ISSN 1464-8997 **1**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 1–21

The mean curvature integral is invariant under bending

Frederic J Almgren Jr
Igor Rivin

Abstract Suppose $\mathcal{M}_t$ is a smooth family of compact connected two dimensional submanifolds of Euclidean space E^3 without boundary varying isometrically in their induced Riemannian metrics. Then we show that the mean curvature integrals

$$\int_{\mathcal{M}_t} H_t \, d\mathcal{H}^2$$

are constant. It is unknown whether there are nontrivial such bendings $\mathcal{M}_t$. The estimates also hold for periodic manifolds for which there are nontrivial bendings. In addition, our methods work essentially without change to show the similar results for submanifolds of H^n and S^n, to wit, if $\mathcal{M}_t = \partial X_t$

$$d \int_{\mathcal{M}_t} H_t \, d\mathcal{H}^2 = -kn - 1dV(X_t),$$

where $k = -1$ for H^3 and $k = 1$ for S^3. The Euclidean case can be viewed as a special case where $k = 0$. The rigidity of the mean curvature integral can be used to show new rigidity results for isometric embeddings and provide new proofs of some well-known results. This, together with far-reaching extensions of the results of the present note is done in the preprint [6]. Our result should be compared with the well-known formula of Herglotz (see [5], also [8] and [2]).

AMS Classification 53A07, 49Q15

Keywords Isometric embedding, integral mean curvature, bending, varifolds

1 Introduction

The underlying idea of this note is the following. Suppose $\mathcal{N}_t$ is a smoothly varying family of polyhedral solids having edges $\{E_t(k)\}_k$, and associated (signed) dihedral angles $\{\theta_t(k)\}_k$. According to a theorem of Schlafli [7]

$$\sum_k \left|E_t(k)\right| \frac{d}{dt}\theta_t(k) = 0.$$

In case edge length is preserved in the family, ie

$$\frac{d}{dt}\left|E_t(k)\right| = 0$$

for each time t and each k, then also (product rule)

$$\frac{d}{dt}\sum_k \left|E_t(k)\right| \theta_t(k) = 0.$$

Should the $\partial\mathcal{N}_t$'s be polyhedral approximations to submanifolds $\mathcal{M}_t$ varying isometrically, one might regard

$$\sum_k \left|E_t(k)\right| \theta_t(k)$$

as a reasonable approximation to the mean curvature integrals

$$\int_{\mathcal{M}_t} H_t \, d\mathcal{H}^2$$

and expect

$$\frac{d}{dt}\left|E_t(k)\right|$$

to be small. Hence it is plausible that the mean curvature integrals of the $\mathcal{M}_t$'s might be constant. In this note we show that that is indeed the case.

Examples such as the isometry pictured on page 306 of volume 5 of [8] show that the mean curvature integral is not preserved under discrete isometries.

Two comments are in order. The first is that it is very likely that there are *no* isometric bendings of hypersurfaces. One reason for the existence of the current work is to produce a tool for resolving this conjecture (as Herglotz' mean curvature variation formula can be used to give a simple proof of Cohn–Vossen's theorem on rigidity of convex hypersurfaces). Secondly, the main theorem can be viewed as a sort of dual bellows theorem (when the hypersurface in question lies in H^n or S^n): as the surface is isometrically deformed, the volume of the *polar dual* stays constant. This should be contrasted with the usual bellows theorem recently proved by Sabitov, Connelly and Walz [4].

2 Terminology and basic facts

Our object in this section is to set up terminology for a family of manifolds varying smoothly through isometries. We consider triangulations of increasing fineness varying with the manifolds. To make possible our mean curvature analysis we associate integral varifolds with both the manifolds and the polyhedral surfaces determined by the triangulations. The mean curvature integral of interest is identified with (minus two times) the varifold first variation associated with the unit normal initial velocity vector field.

2.1 Terminology and facts for a static manifold $\mathcal{M}$

2.1.1 We suppose that $\mathcal{M} \subset \mathbb{R}^3$ is a compact connected smooth two dimensional submanifold of $\mathbb{R}^3$ without boundary oriented by a smooth Gauss mapping $\mathbf{n}\colon \mathcal{M} \to \mathbb{S}^2$ of unit normal vectors.

2.1.2 $H\colon \mathcal{M} \to \mathbb{R}$ denotes half the sum of principal curvatures in direction $\mathbf{n}$ at points in $\mathcal{M}$ so that $H\mathbf{n}$ is the mean curvature vector field of $\mathcal{M}$.

2.1.3 We denote by U a suitable neighborhood of $\mathcal{M}$ in $\mathbb{R}^3$ in which a smooth nearest point retraction mapping $\rho\colon U \to \mathcal{M}$ is well defined. The smooth signed distance function $\sigma\colon U \to \mathbb{R}$ is defined by requiring $p = \rho(p) + \sigma(p)\,\mathbf{n}(\rho(p))$ for each p. We set

$$g = \nabla\sigma\colon U \to \mathbb{R}^3$$

(so that $g|\mathcal{M} = \mathbf{n}$); the vector field g is the initial velocity vector field of the deformation

$$G_t\colon U \to \mathbb{R}^3, \quad G_t(p) = p + t\,g(p) \quad \text{for } p \in U.$$

2.1.4 We denote by

$$V = \mathbf{v}(\mathcal{M})$$

the *integral varifold* associated with $\mathcal{M}$ [1, 3.5]. The first variation distribution of V [1, 4.1, 4.2] is representable by integration [1, 4.3] and can be written

$$\delta V = \mathcal{H}^2 \llcorner \mathcal{M} \wedge (-2H)\mathbf{n}$$

[1, 4.3.5] so that

$$\delta V(g) = \frac{d}{dt}\mathcal{H}^2\big(G_t(\mathcal{M})\big)\Big|_{t=0} = -2\int_{\mathcal{M}} g \cdot H\,\mathbf{n}\,d\mathcal{H}^2 = -2\int_{\mathcal{M}} H\,d\mathcal{H}^2;$$

here $\mathcal{H}^2$ denotes two dimensional Hausdorff measure in $\mathbb{R}^3$.

2.1.5 By a **vertex** p in $\mathcal{M}$ we mean any point p in $\mathcal{M}$. By an **edge** $\langle pq\rangle$ in $\mathcal{M}$ we mean any (unordered) pair of distinct vertexes p, q in $\mathcal{M}$ which are close enough together that there is a unique length minimizing geodesic arc $[\![pq]\!]$ in $\mathcal{M}$ joining them; in particular $\langle pq\rangle = \langle qp\rangle$. For each edge $\langle pq\rangle$ we write $\partial\langle pq\rangle = \{p, q\}$ and call p a vertex of edge $\langle pq\rangle$, etc. We also denote by $\overline{pq}$ the straight line segment in $\mathbb{R}^3$ between p and q, ie the convex hull of p and q. By a *facet* $\langle pqr\rangle$ in $\mathcal{M}$ we mean any (unordered) triple of distinct vertexes p, q, r which are not collinear in $\mathbb{R}^3$ such that $\langle pq\rangle$, $\langle qr\rangle$, $\langle rp\rangle$ are edges in $\mathcal{M}$; in particular, $\langle pqr\rangle = \langle qpr\rangle = \langle rpq\rangle$, etc. For each facet $\langle pqr\rangle$ we write

$\partial\langle pqr\rangle = \big\{\langle pq\rangle,\ \langle qr\rangle,\ \langle rp\rangle\big\}$ and call $\langle pq\rangle$ an edge of facet $\langle pqr\rangle$ and also denote by $\overline{pqr}$ the convex hull of $p,\ q,\ r$ in $\mathbb{R}^3$.

2.1.6 Suppose $0 < \tau < 1$ and $0 < \lambda < 1$. By a τ, λ *regular triangulation* $\mathcal{T}$ *of* $\mathcal{M}$ *of maximum edge length* L we mean

(i) a family $\mathcal{T}_2$ of facets in $\mathcal{M}$, together with

(ii) the family $\mathcal{T}_1$ of all edges of facets in $\mathcal{T}_2$ together with

(iii) the family $\mathcal{T}_0$ of all vertexes of edges in $\mathcal{T}_1$

such that

(iv) $\overline{pqr} \subset U$ for each facet $\langle pqr\rangle$ in $\mathcal{T}_2$

(v) $\mathcal{M}$ is partitioned by the family of subsets

$$\Big\{\rho\big(\overline{pqr} \sim (\overline{pq} \cup \overline{qr} \cup \overline{rq})\big) : \langle pqr\rangle \in \mathcal{T}_2\Big\} \cup \Big\{\rho(\overline{pq}) \sim \{p,\ q\} : \langle pq\rangle \in \mathcal{T}_1\Big\} \cup \Big\{\{p\} : p \in \mathcal{T}_0\Big\}$$

(vi) for facets $\langle pqr\rangle \in \mathcal{T}_2$ we have the uniform nondegeneracy condition: if we set $u = q - p$ and $v = r - p$ then

$$\left| v - \left(\frac{u}{|u|} \cdot v\right) \frac{u}{|u|} \right| \geq \tau |v|$$

(vii) $L = \sup\big\{|p - q| : \langle pq\rangle \in \mathcal{T}_1\big\}$

(viii) for edges in $\mathcal{T}_1$ we have the uniform control on the ratio of lengths:

$$\inf\big\{|p - q| : \langle pq\rangle \in \mathcal{T}_1\big\} \geq \lambda L.$$

2.1.7 Fact [3] It is a standard fact about the geometry of smooth submanifolds that there are $0 < \tau < 1$ and $0 < \lambda < 1$ such that for arbitrarily small maximum edge lengths L there are τ, λ regular triangulations of $\mathcal{M}$ of maximum edge length L. We fix such τ and λ. We hereafter consider only τ, λ regular triangulations $\mathcal{T}$ with very small maximum edge length L. Once L is small the triangles $\overline{pqr}$ associated with $\langle pqr\rangle$ in $\mathcal{T}_2$ are very nearly parallel with the tangent plane to $\mathcal{M}$ at p.

2.1.8 Associated with each facet $\langle pqr\rangle$ in $\mathcal{T}_2$ is the *unit normal vector* $\mathbf{n}(pqr)$ to $\overline{pqr}$ having positive inner product with the normal $\mathbf{n}(p)$ to $\mathcal{M}$ at p.

2.1.9 Associated with each edge $\langle pq\rangle$ in $\mathcal{T}_1$ are exactly two distinct facets $\langle pqr\rangle$ and $\langle pqs\rangle$ in $\mathcal{T}_2$. We denote by

$$\mathbf{n}(pq) = \frac{\mathbf{n}(pqr) + \mathbf{n}(pqs)}{\left|\mathbf{n}(pqr) + \mathbf{n}(pqs)\right|}$$

the *average normal vector* at $\overline{pq}$.

For each $\langle pq\rangle$ we further denote by $\theta(pq)$ the *signed dihedral angle* at $\overline{pq}$ between the oriented plane directions of $\overline{pqr}$ and $\overline{pqs}$ which is characterized by the condition

$$2\sin\left(\frac{\theta(pq)}{2}\right)\mathbf{n}(pq) = V + W$$

where

- V is the unit exterior normal vector to $\overline{pqr}$ along edge $\overline{pq}$, so that, in particular,
$$V\cdot(p-q) = V\cdot\mathbf{n}(pqr) = 0;$$
- W is the unit exterior normal vector to $\overline{pqs}$ along edge $\overline{pq}$.

One checks that

$$\cos\theta(pq) = \mathbf{n}(pqr)\cdot\mathbf{n}(pqs).$$

Finally for each $\langle pq\rangle$ we denote by

$$g(pq) = |p-q|^{-1}\int_{\overline{pq}} g\,d\mathcal{H}^1 \in \mathbb{R}^3$$

the $\overline{pq}$ *average of* g; here $\mathcal{H}^1$ is one dimensional Hausdorff measure in $\mathbb{R}^3$.

2.1.10 Associated with our triangulation $\mathcal{T}$ of $\mathcal{M}$ is the *polyhedral approximation*

$$\mathcal{N}[\mathcal{T}] = \cup\{\overline{pqr} : \langle pqr\rangle \in \mathcal{T}_2\}$$

and the integral varifold

$$V[\mathcal{T}] = \sum_{\langle pqr\rangle\in\mathcal{T}_2}\mathbf{v}\big(\overline{pqr}\big) = \mathbf{v}\big(\mathcal{N}(\mathcal{T})\big)$$

whose first variation distribution is representable by integration

$$\delta V[\mathcal{T}] = \sum_{\langle pq\rangle\in\mathcal{T}_1}\mathcal{H}^1 \llcorner \overline{pq}\wedge\left[2\,\sin\left(\frac{\theta(pq)}{2}\right)\right]\mathbf{n}(pq)$$

[1, 4.3.5] so that

$$\delta V[\mathcal{T}](g) = \sum_{\langle pq\rangle\in\mathcal{T}_1}\Big[|p-q|\Big]\left[2\,\sin\left(\frac{\theta(pq)}{2}\right)\right]\Big[\mathbf{n}(pq)\cdot g(pq)\Big].$$

2.2 Terminology and facts for a flow of manifolds $\mathcal{M}_t$

2.2.1 As in 2.1.1 we suppose that $\mathcal{M} \subset \mathbb{R}^3$ is a compact connected smooth two dimensional submanifold of $\mathbb{R}^3$ without boundary oriented by a smooth Gauss mapping $\mathbf{n}\colon \mathcal{M} \to \mathbb{S}^2$ of unit normal vectors. We suppose additionally that $\varphi\colon (-1,\, 1) \times \mathcal{M} \to \mathbb{R}^3$ is a smooth mapping with $\varphi(0,\, p) = p$ for each $p \in \mathcal{M}$. For each t we set

$$\varphi[t] = \varphi(t,\, \cdot)\colon \mathcal{M} \to \mathbb{R}^3 \quad \text{and} \quad \mathcal{M}_t = \varphi[t](\mathcal{M}).$$

Our principal assumption is that, for each t, the mapping $\varphi[t]\colon \mathcal{M} \to \mathcal{M}_t$ *is an orientation preserving isometric imbedding (of Riemannian manifolds).* In particular, each $\mathcal{M}_t \subset \mathbb{R}^3$ is a compact connected smooth two dimensional submanifold of $\mathbb{R}^3$ without boundary oriented by a smooth Gauss mapping $\mathbf{n_t}\colon \mathcal{M}_t \to \mathbb{S}^2$ of unit normal vectors.

2.2.2 As in 2.1.2, for each t, we denote by $H_t\mathbf{n}_t$ the mean curvature vector field of $\mathcal{M}_t$.

2.2.3 As in 2.1.3, for each t we denote by U_t a suitable neighborhood of $\mathcal{M}_t$ in $\mathbb{R}^3$ in which a smooth nearest point retraction mapping $\rho_t\colon U_t \to \mathcal{M}_t$ is well defined together with smooth signed distance function $\sigma_t\colon U_t \to \mathbb{R}$; also we set $g[t] = \nabla\sigma_t\colon U_t \to \mathbb{R}^3$ as an initial velocity vector field.

2.2.4 By a convenient abuse of notation we assume that we can define a smooth map

$$\varphi\colon (-1,\, 1) \times U_0 \to \mathbb{R}^3,$$

$$\varphi(t,\, p) = \varphi\big(t,\, \rho_0(p) + \sigma_0(p)\mathbf{n}_0(\rho(p))\big) = \varphi\big(t,\, \rho_0(p)\big) + \sigma_0(p)\mathbf{n}_t(\rho_0(p))\big)$$

for each t and p. With $\varphi[t] = \varphi(t,\, \cdot)$ we have $\varphi[0] = \mathbf{1}_{U_0}$ and, additionally, $\sigma_0(p) = \sigma_t\big(\varphi[t](p)\big)$. We further assume that

$$U_t = \varphi[t]\, U_0$$

for each t.

2.2.5 Fact If we replace our initial $\varphi[t]\colon \mathcal{M} \to \mathbb{R}^3$'s by $\varphi[\mu t]$ for large enough μ (equivalently, restrict times t to $-1/\mu < t < 1/\mu$) and decrease the size of U_0 then the extended $\varphi[t]\colon U_0 \to \mathbb{R}^3$'s will exist. Such restrictions do not matter in the proof of our main assertion, since it is local in time and requires only small neighborhoods of the $\mathcal{M}_t$'s.

2.1.6 As in 2.1.4, for each t we denote by

$$V_t = \mathbf{v}(\mathcal{M}_t)$$

the integral varifold associated with $\mathcal{M}_t$.

2.2.7 We fix $0 < \tau < 1/2$ and $0 < \lambda < 1/2$ as in 2.1.7 and fix 2τ, 2λ regular triangulations $\mathcal{T}(1)$, $\mathcal{T}(2)$, $\mathcal{T}(3)$, $\ldots$ of $\mathcal{M}$ having maximum edge lengths $L(1)$, $L(2)$, $L(3)$ $\ldots$ respectively with $\lim_{j\to\infty} L(j) = 0$. For each j, the vertexes of $\mathcal{T}(j)$ are denoted $\mathcal{T}_0(j)$, the edges are denoted $\mathcal{T}_1(j)$, and the facets are denoted $\mathcal{T}_2(j)$. For all large j and each t we have triangulations $\mathcal{T}(1,\,t)$, $\mathcal{T}(2,\,t)$, $\mathcal{T}(3,\,t)$, $\ldots$ of $\mathcal{M}_t$ as follows. With notation similar to that above we specify, for each j and t,

$$\mathcal{T}_0(j,\,t) = \Big\{\varphi[t](p) : p \in \mathcal{T}_0(j)\Big\}, \quad \mathcal{T}_1(j,\,t) = \Big\{\big\langle \varphi[t](p)\,\varphi[t](q)\big\rangle : \langle pq\rangle \in \mathcal{T}_1(j)\Big\},$$

$$\mathcal{T}_2(j,\,t) = \Big\{\big\langle \varphi[t](p)\,\varphi[t](q)\,\varphi[t](r)\big\rangle : \langle pqr\rangle \in \mathcal{T}_2(j)\Big\}.$$

2.2.8 Fact If we replace $\varphi[t]$ by $\varphi[\mu t]$ for large enough μ (equivalently, restrict times t to $-1/\mu < t < 1/\mu$) then $\mathcal{T}(1,\,t)$, $\mathcal{T}(2,\,t)$, $\mathcal{T}(3,\,t)$, $\ldots$ will a sequence of τ,λ regular triangulations of $\mathcal{M}$ with maximum edge lengths $L(j,\,t)$ converging to 0 uniformly in time t as $j\to\infty$. Such restrictions do not matter in the proof of our main assertion, since it is local in time. We assume this has been done, if necessary, and that each of the triangulations $\mathcal{T}(j,\,t)$ is τ,λ regular with maximum edge lengths $L(j,\,t)$ converging to 0 as indicated.

