loops based at p. Define

$$\gamma \cdot \psi(t) = \begin{cases} \gamma(2t) & t \in [0, 1/2] \\ \psi(2t - 1) & t \in [1/2, 1] \end{cases}$$

for all $t \in I$ and

$$[\gamma][\psi] = [\gamma \cdot \psi]$$

Theorem 6.4. *As a binary operation on $\pi(X, p)$, the operation $\cdot$ is well-defined and associative.*

Definition 6.5. Let $p\colon I \to X$ be the constant function $p(t) = p$ (where this second p is the basepoint) for all $t \in I$. Let $\mathbb{1} = [p]$.

Theorem 6.6. *Let γ be a loop based at p. Then $[\gamma] \cdot \mathbb{1} = \mathbb{1} \cdot [\gamma] = [\gamma]$.*

Proof. We'll show that $[\gamma] \cdot \mathbb{1} = [\gamma]$. The proof of the other equality is similar. We must show that $\gamma \cdot \mathbb{1}$ is based homotopic to $\mathbb{1}$. Recall that for all $t \in I$:

$$\gamma \cdot \mathbb{1}(t) = \begin{cases} \gamma(2t) & t \in [0, 1/2] \\ p & t \in [1/2, 1] \end{cases}$$

We think of this as following γ twice as fast as usual and then sitting at the basepoint p for half the time. Our homotopy will slowly decrease the amount of time we sit at p, until we are there only at time $t = 1$. Define

$$F(t, s) = \begin{cases} \gamma(2t(1 - s) + ts) & t \in [0, 1/(2 - s)] \\ p & t \in [1/(2 - s), 1] \end{cases}.$$

⟨ Verify that F is a based homotopy from $\gamma \cdot \mathbb{1}$ to γ ⟩ □

Definition 6.7. Let γ be a loop based at p. Define

$$\overline{\gamma}(t) = \gamma(1 - t)$$

for all $t \in I$ and

$$[\gamma]^{-1} = [\overline{\gamma}].$$

Theorem 6.8. *For every loop γ based at p, $[\gamma]^{-1}$ is well-defined and*

$$[\gamma][\gamma]^{-1} = [\gamma]^{-1} \cdot [\gamma] = \mathbb{1}.$$

Corollary 6.9. *$\pi(X, p)$ is a group (called the **fundamental group** of X.)*

Often all we care about from a group is its isomorphism type.

Definition 6.10. Suppose that G and G' are groups and that $f\colon G \to G'$ is a function. Then f is a **homomorphism** if for all $a, b \in G$

$$f(ab) = f(a)f(b).$$

f is an **isomorphism** if it is a homomorphism and a bijection. An isomorphism from a group to itself is called an **automorphism**.

Theorem 6.11. *Suppose that $p, q \in X$. The groups $\pi(X,p)$ and $\pi(X,q)$ are isomorphic.*

Proof. Let α be a path from p to q. Suppose that γ is a loop based at p. Define

$$f(\alpha) = \alpha \cdot \gamma \cdot \overline{\alpha},$$

and

$$f([\alpha]) = [f(\alpha)].$$

We claim that $f\colon \pi(X,p) \to \pi(X,q)$ is a well-defined group isomorphism.

We prove it is well-defined, first. Suppose that $\gamma \sim \psi$. Let F be a based homotopy from γ to ψ. Define $H\colon I \times I \to X$ by

$$H(t,s) = \begin{cases} \alpha(3t) & t \in [0,1/3] \\ F(3t-1,s) & t \in [1/3,2/3] \\ \overline{\alpha}(3t-2) & t \in [2/3,1] \end{cases}$$

Then H is a based homotopy from $f(\gamma)$ to $f(\psi)$.

Now suppose that γ, ψ are two (not necessarily equivalent) loops based at p. Note that

$$\begin{aligned} f(\gamma) \cdot f(\psi) &\sim \alpha \cdot \gamma \cdot \overline{\alpha} \cdot \alpha \cdot \psi \cdot \overline{\alpha} \\ &\sim \alpha \cdot \gamma \cdot \psi \cdot \overline{\alpha} \\ &\sim f(\gamma \cdot \psi). \end{aligned}$$

Thus, f is a homomorphism.

To show it is an isomorphism, note that the function $g\colon \pi(X,q) \to \pi(X,p)$ defined by

$$g([\phi]) = [\overline{\alpha} \cdot \phi \cdot \alpha]$$

for all loops ϕ based at q is the inverse of f. □

Fundamental groups are homeomorphism invariants. We prove something stronger.

Theorem 6.12 (Functoriality of Fundamental Group)**.** *Suppose that X and Y are path-connected metric spaces with $p \in X$. Suppose that there exists a map $f\colon X \to Y$. Then there is a group homomorphism*

$$f_*\colon \pi(X,p) \to \pi(Y, f(p))$$

such that if f is homotopic to $g\colon X \to Y$ via a homotopy relative to p, then $f_* = g_*$. Furthermore if $h\colon Y \to Z$ is a map, then

$$(h \circ f)_* = h_* \circ f_*.$$

Proof. Define f_* as follows. Let γ be a loop in X based at p. Note that $f \circ \gamma$ is a loop in Y based at $f(p)$. For all $t \in I$, let $f_*([\gamma]) = [f \circ \gamma]$.

