Topics in Modern Geometry, Part II: Hyperbolicity

Daniel Berlyne

October 26, 2022

Contents

5 Classical 2–dimensional geometry

5.1 History

We begin with some history. Around 300 BCE, the Greek mathematician Euclid wrote his famed treatise *Elements*, in which he laid the foundations for geometry as it was understood at the time. In this treatise, he proposed five postulates which he considered to be reasonable assumptions that must be made in order to prove anything further. Indeed, all of his theorems were deduced from these postulates alone. The postulates are as follows.

1. A straight line segment may be drawn between any two points.

2. Any straight line segment may be extended indefinitely to form a straight line.

3. Given any straight line segment, a circle may be drawn with the line segment as its radius and one of the endpoints as its centre.

4. All right angles are congruent (that is, there is an isometry taking one to the other).

5. Given a line L and a point P not on L, there is a unique line through P that does not intersect L.[1]

The geometry that is based on these postulates is known as *Euclidean geometry*; this is the geometry that most of us know from our childhood education, and indeed for roughly 1500 years the vast majority of mathematicians worked only with Euclidean geometry. The Euclidean plane is typically denoted $\mathbb{E}^2$.

One may notice that the fifth postulate (also known as the *parallel postulate*), seems more complicated than the others and less of an obvious "fact". Indeed, in about 100 AD, Menelaus showed in his treatise *Sphaerica* that one can construct a different geometry, known as *spherical geometry*, by replacing the parallel postulate with the following.

5'. Given a line L and a point P not on L, all lines through P must intersect L.

As the name would suggest, spherical geometry is performed on the surface of a sphere, denoted $\mathbb{S}^2$. Note that in non-Euclidean geometries, a "line" is taken to mean a geodesic; that is, a path of shortest length connecting two points. In spherical geometry, geodesics are arcs of circles obtained by taking the intersection of the sphere in $\mathbb{R}^3$ (centred at the origin) with a plane passing through the origin.

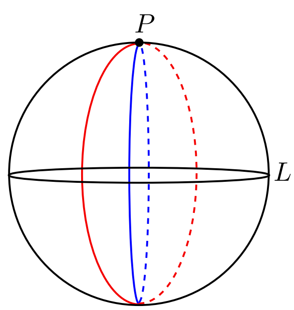

Figure 1: Illustration of 5'.

Around 1820, Nikolai Lobachevsky showed that a third geometry, known as *hyperbolic geometry*, can be constructed by replacing the parallel postulate with the following.

5". Given a line L and a point P not on L, there are infinitely many lines through P that do not intersect L.

In 1907, Henri Poincaré finally put this to rest by conclusively proving that Euclidean, spherical, and hyperbolic geometry are the only possible 2–dimensional geometries[2]. (In three dimensions and higher, the picture is more complicated.) We study classical 2–dimensional hyperbolic geometry in more detail in the next section.

[1] Euclid originally stated the fifth postulate in a slightly different, but equivalent, form. This version is due to John Playfair in 1795.

[2] This is a simplified statement of a result in complex analysis known as the Uniformisation Theorem.

5.2 The hyperbolic plane $\mathbb{H}^2$

There are many ways of modelling 2–dimensional hyperbolic geometry (known as the *hyperbolic plane* and denoted $\mathbb{H}^2$). We focus on the *Poincaré disc model.*

Definition 5.1 (Poincaré disc model)**.** Let $D = \{z \in \mathbb{C} \mid |z| < 1\}$. A *hyperbolic line* is a diameter of D or an arc of a Euclidean circle contained in D that is orthogonal to the boundary $\partial D = \{z \in \mathbb{C} \mid |z| = 1\}$.

Given $z, w \in D$, let γ be the unique hyperbolic line passing through z and w. Let z' and w' be the intersection points of γ with ∂D that are closest to z and w, respectively. Define

$$\mathsf{d}_D(z, w) = \ln\left(\frac{|z' - w| \cdot |w' - z|}{|z' - z| \cdot |w' - w|}\right).$$

We call D equipped with the metric d_D the *Poincaré disc model* of the hyperbolic plane; we write $\mathbb{H}^2 = (D, \mathsf{d}_D)$.

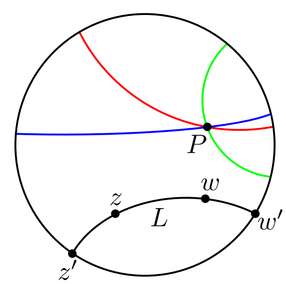

Figure 2: Illustration depicting 5" and the points z' and w'.

Exercise 5.2. Verify that d_D defines a metric on D.

Exercise 5.3. Show that $\mathsf{d}_D(0, z) = \ln\left(\frac{1+|z|}{1-|z|}\right)$.

Note that as z or w approaches the boundary ∂D, the value of $|z - z'|$ or $|w - w'|$ approaches 0, thus $\mathsf{d}_D(z, w)$ tends to infinity. Hence, in order to reach ∂D, one must travel an infinite distance. For this reason, ∂D is often referred to as the *ideal boundary*, since in reality it is not bounding the space in the traditional sense. This is also why lines in $\mathbb{H}^2$ arise as arcs of circles orthogonal to ∂D – since distances blow up to infinity as one approaches the boundary, it is more efficient to move a little towards the centre of D as one travels between two points in $\mathbb{H}^2$.

To study the geometry of $\mathbb{H}^2$ in greater detail, we need to identify what isometries look like. We first define a natural collection of maps of the complex plane, called *Möbius transformations.*

Definition 5.4 (Möbius transformation)**.** A *Möbius transformation* is a map $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ of the form $f(z) = \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbb{C}$ are constants such that $ad - bc \neq 0$. Here we define $f(\infty) = \frac{a}{c}$ and $f(-\frac{d}{c}) = \infty$.

Möbius transformations have many useful properties, as described below.

Proposition 5.5 (Properties of Möbius transformations)**.** *Let $f(z) = \frac{az+b}{cz+d}$ be a Möbius transformation.*

1. *It is always possible to rewrite f so that $ad - bc = 1$ (this is called* normalising *a Möbius transformation).*
2. *There is a unique Möbius transformation taking any three distinct points of $\mathbb{C} \cup \{\infty\}$ to any other three distinct points.*
3. *Euclidean circles in $\mathbb{C} \cup \{\infty\}$ are sent by f to Euclidean circles, where we consider a straight line together with the point ∞ to be a "circle".*
4. *$f(D) = D$ if and only if $d = \bar{a}$ and $c = \bar{b}$ (recall $\bar{z}$ denotes the complex conjugate of z).*
5. *For all $z, w \in D$ there is some f with $d = \bar{a}$ and $c = \bar{b}$ sending z to w.*
6. *If $d = \bar{a}$ and $c = \bar{b}$ then f restricts to an isometry of $\mathbb{H}^2$.*

Proof. We shall prove parts 1 and 2 and leave the remainder as an exercise.

1. Since $ad - bc \neq 0$, we may divide both the numerator and denominator by $\sqrt{ad - bc}$ to give $f(z) = \frac{a'z+b'}{c'z+d'}$ where $a' = \frac{a}{\sqrt{ad-bc}}$, $b' = \frac{b}{\sqrt{ad-bc}}$, $c' = \frac{c}{\sqrt{ad-bc}}$, $d' = \frac{d}{\sqrt{ad-bc}}$. Thus, $a'd' - b'c' = \frac{ad-bc}{ad-bc} = 1$.
2. First note that Möbius transformations are invertible, and moreover the inverse is another Möbius transformation: if $f(z) = \frac{az+b}{cz+d}$, then $f^{-1}(z) = \frac{b-dz}{cz-a}$. Furthermore, it is easy to show the composition of two Möbius transformations is a Möbius transformation. It is therefore sufficient to show that for any $p, q, r \in \mathbb{C} \cup \{\infty\}$, there is a unique $f(z) = \frac{az+b}{cz+d}$ such that $f(0) = p$, $f(1) = q$, $f(\infty) = r$. This is because for any $p', q', r' \in \mathbb{C} \cup \{\infty\}$ we can then find a unique $g(z)$ with $g(0) = p'$, $g(1) = q'$, $g(\infty) = r'$, hence $g^{-1} \circ f$ is the unique Möbius transformation sending p to p', q to q', and r to r'.

   Note that $f(0) = p$ implies $b = pd$ and $f(\infty) = r$ implies $a = rc$. Thus, $f(z)$ is uniquely determined by p, d, r, c. Furthermore, $f(1) = q$ implies $qc + qd = a + b = rc + pd$, so $c = \frac{d(p-q)}{q-r}$. So $f(z)$ is uniquely determined by p, q, r, d. Finally, we can assume $ad - bc = 1$, so $d = \frac{1+bc}{a} = \frac{1+pd}{r}$, thus $d = \frac{1}{r-p}$. Hence, $f(z)$ is uniquely determined by p, q, r, as required. □

Exercise 5.6. (Hard) Prove the rest of Proposition 5.5. Conclude that Möbius transformations of the form $f(z) = \frac{az+b}{\bar{b}z+\bar{a}}$ are isometries of $\mathbb{H}^2$.

Properties (5) and (6) show that $\mathbb{H}^2$ is *homogeneous* (it looks the same everywhere) and *isotropic* (it looks the same in every direction), which are also true of $\mathbb{E}^2$ and $\mathbb{S}^2$. Isotropy follows by considering rotations about the origin, which are Möbius transformations preserving D. We also have:

Fact 5.7. Möbius transformations are continuous and conformal (angle-preserving).

We are able to use Proposition 5.5 to identify the bi-infinite geodesics of $\mathbb{H}^2$.