2.2.9 As in 2.1.8 we associate with each j, t, and $\langle pqr\rangle \in \mathcal{T}_2(j)$ a unit normal vector $\mathbf{n}[t,\,j](pqr)$ to $\overline{\varphi[t](p)\,\varphi[t](q)\,\varphi[t](r)}$. As in 2.1.9 we associate with each j, t, and $\langle pq\rangle \in \mathcal{T}_1(j)$ an average normal vector $\mathbf{n}[t,\,j](pq)$ at $\overline{\varphi[t](p)\,\varphi[t](q)}$ and a signed dihedral angle $\theta[t,\,j](pq)$ at $\overline{\varphi[t](p)\,\varphi[t](q)}$ and the $\overline{\varphi[t](p)\,\varphi[t](q)}$ average $g[t,\,j](pq)$ of $g[t]$.

2.2.10 As in 2.1.10 we associate with each triangulation $\mathcal{T}(j,\,t)$ of $\mathcal{M}_t$ a *polyhedral approximation* $\mathcal{N}[\mathcal{T}(j,\,t)]$ and an integral varifold

$$V[\mathcal{T}(j,\,t)] = \mathbf{v}\big(\mathcal{N}[\mathcal{T}(j,\,t)]\big) = \sum_{\langle pqr\rangle\in\mathcal{T}_1(j)} \mathbf{v}\left(\overline{\varphi[t](p)\,\varphi[t](q)\,\varphi[t](r)}\right)$$

with first variation distribution

$$\delta V[\mathcal{T}(j,\,t)] = \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \mathcal{H}^1 \llcorner \left[\overline{\varphi[t])p)\,\varphi[t](q)}\right] \wedge \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right] \mathbf{n}[t,\,j](pq).$$

so that

$$\begin{aligned}&\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)\\ &= \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]\Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big].\end{aligned}$$

2.2.11 The quantity we wish to show is constant in time is

$$\int_{\mathcal{M}_t} H_t\, d\mathcal{H}^2 = -\left(\frac{1}{2}\right)\delta V_t\big(g[t]\big).$$

Since, for each time t,

$$V_t = \lim_{j\to\infty} V[\mathcal{T}(j,\,t)] \qquad \text{(as varifolds)}$$

we know, for each t,

$$\delta V_t\big(g[t]\big) = \lim_{j\to\infty}\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big).$$

We are thus led to seek to estimate

$$\frac{d}{dt}\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)$$

using the formula in 2.2.10. A key equality it provided by Schlafli's theorem mentioned above which, in the present terminology, asserts for each j and t,

$$\sum_{\langle pq\rangle\in\mathcal{T}_1(j)}\Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\frac{d}{dt}\Big[\theta[t,\,j](pq)\Big] = 0.$$

2.2.12 Fact Since, for each $\langle ppq\rangle$ in $\mathcal{T}_2(j)$, $\partial\langle pqr\rangle$ consists of exactly three edges, and, for each $\langle pq\rangle$ in $\mathcal{T}_1(j)$, there are exactly two distinct facets $\langle pqr\rangle$ in $\mathcal{T}_2(j)$ for which $\langle pq\rangle\in\partial\langle pqr\rangle$ we infer that, for each j,

$$card\big[\mathcal{T}_1(j)\big] = \frac{3}{2}card\big[\mathcal{T}_2(j)\big].$$

We then use the τ,λ regularity of the the $\mathcal{T}(j)$'s to check that that, for each time t and each $\langle ppq\rangle$ in $\mathcal{T}_2(j)$ the following four numbers have bounded ratios (independent of j, t, and $\langle ppq\rangle$) with each other

$$\mathcal{H}^2\left(\overline{\varphi[t](p)\,\varphi[t](q)\,\varphi[t](r)}\right),\qquad \big|\varphi[t](p)-\varphi[t](q)\big|^2,\qquad L(j,\,t)^2,\qquad L(j)^2.$$

Since

$$\lim_{j\to\infty}\mathcal{H}^2\big(\mathcal{N}[j,\,t]\big) = \mathcal{H}^2\big(\mathcal{M}_t\big) = \mathcal{H}^2\big(\mathcal{M}\big),$$

we infer

$$\sup_j \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} L(j)^2 < \infty,\qquad \lim_{j\to\infty}\sum_{\langle pq\rangle\in\mathcal{T}_1(j)} L(j)^3 = 0.$$

3 Modifications of the flow

3.1 Justification for computing with modified flows

As indicated in 2.2, we wish to estimate the time derivatives of

$$\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)$$
$$= \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]\Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big].$$

In each of the $\langle pq\rangle$ summands, each of the three factors

$$\Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big],\quad \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right],\quad \Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big]$$

is an intrinsic geometric quantity (at each time) whose value does not change under isometries of the ambient $\mathbb{R}^3$. With $\langle pqr\rangle$ and $\langle pqs\rangle$ denoting the two facets sharing edge $\langle pq\rangle$, we infer that each of the factors depends at most on the relative positions of $\varphi[t](p)$, $\varphi[t](q)$, $\varphi[t](r)$, $\varphi[t](s)$ and $\varphi[t]\mathcal{M}$. Suppose ψ: $(-1,\,1)\times\mathbb{R}^3\to\mathbb{R}^3$ is continuously differentiable, and for each t, the function $\psi[t]=\psi(t,\,\cdot)$: $\mathbb{R}^3\to\mathbb{R}^3$ is an isometry. Suppose further, we set

$$\varphi^*(t,\,p)=\psi\big(t,\,\varphi(t,\,p)\big),\quad \varphi^*[t]=\varphi^*(t,\,\cdot)$$

for each t and p so that $\varphi^*[t]=\psi[t]\circ\varphi[t]$. If we replace $\mathcal{M}$ by $\mathcal{M}^*=\psi[0]\mathcal{M}$ and φ by φ^* then we could follow the procedures of 2.1 and 2.2 to construct triangulations and polyhedral approximations $\mathcal{T}^*[j,\,t]$ and varifolds V^*, etc. with

$$\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)=\delta V^*[\mathcal{T}^*(j,\,t)]\big(g^*[t]\big).$$

Not only do we have equality in the sum, but, for each $\langle pq\rangle$ the corresponding summands are identical numerically. Hence, in evaluating $\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)$ we are free to (and will) use a different ψ and φ^* for each summand.

3.2 Conventions for derivatives

Suppose W is an open subset of $\mathbb{R}^M$ and $f=\big(f^1,\,f^2,\,\ldots,\,f^N\big)$: $W\to\mathbb{R}^N$ is K times continuously differentiable. We denote by

$$|||D^K f|||$$

the supremum of the partial derivatives

$$\frac{\partial^k f^K}{\partial x_{i(1)}\partial x_{i(2)}\ldots\partial x_{i(K)}}(p)$$

corresponding to all points $p\in W$, all $\big\{i(1),\,i(2),\,\ldots,\,i(K)\big\}\subset\big\{1,\,\ldots,\,M\big\}$ and $k=1,\,\ldots,N$, all choices of orthonormal coordinates $(x_1,\,\ldots,\,x_M)$ for $\mathbb{R}^M$ and all choices of orthonormal coordinates $(y_1,\,\ldots,\,y_N)$ for $\mathbb{R}^N$.

3.3 Conventions for inequalities

In making various estimates we will use use the largest edge length of the jth triangulation, typically called L, and a general purpose constant C. The constant C will have different values in different contexts (even in the same formula). What is implied is that, with $\mathcal{M}$ and φ fixed, the constants C can be chosen independent of the level of triangulation (once it is fine enough) and independent of time t and independent of the various modifications of our flow which are used in obtaining our estimates. As a representative example of our terminology, the expression

$$A = B \pm CL^2$$

means

$$-CL^2 \le A - B \le CL^2.$$

3.4 Fixing a vertex at the origin

Suppose p is a vertex in $\mathcal{M}$ and

$$\varphi_*(-1,\, 1) \times U_0 \to \mathbb{R}^3, \qquad \varphi_*(t,\, q) = \varphi(t,\, q) - \varphi(t,\, p) \quad \text{for each } q.$$

Then $\varphi^*(t,\, p) = (0,\, 0,\, 0)$ for each t. One checks, for $K = 0,\, 1,\, 2,\, 3$ that

$$|||\, D^K \varphi^* \,||| \le 2|||\, D^K \varphi \,|||, \qquad |||\, D^K \varphi^*[t] \,||| = |||\, D^K \varphi[t] \,|||$$

for each t.

3.5 Mapping a frame to the basis vectors

Suppose $(0,\, 0,\, 0) \in \mathcal{M}$ and that $\mathbf{e}_1$ and $\mathbf{e}_2$ are tangent to $\mathcal{M}$ at $(0,\, 0,\, 0)$. Suppose also $\varphi(t,\, 0,\, 0,\, 0) = (0,\, 0,\, 0)$ for each t. Then the mapping φ^* given by setting

$$\varphi^*[t] = \begin{pmatrix} \frac{\partial\varphi^1}{\partial x_1}(t,\, 0,\, 0,\, 0) & \frac{\partial\varphi^2}{\partial x_1}(t,\, 0,\, 0,\, 0) & \frac{\partial\varphi^3}{\partial x_1}(t,\, 0,\, 0,\, 0) \\ \frac{\partial\varphi^1}{\partial x_2}(t,\, 0,\, 0,\, 0) & \frac{\partial\varphi^2}{\partial x_2}(t,\, 0,\, 0,\, 0) & \frac{\partial\varphi^3}{\partial x_2}(t,\, 0,\, 0,\, 0) \\ \frac{\partial\varphi^1}{\partial x_3}(t,\, 0,\, 0,\, 0) & \frac{\partial\varphi^2}{\partial x_3}(t,\, 0,\, 0,\, 0) & \frac{\partial\varphi^3}{\partial x_3}(t,\, 0,\, 0,\, 0) \end{pmatrix} \circ \varphi[t]$$

satisfies

$$\varphi^*[t](0,\, 0,\, 0) = (0,\, 0,\, 0), \quad D\varphi^*[t](0,\, 0,\, 0) = \mathbf{1}_{\mathbf{R}^3}$$

with

$$|||D^K \varphi^*[t]||| = |||D^K \varphi[t]|||$$

for each $K = 1,\, 2,\, 3$ and each t, and

$$\left|\left|\left| \frac{\partial\varphi^*}{\partial t}(t,\, \cdot) \right|\right|\right| \le 3\Big(|||D^0\varphi||| \cdot |||D^2\varphi||| + |||D^1\varphi[t]|||^2 \Big).$$

3.6 Theorem *There is $C < \infty$ such that the following is true for all sufficiently small $\delta > 0$. Suppose $\gamma_0\colon [0, \delta] \to \mathcal{M}$ is an arc length parametrization of a length minimizing geodesic in $\mathcal{M}$ and set*

$$\gamma(s, t) = \varphi[t]\big(\gamma_0(s)\big) \quad \text{for each } s \text{ and } t$$

so that $s \to \gamma(s, t)$ is an arc length parametrization of a geodesic in $\mathcal{M}_t$. We also set

$$r(s, t) = \big|\gamma(0, t) - \gamma(s, t)\big| \quad \text{for each } s \text{ and } t$$

and, for (fixed) $0 < R < \delta$, consider

$$r(R, t) = \big|\gamma(0, t) - \gamma(R, t)\big| \quad \text{for each } t.$$

Then

$$\frac{d}{dt} r(R, t) = \pm C R^2$$

and

$$\lim_{R \downarrow 0} R^{-1} \frac{d}{dt} r(R, t) = 0.$$

Proof We will show

$$\frac{d}{dt} r(R, t)\Big|_{t=0} = \pm C R^2.$$

Step 1 Replacing $\varphi(t, p)$ by $\varphi^*(t, p) = \varphi(t, p) - \varphi(t, \gamma_0(0))$ as in 3.4 if necessary we assume without loss of generality that $\gamma(0, t) = (0, 0, 0)$ for each t.

Step 2 Rotating coordinates if necessary we assume without loss of generality that $\mathbf{e}_1$ and $\mathbf{e}_2$ are tangent to $\mathcal{M}_0$ at $(0, 0, 0)$ and that $\gamma_0'(0) = \mathbf{e}_1$

Step 3 Rotating coordinates as time changes as in 3.5 if necessary we assume without loss of generality that $D\varphi[t](0, 0, 0) = \mathbf{1}_{\mathbf{R}^3}$ for each t.

Step 4 We define

$$X(s, t) = \gamma(s, t) \cdot \mathbf{e}_1, \quad Y(s, t) = \gamma(s, t) \cdot \mathbf{e}_2, \quad Z(s, t) = \gamma(s, t) \cdot \mathbf{e}_3$$

so that

$$\gamma(s, t) = \big(X(s, t), Y(s, t), Z(s, t)\big)$$

and estimate for each s and t:

(a) $X(0, t) = Y(0, t) = Z(0, t) = 0$ (by step 1)

(b) $X_t(0,\,0) = Y_t(0,\,0) = Z_t(0,\,0) = 0$

(c) $X_s(s,\,t)^2 + Y_s(s,\,t)^2 + Z_s(s,\,t)^2 = 1$

(d) $X_s(s,\,t) = \pm 1,\; Y_s(s,\,t) = \pm 1,\; Z_s(s,\,t) = \pm 1$

(e) $1/2 \le r(s,\,t)/|s| \le 1$ (since δ is small)

(f) $X(s,\,0) = \pm Cs,\; Y(s,\,0) = \pm Cs,\; Z(s,\,0) = \pm Cs$

(g) $X_s(0,\,t) = X_s(0,\,0),\; Y_s(0,\,t) = Y_s(0,\,0),\; Z_s(0,\,t) = Z_s(0,\,0)$ (by step 3)

(h) $X_{st}(0,\,0) = Y_{st}(0,\,0) = Z_{st}(0,\,0) = 0$

$$(i) \qquad X_{st}(s,\,0) = X_{st}(0,\,0) + \int_0^s X_{sst}(\eta,\,0)\,d\eta = 0 \pm s \,\sup\big|X_{sst}\big| = \pm Cs,$$
$$Y_{st}(s,\,0) = \pm Cs, \qquad Z_{st}(s,\,0) = \pm Cs$$

$$(j) \qquad X_t(s,\,0) = X_t(0,\,0) + \int_0^s X_{st}(\eta,\,0)\,d\eta = 0 \pm Cs^2,$$
$$Y_t(s,\,0) = \pm Cs^2, \qquad Z_t(s,\,0) = \pm Cs^2$$

(k) $r^2 = X^2 + Y^2 + Z^2$

$$(\ell) \qquad rr_s = XX_s + YY_s + ZZ_s, \quad r_s = \frac{1}{r}\big(XX_s + YY_s + ZZ_s\big)$$

$$(m) \qquad rr_t = XX_t + YY_t + ZZ_t, \quad r_t = \frac{1}{r}\big(XX_t + YY_t + ZZ_t\big)$$

(n) $r_s r_t + r r_{st} = X_s X_t + X X_{st} + Y_s Y_t + Y Y_{st} + Z_s Z_t + Z Z_{st}$

(o) evaluating (n) at $t = 0,\; r > 0$ we see

$$\frac{1}{r(s,\,0)^2}\big((\pm Cs)(\pm 1)\big)\big((\pm Cs)(\pm Cs^2)\big) + r(s,\,0)r_{st}(s,\,0)$$
$$= (\pm 1)(\pm Cs^2) + (\pm Cs)(\pm Cs)$$

(p) $r_{st}(s,\,0) = \pm Cs$

$$(q) \qquad r_t(R,\,0) = r_t(0,\,0) + \int_0^R r_{st}(s,\,0)\,ds = 0 + \int_0^R \pm Cs\,ds = \pm CR^2.$$

□

3.7 Corollary *Suppose triangulation $\mathcal{T}(j)$ has maximum edge length $L = L(j)$ and $\langle pq \rangle$ is an edge in $\mathcal{T}_1(j)$. Then, for each t,*

$$\left| \varphi[t](p) - \varphi[t](q) \right| = \pm CL \quad \text{and} \quad \frac{d}{dt}\left| \varphi[t](p) - \varphi[t](q) \right| = \pm CL^2.$$

3.8 Stabilizing the facets of an edge

Suppose $\mathcal{T}(j)$ is a triangulation with maximum edge length $L = L(j)$ and that $\langle ABC \rangle$, $\langle ACD \rangle$ are facets in $\mathcal{T}_2(j)$ as illustrated

$$\begin{array}{ccccc} & & D = (e, f, 0) & & \\ & \swarrow & & \nwarrow & \\ (0,0,0) = A & & \longleftrightarrow & & C = (d, 0, 0)\,. \\ & \searrow & & \nearrow & \\ & & B = (a, b, c) & & \end{array}$$

Interchanging B and D if necessary we assume without loss of generality the the average normal $\mathbf{n}[0,\, AC]$ to $\mathcal{M}_0$ at A has positive inner product with $(C - A) \times (D - A)$.

