1. Hyperbolic plane

1.1. Intro to $\mathbb{H}$. As a set, $\mathbb{H} := \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\} \subset \hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$, with boundary $\hat{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$.

Recall that **Möbius transformations** are maps $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ of the form $f(z) = \frac{az+b}{cz+d}$ for $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. They are extended to ∞ by continuity (so $f(\infty) = a/c$). This is an action of $GL_2(\mathbb{C})$ by *conformal* maps. One checks that the subgroup $SL_2(\mathbb{R})$ preserves $\mathbb{H}$ by checking that $\mathrm{Im}(z) > 0 \implies \mathrm{Im}(\frac{az+b}{cz+d}) > 0$. Indeed, $\pm A$ act the same, so one might prefer to regard the action as by $PSL_2(\mathbb{R}) := SL_2(\mathbb{R})/\{\pm I\}$. We will say that $PSL_2(\mathbb{R})$ acts by **fractional linear transformations** (FLTs).

Note that every FLT is a composition of affine maps and circle inversion. In particular, for the general f above, we can take $g(z) = c^2 z + cd$, $J(z) = -1/z$, $h(z) = (ad - bc)z + \frac{a}{c}$, and we get $f = h \circ J \circ g$.

The **hyperbolic metric** is often introduced by its distance element $ds = \frac{dz}{y} = \frac{\sqrt{dx^2+dy^2}}{y}$ (then take the length metric: distance is infimal length of a path between points).

Sample calculation: We know that the imaginary axis is a geodesic from any pi to qi with parametrization $\gamma(t) = e^t i$: $\int_\gamma ds = \int_\gamma \frac{|dz|}{y} = \int_p^q \frac{|dy|}{y} = \ln(|q/p|)$, and for any path we have the inequality $|dz| \geqslant |dy|$, so this is shortest-possible.

One can check that length is invariant under FLTs: $\ell(\gamma) = \ell(T \circ \gamma)$ for all $T \in PSL_2(\mathbb{R})$, and do a bit more work to see that indeed $\mathrm{Isom}^+(\mathbb{H}) = PSL_2(\mathbb{R})$.

1.2. Cross-ratio setup. HOWEVER! That was kind of the wrong viewpoint. Instead, we should take the FLT action as essential and build a metric compatible with it. As Möbius transformations, FLTs are not only conformal (preserving circle fields in the tangent space) but actually preserve circles in $\hat{\mathbb{C}}$ (which look like lines and circles in $\mathbb{C}$). Suddenly it is easy to see why FLTs preserve $\mathbb{H}$: since FLTs have real coefficients, $\hat{\mathbb{R}}$ is preserved, so we only need to check check one more point. And being conformal, the action preserves angles. Thus the images of the imaginary axis under FLTs are all **orthocircles**, defined as circles perpendicular to the real axis (some of which look like vertical lines).

FLTs have a complete invariant for their action on $\hat{\mathbb{R}}$ called the **cross-ratio**:

$$[p, q, r, s] := \frac{(r - p)(s - q)}{(q - p)(s - r)}.$$

Cross-ratio extends from $(\hat{\mathbb{R}})^4 \backslash \Delta$ to $(\hat{\mathbb{C}})^4 \backslash \Delta$, where Δ is the fat-diagonal subset with any three points agreeing. To check that cross-ratio is invariant under FLTs, it suffices to check invariance under affine maps and under inversion, both easy.

Then a quite natural definition is to take $d(P, Q) = \ln \left| [\bar{P}, P, Q, \bar{Q}] \right|$, where $\bar{P}, \bar{Q}$ are the endpoints on $\hat{\mathbb{R}}$ of the unique orthocircle through P, Q. Note that for P, Q on the imaginary axis, we recover the same distance as for the previous definition. (Why a metric? Symmetry is immediate. Positive definiteness follows from positivity on the imaginary axis and invariance under FLTs. It's more work to check the triangle inequality, but not too bad.) It's easy to see that the imaginary axis is geodesic and indeed that it's the unique geodesic between any two of its points. Thus our newly redefined hyperbolic metric is uniquely geodesic with orthocircles as its geodesics. And incidentally now it is immediate that FLTs are precisely the isometries.

1.3. Hilbert geometries. There's a completely different construction using cross-ratios on $\hat{\mathbb{R}}$ directly. Hilbert geometry on a convex body: $d(p, q) = \ln[\bar{p}, p, q, \bar{q}]$, where now the line through the points is given a Euclidean numbering with an *arbitrary* 0 and 1. Now the $SL_2(\mathbb{R})$ action by LINEAR transformations sends (Ω, d_Ω) to an isometric $(\Omega', d_{\Omega'})$. Note that for the same reasons as before, straight lines are geodesic. However now there are sometimes alternate geodesics.

Beautiful facts: $(\bigcirc, d_\bigcirc)$ is isometric to $\mathbb{H}$, called the **Klein model**. (Careful! Don't confuse with the Poincaré **disk model**, which is the image of $\mathbb{H}$ under a Möbius transformation sending $\hat{\mathbb{R}}$ to the unit circle.) On the other hand, if Ω is a triangle, the space $(\triangle, d_\triangle)$ is isometric to a Banach space structure on $\mathbb{R}^2$. Furthermore (up to affine maps), $\bigcirc$ is the only Riemannian metric among Hilbert geometries and $\triangle$ is the only Banach space! We'll come back to these examples.

1.4. **Classification of isometries.** Trace classifies $PSL_2(\mathbb{R})$ elements into three types: **hyperbolic** ($|\mathrm{tr}| > 2$), **parabolic** ($|\mathrm{tr}| = 2$), and **elliptic** ($|\mathrm{tr}| < 2$). By basic linear algebra, these are conjugate into the subgroups

$$A = \{\begin{pmatrix} k & 0 \\ 0 & 1/k \end{pmatrix}\}; \quad N = \{\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}\}; \quad K = \{\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\},$$

respectively, and indeed the **Iwasawa decomposition** $PSL_2(\mathbb{R}) = KAN$ is the statement that any element of $PSL_2(\mathbb{R})$ can be expressed (in fact, uniquely) as a product of matrices from these three subgroups.

Note that each of these is a 1–parameter subgroup, so you can consider the orbits of points in $\bar{\mathbb{H}}$ as the parameter is varied. This is how those orbits look in the Poincaré (conformal) disk model:

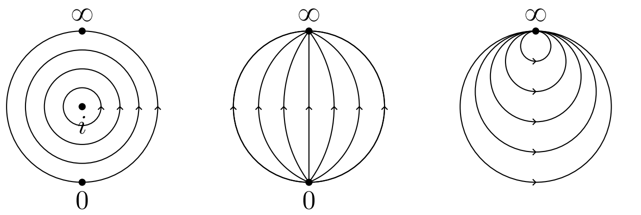


FIGURE 1. Actions on the unit disk by elements of K, A, and N respectively.

Thus this three-way classification corresponds to the fixed-point structure: hyperbolic isometries have two fixed points on the boundary; parabolics have one fixed point on the boundary; and elliptics have one fixed point on the interior of $\mathbb{H}$.

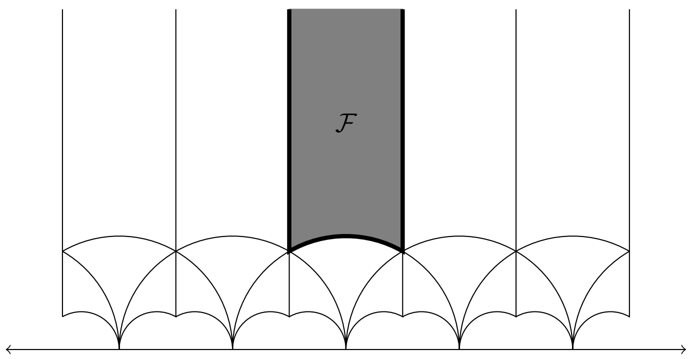
1.5. $SL_2(\mathbb{Z})$ **and Farey graph. Fuchsian groups** are discrete subgroups of $PSL_2(\mathbb{R})$ and the most basic example is $PSL_2(\mathbb{Z})$, sometimes called the **modular group**. If $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is translation and $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is inversion in the unit circle, then $PSL_2(\mathbb{Z}) = \langle T, J \rangle$. You will prove this later in the exercises.

The **Farey graph** is the $PSL_2(\mathbb{Z})$ images of the imaginary axis, and so contains all of $\hat{\mathbb{Q}}$ as its points at infinity. It is not hard to see that p/q and r/s are connected by an edge iff $\det\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \pm 1$, so it is in some sense a geometric recording of the matrix group. Note that for two such fractions, their **mediant** $\frac{p+r}{q+s}$ is adjacent to each.

Finally, let's build a fundamental domain for $PSL_2(\mathbb{Z})$ acting on $\mathbb{H}$: we want a compact tile whose Γ-images cover the plane and for which two tiles can overlap only along their boundary. We can use a **Dirichlet domain**

$$D_p(\Gamma) := \{z \in \mathbb{H} : d(z, p) \leqslant d(z, \gamma(p)) \ \forall \gamma \in \Gamma\}.$$

This is the polygon obtained by intersecting all half-spaces of points closer to p than each $\gamma(p)$. (Note: this construction provides a fundamental domain whenever Γ acts discontinuously and p has a trivial stabilizer.) One sees that $D_{2i}(\Gamma)$ is the triangle with vertices at $\pm\frac{1}{2} + \frac{\sqrt{3}}{2}i$ and ∞. This fundamental domain $\mathcal{F}$ comes up all the time, and its quotient $\mathcal{F}/\Gamma$ is called the **modular surface**.



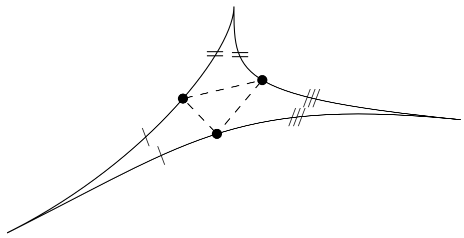
2. TREES, THIN TRIANGLES, INSIZE, AND METRIC TREEISHNESS

2.1. **Trees.** (Simplicial) **trees** are graphs with no cycles, endowed by default with the length metric that gives edges length one. $\mathbb{R}$–trees are generalizations obtained by rescaling: a metric space is an $\mathbb{R}$–tree if for any two points there exists an isometrically embedded $[0, d(p, q)]$ connecting them, and this is the only non-backtracking path between the points. The most general triangle in a tree is a tripod, so there is always (at least) one vertex common to all three sides of a triangle. Furthermore, each side of the triangle is completely contained in the union of the other two. Also, two rays from a common basepoint agree up to a certain point, then diverge completely in the sense that the concatenation of the two rays after the split is a bi-infinite geodesic. Notice that this is an extreme, cartoon version of the behavior of geodesics on a saddle surface (negative curvature).

2.2. **Thin triangles.** The most commonly seen definition of δ–hyperbolicity is that **triangles are thin**. That is, if you have a geodesic triangle (three points, pairwise connected by geodesic segments α, β, γ), it is said to be δ–thin if $\alpha \subset \mathcal{N}_\delta(\beta \cup \gamma)$ and the other two similar inclusions hold.

We have already seen that trees are 0–hyperbolic in this definition.

2.3. **Insize.** Define the **inpoints** of a triangle to be the uniquely determined three points (not necessarily distinct) that split the sides into pairs of equal lengths as in the figure. These must exist (because the triangle inequalities ensure that $a = r + s$, $b = s + t$, $c = r + t$ has a nonnegative solution).