Proposition 5.8. *Bi-infinite geodesics in $\mathbb{H}^2$ are diameters of D or arcs of Euclidean circles orthogonal to ∂D.*

Proof. We first show that the interval $(-1,1)$ on the real axis is a bi-infinite geodesic in $\mathbb{H}^2$. Let $w=0$ and let $z=x$ be some distinct point on $(-1,1)$. Thus, in Definition 5.1 we have $w'=\pm 1$ and $z'=\mp 1$, corresponding to whether x is in $(-1,0)$ or $(0,1)$, respectively. Hence

$$\mathsf{d}_D(z,w)=\ln\left(\frac{|\mp 1-0|\cdot|\pm 1-x|}{|\mp 1-x|\cdot|\pm 1-0|}\right)=\ln\left(\frac{1\mp x}{1\pm x}\right)=\left|\ln\left(\frac{1+x}{1-x}\right)\right|.$$

Note that as x varies from -1 to 1, $g(x):=\ln(\frac{1+x}{1-x})$ varies continuously from $-\infty$ to ∞. Thus, we may define a continuous bijective map $\gamma:\mathbb{R}\to(-1,1)$ by taking $\gamma(y)=\frac{e^y-1}{e^y+1}$, noting that $\gamma(y)=g^{-1}(y)$. This shows that $(-1,1)$ is a bi-infinite geodesic in $\mathbb{H}^2$.

Note that any pair of distinct points on ∂D determines a unique Euclidean circle C in $\mathbb{C}\cup\{\infty\}$ orthogonal to ∂D. Furthermore, Möbius transformations are conformal, i.e. they preserve angles. Thus, by taking a third point on the arc of C inside D, Proposition 5.5 (2),(3) gives us a Möbius transformation f taking $(-1,1)$ to $C\cap D$, which must also take ∂D to ∂D by conformality. Proposition 5.5 (4),(6) then implies f restricts to an isometry of $\mathbb{H}^2$. Thus, all diameters of D and arcs of Euclidean circles orthogonal to ∂D are bi-infinite geodesics of $\mathbb{H}^2$. Furthermore, since all pairs of points of D lie on such a circle, we see that all bi-infinite geodesics are of this form. $\square$

Exercise 5.9. Let $\alpha:[0,L]\to D$ be a smooth path. The length of α in $\mathbb{H}^2$ is given by

$$l_D(\alpha)=\int_0^L \frac{2|\alpha'(t)|}{1-|\alpha(t)|^2}\,dt.$$

Show that the length of a circle of radius r in $\mathbb{H}^2$ is $2\pi\sinh r$. Conclude that distances in $\mathbb{H}^2$ grow exponentially as you move further from the origin.

5.3 Tessellations

One can readily see that angles in $\mathbb{S}^2$ are strictly larger than in $\mathbb{E}^2$, while angles in $\mathbb{H}^2$ are strictly smaller than $\mathbb{E}^2$ (we define the angle between two geodesics in $\mathbb{S}^2$ or $\mathbb{H}^2$ by taking tangent lines at the intersection point of the geodesics and measuring the Euclidean angle between the tangents). For example, this can be seen by considering geodesic triangles in $\mathbb{E}^2$, $\mathbb{S}^2$, and $\mathbb{H}^2$.

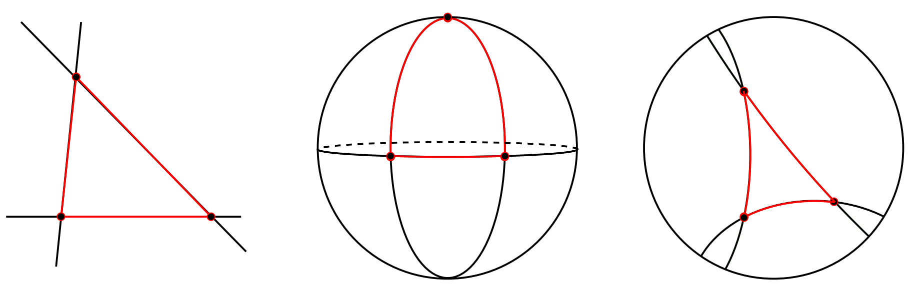

Figure 3: Geodesic triangles in $\mathbb{E}^2$, $\mathbb{S}^2$, and $\mathbb{H}^2$.

In $\mathbb{E}^2$, the sum of the angles in a triangle must equal 180 degrees, or π radians. However, in $\mathbb{S}^2$ the sum is always strictly larger than π, and in $\mathbb{H}^2$ the sum is always strictly smaller than π. In fact, one can observe that as the vertices of a triangle in $\mathbb{H}^2$ approach ∂D, the sum of the angles approaches 0. This follows from the fact that geodesics in $\mathbb{H}^2$ meet ∂D at right angles, meaning if two geodesics meet ∂D at the same point, their tangent lines are equal. A triangle with all its vertices on ∂D is called an *ideal triangle*.

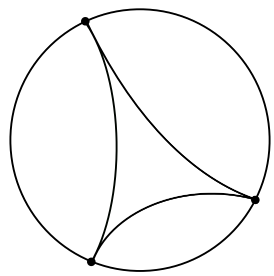

Figure 4: An ideal triangle.

Exercise 5.10. Show that the centre of an ideal triangle in $\mathbb{H}^2$ is at distance $\frac{1}{2}\ln(3)$ from each of its three sides.

This can be generalised to hyperbolic n–gons. In fact, we have something stronger.

Proposition 5.11. *Let $n \in \mathbb{N}_{\geqslant 3}$. Recall that the angles of a Euclidean n–gon are equal to $(1 - \frac{2}{n})\pi$. For all $0 < \theta < (1 - \frac{2}{n})\pi$, there is a regular hyperbolic n–gon with angles equal to θ.*

Proof. Take a regular hyperbolic n–gon centred at the origin in D. Pushing all the vertices out to ∂D at the same rate, the angles of the n–gon decrease continuously to 0. Conversely, if the vertices are all pushed towards the origin at the same rate, the hyperbolic n–gon approaches a Euclidean n–gon, thus the angles increase continuously to $(1 - \frac{2}{n})\pi$. The intermediate value theorem thus implies that there is a hyperbolic n–gon with angles θ for each $0 < \theta < (1 - \frac{2}{n})\pi$. □

We can use this result to show that every regular polygon tessellates either $\mathbb{S}^2$, $\mathbb{E}^2$, or $\mathbb{H}^2$. In fact, the vast majority of tessellations occur in $\mathbb{H}^2$.

Theorem 5.12. *Let $m, n \in \mathbb{N}_{\geqslant 3}$.*

- *If $\frac{1}{m} + \frac{1}{n} > \frac{1}{2}$, there is a tessellation of $\mathbb{S}^2$ by regular n–gons such that m such n–gons meet at each vertex. There are five such tessellations, corresponding to $(m, n) = (3, 3), (3, 4), (3, 5), (4, 3), (5, 3)$.*
- *If $\frac{1}{m} + \frac{1}{n} = \frac{1}{2}$, there is a tessellation of $\mathbb{E}^2$ by regular n–gons such that m such n–gons meet at each vertex. There are three such tessellations, corresponding to $(m, n) = (3, 6), (4, 4), (6, 3)$.*
- *If $\frac{1}{m} + \frac{1}{n} < \frac{1}{2}$, there is a tessellation of $\mathbb{H}^2$ by regular n–gons such that m such n–gons meet at each vertex. There are infinitely many such tessellations, corresponding to all pairs (m, n) other than those listed above.*

Proof. In order for a tessellation by regular n–gons to exist with m such n–gons around each vertex, the interior angles of the n–gons must be $\theta = \frac{2\pi}{m}$, so that the angles around each vertex add up to 2π. Furthermore, notice that the inequalities $\frac{1}{m} + \frac{1}{n} > \frac{1}{2}$, $\frac{1}{m} + \frac{1}{n} = \frac{1}{2}$, and $\frac{1}{m} + \frac{1}{n} < \frac{1}{2}$ are equivalent to $\frac{2\pi}{m} > (1 - \frac{2}{n})\pi$, $\frac{2\pi}{m} = (1 - \frac{2}{n})\pi$, and $\frac{2\pi}{m} < (1 - \frac{2}{n})\pi$. Thus, these three inequalities correspond to the interior angles of the n–gons being larger than that of a Euclidean n–gon, equal to that of a Euclidean n–gon, and smaller than that of a Euclidean n–gon, respectively. Thus, a regular tessellation satisfying one of these three inequalities (if it exists) must be a tessellation of $\mathbb{S}^2$, $\mathbb{E}^2$, or $\mathbb{H}^2$, respectively.

There are only five pairs of integers (m, n) with $m, n \geqslant 3$ that satisfy the inequality $\frac{1}{m} + \frac{1}{n} > \frac{1}{2}$, namely $(m, n) = (3, 3), (3, 4), (3, 5), (4, 3), (5, 3)$. These give tessellations of $\mathbb{S}^2$ corresponding to the five platonic solids: the tetrahedron, octahedron, icosahedron, cube, and dodecahedron, respectively. (The platonic solids can be seen as tessellations of $\mathbb{S}^2$ by bending the faces so they lie on the surface of $\mathbb{S}^2$.)

Similarly, there are only three pairs of integers (m, n) with $m, n \geqslant 3$ satisfying $\frac{1}{m} + \frac{1}{n} = \frac{1}{2}$, namely $(m, n) = (3, 6), (4, 4), (6, 3)$, corresponding to regular tilings of $\mathbb{E}^2$ by triangles, squares, and hexagons, respectively.

On the other hand, there are infinitely many solutions to $\frac{1}{m} + \frac{1}{n} < \frac{1}{2}$. Recall that this inequality corresponds to $0 < \theta < (1 - \frac{2}{n})\pi$, where $\theta = \frac{2\pi}{m}$. Proposition 5.11 now tells us there is a tessellation of $\mathbb{H}^2$ for each such pair (m, n); we can see this by constructing a regular n–gon centred at the origin in $\mathbb{H}^2$ with interior angles $\theta = \frac{2\pi}{m}$ and reflecting the polygon in its edges to produce a tessellation. □

6 Surface groups

In this section we introduce an important class of groups, called *surface groups*. Throughout this section we shall use the notation $[g, h] = ghg^{-1}h^{-1}$ to denote the commutator of two elements g, h in a group G.

