Geometric algorithms for discreteness and faithfulness

Michael Kapovich

Abstract. In this paper I describe two geometric algorithms for certifying discreteness and freeness of finitely generated subgroups of $O(n,1)$, $SL(n,\mathbb{R})$ and, more generally, algorithms for discreteness and faithfulness of certain linear representations of finitely-presented groups.

1. Introduction

The goal of this paper is to describe two algorithms[1] for certifying discreteness of finitely generated subgroups of $SL(n,\mathbb{R})$ and, more generally, algorithms for discreteness and faithfulness of certain representations of finitely-presented groups. The fundamental questions that these algorithms aim to address are:

Question 1.1. *1. Suppose that Γ is a hyperbolic group defined via its finite presentation. Given a homomorphism $\rho : \Gamma \to SL(n,\mathbb{R}) = G$, determine if ρ has finite kernel and discrete image.*

2. Given matrices $A_1, ..., A_k \in G$ determine if the subgroup $\Gamma = \langle A_1, ..., A_k\rangle$ generated by these elements is discrete and/or free of rank k.

There is a separate issue as to what an *algorithm* even means in this setting. One approach is to work with the BSS (Blum–Schub–Smale) or *Real Ram* model of computability over the real numbers. Another approach is to assume that Γ lies in $SL(n,F)$, where F is a number field, e.g. Γ is a subgroup of an arithmetic group. We refer the reader to the papers by Jane Gilman [**G2, G3**] and the author, [**K1**], for discussion of the problems one is facing here. In this paper, we will ignore these foundational issues and concentrate on the geometric side of the problem. We also refer to the paper by Gilman and Maskit [**GM**] as well as other papers by Gilman, [**G1, G2, G3**] for the description and discussion of a very different geometric algorithm for discreteness of subgroups of $PSL(2,\mathbb{R})$.

In the first part of the paper, we discuss geometric algorithms dealing with the case of subgroups of $G = O^+(n,1)$ where much is known. In the second part, we discuss the general case, which is far less studied.

Acknowledgement. I am grateful to the referees of the paper for doing very thorough refereeing job.

2. Basic hyperbolic geometry

All the material of this and the next two sections is standard; proofs can be found for instance in Ratcliffe's book [**Ra**].

We let V be an $n+1$-dimensional real vector space equipped with the Lorentzian bilinear form $\langle \cdot, \cdot \rangle$ of signature $(n,1)$. Concretely, $V = \mathbb{R}^{n+1}$ and

$$\langle \mathbf{x}, \mathbf{y}\rangle = -x_0y_0 + x_1y_1 + ... + x_ny_n.$$

We let V^- denote the subset of V consisting of future-directed (i.e. satisfying $\langle \mathbf{x}, e_0\rangle > 0$) and negative (also known as *time-like*, i.e. satisfying $\langle \mathbf{x}, \mathbf{x}\rangle < 0$) vectors in V. We will be identifying the hyperbolic n-space $\mathbb{H}^n$ with the *imaginary unit sphere* in V^-:

$$\mathbb{H} = \{\mathbf{x} \in V^- : \langle \mathbf{x}, \mathbf{x}\rangle = -1\}.$$

The group $O(n,1) = O(V, \langle \cdot, \cdot \rangle)$ of linear transformations preserving the form $\langle \cdot, \cdot \rangle$ is disconnected. We let $O^+(n,1)$ denote the index two subgroup of $O(n,1)$ preserving the cone V^-. This subgroup also preserves the imaginary unit sphere $\mathbb{H}$ and equals the isometry group of the hyperbolic n-space $\mathbb{H}^n$.

1991 *Mathematics Subject Classification.* Primary 22E40, 20-08; Secondary 20F67, 53C35.

Key words and phrases. Discrete groups, symmetric spaces.

[1] I am not discussing here various versions of the *ping-pong* argument, since this argument is widely known.

By abusing the notation, we call vectors $\mathbf{x}$ satisfying $\langle \mathbf{x}, \mathbf{x}\rangle = -1$, the *unit vectors.* The *tangent space* $T_{\mathbf{x}}\mathbb{H}$ to $\mathbb{H}$ at $\mathbf{x} \in \mathbb{H}$ is defined as the space of vectors orthogonal to $\mathbf{x}$:

$$T_{\mathbf{x}}\mathbb{H} = \{\mathbf{v} \in V : \langle \mathbf{x}, \mathbf{v}\rangle = 0\}.$$

All nonzero vectors in this tangent space are *space-like*, i.e. satisfy

$$\langle \mathbf{v}, \mathbf{v}\rangle > 0.$$

For space-like vectors we have the Lorentzian norm $|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$, while for time-like vectors we also have the "norm" $|\mathbf{v}| = \sqrt{-\langle \mathbf{v}, \mathbf{v}\rangle}$. The *angle* $\alpha = \angle(\mathbf{u}, \mathbf{v})$ between space-like vectors is defined by the usual formula:

$$\langle \mathbf{u}, \mathbf{v}\rangle = \cos(\alpha)|\mathbf{u}| \cdot |\mathbf{v}|.$$

It is also convenient to use for points in $\mathbb{H}^n$ the equivalence classes of vectors in V^-, where two vectors are equivalent if they are multiples of each other. For a vector $\mathbf{x} \in V^-$, we define its 'normalization', the unit vector

$$\bar{\mathbf{x}} = \frac{\mathbf{x}}{|\mathbf{x}|}.$$

Hyperbolic distance. For $\mathbf{x}, \mathbf{y} \in V^-$ we have the following formula for the *hyperbolic distance* $d(\mathbf{x}, \mathbf{y})$:

$$\cosh d(\mathbf{x}, \mathbf{y}) = -\frac{\langle \mathbf{x}, \mathbf{y}\rangle}{|\mathbf{x}| \cdot |\mathbf{y}|}.$$

To be more precise, this formula defines the hyperbolic distance between the normalizations $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in \mathbb{H}$.

The *geodesic segment* $\mathbf{xy}$ between points $\mathbf{x}, \mathbf{y} \in \mathbb{H}$ is defined as

$$\mathbf{xy} = \{\bar{\mathbf{z}} : \mathbf{z} = (1-t)\mathbf{x} + t\mathbf{y}, t \in [0, 1]\}.$$

Accordingly, the *midpoint* $m(\mathbf{x}, \mathbf{y})$ (the point on $\mathbf{xy}$ dividing this segment in two equal parts) is given by

$$m(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}+\mathbf{y}}{|\mathbf{x}+\mathbf{y}|}. \tag{2.1}$$

We also define the (space-like) vector $\mathbf{v} = \mathbf{v}(\mathbf{x}, \mathbf{y})$ as a (not necessarily unit) vector tangent to $\mathbf{xy}$ at the point $\mathbf{x}$:

$$\mathbf{v} = \mathbf{v}(\mathbf{x}, \mathbf{y}) = \mathbf{y} + \langle \mathbf{x}, \mathbf{y}\rangle\mathbf{x}.$$

Then

$$|\mathbf{v}|^2 = \langle \mathbf{x}, \mathbf{y}\rangle^2 - 1.$$

Hyperbolic angle. The *hyperbolic angle* $\alpha = \angle_x(\mathbf{u}, \mathbf{v})$ between two nonzero tangent vectors $\mathbf{u}, \mathbf{v} \in T_{\mathbf{x}}\mathbb{H}$ at $\mathbf{x} \in \mathbb{H}$ is given, as above, by the formula

$$\langle \mathbf{u}, \mathbf{v}\rangle = \cos(\alpha)|\mathbf{u}| \cdot |\mathbf{v}|.$$

Lastly, the hyperbolic angle $\alpha = \angle\mathbf{yxz}$ of a hyperbolic triangle $\mathbf{xyz} \subset \mathbb{H}$ at the vertex $\mathbf{x}$ is defined by the formula

$$\alpha = \angle(\mathbf{v}(\mathbf{x}, \mathbf{y}), \mathbf{v}(\mathbf{x}, \mathbf{z})),$$

i.e.

$$\cos(\alpha) = \frac{\langle \mathbf{v}(\mathbf{x}, \mathbf{y}), \mathbf{v}(\mathbf{x}, \mathbf{z})\rangle}{|\mathbf{v}(\mathbf{x}, \mathbf{y})|\ |\mathbf{v}(\mathbf{x}, \mathbf{z})|} = \frac{\langle \mathbf{y}, \mathbf{z}\rangle + \langle \mathbf{x}, \mathbf{y}\rangle\langle \mathbf{x}, \mathbf{z}\rangle}{(\langle \mathbf{x}, \mathbf{y}\rangle^2 - 1)^{1/2}(\langle \mathbf{x}, \mathbf{z}\rangle^2 - 1)^{1/2}}$$

REMARK 2.1. *Computing (or, at least, estimating from below)* $\sin(\alpha/2)$ *will be useful for the algorithms described in sections 8, 9.*

For instance, $\alpha \geqslant \pi/2$ if and only if

$$\langle \mathbf{y}, \mathbf{z}\rangle + \langle \mathbf{x}, \mathbf{y}\rangle\langle \mathbf{x}, \mathbf{z}\rangle \leqslant 0.$$

Projectivizing the negative cone V^- one obtains the *Klein model* of the hyperbolic space $\mathbb{H}^n$, which is an open ball B in the projective space $\mathbb{R}P^n = PV$: The ball B is the projection $\pi(V^-)$ of V^- to the projective space PV. The boundary sphere S^{n-1} of this ball is the projectivization of the null-cone $\{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{v}\rangle = 0\}$ of *light-like* vectors, also known as *null-vectors.* The hyperboloid $\mathbb{H} \subset V^-$ projects diffeomorphically to B, making B a model of the hyperbolic n-space.

Let $\xi = \pi(\mathbf{v})$, be a point of the boundary sphere S^{n-1} of the ball B. Without loss of generality we can assume that $\mathbf{v}$ is a nonzero vector which belongs to the closure of the cone V^-. We next define *horospheres* $\Sigma \subset \mathbb{H}$ *based* at ξ. Given $\mathbf{v}$, consider the affine hyperplane

$$\{\mathbf{x} \in V : \langle \mathbf{x}, \mathbf{v}\rangle = 1\}.$$

Replacing $\mathbf{v}$ by its positive multiple results in another hyperplane, parallel to the one defined before. Intersecting such hyperplanes with $\mathbb{H}$ yields a foliation of $\mathbb{H}$ by hypersurfaces, called *horospheres* $\Sigma \subset \mathbb{H}$ based at ξ.