Then the **insize** of a triangle is the diameter of the set of inpoints. Then a space is δ–hyperbolic by the insize definition if there is a global bound δ on the insizes of triangles.

In a tree, the three inpoints coincide in the focus point of the tripod, so triangles are 0–hyperbolic in the insize definition as well.

2.4. **Treeishness.** For any metric space X, define

$$K_n(X) := \left\{ \Big(d(x_i, x_j) \Big)_{ij} : (x_1, \ldots, x_n) \in X^n \right\} \subset M_{n\times n}(\mathbb{R}).$$

Let $\mathsf{Tri}_n \subset M_{n\times n}(\mathbb{R})$ be all the matrices corresponding to tuples that three-wise satisfy the triangle inequality, and let $\mathsf{Tree}_n \subset M_{n\times n}(\mathbb{R})$ be the union of $K_n(T)$ over all trees T.

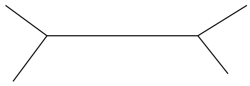
Since triangles behave so differently in different curvature regimes, you might think that K_3 is enough to pick out curvature. But in fact it is easy to see that $K_3(\mathbb{E}^2) = K_3(\mathbb{H}) = K_3(T)$, where T is any tree containing an infinite tripod. In fact, it's worse than that: Misha Kapovich recently proved that almost any unbounded length space has $K_3(X) = \mathsf{Tri}_3$. (The exceptions are coarsely equal to points, rays, and lines!) I'll give a precise statement of this theorem after we have defined quasi-isometry.

So three points tell you nothing (i.e., the sidelengths of a triangle do not detect whether it's thin). However, four points suffice! An alternate definition of δ–hyperbolicity is that X is δ–hyperbolic if $K_4(X) \subset \mathcal{N}_\delta(\mathsf{Tree}_4)$. (You are free to choose your favorite reasonable metric on $M_{4\times 4}$, such as the one given by the norm of a matrix being the sup of the entries.)

The next two definitions will illustrate why four points succeed where three points failed: K_4 suffices to pick out treeish behavior.

2.5. **Pairsums.** For four points x_1, x_2, x_3, x_4, a *pairsum* is $d(x_i, x_j) + d(x_k, x_l)$, where $\{i, j, k, l\} = \{1, 2, 3, 4\}$, so that there are exactly three differently-indexed pairsums for a four-tuple of points. The pairsum (or four-point) definition of δ–hyperbolicity is that for all four-tuples of points in the metric space, the two largest pairsums differ by no more than δ.

It's easy to see that trees are 0–hyperbolic in this definition as well, because the most nondegenerate configuration of four points is



and the two largest pairsums are equal (the sum of all edge-lengths, with the "crossbar" counted twice).

This definition has the enormous advantage that it is applicable in spaces without geodesics. For instance, it is an exercise to show that ultrametric spaces are hyperbolic, which includes the field of p–adic numbers $\mathbb{Q}_p$, a totally disconnected metric space.

2.6. **Gromov product.** There's another way to set up a four-point definition: define the **Gromov product** to be

$$(x \cdot y)_w := \frac{1}{2}\big(d(x,w) + d(y,w) - d(x,y)\big).$$

In a tree, this is easily seen to be the distance from w to the geodesic $\overline{xy}$. More generally, it is proportional to the failure of the triangle inequality to be an equality: if the side lengths of a triangle are called a, b, c, then the triangle inequality ensures that $a + b \geqslant c$, and its **defect** is the value $\Delta = a + b - c \geqslant 0$. The Gromov product is just half of this defect.

In a tree, consider a triangle with vertices x, y, z and an observation point w, and compare the Gromov products $(x \cdot y)_w$, $(y \cdot z)_w$, and $(x \cdot z)_w$. There must be a unique path from w to the tree, intersecting it at some unique point and creating the configuration seen in the figure above. Since that point is on (at least) two of the tree sides of the tripod, there is guaranteed to be a tie for the smallest Gromov product, so we have $(x \cdot y)_w \geqslant \min\big((y \cdot z)_w,\ (x \cdot z)_w\big)$. Loosening this by an additive factor of δ as usual, we define a metric space to be δ–hyperbolic if all four-tuples satisfy

$$(x \cdot y)_w \geqslant \min\big((y \cdot z)_w,\ (x \cdot z)_w\big) - \delta.$$

2.7. **Changing definitions.** We will call a space simply *hyperbolic* if it is δ–hyperbolic in any of the above definitions for any $\delta > 0$. However, note that there is no distinguished value of δ other than zero: the property of being hyperbolic is well-defined, but the value of δ can change between definitions.

2.8. **Examples.** $\mathbb{E}^2$ is of course not hyperbolic because it has fat triangles. In fact appropriate "dilations" produce obstructions to hyperbolicity. Note however that a Euclidean strip *is* hyperbolic.

We can put a 0–hyperbolic metric on $\mathbb{R}^2$ by making it an $\mathbb{R}$–tree. (Indeed it is easy to convince yourself of the so-called Connecting the Dots Lemma: a space is 0–hyperbolic iff it isometrically embeds in an $\mathbb{R}$–tree.) Examples are the SNCF metric or the Broadway metric, but not the taxicab (L^1) metric.

Crucially, the hyperbolic plane is δ–hyperbolic. (You'll compute δ values in the exercises.) Indeed, any manifold with sectional curvature bounded away from zero ($K \leqslant \rho < 0$) is δ–hyperbolic for an appropriately large δ. If it is not pinched, however, negative curvature does *not* suffice.

If the metric on $\mathbb{H}$ is rescaled to make the curvature more negative, the best value of δ shrinks. Indeed: the "asymptotic cone" (rescaling limit) of $\mathbb{H}$ is an (uncountably branching) $\mathbb{R}$–tree, and this is true for any δ–hyperbolic space. Cool theorem due to Kapovich-Kleiner: for finitely presented groups, they are hyperbolic iff some asymptotic cone is a tree.

Note that gluing in spheres (or any other sets) of bounded diameter does not ruin hyperbolicity. It is a *large-scale* negative curvature property.

It is a theorem of Benoist that Hilbert metrics on convex bodies are hyperbolic if and only if the boundary is sufficiently smooth (for instance, real-analytic will do; if there is a nice group action, then C^1 will do). (Theorem of Colbois-Vernicos-Verovic: this happens iff there is a bound on the areas of ideal triangles. Recall that for $\mathbb{H}$ itself, all ideal triangles have area π.)

In low-dimensional topology, one often builds graphs or complexes to record the combinatorics, such as the curve complex (or curve graph) which records intersections of curves on a surface. The curve complex of a surface is 17–hyperbolic, and now there are also δ–hyperbolic complexes associated with the outer automorphism group of a free group. This is useful because just acting nicely on a hyperbolic space gives useful information about a group.

3. QUASI-ISOMETRY, QUASI-GEODESICS

3.1. **Quasi-isometry.** For me a (K, C) quasi-isometric embedding ($K \geqslant 1, C \geqslant 0$) is a map of metric spaces such that if d_1, d_2 are the distances between a pair of points before and after applying the map, then

$$\frac{1}{K}d_1 - C \leqslant d_2 \leqslant Kd_1 + C.$$

Two spaces are quasi-isometric if there's some (K, C)–QI embedding from one to the other whose image is C–dense. (It's easy to see that for such a map, there is a *quasi-inverse* given by taking each point $y \in Y$ to any preimage of y', where y' is the nearest point in $f(X)$. This and a similar transitivity argument explain why $\underset{QI}{\sim}$ is an equivalence relation.)

Note that a biLipschitz map would have just a multiplicative bound, so this should be thought of as a coarse notion of biLipschitz.

Basic examples: $\mathbb{Z} \overset{(1,\frac{1}{2})}{\underset{QI}{\sim}} \mathbb{R}$, $\quad \mathbb{Z}^2 \overset{(\sqrt{2},\frac{\sqrt{2}}{2})}{\underset{QI}{\sim}} \mathbb{R}^2$.

Less standard definition (Druţu-Kapovich): $X \underset{QI}{\sim} Y$ iff there are separated nets A, B in X, Y respectively such that (A, d_X) is biLipschitz to (B, d_Y). This works because when there is a minimum positive distance, the additive constant in a QI can be absorbed by enlarging the multiplicative constant.

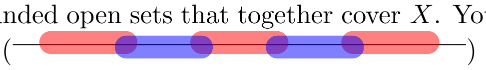
3.2. **QI invariants.**

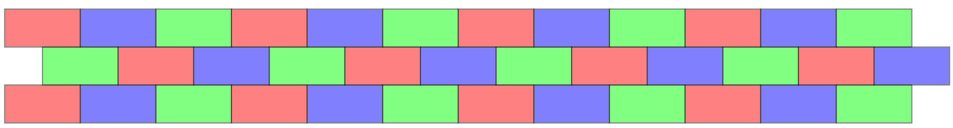
Proposition 1. Hyperbolicity is QI invariant *among geodesic spaces.*

That is, if $X \underset{QI}{\sim} Y$ and X is δ–hyperbolic, then Y is δ'–hyperbolic, where δ' depends on δ, K, and C.

To see what can go wrong in a general metric space, consider the pairsums definition of hyperbolicity. In X, take four points with pairsums $a + b$, $c + d$, and $e + f$, where $P = a + b \geqslant Q = c + d$ are largest. In Y, $P' = a' + b'$ is at most $(Ka + C) + (Kb + C) = KP + 2C$, while similarly $Q' \geqslant \frac{1}{K}Q - 2C$. These can get very different! Knowing that $P - Q \leqslant \delta$ will not help give a uniform bound on $P' - Q'$, which can be about as big as $KP - \frac{1}{K}Q$. Of course, this isn't a proof: a concrete example due to Väisälä is given below. We'll postpone the proof of the proposition until we have the Morse Lemma as a tool.

Other QI invariants:

- Being finitely generated / being finitely presented. In fact, being F_n (having a $K(G, 1)$ with finite n–skeleton) is QI invariant, and these are the $n = 1$, $n = 2$ cases. (See Geoghegan's book for details.) These are examples of *finiteness properties* of groups, which include other QI invariants as well.
- Number of ends: loosely, the number of connected components of $X \backslash B_r$ for large r. Clearly graphs can have any number of ends. However Hopf showed that groups can only have 0, 1, 2, or infinitely many ends. Stallings proved a hugely influential theorem about ends that became a prototype for how to relate coarse geometry to algebraic properties: a group has more than one end iff it splits as a free product with amalgamation or an HNN extension over finite group(s).
- Asymptotic dimension: $\text{asdim}(X) \leqslant n$ iff for any separation constant d there are $n+1$ families of d-separated, uniformly bounded open sets that together cover X. You can see that asdim $\mathbb{R} = 1$ by covering with intervals () and that asdim $\mathbb{R}^2 = 2$ by covering with bricks. (Fatten up the ones in the picture to make overlapping open sets.)



Gromov defined asymptotic dimension, and he noted that asdim $\mathbb{H}$=2, as you'll check in the exercises. Asymptotic dimension is known to be finite for: nilpotent groups (it equals the Hirsch length), hyperbolic groups, mapping class groups, right-angled Artin groups, Coxeter groups, and (strongly) relatively hyperbolic groups. It is infinite for the Grigorchuk group, and for the (finitely presented) Thompson's group F. Osin even gave a finitely presented and boundedly generated example with infinite asdim.

- Asymptotic cone (up to biLipschitz): this is the limit of the metric under a sequence of rescaling constants that go to zero. In full technical glory, you want to take an *ultralimit*. For many purposes, it suffices to consider a somewhat gentler *Gromov-Hausdorff limit*: $X_n \to X$ if for every R, the ball of radius R in X_n and the ball of radius R in X admit finer and finer nets of indexed points so that the pairwise distances converge as $n \to \infty$.