$\langle$ Show that $f_*\colon \pi(X,p) \to \pi(Y, f(p))$ is well-defined. $\rangle$

$\langle$ Show that $(h \circ f)_* = h_* \circ f_*$.$\rangle$

Now suppose that $g\colon X \to Y$ is another map such that f is homotopic to g relative to p via a homotopy $H\colon X \times I \to Y$. Let γ be a loop in X based at p. We want to show that $f \circ \gamma$ is homotopic to $g \circ \gamma$ by a based homotopy. Define

$$H_*(t,s) = H(\gamma(t), s).$$

Notice that for all $t \in I$, $H_*(t,0) = H(\gamma(t),0) = f(\gamma(t))$ and $H_*(t,1) = H(\gamma(t),1) = g(\gamma(t))$. So H_* is a homotopy from $f \circ \gamma$ to $g \circ \gamma$. Since $H(p,s) = p$ for all $s \in I$, the homotopy H_* is a based homotopy. Thus, $f_* = g_*$. □

Remark 6.13. The condition that the homotopy of f to g be relative to p is rather annoying. If $f(p) \neq g(p)$ then $\pi(Y, f(p))$ and $\pi(Y, g(p))$ are isomorphic, but unequal, groups. There is an isomorphism between them that takes the image of f_* to the image of g_*. If $f(p) = g(p)$ but the homotopy H moves p, then f_* and g_* may not be equal, but they will differ by what is called an **inner automorphism** of $\pi_1(Y, f(p))$.

Applying the following corollary to a homeomorphism and its inverse shows that fundamental groups are topological invariants.

Corollary 6.14. *Suppose that X and Y are path-connected metric spaces and that $f\colon X \to Y$ and $g\colon Y \to X$ are maps such that $g \circ f$ is homotopic to the identity map on X and $f \circ g$ is homotopic to the identity map on Y. Let $p \in X$ and $q \in Y$. Then $\pi(X,p)$ and $\pi(Y,q)$ are isomorphic.*

Proof. For simplicity, assume that there is a point $p \in X$ and a homotopy of $g \circ f$ to the identity id_X on X that is relative to p. Also assume that there is a homotopy of $g \circ f$ to the identity id_Y on Y that is relative to $f(p)$. This assumption can be dropped at the expense of making the proof more complicated.

By our hypotheses and the previous simplifying assumptions $(g \circ f)_* = (\mathrm{id}_X)_*$. Since $(\mathrm{id}_X)_*$ is the identity map on $\pi(X,p)$, we have

$$g_* \circ f_* = (g \circ f)_* = \mathrm{id}\colon \pi(X,p) \to \pi(X,p).$$

Similarly, $f_* \circ g_* = \mathrm{id}\colon \pi(Y, f(p)) \to \pi(Y, f(p))$. Thus, f_* and g_* are inverses and so they are each bijections. They are also each group homomorphisms, so they are group isomorphisms. □

Definition 6.15. Let $L \subset S^3$ be a link and $p \in S^3 \setminus L$. The **link group** (or **knot group** if L is a knot) is the group $\pi(L) = \pi(S^3 \setminus L, p)$. Changing p does not change the isomorphism type of $\pi(L)$.

Theorem 6.16. *Suppose that L and L' are equivalent links in S^3. Then $\pi(L)$ is isomorphic to $\pi(L')$.*

Proof. Let H be an isotopy of S^3 taking L to L'. Then $f\colon S^3 \setminus L \to S^3 \setminus L'$ defined by $f(x) = H(x,1)$ for all x is a homeomorphism. Since $f \circ f^{-1}$ aand $f^{-1} \circ f$ equal the identity maps on $S^3 \setminus L$ and $S^3 \setminus L'$ respectively, f_* is a group isomorphism. □

It turns out that if we squeeze a space down onto a subspace, then the fundamental groups are isomorphic. Here is how to make that precise.

Definition 6.17. Let X be a metric space and $A \subset X$. A homotopy $F\colon X \times I \to A$ is a **strong deformation retraction** if $F(x,1) \in A$ for all $x \in X$ and $F(a,s) = a$ for all $a \in A$ and $s \in I$.

Theorem 6.18. *Suppose that X is a metric space and that there is a deformation retraction from X onto $A \subset X$. Let $p \in A$. Then $\pi(X,p)$ is isomorphic to $\pi(A,p)$.*

Proof. Let F be the deformation retraction. Notice that $F(p,s) = p$ for all $s \in I$. Also observe that if γ is a path in X based at p, then $F(\gamma, 1)$ defined by $F(\gamma(t), 1)$ for all $t \in I$ is a path in A for all $t \in I$. It is also based at p. Observe that if γ and γ' are two based paths in X, then $F(\gamma \cdot \gamma', 1) = F(\gamma, 1) \cdot F(\gamma', 1)$.

Define $f\colon \pi(X,p) \to \pi(A,p)$ as follows. For $[\gamma] \in \pi(X,p)$, set $f([\gamma])$ to be the equivalence class of $F(\gamma(t), 1)$. If paths γ and γ' are homotopic in X via a based homotopy H, then considering $F(H(t,u), 1)$ is a based homotopy in A from $F(\gamma, 1)$ to $F(\gamma', 1)$. Thus, f is well-defined. Conversely, observe that if γ and γ' are based paths in A with a based homotopy in A between them, then that based homotopy is also a based homotopy in X. Thus, f is a bijection and also a group homomorphism, so it is an isomorphism. □

6.1. **The fundamental group of the circle.** We take a bit of a detour to prove that $\pi(S^1, p)$ is isomorphic to the additive group $\mathbb{Z}$.