3. Hyperbolic bisectors

Given two distinct points $\mathbf{p}, \mathbf{q} \in \mathbb{H}^n$, the *bisector* $Bis(\mathbf{p}, \mathbf{q})$ of the geodesic segment $\mathbf{pq}$ in $\mathbb{H}^n$ is the collection of all points which are equidistant from $\mathbf{p}$ and $\mathbf{q}$:

$$Bis(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathbb{H}^n : d(\mathbf{p}, \mathbf{x}) = d(\mathbf{q}, \mathbf{x})\}.$$

In terms of Lorentzian geometry, when $\mathbf{p}, \mathbf{q}$ are in the hyperboloid $\mathbb{H}$,

$$Bis(\mathbf{p}, \mathbf{q}) = \mathbb{H} \cap \{\mathbf{x} \in V : \langle \mathbf{p}, \mathbf{x}\rangle = \langle \mathbf{q}, \mathbf{x}\rangle\},$$

with

$$\{\mathbf{x} \in V : \langle \mathbf{p}, \mathbf{x}\rangle = \langle \mathbf{q}, \mathbf{x}\rangle\} = (\mathbf{p} - \mathbf{q})^{\perp},$$

the latter is the Lorentzian orthogonal complement to the time-like vector $\mathbf{p} - \mathbf{q}$. Let us determine when two bisectors $Bis(\mathbf{p}, \mathbf{q}_1), Bis(\mathbf{p}, \mathbf{q}_2)$ have empty intersection in $\mathbb{H}$. The empty intersection condition is equivalent to the property that all nonzero elements of the (typically) codimension 2 linear subspace

$$(\mathbf{p} - \mathbf{q}_1)^{\perp} \cap (\mathbf{p} - \mathbf{q}_2)^{\perp}$$

are null or space-like vectors. Suppose, for a moment, that one of the vectors $\mathbf{x}$ of this intersection is time-like. Then the Gram matrix of the Lorentzian bilinear form restricted to $\mathrm{span}(\mathbf{p} - \mathbf{q}_1, \mathbf{p} - \mathbf{q}_2, \mathbf{x})$, in terms of the basis $\{\mathbf{v}_1 = \mathbf{p} - \mathbf{q}_1, \mathbf{v}_2 = \mathbf{p} - \mathbf{q}_2, \mathbf{x}\}$, equals

$$\begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1\rangle & \langle \mathbf{v}_1, \mathbf{v}_2\rangle & 0 \\ \langle \mathbf{v}_1, \mathbf{v}_2\rangle & \langle \mathbf{v}_2, \mathbf{v}_2\rangle & 0 \\ 0 & 0 & \langle \mathbf{x}, \mathbf{x}\rangle \end{bmatrix},$$

where $\langle \mathbf{x}, \mathbf{x}\rangle < 0, \langle \mathbf{v}_i, \mathbf{v}_i\rangle > 0, i = 1, 2$. Since the signature of the restriction cannot be $(1, 2)$, the submatrix

$$\begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1\rangle & \langle \mathbf{v}_1, \mathbf{v}_2\rangle \\ \langle \mathbf{v}_1, \mathbf{v}_2\rangle & \langle \mathbf{v}_2, \mathbf{v}_2\rangle \end{bmatrix}$$

has to be positive semidefinite, which translates to the inequality

$$\langle \mathbf{v}_1, \mathbf{v}_1\rangle\langle \mathbf{v}_2, \mathbf{v}_2\rangle - \langle \mathbf{v}_1, \mathbf{v}_2\rangle^2 \geqslant 0.$$

The inequality is strict unless $\mathbf{v}_1 = \mathbf{v}_2$, i.e. $\mathbf{q}_1 = \mathbf{q}_2$, which we assume not to be the case. We thus arrive to:

Lemma 3.1. *The intersection* $Bis(\mathbf{p}, \mathbf{q}_1) \cap Bis(\mathbf{p}, \mathbf{q}_2)$ *is empty if and only if*

$$\langle \mathbf{v}_1, \mathbf{v}_1\rangle\langle \mathbf{v}_2, \mathbf{v}_2\rangle - \langle \mathbf{v}_1, \mathbf{v}_2\rangle^2 \leqslant 0.$$

Furthermore, the intersection $(\mathbf{p} - \mathbf{q}_1)^{\perp} \cap (\mathbf{p} - \mathbf{q}_2)^{\perp}$ *consists entirely of space-like vectors (and the vector* 0*) if and only if*

$$\langle \mathbf{v}_1, \mathbf{v}_1\rangle\langle \mathbf{v}_2, \mathbf{v}_2\rangle - \langle \mathbf{v}_1, \mathbf{v}_2\rangle^2 < 0.$$

4. Isometries of the hyperbolic space

Isometries of the hyperbolic space $\mathbb{H}$ are elements of the group $O^+(n, 1)$: They are the linear transformations preserving $\mathbb{H}$. As many things in this world, isometries of the hyperbolic space fall into three groups:

- **Hyperbolic** (also called loxodromic). Every hyperbolic matrix A has two[2] distinct positive real eigenvalues $\lambda > 1, \lambda^{-1} \in (0, 1)$, of multiplicity one, and corresponding eigenvectors $\mathbf{v}_{\pm}$ which are null-vectors in V. These vectors span a 2-dimensional subspace H_A in V, whose intersection with $\mathbb{H}$ is a complete hyperbolic geodesic h_A (a hyperbola in H_A), called *the axis* of A: This geodesic is preserved by A. These isometries will be most important for us.
- **Parabolic**. These isometries A have exactly one (up to scaling) null eigenvector in V and it is fixed by A. (In other words, each parabolic isometry has exactly one fixed point in the boundary sphere S^{n-1} of the Klein model of $\mathbb{H}^n$.) All eigenvalues of A have absolute value 1.
- **Elliptic**. This is everything else, but one can also define these isometries A by the condition that they have fixed vectors in V^- (equivalently, in $\mathbb{H}$). All eigenvalues of A again have absolute value 1.

Taking powers does not change the type of an isometry, but taking products, in general, does. Each parabolic isometry fixes a unique point $\xi \in S^{n-1}$ and can be shown to preserve each horosphere in $\mathbb{H}^n$ based at ξ. In contrast, hyperbolic isometries of $\mathbb{H}^n$ do not preserve any horospheres. An elliptic isometry has an invariant horosphere if and only if it fixes a point in S^{n-1}. It then preserves all horospheres based at that point.

Displacement. Each hyperbolic matrix A acts on its axis h_A as a *translation*[3] by some number τ_A, i.e.

$$d(A\mathbf{x}, \mathbf{x}) = \tau_A$$

[2]Besides these two real eigenvalues there will be other (complex) eigenvalues, but they all have absolute value 1.

[3]After we identify h_A with the real line by the hyperbolic arc-length parameterization.

for all $\mathbf{x} \in h_A$. The number τ_A is computable in terms of the eigenvalues of A:

$$\cosh(\tau_A) = \frac{\lambda + \lambda^{-1}}{2}.$$

Checking hyperbolicity is easy: One computes the eigenvalues and checks if one of them is > 1. This works especially nicely for elements in arithmetic subgroups since there is a uniform lower bound on $\tau(A)$ for hyperbolic elements A of such groups, defined in terms of the arithmetic data of Γ. Conjecturally, for arithmetic subgroups there is a uniform positive lower bound on the displacements, depending only on the dimension:[4]

CONJECTURE 4.1. *For every $n \geqslant 2$ there exists $t(n) > 0$ such that for every arithmetic lattice $\Gamma < O^+(n,1)$, every hyperbolic element $\gamma \in \Gamma$ satisfies $\tau_\gamma > t(n)$.*

Note that this conjecture fails if we do not restrict to arithmetic lattices.

5. Connecting hyperbolic geodesics

In this section I discuss how to connect two hyperbolic geodesics h_1, h_2 in $\mathbb{H}^n$ by a geodesic segment s meeting both orthogonally at its end-points. Such a segment is the shortest segment connecting the two geodesics (unless the geodesics intersect in $\mathbb{H}^n$, which we will assume not to be the case in what follows). The segment s is known to be unique. [5] The segment s exists, unless two geodesics h_1, h_2 are asymptotic to a common point in the boundary sphere of $\mathbb{H}^n$. We will be assuming that geodesics h_i are given as the intersections of $\mathbb{H}$ with the linear 2-dimensional subspaces $W_i = \mathrm{span}(\mathbf{u}_i, \mathbf{v}_i)$, where $\mathbf{u}_i, \mathbf{v}_i$ are (linearly independent) null-vectors in the future null-cone. Then, testing for the existence of a common asymptotic point in S^{n-1} is easy: One simply verifies if two of the four null-vectors $\{\mathbf{u}_1, \mathbf{v}_1, \mathbf{u}_2, \mathbf{v}_2\}$ are multiples of each other. Checking if $h_1 \cap h_2$ is nonempty is also easy: One computes the intersection $W_1 \cap W_2$ (which, generically, if $n \geqslant 3$, will be zero) and checks if this intersection contains a (nonzero) time-like vector.

REMARK 5.1. *Of prime importance for us is the case when $W_i = H_{A_i} = \mathrm{span}(\mathbf{u}_i, \mathbf{v}_i)$, where A_i are hyperbolic elements of $O(n,1)$ and $\mathbf{u}_i, \mathbf{v}_i$ are their light-like future-directed eigenvectors.*

We first work out a condition for orthogonality of two geodesics in $\mathbb{H}^n$ meeting at a common point, as is the case for, say, h_1 and the unique hyperbolic geodesic containing s. Fix $\mathbf{y}_1 = t_1\mathbf{u}_1 + (1-t_1)\mathbf{v}_1$, $t_1 \in (0,1)$, a time-like vector in W_1. (Its normalization $\bar{\mathbf{y}}_1$ is a point in $h_1$and all points in h_1 appear this way.) Then in W_1 we find a (space-like) vector $\mathbf{y}_1^*$ orthogonal to $\mathbf{y}_1$, given by

$$\mathbf{y}_1^* = t_1\mathbf{u}_1 + (t_1 - 1)\mathbf{v}_1.$$

Suppose now that W is the plane spanned by a time-like vector $\mathbf{y}_1 = t_1\mathbf{u}_1 + (1-t_1)\mathbf{v}_1$ (as above) and another time-like vector $\mathbf{y}_2$. Then the hyperbolic geodesics h_1, h in $\mathbb{H}$ defined as $h_1 = W_1 \cap \mathbb{H}$, $h = W \cap \mathbb{H}$, are orthogonal if and only if the vector $\mathbf{y}_1^* \in W_1$ (Lorentzian-orthogonal to $\mathbf{y}_1$) is also Lorentzian-orthogonal to the entire plane W, equivalently, is Lorentzian-orthogonal to $\mathbf{y}_2$. The latter orthogonality condition is

$$\langle t_1\mathbf{u}_1 + (1-t_1)\mathbf{v}_1, \mathbf{y}_2\rangle = 0.$$

This gives us a way to compute the hyperbolic segment s connecting h_1, h_2 orthogonally. Namely, we search for vectors $\mathbf{y}_i = t_i\mathbf{u}_i + (1-t_i)\mathbf{v}_i$, $t_i \in (0,1)$, $i = 1,2$, satisfying the two equations:

$$\langle \mathbf{y}_1^*, \mathbf{y}_2\rangle = \langle \mathbf{y}_2^*, \mathbf{y}_1\rangle = 0,$$

equivalently,

$$\langle t_1\mathbf{u}_1 + (t_1-1)\mathbf{v}_1, t_2\mathbf{u}_2 + (1-t_2)\mathbf{v}_2\rangle = \langle t_2\mathbf{u}_2 + (t_2-1)\mathbf{v}_2, t_1\mathbf{u}_1 + (1-t_1)\mathbf{v}_1\rangle = 0.$$

Then the common perpendicular segment to h_1, h_2 equals $s = \bar{\mathbf{y}}_1\bar{\mathbf{y}}_2 \subset \mathbb{H}$. Searching for vectors $\mathbf{y}_1, \mathbf{y}_2$ amounts to solving the above system of two quadratic equations with the unknowns t_1, t_2.

Lastly, we consider the special case $n = 2$, i.e. the vector space V is 3-dimensional, when the search problem simplifies. Then W_1, W_2 are defined as

$$W_i = \mathbf{p}_i^\perp, i = 1,2,$$

the Lorentzian orthogonal complements to some space-like vectors $\mathbf{p}_1, \mathbf{p}_2$ in V. (In order to determine these vectors, one solves the linear systems $\langle \mathbf{p}_i, \mathbf{u}_i\rangle = \langle \mathbf{p}_i, \mathbf{v}_i\rangle = 0$, $i = 1,2$.) The segment $s = \bar{\mathbf{y}}_1\bar{\mathbf{y}}_2 \subset \mathbb{H}$ is contained in a hyperbolic geodesic h defined as $\mathbb{H} \cap W$,

$$W = \mathbf{p}^\perp,$$

where $\mathbf{p} \in V$ is a nonzero vector satisfying $\langle \mathbf{p}, \mathbf{p}_1\rangle = 0, \langle \mathbf{p}, \mathbf{p}_2\rangle = 0$. Then the (future-directed) vectors $\mathbf{y}_i$ are basis vectors of the lines $\mathbf{p}^\perp \cap W_i$, $i = 1,2$.

[4]This is a special case of the *Lehmer Problem* and *Margulis Conjecture*.