1) Fixing A at the origin Modifying φ if necessary as in 3.4 if necessary we can assume without loss of generality that $\varphi[t](A) = (0,\, 0,\, 0)$ for each t. As indicated there, various derivative bounds are increased by, at most, a controlled amount.

2) Convenient rotations We set $u(t) = \varphi[t](C)$, $\quad v(t) = \varphi[t](D)$ and use the Gramm–Schmidt orthonormalization process to construct

$$U(t) = \frac{u(t)}{|u(t)|}, \quad V(t) = \frac{v(t) - v(t) \cdot U(t)\, U(t)}{|v(t) - v(t) \cdot U(t)\, U(t)|}, \quad W(t) = U(t) \times V(t).$$

One uses the mean value theorem in checking

$$|||D^K U(t)||| \le C \left(\sum_{j=0}^{K+1} |||D^j \varphi||| \right), \quad \text{etc}$$

for each $K = 0,\, 1,\, 2$. We denote by $Q(t)$ the orthogonal matrices having columns equal to $U(t)$, $V(t)$, $W(t)$ respectively (which is the inverse matrix to its transpose). Replacing φ_t by $Q(t) \circ \varphi_t$ if necessary, we assume without loss of generality that there are functions $a(t)$, $b(t)$, $c(t)$, $d(t)$, $e(t)$, $f(t)$, such that

$$\varphi[t](A) = (0,\, 0,\, 0), \quad \varphi[t](B) = (a(t),\, b(t),\, c(t)),$$
$$\varphi[t](C) = (d(t),\, 0,\, 0), \quad \varphi[t](D) = (e(t),\, f(t),\, 0).$$

We assume without loss of generality the existence of functions $F[t]\big(x,\,y\big)$ defined for $(x,\,y)$ near $(0,\,0)$ such that, near $(0,\,0,\,0)$ our manifold $\mathcal{M}_t$ is the graph of $F[t]$. In particular,

$$c(t) = F[t]\big(a(t),\,b(t)\big).$$

We assert that if $|p| \le CL$, then

$$|F[t](p)| \le CL^2, \qquad |\nabla F[t](p)| \le CL. \tag{3.8.1}$$

To see this, first we note that $F[t](A) = F[t](C) = F[t](D) = 0$. Next we invoke Rolle's theorem to conclude the existence of c_1 on segment AD and c_2 on segment CD such

$$\left\langle \frac{D-A}{|D-A|},\, DF[t](c_1)\right\rangle = 0 = \left\langle \frac{D-C}{|D-C|},\, DF[t](c_2)\right\rangle.$$

Since $|p| \le CL$ we infer

$$\left\langle \frac{D-A}{|D-A|},\, DF[t](p)\right\rangle = \pm CL, \qquad \left\langle \frac{D-C}{|D-C|},\, DF[t](p)\right\rangle = \pm CL.$$

In view of 2.1.6(vi)(vii)(viii) and 2.2.7 we infer that $\mathbf{e}_1$ and $\mathbf{e}_2$ are bounded linear combinations of $(D-A)/|D-A|$ and $(D-C)/|D-C|$ from which we conclude that $|\nabla F[t](p)| \le CL$. This in turn implies that $|F[t](p)| \le CL^2$ as asserted.

Since

$$\frac{\partial}{\partial t}F[t](0,\,0) = 0$$

we infer

$$\frac{\partial}{\partial t}F[t](p) = \pm CL \tag{3.8.2}$$

and since

$$\frac{\partial}{\partial t}\left(\varphi[t](A)\cdot \mathbf{e}_3\right) = 0$$

we infer

$$c'(t) = \frac{\partial}{\partial t}F[t](a(t),\,b(t)) = \frac{\partial}{\partial t}\left(\varphi[t](B)\cdot \mathbf{e}_3\right) = \pm CL. \tag{3.8.3}$$

3.9 Proposition *Let L, A, B, C, D, a, b, c, d, e, f be as in 3.8. Then*

(1) $a'(t) = \pm CL^2$

(2) $b'(t) = \pm CL^2$

(3) $c'(t) = \pm CL$

(4) $d'(t) = \pm CL^2$

(5) $e'(t) = \pm CL^2$

(6) $f'(t) = \pm CL^2$.

Proof According to 3.7, if $r(t)$ denotes the distance between the endpoints of an edge of arc length L at time t, then

$$r'(t) = \pm CL^2.$$

(i) We invoke 3.7 directly to infer (4) above.

(ii) We apply 3.7 to the distance between $(0,\,0,\,0)$ and $(e,\,f,\,0)$ to infer

$$\frac{d}{dt}\big(e^2+f^2\big)^{\frac12} = \frac{\big(ee'+ff'\big)}{\big(e^2+f^2\big)^{\frac12}} = \pm CL^2, \qquad ee'+ff' = \pm CL^3.$$

(iii) We apply 3.7 to the distance between $(d,\,0,\,0)$ and $(e,\,f,\,0)$ to infer

$$\frac{d}{dt}\big((e-d)^2+f^2\big)^{\frac12} = \frac{\big(e-d)(e'-d')+ff'\big)}{\big((e-d)^2+f^2\big)^{\frac12}} = \pm CL^2,$$
$$(e-d)(e'-d')+ff' = \pm CL^3.$$

We subtract the first inequality from the second to infer

$$ed'-de'+dd' = \pm CL^3, \qquad de' \pm CL^3, \qquad e' = \pm CL^2.$$

Assertions (5) and (6) follow readily.

(iv) We apply 3.7 to the distance between $(0,\,0,\,0)$ and $(a,\,b,\,c)$ to infer

$$\frac{d}{dt}\big(a^2+b^2+c^2\big)^{\frac12} = \frac{\big(aa'+bb'+cc'\big)}{\big(a^2+b^2+c^2\big)^{\frac12}} = \pm CL^2, \qquad aa'+bb'+cc' = \pm CL^3.$$

(v) We apply 3.7 to the distance between $(d,\,0,\,0)$ and $(a,\,b,\,c)$ to infer

$$\frac{d}{dt}\big((a-d)^2+b^2+c^2\big)^{\frac12} = \frac{\big((a-d)(a'-d')+bb'+cc'\big)}{\big((a-d)^2+b^2+c^2\big)^{\frac12}} = \pm CL^2,$$
$$(a-d)(a'-d')+bb'+cc' = \pm CL^3.$$

We subtract the first inequality form the second to infer

$$ad'-da'+dd' = \pm CL^3, \qquad da' \pm CL^3, \qquad a' = \pm CL^2,$$

which gives assertion (1).

(vi) We estimate from 3.8 that

$$c = F[t](a,\,b) = \pm CL^2, \quad c' = \frac{d}{dt}F[t](a,\,b) + \nabla F[t](a,\,b)\cdot(a',\,b') = \pm CL,$$

which gives (3) above. We have also $cc' = \pm CL^3$. We recall (iv) above and estimate

$$aa'+bb'+cc' = \pm CL^3, \quad bb' = \pm CL^3, \quad b' = \pm CL^2,$$

which is (2) above. ◻

3.10 Proposition *Suppose $\mathcal{T}(j)$ is a triangulation with maximum edge length $L = L(j)$ and $\langle pq \rangle$ is an edge in $\mathcal{T}_1(j)$. Abbreviate $\theta(t) = \theta[t, j](pq)$. Then, for each t,*

$$\theta(t) = \pm CL \tag{1}$$

$$2\sin\left(\frac{\theta(t)}{2}\right) = \pm CL \tag{2}$$

$$\theta'(t) = \pm C \tag{3}$$

$$\frac{d}{dt}\left[2\sin\left(\frac{\theta(t)}{2}\right)\right] = \pm C \tag{4}$$

$$\frac{d}{dt}\left[2\sin\left(\frac{\theta(t)}{2}\right) - \theta\right] = \pm CL^2. \tag{5}$$

Proof Making the modifications of 3.8 if necessary, we assume without loss of generality (in the terminology there) that $\varphi[t](p) = A = (0, 0, 0)$, $\varphi[t](q) = C = (d(t), 0, 0)$, and that there are $\langle pqB_* \rangle, \langle pqD_* \rangle \in \mathcal{T}_2(j)_0$ with $\varphi[t](B_*) = B = (a(t), b(t), c(t))$, $\varphi[t](D_*) = D = (e(t), f(t), 0)$.

The unit normal to $\overline{ACD}$ is $(0, 0, 1)$ while the unit normal to $\overline{ABC}$ is

$$\frac{(0, -c, b)}{(b^2 + c^2)^{\frac{1}{2}}}$$

so that $\cos\theta = \dfrac{b}{(b^2 + c^2)^{\frac{1}{2}}}$,

$$\sin\theta = \pm\left(1 - \cos^2\theta\right)^{\frac{1}{2}} = \pm\left(1 - \frac{b^2}{b^2 + c^2}\right)^{\frac{1}{2}} = \pm\frac{c}{(b^2 + c^2)^{\frac{1}{2}}} = \pm CL$$

in view of 3.8. Assertions (1) and (2) follow. We compute further

$$(\sin\theta)' = \cos\theta\,\theta' = \pm\frac{(b^2 + c^2)^{\frac{1}{2}}c' - c\frac{bb' + cc'}{(b^2+c^2)^{\frac{1}{2}}}}{b^2 + c^2} = \pm C$$

in view of 3.9(1)(2)(3) and 3.8. Assertion (3) and (4) follow. Assertion (5) follows from differentiation and assertions (1) and (3). □

3.11 Proposition *Suppose $\mathcal{T}(j)$ is a triangulation with maximum edge length $L = L(j)$ and $\langle pq\rangle$ is an edge in $\mathcal{T}_1(j)$. Then*

(1) $\mathbf{n}[t,\, j](pq) = \big(0,\, \pm CL,\, 1 \pm CL^4\big)$

(2) $(d/dt)\big(\mathbf{n}[t,\, j](pq)\big) = \big(0,\, \pm C,\, \pm CL\big) + \big(\pm CL,\, \pm CL,\, \pm CL\big)$

(3) $g[t,\, j](pq) = \big(\pm CL,\, \pm CL,\, 1 \pm CL^2\big)$

(4) $(d/dt)g[t,\, j](pq) = \big(\pm C,\, \pm C,\, 0\big) + \big(\pm CL,\, \pm CL,\, \pm CL\big)$

(5) $\mathbf{n}[t,\, j](pq) \cdot g[t,\, j](pq) = 1 \pm CL^2$

(6) $(d/dt)\Bigg(\mathbf{n}[t,\, j](pq) \cdot g[t,\, j](pq)\Bigg) = \pm CL$

(7) $1 - \mathbf{n}[t,\, j](pq) \cdot g[t,\, j](pq) = \pm CL^2$.

Proof We let $A,\ B,\ C,\ D,\ F[t],\ b(t),\ c(t),\ d(t)$ be as in 3.8. We abbreviate $\mathbf{n} = \mathbf{n}[t,\, j](pq)$ and estimate

$$\begin{aligned}\mathbf{n} &= \frac{(0,\,0,\,1) + (0,\,-c,\,b)/(b^2+c^2)^{\frac12}}{\big|(0,\,0,\,1) + (0,\,-c,\,b)/(b^2+c^2)^{\frac12}\big|} \\ &= \frac{\big(0,\,-c,\,b + (b^2+c^2)^{\frac12}\big)}{2^{\frac12}\big(b^2+c^2+b(b^2+c^2)^{\frac12}\big)^{\frac12}}.\end{aligned}$$

The first assertion follows from 3.8.1. We differentiate to conclude $\mathbf{n}' =$

$$\begin{aligned}&\frac{\pm CL\big(0,\,-c',\,b' \pm C(bb'+cc')/L - (L/L)\big(bb'+cc' \pm b'L + \pm C(b/L)(bb'+cc')}{\pm L^2} \\ &\quad = \big(0,\,\pm C,\,\pm CL\big) + \big(\pm CL,\,\pm CL,\,\pm CL\big)\end{aligned}$$

in view of 3.9(2)(3). This is assertion (2).

We abbreviate $g = g[t,\, j](pq)$ and estimate

$$\begin{aligned}g &= \frac{1}{d(t)}\int_0^{d(t)} \frac{\big(-F[t]_x,\,-F[t]_y,\,1\big)}{\big|\big(-F[t]_x,\,-F[t]_y,\,1\big)\big|} \\ &= \frac{1}{d(t)}\int_0^{d(t)} \frac{\big(-F[t]_x,\,-F[t]_y,\,1\big)}{\Bigg(\big(F[t]_x^2 F[t]_y^2 + 1\Bigg)^{\frac12}}.\end{aligned}$$

The third assertion follows from 3.8.1. We differentiate to estimate that dg/dt equals

$$\frac{-d'}{d^2}\int_0^{d(t)} \frac{\big(-F[t]_x,\,-F[t]_y,\,1\big)}{\big(1+F[t]_x^2+F[t]_y^2\big)^{\frac12}} + \frac{d'}{d}\frac{\big(-F[t]_x,\,-F[t]_y,\,1\big)}{\big(1+F[t]_x^2+F[t]_y^2\big)^{\frac12}}$$
$$+\frac{1}{d}\int_0^{d} \frac{\pm CL\big(-F[t]_{tx},\,-F[t]_{ty},\,0\big)}{1+F[t]_x^2+F[t]_y^2}$$
$$-\frac{1}{d}\int_0^{d} \frac{\big(-F[t]_x,\,-F[t]_y,\,1\big)(\pm C/L)\big(F[t]_xF[t]_{tx}+F[t]_yF[t]_{ty}\big)}{1+F[t]_x^2+F[t]_y^2} =$$
$$L\big(\pm C,\,\pm C,\,\pm C\big)+L\big(\pm C,\,\pm C,\,\pm C\big)+\big(\pm C,\,\pm C,\,0\big)+L\big(\pm C,\,\pm C,\,\pm C\big)$$

which gives assertion (4). Assertion (5) follows from assertions (1) and (3). Assertion (6) follows from assertions (1), (2), (3), (4) and integration by parts. Assertion (7) follows from assertions (1) and (3). □

4 Constancy of the mean curvature integral

4.1 The derivative estimates

Suppose triangulation $\mathcal{T}(j)$ has maximum edge length $L = L(j)$. We recall from 2.2.10 that

$$\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)$$
$$= \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\left[2\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]\Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big]$$

and we estimate, for each t that

$$\frac{d}{dt}\Big(\delta V[\mathcal{T}(j)_t]\big(g[t]\big)\Big)$$
$$= \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]'\left[2\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]\Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big]$$
$$+ \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\left[2\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]'\Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big]$$
$$+ \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\left[2\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]\Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big]'.$$

We assert that

$$\frac{d}{dt}\Big(\delta V[\mathcal{T}(j,\,t)]\big(g[t]\big)\Big) = \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \pm CL^3 = \sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \pm CL(j)^3.$$

To see this we will estimate each of the three summands above.

First summand We use 3.7, 3.10(2), 3.11(5) to estimate for each pq,

$$\Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]' \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right] \Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big] \\ = \big(CL^2\big)\big(CL\big)\big(1\pm CL^2\big).$$

Second summand We use 3.10(5), 3.11(7) to estimate for each pq,

$$\begin{aligned}
&\Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big] \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]' \Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big] \\
&= \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\Big[\theta[t,\,j](pq)\Big]' \\
&\quad + \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big] \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right) - \theta[t,\,j](pq)\right]' \\
&\quad + \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big] \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right]' \Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)-1\Big] \\
&= \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\Big[\theta[t,\,j](pq)\Big]' \pm \big(CL\big)\big(CL^2\big) \pm \big(CL\big)\big(C\big)\big(CL^2\big).
\end{aligned}$$

Third summand We use 3.10(2) and 3.11(6) to estimate

$$\Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big] \left[2\,\sin\left(\frac{\theta[t,\,j](pq)}{2}\right)\right] \Big[\mathbf{n}[t,\,j](pq)\cdot g[t,\,j](pq)\Big]' \\ = \big(CL\big)\big(CL\big)\big(CL\big).$$

According to Schlafli's formula [7],

$$\sum_{\langle pq\rangle\in\mathcal{T}_1(j)} \Big[\big|\varphi[t](p)-\varphi[t](q)\big|\Big]\Big[\theta[t,\,j](pq)\Big]' = 0.$$

Our assertion follows.

4.2 Main Theorem

(1) *For each fixed time* t,

$$\lim_{j\to\infty} \delta V[\mathcal{T}(j,\,t)]\big(g[t]\big) = \delta V_t\big(g[t]\big).$$

(2) *For each fixed* j, $\delta V[\mathcal{T}(j)_t]\big(g[t]\big)$ *is a differentiable function of* t *and*

$$\lim_{j\to\infty} \frac{d}{dt}\Big(\delta V[\mathcal{T}(j)_t]\big(g[t]\big)\Big) = 0$$

uniformly in t.

(3) *For each* t

$$\int_{\mathcal{M}_t} H_t\, d\mathcal{H}^2 = \int_{\mathcal{M}} H\, d\mathcal{H}^2.$$

This is the main result of this note.