Definition 6.19. Suppose that X and $\widetilde{X}$ are metric spaces and that $p\colon \widetilde{X} \to X$ is a map such that for all $x \in X$, there exists an $r > 0$ such that for all $\widetilde{x} \in p^{-1}(x)$, there is an open set $U(\widetilde{(x)})$ containing $\widetilde{x}$ so that:

(1) For each $\widetilde{x} \in p^{-1}(x)$, the restriction of p to $U(\widetilde{x})$ is a homeomorphism onto $B_r(x)$.
(2) $p^{-1}(B_r(x)) = \bigcup_{\widetilde{x} \in p^{-1}(x)} U(\widetilde{x})$
(3) If $\widetilde{x}_1, \widetilde{x}_2 \in p^{-1}(x)$ are distinct then $U(\widetilde{x}_1) \cap U(\widetilde{x}_2) = \varnothing$.

Then we say that $p\colon \widetilde{X} \to X$ is a covering map and that $\widetilde{X}$ is a covering space for X.

The next example is not very interesting, so generally we avoid it.

Example 6.20. Let X be any metric space and let $\widetilde{X}$ be some number of disjoint copies of X. Then the natural map from $\widetilde{X} \to X$ which is the identity on each component is a covering map.

Here is the main example we need for now

Example 6.21. Define $p\colon \mathbb{R} \to S^1$ by $p(x) = (\cos 2\pi x, \sin 2\pi x)$. Then p is a covering map.

Theorem 6.22 (Homotopy Lifting Property)**.** *Let $p\colon \widetilde{X} \to X$ be a covering map between path connected spaces. Let $q \in X$ and $\widetilde{q}_0 \in p^{-1}(q)$. Suppose that γ is a loop in X based at q. Then there exists a unique path $\widetilde{\gamma}$ in $\widetilde{X}$ such that $\widetilde{\gamma}(0) = \widetilde{q}_0$ and $\gamma = p \circ \widetilde{\gamma}$. The path $\widetilde{\gamma}$ is the **lift** of γ based at $\widetilde{q}_0$. Furthermore, if ψ is another such loop with $\widetilde{\psi}$ its lift based at $\widetilde{q}_0$, then $\widetilde{\gamma}$ and $\widetilde{\psi}$ are homotopic relative to their endpoints if and only if γ and ψ are based homotopic.*

Proof Sketch. For each $t \in I$, there exists $r_t > 0$ such that $B_{r_t}(\gamma(t)) \subset X$ is one of the open balls appearing in the definition of covering map. By a certain property of the interval, called "compactness", there exist

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = 1$$

such that $\operatorname{im} \gamma \subset \bigcup_{i=0}^{n} B_{r_{t_i}}(\gamma(t_i))$. It is also possible (by a result called the Lebesgue Covering Lemma) to subdivide I into small open intervals such that no more than two of the open intervals overlap at a time and each of the intervals is contained in a unique $B_{r_{t_i}}(\gamma(t_i))$. Number those intervals consecutively as $I_0, I_1, \ldots, I_m$ with $0 \in I_0$ and $1 \in I_m$.

By the definition of covering map, there exists an open set $U(\widetilde{q}_0) \subset \widetilde{X}$ that projects homeomorphically onto $B_{t_0}(\gamma(0))$ and contains $\widetilde{q}_0$. Let $p_0\colon U(\widetilde{q}_0) \to B_{t_0}(\gamma(0))$ be the homeomorphism. Let $\widetilde{\gamma}_0 = p_0^{-1} \circ \gamma|_{I_0}$. This produces the initial segment of our path. We then inductively define $\widetilde{\gamma}$ on each of the subintervals I_i by using the fact that the covering map is a homeomorphism.

The proof that if γ and ψ are based homotopic then their lifts are homotopic relative to their endpoints is a 2-dimensional version of the previous argument. A homotopy between paths is a map from $I \times I$ into X. We subdivide the square $I \times I$ into small little subsquares and inductively define the lift of the homotopy across the subsquares. All that is needed is that $I \times I$ is also a compact metric space.

Finally, suppose that $\widetilde{\gamma}$ and $\widetilde{\psi}$ are lifts of γ and ψ that are homotopic relative to their endpoints. Let $\widetilde{H}\colon I \times I \to \widetilde{X}$ be the homotopy. Then $H\colon I \times I \to X$ defined by $H = p \circ \widetilde{H}$ is a based homotopy of γ to ψ. □

Returning to the circle S^1, for each $n \in \mathbb{Z}$ consider the based loop

$$\gamma_n(t) = (\cos 2\pi n t, \sin 2\pi n t)$$

for all $t \in I$. This is the loop that winds n times counter-clockwise around S^1. (If $n < 0$, this means it winds $|n|$ times clockwise around S^1.)

Corollary 6.23. *Let $q = (1, 0) \in S^1$. The group $\pi(S^1, q)$ is isomorphic to the additive group $\mathbb{Z}$. Each class of based loops has a unique representative γ_n for some $n \in \mathbb{Z}$ and there is an isomorphism to $\mathbb{Z}$ taking $[\gamma_n]$ to n.*

Proof. Let $p\colon \mathbb{R} \to S^1$ be the covering map and observe that $\mathbb{Z} = p^{-1}(q_0)$. Set $\widetilde{q}_0 = 0$. The lift $\widetilde{\gamma}_n$ of γ based at $\widetilde{q}_0$ has its other endpoint at n. Consequently, if $n \neq m$, the based loops γ_n and γ_m are not based homotopic.