[5]This uniqueness comes from the non-existence of rectangles in spaces of negative curvature.

6. Quasigeodesics

While the general definition is more complicated, for the computational purposes, one can think of *quasi-geodesics* in $\mathbb{H}^n$ as certain special piecewise-geodesic paths c in $\mathbb{H}^n$; their advantage over geodesics is that they are more combinatorial objects. Each quasigeodesic comes with a certain constant $\lambda(c) \geqslant 1$, the *quasiisometry constant* of c, defined by the condition that for any two points p, q on c, we have

$$d(p,q) \geqslant \lambda^{-1}|c_{p,q}| - \lambda,$$

where $|c_{p,q}|$ is the length of the subpath of c between p and q.

Remark 6.1. *The above definition of quasigeodesics is not the most general, but it is the most appropriate for computational purposes and suffices when dealing with group-homomorphisms. For the general treatment of quasigeodesics we refer the reader to* [**DK**].

The constant λ measures 'how far c is from being a geodesic.' The magic of hyperbolic geometry is that every quasigeodesic, even an infinite one, is within uniformly bounded distance[6] from some geodesic. (This property is known as the *Morse Lemma*.)

The first paragraph is, in fact totally irrelevant for the computational purposes, it is mostly meant to introduce the terminology 'quasigeodesic.' The following is a practical criterion for something being a quasigeodesic, a proof could be found in [**KLi**]:

Theorem 6.2. *Suppose that c is a piecewise-geodesic path whose angles at the vertices are $\geqslant \alpha > 0$ and whose sides are longer than L, where α and L satisfy*

$$\cosh(L/2)\sin(\alpha/2) \geqslant \nu$$

where $\nu > 1$ is some fixed constant, say, $\sqrt{2}$. Then c is a quasigeodesic. The constant $\lambda(c)$ depends only on L and ν.

Remark 6.3. *The inequality in this theorem takes a particularly simple form if $\alpha \geqslant \pi/2$ (i.e. is obtuse):*

$$\cosh(L/2) \geqslant 2,$$

if we take $\nu = \sqrt{2}$ in this theorem.

The actual geometric requirement in this theorem (stated without invoking angles and lengths) is that for every two consecutive segments $\mathbf{p}_1\mathbf{p}_2, \mathbf{p}_2\mathbf{p}_3$ in the piecewise-geodesic path c, the bisectors $Bis(\mathbf{p}_1, \mathbf{p}_2), Bis(\mathbf{p}_2, \mathbf{p}_3)$ are disjoint and, moreover, are at least some fixed positive distance apart from each other. In view of Lemma 3.1, this condition translates to the language of Lorentzian geometry as:

$$\langle \mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_2 - \mathbf{p}_3\rangle^2 - \langle \mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_2 - \mathbf{p}_1\rangle\langle \mathbf{p}_2 - \mathbf{p}_3, \mathbf{p}_2 - \mathbf{p}_3\rangle \geqslant \epsilon, \tag{6.1}$$

where $\epsilon > 0$ is a fixed positive number.

Note that the inequalities that one needs to check are purely local, we just need to examine consecutive pairs of segments to verify them. This, of course, is still not feasible if there are many (or infinitely many) segments, but for quasigeodesics coming from group theory, there are only finitely many (one can even estimate how large is 'many') options for side-lengths and angles, so the verification becomes a finite problem. This is the 'local-to-global' principle in hyperbolic geometry. This principle has an analogue for higher rank Lie groups such as $SL(n, \mathbb{R})$ (and symmetric spaces they act on) but is much harder to state (and to prove). We will discuss this in section 12.

Example 6.2 (A non-example). *In the upper half-plane model $\{z \in \mathbb{C} : Im(z) > 0\}$ of the hyperbolic plane take the sequence of points*

$$z_k = k + \sqrt{-1}, k \in \mathbb{Z}.$$

Connect each consecutive pair z_k, z_{k+1} by a hyperbolic geodesic segment. The resulting path c has geodesic pieces of the constant length L and constant angle α between the consecutive pieces. These two numbers satisfy the equality

$$\cosh(L/2)\sin(\alpha/2) = 1,$$

but c is within infinite (hyperbolic) distance from any hyperbolic geodesic (it is uniformly close to the Euclidean horizontal line $Im(z) = 1$, a horosphere, but that does not count).

---

[6]this distance is $\leqslant a\lambda^2$, where a is some universal constant

The *KLP algorithm* described later on, uses the following important property of quasigeodesic paths:

Suppose that $c = \mathbf{x}_0\mathbf{x}_1 \star \mathbf{x}_1\mathbf{x}_2 \star \mathbf{x}_3\mathbf{x}_4 \star ...$ is a (finite or infinite) piecewise-geodesic path in $\mathbb{H}^n$, which is a concatenation of the geodesic segments $\mathbf{x}_i\mathbf{x}_{i+1}$. For a natural number N define the piecewise-geodesic path c_N as the concatenation

$$\mathbf{x}_0\mathbf{x}_N \star \mathbf{x}_{2N}\mathbf{x}_{2N}x_{3N} \star ...$$

Then c is λ-quasigeodesic if and only if c_N is $\lambda'(\lambda, N)$-quasigeodesic, for some universal function $\lambda' : \mathbb{R}_+ \times \mathbb{N} \to \mathbb{R}_+$.

7. Group homomorphisms

Suppose that Γ is a finitely generated group with a finite generating set $S = \{s_1, ..., s_k\}$. For concreteness, one can (and, at first, we will) assume that Γ is a free group on S. (But the discussion below will apply to other groups, things just become more complicated.)

A homomorphism

$$\rho : \Gamma \to O(n,1) = O(V, \langle \cdot, \cdot \rangle)$$

is simply a map sending the generators s_i to some matrices $A_i \in O^+(n,1)$ which satisfy the relators of Γ. One can also think of relators of Γ as 'hidden' and all what we have is a set of matrices $A_1, ..., A_k \in O^+(n,1)$. Our task is to 'discover' the hidden relators (or to prove that there is none). The 'freeness problem' is to find a semi-algorithm ensuring that $\rho(\Gamma)$ (the subgroup of $O^+(n,1)$ generated by $A_1, ..., A_k$) is free of rank k on the generators $A_1, ..., A_k$.

Before doing this, we need some terminology. Given a homomorphism ρ (a choice of matrices A_i) and a vector $\mathbf{x} \in \mathbb{H}$ (which is to be chosen wisely to make computations more efficient), one defines the 'orbit map'

$$o_{\mathbf{x}} : \Gamma \to \mathbb{H},$$

sending $\gamma \in \Gamma$ to the vector $\gamma\mathbf{x}$. (Here and in what follows, I will frequently abbreviate $\rho(\gamma)\mathbf{x}$ as $\gamma\mathbf{x}$.) However, what we have is more than just a map of Γ, we also get a map $f = f_{\mathbf{x}}$ of the Cayley graph T of Γ into $\mathbb{H}$ (in the case of a free group of rank k, this graph is the $2k$-valent tree): Send each edge $e = [w_1, w_2]$ of T to the segment

$$(w_1\mathbf{x})(w_2\mathbf{x}) \subset \mathbb{H}.$$

Most importantly, each geodesic path in the tree T (an edge-path without backtracking) is sent to a piecewise-geodesic path in $\mathbb{H}$. (For the computational purposes, one does not need to 'compute' the map f, it is used only to give a geometric explanation of what is happening.)

From the computational viewpoint, such paths in $\mathbb{H}$ are given by their vertex-sequences

$$\mathbf{x} = \mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_N,$$

defined for each reduced word $w = w(A_1^{\pm 1}, ..., A_k^{\pm 1})$ (of the length N) and applying inductively (in order of their appearance in w) the matrices

$$A_1^{\pm 1}, ..., A_k^{\pm 1}$$

to the point $\mathbf{x}$.

Definition 7.1 (Undistorted homomorphisms and subgroups). *A homomorphism ρ is called a* quasiisometric embedding *or, simply,* undistorted, *if the map f_x sends geodesic paths p in T to λ-quasigeodesic paths in $\mathbb{H}$ for some fixed λ independent of p. The image of an undistorted homomorphism is called an* undistorted subgroup *of $O^+(n,1)$.*

One of the many (not so) magic properties of undistorted homomorphisms is that they are **faithful**, i.e. the subgroup generated by $A_1, .., A_k$ is free on this generating set. Moreover, this subgroup is necessarily discrete and contains only hyperbolic matrices (besides the identity). For instance, to see faithfulness, note that if ρ is not faithful then its kernel contains elements γ arbitrarily far from the neutral element of Γ. (Here is one of the few places where we use the assumption that Γ is free. General hyperbolic groups can contain nontrivial finite normal subgroups. But there is always the largest such subgroup.) But, since ρ is undistorted, we have the inequality

$$0 = d(\mathbf{x}, \rho(\gamma)\mathbf{x}) \geqslant \lambda^{-1}|\gamma| - \lambda,$$

where $|\gamma|$ is the distance from γ to the neutral element of Γ. This shows that $|\gamma| \leqslant \lambda^2$, which is a contradiction. A similar argument establishes discreteness of $\rho(\Gamma)$.

Note that the maps $o_{\mathbf{x}}$ and $f_{\mathbf{x}}$ depend on $\mathbf{x}$, but the undistortion property does not. However, an unwise choice of $\mathbf{x}$ will make the quasiisometry constant λ larger and computations longer.

There are *distorted* injective homomorphisms (with discrete images) $\rho : \Gamma \to O^+(n,1)$, even when one restricts to homomorphisms whose targets are arithmetic subgroups of $O^+(n,1)$ such as the subgroup of integer matrices $O(n,1;\mathbb{Z})$. One of the most famous examples of such homeomorphisms comes from the embedding

ρ of the figure 8 knot group into $O^+(3,1)$. Algebraically, this group is a semidirect product of the free group F_2 and the infinite cyclic group. Then $\rho : F_2 \to O^+(3,1)$ is exponentially distorted. It is now known (due to work of Ian Agol) that every lattice in $O^+(3,1)$ contains a finitely generated subgroup which is either free or isomorphic to the fundamental group of a hyperbolic surface, and which is exponentially distorted in $O^+(3,1)$.

The notion of undistorted subgroups is closely related to *geometric finiteness* of subgroups, see [**Bo1, Bo2**]. In particular, every undistorted subgroup of $O^+(n,1)$ has finitely-sided Dirichlet fundamental polyhedra in $\mathbb{H}^n$. (We will discuss Dirichlet domains in more detail in Section 11.) The notion of *geometric finiteness* is a bit more general. For instance, an infinite cyclic subgroup of $O^+(n,1)$ generated by a parabolic element is geometrically finite but exponentially distorted. Geometric finiteness, in turn, is closely related to the property that Dirichlet fundamental polyhedra in $\mathbb{H}^n$ are finitely-sided: The two notions are equivalent if $n \leqslant 3$, but not for $n \geqslant 4$. (See Examples 5 and 6 in Section 12.4 of Ratcliffe's book [**Ra**].) However, for subgroups of lattices in $O^+(n,1)$, geometric finiteness is equivalent to the property that one (equivalently, every) Dirichlet fundamental polyhedron is finitely-sided.

For some reason, not quite clear in general, if one considers finitely generated discrete subgroups of $O^+(n,1)$, geometric finiteness appears to be a *generic property.* If $n = 2$, then every finitely generated discrete subgroup is geometrically finite. However for $n \geqslant 3$, there are finitely generated discrete geometrically infinite subgroups.