Definition 6.1 (Surface group)**.** Let $g \geqslant 1$. The *surface group of genus g*, denoted $\pi_1(S_g)$, is the group with the following presentation:

$$\pi_1(S_g) = \langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \cdot [a_2, b_2] \cdot \ldots \cdot [a_g, b_g] = 1 \rangle.$$

Remark 6.2. These groups appear in topology as the *fundamental groups of closed orientable surfaces*, however we shall not study them from this point of view in this course.

We see that if $g = 1$, then $\pi_1(S_g)$ is quasi-isometric to $\mathbb{E}^2$, while for all $g \geqslant 2$, the surface group $\pi_1(S_g)$ is quasi-isometric to $\mathbb{H}^2$. Thus, almost all surfaces have hyperbolic geometry.

Theorem 6.3. *The Cayley graph of $\pi_1(S_g)$ is quasi-isometric to $\mathbb{E}^2$ if $g = 1$, and quasi-isometric to $\mathbb{H}^2$ if $g \geqslant 2$.*

Proof. Consider the Cayley graph of $\pi_1(S_g)$ with respect to the generating set given in the definition. Denote this Cayley graph by $\Delta(\pi_1(S_g))$. Recall that cycles in the Cayley graph of a group correspond to words w in the generators with $w = 1$. In particular, the relator $[a_1, b_1] \cdot \ldots \cdot [a_g, b_g] = 1$ gives cycles of length $4g$ in $\Delta(\pi_1(S_g))$; see Figure 5. Moreover, since the presentation of $\pi_1(S_g)$ has only one relator, $\Delta(\pi_1(S_g))$ can be constructed entirely by gluing cycles of length $4g$ of the form shown in Figure 5 to each other along edges with the same label.

Furthermore, each vertex of $\Delta(\pi_1(S_g))$ has valence $4g$, with one edge coming out of it for each of the $2g$ generators and their inverses. Thus, by assigning angles of $\frac{2\pi}{4g} = \frac{\pi}{2g}$ to

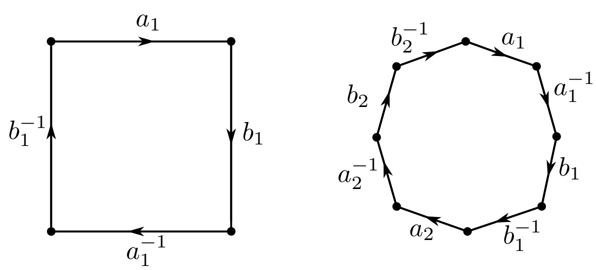

Figure 5: The cycles of length $4g$ for $g = 1$ and $g = 2$.

$\Delta(\pi_1(S_g))$ and applying Theorem 5.12, we can realise $\Delta(\pi_1(S_g))$ as a tessellation by regular $4g$–gons with $4g$ such polygons around each vertex. If $g = 1$, this gives a tessellation of $\mathbb{E}^2$ by squares. If $g \geqslant 2$, this gives a tessellation of $\mathbb{H}^2$. That is, there is an embedding $i : \Delta(\pi_1(S_g)) \to \mathbb{E}^2$ if $g = 1$ and $i : \Delta(\pi_1(S_g)) \to \mathbb{H}^2$ if $g \geqslant 2$.

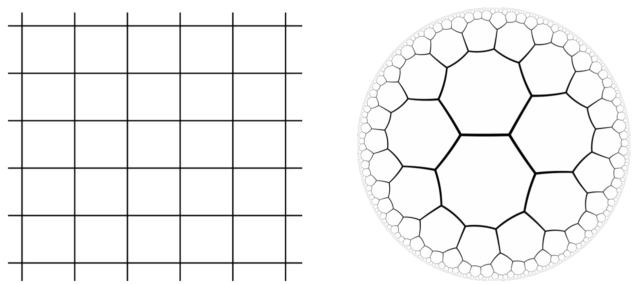

Figure 6: Tessellations of $\mathbb{E}^2$ by squares and $\mathbb{H}^2$ by octagons.

Since we have a regular tessellation, the edges of each polygon have the same finite length L, thus i is an $(L, 0)$–quasi-isometric embedding. Furthermore, each polygon has the same finite area A, so $\mathbb{H}^2$ (or $\mathbb{E}^2$ if $g = 1$) is contained in a finite neighbourhood of $\Delta(\pi_1(S_g))$. Thus, i is a quasi-isometry. $\square$

The groups $\pi_1(S_g)$ are called surface groups because if we take the quotient of $\mathbb{H}^2$ (or $\mathbb{E}^2$ if $g = 1$) by the action of $\pi_1(S_g)$ by translation, we get a surface, obtained by gluing edges of the $4g$–gon labelled by the same letter; see Figure 7 below. By considering this action on $\mathbb{H}^2$ (or $\mathbb{E}^2$) and applying the Milnor-Schwarz lemma, we obtain an alternate proof of Theorem 6.3.

Exercise 6.4. Prove Theorem 6.3 using the Milnor–Schwarz lemma.

Exercise 6.5. The *non-orientable surface group of genus g* is defined as

$$\pi_1(N_g) = \langle c_1, \ldots, c_g \mid c_1^2 c_2^2 \ldots c_g^2 = 1 \rangle.$$

1. Show that $\pi_1(S_1)$ embeds as an index 2 subgroup of $\pi_1(N_2)$. Conclude that $\pi_1(N_2)$ is quasi-isometric to $\pi_1(S_1)$.
2. (Very hard) Show that $\pi_1(S_{g-1})$ embeds as an index 2 subgroup of $\pi_1(N_g)$. Conclude that $\pi_1(N_g)$ is quasi-isometric to $\pi_1(S_{g-1})$.

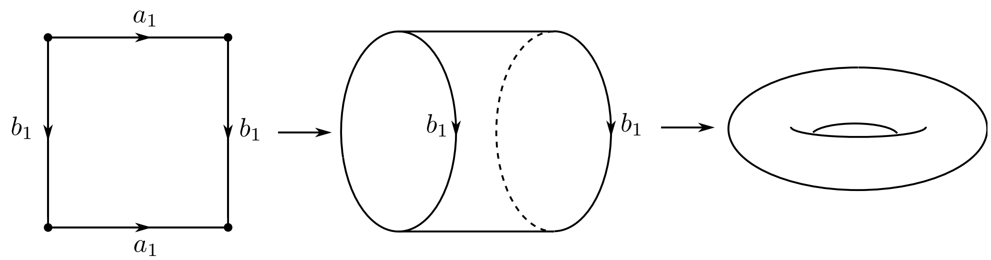

Figure 7: A torus is obtained from the square when the edges labelled by a_1 are glued to each other and the edges labelled by b_1 are glued to each other.

7 Hyperbolic metric spaces

The goal of this section is to define what it means for a general metric space to be hyperbolic, and show that this property is invariant under quasi-isometry. This will then give us a well-defined notion of a hyperbolic group. This form of hyperbolicity is often called "Gromov hyperbolicity", as it was originally developed by Mikhail Gromov in the 1980s.

7.1 Definition of δ–hyperbolicity

As we shall see throughout this course, there are a number of equivalent ways to define a hyperbolic metric space. Some of them are applicable to any metric space (X, d), however for simplicity we shall always assume (X, d) is a *geodesic* metric space; that is, for any two points of X there exists a geodesic between them. This is a natural assumption to make, as our eventual goal is to apply this to Cayley graphs, which are always geodesic metric spaces.

By assuming (X, d) is a geodesic metric space, we can define triangles in (X, d):

Definition 7.1. A *geodesic triangle* T in (X, d) consists of three points $x, y, z \in X$ together with three geodesic segments α, β, γ (one joining each pair of points). We call the three geodesic segments the *sides* of T. We often refer to a geodesic triangle by the geodesic segments it consists of, writing $T = (\alpha, \beta, \gamma)$.

Definition 7.2. Let $\delta \geqslant 0$ and let $T = (\alpha, \beta, \gamma)$ be a geodesic triangle in (X, d). A point $p \in X$ is said to be a *δ–centre* of T if $\max\{\mathsf{d}(p, \alpha), \mathsf{d}(p, \beta), \mathsf{d}(p, \gamma)\} \leqslant \delta$.

Definition 7.3. We say X is *δ–hyperbolic* if every geodesic triangle in X has a δ–centre. We say X is *hyperbolic* if there exists some $\delta \geqslant 0$ such that X is δ–hyperbolic. We call δ the *hyperbolicity constant*.

Notice that the presence of a δ–centre causes triangles in δ–hyperbolic metric spaces to have the same pinched-in appearance of triangles in $\mathbb{H}^2$. See Figure 8.

Warning 7.4. It is important that the same δ works for *all* geodesic triangles. If we were allowed to pick different values of δ for each triangle, then we could always just make δ larger than the diameter of the triangle, making our definition useless!

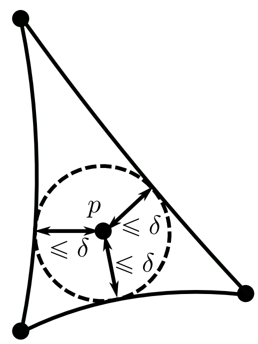

Figure 8: A δ–centre p of a geodesic triangle.

Example 7.5. We list a few basic examples and non-examples of hyperbolic metric spaces.

1. If a metric space has finite diameter d then it is d–hyperbolic.
2. Trees are 0–hyperbolic.
3. $\mathbb{H}^2$ is $(\frac{1}{2}\ln 3)$–hyperbolic.
4. $\mathbb{E}^2$ is not hyperbolic.

