6. Hyperbolic groups.

In this section we will explore some of the basic properties of hyperbolic groups. The notion of a hyperbolic group was introduced by Gromov around 1985. They arise in many different contexts, and there is a sense in which a "generic" finitely presented group is hyperbolic.

For much of the discussion we will just deal with geodesic spaces. One can get quite a long way with just elementary metric space theory as we shall see.

6.1. Definition of a hyperbolic space.

Let (X, d) be a geodesic metric space.

Definition : A *(geodesic) triangle*, T, in X consists of three geodesic segments, (α, β, γ) cyclically connecting three point (called the *vertices* of T). We refer to the geodesics segments as the *sides* of T.

Definition : If $k \geq 0$, a point, $p \in X$ is said to be a *k-centre* for the triangle T if $\max\{d(p,\alpha), d(p,\beta), d(p,\gamma)\} \leq k$.

See Figure 6a. (In the figures in this section, geodesics are often depicted curved inwards, rather than as euclidean straight lines. This is meant to evoke the Poincaré model of the hyperbolic plane, to which hypebolic spaces have a more natural resemblence.)

Definition : We say that X is *k-hyperbolic* if every triangle has a k-centre.

Definition : We say that X is *hyperbolic* if it is k-hyperbolic for some $k \geq 0$. We refer to such a k as a *hyperbolicity constant* for X.

Examples.

(1) Any space of finite diameter, k, is k-hyperbolic.

(2) Any tree is 0-hyperbolic

(3) The hyperbolic plane $\mathbf{H}^2$ is $(\frac{1}{2}\log 3)$-hyperbolic.

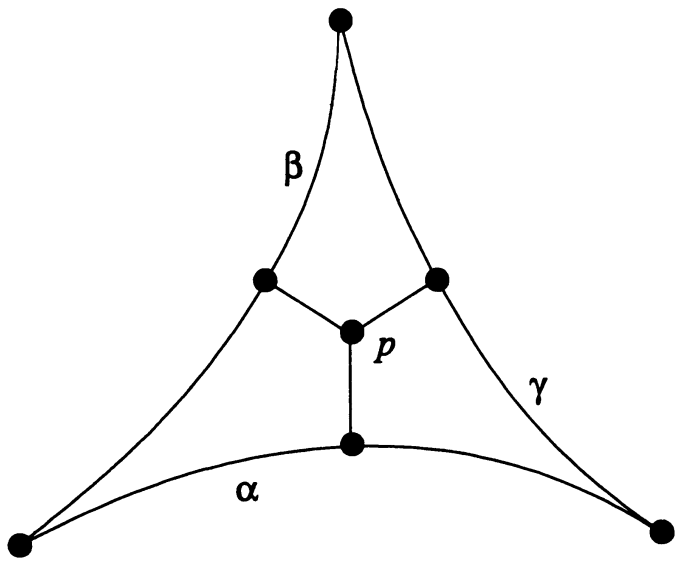

Figure 6a.

(4) In fact, hyperbolic space $\mathbf{H}^n$ of any dimension is $(\frac{1}{2}\log 3)$-hyperbolic: any triangle in $\mathbf{H}^n$ lies in some 2-dimensional plane.

(5) Indeed, any complete simply connected riemanian manifold with curvatures bounded above by some negative constant $-\kappa^2 < 0$ is $(\frac{1}{2\kappa}\log 3)$-hyperbolic. For example, complex and quaternionic hyperbolic spaces are $(\frac{1}{2}\log 3)$-hyperbolic.

In (2) we can consider more general trees than those considered in Section 2. In particular, we can allow any positive length assigned to an edge (rather than just unit length). The result will always be a 0-hyperbolic geodesic space.

Remark: In fact, any 0-hyperbolic geodesic space is a more general sort of tree known as an "$\mathbf{R}$-tree". Here one can allow branch points (i.e. valence ≥ 3 points) to accumulate, so such a tree need not be a graph. (Indeed there are examples where every point is a branch point.) The theory of $\mathbf{R}$-trees was introduced by Morgan and Shalen and developed by Rips and many others, and it is now an important tool in geometric group theory.

Non-examples.

(1) Euclidean space, $\mathbf{R}^n$ for $n \geq 2$ is not hyperbolic.

(2) The 1-skeleton of the regular square tessellation of the plane is not hyperbolic. In fact, this example illustrates a slightly subtle point. It turns out that any three points of this graph can be connected by three geodesics so the triangle formed has a 1-centre (excercise). However not every triangle has this property. In fact, in this graph, we can have two geodesics between the same pair of points which go an arbitrarily long way apart before coming back together again.

6.2. Basic properties.

Before studing properties of a geodesic space, we make a couple of observations that hold in any geodesic metric space.

Let (X, d) be a metric space. We will often abbreviate $d(x, y)$ to xy. Given $x, y, z \in X$, write

$$\langle x, y\rangle_z = \frac{1}{2}(xz + yz - xy).$$

This is sometimes called the "Gromov product". The triangle inequality tells us this is non-negative. One way to think of it is as follows. Set $r = \langle y, z\rangle_x$, $s = \langle z, x\rangle_y$ and $t = \langle x, y\rangle_z$. Then

$$\begin{aligned} xy &= r + s \\ yz &= s + t \\ zx &= t + r. \end{aligned}$$

We can construct a "tripod" consisting of three edges meeting at a vertex of valence three and place the points x, y, z at the other endpoints of these edges (Figure 6b). If we assign the edge lengths r, s, t to these edges, we see that distances between x, y, z in X agree with those in the tripod. (This tripod might not be isometrically embedable in X.)

Another point to note is:

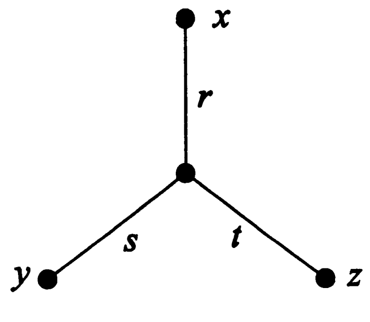

Figure 6b.

Lemma 6.1 : *If α is any geodesic from x to y, then $d(z, \alpha) \geq \langle x, y \rangle_z$.*

Proof : If $a \in \alpha$, then we have

$$\begin{aligned} xy &= xa + ax \\ xz &\leq xa + az \\ yz &\leq ya + az, \end{aligned}$$

and so $az \geq \langle x, y \rangle_z$. ◊

Note also that if z lies on any geodesic from x to y, then $\langle x, y \rangle_z = 0$.

Suppose now that (X, d) is k-hyperbolic. We prove a series of lemmas involving various constants. We aim to provide arguments that are fairly simple, rather than ones that will optimise the constants involved. With more careful arguments, one can probably do better in this regard.

Suppose that $T = (\alpha, \beta, \gamma)$ is a geodesic triangle. If p is any *k-centre*, we can find some $a \in \alpha$ with $ap \leq k$. Such a point a, is then a $2k$-centre for T.