Suppose that γ is any based loop in S^1. Let $\widetilde{\gamma}$ be its left. Recall that $\widetilde{\gamma}(0) = 0$ and that $\widetilde{\gamma}(1) = n$ for some $n \in \mathbb{Z}$. Define $\widetilde{H}\colon I \times I \to \mathbb{R}$ by $\widetilde{H}(t, s) = (1 - s)\widetilde{\gamma}(t) + s\widetilde{\gamma}_n(t)$. Observe that $\widetilde{H}$ is a homotopy from $\widetilde{\gamma}$ to $\widetilde{\gamma}_n$ relative to its endpoints. Thus, γ and γ_n are based homotopic. Consequently, each equivalence class in $\pi(S^1, q)$ is represented by a unique γ_n. Define $f\colon \pi(S^1, q) \to \mathbb{Z}$ by $f([\gamma_n]) = n$. It is easily verified that f is an isomorphism. □

Corollary 6.24. *Let $U \subset S^3$ be the unknot. Then $\pi(U)$ is isomorphic to $\mathbb{Z}$.*

Proof Sketch. It turns out that $S^3 \setminus U$ is homeomorphic to $D \times S^1$ where $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$. Let $H((x, y, \theta), t) = ((1 - t)(x, y), \theta)$. This is a deformation retraction of the open solid torus onto the core circle and so $\pi(U)$ is isomorphic to $\pi(S^1, (1, 0))$, which is isomorphic to $\mathbb{Z}$. □

In 1957, Papakyriakopolous proved that if a knot $K \subset S^3$ has $\pi(K)$ isomorphic to $\mathbb{Z}$, then K is the unknot. More generally, if two prime knots have isomorphic knot groups then they are either equivalent or one is equivalent to the other's mirror image. The square knot and the granny knot have isomorphic groups, so the condition on primeness can't be dropped.

Example 6.25. Prove that $\pi(S^1 \times S^1, q)$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}$. (Hint: find a covering map $\mathbb{R}^2 \to S^1 \times S^1$.)

We conclude this section with a look at how the knot group we've defined here is related to the combinatorial knot group we looked at earlier in the semester.

Consider a diagram D of an oriented knot K. Consider the knot K as coinciding with the diagram D as lying in a $S^2 \subset S^3$ except that at each crossing it dips a bit above the sphere and a bit below the sphere. Pick a point q in $S^3 \setminus D$. At each arc e of D take an oriented loop μ_e encircling e, so that the orientation on K satisfies the right-hand rule. Choose a path α_e from q to μ_e. Let $\gamma_e = \alpha_e \cdot \mu_e \cdot \overline{\alpha}_e$. We give an informal argument that $\{[\gamma_e]\}$ is a generating set for $\pi(K)$.

To see this, let γ be any loop in $S^3 \setminus K$ based at q. By a homotopy, we can assume it is transverse to each region of the diagram D (i.e. not tangent, passing straight through). Imagine the images of the loops γ_e as being made of thin metal rods while γ is made of 2-sided sticky tape. Homotope γ so that any time it passes through a region of the diagram it sticks to a bit of K forming the boundary of the diagram. A further homotopy makes the sticky tape wrap around the loops μ_e. We then crush the stick tape to the rods α_e. As we traverse γ, we then are traversing the loops γ_e and $\overline{\gamma}_e$.

We now examine the relations arising from crossings.

Exercise 6.26. Consider a crossing of D with a the oriented overstrand, b the incoming understrand, and c the outgoing understrand. Suppose that the crossing is right-handed. Draw pictures to show that the loop $\gamma_b \cdot \gamma_a \cdot \overline{\gamma}_b$ is based homotopic to γ_c.

Our informal investigations lead us to suspect that the knot group $\pi(K)$ is isomorphic to the combinatorial knot group defined previously. Indeed this is the case, but giving a rigorous proof is beyond the scope of this course and is best done using the VanKampen Theorem from Algebraic Topology.

7. Path Metrics

Suppose that $X \subset \mathbb{R}^n$ has the property that any two points in X can be joined by a piecewise smooth path.

Definition 7.1. The **euclidean length** of a piecewise-smooth path γ is

$$\ell(\gamma) = \int_0^1 ||\gamma'(t)||dt.$$

This definition is inspired by approximating γ with line segments and measuring their length using the Pythagorean theorem. We can use path lengths to define a metric. Ideally, we would just define the distance between two points in X to be the length of the shortest path between them. The next example shows that this is not always possible.

Example 7.2. Let $X = \mathbb{R}^2 \setminus \{(0,0)\}$ and let $a = (-1,0)$ and $b = (0,1)$. Then there is no shortest path from a to b that lies in X.

In the definition below, we treat $\pm\infty$ as symbols that can be compared with real numbers in the obvious way. We do not allow ourselves to do arithmetic with them as logical contradictions can result.

Definition 7.3. Suppose that $C \subset \mathbb{R}$. A **lower bound** for C is $\lambda \in [-\infty, \infty]$ such that $\lambda \leq c$ for all $c \in C$. Similarly an **upper bound** for C is $\mu \in [-\infty, \infty]$ such that $c \leq \mu$ for all $c \in C$. The **infimum** $\inf C$ of C is the greatest lower bound for C, that is $\lambda \leq \inf C$ for every lower bound λ for C. The **supremum** $\sup C$ of C is the least upper bound for C, that is $\mu \geq \sup C$ for every upper bound μ for C.

The following is a crucial theorem, due to Dedekind. It is logically equivalent to the statment that every monotonic sequence in an interval converges.