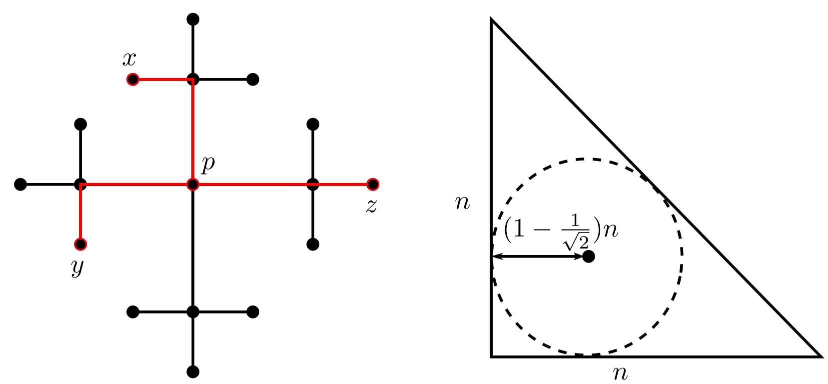

Figure 9: Triangles in trees always look like tripods, and the centre of the tripod is a 0–centre. For $\mathbb{E}^2$, a right-angled triangle with non-hypotenuse sides of length n has at best a $(1-\frac{1}{\sqrt{2}})n$–centre, so no constant δ works for all geodesic triangles.

Exercise 7.6. Prove that $\mathbb{H}^2$ is $(\frac{1}{2}\ln 3)$–hyperbolic:

1. Show that any triple of points $p, q, r \in \partial D$ is taken to any other triple of points $p', q', r' \in \partial D$ by an isometry of $\mathbb{H}^2$. Conclude that all ideal triangles are isometric.
2. Let T be an ideal triangle with its vertices equally spaced along ∂D. Show that T has a $(\frac{1}{2}\ln 3)$–centre.
3. Show that any triangle in $\mathbb{H}^2$ is contained in an ideal triangle. Conclude that every triangle has a $(\frac{1}{2}\ln 3)$–centre, and hence $\mathbb{H}^2$ is $(\frac{1}{2}\ln 3)$–hyperbolic.

7.2 Geodesics in hyperbolic metric spaces

A useful tool in hyperbolic geometry is the *Gromov product*, defined as follows.

Definition 7.7 (Gromov product)**.** Let (X, d) be a δ–hyperbolic metric space. The *Gromov product* of three points $x, y, z \in X$ is

$$\langle x, y\rangle_z = \frac{1}{2}(\mathsf{d}(x,z) + \mathsf{d}(y,z) - \mathsf{d}(x,y)).$$

Exercise 7.8. Use the triangle inequality to show that the Gromov product is always non-negative.

In a hyperbolic metric space X, one can think of the Gromov product as a rough measure of how close the point z is to a geodesic between the points x and y. Indeed, if X is a tree, then every geodesic triangle with vertices x, y, z looks like a tripod, as shown in Figure 9. Denote the centre point of the tripod by p.

One can then see that the geodesic connecting x to z overlaps with the geodesic connecting y to z along the subpath from p to z. Thus,

$$\begin{aligned}
\langle x, y\rangle_z &= \frac{1}{2}(\mathsf{d}(x,z) + \mathsf{d}(y,z) - \mathsf{d}(x,y)) \\
&= \frac{1}{2}(\mathsf{d}(x,p) + \mathsf{d}(p,z) + \mathsf{d}(y,p) + \mathsf{d}(p,z) - \mathsf{d}(x,p) - \mathsf{d}(p,y)) \\
&= \mathsf{d}(p,z),
\end{aligned}$$

which is precisely the distance from z to the geodesic from x to y. In general, we have something slightly weaker:

Lemma 7.9. *If α is a geodesic from x to y, then $\mathsf{d}(z,\alpha) \geqslant \langle x, y\rangle_z$.*

Proof. Let a be a point on α. Then by the triangle inequality,

$$\begin{aligned}
\mathsf{d}(x,y) &= \mathsf{d}(x,a) + \mathsf{d}(a,y) \\
\mathsf{d}(x,z) &\leqslant \mathsf{d}(x,a) + \mathsf{d}(a,z) \\
\mathsf{d}(y,z) &\leqslant \mathsf{d}(y,a) + \mathsf{d}(a,z).
\end{aligned}$$

Putting these into the formula for the Gromov product, we get $\langle x, y\rangle_z \leqslant \mathsf{d}(a,z)$. □

In fact, we also have an inequality in the other direction:

Lemma 7.10. *Let $x, y, z \in X$, and let α be a geodesic from x to y. Then $\mathsf{d}(z,\alpha) \leqslant \langle x, y\rangle_z + 4\delta$.*

Proof. Let $T = (\alpha, \beta, \gamma)$ be a triangle with vertices x, y, z, so that α connects x and y, β connects x and z, and γ connects z and y. Since X is δ–hyperbolic, T has a δ–centre p. Thus, there is a point a on α such that $\mathsf{d}(a,p) \leqslant \delta$. Thus, by the triangle inequality, a is a 2δ–centre for T. That is, there is a point q on β at most 2δ from a, and there is a point r on γ at most 2δ from a. See Figure 10. In particular, by the triangle inequality we have

$$\begin{aligned}
&\mathsf{d}(x,a) + \mathsf{d}(z,a) \leqslant \mathsf{d}(x,q) + 2\delta + \mathsf{d}(z,q) + 2\delta = \mathsf{d}(x,z) + 4\delta \\
&\mathsf{d}(y,a) + \mathsf{d}(z,a) \leqslant \mathsf{d}(y,r) + 2\delta + \mathsf{d}(z,r) + 2\delta = \mathsf{d}(y,z) + 4\delta \\
&\mathsf{d}(x,a) + \mathsf{d}(y,a) = \mathsf{d}(x,y).
\end{aligned}$$

Add the first two inequalities and subtract the third to give $2\mathsf{d}(z,a) \leqslant 2\langle x, y\rangle_z + 8\delta$, i.e. $\mathsf{d}(z,a) \leqslant \langle x, y\rangle_z + 4\delta$, as required. □

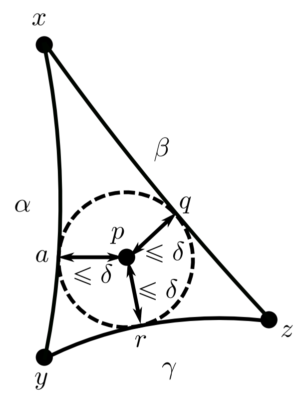

Figure 10: Illustration of Lemma 7.10.

Putting these two lemmas together, we see that $\langle x, y\rangle_z \leqslant \mathsf{d}(z, \alpha) \leqslant \langle x, y\rangle_z + 4\delta$. Thus, the Gromov product $\langle x, y\rangle_z$ is equal to the distance from z to a geodesic α connecting x and y, up to an error of 4δ. Another way of saying this is that they are equal up to applying a $(0, 4\delta)$–quasi-isometry.

We can use this to show that in a hyperbolic space, any two geodesics with the same endpoints travel alongside each other at a bounded distance.

Corollary 7.11. *Let α and β be two geodesics connecting the same pair of points. Then α is contained in the 4δ–neighbourhood of β and vice versa.*

Proof. Denote the endpoints by x and y, and let z be a point on α. Then $\langle x, y\rangle_z \leqslant \mathsf{d}(z, \alpha) = 0$ by Lemma 7.9, so $\mathsf{d}(z, \beta) \leqslant \langle x, y\rangle_z + 4\delta = 4\delta$ by Lemma 7.10. Similarly, if we take z' to be a point on β then we obtain $\mathsf{d}(z', \alpha) \leqslant 4\delta$. □

7.3 The slim triangles definition

In this section, we give a different but equivalent definition of δ–hyperbolicity that is commonly used in the literature.

Notation 7.12. If α is a path and a, b are two points on α, then we write $\alpha[a, b]$ to denote the subpath of α between a and b.

Definition 7.13 (Taut)**.** Let α be a path with endpoints x, y. We say α is t–*taut* if $\text{length}(\alpha) \leqslant \mathsf{d}(x, y) + t$.

Tautness measures how close a path is to being a geodesic. In particular, 0–taut paths are geodesics.

Exercise 7.14. Show that any subpath of a t–taut path is t–taut.

We see that Corollary 7.11 can be generalised to t–taut paths, with the caveat that the bound on the distance between the paths depends on t.

Lemma 7.15. *Let α be a geodesic and let β be a t–taut path with the same endpoints as α.*

1. *β is contained in the $(\frac{1}{2}t + 4\delta)$–neighbourhood of α.*
2. *α is contained in the $(t + 8\delta)$–neighbourhood of β.*

Proof. Let x, y denote the endpoints of α and β.

1. Let z be a point on β. Then
$$\langle x, y\rangle_z = \frac{1}{2}(\mathsf{d}(x,z) + \mathsf{d}(y,z) - \mathsf{d}(x,y)) \leqslant \frac{1}{2}(\text{length}(\beta) - \mathsf{d}(x,y)) \leqslant \frac{1}{2}t$$
by t–tautness of β. Thus, by Lemma 7.10, $\mathsf{d}(z,\alpha) \leqslant \frac{1}{2}t + 4\delta$.