Lemma 6.2 : *Suppose $x, y, z \in X$, and α is any geodesic connecting x to y. Then $d(z, \alpha) \leq \langle x, y \rangle_z + 4k$.*

Proof : Let $t = \langle x, y\rangle_z$. Let β, γ be geodesics from z to x and y respetively (Figure 6c). Let $a \in \alpha$ be a $2k$-centre for the triangle (α, β, γ). Thus

$$xa + az \leq xz + 4k$$
$$ya + az \leq yz + 4k$$
$$xa + ay = xy.$$

Adding the first two of these and subtracting the third, we get $2az \leq 2t + 8k$, and so $az \leq t + 4k$ as required. ◇

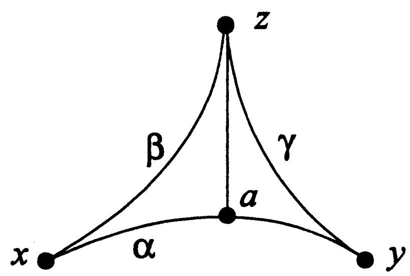

Figure 6c.

Corollary 6.3 : *If α and β are two geodesics connecting the same pair of points, then $\alpha \subseteq N(\beta, 4k)$ and $\beta \subseteq N(\alpha, 4k)$.*

Proof : Let the common endpoints be x and y, and suppose $z \in \beta$. Then $\langle x, y\rangle_z = 0$ and so by Lemma 6.2, $d(z, \alpha) \leq 4k$. This proves the first inclusion, and the other follows by symmetry. ◇

Thus in a hyperbolic space, any two geodesics with the same endpoints remain a bounded distance apart. We also see that, up to an additive constant, we can think of the Gromov product, $\langle x, y\rangle_z$ as the distance between z and any geodesic from x to y.

Notation: Given any path α and points a, b on α, we write $\alpha[a, b]$ for the subpath of α between a and b.

The following terminology is not standard, but will be useful for our purposes.

Definition : A path α is *t-taut* if length$(\alpha) \leq xy + t$, where x, y are the endpoints of α.

Thus a geodesic is a 0-taut path. Also (exercise) any subpath of a t-taut path is t-taut.

We have the following generalisation of Lemma 6.3:

Lemma 6.4 : *Suppose α is a geodesic and β is a t-taut path with the same endpoints. Then:*
(1) $\beta \subseteq N(\alpha, \frac{1}{2}t + 4k)$, and
(2) $\alpha \subseteq N(\beta, t + 8k)$.

Proof : Let x, y be the endpoints of α.
(1) If $z \in \beta$, then $\langle x, y \rangle_z \leq t/2$, and so by Lemma 6.2, $d(z, \alpha) \leq \frac{1}{2}t + 4k$.
(2) Suppose $w \in \alpha$. By a connectedness argument using part (1), we can find some $z \in \beta$ a distance at most $\frac{1}{2}t + 4k$ from points a and b in α, on different sides of w. (Consider the closed subsets, $\beta \cap N(\alpha[x, w], \frac{1}{2}t + 4k)$ and $\beta \cap N(\alpha[y, w], \frac{1}{2}t + 4k)$. By (1) these cover β and so must intersect.) Thus $ab \leq t + 8k$ and $w \in \alpha[a, b]$, so w is a distance at most $\frac{1}{2}t + 4k$ from one of the points a or b (Figure 6d).

It follows that $wz \leq t + 8k$ as required. ◇

Lemma 6.5 : *If (α, β, γ) is a geodesic triangle, then $\alpha \subseteq N(\beta \cup \gamma, 6k)$.*

Proof : Let $a \in \alpha$ be a 2k-centre of (α, β, γ). This cuts α into two segments $\alpha[a, x]$ and $\alpha[a, y]$. Let δ be any geodesic from z to a. Since $d(a, \beta) \leq 2k$, the path $\delta \cup \alpha[a, x]$ is $4k$-taut and so by Lemma 6.4(1), $\alpha[a, x] \subseteq N(\beta, 6k)$. Similarly, $\alpha[a, y] \subseteq N(\gamma, 6k)$. ◇

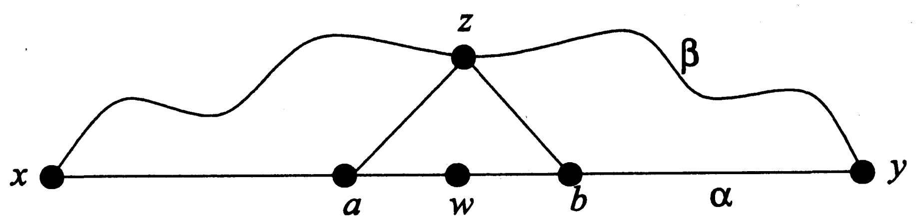

Figure 6d.

Remark: The concusion of Lemma 6.5 gives us an alternative way of defining hyperbolicity. Suppose (α, β, γ) is a geodesic triangle with $\alpha \subseteq N(\beta \cup \gamma, k')$ for some $k' \geq 0$, then by a connectedness argument (similiar to that for proving Lemma 6.4(2)), we can find some point $a \in \alpha$ a distance at most k' for both β and γ. This a will be a k'-centre from (α, β, γ). Thus we can define a space to be hyperbolic if for every geodesic triangle, each edge is a bounded distance from the union of the other two. This definition is equivalent to the one we have given, though the hyperbolicity constants involved may differ by some bounded multiple.

6.3. Projections.

Suppose $x, y, z \in X$ and α is a geodesic connecting x to y. We describe a few different, but essentially equivalent ways of thinking of the notion of a "projection" of z to α.

(P1) One way, we have already seen, is to take geodesics β, γ from z to x and y respectively, and let $a \in \alpha$ be a $2k$-centre for the triangle (α, β, γ). A-priori, this might depend on the choice of β and γ. Here are another two constructions.

(P2) Let $b \in \alpha$ be the unique point so that $xb = \langle y, z \rangle_x$. It follows that $yb = \langle x, z \rangle_y$.

(P3) Choose some $c \in \alpha$ so as to minimise zc. This "neasest point" construction is the closest to what one normally thinks of as projection.

We want to show that these three constructions agree up to bounded distance.

To see this, first note that

$$\begin{aligned} xz &\leq xa + az \leq xz + 4k \\ yz &\leq ya + az \leq yz + 4k \\ xy &= xa + ay, \end{aligned}$$

and so we get $xa - 2k \leq \langle y, z\rangle_x \leq xa + 2k$. It follows that $ab \leq 2k$.