  So the asymptotic cone of $\mathbb{E}^d$ is $\mathbb{E}^d$, because $\mathbb{E}^d$ is invariant under rescaling of the metric. The asymptotic cone of $\mathbb{Z}^d$ with the standard generators is $\mathbb{R}^d$ with the L^1 metric. If you change the generators, you get $\mathbb{R}^d$ with a different norm, as I will explain further in a later lecture. The asymptotic cone of a δ–hyperbolic space is 0–hyperbolic.
- Growth rate [see Lecture 4]
- Divergence (up to rate) and other "geometry at infinity" [see Lecture 5]
- Filling functions (up to rate) [see Lecture 7]
- Boundary (up to homeomorphism) and other "topology at infinity" [see Lecture 8]
- L^p cohomology [go read about it somewhere else!]

Now we can go back and state the theorem of M. Kapovich on the uselessness of K_3 somewhat more carefully. It is true for all X that $K_3(X) \subset \mathsf{Tri}_3$. Clearly if X is QI to a point, ray, or line, then K_3 is much smaller than Tri_3. Kapovich shows that if X is a length space which is not QI to those exceptions, then $K_3(X)$ contains the whole interior of Tri_3. If in addition X has arbitrarily long geodesic segments, then $K_3(X) = \mathsf{Tri}_3$ exactly. So K_3 detects almost nothing of the geometry of X.

3.3. **Quasi-geodesics.** Quasigeodesics are QI embeddings of the real line. They need *not* be continuous images (paths). The condition only requires that there are constants K, C such that

$$\frac{1}{K} \cdot |t_2 - t_1| - C \leqslant d(\gamma(t_1), \gamma(t_2)) \leqslant K \cdot |t_2 - t_1| + C.$$

What we will see is that quasigeodesics are very tame in hyperbolic spaces, but wild in flat spaces.

Example 2. Take two rays from the origin pointing in the first quadrant, say, and create a piecewise straight path between them that travels horizontally and vertically, turning when it hits a ray. It is at worst $\sqrt{2}$ off from being geodesic.

Example 3 (Väisälä)**.** Fancier example. Let $V = \{(x, |x|)\} \subset \mathbb{R}^2$ with the induced metric from Euclidean space, which you will note is NOT a length metric! Then $x \mapsto (x, |x|)$ is a quasi-isometric embedding $\mathbb{R} \to V$ which is surjective, so V is a quasigeodesic. Incidentally this serves as an example to show that quasi-isometry need not preserve hyperbolicity among non-geodesic spaces. Take any four points on the V with top two pairsums not equal. By dilating the picture, you can make the difference between pairsums arbitrarily big.

Example 4. Spirals in the plane can also be quasigeodesic! You'll work this out in the exercises. This illustrates that quasigeodesics are terribly wild in Euclidean geometry, which is a sign that quasi-isometries can behave very violently on flat spaces, even "wrapping them all the way around many times" to turn a straight line into a spiral.

Example 5. k–local geodesics (paths for which every subpath of length k is geodesic) were introduced in Day 1 exercises, where you showed that in a δ–hyperbolic space, an 8δ–local geodesic must be within 2δ of any true geodesic between the same endpoints. (Note this is horribly false in general: for example, in the L^1 metric on the plane, consider the path $(0,0)$–$(N,0)$–(N,N)–$(0,N)$. It is an N–local geodesic but has points at distance N from the only true geodesic between its endpoints.)

Something further is true: *an 8δ–local geodesic in a δ–hyperbolic space must be quasigeodesic.* This will be an exercise.

Later we'll show a stronger statement of the tameness of quasigeodesics in hyperbolic spaces: all quasigeodesics are close to true geodesics.

4. Groups and growth

4.1. **Cayley graphs.** $\mathsf{Cay}(G, S)$ denotes the Cayley graph of group G with generating set S. This graph has vertex set $V = G$ and edges (marked with an a) between g and ga for $a \in S$. From now on we will make the standing assumption (unless otherwise noted) that gensets are finite and symmetric ($S = S^{-1}$) and these will be undirected graphs with finite degree. So now, endowing the Cayley graph with the simplicial graph metric, we can officially talk about the geometry of groups.

It's actually hard to tell whether a graph is a Cayley graph (though it is necessary that it be vertex transitive), and *really* hard to tell whether a graph is "nearly" a Cayley graph. The graph theory of Cayley graphs is already subtle for finite groups: it's an open question whether every Cayley graph of a finite group has a Hamiltonian cycle, for instance.

Woess asked, in the 1980s: Can you find a (connected, locally finite) vertex transitive graph that is not QI to any Cayley graph? Two computer scientists, Diestel and Leader, constructed a family of graphs hoping to produce examples. The **Diestel-Leader graphs** $DL(m, n)$ are built from a pair of trees (one $m + 1$–regular and the other $n + 1$–regular) by a construction called a horocyclic product. The $DL(n, n)$ are Cayley graphs for certain groups called **lamplighter groups**, $\mathbb{Z}_n \wr \mathbb{Z}$. DL conjectured that for $m \neq n$, these were not QI to Cayley graphs, and this remained open for a long time.

4.2. **Milnor-Schwarz Lemma.** Let's say a **geometric** action is one that is isometric, cocompact, and properly discontinuous (for each compact K, only finitely many group elements fail move K off itself). Then we have the crucially important lemma of Schwarz and, later, Milnor.

Lemma 6 (Fundamental Observation of Geometric Group Theory)**.** *If G acts geometrically on a proper geodesic metric space X, then G is finitely generated and $\mathsf{Cay}(G, S) \underset{QI}{\sim} X$ for any finite genset.*

This is quite easily proved; the distance distortion between two Cayley graphs, for instance, is bounded in terms of the longest spelling length of a generator from one genset with respect to the other. Finite generation of G comes from considering a ball large enough that its G–translates cover X. Then there are only finitely many group elements that fail to move it off of itself, and these generate G.

In particular, since $\mathsf{Cay}(G, S) \underset{QI}{\sim} \mathsf{Cay}(G, S')$ for any two finite gensets: if you are only considering QI invariant properties, then any locally finite Cayley graph will do. To illustrate the subtlety here, all hell breaks loose when we consider infinite S, even for finitely generated groups: It is a tantalizing open question proposed by Rich Schwartz to decide whether $\mathsf{Cay}(\mathbb{Z}, \{2^n\}) \underset{QI}{\sim} \mathsf{Cay}(\mathbb{Z}, \{3^n\})$. See exercises!

Since hyperbolicity is QI invariant, we can now define a **hyperbolic group** to be a finitely generated group with some δ–hyperbolic Cayley graph, such as F_n or $\mathbb{Z}$, because in each case the graph in the standard generators is an actual tree. Important examples come from fundamental groups of compact hyperbolic manifolds, like $\pi_1(\Sigma_g)$ in the surface case. The uniformization theorem tells you that these groups act isometrically by deck transformations on $\mathbb{H}$, and the action is free and cocompact, so the group is QI to the hyperbolic plane itself. $SL_2(\mathbb{Z})$ is a slightly more sophisticated example, because its action on $\mathbb{H}$ is not cocompact, and we will establish its hyperbolicity below.

How about the converse of FOGGT? Only in special cases. For instance, groups QI to $\mathbb{R}$ act geometrically on $\mathbb{R}$. Tukia proved for $n \geqslant 3$: if $G \underset{QI}{\sim} \mathbb{H}^n$, then G acts geometrically on $\mathbb{H}^n$. This is true for $n = 2$ but much harder!

4.3. **QI rigidity.** Gromov set up a program to classify groups by quasi-isometry type. Say that a class of (fin-gen) groups is **QI rigid** if any group QI to a member of the class is *virtually isomorphic* to a member of the class. (Virtual isomorphism is a bit technical: isomorphism between respective quotients of finite-index subgroups by finite groups. It turns out to be the best you can hope for.)

QI rigid classes include: hyperbolic groups; virtually abelian groups; virtually nilpotent groups (by Gromov's Polynomial Growth Theorem, stated below); virtually free groups (central idea: Stallings's Theorem on ends); amenable groups; many classes of lattices in Lie groups; fundamental groups of closed hyperbolic n–manifolds; and lattices in Sol, the 3–dimensional solvable model geometry.

Non-rigid classes include: virtually solvable groups (Erschler); fundamental groups of nonpositively curved (i.e., CAT(0)) manifolds; and groups that act geometrically on a CAT(0) space.

Note: the Eskin-Fisher-Whyte work on Sol also proved the Diestel-Leader conjecture as a corollary of the main approach!

4.4. **Growth functions, growth series.** Growth functions with respect to generating sets count the number of points in the ball of radius n in the Cayley graph: the growth function is $\beta(n) := |B_n|$ and the spherical growth function is $\sigma(n) := |S_n| = \beta(n) - \beta(n-1)$. When necessary we can decorate β with G and S as β_G^S, since as a precise function it certainly depends on S.

One trivially gets an exponential upper bound on growth by expressing any word of length at most n with a string of generators, padded out to length exactly n by the identity: $\beta^S(n) \leqslant (|S|+1)^n$.

Let us define rate equivalence by $f(t) \preceq g(t)$ if $\exists A > 0$ such that $f(t) \leqslant Ag(At + A) + At + A$ for all $t > 0$. Say $f \asymp g$ if $f \preceq g$, $g \preceq f$. Note rate picks out polynomiality (being bounded above by a polynomial) and even degree (though identifies linear/sublinear), also picks out exponentiality (being bounded below by e^t). The additive factor of At might seem strange, but it will be needed later to make certain filling functions QI invariant.

Then one shows that (up to $\asymp$), growth is a QI invariant for groups. This is because if f is a QI then $f(B_r) \subset B_{Kr+C}$. Furthermore by the QI condition only finitely many points can map to any one point (looking back at the definition, $d_2 = 0 \implies d_1 \leqslant KC$), so $\beta_X(n) \leqslant KC \cdot \beta_Y(Kn + C)$ and vice versa. So from now on we can talk about rates of growth without specifying S.

This gives us our first argument that $\mathbb{Z}^m \underset{QI}{\sim} \mathbb{Z}^d \iff m = d$, because $\beta_{\mathbb{Z}^d}(n) \asymp n^d$.

Recall that a (f.g.) **nilpotent group** is one for which nested commutators are eventually trivial. And recall that a group **virtually** has a property if some finite-index subgroup has the property. Now we can state Gromov's theorem: *a finitely generated group is virtually nilpotent iff it has polynomial growth.*

Actually it is useful to note that QI invariance of growth rates works for discrete spaces more generally than Cayley graphs, if the growth function is given with respect to a basepoint. And indeed it works for nets, so it transfers to Riemannian volume: if a f.g. group acts geometrically on a Riemannian manifold (connected, complete, with *bounded geometry*), then $\beta(n) \asymp \mathrm{Vol}(B_n)$.

Theorem 7 (Coornaert)**.** Non-elementary hyperbolic groups have *definite exponential growth*:

$$C_1\alpha^n \leqslant \beta^S(n) \leqslant C_2\alpha^n$$

with respect to any generating set. ($\alpha = \alpha(S)$)

Certainly not all groups of exponential growth are hyperbolic: e.g., $F_2 \times \mathbb{Z}$. Note that the growth rates you can see for compact surface groups are $1, n^2, e^n$, since uniformization tells us that the universal covers are $S^2, \mathbb{E}^2, \mathbb{H}$. For 3–manifolds, the full list turns out to be $1 \asymp n, n^3, n^4, e^n$, by geometrization.