2. Let w be a point on α, and let A_x and A_y denote the $(\frac{1}{2}t + 4\delta)$–neighbourhoods of $\alpha[x,w]$ and $\alpha[w,y]$, respectively. By part (1), $\beta \subseteq A_x \cup A_y$. In particular, since β is connected and contains points of both A_x and A_y, it follows that A_x and A_y must intersect in some point z on β. Thus, there are some points a_x on $\alpha[x,w]$ and a_y on $\alpha[w,y]$ such that $\mathsf{d}(z,a_x) \leqslant \frac{1}{2}t + 4\delta$ and $\mathsf{d}(z,a_y) \leqslant \frac{1}{2}t + 4\delta$. Thus, $\mathsf{d}(a_x,a_y) \leqslant \mathsf{d}(a_x,z) + \mathsf{d}(z,a_y) \leqslant t + 8\delta$ by the triangle inequality. Since w lies on $\alpha[a_x,a_y]$, it follows that either $\mathsf{d}(w,a_x) \leqslant \frac{1}{2}t + 4\delta$ or $\mathsf{d}(w,a_y) \leqslant \frac{1}{2}t + 4\delta$. Since a_x and a_y are both distance at most $\frac{1}{2}t + 4\delta$ from z, it follows that $\mathsf{d}(w,\beta) \leqslant t + 8\delta$, as required. □

We are now ready to set up our alternate definition of δ–hyperbolicity, which is stated in terms of *slim triangles*.

Definition 7.16 (Slim)**.** We say a triangle $T = (\alpha,\beta,\gamma)$ is *δ–slim* if each side is contained in the δ–neighbourhood of the union of the other two sides.

Theorem 7.17. *If a geodesic triangle has a δ–centre, then it is 6δ–slim. Conversely, if a geodesic triangle is δ'–slim, then it has a δ'–centre.*

Proof. Let $T = (\alpha,\beta,\gamma)$ be a geodesic triangle and let x,y,z be the vertices of T, where x,y are the endpoints of α, y,z are the endpoints of β, and z,x are the endpoints of γ. Suppose T has a δ–centre p. Then there are some points a,b,c on α,β,γ respectively such that $\mathsf{d}(a,p) \leqslant \delta$, $\mathsf{d}(b,p) \leqslant \delta$, $\mathsf{d}(c,p) \leqslant \delta$. Let ξ be a geodesic from a to z. Then $\text{length}(\xi) \leqslant 2\delta + \mathsf{d}(c,z)$ and $\text{length}(\alpha[a,x]) \leqslant 2\delta + \mathsf{d}(c,x)$. Because c lies on the geodesic γ from x to z, we therefore have $\text{length}(\xi \cup \alpha[a,x]) \leqslant \mathsf{d}(x,z) + 4\delta$, so the path $\xi \cup \alpha[a,x]$ is 4δ–taut. By Lemma 7.15, $\alpha[a,x] \subseteq \xi \cup \alpha[a,x]$ is contained in the 6δ–neighbourhood of γ. Similarly, $\alpha[a,y]$ is contained in the 6δ–neighbourhood of β. Thus, α is contained in the 6δ–neighbourhood of $\beta \cup \gamma$. The proofs for the other two sides proceed similarly, showing that T is 6δ–slim.

Conversely, suppose T is δ'–slim. Then α is contained in the δ'–neighbourhood of $\beta \cup \gamma$. Let A_β and A_γ denote the δ'–neighbourhoods of β and γ, respectively, so that $\alpha \subseteq A_\beta \cup A_\gamma$. Since α is connected and contains points of both A_β and A_γ, it follows that A_β and A_γ must intersect in some point p on α. Thus, $\mathsf{d}(p,\alpha) = 0$, $\mathsf{d}(p,\beta) \leqslant \delta'$, $\mathsf{d}(p,\gamma) \leqslant \delta'$, and so p is a δ'–centre of T. □

Thus, we have the following equivalent definition of δ–hyperbolicity. Be careful to state which definition you are using, because as we saw in the above theorem, the constants involved may differ by a factor of 6 depending which definition is used.

Definition 7.18 (Slim triangles definition)**.** A geodesic metric space X is said to be δ'–*hyperbolic* if all geodesic triangles are δ'–slim.

Exercise 7.19. We say that a geodesic n–gon $P = (\alpha_1, \ldots, \alpha_n)$ is δ'–slim if each side is contained in the δ'–neighbourhood of the union of the other $n-1$ sides. Use the slim triangles definition to show that every geodesic n–gon in a δ'–hyperbolic space X is $(n-2)\delta'$–slim.

7.4 Exponential growth

We show that δ–hyperbolic spaces exhibit exponential growth in a similar way to $\mathbb{H}^2$ (cf. Exercise 5.9). That is to say, distances grow exponentially as you move further away from a fixed point.

Proposition 7.20. *Let X be a δ–hyperbolic space, and let α be a path between two points $p, q \in X$. If γ is a geodesic between p and q, then for every point x on γ,*

$$\mathsf{d}(x, \alpha) \leqslant \delta |\log_2(\text{length}(\alpha))| + 2.$$

Proof. Let $n \in \mathbb{Z}$ such that $2^n \leqslant \text{length}(\alpha) \leqslant 2^{n+1}$ and subdivide α into 2^n segments of length $\frac{1}{2^n}\text{length}(\alpha)$. Use this subdivision to construct geodesic triangles as in Figure 11. Then, using the slim triangles definition of δ–hyperbolicity, we can construct a path from x to a point y on α via points $y_1, \ldots, y_n$ on the geodesic triangles as shown in Figure 11, where $\mathsf{d}(x, y_1) \leqslant \delta$ and $\mathsf{d}(y_i, y_{i+1}) \leqslant \delta$ for all i. Furthermore, $\mathsf{d}(y_n, y) \leqslant 2$ because $\mathsf{d}(y_n, y) \leqslant \frac{1}{2^n}\text{length}(\alpha) \leqslant 2$. Thus, $\mathsf{d}(x, \alpha) \leqslant n\delta + 2 \leqslant \delta|\log_2(\text{length}(\alpha))| + 2$. □

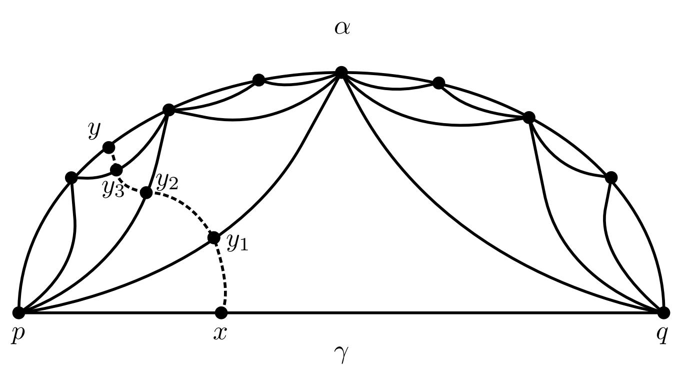

Figure 11: Illustration of Proposition 7.20.

Fix a point $p \in X$. Write $S(p, r) = \{x \in X \mid \mathsf{d}(p, x) = r\}$, $B(p, r) = \{x \in X \mid \mathsf{d}(p, x) \leqslant r\}$, and $\mathring{B}(p, r) = \{x \in X \mid \mathsf{d}(p, x) < r\}$. We see that the lengths of paths in X connecting two points $x, y \in S(p, r)$ and avoiding $\mathring{B}(p, r)$ grows exponentially with $\mathsf{d}(x, y)$, and thus with r.

Proposition 7.21. *Let α be a path in $X \smallsetminus \mathring{B}(p, r)$ connecting $x, y \in S(p, r)$. Then*

$$\text{length}(\alpha) \geqslant 2^{\frac{\mathsf{d}(x,y) - 8\delta - 4}{2\delta}} - 1.$$

Proof. Suppose $\delta = 0$. Then no such paths α exist, thus the result holds vacuously. We may therefore assume $\delta > 0$ without loss of generality.

Suppose length$(\alpha) \leqslant 1$. Then $\mathsf{d}(x,y) \leqslant 1$, so $2^{\frac{\mathsf{d}(x,y)-8\delta-4}{2\delta}} - 1 < 0$. Thus, the inequality holds trivially.

Now suppose length$(\alpha) > 1$. Let γ be a geodesic connecting x and y. By Proposition 7.20, length$(\alpha) \geqslant 2^{\frac{\mathsf{d}(z,\alpha)-2}{\delta}}$ for any point z on γ. By Lemma 7.10, we have

$$\begin{aligned}\mathsf{d}(p,\gamma) &\leqslant \langle x,y\rangle_p + 4\delta \\ &= \frac{1}{2}(\mathsf{d}(x,p) + \mathsf{d}(y,p) - \mathsf{d}(x,y)) + 4\delta \\ &= r - \frac{1}{2}\mathsf{d}(x,y) + 4\delta.\end{aligned}$$

Thus, there exists some point z on γ with $\mathsf{d}(p,z) \leqslant r - \frac{1}{2}\mathsf{d}(x,y) + 4\delta$. Suppose $\mathsf{d}(x,y) > 8\delta$, so that $\gamma \subseteq B(p,r)$. Since α avoids $\mathring{B}(p,r)$, we have $\mathsf{d}(z,\alpha) \geqslant r - \mathsf{d}(z,p) \geqslant \frac{1}{2}\mathsf{d}(x,y) - 4\delta$. Thus,

$$\text{length}(\alpha) \geqslant 2^{\frac{\mathsf{d}(x,y)-8\delta-4}{2\delta}}.$$

On the other hand, if $\mathsf{d}(x,y) \leqslant 8\delta$, then $2^{\frac{\mathsf{d}(x,y)-8\delta-4}{2\delta}} < 1 < \text{length}(\alpha)$. □

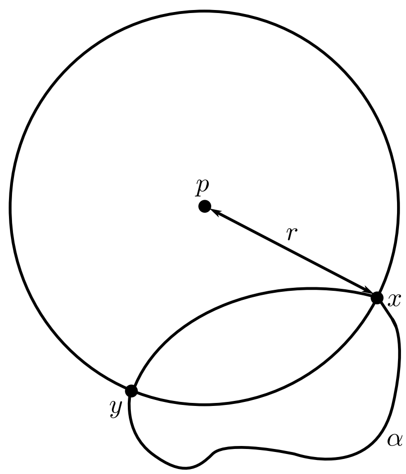

Figure 12: Illustration of Proposition 7.21.