Now note that $zc = d(z, \alpha) = d(z, \alpha[c, x])$. Applying Lemma 6.2 (with α replaced by $\alpha[c, x]$), we see that $zc \leq \langle z, x\rangle_c + 4k$, and so $2zc \leq (zc + zx - cx) + 8k$ giving $zc + cx \leq zx + 8k$. Thus $\langle x, z\rangle_c \leq 4k$ and so by Lemma 6.2 again (with z replaced by c and α replaced by β) we get $d(c, \beta) \leq 4k + 4k \leq 8k$. Similarly, $d(c, \gamma) \leq 8k$. In other words, c is an $8k$-centre for (α, β, γ). We can now apply the argument of the previous paragraph again. The constants have got a bit bigger, and this time we get $bc \leq 8k$.

This shows the above three definitions of projections agree up to bounded distance, depending only on k. It is also worth noting that there is some flexibility in the definitions. For example, if we took a to be any t-centre, or chose $c \in \alpha$ to be any point with $d(z, c) \leq d(z, \alpha) + t$, then we get similar bounds depending only on t and k.

One consequence of this construction is the following:

Lemma 6.6 : *Suppose that $x, y, z \in X$. Then a, b are t-centres of triangles with vertices x, y, z, then ab is bounded in terms of t and k.*

Proof : By Corollary 6.3, a t-centre of one triangle will be a $t + 4k$ centre of any other with the same vertices. We can therefore assume that a and b are centres of the same triangle. We can also assume that they lie on some edge, say α, of this triangle (replacing t by $2t$). The situation is therefore covered by the above discussion. $\diamondsuit$

Of course, one can explicitly calculate the bound in terms of t and k, though such calculations eventually become tedious, and for most purposes it is enough to observe that some formula exists.

We also note that a centre, a, for x, y, z, is decribed up to bounded distance by saying that $ax \leq \langle y, z\rangle_x + t$, $ay \leq \langle z, x\rangle_y + t$ and $az \leq \langle x, y\rangle_z + t$, for some constant t.

Here is another consequence worth noting. Given $x, y, z \in X$, let a be a centre for x, y, z. Let δ, ϵ, ζ be geodesics connecting a to x, y and z respectively. Let τ be the "tripod" $\delta \cup \epsilon \cup \zeta$. This is a tree in X, with extreme points (valence 1 vertices) x, y, z. Note that distances in τ agree with distances in X up to a bounded constant. This is an instance of a much more general result about the "treelike" nature of hyperbolic spaces.

Notation. Given $x, y \in X$ we shall write $[x, y]$ for some choice of geodesic between x and y. If $z, w \in [x, y]$, we will assume that $[z, w] \subseteq [x, y]$.

Of course this involves making a choice, but since any two such geodesics remain a bounded distance apart, in practice this will not matter much. This is just for notational convenience. Formally we can always rephrase any statement to refer to a particular geodesic.

6.4. Trees in hyperbolic spaces.

The following expresses the "treelike" nature of a hyperbolic space:

Proposition 6.7 : *There is a function $h : \mathbf{N} \longrightarrow [0, \infty)$ such that if $F \subseteq X$ with $|F| = n$, then there is a tree, τ, embedded in X, such that for all $x, y \in F$, $d_\tau(x, y) \leq xy + kh(n)$.*

Here d_τ is distance measured in the tree τ. Note that we can assume that all the edges of τ are geodesic segments. We can also assume that every extreme (i.e. valence 1) point of τ lies in F. In this case, τ will be $kh(n)$-taut, in the following sense:

Definition : A tree $\tau \subseteq X$ is *t-taut* if every arc in τ is t-taut.

We will refer to a such a tree, τ, as a "spanning tree" for F.
To prove Propostion 6.7, we will need the following lemma:

Lemma 6.8 : *Suppose $x, y, z \in X$. Suppose that β is a t-taut path from x to y and that y is the nearest point on β to z. Then $\beta \cup [y, z]$ is $(3t + 24k)$-taut.*

Proof : Let α be any geodesic from x to y (Figure 6e). By Lemma 6.5(2), $\alpha \subseteq N(\beta, t + 8k)$. By hypothesis, $d(z, \beta) = yz$, and so $d(z, \alpha) \geq yz - (t + 8k)$. Thus, by Lemma 6.2,

$$\begin{aligned} \langle x, y \rangle_z &\geq d(z, \alpha) - 4k \\ &\geq yz - t - 12k. \end{aligned}$$

That is, $xz + yz - xy \geq 2yz - 2t - 24k$, and so $xy + yz \leq xz + 2t + 24k$. It follows that

$$\begin{aligned} \text{length}(\beta \cup [y, z]) &\leq (xy + t) + yz \\ &\leq xz + 3t + 24k. \end{aligned}$$

◊

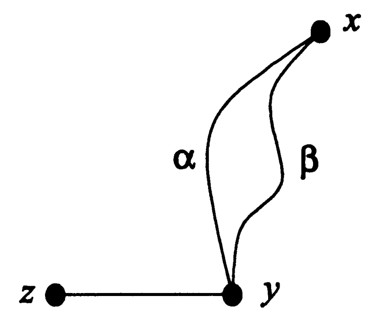

Figure 6e.

Corollary 6.9 : *Suppose that τ is t-taut tree and $z \in X$. Let $y \in \tau$ be a nearest point to z. Then $\tau \cup [y, z]$ is $(3t + 24k)$-taut.* ◇

Proof of Proposition 6.7: Let $F = \{x_1, x_2, \ldots, x_n\}$. Construct τ inductively. Set $\tau_2 = [x_1, x_2]$, and define $\tau_i = \tau_{i-1} \cup [y, x_i]$, where y is a nearest point to x_i in τ_{i-1} (Figure 6f). We now apply Corollary 6.9 inductively, and set $\tau = \tau_n$. ◇

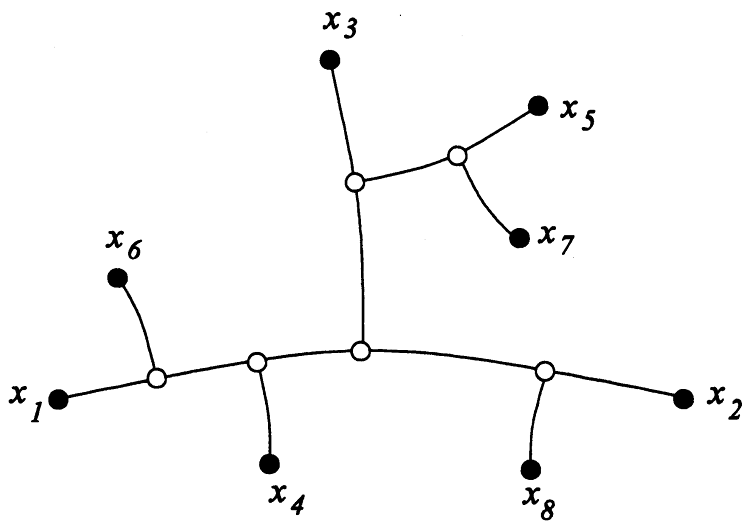

Figure 6f.