One might well wonder: Are there f.g. groups of *intermediate* growth between polynomial and exponential? A famous theorem of Grigorchuk constructs such a group. (However it is not finitely presented and it's not known if that's possible.)

4.5. **Growth series.** Given a growth function, consider the series given by the associated generating function $F(x) = \sum \beta(n)x^n$. One might wonder whether the series is rational, i.e., $F(x)$ is a ratio of polynomials.[1] Note that rational growth can occur in polynomial or exponential growth regimes: you can check that so $1/(1-x)^k$ forms a series whose coefficients are polynomial of degree k. On the other hand, $1/(1-2x) = \sum 2^n x^n$ has exponentially growing coefficients.

Some theorems:

Benson: virtually abelian groups have rational growth with respect to all gensets.

Cannon: hyperbolic groups have rational growth with respect to all gensets.

Shapiro: the Heisenberg group has rational growth with respect to the *standard* genset.

Stoll: the 5D Heisenberg group does not!! It has rational growth wrt a special ("cubical") genset, but transcendental in the standard generators.

Duchin–Shapiro: nevertheless, $H(\mathbb{Z})$ has rational growth wrt all gensets.

[1]The reason that rational growth is interesting is that it gives a recursive relationship among the $\beta(n)$ values, so you can generate all values of the β function from the knowledge of finitely many, and this turns out to be equivalent to being able to draw the full Cayley graph from the information in a finite portion.

5. Morse lemma, divergence, contraction

It is immediate to see that in a hyperbolic space, two geodesic segments between the same two endpoints must be δ–close (for the δ in the thin triangles definition). That's just because you can form a geodesic triangle by splitting one of the segments in half, and every point on the other segment must be δ–close to one of the halves. You can think of this as **geodesic stability**.

It takes a bit more work to establish **quasigeodesic stability**.

5.1. Detours.

We begin with a crucial observation: to avoid a ball in a hyperbolic space, you have to take a detour of exponential length. For instance in $\mathbb{H}$, circumferences of metric circles have length $\sinh r \sim e^r$ compared to their diameter $2r$, while in trees there is no such detouring path at all!

Bridson-Haefliger sets this up very elegantly. Draw your own pictures to track the arguments.

Lemma 8. If a path between points on a geodesic avoids some ball $B_D(x)$ centered at any x on the geodesic, then its length ℓ satisfies $\ell \geqslant 2^{\frac{D-1}{\delta}}$.

Proof. Suppose γ is your avoidant path and α is the geodesic being avoided. We'll take an arbitrary point x on α and show that $d = d(x, \gamma)$ satisfies $\ell(\gamma) \geqslant 2^{\frac{d-1}{\delta}}$.

Let $\ell = \ell(\gamma)$ and fix the n so that $\frac{\ell}{2^{n+1}} \leqslant 1 \leqslant \frac{\ell}{2^n}$. Now subdivide γ into 2^n equal-length subsegments, so that each has length $\ell/2^n$. Consider the geodesic triangle formed by α and the midpoint of γ. This has two neighboring geodesic triangles formed by each new side together with the points 1/4 and 3/4 of the way along γ, and each of these has two neighboring geodesic triangles with the 1/8 points, and so on. Let x_1 be the closest point to x on either of the new sides of the first triangle; let x_2 be the closest point to x_1 on either of the sides of a successive triangle, and so on. By thin triangles, we have $d(x_n, x) \leqslant n\delta$. On the other hand, x_n is on a geodesic segment of length $\ell/2^n$, so if we let y be the nearer endpoint on γ, we have $d(x, \gamma) \leqslant n\delta + \ell/2^{n+1} \leqslant n\delta + 1$. From $d \leqslant n\delta + 1$, we get $\frac{d-1}{\delta} \leqslant n$, so $2^{\frac{d-1}{\delta}} \leqslant 2^n \leqslant \ell$, as desired. We've bounded d from above relative to ℓ, so taking x to be a point on α which is at least D from γ, we also get a bound on $\ell(\gamma)$ from below relative to D. □

5.2. Quasigeodesic stability.

Theorem 9 (Morse Lemma)**.** In a δ–hyperbolic space, for every K and C there is an M such that every (K, C)–quasigeodesic segment is M–close to every true geodesic between the same endpoints.

Still following Bridson-Haefliger, I will sketch a proof in the case that β is a (K, C)–quasigeodesic segment that is a path (a continuous embedding of an interval). (This is not too much of a worry, because for any quasigeodesic one can find a nearby continuous path.) Let α be the true geodesic between the same endpoints. Let x be the point on α that is farthest from β, and let that distance be D. We begin by finding an upper bound on D. To do this, just note that all of β must avoid the (open) D–ball centered at x, just because x never gets closer than D to β. Let p, q be points on α just outside that D–ball, say at distance $2D$ from x on either side. (Really anything larger than D will do, so take the endpoints if they are closer than $2D$ from x.) Let their closest point projections to β be called p', q', and let γ_0 be the subsegment of β from p' to q' and γ be the path from p to q obtained by $\overline{pp'}$ followed by γ_0 followed by $\overline{q'q}$. Since γ avoids $B_D(x_0)$, its length satisfies $2^{\frac{D-1}{\delta}} \leqslant \ell(\gamma)$. On the other hand, there is a path between the endpoints of γ_0 of length at most $6D$ because $d(p, p'), d(q, q') \leqslant D$ by definition of D, and $d(p, q) = 4D$. By quasigeodesity, $\ell(\gamma_0) \leqslant 6DK + C$, and thus $\ell(\gamma) \leqslant 6DK + C + 2D$. So we have

$$2^{\frac{D-1}{\delta}} \leqslant 6DK + C + 2D,$$

which fails for large D, so there is an upper bound D_0 on the possible values of D. We are now almost done. The last thing to worry about is the possibility that β might take a long sojourn away from α, even though α is always uniformly close to some point of β. But for this to happen, consider the subsegment of β consisting of the sojourn outside of $\mathcal{N}_D(\alpha)$. It begins and ends at points that are at most $2D$ from each other (since they are at most D from some point on α), so the quasigeodesity condition ensures that its length is bounded, and this completes the argument.

Consequence: **fellow-traveling**. If two quasigeodesics have close starting points and close ending points, then there is a bound on how far apart they are in terms of K, C, and the meaning of "close."

5.3. **Divergence.** We recall that in a tree, rays from a common basepoint agree exactly until they diverge completely. From this and from the qualitative behavior of negatively curved spaces, we should expect that hyperbolic spaces have fast, probably exponential, divergence of geodesics. However, it won't do to take a pair of rays γ_1, γ_2 and simply measure the separation $d(\gamma_1(t), \gamma_2(t))$, because the triangle inequality ensures that this is always $\leqslant 2t$.

Instead, we can define the divergence of geodesics as follows: if γ_1 and γ_2 are rays from a common basepoint x_0, then let $div(\gamma_1, \gamma_2, t)$ be the length of the shortest path from $\gamma_1(t)$ to $\gamma_2(t)$ avoiding $B_t(x_0)$ (if such a path exists).

Then let γ_1, γ_2 be the rays pointing opposite directions along a bi-infinite geodesic in a δ–hyperbolic space. The detour estimate above immediately gives us $div \geqslant 2^{(t-1)/\delta} \succeq e^t$.

By contrast, for any two rays in the Euclidean plane, $div = 2\pi\theta t$, an exactly linear function with coefficient given by the angle between the rays. So divergence seems to successfully distinguish curvature regimes.

Proposition 10. In any δ–hyperbolic space, if two rays γ_1, γ_2 in the same end eventually separate by more than 2δ, then $div(\gamma_1, \gamma_2, t) \asymp e^t$.

Indeed, this fact that *past a threshold separation, divergence is exponential* turns out to be an alternate definition of δ–hyperbolicity.

Now let us try to use divergence to attach an invariant to spaces and groups. We need to be a bit careful to make it a QI invariant, which we'll do following the setup of Gersten. Define $div(X)$, the divergence rate of a space X, as follows. First define functions

$$f_\rho(t) := \sup_{x,y \in S_t(x_0)} \inf \ell(\gamma)$$

for $0 < \rho \leqslant 1$, where the inf is over all paths γ from x to y in $X \backslash B_{\rho t}(x_0)$. (The role of ρ is to allow the filling path to dip some fixed proportion inside B_t, which is necessary for QI invariance.) Then you get a family $\{f_\rho\}_\rho$ of functions of t. We'll say that $div(X) \asymp t^n$ if $f_\rho(t) \asymp t^n$ for all sufficiently small ρ. This is the meaning of polynomial divergence, and exponential divergence is defined similarly. So we see that $div(\mathbb{E}^d) \asymp t$ and of course $div(\mathbb{H}) \asymp e^t$.

Quick facts: of course, nonelementary hyperbolic groups have exponential divergence; nilpotent groups have linear divergence (see exercises); NPC symmetric spaces have linear or exponential divergence; right-angled Artin groups have linear or quadratic divergence, mapping class groups have quadratic divergence; and there are CAT(0) groups, and even right-angled Coxeter groups, with every polynomial degree of divergence. This last fact is a recent theorem of Dani-Thomas.

WARNING! Many authors (Bridson-Haefliger, Papasoglu, Short et al) use a different definition of divergence for spaces that is problematic because it asks for a precise divergence function bounding all pairs of rays rather than a global rate like this. With their definition, it is not even the case that the divergence of flat spaces is linear. (Since the divergence function of a pair of rays in the flat plane is $2\pi\theta t$, there is no one linear function that lower-bounds them all!) This is how it is possible to find statements in the literature such as "unbounded divergence implies exponential divergence."

5.4. **Contraction.** We've already seen that straying far from a geodesic forces you to be long; on the other hand, projection to a geodesic makes you teeny.

The strong contraction property sounds so strong you might have to check that you read it right. *In a δ–hyperbolic geodesic space, there exists $M > 0$ depending on δ such that for any geodesic γ, and any ball $B_r(x)$ of arbitrary radius that is disjoint from γ, the closest-point projection of the ball to the geodesic has diameter at most M.*

Alternate statement of strong contraction: *there exists $M > 0$ so that if any geodesics α, β satisfy $\beta \cap \mathcal{N}_{2\delta}(\alpha) = \varnothing$, then the closest-point projection* $\mathrm{proj}_\alpha(\beta)$ *has length* $\leqslant M$.

These are not-too-hard consequences of hyperbolicity, left to the exercises. The fact that all geodesics are contracting in this way is yet another alternate definition of hyperbolicity, but this is interesting partly because some geodesics even in non-hyperbolic spaces behave this way. We can call these *contracting geodesics.*

6. Algorithmic problems, Dehn's algorithm

There are three classical algorithmic decision problems for groups. First is the **word problem**, which asks for an algorithm to tell if two words represent the same group element. Second is the **conjugacy problem**, which takes two words and asks if they are in the same conjugacy class. Finally, the **isomorphism problem** is to decide whether two different presentations represent the same abstract group. More recently a lot of attention has turned to the **equation problem**: If I give you a word in group elements and variables and ask if there are values of the variables for which $w = 1$, is there an algorithm to decide yes or no? (Variations: systems of equations; systems of equations and inequations.)

Interestingly, there are long-known classes of groups for which each of these problems is undecidable, since Novikov 1955.

Hyperbolic groups have a decidable word problem by an easy algorithm due to Dehn, discussed below. Massively hard work due to Rips, Sela, Dahmani-Guirardel, and others shows that hyperbolic groups also have decidable conjugacy, isomorphism, and equation problems (even for systems of equations and inequations).