7.5 Quasi-geodesics

In geometric group theory, we are interested in properties that are invariant under quasi-isometry. Notice that geodesics may no longer be geodesics once a quasi-isometry has been applied! We therefore introduce the notion of a *quasi-geodesic*. This gives us a useful class of paths that behave nicely in hyperbolic spaces.

Definition 7.22 (Quasi-geodesic)**.** An (A,B)–*quasi-geodesic* in a metric space X is an (A,B)–quasi-isometric embedding $\gamma : I \to X$, where I is a (possibly unbounded) interval of $\mathbb{R}$. We say $\gamma : I \to X$ is a *quasi-geodesic* if it is an (A,B)–quasi-geodesic for some A, B.

We often abuse notation by using the term quasi-geodesic when we really mean its image as a subset of X.

Remark 7.23. Note that an (A, B)–quasi-geodesic $\gamma : I \to X$ is not necessarily connected, since two points that are next to each other in I may be sent to points in X that are up to a distance B from each other. In particular, γ is not necessarily a path. However, it is always possible to turn γ into a path γ' by connecting the endpoints of consecutive connected components of γ with geodesic segments. Furthermore, γ' is contained in the B–neighbourhood of γ and in particular is itself an (A, B)–quasi-geodesic.

Note that by composing a quasi-geodesic path γ with an appropriate rescaling map from $\mathbb{R}$ to $\mathbb{R}$, we may assume that for all $t_1, t_2 \in I$, we have $\text{length}(\gamma([t_1, t_2])) = |t_1 - t_2|$. Working under this assumption, we therefore have:

Proposition 7.24. *Let γ be an (A, B)–quasi-geodesic path. Then for all points x, y on γ,*

$$\text{length}(\gamma[x, y]) \leqslant A\mathsf{d}(x, y) + B.$$

If X is δ–hyperbolic, then the following result tells us that quasi-geodesics always travel close to geodesics.

Lemma 7.25 (Morse Lemma)**.** *Let α be a geodesic in a δ–hyperbolic space X and let β be an (A, B)–quasi-geodesic with the same endpoints as α. Then there exists some $R \geqslant 0$, depending only on A, B, δ, such that β is contained in the R–neighbourhood of α and vice versa.*

Proof. Let a, b denote the endpoints of α, and let p be the point on α furthest from β. Let $t = \mathsf{d}(p, \beta)$. In particular, $\mathsf{d}(p, a) \geqslant t$. There is therefore a point a_0 on $\alpha[a, p]$ with $\mathsf{d}(p, a_0) = t$. Define a_1 to be the point on $\alpha[a, p]$ with $\mathsf{d}(p, a_1) = 2t$, if it exists. If such a point does not exist, define $a_1 = a$. By definition of p, we must have $\mathsf{d}(a_1, \beta) \leqslant t$. Thus, there is some point a_2 on β with $\mathsf{d}(a_1, a_2) \leqslant t$. If $a_1 = a$, set $a_2 = a$ too. Define points b_0, b_1, b_2 similarly by considering $\alpha[p, b]$ instead. See Figure 13 for an illustration.

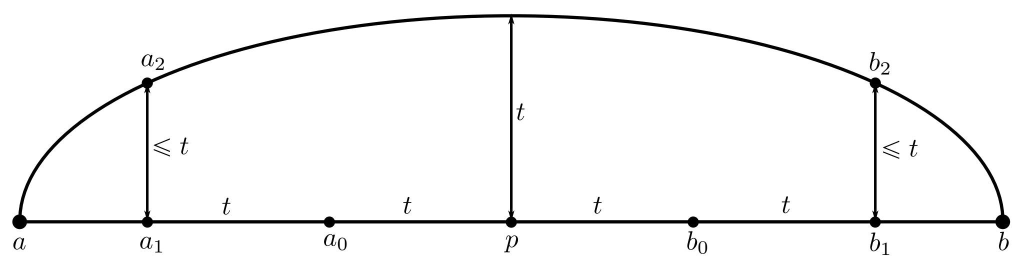

Figure 13: Illustration of the proof of the Morse Lemma.

Let β' be the (A, B)–quasi-geodesic path corresponding to β, as constructed in Remark 7.23. Let $\gamma = \alpha[a_0, a_1] \cup [a_1, a_2] \cup \beta'[a_2, b_2] \cup [b_2, b_1] \cup \alpha[b_1, b_0]$. Note that $\mathsf{d}(a_2, b_2) \leqslant 6t$ by the triangle inequality, so by applying Proposition 7.24 to β', we have

$$\begin{aligned}
\text{length}(\beta'[a_2, b_2]) &\leqslant A\mathsf{d}(a_2, b_2) + B \\
&\leqslant 6At + B, \\
\text{length}(\gamma) &\leqslant 4t + \text{length}(\beta'[a_2, b_2]) \\
&\leqslant (6A + 4)t + B.
\end{aligned}$$

Note that by construction $\gamma \cap \mathring{B}(p,t) = \varnothing$, and $a_0, b_0 \in S(p,t)$. Thus, by Proposition 7.21, we have

$$\text{length}(\gamma) \geqslant 2^{\frac{2t-8\delta-4}{2\delta}} - 1 = 2^{\frac{t-4\delta-2}{\delta}} - 1.$$

Putting these together, we have

$$2^{\frac{t-4\delta-2}{\delta}} \leqslant (6A+4)t + B + 1.$$

This gives an upper bound R on t that depends only on A, B, δ. By definition of t, we have shown that α is contained in the R–neighbourhood of β.

Now let q be a point on β and let B_a and B_b denote the R–neighbourhoods of $\beta[a,q]$ and $\beta[q,b]$, respectively. We know that $\alpha \subseteq B_a \cup B_b$. In particular, since α is connected and contains points of both B_a and B_b, it follows that B_a and B_b must intersect in some point x on α. Thus, there are some points y_a on $\beta[a,q]$ and y_b on $\beta[q,b]$ such that $\mathsf{d}(x, y_a) \leqslant R$ and $\mathsf{d}(x, y_b) \leqslant R$. Thus, $\mathsf{d}(y_a, y_b) \leqslant 2R$ by the triangle inequality. Since q lies on $\beta[y_a, y_b]$, it follows that either $\mathsf{d}(q, y_a) \leqslant R$ or $\mathsf{d}(q, y_b) \leqslant R$. Since y_a and y_b are both distance at most R from x, it follows that $\mathsf{d}(q, \alpha) \leqslant 2R$. Thus, β is contained in the $2R$–neighbourhood of α, as required. ☐

Exercise 7.26. (a) Show that the logarithmic spiral $\gamma : (0, \infty) \to \mathbb{E}^2$ given by

$$\gamma(t) = (t\cos(\ln t), t\sin(\ln t))$$

is a quasi-geodesic. (Hint: to get the upper bound, use the mean value theorem.)

(b) Show that the logarithmic spiral γ does not satisfy the Morse Lemma; that is, show there is no $R \geqslant 0$ such that for all subsegments $\gamma[a,b]$, the geodesic α in $\mathbb{E}^2$ with endpoints a, b is contained in the R–neighbourhood of $\gamma[a,b]$ and vice versa. This provides another proof that $\mathbb{E}^2$ is not hyperbolic.

We can use the Morse Lemma to show that the definition of hyperbolicity could equally well have been formulated in terms of (A,B)–quasi-geodesic triangles.

Lemma 7.27. *Any (A,B)–quasi-geodesic triangle $T = (\alpha, \beta, \gamma)$ has a δ'–centre, where δ' depends only on A, B, δ.*

Proof. Let $T' = (\alpha', \beta', \gamma')$ be a geodesic triangle with the same vertices as T. Then T' has a δ–centre p. Applying Lemma 7.25, we see that the distance from p to α, β, and γ is at most $\delta + R$. Thus, p is a $(\delta + R)$–centre for T. ☐

7.6 Quasi-isometry invariance of δ–hyperbolicity

We conclude with the most important result of this section, which allows us to define a hyperbolic group.

Theorem 7.28. *Suppose that X and X' are geodesic spaces and let $f : X' \to X$ be an (A,B)–quasi-isometric embedding. If X is δ–hyperbolic then X' is δ'–hyperbolic, where δ' depends only on A, B, δ.*

Proof. Suppose X is δ–hyperbolic and let $f : X' \to X$ be an (A,B)–quasi-isometric embedding. Let $T' = (\alpha', \beta', \gamma')$ be a geodesic triangle in X' and note that $T = (f(\alpha'), f(\beta'), f(\gamma'))$

is a quasi-geodesic triangle in X. Thus, by Lemma 7.27, T has a δ''–centre p, where δ'' depends only on A, B, δ. That is, there exist points $f(a'), f(b'), f(c')$ on $f(\alpha'), f(\beta'), f(\gamma')$, respectively, such that

$$\mathsf{d}(p, f(a')), \mathsf{d}(p, f(b')), \mathsf{d}(p, f(c')) \leqslant \delta''.$$

Thus,

$$\mathsf{d}(f(a'), f(a')), \mathsf{d}(f(a'), f(b')), \mathsf{d}(f(a'), f(c')) \leqslant 2\delta'',$$

so $f(a')$ is a $2\delta''$–centre for T. Since f is a quasi-isometric embedding, we have

$$\begin{aligned} \mathsf{d}'(a', b') &\leqslant A\mathsf{d}(f(a'), f(b')) + B \leqslant 2A\delta'' + B \\ \mathsf{d}'(a', c') &\leqslant A\mathsf{d}(f(a'), f(c')) + B \leqslant 2A\delta'' + B, \end{aligned}$$

so a' is a $(2A\delta'' + B)$–centre for T'. Thus, X' is δ'–hyperbolic, where $\delta' = 2A\delta'' + B$. □

Corollary 7.29. *Suppose that X and X' are geodesic spaces with X quasi-isometric to X'. Then X is hyperbolic if and only if X' is.*

Proof. If X is quasi-isometric to X' then there exists a quasi-isometric embedding $f : X' \to X$ and a quasi-isometric embedding $g : X \to X'$. Thus, by Theorem 7.28, X is hyperbolic if and only if X is. □

This theorem gives us a tool for showing two spaces are not quasi-isometric to each other, by showing that one space is hyperbolic and the other is not. For example, we have the following immediate consequences:

1. If $n \geqslant 2$, then $\mathbb{R}^n$ is not quasi-isometric to a tree.
2. $\mathbb{R}^m \not\sim \mathbb{H}^n$ for all $m, n \geqslant 2$.