We remark that this argument gives $h(n)$ exponential in n. In fact, one can show that the same construction gives $h(n)$ linear in n, but this is more subtle. One cannot do better than linear for an arbitrary ordering of the points of F (exercise). A different construction can be used to give a tree with $h(n) = O(\log n)$. This is the best possible:

Exercise: Let F be a set of n equally spaced points around a circle of radius $r \geq \log(n)$ in $\mathbf{H}^2$. Then no spanning tree can be better than t-taut, where $t = O(r) = O(\log n)$. (Use the fact that the length of a circle of radius r is $2\pi \sinh r$.)

As far as I know, the following question remains open (even for $\mathbf{H}^2$):

Question: In the construction of the tree in Proposition 6.7, can one choose the order of the points $(x_i)_i$ so as always to give a tree with $h(n) = O(\log n)$?

Proposition 6.7 is very useful. We are frequently in a situation where we are dealing just with a bounded number of points. If we are only interested in estimating something up to an additive constant (depending on k), then we can assume we are working in a tree.

For many applications, it is enough to embed our set F is some tree τ, and do not need to know that τ is actually embedded in X. It is possible to construct such a tree by a more direct argument, though we won't describe the construction here.

6.5. The four-point condition.

Let us suppose that τ is a tree containing four points $x, y, z, w \in \tau$. One can see that, measuring distances in τ, we have

$$xy + zw \leq \max\{xz + yw, xw + yz\}.$$

Suppose, for example, that the arcs from x to y and from z to w meet in at most one point (Figure 6g).

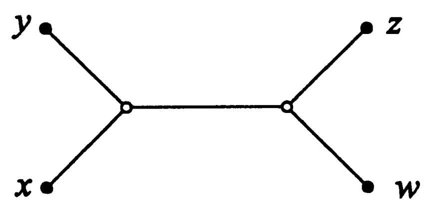

Figure 6g.

In this case, we write $(xy|zw)$, and this situation we see that $xy + zw \leq xz + yw = xw + yz$. Whatever the arrangement of the three points, it is easily seen that at least one of $(xy|zw)$, $(xz|yw)$ or $(xw|yz)$ must hold, thereby verifying the above inequality.

From this we can deduce:

Lemma 6.10 : *Given $k \geq 0$, there is some $k' \geq 0$ such that if X is a k-hyperbolic geodesic spaces, and $x, y, z, w \in X$, then*

$$xy + zw \leq \max\{xz + yw, xw + yz\} + k'.$$

Proof : By Proposition 6.7, we can find a tree τ, containing x, y, z, w, so that distances in τ agree with distances in X up to an additive constant $kh(4)$. We can now apply the above observation. ◇

As usual, k' is some particular multiple of k, which we could calculate explicitly (exercise). (In fact, there are more direct routes to this particular result that would probably give better constants.)

It turns out that hyperbolicity is characterised by this property. Let us suppose, for the moment, that (X, d) is any geodesic space and $k' \geq 0$ is some constant. We suppose:

$$(*) \qquad (\forall x, y, z, w \in X)(xy + zw \leq \max\{xz + yw, xw + yz\} + k').$$

Given $x, y, z \in X$ and a geodesic α from x to y, let $a \in \alpha$ be the point with $xa = \langle y, z\rangle_x$ (cf. the earlier discussion of projections).

Lemma 6.11 : *$xa + az \leq xz + k'$ and $yz + az \leq yz + k'$.*

Proof : Let $r = \langle y, z\rangle_x$, $s = \langle z, x\rangle_y$ and $t = \langle x, y\rangle_z$. Thus,

$$\begin{aligned} xy &= r + s \\ yz &= s + t \\ zx &= t + r \\ xa &= r \\ ya &= s. \end{aligned}$$

Let

$$za = u.$$

We now apply $(*)$ to $\{x, y, z, a\}$. The three distance sums in $(*)$ are

$$\begin{gathered} r + s + u \\ r + s + t \\ r + s + t, \end{gathered}$$

and so $u \leq t + k'$. But now $xa + az = r + u \leq r + t + k' = xz + k'$ and $ya + az = s + u \leq s + t + k' = yz + k'$. ♢

Lemma 6.12 : *In the above situation, let β and γ be geodesics from z to x and from z to y repectively. Then a is a $(3k'/2)$-centre for the triangle (α, β, γ).*

Proof : Let b be the projection of a to β in the above sense. Applying Lemma 6.11 to a and β (in place of z and α) we see that

$$\begin{aligned} ab + bx &\leq ax + k' \\ ab + bz &\leq az + k'. \end{aligned}$$

Adding we get

$$\begin{aligned} 2ab + (xb + bz) &\leq xa + az + 2k' \\ 2ab + xz &\leq xa + az + 2k' \\ &\leq xz + 3k' \end{aligned}$$

applying Lemma 6.11 again to z and α. We see that $ab \leq 3k'/2$. We have shown that $d(a, \beta) \leq 3k'/2$.

Similarly $d(a, \gamma) \leq 3k'/2$ as required. ♢

We have shown that under the assumption $(*)$ every triangle has a $(3k'/2)$-centre. Putting this together with Lemma 6.10, we get:

Proposition 6.13 : *For a geodesic metric space, the condition $(*)$ is equivalent to hyperbolicity.* ♢

We remark that $(*)$ makes no reference to geodesics, and so, in principle, makes sense for any metric space. Its main application, however is to geodesic spaces.

Remark: The "four point" condition $(*)$ is frequently given in the following equivalent form:

$$(\forall x, y, z, w \in X)(\langle x, y\rangle_w \geq \min\{\langle x, z\rangle_w, \langle y, z\rangle_w\} - k'')$$

(where $k'' = k'/2$). Indeed this was the first definition of hyperbolicity given in Gromov's original paper on the subject.

6.6. Exponential growth of distances.

We observed in Section 5 that the length of a hyperbolic circle grows exponentally in the diameter. The following can be viewed as a more general expression of this phenomenon.

We fix a basepoint $p \in X$. We write $N(x, r) = \{y \in X \mid d(x, y) \leq r\}$, and $S(x, r) = \{y \in X \mid d(x, y) = r\}$. We write $N^0(x, r) = N(x, r) \setminus S(x, r)$.

Proposition 6.14 : *There are constants $\mu > 0$ and $K \geq 0$ such that for all $r \geq 0$, if α is a path in $X \setminus N^0(p, r)$ connecting $x, y \in S(x, r)$, then* length $\alpha \geq e^{\mu d(x,y)} - K$.

See Figure 6h.

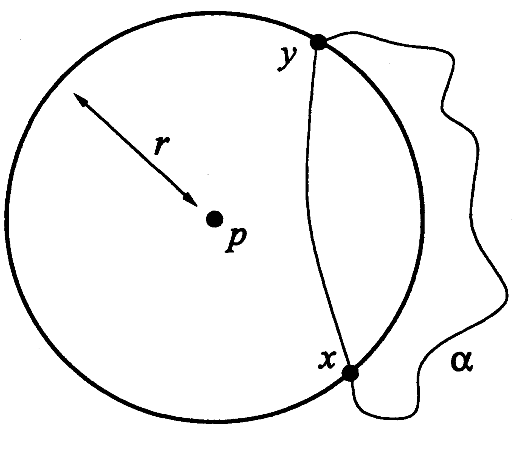

Figure 6h.