Theorem 7.4. *Suppose* $C \subset \mathbb{R}$*. Then* $\inf C$ *and* $\sup C$ *exist. Furthermore, if* $C \neq \varnothing$ *and there exists an upper bound* $M \in \mathbb{R}$ *for* C*, then* $\sup C \in \mathbb{R}$*. If* $C \neq \varnothing$ *and there exists a lower bound* $m \in \mathbb{R}$ *for* C*, then* $\inf C \in \mathbb{R}$*.*

Definition 7.5. Let $X \subset \mathbb{R}^n$ have the property that any two points can be joined by a piecewise smooth path lying entirely in X. For $x, y \in X$, define $d_{path}(x, y) = \inf_{\gamma} \ell(\gamma)$, where the infimum is taken over all piecewise smooth paths joining x to y.

Theorem 7.6. d_{path} *is a metric, called the* ***path metric*** *on* X*.*

Proof. Let $x, y \in X$. Since the length of any piecewise smooth path between x and y is non-negative, $d_{path}(x, y) \geq 0$. If $x = y$, then the constant path is a path of length 0 from x to y, so $d_{path}(x, y) = 0$ if $x = y$. We tackle the proof that if $x \neq y$, then $d_{path}(x, y) > 0$ last.

For $x, y \in X$, let $P(x, y)$ be the set of all piecewise smooth paths from x to y. If $\gamma \in P(x, y)$ then the path $\overline{\gamma}$ defined by $\overline{\gamma}(t) = \gamma(1-t)$ for all $t \in I$ is a path from y to x. Observe that $\ell(\overline{\gamma}) = \ell(\gamma)$. Thus the set of lengths of paths from x to y is the equal to the set of lengths of paths from y to x and so $d_{path}(x, y) = d_{path}(y, x)$.

Suppose that $x, y, z \in X$. Consider a path $\gamma \in P(x, y)$ and a path $\psi \in P(y, z)$. Define

$$\gamma \cdot \psi(t) = \begin{cases} \gamma(2t) & t \in [0, 1/2] \\ \psi(2t-1) & t \in [1/2, 1] \end{cases}.$$

Observe that $\gamma \cdot \psi \in P(x, z)$ and that $\ell(\gamma \cdot \psi) = \ell(\gamma) + \ell(\psi)$. Thus, $d(x, z) \leq \ell(\gamma) + \ell(\psi)$. This is almost enough to prove the triangle inequality. We just have to deal with the infimum.

By the definition of infimum, for any $\epsilon > 0$, there is a path $\gamma \in P(x, y)$ such that $\ell(\gamma) \leq d(x, y) + \epsilon/2$. Likewise, there exists a path $\psi \in P(y, z)$ such that $\ell(\psi) \leq d(y, z) + \epsilon/2$. Thus, $d(x, z) \leq d(x, y) + d(y, z) + \epsilon$, no matter what $\epsilon > 0$ is. But since $d(x, z)$ and $d(x, y)$ and $d(y, z)$ are all fixed numbers, the only way this is possible is if $d(x, z) \leq d(x, y) + d(y, z)$.

Finally, we show that if $x \neq y$, then $d(x, y) > 0$. Let $\gamma \in P(x, y)$. The key idea is that in $\mathbb{R}^n$ the shortest distance between any two points is a straight line. We'll prove that. Notice that the function $F \colon \mathbb{R}^n \to \mathbb{R}^n$ defined by $f(p) = p - x$ does not change distances. Another calculation shows that a rotation about the origin of $\mathbb{R}^n$ does not change distances. Thus, we may assume that $x = (0, \ldots, 0)$ and that $y = (0, \ldots, 0, y_n)$ with $y_n > 0$. Notice that the length of the line segment betweeen x and y is y_n. For simplicity, we'll assume that the path γ is smooth, rather than just piecewise smooth.

For each $t \in I$, $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$. By standard facts from calculus:

$$\begin{array}{rcl} \ell(\gamma) & = & \int_0^1 \sqrt{\gamma_1'(t)^2 + \cdot + \gamma_n^2(t)}\, dt. \\ & \geq & \int_0^1 |\gamma_n'(t)|\, dt \\ & \geq & |\int_0^1 \gamma_n'(t)\, dt \\ & = & \gamma_n(1) - \gamma_n(0) \\ & = & y_n \end{array}.$$

Thus, not only is $d(x, y) > 0$, but it is at least the length of the line segment joining x to y. ☐

Before moving on, we make an aside to define another knot invariant. It is the subject of active research.

Definition 7.7. Let $K \subset \mathbb{R}^3$ be a knot. For $p \neq q \in K$, define $\delta(p,q) = \frac{d_{path}(p,q)}{d(p,q)}$, where $d_{path}(p,q)$ is the distance between p and q along K and $d(p,q)$ is the euclidean distance. The **distortion** of the knot K is $\delta(K) = \sup_{p \neq q} \delta(p,q)$ and the **piecewise smooth distortion** of the knot type of K is $\inf_K \delta(K)$ where the infimum is taken over all piecewise smooth representatives K of the knot type.

In 1978 Mikhail Gromov defined distortion. He showed that the distortion of a loop is at least $\pi/2$ is equal $\pi/2$ if and only if the loop is a round circle. Gromov asked if there were knot types where the distortion of the knot type was arbitrarily large. Denne and Sullivan showed that if K is nontrivial, then $\delta(K) \geq 5\pi/3$. In his undergraduate thesis ($\sim$ 2011), John Pardon showed that the torus knots can have distortion that is arbitrarily large. A few years ago, collaborators and I adopted Pardon's methods to use bridge number and another invariant "bridge distance" to give lower bounds for distortion. One of the great things about Pardon's paper is that he can work with a more general class of knots then just piecewise smooth ones - the only thing that is required is that they be "rectifiable" – that is, that length be a well-defined quantity.