Proof To prove the first assertion, we check that

$$(\rho_t)_\sharp V[\mathcal{T}(j,\,t)] = V_t$$

for each t and all large j. Indeed, the τ regularity of our triangulations implies that the normal directions of the $\mathcal{N}[\mathcal{T}(j)_t]$ are very nearly equal to the normal directions of nearby points on $\mathcal{M}_t$ and that the restriction of $D\rho_t$ to the tangent planes of the $\mathcal{N}[\mathcal{T}(j)_t]$ is very nearly an orthogonal injection. The first assertion follows with use of the first variation formula given in [14.1, 4.2]. Assertion (2) follows from 4.1 since

$$\sum_{\langle pq\rangle\in\mathcal{T}_1(j)} L(j)^2$$

is dominated by the area of $\mathcal{M}$ (see 2.2.12) and $\lim_{j\to\infty} L(j) = 0$. Assertion (3) follows from assertions (1) and (2) and our observation in 2.1.4. □

Acknowledgements Fred Almgren tragically passed away shortly after this note was written. Since then, the main result for smooth surfaces has been reproved in an easier way and generalized to the setting of Einstein manifolds by J-M Schlenker together with the second author of the current paper [6]. Nonetheless, it seems clear that the methods used here can be used to extend these results in other directions.

References

[1] **W K Allard**, *On the first variation of a varifold*, Annals of Math. 95 (1972) 417–491

[2] **M Berger**, **B Gostiaux**, *Geometrie differentielle: varietes, courbes et surfaces*, Presses Universitaires de France, III (1987)

[3] **J Cheeger**, **W Müller**, **R Schrader**, *On the curvature of piecewise flat spaces*, Comm. Math. Phys. 92 (1984) 405–454

[4] **R Connelly**, **I Sabitov**, **A Walz**, *The bellows conjecture*, Beiträge Algebra Geom. 38 (1997) 1–10

[5] **G Herglotz**, *Ueber der Starrheit der Eiflachen*, Abh. Math. Semin. Hansische Univ. 92 (1943) 127–129

[6] **I Rivin**, **J-M Schlenker**, *Schläfli formula and Einstein manifolds*, IHES preprint (1998)

[7] **D V Alekseevsky**, **È B Vinberg**, **A S Solodovnikov**, *Geometry of spaces of constant curvature*, from: "Geometry II", Encyclopaedia Math. Sci. 29, Springer–Verlag, Berlin (1993)

[8] **M Spivak**, *A Comprehensive Introduction to Differential Geometry*, (Second Edition) Publish or Perish, Berkeley (1979)

Mathematics Institute, University of Warwick
Coventry, CV4 7AL, UK

Email: `igor@maths.warwick.ac.uk`

Received: 10 May 1998

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997 **23**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 23–49

A brief survey of the deformation theory of Kleinian groups

James W Anderson

Abstract We give a brief overview of the current state of the study of the deformation theory of Kleinian groups. The topics covered include the definition of the deformation space of a Kleinian group and of several important subspaces; a discussion of the parametrization by topological data of the components of the closure of the deformation space; the relationship between algebraic and geometric limits of sequences of Kleinian groups; and the behavior of several geometrically and analytically interesting functions on the deformation space.

AMS Classification 30F40; 57M50

Keywords Kleinian group, deformation space, hyperbolic manifold, algebraic limits, geometric limits, strong limits

Dedicated to David Epstein on the occasion of his 60th birthday

1 Introduction

Kleinian groups, which are the discrete groups of orientation preserving isometries of hyperbolic space, have been studied for a number of years, and have been of particular interest since the work of Thurston in the late 1970s on the geometrization of compact 3–manifolds. A Kleinian group can be viewed either as an isolated, single group, or as one of a member of a family or continuum of groups.

In this note, we concentrate our attention on the latter scenario, which is the deformation theory of the title, and attempt to give a description of various of the more common families of Kleinian groups which are considered when doing deformation theory. No proofs are given, though it is hoped that reasonable coverage of the current state of the subject is given, and that ample references have been given for the interested reader to venture boldly forth into the literature.

It is possible to consider the questions raised here in much more general settings, for example for Kleinian groups in n–dimensions for general n, but that is beyond the scope of what is attempted here. Some material on this aspect of the question can be found in Bowditch [23] and the references contained therein.

The author would like to thank Dick Canary, Ed Taylor, and Brian Bowditch for useful conversations during the preparation of this work, as well as the referee for useful comments.

2 The deformation spaces

We begin by giving a few basic definitions of the objects considered in this note, namely Kleinian groups. We go on to define and describe the basic structure of the deformation spaces we are considering herein.

A *Kleinian group* is a discrete subgroup of $\mathrm{PSL}_2(\mathbf{C}) = \mathrm{SL}_2(\mathbf{C})/\{\pm \mathrm{I}\}$, which we view as acting both on the Riemann sphere $\overline{\mathbf{C}}$ by Möbius transformations and on real hyperbolic 3–space $\mathbf{H}^3$ by isometries, where the two actions are linked by the Poincaré extension.

The action of an infinite Kleinian group Γ partitions $\overline{\mathbf{C}}$ into two sets, the *domain of discontinuity* $\Omega(\Gamma)$, which is the largest open subset of $\overline{\mathbf{C}}$ on which Γ acts discontinuously, and the *limit set* $\Lambda(\Gamma)$. If $\Lambda(\Gamma)$ contains two or fewer points, Γ is *elementary*, otherwise Γ is *non-elementary*. For a non-elementary Kleinian group Γ, the limit set $\Lambda(\Gamma)$ can also be described as the smallest non-empty closed subset of $\overline{\mathbf{C}}$ invariant under Γ. We refer the reader to Maskit [68] or Matsuzaki and Taniguchi [71] as a reference for the basics of Kleinian groups.

An isomorphism $\varphi\colon \Gamma \to \Phi$ between Kleinian groups Γ and Φ is *type-preserving* if, for $\gamma \in \Gamma$, we have that γ is parabolic if and only if $\varphi(\gamma)$ is parabolic.

A Kleinian group is *convex cocompact* if its convex core is compact; recall that the *convex core* associated to a Kleinian group Γ is the minimal convex submanifold of $\mathbf{H}^3/\Gamma$ whose inclusion is a homotopy equivalence. More generally, a Kleinian group is *geometrically finite* if it is finitely generated and if its convex core has finite volume. This is one of several equivalent definitions of geometric finiteness; the interested reader is referred to Bowditch [22] for a complete discussion.

A Kleinian group Γ is *topologically tame* if its corresponding quotient 3–manifold $\mathbf{H}^3/\Gamma$ is homeomorphic to the interior of a compact 3–manifold. Geometrically finite Kleinian groups are topologically tame. It was conjectured by Marden [64] that all finitely generated Kleinian groups are topologically tame.

A compact 3–manifold M is *hyperbolizable* if there exists a hyperbolic 3–manifold $N = \mathbf{H}^3/\Gamma$ homeomorphic to the interior of M. Note that a hyperbolizable 3–manifold M is necessarily orientable; *irreducible*, in that every embedded 2–sphere in M bounds a 3–ball in M; and *atoroidal*, in that every embedded torus T in M is homotopic into ∂M. Also, since the universal cover $\mathbf{H}^3$ of N is contractible, the fundamental group of M is isomorphic to Γ. For a discussion of the basic theory of 3–manifolds, we refer the reader to Hempel [48] and Jaco [49].

Keeping to our viewpoint of a Kleinian group as a member of a family of groups, throughout this survey we view a Kleinian group as the image $\rho(G)$ of a representation ρ of a group G into $\mathrm{PSL}_2(\mathbf{C})$. Unless explicitly stated otherwise, we assume that G is finitely generated, torsion-free, and non-abelian, so that in particular $\rho(G)$ is non-elementary.

2.1 The representation varieties $\mathcal{HOM}(G)$ and $\mathrm{R}(G) = \mathcal{HOM}(G)/\mathrm{PSL}_2(\mathbf{C})$

The most basic of the deformation spaces is the *representation variety* $\mathcal{HOM}(G)$ which is the space of all representations of G into $\mathrm{PSL}_2(\mathbf{C})$ with the following topology. Given a set of generators $\{g_1, \ldots, g_k\}$ for G, we may naturally view $\mathcal{HOM}(G)$ as a subset of $\mathrm{PSL}_2(\mathbf{C})^k$, where a representation $\rho \in \mathcal{HOM}(G)$ corresponds to the k–tuple $(\rho(g_1), \ldots, \rho(g_k))$ in $\mathrm{PSL}_2(\mathbf{C})^k$. The defining polynomials of this variety are determined by the relations in G. In particular, if G is free, then $\mathcal{HOM}(G) = \mathrm{PSL}_2(\mathbf{C})^k$. It is easy to see that $\mathcal{HOM}(G)$ is a closed subset of $\mathrm{PSL}_2(\mathbf{C})^k$.

The representations in $\mathcal{HOM}(G)$ are *unnormalized*, in the sense that there is a natural free action of $\mathrm{PSL}_2(\mathbf{C})$ on $\mathcal{HOM}(G)$ by conjugation. Depending on the particular question being addressed, it is sometimes preferable to remove the ambiguity of this action and form the quotient space $\mathrm{R}(G) = \mathcal{HOM}(G)/\mathrm{PSL}_2(\mathbf{C})$.

Though a detailed description is beyond the scope of this survey, we pause here to mention work of Culler and Shalen [40], [41], in which a slight variant of the representation variety as defined above plays a fundamental role, and which has inspired further work of Morgan and Shalen [78], [79], [80] and Culler, Gordon, Luecke, and Shalen [39]. The basic object here is not the space $\mathrm{R}(G)$ of all representations of G into $\mathrm{PSL}_2(\mathbf{C})$ as defined above, but instead the related space $\mathrm{X}(G)$ of all representations of G into $\mathrm{SL}_2(\mathbf{C})$, modulo the action of $\mathrm{SL}_2(\mathbf{C})$. The introduction of this space $\mathrm{X}(G)$ does beg the question of when a representation of G into $\mathrm{PSL}_2(\mathbf{C})$ can be lifted to a representation of G into $\mathrm{SL}_2(\mathbf{C})$. We note in passing that this question of lifting representations has been considered by a number of authors, including Culler, Kra, and Thurston, to name but a few; we refer the reader to the article by Kra [61] for exact statements and a review of the history, including references.

By considering the global structure of the variety $\mathrm{X}(G)$ in the case that G is the fundamental group of a compact, hyperbolizable 3–manifold M, and in particular the ideal points of its compactification, Culler and Shalen [40] are able to analyze the actions of G on trees, which in turn has connections with the existence of essential incompressible surfaces in M, finite group actions on M, and has particular consequences in the case that M is the complement of a knot in $\mathbf{S}^3$. We refer the reader to the excellent survey article by Shalen [94], as well as to the papers cited above.

2.2 The spaces $\mathcal{HOM}_{\mathrm{T}}(G)$ and $\mathrm{R}_{\mathrm{T}}(G) = \mathcal{HOM}_{\mathrm{T}}(G)/\mathrm{PSL}_2(\mathbf{C})$ of the minimally parabolic representations

Let $\mathcal{HOM}_{\mathrm{T}}(G)$ denote the subspace of $\mathcal{HOM}(G)$ consisting of those representations ρ for which $\rho(g)$ is parabolic if and only if g lies in a rank two free abelian subgroup of G. We refer to $\mathcal{HOM}_{\mathrm{T}}(G)$ as the space of *minimally parabolic* representations of G. In particular, if G contains no $\mathbf{Z}\oplus\mathbf{Z}$ subgroups, then the image $\rho(G)$ of every ρ in $\mathcal{HOM}_{\mathrm{T}}(G)$ is *purely loxodromic*, in that every non-trivial element of $\rho(G)$ is loxodromic. Set $\mathrm{R}_{\mathrm{T}}(G) = \mathcal{HOM}_{\mathrm{T}}(G)/\mathrm{PSL}_2(\mathbf{C})$.

2.3 The spaces $\mathcal{D}(G)$ and $\mathrm{AH}(G) = \mathcal{D}(G)/\mathrm{PSL}_2(\mathbf{C})$ of discrete, faithful representations

Let $\mathcal{D}(G)$ denote the subspace of $\mathcal{HOM}(G)$ consisting of the discrete, faithful representations of G, that is, the injective homomorphisms of G into $\mathrm{PSL}_2(\mathbf{C})$ with discrete image. For the purposes of this note, the space $\mathcal{D}(G)$ is our universe, as it is the space of all Kleinian groups isomorphic to G. Set $\mathrm{AH}(G) = \mathcal{D}(G)/\mathrm{PSL}_2(\mathbf{C})$.

We note that there exists an equivalent formulation of $\mathrm{AH}(G)$ in terms of manifolds. Given a hyperbolic 3–manifold N, let $\mathrm{H}(N)$ denote the set of all pairs (f, K), where K is a hyperbolic 3–manifold and $f\colon N \to K$ is a homotopy equivalence, modulo the equivalence relation $(f, K) \sim (g, L)$ if there exists an orientation preserving isometry $\alpha\colon K \to L$ so that $\alpha \circ f$ is homotopic to g. The topology on $\mathrm{H}(N)$ is given by noting that, if we let $\Gamma \subset \mathrm{PSL}_2(\mathbf{C})$ be a choice of conjugacy class of the fundamental group of N, then each element (f, K) in $\mathrm{H}(N)$ gives rise to a discrete, faithful representation $\varphi = f_*$ of Γ into $\mathrm{PSL}_2(\mathbf{C})$, with equivalent points in $\mathrm{H}(N)$ giving rise to conjugate representations into $\mathrm{PSL}_2(\mathbf{C})$. Hence, equipping $\mathrm{H}(N)$ with this topology once again gives rise to $\mathrm{AH}(G)$ with $G = \pi_1(N)$.

The following theorem, due to Jørgensen, describes the fundamental property of $\mathcal{D}(G)$, namely that the limit of a sequence of elements of $\mathcal{D}(G)$ is again an element of $\mathcal{D}(G)$.

Theorem 2.1 (Jørgensen [53]) *$\mathcal{D}(G)$ is a closed subset of $\mathcal{HOM}(G)$.*

There is one notable case in which $\mathrm{AH}(G)$ is completely understood, namely in the case that G is the fundamental group of a compact, hyperbolizable 3–manifold M whose boundary is the union of a (possibly empty) collection of tori. In this case, the hyperbolic structure on the interior of M is unique, by the classical Rigidity Theorem of Mostow, for closed manifolds, and Prasad, for manifolds with non-empty toroidal boundary. Rephrasing this statement as a statement about deformation spaces yields the following.

Theorem 2.2 (Mostow [81] and Prasad [91]) *Suppose that G is the fundamental group of a compact, orientable 3–manifold M whose boundary is the union of a (possibly empty) collection of tori. Then, $\mathrm{AH}(G)$ either is empty or consists of a single point.*

Given this result, it will cause us no loss of generality to assume that henceforth all Kleinian groups have infinite volume quotients.

2.4 The spaces $\mathcal{P}(G)$ and $\mathrm{MP}(G) = \mathcal{P}(G)/\mathrm{PSL}_2(\mathbf{C})$ of geometrically finite, minimally parabolic representations

Let $\mathcal{P}(G)$ denote the subset of $\mathcal{D}(G)$ consisting of those representations ρ with geometrically finite, minimally parabolic image $\rho(G)$. In particular, if G contains no $\mathbf{Z} \oplus \mathbf{Z}$ subgroups, then the image $\rho(G)$ of every representation $\rho \in \mathcal{P}(G)$ is convex cocompact. Set $\mathrm{MP}(G) = \mathcal{P}(G)/\mathrm{PSL}_2(\mathbf{C})$, and note that since $\mathrm{PSL}_2(\mathbf{C})$ is connected, the quotient map gives a one-to-one correspondence between the connected components of $\mathcal{P}(G)$ and those of $\mathrm{MP}(G)$.

It is an immediate consequence of the Core Theorem of Scott [93] and the Hyperbolization Theorem of Thurston that if $\mathcal{D}(G)$ is non-empty, then $\mathcal{P}(G)$ is non-empty. For a discussion of the Hyperbolization Theorem, see Morgan [77], Otal and Paulin [90], or Otal [89] for the fibered case.

We note here that, if there exists a geometrically finite, minimally parabolic representation of G into $\mathrm{PSL}_2(\mathbf{C})$, then in general there exist many geometrically finite representations which are not minimally parabolic, which can be constructed as limits of the geometrically finite, minimally parabolic representations. This construction has been explored in detail for a number of cases by Maskit [69] and Ohshika [85].

In the case that G is itself a geometrically finite, minimally parabolic Kleinian group, the structure of $\mathrm{MP}(G)$ is fairly well understood, both as a subset of $\mathrm{AH}(G)$ and in terms of how the components of $\mathrm{MP}(G)$ are parametrized by topological data. We spend the remainder of this section making these statements precise.

We begin with the Quasiconformal Stability Theorem of Marden [64].

Theorem 2.3 (Marden [64]) *If G is a geometrically finite, minimally parabolic Kleinian group, then $\mathrm{MP}(G)$ is an open subset of $\mathrm{R}(G)$.*

As a converse to this, we have the Structural Stability Theorem of Sullivan [97]. We note here that the versions of the Theorems of Marden and Sullivan given here are not the strongest, but are adapted to the point of view taken in this paper. The general statements holds valid in slices of $\mathrm{AH}(G)$ in which a certain collection of elements of G are required to have parabolic image, not just those which belong to $\mathbf{Z}\oplus\mathbf{Z}$ subgroups.

Theorem 2.4 (Sullivan [97]) *Let G be a finitely generated, torsion-free, non-elementary Kleinian group. If there exists an open neighborhood of the identity representation in $\mathrm{R}(G)$ which lies in $\mathrm{AH}(G)$, then G is geometrically finite and minimally parabolic.*

Combining these, we see that $\mathrm{MP}(G)$ is the interior of $\mathrm{AH}(G)$. A natural question which arises from this is whether there are points of $\mathrm{AH}(G)$ which do not lie in the closure of $\mathrm{MP}(G)$.