The idea behind the proof is that projections from a sphere to a smaller concentric sphere will tend to reduce distances by a uniform factor less than 1. Because of the additive constants involved in the definition of hyperbolicity, we will need to express the argument as a discrete process.

Our proof will be to use the following two related observations.

Lemma 6.15 : *For all sufficiently large h in relation to the hyperbolicity constant, if $x, y \in X$ with $d(x, y) \leq h/2$, then $d(x', y') \leq h$ where $x' \in [p, x]$ and $y' \in [p, y]$ with $d(p, x') = d(p, y')$.*

Lemma 6.16 : *For all sufficiently large h in relation to the hyperbolicity constant, if $x, y \in X$ with $d(p, x) = d(p, y)$ and $d(x, y) \leq 2h$, then $d(x', y') \leq h$, where $x' \in [p, x]$ and $y' \in [p, y]$ with $d(x, x') = d(y, y') = h$.*

We leave the proofs as an exercise — for example, either verify the statements in a spanning tree for the five points p, x, y, x', y', or else by a more direct argument by considering the triangle with sides $[p, x], [p, y], [x, y]$.

Proof of Proposition 6.14 : Let $l = \text{length}\, \alpha$. We can assume that $l \geq 4h$. We can thus find some $n \in \mathbf{N}$ so that $2^{m-2}h \leq l \leq 2^{m-1}h$. Let $x = x_0, x_1, \ldots, x_{2^m} = y$ be a sequence of $2^m + 1$ points along α so that $d(x_i, x_{i+1}) \leq h/2$ for all i. Let $y_i \in [p, x_i]$ with $d(p, y_i) = r$. Thus $y_0^0 = x$, $y_{2^m}^0 = y$ and $y_i^0 \in S(r)$ for all i. By Lemma 6.15, $d(y_i^0, y_{i+1}^0) \leq h$ for all i. (Figure 6i, where $m = 3$.)

Now define a sequence $y_0^1, \ldots, y_{2^{m-1}}^1$ as follows. If $r \leq h$ set $y_i^1 = p$ for all i. If $r \geq h$, let $y_i^1 \in [p, y_{2i}^0]$ be the point with $d(y_i^1, y_{2i}^0) = h$. Now $d(y_{2i}^0, y_{2i+2}^0) \leq 2h$, and so by Lemma 6.16, we have $d(y_i^1, y_{i+1}^1) \leq h$. We now proceed inductively, each time eliminating half the points, and moving the others a distance h towards p (or possibly setting them all equal to p). For each $j = 1, 2, \ldots, m$, we get a sequence $y_0^j, \ldots, y_{2^{m-j}}^j$, with $d(y_i^j, y_{i+1}^j) \leq h$ for all i. We end up with a 2-point sequence, y_0^m, y_1^m. Note that $d(y_0^m, y_1^m) \leq h$. Now for all j, $d(y_0^j, y_0^{j+1}) \leq h$ and so $d(y_0^0, y_0^m) \leq mh$. Similarly, $d(y_{2^m}^0, y_1^m) \leq mh$. But $x = y_0^0$ and $y = y_{2^m}^0$, and so $d(x, y) \leq 2mh + h = (2m + 1)h$, and so $m \geq (d(x, y) - h)/2h$.

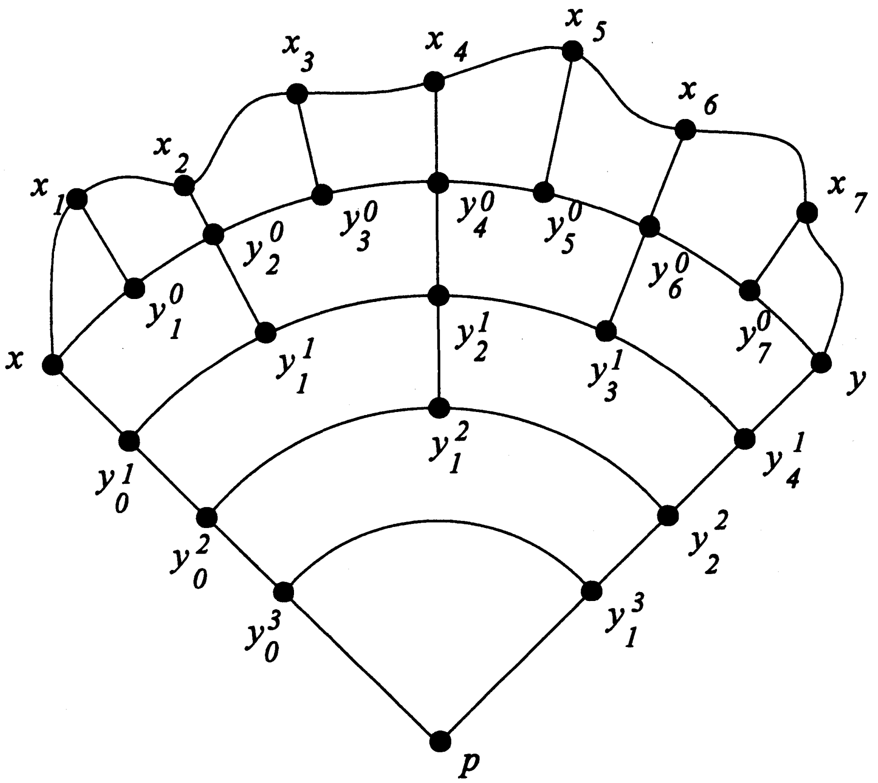

Figure 6i.

We see that $l \geq 2^{m-2}h \geq 2^{(d(x,y)-h)/2h}h/4$. This is under the initial assumption that $l \geq 4h$. Thus, in general, we always get an inequality of the form $l \geq e^{\mu d(x,y)} - K$, where μ and K depend only on h, and hence only on the hyperbolicity constant k, as required. $\diamondsuit$

Remark : It turns out that the exponential growth of distances gives another formulation of hyperbolicity — essentially taking the conclusion of Proposition 6.14 as a hypothesis. We will not give a precise formulation of this here.

6.7. Quasigeodesics.

The notion of a quasigeodsesic path is another fundamental notion in geometric group theory. The following definition will make sense in any metric space, though it is mainly of interest in geodesic spaces. In what follows we shall abuse notion slightly and identify a path in X with its image as a subset of X, (even if the path is not injective). Given two point, x, y in a path α, we shall write $\alpha[x, y]$ for the segment of α beween x and y.

Definition : A path, β, is a (λ, h)-*quasigeodesic*, with respect to constants $\lambda \geq 1$ and $h \geq 0$, if for all $x, y \in \beta$, $\text{length}(\beta[x, y]) \leq \lambda d(x, y) + h$. A *quasigeodesic* is a path that is (λ, h)-quasigeodesic for some λ and h.