8. Testing for undistortion. Part I

If one looks closely at the maps f constructed in the previous section, one observes that f produces piecewise-geodesic paths in $\mathbb{H}$ (images of geodesics in T) that satisfy two 'finiteness' properties:

- The number of possible edge-lengths in these paths is at most k (one for each generator); they are given by the distances
$$L_1 = d(\mathbf{x}, A_1\mathbf{x}), ..., L_k = d(\mathbf{x}, A_k\mathbf{x}).$$
Recall that (see Section 2),
$$\cosh d(\mathbf{x}, A_i\mathbf{x}) = -\langle \mathbf{x}, A_i\mathbf{x}\rangle.$$
- The number of angles between the consecutive segments in such paths is at most $k(2k-1)$: One angle $\alpha_{i,\pm j}$ for each pair of generators $A_i, A_j^{\pm 1}$, where, of course, we do not allow pairs of the form (A_i, A_i) (which would correspond to backtracking in the tree T or, equivalently, nonreduced words w). Here
$$\alpha_{i,\pm j} = \angle(A_i\mathbf{x})\mathbf{x}(A_j^{\pm 1}\mathbf{x}),$$
that is (see Section 2),
$$\cos(\alpha_{i,\pm j}) = \frac{\langle A_i\mathbf{x}, A_j\mathbf{x}\rangle + \langle \mathbf{x}, A_i\mathbf{x}\rangle\langle \mathbf{x}, A_j^{\pm 1}\mathbf{x}\rangle}{(\langle \mathbf{x}, A_i\mathbf{x}\rangle^2 - 1)^{1/2}(\langle \mathbf{x}, A_j^{\pm 1}\mathbf{x}\rangle^2 - 1)^{1/2}}.$$

Now comes our first (and rather dumb) algorithm for testing the undistortion property. Even though it is dumb, it works quite well 'generically' and this is what's behind, say, the Fuchs–Rivin's proof of genericity of free subgroups in arithmetic groups, see [**FR**, Lemma 2.5].

The Dirty Harry Algorithm.[7] Check the inequality
$$\cosh(L/2)\sin(\alpha/2) > 1.$$
where $L = \min(L_1, ..., L_k)$, and
$$\alpha = \min\{\alpha_{i,\pm j} : 1 \leqslant i, j \leqslant k, (i, \pm j) \neq (i, -i)\}.$$
If this inequality holds, then indeed, f sends geodesic paths in the tree T to quasigeodesic paths in $\mathbb{H}$ and, hence, f is undistorted and, hence, injective with discrete image.

The geometric (rather than *coarse-geometric*) meaning of the inequality in the algorithm is that it ensures that the geodesic bisectors of the segments $(\mathbf{x})(A_i^{\pm 1}\mathbf{x})$ are pairwise disjoint in $\mathbb{H}^n$ (or even in the compactified hyperbolic space). If this happens then these bisectors will bound the Dirichlet fundamental domain of $\rho(\Gamma)$ in $\mathbb{H}^n$ centered at $\mathbf{x}$. We discuss Dirichlet fundamental domains and the corresponding *Poincaré* algorithm in detail in Section 11. For now, we simply record the fact that if the Dirty Harry Algorithm succeeds, then so does the KLP-algorithm, in its $O(n,1)$-version, and the *Poincaré* algorithm. Moreover, both of the latter algorithms terminate on their first step, dealing with group elements of word-length 1.

As we discussed in Section 6, instead of computing hyperbolic distances and angles, it is easier to test disjointness of bisectors. In this, more computationally-friendly, form, the Dirty Harry Algorithm works as follows. For each generator $A_i^{\pm 1}$, compute the vector
$$\mathbf{u}_{\pm i} = \mathbf{x} - A_i^{\pm 1}\mathbf{x}.$$

[7] "Do you feel lucky today?"

Then for each pair of different vectors $\mathbf{v}, \mathbf{w} \in \{\mathbf{u}_{\pm i} : i = 1, ..., k\}$ compute the difference

$$D_{\mathbf{v},\mathbf{w}} := \langle \mathbf{v}, \mathbf{w} \rangle^2 - \langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle.$$

If all differences satisfy $D_{\mathbf{v},\mathbf{w}} > 0$, the algorithm succeeds and the representation ρ is discrete, faithful and even undistorted. If all the differences satisfy $D_{\mathbf{v},\mathbf{w}} \geqslant 0$, the algorithm succeeds and the representation ρ is discrete and faithful. Otherwise, i.e. if some difference $D_{\mathbf{v},\mathbf{w}}$ is negative, the algorithm fails.

Even if ρ is undistorted, the Dirty Harry Algorithm might not work. It has a better chance of success provided one makes a 'wise' choice of the point $\mathbf{x}$.

Choosing x wisely (the 2-generator case). I first consider the case of 2-generator groups. Recall that for nondistortion to occur, all nontrivial elements of $\rho(\Gamma)$ have to be hyperbolic. Therefore, one should first check for hyperbolicity of the generators A_1, A_2. If one of them is nonhyperbolic (compute the eigenvalues), one stops and proceeds to try some other matrices. Suppose that A_1, A_2 are hyperbolic.

Let $h_1 = H_{A_1} \cap \mathbb{H}, h_2 = H_{A_2} \cap \mathbb{H}$ be the axes of A_1, A_2 (see section 4). Compute the geodesic segment $s_{h_1,h_2} = \mathbf{x}_1\mathbf{x}_2$ connecting h_1 and h_2 (see section 5). Then the 'wise' choice of $\mathbf{x}$ is the midpoint $m(\mathbf{x}_1\mathbf{x}_2)$ of the segment $\mathbf{x}_1\mathbf{x}_2$ (see (2.1)).

Remark 8.1. *While there are some heuristic reasons why one should be choosing the midpoint as* $\mathbf{x}$*, there is no solid mathematical justification for this.*

The idea of taking a midpoint extends to the case of a larger number of generators, but is computationally a bit more demanding:

Choosing x wisely (the general case). First, as above, for each pair of distinct indices i, j, compute the midpoint m_{ij} of the segment

$$s_{ij} = s_{h_i,h_j}$$

connecting the axes h_i, h_j of A_i, A_j. Then compute

$$\mathbf{b} = \sum_{i,j} m_{ij}$$

where the sum is taken over all pairs distinct elements of the set of midpoints of segments s_{ij},

$$M = \{m_{ij}, 1 \leqslant i < j \leqslant k\}.$$

Lastly, take as the point $\mathbf{x} \in \mathbb{H}$ the normalization of $\mathbf{b}$:

$$\mathbf{x} = \bar{\mathbf{b}} = \frac{\mathbf{b}}{|\mathbf{b}|}.$$

9. Testing for undistortion. Part II

I will now describe a simplified form of the KLP (Kapovich–Leeb–Porti) algorithm, written originally for subgroups of general semisimple real Lie groups, testing for the *Anosov property* of representations of hyperbolic groups; this is an adaptation and simplification in the case of $O^+(n,1)$). I will first do it for free groups and then in general.

This algorithm does not require luck: It terminates if and only if the homomorphism ρ is undistorted, therefore establishing semidecidability of the 'testing for undistortion' problem.

Instead of analyzing just the generators A_i (and their inverses), for $N \geqslant 1$ the $O^+(n,1)$-version of the KLP algorithm explores *radius* N *balls in the Cayley graph* T *centered at the neautral element* 1. If the algorithm provides the desired output for some N, it terminates, otherwise, it runs forever. Suppose that Γ is a free group with free generating set $s_1, ..., s_k$. Then the Cayley graph of Γ with respect to this generating set is a simplicial tree T. As before, we are given a homomorphism $\rho : \Gamma \to O^+(n,1)$, $\rho(s_i) = A_i, i = 1, ..., k$.

First, some terminology. The *radius* R (where R is a natural number) ball $B(R)$ centered at 1 in the vertex-set of the tree T is just the set of reduced words in $s_i^{\pm 1}$ of length at most R. We will be working with *(geodesic)* N*-triples* in such balls: These are triples of reduced words $w_1, 1, w_2$ which lie on a common geodesic segment (connecting w_1 to w_2) in the ball $B(N)$ and satisfy

$$|w_1| = |w_2| = N.$$

Here $|w|$ is the word-length of a reduced word w.

Concretely, being a geodesic N-triple means two things:

(1) $|w_1| = N, |w_2| = N$.
(2) The prefix of the word w_1 is different from the prefix of a word w_2, where the prefix of a word is the first letter (in the alphabet $s_i^{\pm 1}, i = 1, ..., k$) of the word.

Fix a point $\mathbf{x} \in \mathbb{H}$. (As before, it is best if this point is chosen wisely, but, unlike in the Dirty Harry case, the outcome of the KLP algorithm does not depend on the choice.)

Given a (geodesic) triple $\tau = (w_1, 1, w_2)$, compute the difference

$$D_\tau = \langle \mathbf{x} - w_1\mathbf{x}, \mathbf{x} - w_2\mathbf{x}\rangle^2 - \langle \mathbf{x} - w_1\mathbf{x}, \mathbf{x} - w_1\mathbf{x}\rangle\langle \mathbf{x} - w_2\mathbf{x}, \mathbf{x} - w_2\mathbf{x}\rangle.$$

DEFINITION 9.1. *We say that a triple $\tau = (w_1, 1, w_2)$ satisfies the* **qi condition** *if $D_\tau > 0$.*

Note that the triple $(w_1, 1, w_2)$ satisfies the qi condition if and only if the triple $(w_2, 1, w_1)$ does.

Now, we are ready for the actual algorithm (adapted from [**KLP1**]).

The rank one KLP algorithm. For each natural number N, consider all[8] (geodesic) N-triples $(w_1, 1, w_2)$ where w_1, w_2 are reduced words in the generators A_i, A_j^{-1},

$$|w_1| = N, |w_2| = N,$$

and the prefix of w_1 is different from the prefix of w_2.

For every such N-triple, check if it satisfies the **qi condition** as defined above. If all such N-triples pass the qi test, the algorithm stops: This means that the subgroup $\rho(\Gamma) < O^+(n, 1)$ generated by $A_1 = \rho(s_1), ..., A_k = \rho(s_k)$ is undistorted and is free of rank k.

If one of the N-triples $(w_1, 1, w_2)$ fails the test, then stop the analysis of N-triples, increase N by 1 and repeat. As a bonus, once the algorithm stops (if it does!) we can also estimate from above the quasiisometry constant of the orbit map $o_{\mathbf{x}} : \Gamma \to \gamma\mathbf{x} \subset X$.

REMARK 9.1. *Step 1 of the rank one KLP algorithm (i.e. $N = 1$) is nothing but the Dirty Harry Algorithm.*

THEOREM 9.2. *The KLP algorithm terminates if and only if the subgroup $\rho(\Gamma) < O^+(n, 1)$ generated by $A_1, ..., A_k$ is undistorted and is free of rank k.*

PROOF. The proof of this theorem is a special case of the one given in [**KLP1**, section 7]. Namely, suppose that the algorithm terminates. Then the orbit map $o_{\mathbf{x}} : \Gamma \to \mathbb{H}^n$ satisfies the following property: For each geodesic path p in T starting at 1, the restriction of $o_{\mathbf{x}}$ to p_N is λ-quasigeodesic for some λ independent of p, hence, the restriction to p is a λ'-quasigeodesic for some uniform constant λ, see the last paragraph of section 6. In order to conclude that $o_{\mathbf{x}}$ is a quasiisometry, we have to consider general paths p in T, not necessarily starting at the neutral element 1. However, the map $o_{\mathbf{x}}$ is ρ-equivariant:

$$o_{\mathbf{x}}(\gamma z) = \rho(\gamma) o_{\mathbf{x}}(z), z \in T.$$

Since Γ acts transitively on the vertex-set of T and the post-composition with isometries in $\rho(\Gamma)$ does not change the quasiisometry properties of a path, it follows that $o_{\mathbf{x}}$ is a quasiisometry.