Exercise 7.30. Show that two surface groups $\pi_1(S_g)$ and $\pi_1(S_h)$ are quasi-isometric if and only if either $g = h = 1$ or $g, h \geqslant 2$.

8 Hyperbolic groups

We now define a hyperbolic group and prove some nice properties.

8.1 Definition and examples

Definition 8.1. A group G is *hyperbolic* if it is finitely generated and its Cayley graph is hyperbolic.

This is well-defined and does not depend on the choice of Cayley graph – we saw in the first half of the course that all Cayley graphs of a finitely generated group are quasi-isometric to each other, and Corollary 7.29 tells us that if two spaces are quasi-isometric, then one is hyperbolic if and only if the other is.

Applying the Milnor–Schwarz Lemma together with Corollary 7.29, we also have the following result:

Lemma 8.2. *Suppose a group G acts properly discontinuously and cocompactly on a proper hyperbolic space. Then G is hyperbolic.*

Example 8.3. We list some examples and non-examples of hyperbolic groups.

1. All finite groups are hyperbolic, since their Cayley graphs are finite and therefore have bounded diameter.

2. All free groups are hyperbolic, as their Cayley graphs are trees, which are 0–hyperbolic spaces.

3. All surface groups $\pi_1(S_g)$ with $g \geqslant 2$ are hyperbolic, since they are quasi-isometric to $\mathbb{H}^2$.

4. $\mathbb{Z}^n$ is not hyperbolic for any $n \geqslant 2$, since its Cayley graph is quasi-isometric to $\mathbb{E}^n$.

Exercise 8.4. Define the (p,q,r)*–triangle group* $T(p,q,r)$ to be the group with the following presentation:

$$T(p,q,r) = \langle a,b,c \mid a^2 = b^2 = c^2 = (ab)^p = (ac)^q = (bc)^r = 1\rangle.$$

Show that $T(p,q,r)$ is hyperbolic if and only if $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$. (Hint: Use tessellations.)

Exercise 8.5. Define the *Baumslag–Solitar group* $BS(m,n)$ to be the group with the following presentation:

$$BS(m,n) = \langle a,b \mid a^m = ba^nb^{-1}\rangle.$$

(a) Show that $BS(1,n)$ is not hyperbolic for any $n \geqslant 1$.

(b) (Hard) Show that $BS(m,n)$ is not hyperbolic for any $m,n \geqslant 1$.

8.2 Subgroups of hyperbolic groups

We examine which groups can/cannot appear as subgroups of hyperbolic groups. We start by introducing a nice class of subgroups called *quasi-convex* subgroups.

Definition 8.6 (Quasi-convex)**.** A subspace Y of a geodesic metric space X is said to be *quasi-convex* if there exists a constant $K \geqslant 0$ such that for all $y_1, y_2 \in Y$, each geodesic in X joining y_1 to y_2 is contained in the K–neighbourhood of Y.

We consider a subgroup $H \leqslant G$ to be quasi-convex if it is quasi-convex as a subspace of the Cayley graph. We show that quasi-convex subgroups always quasi-isometrically embed.

Lemma 8.7. *Let G be a group with finite generating set S and let H be a subgroup of G. If H is quasi-convex in $\Delta(G,S)$, then H is finitely generated and quasi-isometrically embeds in G via the inclusion map, with respect to any word metric.*

Proof. Let K be the quasi-convexity constant for H in $\Delta(G,S)$. Given $h \in H$, take a geodesic γ from 1 to h in $\Delta(G,S)$, and suppose it is labelled by $a_1,\ldots,a_n$. Since H is K–quasi-convex, γ is contained in the K–neighbourhood of H. Thus, for each $1 \leqslant i \leqslant n$, $\mathsf{d}(a_1\ldots a_i, H) \leqslant K$. Thus, for each i there is some word u_i of length at most K such that $a_1\ldots a_iu_i \in H$. In particular, by multiplying $a_1\ldots a_iu_i$ on the left by $(a_1\ldots a_{i-1}u_{i-1})^{-1}$, we see that $h_i := u_{i-1}^{-1}a_iu_i \in H$ for each $1 \leqslant i \leqslant n$ (define $u_0 = 1$ and note that $u_n = 1$ too). See Figure 14.

We see that $h = h_1\ldots h_n$, and moreover each h_i has word length at most $2K+1$ since each u_i has word length at most K and a_i is a generator. Thus, we have shown that any

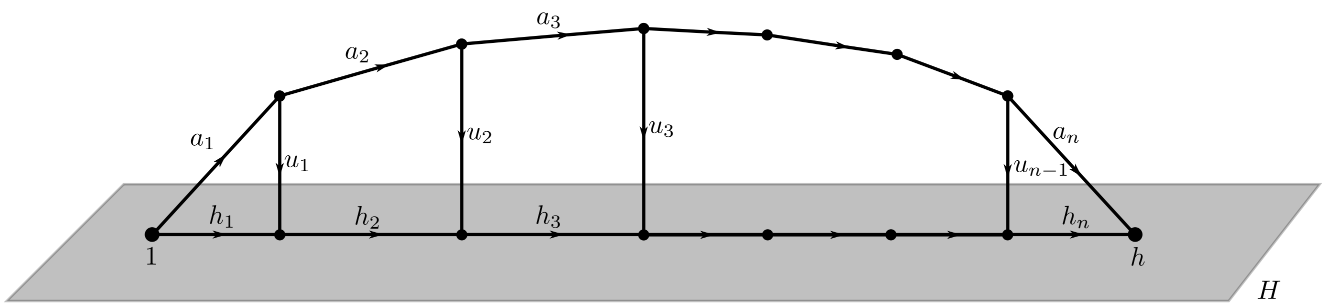

Figure 14: Illustration of the proof of Lemma 8.7.

element of H can be written as a word in elements of H that lie in the ball of radius $2K+1$ in $\Delta(G,S)$. That is, H is generated by the (finite) set of elements that lie in this ball. Furthermore, the distance from 1 to h in the word metric associated to this generating set is n, which is also the distance from 1 to h in $\Delta(G,S)$. Thus, the inclusion map from H to $\Delta(G,S)$ is a quasi-isometric embedding. □

Corollary 8.8. *Let G be a hyperbolic group with finite generating set S and let $H \leqslant G$ be quasi-convex in $\Delta(G,S)$. Then H is hyperbolic.*

Proof. By Lemma 8.7, H is finitely generated and quasi-isometrically embeds in G. Thus, by Theorem 7.28, H is hyperbolic. □

Recall the definition of the *centraliser* of an element and the *centre* of a group:

Definition 8.9 (Centraliser and centre)**.** For $g \in G$, the *centraliser* of g is

$$C_G(g) = \{h \in G \mid gh = hg\}.$$

The *centre* of G is

$$Z(G) = \{h \in G \mid \forall g \in G,\, gh = hg\}.$$

We take the following two results as facts.

Proposition 8.10. *If H_1 and H_2 are quasi-convex subgroups of G with finite generating sets, then $H_1 \cap H_2$ is also quasi-convex.*

Proposition 8.11. *Let G be a hyperbolic group. The centraliser $C_G(g)$ of each element $g \in G$ is a quasi-convex subgroup. In particular, $C_G(g)$ is hyperbolic.*

We can now show that infinite order elements in a hyperbolic group generate quasi-geodesics in the Cayley graph:

Lemma 8.12. *Let G be a hyperbolic group with finite generating set S and let $g \in G$ be an infinite order element. Then $\langle g \rangle$, considered as a subset of the Cayley graph $\Delta(G,S)$, is a quasi-geodesic.*

Proof. We already saw that there exists a quasi-isometry $f : \mathbb{R} \to \mathbb{Z}$. Define $\gamma = \gamma' \circ f$, where $\gamma' : \mathbb{Z} \to \Delta(G,S)$ is defined as $\gamma'(n) = g^n$. It suffices to show that γ' is a quasi-isometric embedding.

Recall that the centraliser $C_G(g)$ is quasi-convex by Proposition 8.11, thus $C_G(g)$ is finitely generated and hyperbolic. In particular, we can apply Proposition 8.11 again to show that $C_{C_G(g)}(h)$ is quasi-convex in $C_G(g)$ for any $h \in C_G(g)$. Note that the centre of a group can be expressed as the intersection of the centralisers of the generators; this can be seen from the definitions. Since $C_G(g)$ is finitely generated, $Z(C_G(g))$ can therefore be expressed as a finite intersection of subgroups of the form $C_{C_G(g)}(h)$. As these are all quasi-convex, Proposition 8.10 implies $Z(C_G(g))$ is quasi-convex in $C_G(g)$, hence finitely generated and hyperbolic by Corollary 8.8.