6.1. **Dehn presentations.** Let us say that $G = \langle a_1, \ldots, a_n \,|\, r_1, \ldots, r_m \rangle$ is a **Dehn presentation** of G if the following very special set of circumstances is in place:

- There is a set of strings $u_1, v_1, \ldots, u_m, v_m$ and each relator r_i is of the form $r_i = u_i v_i^{-1}$. (Relator r_i encodes the equivalence in the group of u_i and v_i.)
- For each i, the spelling length of v_i is shorter than the spelling length of u_i.
- For any nonempty string w in the alphabet $S = \{a_i\}$ that represents the identity element, if w has been reduced by canceling all occurrences of $a_i a_i^{-1}$, then at least one of the u_i or u_i^{-1} must appear as a substring.

This last bullet point is a lot to ask! In general, given a presentation, the word problem is hard: you can look at a long string of generators, and there's no good way to see whether it simplifies, because you might have to first use one relation in the group that makes the string longer before you can use another relation that makes it shorter. A Dehn presentation is special because any string can be simplified by using one of finitely many moves, each of which makes the string strictly shorter. Thus the number of replacements required is less than the initial length of the string.

As a trivial example, $F_2 = \langle a, b \,|\, \varnothing \rangle$ is a Dehn presentation with $m = 0$. This works because there are no spellings of the identity left after words are reduced.

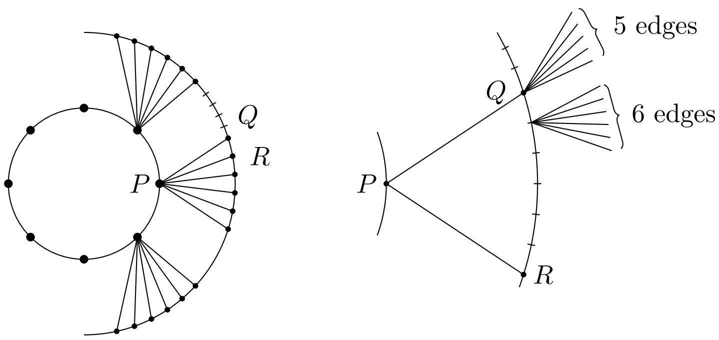
As a less trivial example, consider the presentation $PSL_2(\mathbb{Z}) = \langle J, B \,|\, J^2, B^3 \rangle$. The Cayley graph is nearly a 3–regular tree: it has triangles labeled by B–edges, mutually connected by J–edges. We can form a Dehn presentation with $u_1 = B^2$, $v_1 = B^{-1}$, $u_2 = J^2$, and $v_2 = I$, clearly satisfying the first two bullet points. For the third, we see from the Cayley graph that any nontrivial loop with no J^2 or JJ^{-1} must contain a path around one of the B–triangles. Thus it has a B^2 or B^{-2} substring.

A similar idea works for surface groups; consider for example $\pi_1(\Sigma_2)$, the fundamental group of the genus-two surface, which by Poincaré's Theorem has the presentation

$$\langle a, b, c, d : \; [a, b]\cdot[c, d] \; \rangle.$$

The idea is that any loop includes some combination of backtracking and at least six letters of the length-eight relator. So the shortening method works much the same as when dealing with the triangles in the $PSL_2(\mathbb{Z})$ case: begin with the length-eight string $aba^{-1}b^{-1}cdc^{-1}d^{-1}$ and consider all of its cyclic permutations ($ba^{-1}b^{-1}cdc^{-1}d^{-1}a$, etc). Put all of the six-letter initial substrings of these as the u_i and their two-letter counterparts as the v_i.

Now we need to convince ourselves that this is indeed Dehn. We know from the classical Poincaré theorem that the fundamental domain for the action of this group on $\mathbb{H}$ is an octagon in the hyperbolic plane, and we can draw the Cayley graph as the dual graph to this octagonal tiling. The graph will therefore have an octagon corresponding to the basic relator, enclosing a vertex around which there are eight tiles. By symmetry, the whole Cayley graph can be drawn in the plane as a graph of octagons. A nice way to visualize this combinatorics precisely is to draw a schematic that organizes the octagons by their distance from the central octagon. I am adapting this visualization directly from Stillwell's



Classical Topology and Combinatorial Group Theory. It is an excellent way to understand the sense in which the structure of the genus two surface group is tree-like but not a tree.

Begin with an arbitrary octagon in the tiling and construct a diagram by arranging its vertices and edges around the unit circle C_1 in the plane. Then arrange the neighboring octagons radially around it, with the next ring being formed by those that share either an edge or just a vertex with the center tile, and all of those lying between the circles C_1 and C_2; continue this way. Every octagon appears exactly once in this arrangement, and—aside from the central one—lies between the circle C_k and the circle C_{k+1} for some k. Any such octagon either has 0 or 1 edges on C_k, and accordingly it has either 6 or 5 edges on C_{k+1}. (The octagon containing the vertices marked P, Q, and R in the figure is the first case, and the one on the other side of edge PQ is the second.) Now consider an arbitrary loop along edges in the diagram. It touches some outermost circle C. If the loop ever traverses an edge and then immediately backtracks, then that represents a trivial cancellation $a_i a_i^{-1}$. Otherwise it reaches C along a radial edge, travels along C, and then returns along a different edge. But then that word contains 6 of the edges around a single octagon (the radial edge plus at least 5 edges on the outer circle).

Thus we can see that the tiling is rapidly branching, and though it is not a tree (there is more than one way to reach a vertex), alternate paths are costly.

6.2. **Dehn's algorithm.**

Theorem 11. Hyperbolic groups admit Dehn presentations.

Proof. Fix any $K > 8\delta$; if you know δ, then this is a constructive proof; if you don't, then it's just an existence proof! Then consider an arbitrary (finite) generating set $S = \{a_i\}$ for G and form *all* freely reduced spellings t_i with spelling length at most K. There are a lot of these, but only finitely many. The group was given to you either by a Cayley graph or by some presentation, which you can use to build out finitely much of the Cayley graph in finite time. So now you can check which of the t_i represent the same group element by just following them in the graph. Let the u_i be the non-geodesic spellings from that list, and for each u_i, let v_i be some geodesic spelling of the same group element, so it is guaranteed to be strictly shorter. Then put $R = \{r_i = u_i v_i^{-1}\}$, and I claim $G = \langle S|R\rangle$ is a Dehn presentation. The first two bullet points are satisfied by construction. For the third, note that there are no 8δ–local geodesic loops of length at least 8δ in a δ–hyperbolic space. (This follows from the fact that any 8δ–local geodesic stays within 2δ of the true geodesic between its endpoints.) So any loop longer than 8δ has a non-geodesic subsegment of length $\leqslant K$, which is one of our u_i above. If the loop is shorter than 8δ, then it *is* one of the u_i. This verifies the last condition. □

This gives us **Dehn's algorithm** for solving the word problem in hyperbolic groups: starting with a word that may or may not represent the empty word, we seek these u_i subwords. Each time we can replace one with its corresponding v_i, we have shortened our word. If we find no u_i and no trivial cancellations $a_i a_i^{-1}$, we halt, and if there are any letters left, our original word was nontrivial in the group. This actually solves the word problem in linear time.

You may not be surprised to hear: a group is hyperbolic if and only if it has a Dehn presentation. However, there are other groups with fast algorithms for the word problem, such as free abelian groups! (Just check if each exponent sum is zero.)

7. Filling

7.1. **The isoperimetric problem.** The classical isoperimetric problem in the plane asks how to extremize the relationship between the length of a loop and the area it encloses: in other words, find $\sup_\gamma \inf_D A(D)$, where A stands for area, the sup is over loops γ of length at most 1, say, and the inf is over fillings, i.e., topological disks with boundary γ. As we know, this sup is achieved for circles, where the value is $1/4\pi$. By dilating, we know that circles give the best answer on every scale. We can set up a generalization by defining an isoperimetric function $f(\ell) = \sup_{\ell(\gamma)\leqslant \ell} \inf_D A(D)$, and now of course we get $f(\ell) = \ell^2/4\pi$, a quadratic function of length.

In general it will be really hard to compute this minimax problem exactly, so we will be content with finding rates of growth; here $f(\ell) \asymp \ell^2$.

What about the hyperbolic plane? It will turn out that round circles are still optimal; but here, for a circle of radius r, the circumference and area are both on the order of e^r, so we get $f(\ell) \asymp \ell$. Fillings are actually "cheaper" in hyperbolic spaces! (Though maybe this isn't so surprising, since loops themselves are so expensive relative to their diameter.)

We could have set up this definition in any Riemannian manifold or symmetric space, using length and 2–volume.

7.2. **Dehn functions.** To define filling functions in groups, we need a notion of area. We will build a *Cayley 2–complex* associated to a presentation: begin with a Cayley graph; paste in 2–cells corresponding to the relators in the presentation; and add 2–cells for all of their conjugates. This is the universal cover of the *presentation 2–complex* built on a single vertex with loops for generators and 2–cell for relators. (Note: You can always add higher cells to the presentation 2–complex to get a $K(G,1)$ for the group.)

If a word represents the identity element, then of course it corresponds to a loop in the Cayley graph, and since our Cayley 2–complex is simply connected, that loop is null-homotopic. Define a filling of the loop to be the cells crossed in a null-homotopy; then area of a loop is the least number of 2–cells you must slide the loop across to trivialize it. Example: $\mathbb{Z}^2 = \langle a, b : [a, b] \rangle$, the word $a^n b^n a^{-n} b^{-n}$ has area n^2, though of course it also has less efficient fillings corresponding to homotopies that enlarge it before shrinking it.

The Dehn function is then the function $f(n)$ as above. One shows that this does not depend on the presentation (up to $\asymp$) and indeed that if we take a slightly coarser notion of area we get a QI invariant, so the Dehn function of $\mathbb{Z}^2$ is precisely n^2, and the Dehn function of $\mathbb{Z}^d$ is still n^2 for any $d \geqslant 2$.

Loosely, we can get coarse filling area in metric spaces by defining a filling to be a map from a disk in $\mathbb{R}^2$ to X that takes the boundary to the loop to be filled, and takes the unit grid of points in the disk to a net of bounded mesh in the space. Then the least Euclidean area of such a disk can be taken to be the filling area of the loop. Also this means that for more generality we could have worked in any model space on which G acts geometrically. But one has to be very careful about definitions and there are many subtleties in the QI invariance arguments in different settings.

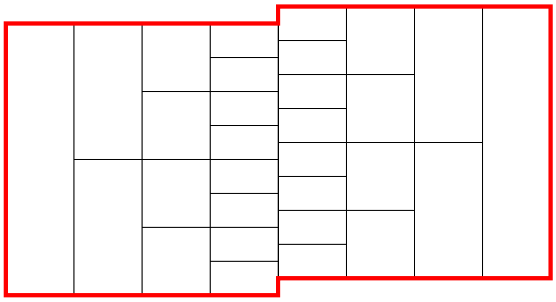
7.3. **Hyperbolic groups.**

Proposition 12. Hyperbolic groups have linear Dehn functions.

Proof. Dehn's algorithm! In the presentation complex for the Dehn presentation, each replacement of u_i with v_i slides across a single cell, and this only needs to be done at most n times. □

The two other pieces of this story are remarkable: anything with a linear Dehn function is hyperbolic, and anything that is not hyperbolic has at least a quadratic Dehn function. (These can be proved by clever but basically standard hyperbolicity arguments.) So there is a gap between linear and quadratic rates of filling—for instance, you can't engineer a group with $n^{3/2}$ or $n \log n$ Dehn function. And, taken together, these facts mean that linear (or indeed subquadratic) filling functions provide yet another alternative definition of hyperbolic spaces.