For the opposite implication, we refer the reader to [**KLP1**]. The key is the following property: Since ρ is undistorted, for each geodesic path p in T starting at the neutral element, the vertices in $\mathbb{H}^n$ of the image c_N of p_N are uniformly close to the hyperbolic geodesic connecting $\mathbf{x}$ to the terminal point of c_N. At the same time, the distance between the consecutive points of c_N diverges to ∞ (this again uses nondistortion). Therefore, for each triple of consecutive points of c_N, say,

$$\mathbf{x}, \mathbf{x}_1 = c(N), \mathbf{x}_2 = c(2N),$$

the distances $d(\mathbf{x}, \mathbf{x}_1), d(\mathbf{x}_1, \mathbf{x}_2)$ grow arbitrarily large (as $N \to \infty$), while the angle $\angle \mathbf{x}\mathbf{x}_1\mathbf{x}_2$ is uniformly bounded away from zero. From this, one sees that the triple

$$\mathbf{x}, \mathbf{x}_1 = c(N), \mathbf{x}_2 = c(2N)$$

satisfies the qi condition if N is sufficiently large. □

One can use the 'wise choice' of $\mathbf{x}$ as described in section 8. I will describe alternatives in the next section.

One of important modifications in $\mathbb{H}^n$ of the KLP test is that it suffices to work with triples rather than with *quadruples* as it is done in [**KLP1**]. In a sense, the quadruple test from [**KLP1**] is more efficient, the drawback, however, is that one has to explore significantly larger balls in the Cayley graph. Here is the description of the quadruple test from [**KLP1**], again adapted to the case of the hyperbolic space.

Instead of triples, one works with *quadruples* of reduced words $w_0 = 1, w_1, w_2, w_3$ which lie on a common geodesic segment, satisfying $d(w_i, w_{i+1}) = N$, $i = 0, 1, 2$.

Given a (geodesic) quadruple $(1, w_1, w_2, w_3)$, compute the quadruple of vectors

$$\mathbf{x}_0 = \mathbf{x}, \mathbf{x}_1 = w_1(\mathbf{x}), \mathbf{x}_2 = w_2(\mathbf{x}), \mathbf{x}_3 = w_3(\mathbf{x})$$

in $\mathbb{H}$. For each segment $\mathbf{x}_0\mathbf{x}_1, \mathbf{x}_1\mathbf{x}_2, \mathbf{x}_2\mathbf{x}_3$ compute its midpoint

$$m_1 = m(\mathbf{x}_0, \mathbf{x}_1), m_2 = m(\mathbf{x}_1, \mathbf{x}_2), m_3 = m(\mathbf{x}_2, \mathbf{x}_3).$$

[8] Actually, from two triples $(w_1, 1, w_2)$, $(w_2, 1, w_1)$ it suffices to check just one.

DEFINITION 9.2. *Fix $\epsilon > 0$. We say that a quadruple $(1, w_1, w_2, w_3)$ satisfies the ϵ-**midpoint condition** if the triple of midpoints $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3) = (m_1, m_2, m_3)$ satisfies the inequality (6.1) from Section 6.*

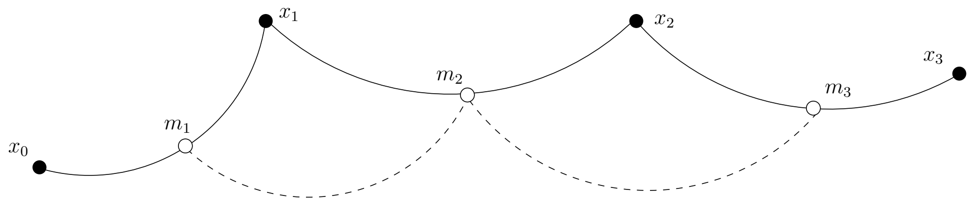

FIGURE 1. Midpoints

Then the KLP algorithm amounts to checking the midpoint condition for all (geodesic) N-quadruples $(1, w_1, w_2, w_3)$. Because one uses midpoints, in the undistorted case, the angles α are not just bounded away from zero as $N \to \infty$, but actually converge to π. (This convergence to π is critical in the higher rank case.) In particular, the products $\cosh(L/2)\sin(\alpha/2)$ diverge to infinity faster than in the qi test for triples.

10. Testing for nondistortion. Part III

Below is a version of the KLP algorithm for non-free subgroups. The algorithm is testing for the following:

Let $\Gamma = \langle a_1, ..., a_k | r_1, ..., r_s\rangle$ be a *word-hyperbolic group* given by its finite presentation. Let $\rho : \Gamma \to O^+(n, 1)$ be a homomorphism. The KLP algorithm determines if ρ is undistorted. If ρ is undistorted, it might fail to be injective, but, in the worst case, its kernel is finite. 'Most' examples of infinite hyperbolic groups have no nontrivial finite normal subgroups, thus, the nondistortion property *effectively* implies injectivity.

The only difference with section 9 is that the Cayley graph of Γ is no longer a tree and it is harder to test if a triple $(w_1, 1, w_2)$ lies on a geodesic. The right condition is

$$|w_1|_\Gamma = |w_2|_\Gamma = N, |w_1^{-1} w_2|_\Gamma = 2N$$

where $|w|_\Gamma$ is the length of the shortest word representing the same element of Γ as w. However, for hyperbolic groups (with a fixed presentation) there are practical algorithms for computing $|w|_\Gamma$ which can be used, see e.g. [**EH**]. Other than that, the KLP algorithm is the same as before.

Note that general word-hyperbolic groups, such as the triangle group $\Delta(p, q, r)$, could have generators a_i of finite order. Hence, **one cannot use** the procedure from section 8 in order to make a 'wise' choice of the vector $\mathbf{x}$. Here are the best alternatives I know:

Given matrices $A_1, ..., A_k$ in $O^+(n, 1)$, define the function

$$D(\mathbf{y}) := \max(d(A_1^{\pm 1}\mathbf{y}, \mathbf{y}), ..., d(A_k^{\pm 1}\mathbf{y}, \mathbf{y}))$$

on $\mathbb{H}$. This function is *convex* (actually, strictly convex in most examples). Then choose $\mathbf{x}$ to be the point of minimum of $D(\mathbf{y})$. In fact, it suffices just to be 'not too far' from the minimum, in any reasonable sense. From the linear algebra viewpoint, this is an unpleasant min-max problem since $D(\mathbf{y})$ is very nonlinear. Here is a practical replacement of the above min-max problem:

The min-max problem. Define the function

$$M(\mathbf{y}) := \max(-\langle A_1^{\pm 1}\mathbf{y}, \mathbf{y}\rangle, ..., -\langle A_k^{\pm 1}\mathbf{y}, \mathbf{y}\rangle)$$

on the open convex cone V^-. This function is piecewise-quadratic. Now, minimize this function over the hyperboloid $\mathbb{H} \subset V^-$. Choose $\mathbf{x}$ to be its minimum (even approximate one in any reasonable sense).

More alternatives. Instead of $M(\mathbf{y})$, you can take your favorite norm $|| \cdot ||$ of the k-tuple

$$(-\langle A_1^{\pm 1}\mathbf{y}, \mathbf{y}\rangle, ..., -\langle A_k^{\pm 1}\mathbf{y}, \mathbf{y}\rangle),$$

or even take the 'energy' (which is a quadratic function in $\mathbf{y}$)

$$E(\mathbf{y}) := -\sum_{i=1}^{k}\langle A_i\mathbf{y}, \mathbf{y}\rangle - \sum_{i=1}^{k}\langle A_i^{-1}\mathbf{y}, \mathbf{y}\rangle.$$

Now, minimize the norm, or $E(\mathbf{y})$, over the hyperboloid $\mathbb{H}$.

Other rank one Lie groups. What is described above works just as well when instead of the group $O^+(n, 1)$ of isometries of the hyperbolic n-space one considers isometry groups of other negatively curved symmetric spaces X, e.g. the group $PU(n, 1)$ of biholomorphic isometries of the complex-hyperbolic n-space. The inequality in Theorem 6.2 still implies the quasigeodesic condition provided one normalizes the Riemannian metric of X to have the upper curvature bound -1.

11. Selberg's higher rank generalization of Dirichlet domain and the Poincaré Algorithm

We begin by defining Dirichlet domains for discrete group actions on general metric spaces. Let (X, d) be a metric space and Γ a discrete isometry group of X, where discreteness is understood in the sense that for one (equivalently, every) $x \in X$ and every sequence of distinct elements $\gamma_i \in \Gamma$, we have

$$\lim_{i\to\infty} d(x, \gamma_i x) = \infty.$$

Suppose that $o \in X$ is a point which is fixed only by the neutral element of Γ. (Without this assumption, we will not obtain a fundamental domain by applying the Dirichlet construction.) Then define

$$D(\Gamma, o) = D(o) := \{x \in X : d(o, x) \leqslant d(\gamma o, x), \quad \forall \gamma \in \Gamma\}.$$

In order to understand where this definition comes from, consider the orbit, $\Gamma o \subset X$. The discreteness condition on Γ implies that this orbit is a discrete closed subset of X (moreover, each metric ball in X contains only finitely many orbit points). Thus, one defines the *Voronoi tiling* of X corresponding to this orbit:

$$D(\gamma o) = \{x \in X : d(\gamma o, x) \leqslant d(o', x) \quad \forall o' \in \Gamma o\},$$

is the tile labeled by the point γo. It is clear that every point of X belongs to one of the tiles and that Γ permutes the tiles simply-transitively. The *open tile* $\mathring{D}(\gamma o)$ is defined by

$$\mathring{D}(\gamma o) = \{x \in X : d(\gamma o, x) < d(o', x) \quad \forall o' \in \Gamma o \backslash \{\gamma o\}\}.$$

In view of continuity of the distance function $d : X^2 \to \mathbb{R}$, each $\mathring{D}(\gamma o)$ is an open subset of X. In general, however, the closure of an open tile need not be the corresponding closed tile $D(\gamma o)$. For instance, if the metric d is discrete, this will not be the case as the open tile is the singleton $\{\gamma o\}$. Moreover, the triangle inequalities imply that $\mathring{D}(\gamma o) \cap \mathring{D}(\gamma' o) \neq \emptyset$ if and only if $\gamma o = \gamma' o$. Suppose now that, additionally, (X, d) is a *geodesic space*, i.e. for any two points $x, y \in X$, $d(x, y)$ is the length of the shortest (geodesic) path in X connecting x and y. Then it is not hard to see that $\mathring{D}(\gamma o)$ is dense in $D(\gamma o)$, see [**K2**]. Thus, under this extra assumption, if $\gamma o \neq \gamma' o$, then

$$D(\gamma o) \cap D(\gamma' o) \subset \partial D(\gamma o) \cap \partial D(\gamma' o).$$

Discreteness of Γ implies that every bounded subset of X has nonempty intersection only with finitely many tiles. Thus, $D(\Gamma, o)$ serves as a fundamental domain for the action of Γ on X. We refer to [**K2**] for details.

For general metric spaces and even general Riemannian manifolds, very little can be said about geometry of Dirichlet domains and even of the *bisectors*

$$Bis(o, \gamma o) = \{x : d(o, x) = d(\gamma o, x)\}$$

bounding these domains. There is one case, however, when bisectors and, accordingly, Dirichlet domains, have particularly nice structure, namely, when (X, d) is the hyperbolic n-space $\mathbb{H}$. Then each bisector

$$Bis(\mathbf{p}, \mathbf{q}) = \{\mathbf{x} \in \mathbb{H} : \langle \mathbf{p}, \mathbf{x} \rangle = \langle \mathbf{q}, \mathbf{x} \rangle\}, \mathbf{p} \neq \mathbf{q},$$

is the intersections of the hyperboloid $\mathbb{H}$ with the linear hyperplane

$$\{\mathbf{x} \in V : \langle \mathbf{p} - \mathbf{q}, \mathbf{x} \rangle = 0\}.$$

Accordingly, $D(\Gamma, \mathbf{p})$ is the intersection of a (possibly infinitely-sided) convex polyhedral cone with $\mathbb{H}$. This polyhedral structure of $D(\Gamma, \mathbf{p})$ makes it amenable to algorithmic computations. The corresponding *Poincaré Algorithm* was described first, to my knowledge, by Riley in [**R1**] in the case $n = 3$ (who even wrote a code, in Fortran), and, in greater detail (but without actual computer implementation), by Epstein and Petronio in [**EP**].