In other words, it takes the shortest route to within certain linear bounds. Note that a $(1, h)$-quasigeodesic is the same as an h-taut path.

We now suppose that (X, d) is k-hyperbolic again. A basic fact about quasigeodesiscs is that they remain a bounded distance apart (cf. Lemma 6.4).

Proposition 6.17 : *Suppose that α is a geodesic, and β is a (λ, h)-quasigeodesic with the same endpoints. Then $\beta \subseteq N(\alpha, r)$ and $\alpha \subseteq N(\beta, r)$ where r depends only on λ, h, and the hyperbolicity constant k.*

Proof : We first show that α lies a bounded distance from β. (In other words, we proceed in the opposite order from Lemma 6.4.) Let a, b be the endpoints of α.

Choose $p \in \alpha$ so as to maximise $d(p, \beta) = t$, say. Let $a_0, a_1 \in [a, p]$ be points with $d(p, a_0) = t$ and $d(p, a_1) = 2t$. The point a_0 certainly exists, since $d(p, a) \geq t$. If $d(p, a) < 2t$, we set $a_1 = a$ instead. Now $d(a_1, \beta) \leq t$, and so there is some point $a_2 \in \beta$ with $d(a_1, a_2) \leq t$. If $a_1 = a$, we set $a_2 = a$. We similarly define points b_0, b_1, b_2 (Figure 6j).

Note that $d(a_2, b_2) \leq 6t$. Let $\delta = \beta[a_2, b_2]$, and let $\gamma = [a_0, a_1] \cup [a_1, a_2] \cup \delta \cup [b_2, b_1] \cup [b_1, b_0]$. Note that $\gamma \cap N^0(p, t) = \emptyset$. Since β is

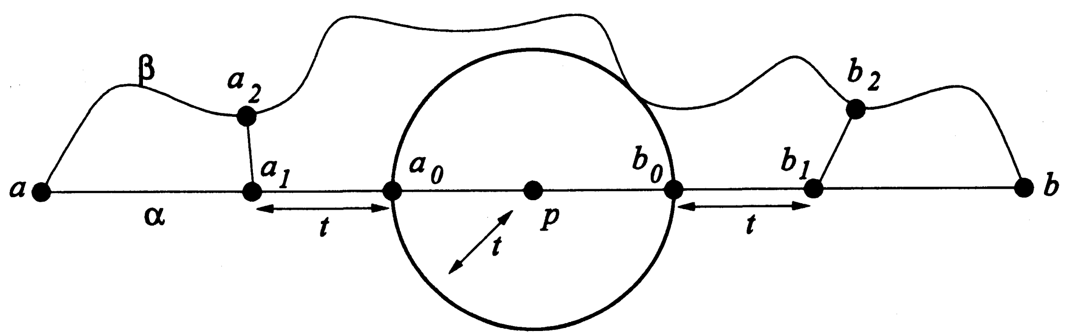

Figure 6j.

quasigeodesic,

$$\begin{aligned}\operatorname{length}\delta &\leq \lambda d(a_2, b_2) + h\\ &\leq 6\lambda t + h,\end{aligned}$$

and so

$$\begin{aligned}\operatorname{length}\gamma &\leq 4t + \operatorname{length}\delta\\ &\leq (6\lambda + 4)t + h.\end{aligned}$$

On the other hand, $d(a_0, b_0) = 2t$, and γ does not meet $N^0(p,t)$. Thus applying Proposition 6.14, we get

$$\operatorname{length}\gamma \geq e^{\mu(2t)} - K.$$

Putting these together we get

$$e^{2\mu t} \leq (6\lambda + 4)t + h + K$$

which places an upper bound of t in terms of λ, h, μ, K, and hence in terms of λ, h and k.

To show that β lies in a bounded neighbourhood of α, one can now use a connectedness argument similar to that use in Lemma 6.4 (with the roles of α and β interchanged). ◇

Of course (after doubling the constant r) Propositon 6.17 applies equally well to two quasigeodesics, α and β with the same endpoints.

Using Proposition 6.17, we see that we can formulate hyperbolicity equally well using quasigeodesic triangles, that is where α, β, γ

are assumed quasigeodesic with fixed constants. In particular, we note:

Lemma 6.18 : *Any (λ, k)-quasigeodesic triangle (α, β, γ) has a t-centre, where t depends only on λ, h and k.*

Proof : Let $(\alpha', \beta', \gamma')$ be a geodesic triangle with the same vertices. Applying Proposition 6.17, we see that any k-centre of $(\alpha', \beta', \gamma')$ will be a $(k + r)$-centre for (α, β, γ). ◊

6.8. Hausdorff distances.

Before continuing we make the following useful definition.

Definition : Suppose P, Q are subsets of a metric space (X, d). We define the *Hausdorff distance* between P and Q as the infimum of those $r \in [0, \infty]$ for which $P \subseteq N(Q, r)$ and $Q \subseteq N(P, r)$.

Exercise: This is a pseudometric on the set of all bounded subsets of X. (It is only a pseudometric, since the Hausdorff distance between a set and its closure is 0.) Restricted to the set of closed subsets of X, this is a metric.

Note that Proposition 6.17 implies that the Hausdorff distance between two quasigeodesics with the same endpoints is bounded in terms of the quasigeodesic and hyperbolicity constants.

6.9. Quasi-isometry invariance of hyperbolicity.

This is the key fact that makes the theory of hyperbolic groups work.

Suppose that (X, d) and (X', d') are geodesic spaces and that $\phi : X \longrightarrow X'$ is a quasi-isometry. We would like to say that the image of a geodesic is a quasi-geodesic, but this is complicated by the fact that quasi-isometries are not assumed continuous. The following technical discussion is designed to get around that point.

Fix some $h > 0$. Suppose that α is a geodesic in X from x to y. Choose points $x = x_0, x_1, \ldots, x_n = y$ along α so that $d(x_i, x_{i+1}) \leq h$ and $n \leq l/h \leq n+1$. Let $y_i = \phi(x_i) \in X'$. Let $\bar{\alpha} = [y_0, y_1] \cup [y_1, y_2] \cup \cdots \cup [y_{n-1}, y_n]$.

Exercise: If α is a geodesic in X and $\bar{\alpha}$ constructed as above, then $\bar{\alpha}$ is quasigeodesic, and the Hausdorff distance between $\bar{\alpha}$ and $\phi(\alpha)$ is bounded. As usual, the statement is *uniform* in the sense that the constants of the conclusion depend only on those of the hypotheses and our choice of h.

We are free to choose h however we wish, though it may be natural to choose it in relation to the other constants of a given argument, such as the constant of hyperbolicity.