Definition 7.8. The path metric on S^n for $n \geq 1$ is called the **spherical metric**.

We can obtain other metrics by varying the construction. Here is an extremely important example.

Definition 7.9. Let $\mathbb{H}^n = \mathbb{R}^n_+ = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_n > 0\}$. Define the **hyperbolic length** $\ell_{hyp}(\gamma)$ of a piecewise smooth path γ to be

$$\ell_{hyp}(\gamma) = \int_0^1 \frac{||\gamma'(t)||}{x_n(t)}\, dt$$

where $x_n(t)$ is the last coordinate of $\gamma(t)$. The **hyperbolic metric** $d_{hyp}(a,b)$ between two points $a, b \in \mathbb{H}^n$ is defined to be $\inf_\gamma \ell_{hyp}(\gamma)$ where the infimum is over all piecewise smooth points joining p to q.

Exercise 7.10. Show that d_{hyp} is a pseudometric on $\mathbb{H}^n$.

In the remainder, we show that d_{hyp} is a metric and that there is a geodesic (i.e. shortest path) between any two points. We also give an explicit description of these geodesics.

Lemma 7.11. *The following functions $T\colon \mathbb{H}^n \to \mathbb{H}^n$ are isometries:*

- *Horizontal translation: $T(x_1, \dots, x_n) = (x_1 + a_1, x_2 + a_2, \dots, x_{n-1} + a_{n-1}, x_n)$ for some $(a_1, \dots, a_{n-1}) \in \mathbb{R}^{n-1}$.*
- *Dilation: $T(x_1, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$ for some $\lambda > 0$.*
- *Horizontal rotation: rotation around the line $\{(0, \dots, t) \in \mathbb{H}^n : t > 0\}$.*
- *Sphere Inversion: $T(x) = x/||x||^2$*

Lemma 7.12. *Let $x = (c_1, \dots, c_{n-1}, a)$ and $y = (c_1, \dots, c_{n-1}, b)$ be two points in $\mathbb{H}^n$ lying on the same vertical line, with $a < b$. Let*

$$\gamma(t) = (1-t)x + ty = (c_1, \dots, c_{n-1}, (1-t)a + tb)$$

for $t \in [0,1]$. Then γ is a geodesic from x to y and any other geodesic between them has the same image.

Exercise 7.13. Show that for the γ in the statement of the lemma $\ell_{hyp}(\gamma) = \ln(b/a)$.

Proof of Lemma 7.12. Let $\psi(t) = (x_1(t), x_2(t), \dots, x_n(t))$ be a path in $\mathbb{H}^n$ from x to y. Notice that

$$\frac{||\psi'(t)||}{x_n(t)} = (\frac{x_1'(t)^2 + \cdots + x_{n-1}'(t)^2 + x_n'(t)^2}{x_n(t)^2})^{1/2} \geq (\frac{x_n'(t)^2}{x_n(t)^2})^{1/2}.$$

Since both sides are non-negative,

$$\ell(\psi) = \int_0^1 \frac{||\psi'(t)||}{x_n(t)}\, dt \geq \int_0^1 \frac{|x_n'(t)|}{x_n(t)}\, dt.$$

If equality holds,

$$\int_0^1 \frac{||\psi'(t)|| - |x_n'(t)|}{x_n(t)}\, dt = 0$$

Observe that the integrand is nonnegative, by our previous work. Since ψ is piecewise smooth, the integrand is piecewise continuous. Thus, except at finitely many points, $||\psi'(t)|| = |x_n'(t)|$. But this means that, except at finitely many points, $(x_1'(t), \dots, x_{n-1}'(t)) = (0, \dots, 0)$. Thus, except at finitely many points, $(x_1(t), \dots, x_{n-1}(t))$ is constant. But ψ is continuous, so this would imply that it is, in fact, constant. That is, if equality holds, then the image of ψ lies on the vertical line passing through x and y.

Observe that

$$\ell_{hyp}(\psi) \geq \int_0^1 \frac{|x_n'(t)|}{x_n(t)}\, dt \geq \int_0^1 \frac{x_n'(t)}{x_n(t)}\, dt = \ell_{hyp}(\gamma).$$

Thus, γ is a geodesic from x to y. Furthermore, equality holds if and only if $|x_n'(t)| = x_n'(t)$ for all values of t where ψ is smooth. Since ψ is piecewise smooth, $x_n'(t)$ is piecewise continuous, so x_n' is non-negative except, possibly, at finitely many points. Since ψ is continuous, x_n is non-decreasing. Consequently, if $\ell_{hyp}(\psi) = \ell_{hyp}(\gamma)$, then ψ has the same image as γ and does not backtrack. □

Theorem 7.14. *Geodesics in $\mathbb{H}^n$ are portions of vertical lines and portions of circles perpendicular to the plane $\mathbb{R}^{n-1} = \{(x_1, \dots, x_{n-1}, 0) \in \mathbb{R}^n\}$. There is a unique hyperbolic geodesic between any two distinct points of $\mathbb{H}^n$.*

Proof. Notice that if γ is a geodesic in $\mathbb{H}^n$ and if f is an isometry of $\mathbb{H}^n$, then $f \circ \gamma$ is also a geodesic in $\mathbb{H}^n$. Let $x, y \in \mathbb{H}^n$. We have already seen that if x and y lie on the same vertical line then the line segment between them is the unique image of a geodesic between them. Suppose now that x and y do not lie on the same vertical line. Let C be the half circle in $\mathbb{H}^n$, with endpoints on $\mathbb{R}^{n-1}$, passing through them which and perpendicular to $\mathbb{R}^{n-1}$. (Why is there such?) Recall $S^{n-1} = \{z \in \mathbb{R}^n : ||z|| = 1\}$