We now specialize to the case of discrete subgroups of $G = SL(n, \mathbb{R})$, focusing on computational aspects. The group G acts naturally on the vector space V of symmetric $n \times n$ matrices, $M \mapsto g^T M g$, for $M \in V$. This action corresponds to the change of variables in the quadratic form defined by M. The group G also preserves the open cone $P \subset V$ of (strictly) positive-definite matrices and the hypersurface $X \subset P$ consisting of matrices of unit determinant. Moreover, G acts transitively on X with the stabilizer of the identity matrix equal to $SO(n)$. The hypersurface X does have a G-invariant Riemannian metric which, on the tangent space at the identity matrix I, equals

$$\langle A, B \rangle = tr(AB).$$

Given this metric, one defines the associated Riemannian distance function d. For instance,

$$d(I, A) = \left(\sum_i \log^2(\lambda_i)\right)^{1/2},$$

where λ_i's are the eigenvalues of $A \in X$.

Then, given a discrete subgroup $\Gamma < G$ which has trivial intersection with $SO(n)$, one defines the *Dirichlet domain* as above by
$$D(\Gamma, I) = \{C \in X : d(I, C) \leqslant d(\gamma^T \gamma, C), \forall \gamma \in \Gamma\}.$$
The trouble is that such domains are bounded by pieces of *geodesic bisectors* for the metric d, which are hard to compute (unlike in the case of the Lorentzian model of the hyperbolic space, where bisectors are linear).

Below, we describe a 2-point invariant $s(A, B)$ due to Selberg, [**Se**], which, while not a metric, can be used in lieu of one to define Dirichlet domains in P (and in X). This use is also due to Selberg but appears to be relatively unknown. The advantage of $s(A, B)$ is that it is easy to compute and the corresponding bisectors are linear.

Assuming that $A \in V$ is an invertible matrix and $B \in V$ is an arbitrary matrix, we define $s(A, B) := tr(A^{-1}B)$. Assuming further that $A, B \in P$ (at this point, we do not yet impose the condition $A, B \in X$), we set
$$\sigma(A, B) := \log\left(\frac{1}{n} s(A, B)\right) = \log\left(\frac{1}{n} tr(A^{-1}B)\right).$$
Then s, and, hence, σ, is G-invariant because trace is conjugacy-invariant:
$$s(g^T A g, g^T B g) = tr(g^{-1}A^{-1}(g^T)^{-1} g^T B g) = tr(g^{-1}A^{-1}Bg) = tr(A^{-1}B) = s(A, B).$$
The normalization (in the definition of σ) is chosen so that for $A, B \in X$ (i.e. $\det(A) = \det(B) = 1$), $\sigma(A, B) \geqslant 0$ with equality if and only if $A = B$, making σ a *premetric*: Indeed, without loss of generality, we may assume that both matrices A, B are diagonal,
$$A = Diag(a_1, ..., a_n), \quad B = Diag(b_1, ..., b_n).$$
Then, by the AM–GM inequality,
$$\frac{1}{n} s(A, B) = \frac{1}{n}\sum_{i=1}^n \frac{b_i}{a_i} \geqslant \left(\prod_{i=1}^n \frac{b_i}{a_i}\right)^{1/n} = 1,$$
with equality if and only if $b_i = a_i$ for all i.

However, in general, $\sigma(A, B) \neq \sigma(B, A)$ and σ fails the triangle inequality. Nevertheless, we will pretend that σ is a metric. For $A_1, A_2 \in X$, the σ-bisectors,
$$Bis(A_1, A_2) = \{B \in V : \sigma(A_1, B) = \sigma(A_2, B)\} = \{B \in V : tr(A_1^{-1}B) = tr(A_2^{-1}B)\}$$
are defined by an equation which is linear in the variable B. Hence, σ-bisectors are linear. Clearly,
$$\{B \in V : \sigma(A_1, B) < \sigma(A_2, B)\} \cap \{B \in V : \sigma(A_1, B) > \sigma(A_2, B)\} = \varnothing. \tag{11.1}$$

Another useful property of the 2-point invariant σ is that the function $B \mapsto \sigma(I, B)$ is proper when restricted to X. This is so because of the comparison to the invariant Finsler metric d_{max} on X:
$$\sigma(I, A) \leqslant d_{max}(I, A) \leqslant \sigma(I, A) + \log(n). \tag{11.2}$$
Here $d_{max}(I, A)$ is the logarithm of the largest eigenvalue of the matrix $A \in P$. From this, it follows that
$$\lim_{||A|| \to \infty, A \in X} d_{max}(I, Bis(I, A) \cap X) = \infty.$$

It is also instructive to consider the invariant $s(A, B)$ in the special case $n = 2$, i.e. $G = SL(2, \mathbb{R})$, when the vector space V is 3-dimensional. Up to the harmless multiplicative factor $-1/\det(A)$, the 2-point invariant $s(A, B)$ equals $(A, B) = -tr(adj(A)B)$, where
$$adj\begin{pmatrix} a & b \\ b & c \end{pmatrix} = \begin{pmatrix} c & -b \\ -b & a \end{pmatrix}.$$
Thus, (A, B) is a bilinear form on V, still invariant under the action of G. A direct computation shows that this form is symmetric and has signature $(2, 1)$. Thus, the vector space V equipped with the form $(\cdot, \cdot)$ is a 3-dimensional Lorentzian vector space. The group G acts on V with the kernel $\{\pm I\}$, hence, through a group isomorphic to $SO^+(2, 1)$, making it the identity component of the group of all linear automorphisms of $(\cdot, \cdot)$. Let us compute the quadratic form corresponding to our bilinear form:
$$(A, A) = -tr\left\{\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix}\right\} = b^2 - ac.$$
The matrix A is positive-definite if and only if $a > 0$ and $(A, A) < 0$. Hence the convex cone $P \subset V$ is a component (given by the inequality $a > 0$) of the set of time-like vectors in this Lorentzian space. The σ-bisectors in V are nothing but the Lorentzian bisectors discussed in Section 3.

With all these geometric preliminaries out of the way, we now return to the discussion of fundamental domains of discrete subgroups of G. Given a discrete subgroup $\Gamma < G$ as above, one defines the *Dirichlet–Selberg fundamental domain* of Γ in P centered at the identity matrix I:

$$DS(\Gamma, I) = \{C \in P : s(I, C) \leqslant s(\gamma^T \gamma, C), \forall \gamma \in \Gamma\}.$$

REMARK 11.1. *With a minor modification, this definition generalizes to fundamental domains centered at non-identity matrices $p \in X$,*

$$DS(\Gamma, p) = \{C \in P : s(p, C) \leqslant s(\gamma^T p \gamma, C), \forall \gamma \in \Gamma\}.$$

The fact that for $\gamma, \gamma' \in \Gamma$, the interiors of $\gamma DS(\Gamma, p), \gamma' DS(\Gamma, p)$ intersect if and only if $\gamma p = \gamma' p$ follows from (11.1).

LEMMA 11.2. *The collection of domains $DS(\Gamma, \gamma p), \gamma \in \Gamma$, is locally-finite in P, i.e. every compact in P intersects only finitely many cones $DS(\Gamma, \gamma p)$.*

PROOF. It suffices to prove the claim for the intersections of Dirichlet–Selberg domains with X. Suppose $R < \infty$ is such that for the R-ball $B(p, R) \subset X$ (with respect to the Finsler metric d_{max} on X) the σ-bisector $Bis(\gamma' p, \gamma p)$ bounding $D(\Gamma, \gamma p)$ intersects $B(p, R)$. Thus, there exists $x \in X$ such that

$$\sigma(\gamma p, x) \leqslant \sigma(p, x) \leqslant d_{max}(p, x) \leqslant R,$$

which (cf. (11.2)) implies that

$$d_{max}(\gamma p, x) - \log(n) \leqslant \sigma(\gamma p, x) \leqslant R.$$

In view of the proper discontinuity of the action of Γ on X, the number of such elements $\gamma \in \Gamma$ is finite. Thus, only finitely many Selberg-bisectors $Bis(\gamma' p, \gamma p)$ can intersect $B(p, R)$. ☐

In particular, each compact subset of P intersects only finitely many bisectors bounding DS and, hence, linearity of bisectors implies that $DS(\Gamma, I)$ is a convex polyhedral cone in P. (This cone might have infinitely many faces.) It also follows that $DS(\Gamma, p)$ satisfies all the properties of a fundamental domain of a discrete group action (cf. [**Ra**]), which justifies the name *Dirichlet–Selberg fundamental domain*.

The definition of $DS(\Gamma, p)$ suggests a slew of open questions. For instance:

QUESTION 11.3. *1. Which discrete subgroups have finitely-sided Dirichlet–Selberg domains?*

2. Uniform lattices do have finitely-sided Dirichlet–Selberg domains, but what about non-uniform lattices?

3. Do Anosov subgroups[9] of G have finitely-sided Dirichlet–Selberg domains (at least for some choice of base-points p)?

In contrast to discrete subgroups of $O^+(n, 1)$, it is quite unclear how *generic* are subgroups with finitely-sided Dirichlet–Selberg domains among discrete finitely generated subgroups of $SL(n, \mathbb{R})$.

By analogy with the Poincaré Fundamental Polyhedron Theorem in hyperbolic geometry (see [**EP**], and [**Ra**, Section 13.5]), one obtains a similar theorem in P, working with (relatively) closed convex polyhedral cones $C \subset P$ bounded by σ-bisectors. The paper [**EP**] is especially useful here, since it focuses on algorithmic aspects of the Poincaré Fundamental Polyhedron Theorem in the BSS computability model. Below is a review of the formulation of this theorem, adopted to the setting of domains bounded by σ-bisectors.

Suppose that C is a finitely-sided convex polyhedral cone in P, bounded by σ-bisectors of the form $Bis(p, \gamma_i p)$, where p is a chosen point in X and $\gamma_1, ..., \gamma_q$ are certain elements of G. We further assume that the facets in C are matched in pairs F_i, F_i' by the elements $\gamma_i \in G, i = 1, ..., q$, so that $\gamma_i : F_i \to F_i'$, $\gamma_i^{-1} : F_i' \to F_i$, and

$$F_i' \subset Bis(p, \gamma_i p), \quad F_i \subset Bis(p, \gamma_i^{-1} p), i = 1, ..., q.$$

Then, similarly to the case of the hyperbolic space, one defines *ridge-cycles* in ∂C corresponding to the *ridges*, which are codimension 2 faces E of C. Every such ridge-cycle is a finite sequence of elements $\gamma_i^{\pm 1}$,

$$(\gamma_{i_1}^{\pm 1}, ..., \gamma_{i_\ell}^{\pm 1}).$$

Specifically, assuming that the ridge E is the intersection of, say, facets F_{i_1}, F_{i_ℓ}, one starts the cycle with the generator γ_{i_1} pairing F_{i_1} and F_{i_1}'. The image $\gamma_{i_1}(E)$ is a ridge of the facet F_{i_1}', hence, is the intersection of F_{i_1}' and F_{i_2}. Then the second element in the cycle is the generator γ_{i_2} pairing F_{i_2} and F_{i_2}', etc. Since the cone C has only finitely many faces, eventually, we come back to the original ridge E, completing the cycle. This yields the product

$$\beta_{E,\ell} = \gamma_{i_\ell}^{\pm 1} \circ ... \circ \gamma_{i_1}^{\pm 1}.$$

[9]See section 12.