Theorem 6.19 : *Suppose that X and X' are geodesic spaces with $X \sim X'$, then X is hyperbolic if and only if X' is.*

Proof : Let $\phi : (X, d) \longrightarrow (X', d')$ be a quasi-isometry and suppose that X' is k-hyperbolic. Let (α, β, γ) be a geodesic triangle in X. Let $\bar{\alpha}, \bar{\beta}, \bar{\gamma}$ be the quasigeodesics a bounded distance from $\phi(\alpha), \phi(\beta), \phi(\gamma)$ as constructed above. By Lemma 6.18, $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ has a t-centre, q, where t depends only on k and the quasi-geodesics constants. Since $\phi(X)$ is cobounded, there is some $p \in X$ with $\phi(p)$ a bounded distance from q. Now $\phi(p)$ is a bounded distance from each of $\phi(\alpha)$, $\phi(\beta)$ and $\phi(\gamma)$. It now follows that p is a bounded distance from each of α, β, γ. In other words p is a centre for the triangle (α, β, γ). ◊

In fact, we see that the hyperbolicity constant of X depends only on that of X' and the quasi-isometry constants. (In the construction of $\bar{\alpha}, \bar{\beta}, \bar{\gamma}$ it is natural to take $h = k$. In this way, we get linear bounds between the hyperbolicity contants.)

Theorem 6.19 has some immediate consequences. For example we see:

(1) If $m, n \geq 2$, then $\mathbf{R}^m \not\sim \mathbf{H}^n$.

(2) If $n \geq 2$, then $\mathbf{R}^n$ is not quasi-isometric to any tree.

In particular, we get another proof that $\mathbf{R}^2 \not\sim \mathbf{R}$ and that $\mathbf{R}^2 \not\sim [0, \infty)$.

6.10. Hyperbolic groups.

We are finally ready for the following fundamental notion:

Definition : A group Γ is *hyperbolic* if it is finitely generated and its Cayley graph $\Delta(\Gamma)$ is hyperbolic.

By Theorem 3.3 and Theorem 6.19 this is well defined — it doesn't matter which finite generating set we take to construct the Cayley graph.

We note:

Lemma 6.20 : *Suppose that Γ acts properly discontinously cocompactly on a proper hyperbolic (geodesic) space, then Γ is hyperbolic.*

Proof : By Theorem 3.5, Theorem 3.6 and Theorem 6.19. $\diamond$

Examples:

(1) Any finite group.

(2) Any virtually free group.

(3) The fundamental group of any compact hyperbolic manifold. Note that if $\Gamma = \pi_1(M)$, where M is compact hyperbolic, then Γ acts properly discontinously cocompactly on $\mathbf{H}^n$.

(4) In particular, if Σ is any compact (orientable) surface of genus at least 2, then $\pi_1(\Sigma)$ is hyperbolic.

Non-examples:

(1) $\mathbf{Z}^n$ for any $n \geq 2$.

(2) It turns out that a hyperbolic group cannot contain any $\mathbf{Z}^2$ subgroup, so this fact provides many more non-examples. For example,

many matrix groups $SL(n, \mathbf{Z})$ etc., knot groups (fundamental groups of knot complements), mapping class groups, braid groups etc. This is not the only obstruction, however.

6.11. Some properties of hyperbolic groups.

This is all we have time to deal with systematically in this course. We shall finish off by listing a few interesting directions currently being pursued. This list is by no means complete.

6.11.1. Subgroups.

(S1) Suppose g is an infinite order element of a hyperbolic group, Γ, so that $\langle g \rangle \cong \mathbf{Z}$. Then $\langle g \rangle$ is a "quasiconvex subgroup". Here this means that if $x \in \Delta(\Gamma)$, then the bi-infinite path $\alpha = \bigcup_{n \in \mathbf{Z}} [g^n x, g^{n+1} x]$ is quasi-geodesic. This is, in fact, the same as saying that the "stable length" $||g|| = \lim_{n \to \infty} \frac{1}{n} d(x, g^n x)$ is positive. (There are many examples of non-hyperbolic groups where this fails.)

(S2) Suppose that, in (S1), $h \in \Gamma$ is another element commuting with g. The bi-infinite path $h\alpha = \bigcup_{n \in \mathbf{Z}} [g^n hx, g^{n+1} hx]$ is a finite Hausdorff distance from α. In fact, using Proposition 6.17, one can show that the Hausdorff distance between α and $h\alpha$ is uniformly bounded, that is, it depends only on the quasi-geodesic constants of α, and not on h. But since $\Delta(\Gamma)$ is locally finite, there are only finitely many possibilities for $h\alpha$. As a result, one can show:

A hyperbolic group cannot contain any subgroup isomorphic to $\mathbf{Z}^2$.

(S3) A "Baumslag-Solitar group" is a group of the form $B(m, n) = \langle g, h \mid g^m h = h g^n \rangle$, where $m, n \geq 1$. (Note that $B(1, 1) = \mathbf{Z}^2$.) By a similar argument to (S2), one can show, in fact, that:

A hyperbolic group cannot contain any Baumslag-Solitar subgroup.

(Indeed if $(m, n) \neq (1, 1)$, this can also be seen from (S1) since, in that case, the stable length of h in $B(m, n)$ can be shown to be 0.)

(S4) Any hyperbolic group Γ that is not finite or virtually cyclic contains a free subgroup of rank 2, and hence free subgroups of any countable rank. (In particular, Γ has "exponential growth".) The usual way to construct such subgroups uses the so called "small cancellation theory", which long predates the invention of hyperbolic groups (see (F5) below).

(S5) It is a non-trivial question as to which groups can be embedded in hyperbolic groups. In (S4) we saw examples of non–f.g. subgroups of hyperbolic groups: free groups of infinite rank. One can construct hyperbolic groups that contain f.g. subgroups that are not f.p. One can also construct f.p. subgroups which are not hyperbolic, though these constructions become increasingly complicated.

(S6) A hyperbolic group contains only finitely many conjugacy classes of finite subgroups. To see this one can argue as follows. Suppose G is a finite subgroup of the hyperbolic group, Γ. Let a be any vertex of the Cayley graph $\Delta(\Gamma)$. Let $r \in \mathbf{N}$ be minimal such that the orbit, Ga, is contained in $N(b,r)$ for some vertex b of $\Delta(\Gamma)$. Now let B be the set of all vertices b with $Ga \subseteq N(b,r)$. Thus B is G-invariant, and an exercise in hyperbolicity shows that it has diameter bounded in terms of the hyperbolicity constant. There are thus only finitely many possiblities for B up to the Γ-action on $\Delta(\Gamma)$, and it follows that there are only finitely many possibilities for G up to conjugacy.

(S7) It is an open question as to whether every hyperbolic group is virtually torsion-free.

6.11.2. Finiteness and computablility properties.

(F1) By hypothesis, a hyperbolic group is f.g. One can show that any hyperbolic group is f.p.