Notice that horizontal translations, horizontal rotations, and dilations do not change whether a path has an image that is a vertical line segment or a portion of a circle perpendicular to $\mathbb{R}^{n-1}$. Let p_0 be the endpoint of C closest to x. Use horizontal translations to make p_0 lie at the origin. Use a horizontal rotation to ensure that C lies in the x_1, x_n-plane. Henceforth, we may assume that $n = 2$. Apply the sphere inversion to C to obtain C'. This takes the endpoint at the origin to infinity and leaves the other endpoint of C fixed. We claim that C' is a vertical line. Let x' and y' be the images of x and y on C'.

It is easier to show this proceeding the other way. Let L be a vertical line in $\mathbb{H}^2$ with one endpoint P at $(p, 0)$ for some $0 < p$. The sphere inversion takes P to the point $(1/p, 0)$. Let $m = (1/2p, 0)$. Parameterize L as (p, t) for $t > 0$. The sphere inversion converts it to the curve $\alpha(t) = \frac{1}{p^2+t^2}(p, t)$. A straightforward calculation shows that the Euclidean distance from $\alpha(t)$ to m is equal to $1/2p$, which is a constant. As $t \to \infty$, $\alpha(t)$ approaches $(0, 0)$. Thus, $\alpha(t)$ is a parameterization of the half circle with center m. Since sphere inversion is its own inverse, sphere inversion takes the half circle with center $(1/2p, 0)$ to L. Choosing p so that m is the center of C, we see that C' is a vertical line. Thus, the geodesic from x' to y' lies on C' and so the geodesic from x to y lies on C. □

Exercise 7.15. Show that given $a, a', b, b', c, c' \in \mathbb{H}^2$, there is a hyperbolic isometry taking a to a', b to b', and c to c'. (Hint: Use the techniques in the previous proof)

8. Gluing

Gluing spaces together is a fundamental part of topology, although giving a formal definition is a little tricky. It either requires abandoning metric spaces altogether or some somewhat finicky work involving metric spaces. Since metric spaces are important in their own right, we choose the metric space approach.

Let Z be a set and $d_Z \colon Z \times Z \to [0, \infty]$ a function such that:

(1) $d_Z(a, a) = 0$ for all $a \in Z$
(2) $d_Z(a, b) = d_Z(b, a)$ for all $a, b \in Z$

Suppose that $\sim$ is an equivalence relation on Z such that each equivalence class has only finitely many elements. For $a, b \in Z$, a **chain** from a to b is a sequence

$$a = z_0, z_0', z_1, z_1', \ldots, z_n, z_n' = b$$

such that all $z_i, z_i' \in Z$ and for each i, $z_i \sim z_i'$. The **length** of the chain C is

$$\ell(C) = \sum i = 1^{n-1} d_Z(z_i', z_{i+1})$$

We imagine the chain as a series of walks and teleportations. Going from z_i to z_i' is a teleportation and from z_i' to z_{i+1} is a walk. The length of the chain is the total distance of the walks. For $a, b \in Z$, let $d(a, b) = \inf_C \ell(C)$ where the infimum is taken over all chains from a to b.

Theorem 8.1. *The function d has the following properties:*

(1) For all $a \in Z$, $d(a, a) = 0$
(2) For all $a, b \in Z$, $d(a, b) = d(b, a)$
(3) For all $a, b, c \in Z$, $d(a, c) \leq d(a, b) + d(b, c)$

Proof. We prove only the triangle inequality. By the definition of infimum, for all $\epsilon > 0$, there exists a chain $C(a, b)$ from a to b such that $\ell(C(a, b)) \leq d(a, b) + \epsilon/2$. Similarly, there exists a chain $C(b, c)$ from b to c such that $\ell(C(b, c)) \leq d(b, c) + \epsilon/2$. The chain $C(a, b)$ is a sequence

$$a = z_0, z_0', z_1, z_1', \ldots, z_n, z_n' = b$$

such that $z_i \sim z_i'$ for all i. The chain $C(b, c)$ is a sequence

$$b = w_0, w_0', w_1, w_1', \ldots, w_m, w_m' = c$$

Since $z_n \sim z_n' = w_0 \sim w_0'$, the sequence

$$a = z_0, z_0', z_1, z_1', \ldots, z_n, w_0', w_1, w_1', \ldots, w_m' = c$$

is a chain from a to c. Its length is $\ell(C(a, b)) + \ell(C(b, c))$. Thus,

$$d(a, c) \leq \ell(C(a, b)) + \ell(C(b, c)) = d(a, b) + d(b, c) + \epsilon.$$

Since this is true for all $\epsilon > 0$, $d(a, c) \leq d(a, b) + d(b, c)$. □

Corollary 8.2. *Suppose that Z is a set and d_Z is as above. Suppose also that $\sim$ is an equivalence relation on Z such that between any two points there is a chain of finite length. Define d as above and let $d([a],[b]) = d(a,b)$ for all $[a],[b] \in Z/\sim$. Then d is a pseudometric on $Z/\sim$.*