Conjecture 2.5 (Density conjecture) *$\mathrm{AH}(G)$ is the closure of $\mathrm{MP}(G)$.*

This Conjecture is due originally to Bers in the case that G is the fundamental group of a surface, see Bers [14], and extended by Thurston to general G.

There has been a good deal of work in the past couple of years on the global structure of $\mathrm{MP}(G)$ and its closure. We begin with an example to show that there exist groups G for which $\mathrm{MP}(G)$ is disconnected; the example we give here, in which $\mathrm{MP}(G)$ has finitely many components, comes from the discussion in Anderson and Canary [6].

Let T be a solid torus and for large k, let $A_1, \ldots, A_k$ be disjoint embedded annuli in ∂T whose inclusion into T induces an isomorphism of fundamental groups. For each $1 \leq j \leq k$, let S_j be a compact, orientable surface of genus j with a single boundary component, and let $Y_j = S_j \times I$, where I is a closed interval. Construct a compact 3–manifold M by attaching the annulus $\partial S_j \times I$ in ∂Y_j to the annulus A_j in ∂T. The resulting 3–manifold M is compact and hyperbolizable 3–manifold and has fundamental group G. This 3–manifold is an example of a *book of I–bundles*. Let ρ be an element of $\mathrm{MP}(G)$ for which the interior of M is homeomorphic to $\mathbf{H}^3/\rho(G)$.

Let τ be a permutation of $\{1, \ldots, k\}$, and consider now the manifold M_τ obtained by attaching the annulus $\partial S_j \times I$ in ∂Y_j to the annulus $A_{\tau(j)}$ in ∂T. By construction, M_τ is compact and hyperbolizable, and has fundamental group G; let ρ_τ be an element of $\mathrm{MP}(G)$ for which the interior of M_τ is homeomorphic to $\mathbf{H}^3/\rho_\tau(G)$. Since M and M_τ have isomorphic fundamental groups, they are homotopy equivalent. However, in the case that τ is not some power of the cycle $(12 \cdots k)$, then there does not exist an orientation preserving homeomorphism between M and M_τ, and hence ρ and ρ_τ lie in different components of $\mathrm{MP}(G)$.

In the general case that G is finitely generated and does not split as a free product, there exists a characterization of the components of both $\mathrm{MP}(G)$ and its closure $\overline{\mathrm{MP}(G)}$ in terms of the topology of a compact, hyperbolizable 3–manifold M with fundamental group G. This characterization combines work of Canary and McCullough [33] and of Anderson, Canary, and McCullough [10]. We need to develop a bit of topological machinery before discussing this characterization.

For a compact, oriented, hyperbolizable 3–manifold M with non-empty, incompressible boundary, let $\mathcal{A}(M)$ denote the set of *marked homeomorphism types of compact, oriented* 3*–manifolds homotopy equivalent to* M. Explicitly, $\mathcal{A}(M)$ is the set of equivalence classes of pairs (M', h'), where M' is a compact, oriented, irreducible 3–manifold and $h'\colon M \to M'$ is a homotopy equivalence, and where two pairs (M_1, h_1) and (M_2, h_2) are equivalent if there exists an orientation preserving homeomorphism $j\colon M_1 \to M_2$ such that $j \circ h_1$ is homotopic to h_2. Denote the class of (M', h') in $\mathcal{A}(M)$ by $[(M', h')]$.

There exists a natural map $\Theta\colon \mathrm{AH}(\pi_1(M)) \to \mathcal{A}(M)$, defined as follows. For $\rho \in \mathrm{AH}(\pi_1(M))$, let M_ρ be a compact core for $N_\rho = \mathbf{H}^3/\rho(\pi_1(M))$ and let $r_\rho\colon M \to M_\rho$ be a homotopy equivalence such that $(r_\rho)_*\colon \pi_1(M) \to \pi_1(M_\rho)$ is equal to ρ. Set $\Theta(\rho) = [(M_\rho, h_\rho)]$. It is known that the restriction of Θ to $\mathrm{MP}(\pi_1(M))$ is surjective, and that two elements ρ_1 and ρ_2 of $\mathrm{MP}(\pi_1(M))$ lie in the same component of $\mathrm{MP}(\pi_1(M))$ if and only if $\Theta(\rho_1) = \Theta(\rho_2)$. Hence, Θ induces a one-to-one correspondence between the components of $\mathrm{MP}(\pi_1(M))$ and the elements of $\mathcal{A}(M)$; the reader is directed to Canary and McCullough [33] for complete details.

Given a pair M_1 and M_2 of compact, hyperbolizable 3–manifolds with non-empty, incompressible boundary, say that a homotopy equivalence $h\colon M_1 \to M_2$ is a *primitive shuffle* if there exists a finite collection $\mathcal{V}_1$ of primitive solid torus components of the characteristic submanifold $\Sigma(M_1)$ and a finite collection $\mathcal{V}_2$ of solid torus components of $\Sigma(M_2)$, so that $h^{-1}(\mathcal{V}_2) = \mathcal{V}_1$ and so that h restricts to an orientation preserving homeomorphism from $\overline{M_1 - \mathcal{V}_1}$ to $\overline{M_2 - \mathcal{V}_2}$; we do not define the *characteristic submanifold* here, but instead refer the reader to Canary and McCullough [33], Jaco and Shalen [50], or Johannson [51].

Let $[(M_1, h_1)]$ and $[(M_2, h_2)]$ be two elements of $\mathcal{A}(M)$. Say that $[(M_2, h_2)]$ is *primitive shuffle equivalent* to $[(M_1, h_1)]$ if there exists a primitive shuffle $\varphi\colon M_1 \to M_2$ such that $[(M_2, h_2)] = [(M_2, \varphi \circ h_1)]$. We note that when M is hyperbolizable, this gives an equivalence relation on $\mathcal{A}(M)$, where each equivalence class contains finitely many elements of $\mathcal{A}(M)$; let $\widehat{\mathcal{A}}(M)$ denote the set of equivalence classes. By considering the composition $\widehat{\Theta} = q \circ \Theta$ of Θ with the quotient map $q\colon \mathcal{A}(M) \to \widehat{\mathcal{A}}(M)$, we obtain the following complete enumeration of the components of $\overline{\mathrm{MP}(\pi_1(M))}$.

Theorem 2.6 (Anderson, Canary, and McCullough [10]) *Let M be a compact, hyperbolizable 3–manifold with non-empty, incompressible boundary, and let $[(M_1, h_1)]$ and $[(M_2, h_2)]$ be two elements of $\mathcal{A}(M)$. The associated components of $\mathrm{MP}(\pi_1(M))$ have intersecting closures if and only if $[(M_2, h_2)]$ is primitive shuffle equivalent to $[(M_1, h_1)]$. In particular, $\widehat{\Theta}$ gives a one-to-one correspondence between the components of $\overline{\mathrm{MP}(\pi_1(M))}$ and the elements of $\widehat{\mathcal{A}}$.*

Before closing this section, we highlight two consequences of the analysis involved in the proof of Theorem 2.6. The first involves the accumulation, or, more precisely, the lack thereof, of components of $\mathrm{MP}(\pi_1(M))$.

Proposition 2.7 (Anderson, Canary, and McCullough [10]) *Let M be a compact, hyperbolizable 3–manifold with non-empty, incompressible boundary. Then, the components of* $\mathrm{MP}(\pi_1(M))$ *cannot accumulate in* $\mathrm{AH}(\pi_1(M))$. *In particular, the closure* $\overline{\mathrm{MP}(\pi_1(M))}$ *of* $\mathrm{MP}(\pi_1(M))$ *is the union of the closures of the components of* $\mathrm{MP}(\pi_1(M))$.

The second involves giving a complete characterization, in terms of the topology of M, as to precisely when $\overline{\mathrm{MP}(\pi_1(M)))}$ has infinitely many components. Recall that a compact, hyperbolizable 3–manifold M with non-empty, incompressible boundary has *double trouble* if there exists a toroidal component T of ∂M and homotopically non-trivial simple closed curves C_1 in T and C_2 and C_3 in $\partial M - T$ such that C_2 and C_3 are not homotopic in ∂M, but C_1, C_2 and C_3 are homotopic in M.

Theorem 2.8 (Anderson, Canary, and McCullough [10]) *Let M be a compact, hyperbolizable 3–manifold with non-empty, incompressible boundary. Then,* $\overline{\mathrm{MP}(\pi_1(M))}$ *has infinitely many components if and only if M has double trouble. Moreover, if M has double trouble, then* $\mathrm{AH}(\pi_1(M))$ *has infinitely many components.*

2.5 The spaces $\mathcal{QC}(G)$ and $\mathrm{QC}(G) = \mathcal{QC}(G)/\mathrm{PSL}_2(\mathbf{C})$ of quasiconformal deformations

In the case that G is itself a finitely generated Kleinian group, the classical deformation theory of G consists largely of the study of the space of *quasiconformal deformations* of G, which consists of those representations of G into $\mathrm{PSL}_2(\mathbf{C})$ which are induced by a quasiconformal homeomorphism of the Riemann sphere $\overline{\mathbf{C}}$.

We do not give a precise definition here, but roughly, a *quasiconformal homeomorphism* ω of $\overline{\mathbf{C}}$ is a homeomorphism which distorts the standard complex structure on $\overline{\mathbf{C}}$ by a bounded amount; the interested reader is referred to Ahlfors [2] or to Lehto and Virtanen [63] for a thorough discussion of quasiconformality. We do note that a quasiconformal homeomorphism $\omega\colon \overline{\mathbf{C}} \to \overline{\mathbf{C}}$ is completely determined (up to post-composition by a Möbius transformation) by the measurable function $\mu = \omega_{\overline{z}}/\omega_z$, and that to every measurable function μ on $\overline{\mathbf{C}}$ with $\| \mu \|_\infty < 1$ there exists a quasiconformal homeomorphism ω of $\overline{\mathbf{C}}$ which solves the Beltrami equation $\mu\omega_z = \omega_{\overline{z}}$.

Set $\mathcal{QC}(G)$ to be the space of those representations ρ of G into $\mathrm{PSL}_2(\mathbf{C})$ which are induced by a quasiconformal homeomorphism of $\overline{\mathbf{C}}$, so that $\rho \in \mathcal{QC}(G)$ if there exists a quasiconformal homeomorphism ω of $\overline{\mathbf{C}}$ so that $\rho(g) = \omega \circ g \circ \omega^{-1}$ for all $g \in G$. By definition, we have that $\mathcal{QC}(G)$ is contained in $\mathcal{D}(G)$. Set $\mathrm{QC}(G) = \mathcal{QC}(G)/\mathrm{PSL}_2(\mathbf{C})$.

It is known that $\mathrm{QC}(G)$ is a complex manifold, and is actually the quotient of the Teichmüller space of the (possibly disconnected) quotient Riemann surface $\Omega(G)/G$ by a properly discontinuous group of biholomorphic automorphims. This result, in its full generality, follows from the work of a number of authors, including Maskit [70], Kra [62], Bers [16], and Sullivan [95].

We note here, in the case that G is a geometrically finite, minimally parabolic Kleinian group, that it follows from the Isomorphism Theorem of Marden [64] that $\mathrm{QC}(G)$ is the component of $\mathrm{MP}(G)$ containing the identity representation.

Sullivan [95] has shown, for a finitely generated Kleinian group G, if there exists a quasiconformal homeomorphism ω of $\overline{\mathbf{C}}$ which conjugates G to a Kleinian group and which is conformal on $\Omega(G)$, then ω is necessarily a Möbius transformation. In other words, if ω conjugates G to subgroup of $\mathrm{PSL}_2(\mathbf{C})$, then $\mu = \omega_{\overline{z}}/\omega_z$ is equal to 0 on $\Lambda(G)$.

In particular, if $\Omega(G)$ is empty, then $\mathrm{QC}(G)$ consists of a single point, namely the identity representation. This can be viewed as a generalization of Theorem 2.2, as Sullivan's result also holds for an infinite volume hyperbolic 3–manifold N whose uniformizing Kleinian group G has limit set the whole Riemann sphere.

We note here that the study of quasiconformal deformations of finitely Kleinian groups is the origin of the Ahlfors Measure Conjecture. In [3], Ahlfors raises the question of whether the limit set of a finitely generated Kleinian group with non-empty domain of discontinuity necessarily has zero area. If this conjecture is true, then it would be impossible for a quasiconformal deformation of a finitely generated Kleinian group G to be supported on the limit set of G. The result of Sullivan mentioned above implies that no such deformation exists, though without solving the Measure Conjecture, which has not yet been completely resolved. It is known that the Measure Conjecture holds in a large number of cases, in particular it holds for all topologically tame groups. For a discussion of this connection, we refer the reader to Canary [30] and the references contained therein.

There are several classes of Kleinian groups for which $\mathrm{QC}(G)$ has been extensively studied, which we discuss here.

A *Schottky group* is a finitely generated, purely loxodromic Kleinian group G which is free on g generators and whose domain of discontinuity is non-empty;

the number of generators is sometimes referred to as the *genus* of the Schottky group. This is not the original definition, but is equivalent to the usual definition by a theorem of Maskit [67]. In particular, a Schottky group is necessarily convex cocompact. Chuckrow [37] shows that any two Schottky groups of the same rank are quasiconformally conjugate, so that $\mathrm{QC}(G)$ is in fact equal to the space $\mathrm{MP}(G)$ of all convex cocompact representations of a group G which is free on g generators into $\mathrm{PSL}_2(\mathbf{C})$.

In the same paper [37], Chuckrow also engages in an analysis of the closure of $\mathrm{QC}(G)$ in $\mathrm{R}(G)$ for a Schottky group G of genus g. In particular, she shows that every point ρ in $\partial\mathrm{QC}(G)$ has the property that $\rho(G)$ is free on g generators, and contains no elliptic elements of infinite order. However, this in itself is not enough to show that $\rho(G)$ is discrete, as Greenberg [46] has constructed free, purely loxodromic subgroups of $\mathrm{PSL}_2(\mathbf{C})$ which are not discrete.

More generally, Chuckrow also shows that the limit ρ of a convergent sequence $\{\rho_n\}$ of type-preserving faithful representations in $\mathcal{HOM}(G)$ is again a faithful representation of G, and that $\rho(G)$ contains no elliptic elements of infinite order.

Jørgensen [53] credits his desire to generalize the results of Chuckrow [37] to leading him to what is now commonly referred to as Jørgensen's inequality, which states that if γ and φ are elements of $\mathrm{PSL}_2(\mathbf{C})$ which generate a non-elementary Kleinian group, then $|\mathrm{tr}^2(\gamma)-4|+|\mathrm{tr}([\gamma,\varphi])-2|\geq 1$, where $\mathrm{tr}(\gamma)$ is the trace of a matrix representative of γ in $\mathrm{SL}_2(\mathbf{C})$. The proof of Theorem 2.1 is a direct application of this inequality.

For a Schottky group G, it is known that $\mathrm{AH}(G)$ is not compact. There is work of Canary [27] and Otal [88] on a conjecture of Thurston which gives conditions under which sequences in $\mathrm{QC}(G)$ have convergent subsequences; we do not give details here, instead referring the interested reader to the papers cited above.

We also mention here the work of Keen and Series [59] on the Riley slice of the space of 2–generator Schottky groups, in which they introduce coordinates on the Riley slice and study the cusp points on the boundary of the Riley slice.

A *quasifuchsian* group is a finitely generated Kleinian group whose limit set is a Jordan curve and which contains no element interchanging the two components of its domain of discontinuity. Consequently, every quasifuchsian group is isomorphic to the fundamental group of a surface. It is known that any two isomorphic purely loxodromic quasifuchsian groups are quasiconformally conjugate, by work of Maskit [66], and hence for a purely loxodromic quasifuchsian group G we have that $\mathrm{MP}(G)=\mathrm{QC}(G)$.

This equality does not hold for quasifuchsian groups uniformizing punctured surfaces, for several reasons. First, the quasifuchsian groups uniformizing the three-times punctured sphere and the once-punctured torus are isomorphic, namely the free group of rank two, but cannot be quasiconformally conjugate, as the surfaces are not homeomorphic. Second, as every quasifuchsian group isomorphic to the free group G of rank two contains parabolic elements, no quasifuchsian group isomorphic to G lies in $\mathrm{MP}(G)$.

It is known that $\mathrm{QC}(G)$ is biholomorphically equivalent to the product of Teichmüller spaces $\mathcal{T}(S) \times \mathcal{T}(\overline{S})$, where S is one of the components of $\Omega(G)/G$ and $\overline{S}$ is its complex conjugate.

A *Bers slice* of $\mathrm{QC}(G)$ for a quasifuchsian group G is a subspace of $\mathrm{QC}(G)$ of the form $B(s_0) = \mathcal{T}(S) \times \{s_0\}$. The structure of the closure of $B(s_0)$ in $\mathrm{AH}(G)$ has been studied by a number of authors, including Bers [14], Kerckhoff and Thurston [60], Maskit [66], McMullen [74], and Minsky [76]. In particular, Bers [14] showed that the closure $\overline{B(s_0)}$ of $B(s_0)$ is compact, and Kerckhoff and Thurston [60] have shown that the compactification $\overline{B(s_0)}$ depends on the basepoint s_0, and so there are actually uncountably many such compactifications. Among other major results, Minsky [76] has shown that every punctured torus group lies in the boundary of $\mathrm{QC}(G)$, where G is a quasifuchsian group uniformizing a punctured torus and where a *punctured torus group* is a Kleinian group generated by two elements with parabolic commutator. In particular, this shows that the relative version of the Density Conjecture holds for punctured torus groups.