7.4. **A few other examples.** The Dehn function of the Baumslag-Solitar group $BS(1,2) = \langle a, b : bab^{-1} = a^2 \rangle$ is easily seen to be exponential by taking a loop that's half in one plane-of-bricks and half in another. To be explicit: one considers the word $ab^{-k}ab^k a^{-1} b^{-k} a^{-1} b^k$, shown in red for $k = 4$.



The length of this word is linear in k (namely $4k + 4$) but any filling contains the cells that are shown, and there are $2^{k+1} - 2$ of them.

The Heisenberg group has cubic Dehn function, but Allcock showed that the higher Heisenberg groups have quadratic Dehn function.

For the linear groups $SL_n(\mathbb{Z})$, we know that the Dehn function is linear for $n = 2$ because the group is hyperbolic, and Thurston showed that it is exponential for $n = 3$. He conjectured that it then drops to quadratic for all $n \geqslant 4$! This was recently proven by Robert Young for $n \geqslant 5$ and is still an open problem for $n = 4$.

Automatic groups have quadratic Dehn function, and this includes mapping class groups. However, $\mathrm{Out}(F_n)$ and $\mathrm{Aut}(F_n)$ both have exponential Dehn function for $n \geqslant 3$.

7.5. **Higher Dehn functions, higher divergence functions.** Isoperimetric inequalities can be studied in higher dimension, for instance by filling spheres S^k of at most a certain k–volume by balls B^{k+1}. (One could also loosen up the topology: for instance, instead of a disk, a loop can be filled with any surface with circle boundary.) Or one can do things in the homological category, filling cycles by chains. If you do this very carefully, you can get QI invariants for each dimension k. And for reasonable definitions of higher filling functions, hyperbolic groups have linear Dehn functions in every dimension (see Mineyev/Lang).

Briefly, let me note that the divergence of geodesics, which studied how hard it can be to connect two points with an "avoidant" path, admits a nice generalization to higher dimension similarly. A pair of points is a 0–sphere, and a path is a 1–ball. Instead, we can study the filling of avoidant k–spheres with avoidant $(k + 1)$–balls.

As motivation for why you might want to do this, note that you get some QI invariants this way which distinguish more different kinds of spaces. For instance, higher divergence functions detect the rank of a symmetric space (Hindawi/Leuzinger), and they give finer distinctions between right-angled Artin groups than were previously available (giving new results in the QI classification of RAAGs).

7.6. **The Dehn spectrum and Wenger's results.** What are the possible growth rates of Dehn functions? First, consider the powers of growth, defining the "isoperimetric profile" to be

$$\mathrm{IP} = \{\alpha \in [1, \infty) : \delta_G(n) \asymp n^\alpha \text{ for some f.g. } G\}.$$

Then $\overline{\mathrm{IP}} = \{1\} \cup [2, \infty)$ and $\mathbb{Q} \cap [2, \infty) \subset \mathrm{IP}$. Amazingly, for higher Dehn functions, there is no gap! In all dimensions bigger than 1, $\mathbb{Q} \cap [1, \infty) \subset \mathrm{IP}$. This combines work of Brady, Bridson, Forester, Shankar, and others.

What else can you see besides power growth and exponential growth? Iterated HNN extensions of $BS(1, 2)$ have Dehn functions that are towers of exponentials of any height! And this can be done for higher Dehn functions as well.

There is also non-power growth in the polynomial range. For k–step nilpotent groups, the Dehn function is $\preceq n^{k+1}$. Wenger constructs a class of 2–step nilpotent groups in which $n^2 \precneqq \delta(n) \preceq n^2 \log n$.

Wenger also proves a remarkable sharpening of Gromov's gap: not only is there a gap between linear and quadratic rates of growth, but indeed if the isoperimetric function in a geodesic space satisfies $f(\ell) \leqslant (1 - \epsilon)\ell^2/4\pi$ for any $\epsilon > 0$, i.e., if it is even the slightest bit less than the Euclidean rate, then the space is hyperbolic.

8. Boundaries

8.1. **Bordification.** A **bordification** of a metric space X will be some Hausdorff space $\bar{X}$ in which X embeds as an open, dense subset. For a given bordification, the **boundary** of X is $\partial X := \bar{X} \backslash X$. This may or may not be a compactification.

A basic example is the one-point compactification: take $\bar{X}$ to be X plus a single point, and take basic open neighborhoods of the point to be itself plus complements of closed balls in X. But this is a special case of the **ends compactification**: one can take $\bar{X} = X \cup \text{Ends}(X)$. Earlier, ends were informally defined as connected components of $X \backslash B_r(x_0)$ as r gets large. More formally, ends are equivalence classes of proper paths based at x_0: two paths are equivalent if for every r there exists some time after which the paths are in the same connected component of $X \backslash B_r$. Similarly a sequence of points x_n in X converges to an end $[\gamma]$ if for each large r there exists an n after which the x_n are in the γ component of $X \backslash B_r$. Clearly for one-ended spaces (like $\mathbb{R}^2$), this recovers the one-point compactification.

8.2. **Visual boundary.** Define the **visual boundary** of a geodesic space X to be

$$\partial_\infty(X) = \partial_\infty(X, x_0) = \{\text{geodesic rays } \gamma \text{ based at } x_0\}\big/_{\sim} \ ,$$

where $\gamma \sim \gamma'$ if $\gamma' \subset \mathcal{N}_m(\gamma)$ for some $m < \infty$ (i.e., if they stay within some bounded distance forever, in which case we call them *asymptotic*). If $\xi = [\gamma] \in \partial_\infty X$, we also say that γ is asymptotic to ξ. We think of $\partial_\infty X$ as the "lines of sight" from x_0.

This boundary has a nice topology: two rays are *close* if they stay within distance m for a long time, so a sequence of rays γ_n converges to γ if the length of time for which γ_n fellow-travels γ to within some m goes to infinity as $n \to \infty$. (That is, use the compact-open topology.)

In $\mathbb{E}^2$, as in $\mathbb{H}$, no distinct rays emanating from x_0 are asymptotic, so the visual boundary is just the set of directions, parametrized by angle $\theta \in S^1$. The topology is just the standard topology on the circle, so $\partial_\infty(\mathbb{E}^2) \cong \partial_\infty(\mathbb{H}) \cong S^1$. In a tree it is also true any distinct rays give different boundary points. In the four-regular tree that gives the standard Cayley graph for the free group F_2, it's easy to see that the boundary is uncountable, with the topology of a Cantor set.

This construction can produce funny fractal boundaries: Kapovich-Kleiner show that if the visual boundary of a hyperbolic group is 1–dimensional, connected, with no local cut points, then it is homeomorphic to either a Sierpinski carpet or a Menger sponge.

8.3. **Visual boundaries of hyperbolic spaces, and visual metrics.** For hyperbolic geodesic spaces, the visual boundary is a quasi-isometry invariant, up to homeomorphism. That is, if $f : X \to Y$ is a QI embedding, then f extends continuously to $\partial_\infty X$. The proof is essentially what you would expect: geodesics in X are mapped to quasigeodesics in Y, which are close to geodesics that are all mutually asymptotic, so there is a well-defined point of $\partial_\infty Y$ associated to each point of $\partial_\infty X$. Continuity is because if a sequence of boundary points $\xi_n \in \partial_\infty X$ converges to $\xi \in \partial_\infty X$, then there are corresponding geodesic rays that fellow travel one ray for longer and longer times, and after applying the quasi-isometry we can arrange that this is still true.

Thus $\mathbb{H}^m \underset{QI}{\sim} \mathbb{H}^d \iff m = d$, and cocompact lattices $\Lambda \leqslant SO(m,1)$ and $\Gamma \leqslant SO(d,1)$ are QI iff $m = d$. This is just because $\partial_\infty(\mathbb{H}^d) \cong S^{d-1}$.

Recall the Gromov product $(y \cdot z)_w$, which was defined as half the defect in the triangle inequality, and comes close to measuring the distance of w to $\overline{yz}$. This extends to a Gromov product on the boundary by defining $(\xi \cdot \nu)_{x_0} = \lim_{t\to\infty}(\alpha(t) \cdot \beta(t))_{x_0}$ where $\xi = [\alpha], \nu = [\beta]$. If rays fellow-travel a long time, then the Gromov product is large. So we can put a **visual metric** on the boundary defined by $d(\xi, \nu) := e^{-(\xi \cdot \nu)}$, and this will be compatible with the topology already defined.

Notice how similar this is to the p–adic norm: things are p–adically close if they agree for a long time in their p–adic expansions. (Yet more evidence that $\mathbb{Q}_p$ is legitimately hyperbolic.)

8.4. **Basepoint independence and visibility.** In the hyperbolic setting, for any basepoint $x_0 \in X$ and boundary point $\xi = [\gamma] \in \partial_\infty X$, there is a geodesic ray based at any other point $p \in X$ and asymptotic to γ. This is found by considering the sequence of geodesic segments from p to $\gamma(n)$ and applying Arzelà-Ascoli. So we really didn't need to fuss about basepoints.

Similarly, there is a bi-infinite geodesic between any two points on the boundary: consider the geodesic segments from $\alpha(n)$ to $\beta(n)$. By thin triangles, these all dip back to some neighborhood of where α and β separated. But then we can again apply Arzelà-Ascoli to get a limit. When any two points on the boundary can "see" each other in this way, we call the space a **visibility space**.

Note that $\mathbb{E}^2$ badly fails to be a visibility space: only antipodes on the boundary are connected by a geodesic line. If we tried to mimic the construction from the last paragraph for non-antipodal boundary points, we'd find that the slope of the connecting segment depended completely on the sequences of endpoints chosen on the rays.

8.5. **Contracting boundary.** Some of the characterizations of hyperbolicity we have seen above were that all geodesics are contracting and stable. That is part of what gives the visual boundary such nice properties. Recall that even in a non-hyperbolic space, some rays may possess this contraction property. For instance, Minsky shows that in Teichmüller space, geodesics that stay in the so-called *thick part* are contracting. It is understood which kinds of geodesics in RAAGs are contracting, based on the amount of time that they spend in flats.

Recent work of Charney and Sultan defines a boundary for more general spaces by just considering classes of contracting rays. This does successfully produce a QI-invariant boundary for larger classes of groups: CAT(0), so far.

8.6. **Horofunctions.** The simplest and most natural description of the horofunction boundary is this: we want to build a bordification directly from the distance function. We note that distance gives a map from X to the function space $C(X)$ via $x \mapsto f_x := d(x, \cdot)$. But this won't give any nice limiting behavior as x goes to infinity in X because for any fixed y the $f_x(y)$ will diverge, so the f_x don't converge to a nice function. We fix this by a normalization:

$$F_x := d(x, \cdot) - d(x, x_0), \quad \text{so that} \quad F_{\{x_n\}}(y) = \lim_{n\to\infty} d(x_n, y) - d(x_n, x_0)$$

may possibly converge to a function for an appropriate sequence of points x_n.