Ridge-cycles are defined so that ℓ is the least natural number such that $\beta_{E,\ell}(E) = E$. The conditions of the Poincaré Fundamental Polyhedron Theorem require that each β_ℓ has finite order N_E and fixes E pointwise. Moreover, let U_s denote a small neighborhood of the ridge $\beta_{E,s}(E)$ in the cone C. Set

$$F_E := \bigcup_{s=1}^{\ell-1} \beta_{E,s}(U_s).$$

Then the ridge-cycle condition also requires F_E together with its images

$$\beta_{E,\ell}(F_E), \beta^2_{E,k}(F_E), ..., \beta^{N_E-1}_{E,k}(F_E)$$

to form a perfect tiling of a neighborhood of E in P. This tiling condition can be reformulated in terms of Riemannian angles. Pick a point $x = x_0 \in E$. For each $s = 0, 1, ..., \ell-1$, we set $x_s = \beta_{E,s}(x)$. Let $\alpha_s(x)$ denote the Riemannian (with respect to the G-invariant Riemannian metric on X) dihedral angle between the facets of $C \cap X$ at $x_s \in \beta_{E,s}(E) \cap X$. Then one requires

$$\alpha_E := \sum_{s=1}^{\ell} \alpha_s = \frac{2\pi}{N_E}. \tag{11.3}$$

(If this holds for one choice of x, then it holds for all choices.)

Remark 11.4. *Yukun Du,* [**D**]*, recently defined (for generic generators γ_i) non-Riemannian analogues of angles between the above bisectors, which are G-invariant, satisfy the natural additivity property and also the property that the* neighborhood tiling *above is equivalent to the angle-sum condition* (11.3)*. These "angles" do not depend on choices of points $x \in E$ and are defined in terms of linear algebra. Hence, they are more amenable to computations than the Riemannian angles.*

A polyhedral cone C as above is a *pre-Dirichlet–Selberg domain* for the subgroup $\Gamma < G$ generated by $\gamma_1, ..., \gamma_q$: If Γ is discrete (which is, a priori, unclear), its actual Dirichlet–Selberg domain $DS(\Gamma, p)$ is contained in the cone C.

The last condition of the Poincaré Fundamental Polyhedron Theorem is the least pleasant one (it is void if $D = C \cap X$ is compact). The face-pairing transformations γ_i of $D := C \cap X$ define an equivalence relation $\sim$ on D generated by

$$x \sim \gamma_i(x), x \in F_i \cap X.$$

The ridge-cycle conditions above imply that the quotient-space $D/\sim$ has natural structure of a Riemannian orbifold modeled on the symmetric space X. Then the last condition requires this orbifold to be metrically complete. In the setting of the hyperbolic space, this metric completeness requirement can be replaced by a more computable "ideal vertex cycle" condition that can be found in [**EP**], [**Ra**] (see Theorems 13.4.5 and Theorems 13.4.7 in Ratcliffe's book). I currently do not know how to formulate a similar condition in the setting of convex cones in P as above. Below, is a review of the "ideal vertex cycle" condition in the context of hyperbolic spaces. Following [**Ra**], a *cusp point* of a finitely-sided convex polyhedron $Q \subset \mathbb{H}^n$ is a point v of the closure of Q in $\overline{\mathbb{H}^n} = \mathbb{H}^n \cup S^{n-1}$ (here we use the Klein model of the hyperbolic space) which equals to the intersection of closures in $\overline{\mathbb{H}^n}$ of all faces of Q whose closures contain v. Then, similarly, to the ridge-cycles, one defines ideal vertex cycles of such cusp points v. Let ℓ be the least integer such that the product $\beta_{v,\ell}$ sends v to itself. The *ideal vertex cycle* condition then is that every such $\beta_{v,\ell}$ is either elliptic or parabolic.

Lastly, returning to the pre-Dirichlet–Selberg domains C, one has:

Theorem 11.4. *The above ridge-cycles and completeness conditions are necessary and sufficient for C to be a fundamental domain for the action on P of the subgroup $\Gamma < G$ generated by the elements $\gamma_1, \gamma_2, ...$. Moreover, Γ has the presentation in the above generating set, where the relators are the products $\beta_{E,\ell}$ of the ridge-cycles, and E runs through the set of equivalence classes of the ridges in C.*

The proof of this theorem is exactly the same as in the hyperbolic case, see [**EP, Ra**]. The hardest part in a computational implementation of this theorem is the completeness condition and presently, we do not know how to deal with the issue (see, however, Conjecture 11.5 below). However, in the case of a compact fundamental domain the completeness is automatic and one obtains an algorithm for computing a fundamental domain and, hence, a finite presentation of a uniform lattice in G. Let $\Gamma < SL(n, \mathbb{R})$ be a uniform arithmetic lattice, given by its arithmetic data.

1. For each $N \in \mathbb{N}$, compute the subset Γ_N matrices $A \in \Gamma$ such that $s(I, A) \leqslant N$, which is a finite search.
2. Compute the intersection C_N of closed half-spaces $\{s(I, x) \leqslant s(A^T A, x)\}$ (in the vector space V of symmetric matrices) for $A \in \Gamma_N$. The intersection $C_N \cap P$ will be called a *partial Dirichlet–Selberg domain* of Γ.
3. Check if this intersection is contained in the cone P of positive-definite matrices.

(4) If it is not, then increase N to $N+1$ and repeat.
(5) Suppose, yes, then compute the *Selberg-radius* $\delta = \max\{\sigma(I,x) : x \in D_N\}$ of $D_N := C_N \cap X$ (imposing the extra condition det $= 1$).
(6) Next, find all $A \in \Gamma$ such that $N \leqslant \sigma(I, A^T A) \leqslant 2\delta + 2\log(n)$.
(7) $D_{2\delta+2\log(n)}$ will be the Dirichlet–Selberg domain of Γ.

The reader can find examples (my list is far from exhaustive) of similar algorithms in [**R1, R2**] (in the case of subgroups of $PSL(2,\mathbb{C})$ and $PSL(2,\mathbb{R})$ respectively), in [**M**] (in the case of subgroups of $PSL(2,\mathbb{R})$), in [**Li, P, Si**] (in the case of subgroups of $PSL(2,\mathbb{C})$), in [**C**] (in the case of groups acting on the bidisk), [**CS**] (in the case of groups acting on the complex 2-ball) and [**EP**] for subgroups of $O(n,1)$. See also [**J**] for a discussion of Siegel fundamental domains for the subgroup $Sp(2n,\mathbb{Z}) < Sp(2n,\mathbb{R})$.

Lastly, for general discrete subgroups $\Gamma < G$, we have:

Conjecture 11.5. *Each pre-Dirichlet–Selberg domain of* Γ *(regardless of its compactness!) satisfies the completeness condition and, hence, is a fundamental domain of* Γ.

One reason to be optimistic is that the analogous statement does hold for subgroups of $O(n,1)$ and pre-Dirichlet domains of such subgroups. The key reason is that if Q is a pre-Dirichlet domain in $\mathbb{H}^n$ centered at a point $p \in \mathbb{H}^n$, then for each point $v \in S^{n-1}$ in the closure of $Bis(p, \gamma_i^{\pm 1} p)$, the points $p, \gamma_i^{\pm 1}(p)$ lie on the same *horosphere* in $\mathbb{H}^n$ based at v. For each cusp-point v of Q we take the horosphere Σ_v based at v and passing through p. Suppose that

$$(\gamma_{i_1}^{\pm 1}, ..., \gamma_{i_\ell}^{\pm 1})$$

is an ideal vertex cycle of a cusp point v of Q and

$$\gamma_{i_1}^{\pm 1}(v) = v_1, \gamma_{i_2}^{\pm 1}(v_1) = v_2, ..., \gamma_{i_\ell}^{\pm 1}(v_{\ell-1}) = v.$$

The above observation about boundary points of bisectors then implies that

$$\gamma_{i_1}^{\pm 1} : \Sigma_v \to \Sigma_{v_1}, \quad \gamma_{i_2}^{\pm 1} : \Sigma_{v_1} \to \Sigma_{v_2}, ..., \quad \gamma_{i_\ell}^{\pm 1} : \Sigma_{v_{\ell-1}} \to \Sigma_v.$$

Applying this to the product

$$\beta_{v,\ell} = \gamma_{i_\ell}^{\pm 1} ... \circ \gamma_{i_1}^{\pm 1},$$

we conclude that it sends the horosphere Σ_v back to itself. Hence, $\beta_{v,\ell}$ is elliptic or parabolic, verifying the ideal vertex cycle condition.

12. Computational aspects of Anosov subgroups

We will again limit the discussion to the case of discrete subgroups of $G = SL(n,\mathbb{R})$. While general *undistorted* finitely generated subgroups of G are rather poorly-behaved (for instance, they need not be finitely presentable), the *Anosov* condition below eliminates various pathologies and results in a class of subgroups which share many desirable properties with undistorted subgroups of $O^+(n,1)$. The *Anosov property* was originally formulated for discrete subgroups Γ (of semisimple Lie groups) by Labourie, Guichard and Wienhard, see [**La**], [**GW**]. While it is defined *relative* to a certain parabolic subgroup P of G, for the sake of simplicity, I limit myself to the discussion when P is a minimal parabolic subgroup (i.e. Borel subgroup) of $G = SL(n,\mathbb{R})$, i.e. the subgroup of upper triangular matrices. To simplify the terminology, we will refer to such subgroups simply as *Anosov*. The following is a treatment of Anosov subgroups following our work with Leeb and Porti, [**KLP2**] and [**KL**].

First, let us revisit the notion of discreteness for subgroups of $G = SL(n,\mathbb{R})$: A subgroup $\Gamma < G$ is discrete if every sequence of distinct matrices γ_i in Γ diverges to infinity, $||\gamma_i|| \to \infty$. Equivalently, the sequence of highest singular values of γ_i's diverges to infinity.

Looking at the singular values of these matrices arranged in the decreasing order,

$$\sigma_1(\gamma_i) \geqslant \sigma_2(\gamma_i) \geqslant ... \geqslant \sigma_n(\gamma_i),$$

and their asymptotics as $i \to \infty$, one realizes that this divergence to infinity can happen in quantitatively different ways. For instance, a sequence of matrices is called *regular* if each sequence of successive quotients $\frac{\sigma_k(\gamma_i)}{\sigma_{k+1}(\gamma_i)}$ diverges to infinity. Accordingly, *regularity* of a discrete subgroup Γ means that every unbounded sequence in it is regular. (The regularity condition can be weakened to *partial regularity* by looking at the ratios of *some* of the successive singular values, leading to an interesting theory as well: This corresponds to the notion of P-Anosov subgroups for general parabolic subgroups $P < G$.) Regularity is equivalent to discreteness if $n = 2$ but not for $n \geqslant 3$. For instance, the subgroup of matrices with integer coefficients, $SL(n,\mathbb{Z})$, is discrete but is not even partially regular if $n \geqslant 3$. (One way to see this lack of regularity is to observe that $SL(n,\mathbb{Z})$ contains diagonalizable subgroups isomorphic to $\mathbb{Z}^{n-1}$ and such subgroups are easily seen to be non-regular with respect to any parabolic subgroup $P < G$.)

So far, our discussion was in terms of linear algebra; in order to get the actual Anosov condition, one connects linear algebra with the geometry of Γ itself, assuming that Γ is finitely generated, equipped with a word-metric d_Γ. It is not hard to see that the ratios of singular values as above cannot diverge to infinity at rate faster than exponential with respect to $d_\Gamma(1,\gamma_i)$, but they *can* diverge to infinity subexponentially, even linearly. (This happens, for instance, in the case of $SL(2,\mathbb{Z})$ when we consider the sequence of powers of a unipotent matrix.) This observation leads to a definition, which (in a more geometric form) first appeared in our work with Leeb and Porti:

DEFINITION 12.1. *A (discrete) finitely generated subgroup* $\Gamma < G = SL(n,\mathbb{R})$ *is called* URU *if there exists* $A > 0$ *such that for every* $\gamma \in \Gamma$,

$$\frac{\sigma_k(\gamma)}{\sigma_{k+1}(\gamma)} \geqslant A^{-1}\exp(A\cdot d_\Gamma(1,\gamma)), k = 1, ..., n-1.$$

In particular, this definition includes the property that Γ is undistorted in G and is a regular subgroup. However, the regularity condition appearing in this definition is a bit stronger than the one formulated above, it is called *uniform regularity* in [**KLP1, KLP2**]. We will not define it here (as it will not be needed), but only note that URU stands for *uniformly regular undistorted.* It is proven in [**KLP2**] that every URU subgroup is word-hyperbolic.