There are other slices of $\mathrm{QC}(G)$ which have been extensively studied. There is the extensive work of Keen and Series, see for instance [56], [57], and [58], inspired in part by unpublished work of Wright [103], on the Maskit slice of the Teichmüller space of a punctured torus in terms of *pleating coordinates*, which are natural and geometrically interesting coordinates on the Teichmüller space of the punctured torus which are given in terms of the geometry of the corresponding hyperbolic 3–manifolds.

In the case that G is a Kleinian group for which the corresponding 3–manifold $M = (\mathbf{H}^3 \cup \Omega(G))/G$ is a compact, acylindrical 3–manifold with non-empty, incompressible boundary, then every representation in $\mathrm{MP}(G)$ in fact lies in $\mathrm{QC}(G)$; this follows from work of Johannson [51]. In addition, Thurston [99] has shown that $\mathrm{AH}(G)$ is compact for such G; another proof is given by Morgan and Shalen [80].

2.6 The spaces of $\mathcal{TT}(G)$ and $\mathrm{TT}(G) = \mathcal{TT}(G)/\mathrm{PSL}_2(\mathbf{C})$ of topologically tame representations

There is one last class of deformations which we need to define, before beginning our discussion of the relationships between these spaces. We begin with a topological definition. A compact submanifold M of a hyperbolic 3–manifold N is a *compact core* if the inclusion of M into N is a homotopy equivalence. The Core Theorem of Scott [93] implies that every hyperbolic 3–manifold with finitely generated fundamental group has a compact core. Marden [64] asked whether every hyperbolic 3–manifold N with finitely generated fundamental group is necessarily *topologically tame*, in that N is homeomorphic to the interior of its compact core.

Set $\mathcal{TT}(G)$ to be the subspace of $\mathcal{D}(G)$ consisting of the representations ρ with minimally parabolic, topologically tame image $\rho(G)$.
Set $\mathrm{TT}(G) = \mathcal{TT}(G)/\mathrm{PSL}_2(\mathbf{C})$.

There is a notion related to topological tameness, namely *geometric tameness*, first defined by Thurston [102]. We do not discuss geometric tameness here; the interested reader should consult Thurston [102], Bonahon [18], or Canary [29]. Thurston [102] showed that geometrically tame hyperbolic 3–manifolds with freely indecomposible fundamental group are topologically tame and satisfy the Ahlfors Measure Conjecture. Bonahon [18] showed that if every non-trivial free product splitting of a finitely generated Kleinian group Γ has the property that there exists a parabolic element of Γ not conjugate into one of the free factors, then Γ is geometrically tame. Canary [29] extended the definition of geometrically tame to all hyperbolic 3–manifolds, proved that topologically tame hyperbolic 3–manifolds are geometrically tame, and proved that topological tameness has a number of geometric and analytic consequences; in particular, he established that the Ahlfors Measure Conjecture holds for topologically tame Kleinian groups.

3 Geometric limits

There is a second notion of convergence for Kleinian groups which is distinct from the topology described above, which is equally important in the study of deformations spaces.

A sequence $\{\Gamma_n\}$ of Kleinian groups converges *geometrically* to a Kleinian group $\widehat{\Gamma}$ if two conditions are met, namely that every element of $\widehat{\Gamma}$ is the limit of a sequence of elements $\{\gamma_n \in \Gamma_n\}$ and that every accumulation point of every sequence $\{\gamma_n \in \Gamma_n\}$ lies in $\widehat{\Gamma}$. Note that, unlike the topology of algebraic convergence described above, the geometric limit of a sequence of isomorphic Kleinian groups need not be isomorphic to the groups in the sequence, and indeed need not be finitely generated. However, it is known that the geometric limit of a sequence of non-elementary, torsion-free Kleinian groups is again torsion-free.

We note here that it is possible to phrase the definition of geometric convergence in terms of the quotient hyperbolic 3–manifolds. Setting notation, let 0 denote a choice of basepoint for $\mathbf{H}^3$, and let $p_j\colon \mathbf{H}^3 \to N_j = \mathbf{H}^3/\rho_j(G)$ and $p\colon \mathbf{H}^3 \to \widehat{N} = \mathbf{H}^3/\widehat{\Gamma}$ be the covering maps. Let $B_R(0) \subset \mathbf{H}^3$ be a ball of radius R centered at the basepoint 0.

Lemma 3.1 *A sequence of torsion-free Kleinian groups $\{\Gamma_n\}$ converges geometrically to a torsion-free Kleinian group $\widehat{\Gamma}$ if and only if there exists a sequence $\{(R_n, K_n)\}$ and a sequence of orientation preserving maps $\tilde{f}_n\colon B_{R_n}(0) \to \mathbf{H}^3$ such that the following hold:*

1) *$R_n \to \infty$ and $K_n \to 1$ as $i \to \infty$;*
2) *the map $\tilde{f}_n$ is a K_n–bilipschitz diffeomorphism onto its image, $\tilde{f}_n(0) = 0$, and $\{\tilde{f}_n|_A\}$ converges to the identity for any compact set A; and*
3) *$\tilde{f}_n$ descends to a map $f_n\colon Z_n \to \widehat{N}$, where $Z_n = B_{R_n}(0)/\Gamma_n$ is a submanifold of N_n; moreover, f_n is also an orientation preserving K_n–biLipschitz diffeomorphism onto its image.*

For a proof of this Lemma, see Theorem 3.2.9 of Canary, Epstein, and Green [32], and Theorem E.1.13 and Remark E.1.19 of Benedetti and Petronio [13].

A fundamental example of the difference between algebraic and geometric convergence of Kleinian groups is given by the following explicit example of Jørgensen and Marden [55]; earlier examples are given in Jørgensen [52]. Choose ω_1 and ω_2 in $\mathbf{C} - \{0\}$ which are linearly independent over $\mathbf{R}$, and for each $n \geq 1$ set $\omega_{1n} = \omega_1 + n\omega_2$, $\omega_{2n} = \omega_2$, and $\tau_n = \omega_{2n}/\omega_{1n}$. Consider the loxodromic elements $L_n(z) = \exp(-2\pi\mathbf{i}\tau_n)z + \omega_2$. Then, as $n \to \infty$, L_n converges to $L(z) = z + \omega_2$, and so $\langle L_n \rangle$ converges algebraically to $\langle L \rangle$. However, note that $L_n^{-n}(z)$ converges to $K(z) = z + \omega_1$ as $n \to \infty$. Hence, $\langle L_n \rangle$ converges geometrically to $\langle L, K \rangle = \mathbf{Z} \oplus \mathbf{Z}$.

This example of the geometric convergence of loxodromic cyclic groups to rank two parabolic groups underlies much of the algebra of the operation of *Dehn surgery*, which we describe here.

Let M be a compact, hyperbolizable 3–manifold, let T be a torus component of ∂M, and choose a meridian–longitude system (α, β) on T. Let P be a solid torus and let c be a simple closed curve on ∂P bounding a disc in P. For each pair (m, n) of relatively prime integers, let $M(m, n)$ be the 3–manifold by attaching ∂P to T by an orientation-reversing homeomorphism which identifies c with $m\alpha + n\beta$; we refer to $M(m, n)$ as the result of (m, n) Dehn surgery along T. The following Theorem describes the basic properties of this operation; the version we state is due to Comar [38].

Theorem 3.2 (Comar [38]) *Let M be a compact, hyperbolizable 3–manifold and let $T = \{T_1, \ldots, T_k\}$ be a non-empty collection of tori in ∂M. Let $\hat{N} = \mathbf{H}^3/\Gamma$ be a geometrically finite hyperbolic 3–manifold and let $\psi\colon \mathrm{int}(M) \to N$ be an orientation preserving homeomorphism. Further assume that every parabolic element of Γ lies in a rank two parabolic subgroup. Let (m_i, l_i) be a meridian–longitude basis for T_i. Let $\{(\mathbf{p}_n, \mathbf{q}_n) = ((p_n^1, q_n^1), \ldots, (p_n^k, q_n^k))\}$ be a sequence of k–tuples of pairs of relatively prime integers such that, for each i, $\{(p_n^i, q_n^i)\}$ converges to ∞ as $n \to \infty$.*

Then, for all sufficiently large n, there exists a representation $\beta_n\colon \Gamma \to \mathrm{PSL}_2(\mathbf{C})$ with discrete image such that

1) *$\beta_n(\Gamma)$ is geometrically finite, uniformizes $M(\mathbf{p}_n, \mathbf{q}_n)$, and every parabolic element of $\beta_n(\Gamma)$ lies in a rank two parabolic subgroup;*
2) *the kernel of $\beta_n \circ \psi_*$ is normally generated by $\{m_1^{p_n^1} l_1^{q_n^1}, \ldots, m_k^{p_n^k} l_k^{q_n^k}\}$; and*
3) *$\{\beta_n\}$ converges to the identity representation of Γ.*

The idea of Theorem 3.2 is due to Thurston [102] in the case that the hyperbolic 3–manifold N has finite volume, so that ∂M consists purely of tori. In this case, it is also known that $\mathrm{volume}(\mathbf{H}^3/\beta_n(\Gamma)) < \mathrm{volume}(\mathbf{H}^3/\Gamma)$ for each n, and that $\mathrm{volume}(\mathbf{H}^3/\beta_n(\Gamma)) \to \mathrm{volume}(\mathbf{H}^3/\Gamma)$ as $n \to \infty$. For a more detailed discussion of this phenomenon, we refer the reader to Gromov [47] and Benedetti and Petronio [13]. The generalization to the case that N has infinite volume is due independently to Bonahon and Otal [21] and Comar [38]. Note that the $\beta_n(\Gamma)$ are not isomorphic, and hence there is no notion of algebraic convergence for these groups.

In the case that we have a sequence of representations in $\mathcal{D}(G)$, the following result of Jørgensen and Marden is extremely useful.

Proposition 3.3 (Jørgensen and Marden [55]) *Let $\{\rho_n\}$ be a sequence in $\mathrm{AH}(G)$ converging to ρ; then, there is a subsequence of $\{\rho_n\}$, again called*

$\{\rho_n\}$, so that $\{\rho_n(G)\}$ converges geometrically to a Kleinian group $\widehat{\Gamma}$ containing $\rho(G)$.

A sequence $\{\rho_n\}$ in $\mathcal{D}(G)$ converges *strongly* to ρ if $\{\rho_n\}$ converges algebraically to ρ and if $\{\rho_n(G)\}$ converges geometrically to $\rho(G)$. Note that we may consider $\mathcal{D}(G)$, and $\mathrm{AH}(G)$, to be endowed with topology of strong convergence, instead of the topology of algebraic convergence. We also refer the reader to the recent article of McMullen [75], in which a variant of the notion of strong convergence is explored in a somewhat more general setting.

Generalizing the behavior of the sequence of loxodromic cyclic groups described above, examples of sequences $\{\rho_n\}$ in $\mathcal{D}(G)$ which converge algebraically to ρ and for which $\{\rho_n(G)\}$ converges geometrically to a Kleinian group Γ properly containing $\rho(G)$ have been constructed by a number of authors, including Thurston [102], [100], Kerckhoff and Thurston [60], Anderson and Canary [6], Ohshika [84], and Brock [26], [25], among others.

Jørgensen and Marden [55] carry out a very detailed study of the relationship between the algebraic limit and the geometric limit in the case when the geometric limit is assumed to be geometrically finite. In general, not much is known about the relationship between the algebraic and geometric limits of a sequence of isomorphic Kleinian groups. We spend the remainder of this section discussing this question.

A fundamental point in understanding how algebraic limits sit inside geometric limits is the following algebraic fact, which is an easy application of Jørgensen's inequality.

Proposition 3.4 (Anderson, Canary, Culler, and Shalen [9]) *Let $\{\rho_n\}$ be a sequence in $\mathcal{D}(G)$ which converges to ρ and for which $\{\rho_n(G)\}$ converges geometrically to a Kleinian group $\widehat{\Gamma}$ containing $\rho(G)$. Then, for each $\gamma \in \widehat{\Gamma} - \rho(G)$, the intersection $\gamma\rho(G)\gamma^{-1}$ is either trivial or parabolic cyclic.*

One of the first applications of this result, also in [9], was to show, when the algebraic limit is a maximal cusp, that the convex hull of the quotient 3–manifold corresponding to the algebraic limit embeds in the quotient 3–manifold corresponding to the geometric limit. This was part of a more general attempt to understand the relationship between the volume and the rank of homology for a finite volume hyperbolic 3–manifold.

Another application was given by Anderson and Canary [7]. Before stating the generalization, we need to give a definition. Given a Kleinian group Γ, consider its associated 3–manifold $M = (\mathbf{H}^3 \cup \Omega(\Gamma))/\Gamma$, where $\Omega(\Gamma)$ is the domain of discontinuity of Γ. Then, Γ has connected limit set and no accidental parabolics if and only if every closed curve γ in ∂M which is homotopic to a curve of arbitrarily small length in the interior of M with the hyperbolic metric, is homotopic to a curve of arbitrarily small length in ∂M, with its induced metric.

Theorem 3.5 (Anderson and Canary [7]) *Let G be a finitely generated, torsion-free, non-abelian group, let $\{\rho_n\}$ be a sequence in $\mathcal{D}(G)$ converging to ρ, and suppose that $\{\rho_n(G)\}$ converges geometrically to $\widehat{\Gamma}$. Let $N = \mathbf{H}^3/\rho(G)$ and $\widehat{N} = \mathbf{H}^3/\widehat{\Gamma}$, and let $\pi\colon N \to \widehat{N}$ be the covering map. If $\rho(G)$ has non-empty domain of discontinuity, connected limit set, and contains no accidental parabolics, then there exists a compact core M for N such that π is an embedding restricted to M.*

One can apply Theorem 3.5 to show that certain algebraically convergent sequences are actually strongly convergent. This is of interest, as it is generally much more difficult to determine strong convergence of a sequence of representations than to determine algebraic convergence.

Theorem 3.6 (Anderson and Canary [7]) *Let G be a finitely generated, torsion-free, non-abelian group and let $\{\rho_n\}$ be a sequence in $\mathcal{D}(G)$ converging to ρ. Suppose that $\rho_n(G)$ is purely loxodromic for all n, and that $\rho(G)$ is purely loxodromic. If $\Omega(\rho(G))$ is non-empty, then $\{\rho_n(G)\}$ converges strongly to $\rho(G)$. Moreover, $\{\Lambda(\rho_n(G))\}$ converges to $\Lambda(\rho(G))$.*

Theorem 3.7 (Anderson and Canary [7]) *Let G be a finitely generated, torsion-free, non-abelian group and let $\{\rho_n\}$ be a sequence in $\mathcal{D}(G)$ converging to ρ. Suppose that $\rho_n(G)$ is purely loxodromic for all n, that $\rho(G)$ is purely loxodromic, and that G is not a non-trivial free product of (orientable) surface groups and cyclic groups, then $\{\rho_n(G)\}$ converges strongly to $\rho(G)$. Moreover, $\{\Lambda(\rho_n(G))\}$ converges to $\Lambda(\rho(G))$.*

Both Theorem 3.6 and Theorem 3.7 have been generalized by Anderson and Canary [8] to Kleinian groups containing parabolic elements, under the hypothesis that the sequences are type-preserving.

One reason that strong convergence is interesting is that strongly convergent sequences of isomorphic Kleinian groups tend to be extremely well behaved, as one has the geometric data coming from the convergence of the quotient

3–manifolds as well as the algebraic data coming from the convergence of the representations. For instance, there is the following Theorem of Canary and Minsky [34]. We note that a similar result is proven independently by Ohshika [86].

Theorem 3.8 (Canary and Minsky [34]) *Let M be a compact, irreducible 3–manifold and let $\{\rho_n\}$ be a sequence in $\mathrm{TT}(\pi_1(M))$ converging strongly to ρ, where each $\rho_n(\pi_1(M))$ and $\rho(\pi_1(M))$ are purely loxodromic. Then, $\rho(\pi_1(M))$ is topologically tame; moreover, for all sufficiently large n, there exists a homeomorphism $\varphi_n\colon \mathbf{H}^3/\rho_n(\pi_1(M)) \to \mathbf{H}^3/\rho(\pi_1(M))$ so that $(\varphi_n)_* = \rho\circ\rho_n^{-1}$.*

By combining the results of Anderson and Canary [7] and of Canary and Minsky [34] stated above, one may conclude that certain algebraic limits of sequences of isomorphic topologically tame Kleinian groups are again topologically tame.

There is also the following result of Taylor [98].

Theorem 3.9 (Taylor [98]) *Let G be a finitely generated, torsion-free, nonabelian group, and let $\{\rho_n\}$ be a sequence in $\mathcal{D}(G)$ converging strongly to ρ, where each $\rho_n(G)$ has infinite co-volume. If $\rho(G)$ is geometrically finite, then $\rho_n(G)$ is geometrically finite for n sufficiently large.*

The guiding Conjecture in the study of the relationship between algebraic and geometric limits, usually attributed to Jørgensen, is stated below.

Conjecture 3.10 (Jørgensen) *Let Γ be a finitely generated, torsion-free, non-elementary Kleinian group, let $\{\rho_n\}$ be a sequence in $\mathcal{D}(\Gamma)$ converging to ρ, and suppose that $\{\rho_n(\Gamma)\}$ converges geometrically to $\widehat{\Gamma}$. If ρ is type-preserving, then $\rho(\Gamma)=\widehat{\Gamma}$.*

As we have seen above, this conjecture has been shown to hold in a wide variety of cases, including the case in which the sequence $\{\rho_n\}$ is type-preserving and the limit group $\rho(\Gamma)$ either has non-empty domain of discontinuity or is not a non-trivial free product of cyclic groups and the fundamental groups of orientable surfaces.