Here is a special case of this construction. Suppose that (X, d_X) and (Y, d_Y) are metric spaces with subspaces $A \subset X$ and $B \subset Y$. Assume that either $X \cap Y = \varnothing$ or $(X, d_X) = (Y, d_Y)$. In the latter case, suppose that $A \cap B = \varnothing$. Suppose also that there is a homeomorphism $h\colon A \to B$. Let $Z = X \cup Y$. If $X \neq Y$ and $x \in X$ and $y \in Y$, let $d_Z(x,y) = d_Z(y,x) = \infty$. Otherwise, let $d_Z(x,y)$ equal either $d_X(x,y)$ or $d_Y(x,y)$ depending on whether or not $x, y \in X$ or $x, y \in Y$. Define $\sim$ on Z by declaring $a \sim f(a)$ for all $a \in A$ and then extending $\sim$ to be an equivalence relation on Z. Let $Z/\sim$ be the set of equivalence classes. Notice that between any two points there is a chain of finite length and so the functon d defined above is a pseudometric on $Z/\sim$. If $X \neq Y$, we let $X \cup_h Y = Z/\sim$.

Example 8.3. Let $Z = I \times I$. Let $A = \{(1/n, 0) : n \in \mathbb{N}\}$ and $B = \{(1/n, 1) : n \in \mathbb{N}\}$. Define $h\colon A \to B$ by declaring $h((1/n,0)) = (1/n, 1)$ for all $n \in \mathbb{N}$. Let C_n be the chain

$$(0,0), (0,0), (1/n, 0), (1/n, 1), (0,1), (0,1).$$

Then C_n is a chain of length $2/n$ from $a = (0,0)$ to $b = (0,1)$. Thus, $d(a,b) = 0$. Since $a \not\sim b$, d is not a pseudometric on $Z/\sim$.

Theorem 8.4. *Let $Z = [0, 2\pi]$. Let $A = \{0\}$ and $B = \{2\pi\}$ and $h(0) = 2\pi$. Then d is a metric on $Z/\sim$ and $Z/\sim$ is isometric to S^1 (with the path metric).*

Proof. For convenience, let d_Z be the Euclidean metric on Z. Let $[a],[b] \in Z/\sim$ be distinct. Without loss of generality, we may assume that $[a] = \{a\}$ and either $[b] = \{b\}$ or $[b] = \{0, 2\pi\}$. Let

$$C : a, a, z_1, z_1', \ldots, z_n, z_n' = b$$

be a chain from a to b. As we are seeking to minimize the length of chains, we may as well assume that there do not exist distinct i, j with $z_i \sim z_j$. Suppose, first, that $[b] = \{0, 2\pi\}$. Then, for all $i < n$, $[z_i] = \{z_i\}$. Thus, the chain is of the form:

$$a, a, z_1, z_1, \ldots, z_{n-1}, z_{n-1}, z_n, z_n' = b$$

Its length is

$$d_Z(a, z_1) + d_Z(z_1, z_2) + d(z_2, z_3) + \cdots + d_Z(z_{n-1}, z_n) \geq d_Z(a, z_n).$$

Thus, $d([a], \{0, 2\pi\}) = \min\{d_Z(a, 0), d_Z(a, 2\pi)\}$.

A similar argument shows that if $[b] = \{b\}$, then $d([a],[b]) = \min\{d_Z(a,0) + d_Z(2\pi, b), d_Z(a, 2\pi) + d_Z(0,b)\}$. In particular, d is a metric on $Z/\sim$.

Define $f\colon Z/\sim \to S^1$ by $f([a]) = (\cos a, \sin a)$. Since cos and sin are 2π-periodic, f is well-defined. Standard trigonmetric facts show it is a bijection, as well. To see that f is an isometry, recall that if two points on the circle are separated by an angle of $\theta \leq \pi$, then they are distance θ apart on the circle. □

Exercise 8.5. Let Q be a parallelogram in $\mathbb{R}^3$. Let T be the result of gluing opposite sides together, without a twist. Then T is homeomorphic to $S^1 \times S^1$.

Example 8.6. Let $K \subset S^3$ be a tame knot. For small enough $r > 0$, the subspace $N_r(K) = \{x \in S^3 : d(x, K) \leq r\}$ is homeomorphic to $V = D^2 \times S^1$. Let $\mathring{V} = \{x \in S^3 : d(x,k) < r\}$. The **exterior** of K is $X(K) = S^3 \setminus \mathring{V}$. The boundary $\partial X(K)$ is the torus $X(K) \cap N_r(K)$. Choose a homeomorphism $h\colon \partial V \to \partial X(K)$. We say that the space $M(K,h) = V \cup_h X(K)$ is obtained by **Dehn surgery** on K. There is one choice of h such that $V \cup_h X(K) = S^3$, but in almost all other cases $M(K, h)$ is not homeomorphic to S^3, but is instead some other 3-dimensional space. Lickorish and Wallace proved in the 1960s that every 3-dimensional space of a certain fairly generic type can be obtained by Dehn surgeries on the components of a link.

Example 8.7. For some $g \geq 0$, let Γ_g be a graph in $\mathbb{R}^2 \subset \mathbb{R}^3$ consisting of a single vertex and g loops. Let $H_g = \{x \in \mathbb{R}^3 : d(x, \Gamma_g) \leq r\}$ for some small value of $r > 0$. H_g is called a **genus** g **handlebody**. Its boundary is the set of points where equality holds. Let H_g' be another copy of H_g. Choose a homeomorphism $h\colon \partial H_g \to \partial H_g$. Then $M(g, h) = H_g \cup_h H_g'$ is a 3-dimensional space. Every 3-dimensional space of a certain fairly generic type can be obtained this way for some choice of g and h.