Given a hyperbolic group Γ and a homomorphism $\rho : \Gamma \to G$, one says that ρ *is Anosov* if ρ has finite kernel and Anosov image.

It was proven in [**KLP1**] that the Anosov property for group homomorphisms is semidecidable. The KLP algorithm for testing the Anosov property is very similar to the one described in Section 9 for representations to $O^+(n,1)$ using the *midpoint test.* The main differences with the hyperbolic case are:

(1) One adds a regularity test for the geodesic segments $m_j m_{j+1}$ connecting the midpoints of the geodesic segments $\rho(w_j)(x)\rho(w_{j+1})(x)$ in the space X of positive definite matrices with unit determinant. (Here w_j, w_{j+1} are the words appearing in geodesic quadruples in the midpoint test described in Section 9.) For the geodesic segment connecting the identity matrix to a matrix $m \in X$, the regularity condition amounts to checking that m satisfies the eigenvalue inequalities

$$\frac{\lambda_k(m)}{\lambda_{k+1}(m)} \geqslant r > 1, \quad k = 1, 2, ..., n-1.$$

(2) The Riemannian angles α appearing in Section 9 are replaced with certain ζ-angles, which I will discuss below.

(3) In the analogue of the inequality in Theorem 6.2, the distances L and ζ-angles α are *decoupled*: One requires that $L \geqslant L_n(r)$ and $\alpha \geqslant \pi - \epsilon_n(r)$, where $L_n(r)$ and $\epsilon_n(r)$ are certain functions.

In [**KLP1**] the existence of the functions $L_n = L_n(r)$ and $\epsilon_n = \epsilon_n(r)$ (for which the algorithm works) was established by certain continuity arguments. Max Riestenberg in his PhD thesis, [**Ri**], computed these functions explicitly, making it, in theory, possible, to test if the given representation is Anosov, in particular, has finite kernel and discrete image. The KLP algorithm then runs essentially as in Section 9, except, in addition to increasing N, one also decreases $r > 1$, taking it equal, say, to $1 + \frac{1}{N}$. The algorithm terminates if and only if ρ is Anosov.

Below is a definition of ζ-angles adapted to the setting of the symmetric space X of the group $SL(n,\mathbb{R})$. We first define ζ-angles between tangent directions $\mathbf{u}, \mathbf{v}$, i.e. nonzero vectors in the tangent space $T_I X$ at the identity matrix $I \in X$. This tangent space is nothing but the space of traceless symmetric matrices. The ζ-angle is defined only between *regular* matrices $\mathbf{u}, \mathbf{v} \in T_I X$, meaning that the eigenvalues of $\mathbf{u}$ and of $\mathbf{v}$ are pairwise distinct.

The ζ-angle is defined with respect to a fixed diagonal matrix

$$\zeta = Diag(\zeta_1, ..., \zeta_n)$$

satisfying the following conditions:

(1) $\zeta_i = -\zeta_{n-i}$, $i = 1, 2....$
(2) $\sum_{i=1}^n \zeta_i = 0$.
(3) $\zeta_i > \zeta_{i+1}$ for all $i \leqslant n-1$.

There is no canonical choice of such vectors, one can take, for instance, ζ coming from the *sum of positive coroots* of the root system of type A, namely,

$$\zeta_1 = (n-1), \quad \zeta_{i+1} = \zeta_i - 2, i = 1, 2, ..., n-1.$$

In particular (with this choice), if n is odd, then (for $n-1=2k$),

$$||\zeta||^2 = 2\sum_{i=1}^{k}(2i)^2,$$

while if $n=2k$ is even, then

$$||\zeta||^2 = 2\sum_{i=1}^{k}(2i-1)^2.$$

The most important property that ζ has, is that it belongs to the *open Weyl chamber* defined by the inequalities

$$\zeta_1 > \zeta_2 > ... > \zeta_n.$$

Next, given a matrix $\mathbf{u} \in T_I X$ with the eigenvalues $\lambda_1 > \lambda_2 > ... > \lambda_n$, let $Q \in O(n)$ be the matrix which diagonalizes $\mathbf{u}$, i.e.

$$\mathbf{u} = Q^T Diag(\lambda_1, ..., \lambda_n) Q.$$

Then define the ζ*-direction* of $\mathbf{u}$ as

$$\mathbf{u}_\zeta := Q^T \zeta Q.$$

Recall that $T_I X$ is equipped with the Riemannian metric $\langle \mathbf{x}, \mathbf{y} \rangle = tr(\mathbf{xy})$. The ζ-angle between $\mathbf{u}, \mathbf{v}$ is defined as the Riemannian angle between the directions $\mathbf{u}_\zeta$ and $\mathbf{v}_\zeta$. In other words,

$$\cos(\angle_I^\zeta(\mathbf{u}, \mathbf{v})) = \frac{tr(\mathbf{u}_\zeta \mathbf{v}_\zeta)}{||\zeta||^2}.$$

By the construction, such angles are invariant under the action of $O(n)$ on the tangent space $T_I X$.

We next define ζ-angles for triangles $\Delta I M_1 M_2$ in X, by measuring the angle at the corner I of the triangle. Define matrices $\mathbf{m}_i \in T_I X$ by

$$\mathbf{m}_i = \log(M_i), i = 1, 2.$$

Then set

$$\angle_I^\zeta(M_1, M_2) := \angle_I^\zeta(\mathbf{m}_1, \mathbf{m}_2) = \angle_I^\zeta(\log M_1, \log M_2).$$

Lastly, in order to define ζ-angles of general triangles $\Delta M M_1 M_2$ in X, we impose the G-invariance of such angles. Suppose that $M = g^T g, g \in G$. Then set

$$\angle_M^\zeta(M_1, M_2) := \angle_I^\zeta((g^{-1})^T M_1 g^{-1}, (g^{-1})^T M_2 g^{-1}).$$

We will not define the the functions $L_n = L_n(r)$ and $\epsilon_n = \epsilon_n(r)$ and refer instead to the work of Max Riestenberg, [**Ri**]. Note that the computational feasibility of the KLP algorithm is currently unclear even in the case of subgroups of $O(n,1)$, which should be addressed first (before the case of discrete subgroups of $SL(n,\mathbb{R})$ is discussed). To the best of our knowledge, this was never done.

References

[Bo1] B. H. Bowditch, *Geometrical finiteness for hyperbolic groups*, J. Funct. Anal. **113** (1993), no. 2, pp. 245–317.

[Bo2] B. H. Bowditch, *Geometrical finiteness with variable negative curvature*, Duke Math. J. **77** (1995) pp. 229–274.

[C] H. Cohn, *Some computer-assisted topological models of Hilbert fundamental domains*, Math. Comp. **23** (1969), pp. 475–487.

[CS] D. I. Cartwright and T. Steger, *Finding generators and relations for groups acting on the hyperbolic ball*, Preprint, 2017. arXiv:1701.02452. Software documentation: `https://www.maths.usyd.edu.au/u/donaldc/fakeprojectiveplanes/`

[DK] C. Drutu and M. Kapovich, "Geometric group theory." AMS series "Colloquium Publications", 2018.

[D] Y. Du, *Hyperplanes in the symmetric space* $SL(n,\mathbb{R})/SO(n)$, Preprint.

[EH] D. B. A. Epstein and D. Holt, *Computation in word-hyperbolic groups*, International Journal of Algebra and Computation, **11** (2001), no. 4, pp. 467–487.

[EP] D. B. A. Epstein and C. Petronio, *An exposition of Poincaré's polyhedron theorem*, Enseign. Math. (2) **40** (1994), no. 1–2, pp. 113–170.

[FR] E. Fuchs and I. Rivin, *Generic thinness in finitely generated subgroups of* $SL_n(\mathbb{Z})$. Int. Math. Res. Not. IMRN 2017, no. **17**, pp. 5385–5414.

[G1] J. Gilman, *Two-generator discrete subgroups of* $PSL(2,\mathbb{R})$, Mem. Amer. Math. Soc., **117** (1995) no. 561.

[G2] J. Gilman, Algorithms, complexity and discreteness criteria in $PSL(2,\mathbb{C})$, *J. Anal. Math.* **73** (1997), pp. 91–114.

[G3] J. Gilman, *Computability models: Algebraic, topological and geometric algorithms.* These proceedings.

[GM] J. Gilman, B. Maskit, *An algorithm for 2-generator Fuchsian groups*, Michigan Math. J., **38** (1991), no. 1, pp. 13–32.

[GW] O. Guichard, A. Wienhard, *Anosov representations: Domains of discontinuity and applications*, Invent. Math. **190** (2012) pp. 357–438.

[J] C. Jaber, *Algorithmic approaches to Siegel's fundamental domain*, PhD Thesis, Université de Bourgogne, 2017. `https://tel.archives-ouvertes.fr/tel-01813184/document`

[K1] M. Kapovich, *Discreteness is undecidable*, International Journal of Algebra and Computation, **26** (2016) pp. 467–472.

[K2] M. Kapovich, *A note on properly discontinuous actions*, Preprint, 2022. `https://www.math.ucdavis.edu/~kapovich/EPR/prop-disc.pdf`

[KL] M. Kapovich and B. Leeb, *Discrete isometry groups of symmetric spaces*, Volume IV of Handbook of Group Actions. The ALM series, International Press, Eds. L.Ji, A.Papadopoulos, S-T.Yau. (2018) Chapter 5, pp. 191–290.

[KLP1] M. Kapovich, B. Leeb and J. Porti, *Morse actions of discrete groups on symmetric spaces*, arXiv e-print, 2014. arXiv:1403.7671.

[KLP2] M. Kapovich, B. Leeb and J. Porti, *A Morse Lemma for quasigeodesics in symmetric spaces and euclidean buildings*, Geometry and Topology, **22** (2018) pp. 3827–3923.

[KLi] M. Kapovich and B. Liu, *Geometric finiteness in negatively pinched Hadamard manifolds*, Ann. Acad. Sci. Fennicae, **44** (2019) 2, pp. 841–875.

[La] F. Labourie, *Anosov flows, surface groups and curves in projective space*, Invent. Math. **165** (2006) pp. 51–114.

[Li] M. Lipyanskiy, *Computer-assisted application of Poincare's fundamental polyhedron theorem*, Preprint, 2001, `http://www.math.columbia.edu/~ums/pdf/poincare.pdf`.

[M] G. Muller, *Computing a generating set of arithmetic Kleinian groups*. In: "Teichmüller theory and moduli problem," pp. 513–517, Ramanujan Math. Soc. Lect. Notes Ser., **10**, Ramanujan Math. Soc., Mysore, 2010.

[P] A. Page, *Computing arithmetic Kleinian groups*, Math. Comp. **84** (2015), no. 295, pp. 2361–2390.

[Ra] J. Ratcliffe, "Foundations of hyperbolic manifolds." Third edition. Graduate Texts in Mathematics, **149**, Springer Verlag, 2019

[Ri] J. M. Riestenberg, *A quantified local-to-global principle for Morse quasigeodesics*, Preprint, 2021. arXiv:2101.07162.

[R1] R. Riley, *Applications of a computer implementation of Poincaré's theorem on fundamental polyhedra*, Math. of Computation, **40** (1983) pp. 607–632.

[R2] R. Riley, *Nielsen's algorithm to decide whether a group is Fuchsian*. "In the tradition of Ahlfors and Bers (Stony Brook, NY, 1998)," pp. 255–270.

[Se] A. Selberg, *On discontinuous groups in higher-dimensional symmetric spaces*. In: "Contributions to Function Theory," Tata Institute of Fundamental Research, Bombay (1960), pp. 147–164.

[Si] K. Siegel, *Applying Poincare's polyhedron theorem to groups of hyperbolic isometries*, Preprint, `https://kylersiegel.xyz/Poincare.pdf`.

Department of Mathematics, University of California, Davis, CA 95616

Current address: Department of Mathematics, University of California, Davis, CA 95616

E-mail address: `kapovich@ucdavis.edu`