(F2) If Γ is any hyperbolic group, then one can construct a locally finite contractible simplicial complex, K, (the "Rips complex") such that Γ acts properly discontinuously cocompactly on K. This is a strong finiteness condition. For example, a group acts p.d.c. on a locally finite connected complex if and only if it is f.g. (Here we could take K to be a graph. The case of free p.d.c. actions was discussed in Section 2.) Moreover, a group acts freely p.d.c. on a locally finite simply-connnected complex if and only if it is f.p. Here we are assuming that K is contractible, which is equivalent to saying that the homototopy groups, $\pi_n(K)$, are trivial for all $n \in \mathbf{N}$.

(F3) Suppose that Γ is a group which acts properly discontinuously cocompactly on a finite dimensional locally-finite contractible complex, and that Γ has no Baumslag-Solitar subgroups. Is Γ hyperbolic? This seems to be an open question (due to Bestvina). In other words are the conditions (S3) and (F2) together the only obstructions to hyperbolicity?

(F4) The fact that Γ is finitely presented can be strengthenned in another direction to a "linear isoperimetric inequality". One can construct a p.d.c. action of Γ on a locally finite 2-dimensional simplicial complex K with the property that if α is a curve in the 1-skeleton of length n, then α bounds a disc in K (not nescessarily embedded) meeting at most $f(n)$ 2-simplices, where f is a linear function. One can give an equivalent algebraic statement. Fix any finite presentation of Γ. Suppose that w is a word in the generators and their inverses representing the identity in Γ. Then we can reduce w to the trivial word (of length 0) by repeatedly applying the relations. A linear isoperimetric inequality says that we only need to do this at most $f(n)$ times, where n is the length of w and $f : \mathbf{N} \longrightarrow \mathbf{N}$ is linear. It turns out that:

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

(In fact, a subquadratic isoperimetric inequality is sufficient.)

(F5) We can make the following additional remarks. It turns out that isoperimetic inequalities of this sort are q.i. invariants (thereby giving a different proof that hyperbolicity is q.i. invariant.) The group $\mathbf{Z}^2$ has a quadratic, but not a linear isoperimetric inquality. The Heisenberg group has a cubic inequality. Other groups have exponential inqualities (or worse).

One can show that, for a f.p. group, a (sub)computable isoperimetric inequality is equivalent to solvability of the word problem. (It puts a computable bound on the work we need to do to check whether or not a word can be reduced to the trival word.) This shows that a solvable word problem is q.i. invariant. It also shows that the word problem in a hyperbolic group is solvable. In fact (though this is not an immediate consequence) it can be solved in linear time. The linear time algorithm is the "Dehn algorithm". Back in the 1920s

Dehn used ideas of hyperbolic geometry to show that the word problem in a surface group is solvable. This was the beginning of "small cancellation theory" referred to in (S4). The same basic idea applies to general hyperbolic groups.

(F6) A lot more can be said in relation to computablity. For example, it turns out that a hyperbolic group is "automatic" in the sense of Thurston. This was essentially shown by Cannon, before the either of the notions "hyperbolic" or "automatic" were formally defined. Automaticity is a formal criterion which implies that many calculations can be carried out very efficiently. In particular, the word problem is solvable.

6.11.3. Boundaries.

(B1) Let X be a proper hyperbolic space. A "ray" in X is a semi-infinite geodesic. We say that two rays are "parallel" if the Hausdorff distance between them is finite. This is an equivalence relation, and we write ∂X for the set of equivalence classes. (We can equivalently use quasi-geodesic rays.) One can put a topology on ∂X — informally two rays are close in this topology if they remain close over a long distance in X. It turns out that ∂X is compact and metrisable. We refer to ∂X as the "(Gromov) boundary" of X. Any q.i. $\phi : X \longrightarrow Y$ induces a homeomorphism $\partial X \longrightarrow \partial Y$. Thus the homeomorphism type of ∂X is a quasi-isometry invariant. In particular, it makes sense to talk about the boundary, $\partial\Gamma$, of a hyperbolic group Γ.

(B2) The boundary of a compact space is empty. Thus ∂(finite group) $= \emptyset$.

(B3) The real line has two boundary points — one for each end. The boundary of the group $\mathbf{Z}$, or any virtually cyclic group, is thus the two-point space.

(B4) If $p \geq 3$, then ∂T_p is a cantor set. Thus ∂F_n is a cantor set for all $n \geq 2$.

(B5) $\partial \mathbf{H}^n$ can be identified with the ideal boundary we have already defined — the boundary of the Poincaré disc. It is thus homeomorphic to S^{n-1}. It follows that:

If $\mathbf{H}^m \sim \mathbf{H}^n$ then $m = n$.

Note in particular that the fundamental group of a compact surface cannot be q.i. to the fundamental group of a hyperbolic 3-manifold.

(B6) The work of Tukia, Gabai, Casson and Jungreis referred to earlier shows that if the boundary of a hyperbolic group is homeomorphic to a circle, then the group is a virtual surface group. Cannon asked whether a hyperbolic group with boundary a 2-sphere is a virtual hyperbolic 3-manifold group. This question remains open. The analogous assertion certainly fails in higher dimensions. For example the fundamental group of a compact complex hyperbolic 4-manifold is hyperbolic and has boundary a 3-sphere, but it does not admit any p.d.c. action on (real) hyperbolic 4-space.

(B7) A hyperbolic group Γ acts by homeomorphism on the boundary $\partial\Gamma$. In fact this has a particular dynamical property: it is a "uniform convergence group". This means that induced action on the space of distinct triples (the configuration space of 3-element subsets of the boundary) is p.d.c. The notion of a convergence group was introduced by Gehring and Martin and explored by a number of people, such as Tukia. It turns out (Bowditch) that one can characterise hyperbolic groups in these terms:

If a group acts as a uniform convergence group on a compact metrisable space with no isolated points, then the group is hyperbolic, and the space is equivariantly homeomoprhic to the boundary.

(B8) It turns out that any compact metrisable topological space is homeomorphic to the boundary of some proper hyperbolic space. However, there are constraints on what kinds of spaces can arise as boundaries of hyperbolic groups. In some sense, the "generic boundary" is a Menger curve, but there are many other examples. One can get a lot of information about the algebraic structure of a hyperbolic group from the topology of its boundary.

6.11.4. Other directions.

There are many other directions in the study of hyperbolic groups which I have not had time to mention. The "JSJ splitting" theory of Sela, for example, which has inspired similar results for much wider classes of groups. The Markov property of the boundary, the "geodesic flow" of a hyperbolic group, bounded co-homology, the Novikov conjecture etc.

The subject of "relatively hyperbolic" groups is very fashionable at the moment. These include fundamental groups of finite-volume hyperbolic manifolds, amalgamated free products of groups over finite groups, the "limit groups" defined by Sela in relation to the Tarski problem etc.

There are many naturally arising spaces that are hyperbolic that do not stem directly from groups. For example the Harvey curve complex associated to a surface was shown to be hyperbolic by Masur and Minsky. This has many implications, for example to the study of the mapping class groups.