4 Functions on deformation spaces

There are several numerical quantities associated to a Kleinian group Γ; one is the *Hausdorff dimension* $\mathrm{D}(\Gamma)$ of the limit set $\Lambda(\Gamma)$ of Γ, another is the smallest positive eigenvalue L(Γ) of the Laplacian on the corresponding hyperbolic 3–manifold $\mathbf{H}^3/\Gamma$. These two functions are closely related; namely, if Γ is topologically tame, then $\mathrm{L}(\Gamma) = \mathrm{D}(\Gamma)(2 - \mathrm{D}(\Gamma))$ when $\mathrm{D}(\Gamma) \geq 1$, and $\mathrm{L}(\Gamma) = 1$ when $\mathrm{D}(\Gamma) \leq 1$. The relationship between these two quantities has been studied by a number of authors, including Sullivan [96], Bishop and Jones [17], Canary [31], and Canary, Minsky, and Taylor [35] (from which the statement given above is taken). It is natural to consider how these functions behave on the spaces we have been discussing in this note.

We begin by giving a few topological definitions. A compact, hyperbolizable 3–manifold with incompressible boundary is a *generalized book of* I*–bundles* if there exists a disjoint collection A of essential annuli in M so that each component of the closure of the complement of A in M is either a solid torus, a thickened torus, or an I–bundle whose intersection with ∂M is the associated ∂I–bundle.

An *incompressible core* of a compact hyperbolizable 3–manifold is a compact submanifold P, possibly disconnected, with incompressible boundary so that M can be obtained from P by adding 1–handles.

We begin with a pair of results of Canary, Minsky, and Taylor [35] which relates the topology of M to the behavior of these functions on a well-defined subset of $\mathrm{AH}(\pi_1(M))$, and show that they are in a sense dual to one another.

Theorem 4.1 (Canary, Minsky, and Taylor [35]) *Let M be a compact, hyperbolizable 3–manifold. Then, $\sup L(\rho(\pi_1(M))) = 1$ if and only if every component of the incompressible core of M is a generalized book of I–bundles; otherwise, $\sup L(\rho(\pi_1(M))) < 1$. Here, the supremum is taken over all ρ in $\mathrm{AH}(\pi_1(M))$ for which $\mathbf{H}^3/\rho(\pi_1(M))$ is homeomorphic to the interior of M.*

Theorem 4.2 (Canary, Minsky, and Taylor [35]) *Let M be a compact, hyperbolizable 3–manifold which is not a handlebody or a thickened torus. Then, $\inf D(\rho(\pi_1(M))) = 1$ if and only if every component of the incompressible core of M is a geneneralized book of I–bundles; otherwise, $\inf D(\rho(\pi_1(M))) > 1$. Here, the infimum is taken over all ρ in $\mathrm{AH}(\pi_1(M))$ for which $\mathbf{H}^3/\rho(\pi_1(M))$ is homeomorphic to the interior of M.*

It is also possible to consider how these quantities behave under taking limits. We note that results similar to Theorem 4.3 have been obtained by McMullen [75], who also shows that the function D is not continuous on $\mathcal{D}(\pi_1(M))$ in the case that M is a handlebody.

Theorem 4.3 (Canary and Taylor [36]) *Let M be a compact, hyperbolizable 3–manifold which is not homeomorphic to a handlebody. Then $D(\rho)$ is continuous on $\mathcal{D}(\pi_1(M))$ endowed with the topology of strong convergence.*

Recently, Fan and Jorgenson [44] have made use of the heat kernel to prove the continuity of small eigenvalues and small eigenfunctions of the Laplacian for sequences of hyperbolic 3–manifolds converging to a geometrically finite limit manifold, where the convergence is the variant of strong convergence considered by McMullen [75].

There are several functions on $\mathrm{QC}(G)$ which have been studied by Bonahon. In order to keep definitions to a minimum, we state his results for geometrically finite G, though we note that they hold for a general finitely generated Kleinian group G. Given a representation ρ in $\mathrm{QC}(G)$, recall that the *convex core* C_ρ of $\mathbf{H}^3/\rho(G)$ is the smallest convex submanifold of $\mathbf{H}^3/\rho(G)$ whose inclusion is a homotopy equivalence. By restricting the hyperbolic metric on $\mathbf{H}^3/\rho(G)$ to ∂C_ρ, we obtain a map μ from $\mathrm{QC}(G)$ to the Teichmüller space $\mathcal{T}(\Omega(G)/G)$ of the Riemann surface $\Omega(G)/G$.

Theorem 4.4 (Bonahon [20]) *For a geometrically finite Kleinian group G, the map $\mu\colon \mathrm{QC}(G) \to \mathcal{T}(\Omega(G)/G)$ is continuously differentiable.*

Another function on $\mathrm{QC}(G)$ studied by Bonahon, by developing an analog of the Schläfli formula for the volume of a polyhedron in hyperbolic space, is the function $\mathrm{vol}\colon \mathrm{QC}(G) \to [0,\infty)$, which associates to $\rho \in \mathrm{QC}(G)$ the volume $\mathrm{vol}(\rho)$ of the convex core C_ρ of $\mathbf{H}^3/\rho(G)$.

Theorem 4.5 (Bonahon [19]) *Let G be a geometrically finite Kleinian group. If the boundary ∂C_ρ of the convex core C_ρ of $\mathbf{H}^3/\rho(G)$ is totally geodesic, then ρ is a local minimum of* $\mathrm{vol}\colon \mathrm{QC}(G) \to [0,\infty)$.

It is known that the Hausdorff dimension of the limit set is a continuous function on $\mathrm{QC}(\Gamma)$, using estimates relating the Hausdorff dimension and quasiconformal dilitations due to Gehring and Väisälä [45]. In some cases, it is possible to obtain more analytic information.

Theorem 4.6 (Ruelle [92]) *Let Γ be a convex cocompact Kleinian group whose limit set supports an expanding Markov partition. Then, the Hausdorff dimension of the limit set is a real analytic function on* $\mathrm{QC}(\Gamma)$.

Earlier work of Bowen [24] shows that quasifuchsian and Schottky groups support such Markov partitions. The following Theorem follows by combining these results of Bowen and Ruelle with a condition which implies the existence of an expanding Markov partition, namely that there exists a fundamental polyhedron in $\mathbf{H}^3$ for the Kleinian group G which has the *even cornered property*, together with the Klein Combination Theorem.

Theorem 4.7 (Anderson and Rocha [11]) *Let G be a convex cocompact Kleinian group which is isomorphic to the free product of cyclic groups and fundamental groups of 2–orbifolds. Then, the Hausdorff dimension of the limit set is a real analytic function on* $\mathrm{QC}(G)$.

We note here that it is not yet established that all convex cocompact Kleinian groups support such Markov partitions.

Another function one can consider is the injectivity radius of the corresponding quotient hyperbolic 3–manifold. For a hyperbolic 3–manifold N, the *injectivity radius* $\mathrm{inj}_N(x)$ at a point $x \in N$ is one-half the length of the shortest homotopically non-trivial closed curve through x. The following Conjecture is due to McMullen.

Conjecture 4.8 *Let G be a finitely generated group with g generators. Then, there exists a constant $C = C(g)$ so that, if N is a hyperbolic 3–manifold with fundamental group isomorphic to G and if x lies in the convex core of N, then* $\mathrm{inj}_N(x) \le C$.

Kerckhoff and Thurston [60] show that, if M is the product of a closed, orientable surface S of genus at least 2 with the interval, then there exists a constant $C = C(M)$ so that if N is a hyperbolic 3–manifold which is homeomorphic to the interior of M and if N has no cusps, then the injectivity radius on the convex core of N is bounded above by C. Fan [42] generalizes this to show that, if M is a compact, hyperbolizable 3–manifold which is either a book of I–bundles or is acylindrical, then there exists a constant $C = C(M)$ so that, if N is any hyperbolic 3–manifold homeomorphic to the interior of M, then the injectivity radius on the convex core of N is bounded above by C.

We close by mentioning recent work of Basmajian and Wolpert [12] concerning the persistance of intersecting closed geodesics. Say that a Kleinian group Γ has the *SPD property* if all the closed geodesics in $\mathbf{H}^3/\Gamma$ are simple and pairwise disjoint.

Theorem 4.9 (Basmajian and Wolpert [12]) *Let G be a torsion-free, convex co-compact Kleinian group, and let U be the component of* $\mathrm{MP}(G)$ *containing the identity representation. Then, either*

1) *there exists a subset V of U, which is the intersection of a countably many open dense sets, so that $\rho(G)$ has the SPD property for every $\rho \in V$, or*

2) *there exists a pair of loxodromic elements α and β of G so that the closed geodesics in $\mathbf{H}^3/\rho(G)$ corresponding to loxodromic elements $\rho(\alpha)$ and $\rho(\beta)$ intersect at an angle constant over all $\rho \in U$; in particular, there is* no *element $\rho \in U$ so that $\rho(G)$ has the SPD property.*

They also show that the first possibility holds in the case that G is a purely loxodromic Fuchsian group.

References

[1] **W Abikoff**, **B Maskit**, *Geometric decompositions of Kleinian groups*, Amer. J. Math. 99 (1977) 687–697

[2] **L V Ahlfors**, *Lectures on Quasiconformal Mappings*, Wadsworth and Brooks–Cole (1987)

[3] **L V Ahlfors**, *Finitely generated Kleinian groups*, Amer. J. Math. 86 (1964) 413–29

[4] **L V Ahlfors**, *The structure of a finitely generated Kleinian group*, Acta Math. 122 (1969) 1–17

[5] **L V Ahlfors**, **L Bers**, *Riemann's mapping theorem for variable metrics*, Annals of Math. 72 (1960) 385–404

[6] **J W Anderson**, **R D Canary**, *Algebraic limits of Kleinian groups which rearrange the pages of a book*, Invent. Math. 126 (1996) 205–214

[7] **J W Anderson**, **R D Canary**, *Cores of hyperbolic 3–manifolds and limits of Kleinian groups*, Amer. J. Math. 118 (1996) 745–779

[8] **J W Anderson**, **R D Canary**, *Cores of hyperbolic 3–manifolds and limits of Kleinian groups II*, Journal London Math. Soc. (to appear)

[9] **J W Anderson**, **R D Canary**, **M Culler**, **P Shalen**, *Free Kleinian groups and volumes of hyperbolic 3–manifolds*, J. Differential Geometry 43 (1996) 738–782

[10] **J W Anderson**, **R D Canary**, **D McCullough**, *On the topology of deformation spaces of Kleinian groups*, preprint

[11] **J W Anderson**, **A C Rocha**, *Analyticity of Hausdorff dimension of limit sets of Kleinian groups*, Ann. Acad. Sci. Fenn. 22 (1997) 349–364

[12] **A Basmajian**, **S A Wolpert**, *Hyperbolic 3–manifolds with non-intersecting closed geodesics*, in preparation

[13] **R Benedetti**, **C Petronio**, *Lectures on Hyperbolic Geometry*, Springer–Verlag Universitext, (1992)

[14] **L Bers**, *On boundaries of Teichmüller spaces and on Kleinian groups*, Annals of Math. 91 (1970) 570–600

[15] **L Bers**, *Simultaneous uniformization*, Bull. Amer. Math. Soc. 66 (1960) 94–97

[16] **L Bers**, *Spaces of Kleinian groups*, from: "Maryland Conference in Several Complex Variables I", Springer–Verlag Lecture Notes in Math. vol. 155 (1970) 9–34

[17] **C J Bishop**, **P W Jones**, *Hausdorff dimension and Kleinian groups*, Acta Math. 197 (1997) 1–39

[18] **F Bonahon**, *Bouts des variétés hyperboliques de dimension 3*, Annals of Math. 124 (1986) 71–158

[19] **F Bonahon**, *A Schläfli-type formula for convex cores of hyperbolic* 3*–manifolds*, preprint

[20] **F Bonahon**, *Variations of the boundary geometry of* 3*–dimensional hyperbolic convex cores*, preprint

[21] **F Bonahon**, **J P Otal**, *Variétés hyperboliques à géodésiques arbitrairement courtes*, Bull. London Math. Soc. 20 (1988) 255–261

[22] **B H Bowditch**, *Geometrical finiteness for hyperbolic groups*, J. Funct. Anal. 113 (1993) 245–317

[23] **B H Bowditch**, *Spaces of geometrically finite representations* Ann. Acad. Sci. Fenn. (to appear)

[24] **R Bowen**, *Hausdorff dimension of quasi-circles*, IHES Publ. Math. 50 (1979) 1–25

[25] **J F Brock**, *Iteration of mapping classes on a Bers' slice: examples of algebraic and geometric limits of hyperbolic* 3*–manifolds*, Proceedings of the Bers colloquium (1997) (to appear)

[26] **J F Brock**, *Iteration of mapping classes and limits of hyperbolic* 3*–manifolds*, PhD Thesis UC Berkeley (1997)

[27] **R D Canary**, *Algebraic convergence of Schottky groups*, Trans. Amer. Math. Soc. 337 (1993) 235–258

[28] **R D Canary**, *Covering theorems for hyperbolic* 3*–manifolds*, Proceedings of Low-Dimensional Topology, International Press (1994) 21–30

[29] **R D Canary**, *Ends of hyperbolic* 3*–manifolds*, Journal Amer. Math. Soc. 6 (1993) 1–35

[30] **R D Canary**, *Geometrically tame hyperbolic* 3*–manifolds*, Proceedings of Symposia in Pure Mathematics 54 (1993) 99–109

[31] **R D Canary**, *On the Laplacian and the geometry of hyperbolic* 3*–manifolds*, J. Differential Geometry 36 (1992) 349–367

[32] **R D Canary**, **D B A Epstein**, **P Green**, *Notes on notes of Thurston*, from: "Analytical and Geometrical Aspects of Hyperbolic Spaces", Cambridge University Press, (1987) 3–92

[33] **R D Canary**, **D McCullough**, *Homotopy equivalences of 3–manifolds and deformation theory of Kleinian groups*, in preparation

[34] **R D Canary**, **Y N Minsky**, *On limits of tame hyperbolic 3–manifolds*, J. Differential Geometry 43 (1996) 1–41

[35] **R D Canary**, **Y N Minsky**, **E C Taylor**, *Spectral theory, Hausdorff dimension and the topology of hyperbolic 3–manifolds*, J. Geom. Analysis (to appear)

[36] **R D Canary**, **E C Taylor**, *Hausdorff dimension and limits of Kleinian groups*, preprint

[37] **V Chuckrow**, *On Schottky groups with applications to Kleinian groups*, Annals of Math. 88 (1968) 47–61

[38] **T D Comar**, *Hyperbolic Dehn surgery and convergence of Kleinian groups*, PhD Thesis University of Michigan (1996)

[39] **M Culler**, **C McA Gordon**, **J Luecke**, **P B Shalen**, *Dehn surgery on knots*, Annals of Math. 125 (1987) 237–300

[40] **M Culler**, **P B Shalen**, *Varieties of group representations and splittings of 3–manifolds*, Annals of Math. 117 (1983) 109–146

[41] **M Culler**, **P. B. Shalen**, *Bounded separating incompressible surfaces in knot manifolds*, Invent. Math. 75 (1984) 537–545

[42] **C E Fan**, *Injectivity radius bounds in hyperbolic convex cores*, PhD Thesis University of Michigan (1997)

[43] **C E Fan**, *Injectivity radius bounds in hyperbolic I-bundle convex cores*, preprint

[44] **C E Fan**, **J Jorgenson**, *Small eigenvalues and Hausdorff dimension of sequences of hyperbolic 3–manifolds*, preprint

[45] **F W Gehring**, **J Väisälä**, *Hausdorff dimension and quasiconformal mappings*, Journal London Math. Soc. 6 (1973) 504–512

[46] **L Greenberg**, *Discrete subgroups of the Lorentz group*, Math. Scand. 10 (1962) 85–107

[47] **M Gromov**, *Hyperbolic manifolds after Jørgensen and Thurston*, from: "Séminaire Bourbaki", Lecture Notes in Math. 842, Springer–Verlag (1981) 40–53

[48] **J Hempel**, *3–manifolds*, Annals of Math. Studies 86, Princeton University Press (1976)

[49] **W Jaco**, *Lectures on three-manifold topology*, Conference board of the mathematical sciences, regional conference series in mathematics no. 43, AMS (1980)

[50] **W Jaco**, **P B Shalen**, *Seifert fibered spaces in 3–manifolds*, Memoirs AMS 220 (1979)

[51] **K Johannson**, *Homotopy equivalences of 3–manifolds with boundaries*, Lecture notes in Math. vol. 761, Springer–Verlag (1979)

[52] **T Jørgensen**, *On cyclic groups of Möbius transformations*, Math. Scand. 33 (1973) 250–260

[53] **T Jørgensen**, *On discrete groups of Möbius transformations*, Amer. J. Math. 98 (1976) 739–749

[54] **T Jørgensen**, **P Klein**, *Algebraic convergence of finitely generated Kleinian groups*, Quart. J. Math. Oxford 33 (1982) 325–332

[55] **T Jørgensen**, **A Marden**, *Algebraic and geometric convergence of Kleinian groups*, Math. Scand. 66 (1990) 47–72

[56] **L Keen**, **C Series**, *Pleating coordinates for the Teichmüller space of a punctured torus*, Bull. Amer. Math. Soc. 26 (1992) 141–146

[57] **L Keen**, **C Series**, *Pleating coordinates for the Maskit embedding of the Teichmüller space of a punctured torus*, Topology 32 (1993) 719–749

[58] **L Keen**, **C Series**, *Pleating invariants for punctured torus groups*, preprint

[59] **L Keen**, **C Series**, *The Riley slice of Schottky space*, Proc. London Math. Soc. 69 (1994) 72–90