In particular, consider x_n exiting to infinity along a geodesic ray. Then $\overline{x_n y}$ and $\overline{x_n x_0}$ are two sides of a geodesic triangle whose third side is $\overline{y x_0}$; if the sidelengths of a triangle are a, b, c, then $b + c \geqslant a \implies b - a \geqslant -c$, so $F_{x_n}(y) \geqslant -d(x_0, y)$. Also, the difference $F_{x_{n+1}}(y) - F_{x_n}(y)$ is equal to $d(x_{n+1}, y) - d(x_{n+1}, x_n) - d(x_n, y)$, and the triangle inequality again ensures that this is nonpositive, so the sequence of numbers $F_{x_n}(y)$ is non-increasing and bounded below, and thus converges!

Example, which you should check: if $\{x_n\}$ goes up along the y–axis in $\mathbb{E}^2$, then the associated horofunction evaluates at $y = (p, q)$ as $F(y) = -q$. Thus its level sets are horizontal lines.

So if we let $\iota : X \to C(X)$ be given by this $\iota(x) = F_x$, we can let the horofunction bordification be its closure, and the horofunction boundary $\partial_h(X)$ be all new limit points (**horofunctions**) obtained in this process. The functions induced by geodesic rays are called **Busemann functions**, but there can be others. For proper metric spaces this is always a compactification.

For CAT(0) spaces, Busemann functions are the only horofunctions, and there is a nice Busemann map $\partial_\infty \to \partial_h$ that is a homeomorphism. In general, there can certainly be Busemann functions that are not horofunctions, even in δ–hyperbolic spaces.

Horocycles are defined to be level sets of horofunctions, and **horoballs** are sub-level sets. This recovers the usual definitions of horoballs and horocycles in the hyperbolic plane: balls and disks tangent to the boundary.

8.7. **Good boundaries for groups?** These are useful constructions, but they have some issues. The visual boundary need not even be a bordification if G is not hyperbolic: in particular, you'll show in the exercises that $\partial_\infty \mathbb{Z}^2$ is an uncountable set with the trivial topology, meaning that undistorted flat subgroups produce useless, horribly-non-Hausdorff visual boundaries. The horofunction boundary isn't QI invariant and indeed for Cayley graphs it depends heavily on the choice of generating set.

Neumann and Shapiro posed a question in one of their papers. *Find a notion of boundary for finitely generated groups that meets just three desiderata: intrinsic to the group itself; the boundary of a free group is a Cantor set; the boundary of $\mathbb{Z}^d$ is a sphere.* There is no known construction that works.

9. Dynamics

9.1. **Quasi-axes and classification of isometries.** Recall the classification of isometries of $\mathbb{H}$: in the hyperbolic case, there was a geodesic axis, and the action of g pushed points along that axis from a global repelling point on the boundary to a global attracting point. It will turn out that in hyperbolic groups, all infinite order elements behave roughly this way: they have a quasigeodesic serving as an axis.

Proposition 13. For any infinite-order element g in a hyperbolic group G there is a (K, C)–quasigeodesic that is g–invariant. (K, C depending only on δ.)

Note that g can't fix a point, because $gh = h \implies g = e$. Choose some vertex h of the Cayley graph for which $d(h, gh)$ is minimal, which is possible since it takes integer values. Let $d = d(h, gh)$; if $d \geqslant 8\delta$, then let γ be the concatenation of the g–images of a geodesic from h to gh. By construction, this is an 8δ–local geodesic, so it is quasigeodesic.

Now suppose some point is moved by less than 8δ. Since g has infinite order, its orbits can't remain in any bounded set, so suppose $d(h, g^n h) > m$, where n and m are large. Then one can find a point h' on the geodesic from h to $g^n h$ such that connecting the powers $g^k h'$ by geodesic segments produces a quasigeodesic. (Proof omitted: try it.)

We will call this invariant quasigeodesic a **quasi-axis** for g. Note that it is asymptotic to a pair of points on the boundary, and by visibility there is also a geodesic between those points. So one can choose between a quasi-geodesic that is g–invariant and a true geodesic that is nearly g–invariant.

Proposition 14. Any two quasi-axes for g are bounded distance apart (for a bound depending on δ).

Proof. Consider two quasi-axes, and let d be the minimum distance from a point on α to a point on β. Then the Hausdorff distance between the axes is at most $d + 2C$, where the points $g^n p$ are C–dense on the axes. Now consider very distant points s, t on α and their closest points s', t' on β. The geodesic quadrilateral on s, s', t', t is 2δ–thin, and it is easy to see that most of the long sides must be 2δ–close to each other. But α, β, are quasigeodesic, so are close to the true geodesics between any of their points. □

It's too much to hope for quasi-axes without hyperbolicity, but you can generalize the idea somewhat by defining an **asymptotic translation length**: for any metric space and any isometry, set $\tau(g) := \lim_{n\to\infty} \frac{d(p, g^n p)}{n}$. (It follows from the triangle inequality that this exists and is independent of p.)

Gromov observed that in hyperbolic groups, the translation lengths are discrete; indeed, there is an N depending on δ so that all $\tau(g)$ are in $\frac{1}{N}\mathbb{Z}$. A simple example of non-integer translation lengths is given by taking $\mathbb{Z}$ with the non-standard generating set $\pm\{1, N\}$. Then $d(0, n) \approx n/N$ for large n, so $\tau(1) = \lim \frac{n/N}{n} = \frac{1}{N}$. Similarly, $\tau(j) = \frac{|j|}{N}$. Thanks to Chris Leininger for this example.

Proposition 15. Hyperbolic groups have no subgroups isomorphic to $\mathbb{Z}^2$.

Very loosely, the reason is that if $[g, h] = 1$, then h permutes the quasi-axes of g, so it nearly fixes a quasi-axis, so it nearly behaves like g^n. One can make this precise to see that the centralizer $C(g)$ is virtually $\langle g \rangle$.

This is very strong! Note that hyperbolic groups can certainly have non-hyperbolic subgroups, if they are embedded in a very distance-distorting way. However, $\mathbb{Z}^2$ is a forbidden subgroup, no matter how embedded.

What about the converse? Moussong proved that a Coxeter group is hyperbolic iff it contains no $\mathbb{Z}^2$ subgroup.

The **Bestvina-Bridson wager** asks if the same is true for fundamental groups of compact nonpositively curved manifolds. (Note: this is now established in dimension three, by geometrization, but open in higher dimension.)

So we have found that the action of a non-torsion element on the group has **North-South dynamics**: a quasi-axis connecting a repelling to an attracting fixed point on the boundary.

The three-fold classification of isometries of $\mathbb{H}$ has also been exported to the non-hyperbolic setting. For instance, consider the **mapping class group** of a surface: the discrete group of (orientation-preserving) diffeomorphisms, up to isotopy. (It turns out to be finitely presented.) This has a nice isometric action on a topological ball called the Teichmüller space. Then there is a classic theorem of Thurston's: any mapping class is either finite-order (corresponding to elliptic case), reducible (corresponding to parabolic case), or has a quasi-axis and North-South dynamics. This last kind is called **pseudo-Anosov**. These are classified by their fixed points on an appropriate boundary sphere.

Question (de la Harpe): Can $SL_3(\mathbb{Z})$ act on anything with N-S dynamics?

9.2. **Applications of N-S dynamics.** One important application is **ping pong lemmas**: having N-S dynamics allows you to find free subgroups of your isometry group. The classic case is again in the hyperbolic plane. One takes two different hyperbolic isometries, such that the boundary fixed points $\alpha^+, \alpha^-, \beta^+, \beta^-$ are all distinct. Then, by passing to high enough powers, there are neighborhoods A^+, A^-, B^+, B^- such that one isometry takes everything outside A^- into A^+ and the other takes everything outside B^- into B^+. Once you're in this setting, it's easy to see that any nontrivial string in these isometries a and b can't equal the identity: it either ends in a or b, so it either maps the whole space into A^+ or B^+. Thus $\langle a, b\rangle$ is a free subgroup of isometries.

9.3. **Ray approximation.** The important **Multiplicative Ergodic Theorem** says that for a random walk on a symmetric space with finite first moment, there is a value A (called the first **Lyapunov exponent**) such that for almost every sample path $\omega \in \Omega$, $\lim_{n\to\infty} \frac{1}{n} d(x_0, g_n x_0) = A$; if $A > 0$ (such random walks are called **ballistic**), then for a.e. ω there is a geodesic ray γ with

$$\lim_{n\to\infty} \frac{1}{n} d(g_n x_0, \gamma(An)) = 0$$

for $g_n = g(\omega)g(T\omega)\cdots g(T^{n-1}\omega)$, the product of random group elements. This says that for this dominant speed A, almost every sequence of points in the random walk has some geodesic ray such that the points, though they may seem to be scattered all about, actually are sublinearly well-approximated by steady progress along γ with speed A. Let us call this **geodesic tracking** or **ray approximation**. (See figure below.) This was initially proved for the $GL_n(\mathbb{R})$ case by Oseledets in the 1960s, then interpreted geometrically and extended to general symmetric spaces by Kaimanovich.

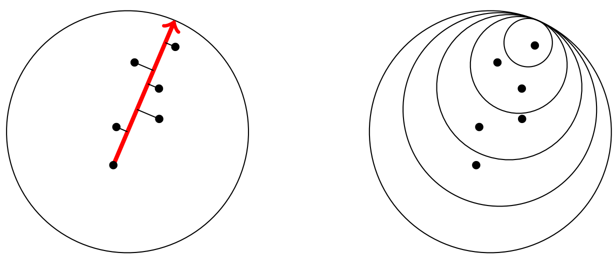
Kaimanovich extends this to random walks on hyperbolic groups. Karlsson-Margulis extend to non-positive curvature. And it works as well for the mapping class group action on Teichmüller space (Duchin, Tiozzo).

9.4. **Ergodic theorem for isometries.** Here is a massively more general setup that should be thought of as similar in flavor. Even when your random walk isn't following a geodesic, it is regularly traversing the level sets of some horofunction.

Theorem 16 (Karlsson-Ledrappier)**.** Let G be a locally compact group acting by isometries on a proper space X, and consider a random walk with finite first moment. Then for almost every sample path $\omega \in \Omega$, there is some horofunction h_ω such that

$$\lim_{n\to\infty} \frac{1}{n} d(x_0, g_n x_0) = -\lim_{n\to\infty} \frac{1}{n} h_\omega(g_n x_0).$$

Here's how you should think about this: in the ray approximation scenario above, the points of the random walk were falling sublinearly close to a geodesic ray, so there was a point on the visual boundary towards which they were making steady progress. This theorem says that even if you're in a metric space with a terrible visual boundary, there is still some point in the horofunction boundary towards which you are making steady progress.



10. More randomness

10.1. **Random group elements.** Let us define various notions of the **asymptotic density** of a property P, associated to corresponding ways to pick out a random element of a group.

Given (G, S), suppose G acts on X and identify G with its orbit. Identify property P with a subset of G (for instance, P could be "finite-order," and then that subset would contain the torsion elements). Define

$$\mathrm{Prob}(P) := \lim_{n\to\infty} \frac{\#(B_n^S \cap P)}{\# B_n^S}; \qquad \mathrm{GProb}(P) := \lim_{n\to\infty} \frac{\#(\mathrm{Ball}_n^X \cap P)}{\#(\mathrm{Ball}_n^X \cap G)}.$$

At least a priori (and usually in fact), Prob depends on the genset S and GProb depends on X. A third way to randomly pick group elements, rather than uniformly in balls in the Cayley graph or model space, is via random walks. We can let $\mathrm{RProb}(P) := \lim_{n\to\infty} \mathbb{P}(g_n \in P)$, measuring the probability that the nth sample point in the random walk has property P.