Since $Z(C_G(g))$ is abelian by definition, it must contain $\mathbb{Z}^n$ as a finite-index subgroup for some $n \geqslant 0$. Thus, $Z(C_G(g))$ is quasi-isometric to $\mathbb{Z}^n$. But we know that $\mathbb{Z}^n$ is not hyperbolic for $n \geqslant 2$. We must therefore have $n \leqslant 1$. Moreover, since $\langle g \rangle \cong \mathbb{Z}$ is a subgroup of $Z(C_G(g))$, it follows that $Z(C_G(g))$ must be quasi-isometric to $\langle g \rangle$.

Recall that we showed $Z(C_G(g))$ is quasi-convex in $C_G(g)$, which is quasi-convex in G. Applying Lemma 8.7, we therefore have the following sequence of quasi-isometric embeddings induced by the inclusion maps:

$$\langle g \rangle \to Z(C_G(g)) \to C_G(g) \to \Delta(G, S).$$

We have therefore shown that γ' is a quasi-isometric embedding, as required (noting that $\mathbb{Z} \cong \langle g \rangle$ via the map $n \mapsto g^n$). $\square$

We now look at some pathological subgroups that provide obstructions to hyperbolicity.

Proposition 8.13. *A hyperbolic group cannot contain any subgroup isomorphic to $\mathbb{Z}^2$.*

Proof. Let G be a δ–hyperbolic group with respect to a finite generating set S and let $g \in G$ be an infinite order element. We wish to show there is no other infinite order element $h \in G$ that commutes with g, meaning g is not a generator of a $\mathbb{Z}^2$ subgroup of G. Suppose such an h existed. Then $g, h \in C_G(g)$ and in particular the $\mathbb{Z}^2$ subgroup generated by g and h is contained in $C_G(g)$. Thus, it is sufficient to show $C_G(g)$ is contained in a bounded neighbourhood of $\langle g \rangle \cong \mathbb{Z}$ and hence cannot contain a $\mathbb{Z}^2$ subgroup.

Suppose $h^{-1} g^p h = g^q$ for some $h \in G$, $p, q \in \mathbb{Z}$. Without loss of generality, suppose $|p| \leqslant |q|$. Then $h^{-1} g^{p^m} h = g^{q^m}$ for all $m \in \mathbb{Z}$, so

$$\mathsf{d}(1, g^{q^m}) \leqslant 2\mathsf{d}(1, h) + \mathsf{d}(1, g^{p^m}) = 2\mathsf{d}(1, h) + \frac{|p|^m}{|q|^m} \mathsf{d}(1, g^{q^m}).$$

Since $\langle g \rangle$ is a quasi-geodesic by Lemma 8.12, we must have $|p| = |q|$. Thus, the positive powers of g lie in distinct conjugacy classes. In particular, all conjugates of g lie at least some distance ε from 1. Replacing g with a suitable power if necessary, we may therefore assume no conjugate of g is at distance $\leqslant 5\delta$ from 1.

We claim that if $h \in G$ commutes with g, then $\mathsf{d}(h, \langle g \rangle) \leqslant 2\mathsf{d}(1, g) + 4\delta =: K$. This would then imply $C_G(g)$ is contained in the K–neighbourhood of $\langle g \rangle$, completing the proof. Suppose this is not the case and let $r \in \mathbb{Z}$ such that $\mathsf{d}(h, \langle g \rangle) = \mathsf{d}(h, g^r)$. Note that if h commutes with g, then so does $g^{-r}h$, and furthermore $\mathsf{d}(g^{-r}h, 1) = \mathsf{d}(g^{-r}h, g^{-r}g^r) = \mathsf{d}(h, g^r) = \mathsf{d}(h, \langle g \rangle) > K$. Thus, replacing h with $g^{-r}h$, we may assume $\mathsf{d}(h, 1) > K$. Let Q be a geodesic quadrilateral in $\Delta(G, S)$ with vertices 1, g, $gh = hg$, h and sides $[1, g], [g, gh] = g[1, h], [hg, h] = h[g, 1], [h, 1]$, and let h_m be the point on $[h, 1]$ at distance $\mathsf{d}(1, g) + 3\delta$ from 1, as shown in Figure 15, so that $\mathsf{d}(h_m, 1) > \frac{1}{2}K = \mathsf{d}(1, g) + 2\delta$ and $\mathsf{d}(h_m, h) > \frac{1}{2}K = \mathsf{d}(1, g) + 2\delta$.

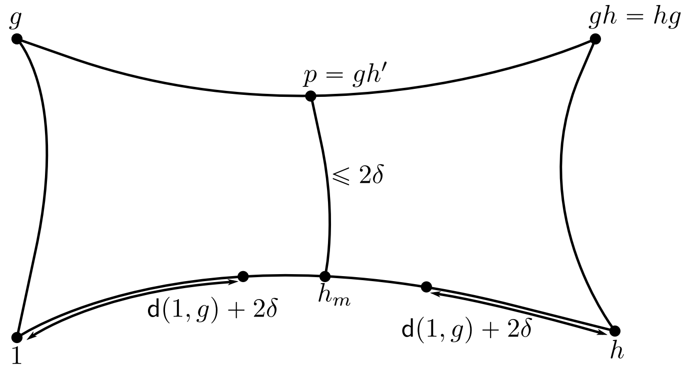

Figure 15: Illustration of the proof of the Proposition 8.13.

Since $\Delta(G,S)$ is δ–hyperbolic, $\mathsf{d}(h_m,p) \leqslant 2\delta$ for some point p on one of the other three sides of Q by the slim triangles definition. Moreover, since $\mathsf{d}(h_m,1) > \mathsf{d}(1,g)+2\delta$ and $\mathsf{d}(h_m,h) > \mathsf{d}(1,g)+2\delta$, we see that p cannot lie on $[1,g]$ or $[hg,h]$, otherwise it would violate the triangle inequality. Thus, p lies on $[g,gh]$ and therefore has the form $p = gh'$ for some h' on $[1,h]$. Recall that we chose h so that $\mathsf{d}(h,\langle g\rangle) = \mathsf{d}(h,1)$. Thus, $\mathsf{d}(h_m,\langle g\rangle) = \mathsf{d}(h_m,1)$ too. Since $\mathsf{d}(h_m,1) > \mathsf{d}(1,g)+2\delta$ and $\mathsf{d}(h_m,g) \leqslant \mathsf{d}(h_m,gh') + \mathsf{d}(gh',g) \leqslant \mathsf{d}(gh',g)+2\delta$, we therefore see that we must have $\mathsf{d}(h',1) = \mathsf{d}(gh',g) > \mathsf{d}(1,g)$. This implies that $\mathsf{d}(h',h_m) \leqslant 3\delta$, so $\mathsf{d}(h',gh') \leqslant 5\delta$ by the triangle inequality. But $\mathsf{d}(h',gh') = \mathsf{d}(1,(h')^{-1}gh')$, so this contradicts our assumption that no conjugate of g is at distance $\leqslant 5\delta$ from 1. $\square$

In fact, this implies that a hyperbolic group cannot contain a $\mathbb{Z}^n$ subgroup for any $n \geqslant 2$, since $\mathbb{Z}^2$ appears as a subgroup of all of these.

Recall that the *Baumslag–Solitar group* $BS(m,n)$ is defined as

$$BS(m,n) = \langle a,b \mid a^m = ba^nb^{-1}\rangle.$$

We see that Baumslag–Solitar groups also provide an obstruction to hyperbolicity. In fact, since $BS(1,1) = \mathbb{Z}^2$, this is a strict generalisation of Proposition 8.13. We shall not include a proof here, but we note that this can be proved using centralisers in a similar manner to Proposition 8.13.

Proposition 8.14. *A hyperbolic group cannot contain any Baumslag–Solitar subgroup.*

Note that there are hyperbolic groups G that contain non-hyperbolic subgroups H (this is highly non-trivial!), so the converse is not true; if H contained a Baumslag–Solitar subgroup, then G would also contain a Baumslag–Solitar subgroup, contradicting Proposition 8.14.

Exercise 8.15. Show that in $BS(1,n) = \langle a,b \mid a = ba^nb^{-1}\rangle$, the cyclic subgroup $\langle a\rangle$ is not a quasi-geodesic.

8.3 Other properties of hyperbolic groups

We conclude by listing some other nice properties of hyperbolic groups, without proof.

Recall that by definition, hyperbolic groups are finitely generated. In fact, they are finitely presented.

Theorem 8.16. *Every hyperbolic group is finitely presented.*

Hyperbolic groups always contain free groups:

Theorem 8.17. *Every hyperbolic group that is not finite or virtually $\mathbb{Z}$ contains a free subgroup of rank 2, and hence free subgroups of any countable rank.*

Hyperbolicity is equivalent to having a *linear isoperimetric function*:

Definition 8.18. Let G be a finitely presented group and fix a finite presentation of G. Suppose w is a word in the generators and their inverses representing the identity in G. Then we can reduce w to the trivial word by repeatedly applying the relations. We say G has a *linear isoperimetric function* if we only need to do this at most $f(n)$ times, where n is the length of w and $f : \mathbb{N} \to \mathbb{N}$ is linear.

Theorem 8.19. *A group is hyperbolic if and only if it has a linear isoperimetric function.*

In fact, the order of an isoperimetric function is invariant under quasi-isometry. For example, $\mathbb{Z}^n$ has a quadratic isoperimetric function for all $n \geqslant 2$.

Hyperbolic groups also have nice algorithmic properties. For example, one can ask the following question: for a finitely presented group $G = \langle S \mid R \rangle$, is there an algorithm to decide whether a word w in the generators S satisfies $w = 1$? We say the group G has *solvable word problem* if the answer is yes.

Theorem 8.20. *Hyperbolic groups have solvable word problem. In fact, the word problem can be solved in linear time (with respect to the length of the word).*

We can ask other similar questions; for example, is there an algorithm to decide whether one word is conjugate to another? This is known as the *conjugacy problem.*

Theorem 8.21. *The conjugacy problem is solvable in linear time for hyperbolic groups.*