The literature contains many beautiful theorems about RProb. For actions that are well-understood, there are also results on GProb. There are many results on Prob for free groups, where it is not that different from RProb, but fairly few other than that. This is because for complicated groups it is very hard to build up the Cayley graphs finely enough to solve these counting problems.

For instance, it has long been suspected that in the Thurston classification of mapping classes, "almost every mapping class is pseudo-Anosov." (This is corroborated by experimental evidence of Dunfield and D. Thurston, which shows that with respect to a finite set of Dehn twist generators, it typically only takes a product of about length three before you can expect to see a pseudo-Anosov!) The genericity of pseudo-Anosovs is now a theorem of Maher for RProb; it is established by Maher and by Masur for GProb with respect to the action on Teichmüller space; and it is still wide open for Prob.

10.2. **Nilpotent groups and limit metrics.** An exception to the rule that Prob is too hard to work with is in nilpotent groups. For example, a theorem of P. Dani states that in virtually nilpotent groups, Prob(torsion) is independent of S and ranges over all of $\mathbb{Q} \cap [0, 1)$ as the group varies!

A major tool for asymptotic density results in nilpotent groups is found in work of Pansu, which says that the large-scale structure of the discrete group is the same as a certain metric on an ambient Lie group. Thus for appropriately homogeneous properties P, Prob in the discrete group agrees with GProb in that Lie group with respect to the limit metric.

A special case where things work out particularly nicely is for free abelian groups $\mathbb{Z}^d$. In that case, if S is the (symmetric) generating set, then we can let Q be its convex hull in $\mathbb{R}^d$ and let L be its boundary. Then one can show that there is a constant K depending on S so that $(n - K)Q \cap \mathbb{Z}^d \subseteq B_n \subseteq nQ \cap \mathbb{Z}^d$. In other words, the ball of radius n is very nearly the same as the set of lattice points in the polytope nQ. So we can see a *limit shape* emerging: if we take dilates $\frac{1}{n} B_n$, they converge to the polytope Q, and accordingly the word metric is well-approximated by the norm on $\mathbb{R}^d$ for which Q is the unit ball. Furthermore, this also ensures that the counting measure on $\frac{1}{n} B_n$ converges to the volume measure on Q. Therefore if P itself is a property that behaves well under rescaling, we find that Prob in the group is just GProb in the normed space. In fact, these probabilities will depend just on the limit shape Q. Small example to illustrate the point: in $(\mathbb{Z}^2, S)$, the probability that a group element has a geodesic spelling using exactly two generators equals $r/2A$, where r is the number of sides of the polygon Q and A is its area!

10.3. **Statistical stability of geodesics.** In a geodesic space X, with basepoint x_0, let $\mathcal{G}(x)$ the set of all geodesics $\overline{x_0 x}$, where the distance between two geodesics is their Hausdorff distance, which is always at most $d(x, x_0)/2$. Define an *instability function* $\mathcal{I}(x) := 2 \cdot \frac{diam\, \mathcal{G}(x)}{d(x,x_0)}$, normalized so that $0 \leqslant \mathcal{I}(x) \leqslant 1$ for all x. For example, in $(\mathbb{Z}^2, \mathsf{std})$, we have $\mathcal{I}(a, 0) = 0$ but $\mathcal{I}(a, a) = 1$.

Then let the **stability index** of a group G with respect to a genset S be

$$\mathsf{Stab}(G, S) := \lim_{\epsilon\to 0} \mathrm{Prob}\big(\mathcal{I}(x) \leqslant \epsilon\big).$$

This measures the proportion of points for which the distance between possible geodesics from the basepoint is less than any linear function of their length.

Theorem 17. With respect to any finite generating set, the density of sublinearly stable points is

$$\mathsf{Stab}(G,S) = \begin{cases} 0, & G = \mathbb{Z}^d, \\ p/q, & G = H(\mathbb{Z}), \\ 1, & G \text{ unbounded, } \delta\text{-hyperbolic.} \end{cases}$$

Here, p/q is a rational number that depends nontrivally on S and can be computed precisely. For the standard generators, the stability index is 19/31. (It is easy to find generating sets for which the stability index gets arbitrarily close to 1. Conjecturally, 19/31 is minimal.)

10.4. **Random groups.** My treatment here follows Ollivier's survey, but I've tried to update the summary of density results in the following chart to be current with the literature.

Fix an m and consider groups generated by $S = \{a_1, \dots, a_m\}$. Then in the free group F_m, the size of the ℓ–sphere is $S_\ell \approx (2m-1)^\ell$. For any *density* $0 \leqslant d \leqslant 1$, we'll choose uniformly at random a subset $R \subset S_\ell$ having exactly $(2m-1)^{d\ell}$ elements. (So if $d = 1/2$, we took the square root of the total possible number.) (Note we could use B_ℓ instead of S_ℓ and it wouldn't be much different, since growth is exponential.)

Then a **random group** is $G = \langle a_1, \dots, a_m \mid R\rangle$ for a random relator set R. We can now talk about statistical properties of random groups by letting $\ell \to \infty$.

Proposition 18. Let R be a random set of relators at density d, at length ℓ. Fix $0 \leqslant \alpha < d$. Then with probability 1, every reduced word of length $\alpha\ell$ appears as a subword of some word in R.

This relates in a nice way to the **small cancellation conditions** that have been studied in combinatorial group theory for a very long time: fix $0 \leqslant \alpha \leqslant 1$ and for a presentation $\langle S|R\rangle$, replace R with its closure under cyclic rewriting. For every pair of relators, let u_{ij} be the maximal initial string on which they agree; call these the *pieces* of the presentation. Then the presentation is $C'(\alpha)$ if all pieces are less than α proportion of the relators in which they appear. Then it is a classical result that $C'(1/6)$ implies that the given presentation is a Dehn presentation, which as we have seen means that G is hyperbolic. So the proposition and a bit more work shows that at density 1/12 or less, the group should be hyperbolic. But in fact one can do better.

Proposition 19 (Probabilistic pigeon-hole principle)**.** For any $\epsilon > 0$, if $N^{\frac{1}{2}+\epsilon}$ balls are randomly put into N boxes, then there is almost sure to be a box with at least two balls, as $N \to \infty$.

This makes it almost sure that we have two of the same relator, and similarly almost sure that we have a relator and another differing from it in one place. If $r_1 = wa_1$ and $r_2 = wa_2$ are both relators, then $a_1 = a_2$. Eventually, this will cause all generators to be identified which each other and their inverses, so G is trivial or $\mathbb{Z}/2\mathbb{Z}$.

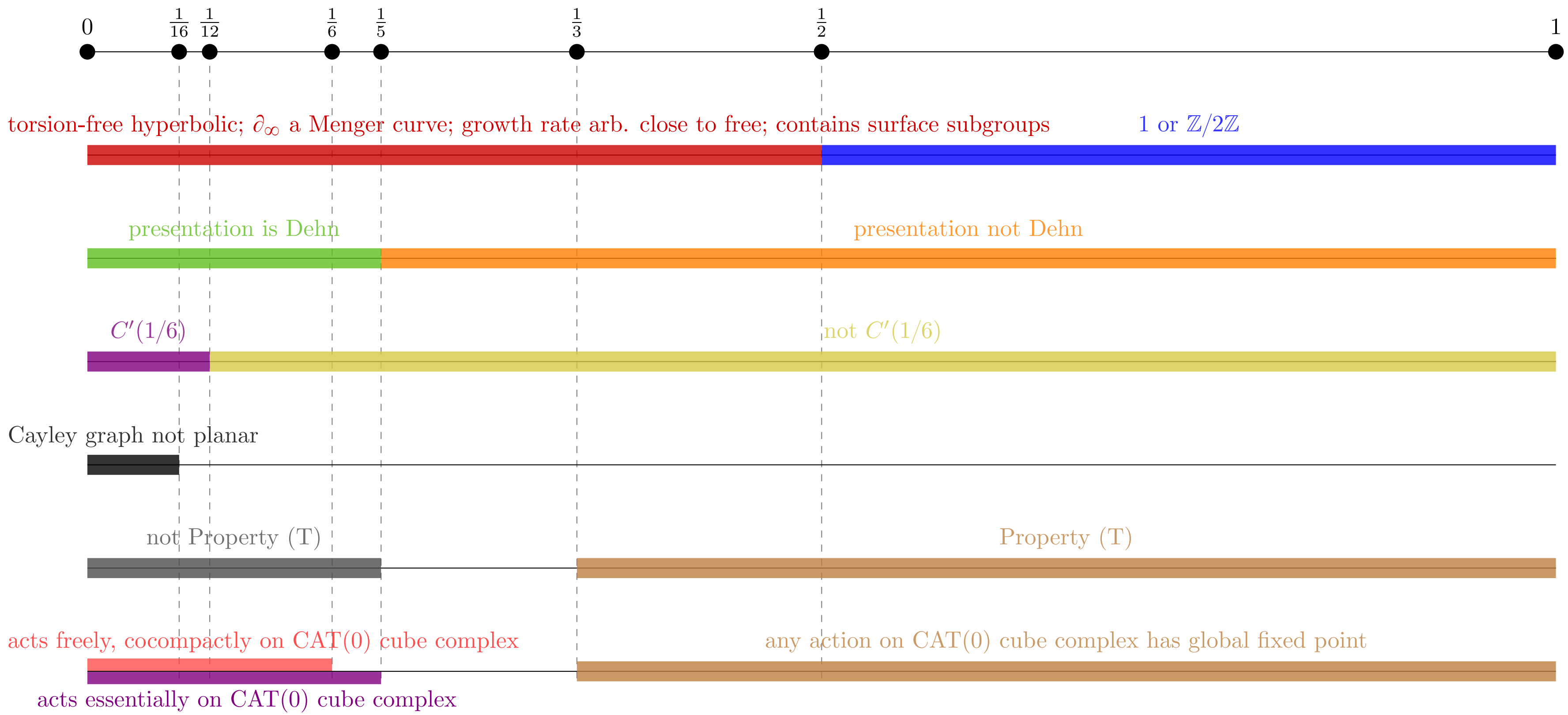
Theorem 20 (Gromov)**.** For random groups with $d > 1/2$, almost surely $G = 1$ or $\mathbb{Z}/2\mathbb{Z}$. For random groups with $d < 1/2$, almost surely G is (infinite) torsion-free hyperbolic, of geometric dimension 2.

The geometric dimension is the smallest dimension of a $K(G,1)$ complex, so in particular these random groups are not elementary (virtually $\mathbb{Z}$). And now a great deal more is known about these random groups at density $< 1/2$, such as results about their growth rates, spectral gaps, boundaries, subgroups, and so on. There is even a bound on the hyperbolicity constant: $\delta \leqslant 4\ell/(1-2d)$ in the thin triangles definition.

Right now the most mysterious density range is $1/5 < d < 1/3$, where there is a conspicuous gap in understanding how G can act. (See below.) Another question of some interest is to characterize what happens at $d = 1/2$. Here there is more sensitive dependence on features of the model such as the choice of m (the number of generators) and whether relators are selected from S_ℓ or B_ℓ, neither of which matters at other densities. For instance, the probability that $G = 1$ at density 1/2 is strictly between 0 and 1 and depends on m, which is unpleasant. To make this situation nicer one can "tune" $d \to 1/2$ with $\ell \to \infty$ and try to find interesting limiting properties. There is still a lot to do in the study of random groups....

Density diagram

Almost sure properties of random groups at densities $0 \leqslant d \leqslant 1$.



(Attributions: some combinations of Gromov, Ollivier, Wise, Żuk, Calegari, Walker, Arzhantseva, Cherix, Przytyczki, Dahmani, Guirardel...)