[60] **S P Kerckhoff**, **W P Thurston**, *Non-continuity of the action of the mapping class group at Bers' boundary of Teichmüller space*, Invent. Math. 100 (1990) 25–47

[61] **I Kra**, *On lifting Kleinian groups to* $\mathrm{SL}_2(\mathbf{C})$, from: "Differential Geometry and Complex Analysis", I Chavel and H Farkas (editors), Springer–Verlag (1985) 181–193

[62] **I Kra**, *On spaces of Kleinian groups*, Comm. Math. Helv. 47 (1972) 53–69

[63] **O Lehto**, **K I Virtanen**, *Quasiconformal mappings in the plane*, Springer–Verlag (1973)

[64] **A Marden**, *The geometry of finitely generated Kleinian groups*, Annals of Math. 99 (1974) 383–462

[65] **A Marden**, *Geometric relations between homeomorphic Riemann surfaces*, Bull. Amer. Math. Soc. 3 (1980) 1001–1017

[66] **B Maskit**, *On boundaries of Teichmüller spaces and on Kleinian groups II*, Annals of Math. 91 (1970) 607–639

[67] **B Maskit**, *A characterization of Schottky groups*, J. d'Analyse Math. 19 (1967) 227–230

[68] **B Maskit**, *Kleinian groups*, Springer–Verlag (1988)

[69] **B Maskit**, *Parabolic elements in Kleinian groups*, Annals of Math. 117 (1983) 659–668

[70] **B Maskit**, *Self-maps of Kleinian groups*, Amer. J. Math. 93 (1971) 840–856

[71] **K Matsuzaki**, **M Taniguchi**, *Hyperbolic Manifolds and Kleinian Groups*, Clarendon Press, (1998)

[72] **D McCullough**, *Compact submanifolds of* 3*–manifolds with boundary*, Quart. J. Math. Oxford 37 (1986) 299–307

[73] **D McCullough**, **A Miller**, **G A Swarup**, *Uniqueness of cores of non-compact* 3*–manifolds*, Journal London Math. Soc. 32 (1985) 548–556

[74] **C T McMullen**, *Cusps are dense*, Annals of Math. 133 (1991) 217–247

[75] **C T McMullen**, *Hausdorff dimension and conformal dynamics I: Strong convergence of Kleinian groups*, preprint

[76] **Y N Minsky**, *The classification of punctured-torus groups*, preprint

[77] **J W Morgan**, *On Thurston's uniformization theorem for three-dimensional manifolds*, from: "The Smith Conjecture", J Morgan and H Bass (editors), Academic Press (1984) 37–125

[78] **J Morgan**, **P B Shalen**, *Valuations, trees, and degenerations of hyperbolic structures I*, Annals of Math. 120 (1984) 401–476

[79] **J Morgan**, **P B Shalen**, *Degenerations of hyperbolic structures II: Measured laminations in* 3*–manifolds*, Annals of Math. 127 (1988) 403–456

[80] **J Morgan**, **P B Shalen**, *Degenerations of hyperbolic structures III: Actions of* 3*–manifold groups on trees and Thurston's compactification theorem*, Annals of Math. 127 (1988) 457–519

[81] **G D Mostow**, *Strong rigidity of locally symmetric spaces*, Annals of Math. Studies 78, Princeton University Press (1973)

[82] **K Ohshika**, *Divergent sequences of Kleinian groups*, preprint

[83] **K Ohshika**, *Ending laminations and boundaries for deformation spaces of Kleinian groups*, Journal London Math. Soc. 42 (1990) 111–121

[84] **K Ohshika**, *Geometric behaviour of Kleinian groups on boundaries for deformation spaces*, Quart. J. Math. Oxford 43 (1992) 97–111

[85] **K Ohshika**, *Geometrically finite Kleinian groups and parabolic elements*, preprint

[86] **K Ohshika**, *Kleinian groups which are limits of geometrically finite groups*, preprint

[87] **K Ohshika**, *Strong convergence of Kleinian groups and Caratheodory convergence of domains of discontinuity*, Math. Proc. Camb. Phil. Soc. 112 (1992) 297–307

[88] **J P Otal**, *Sur la degenerescence des groupes de Schottky*, Duke Math. J. 74 (1994) 777–792

[89] **J P Otal**, *Le Théoréme d'hyperbolisation pour les variétés fibrées de dimension 3*, Astérisque 235 (1996)

[90] **J P Otal**, **F Paulin**, *Géométrie hyperbolique et groupes kleiniens*, in preparation

[91] **G Prasad**, *Strong rigidity in* $\mathbf{Q}$*–rank* 1 *lattices*, Invent. Math. 21 (1973) 255–286

[92] **D Ruelle**, *Repellers for real analytic maps*, Ergodic Theory and Dynam. Sys. 2 (1982) 99–107

[93] **G P Scott**, *Compact submanifolds of* 3*–manifolds*, Journal London Math. Soc. 7 (1973) 246-250

[94] **P B Shalen**, *Representations of* 3*–manifold groups and applications in topology*, Proceedings of the International Congress of Mathematicians (1986) 607–614

[95] **D P Sullivan**, *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, from: "Riemann surfaces and related topics", I Kra and B Maskit (editors), Annals of Math. Studies 97, Princeton University Press (1981) 465–496

[96] **D P Sullivan**, *Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups*, Acta Math. 153 (1984) 259–277

[97] **D P Sullivan**, *Quasiconformal homeomorphisms and dynamics II: Structural stability implies hyperbolicity of Kleinian groups*, Acta Math. 155 (1985) 243–260

[98] **E C Taylor**, *Geometric finiteness and the convergence of Kleinian groups*, Comm. Analysis and Geometry, (to appear)

[99] **W P Thurston**, *Hyperbolic structures on 3–manifolds I: Deformation of acylindrical manifolds*, Annals of Math. 124 (1986) 203–246

[100] **W P Thurston**, *Hyperbolic structures on 3–manifolds II: Surface groups and 3–manifolds which fiber over the circle*, preprint

[101] **W P Thurston**, *Three-dimensional manifolds, Kleinian groups, and hyperbolic geometry*, Bull. Amer. Math. Soc. 6 (1982) 357–381

[102] **W P Thurston**, *The geometry and topology of 3–manifolds*, Lecture notes Princeton University (1978)

[103] **D J Wright**, *The shape of the boundary of Maskit's embedding of the Teichmüller space of once punctured tori*, preprint

Faculty of Mathematical Studies
University of Southampton
Southampton, SO17 1BJ, England

Email: `jwa@maths.soton.ac.uk`

Received: 15 November 1997

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997 **51**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 51–97

Boundaries of strongly accessible hyperbolic groups

B H Bowditch

Abstract We consider splittings of groups over finite and two-ended subgroups. We study the combinatorics of such splittings using generalisations of Whitehead graphs. In the case of hyperbolic groups, we relate this to the topology of the boundary. In particular, we give a proof that the boundary of a one-ended strongly accessible hyperbolic group has no global cut point.

AMS Classification 20F32

Keywords Boundary, accessibility, hyperbolic group, cutpoint, Whitehead graph

Dedicated to David Epstein in celebration of his 60th birthday.

0 Introduction

In this paper, we consider splittings of groups over finite and two-ended (ie virtually cyclic) groups. A "splitting" of a group, Γ, over a class of subgroups may be viewed a presentation of Γ as a graph of groups, where each edge group lies in this class. The splitting is "non-trivial" if no vertex group equals Γ. It is said to be a splitting "relative to" a given set of subgroups, if every subgroup in this set can be conjugated into one of the vertex groups. Splittings of a given group are often reflected in its large scale geometry. Thus, for example, Stallings's theorem [27] tells us that a finitely generated group splits non-trivially over a finite group if and only if it has more than one end. Furthermore, splittings of a hyperbolic groups over finite and two-ended subgroups can be seen in the topology of its boundary. An investigation of this phenomenon will be one of the main objectives of this paper.

The extent to which a group can be split indefinitely over a certain class of subgroups is described by the notion of "accessibility". Suppose, Γ is a group, and $\mathcal{C}$ is a set of subgroups of Γ. We say that Γ is *accessible* over $\mathcal{C}$ if it can be represented as a finite graph of groups with all edge groups lying in $\mathcal{C}$, and such that no vertex groups splits non-trivially relative to the incident edge groups. Dunwoody's theorem [10] tells us that any finitely presented group is accessible over all finite subgroups. The result of [1] generalises this to "small" subgroups.

There are also stronger notions of accessibility, which have been considered by Swarup, Dunwoody and others. One definition is as follows. Let $\mathcal{C}$ be a set of subgroups of Γ. Any subgroup of Γ which does not split non-trivially over $\mathcal{C}$ is deemed to be "strongly accessible" over $\mathcal{C}$. Then, inductively, any subgroup which can be expressed as a finite graph of groups with all edge groups in $\mathcal{C}$ and all vertex groups strongly accessible is itself deemed to be "strongly accessible". Put another way, Γ is strongly accessible if some sequence of splittings of Γ must terminate in a finite number of steps ending up with a finite number of groups which split no further. (Of course, this definition leaves open the possibility that there might be a different sequence of splittings which does not terminate.) If $\mathcal{C}$ is the set of finite subgroups, then strong accessibility coincides with the standard notion of accessibility, and is thus dealt with by Dunwoody's theorem in the case of finitely presented groups. Recently Delzant and Potyagailo [8] have shown that any finitely presented group is strongly accessible over any elementary set of subgroups. (A set $\mathcal{C}$ of subgroups is "elementary" if no element of $\mathcal{C}$ contains a non-cyclic free subgroup, each infinite element of $\mathcal{C}$ is contained in a unique maximal element of $\mathcal{C}$, and each maximal element of $\mathcal{C}$ is equal to its normaliser in Γ.)

If Γ is hyperbolic in the sense of Gromov [15], then the set of all finite and two-ended subgroups is elementary. Thus, the result of [8] tells us that Γ is strongly accessible. (In the context of hyperbolic groups, we shall always take "strongly accessible" to mean strongly accessible over finite and two-ended subgroups.)

The boundary, $\partial\Gamma$, of Γ is a compact metrisable space, and is connected if and only if Γ is one-ended. In this case, it was shown in [3] that $\partial\Gamma$ is locally connected provided it has no global cut point. In this paper, we show (Theorem 9.3):

Theorem *The boundary of a one-ended strongly accessible group has no global cut point.*

Thus, together with [8] and [3], we arrive at the conclusion that the boundary of every one-ended hyperbolic group is locally connected. This was already obtained by Swarup [28] using results from [4,6,19] shortly after the original draft of this paper was circulated (and prior to the result of [8]). An elaboration of the argument was given shortly afterwards in [7].

One consequence of this local connectedness is the fact that every hyperbolic group is semistable at infinity [21]. (It has been conjectured that every finitely presented group has this property.) This implication was observed by Geoghegen and reported in [3]. I am indebted to Ross Geoghegen for the following elaboration of how this works. The semistability of an accessible group is equivalent to the semistability of each of its maximal one-ended subgroups. Suppose, then, that Γ is a one-ended hyperbolic group. It was shown in [3] that $\partial\Gamma$ naturally compactifies the Rips complex, so as to give a contractable ANR, with $\partial\Gamma$ embedded as a Z–set. It follows that semistability at infinity for Γ is equivalent to $\partial\Gamma$ being pointed 1–movable, the latter property being intrinsic to $\partial\Gamma$. Moreover, it was shown in [18] that a metrisable continuum is pointed 1–movable if and only if it has the shape of a Peano continuum (see also [12]). It follows that if Γ is one-ended hyperbolic, then $\partial\Gamma$ is semistable at infinity if and only of $\partial\Gamma$ has the shape of a Peano continuum. (We remark that an alternative route to semistability for a hyperbolic group would be to use the result of [22] in place of Theorem 8.1 of this paper, together with the results of [4,6].)

We shall carry out much of our analysis of splitting in a fairly general context. We remark that any one-ended finitely presented group admits a canonical splitting over two ended subgroups, namely the JSJ splitting (see [24,11,13], or in the context of hyperbolic groups [25,5]). The vertex group are again finitely presented, and so we can split them over finite subgroups as necessary and iterate the process, discarding any finite vertex groups that arise along the way. This eventually leads to a canonical decomposition of the group into one-ended subgroups, none of which split over any two-ended subgroup. Further discussion of this procedure will be given in Section 9. We shall not make any explicit use of the JSJ splitting in this paper.

In this paper, we shall be considering in some detail the general issue of splittings over two-ended subgroups. One point to note (Theorem 2.3) is the following:

Theorem *The fundamental group of a finite graph of groups with two-ended edge groups is one-ended if and only if no vertex group splits over a finite subgroup relative to the incident edge groups.*

(The case where the vertex groups are all free or surface groups is dealt with in [20].)

To find a criterion for recognising whether a given group splits over a finite group relative to a given finite set of two-ended subgroups, we shall generalise work of Whitehead and Otal in the case of free groups. Given a free group, F, and a non-trivial element, $\gamma \in F$, we say that γ is "indecomposable" in F, if it cannot be conjugated into any proper free factor of F.

This can be interpreted topologically. Note that the boundary, ∂F, of F is a Cantor set. We define an equivalence relation, $\approx$, on ∂F, by deeming that $x \approx y$ if and only if either $x = y$ or x and y are the fixed points of some conjugate of γ. Now, it's easily verified that this relation is closed, and so the (equivariant) quotient, $\partial F/\approx$ is compact hausdorff. It was shown in [23] that γ is indecomposable if and only if $\partial F/\approx$ is connected (in which case, $\partial F/\approx$ is locally connected and has no global cut point).

A combinatorial criterion for indecomposability is formulated in [30]. Let $a_1, a_2, \ldots, a_n$ be a system of free generators for F. Let w be a reduced cyclic word in the a_i's and their inverses representing (the conjugacy class of) γ. Let $\mathcal{G}$ be the graph (called the "Whitehead graph") with vertex set $a_1, \ldots, a_n, a_1^{-1}, \ldots, a_n^{-1}$, and with $a_i^{\epsilon_i}$ deemed to be adjacent to $a_j^{\epsilon_j}$ if and only if the string $a_i^{\epsilon_i} a_j^{-\epsilon_j}$ occurs somewhere in w (where $\epsilon_i, \epsilon_j \in \{-1, 1\}$). Suppose we choose the generating set so as to minimise the length of the word w. Then (a simple consequence of) Whitehead's lemma tells us that γ is indecomposable if and only if $\mathcal{G}$ is connected. (Moreover in such a case, $\mathcal{G}$ has no cut vertex.)

This can be reinterpreted in terms of what we shall call "arc systems". Let T be the Cayley graph of F with respect to free generators $a_1 \ldots a_n$. Thus, T is a simplicial tree, whose ideal boundary, ∂T, may be naturally identified with ∂F. The element γ determines a biinfinite arc, β, in T, namely the axis of γ. Let $\mathcal{B}$ be the set of images of β under Γ. We refer to $\mathcal{B}$ as a (Γ–invariant) "arc system". We can reconstruct the Whitehead graph, as well as the equivalence relation $\approx$, from this arc system in a simple combinatorial fashion, as described in Section 3. The above discussion applies equally well if we replace γ by a finite set, $\{\gamma_1, \ldots, \gamma_p\}$, of non-trivial elements of Γ.

One can generalise these notions to an arbitrary hyperbolic group, Γ. Suppose that $\{H_1, \ldots, H_p\}$ is a finite set of two-ended subgroups of Γ. We define an equivalence relation, $\approx$, on $\partial\Gamma$ by identifying the two endpoints of each conjugate to each H_i. Thus, as before, $\partial\Gamma/\approx$ is hausdorff. We shall see (Theorem 5.2) that:

Theorem *$\partial\Gamma/\approx$ is connected if and only if Γ does not split over a finite group relative to $\{H_1, \ldots, H_p\}$.*

We can also give a combinatorial means of recognising if Γ splits in this way. We can decompose its boundary, $\partial\Gamma$, as a disjoint union of two Γ–invariant sets, $\partial_0\Gamma$ and $\partial_\infty\Gamma$, where $\partial_\infty\Gamma$ is the set of singleton components of $\partial\Gamma$. Algebraically this corresponds the action of Γ on a simplicial tree, T, without edge inversions, with finite quotient, and with finite edge stabilisers and finite or one-ended vertex stabilisers. Such an action is given by the accessibility theorem [10]. Each of the vertex groups is quasiconvex, and hence intrinsically hyperbolic. Now, $\partial_\infty\Gamma$ can be canonically identified with ∂T, and the connected components of $\partial_0\Gamma$ are precisely the boundaries of the infinite vertex stabilisers. The infinite vertex stabilisers are, in fact, precisely the maximal one-ended subgroups of Γ. (Note that Γ is virtually free if and only of $\partial_0\Gamma = \emptyset$.) We can construct an analogue of the Whitehead graph by considering the arc system on T, consisting of all the translates of the axes of those H_i which do not fix any vertex of T.

This combinatorial construction can be carried out for any group which is accessible over finite subgroups. Put together with Theorem 2.3, this gives a combinatorial criterion for recognising when a finitely presented group represented as graph of groups with two-ended edge groups is one-ended. This generalises work of Martinez [20]. It is also worth remarking that the result of [2] tells us that such a group is hyperbolic if and only if all the vertex groups are hyperbolic, and there is no Baumslag–Solitar (or free abelian) subgroup.

The structure of this paper is roughly as follows. In Section 1, we explore some general facts about groups accessible over finite groups. In Section 2, we give a criterion (Theorem 2.3) for a finite graph of groups with two-ended edge groups to be one-ended. In Section 3, we study arc systems on trees and their connections to Whitehead graphs. In Section 4, we give an overview of some general facts about quasiconvex splittings. In Section 5, we look at certain quotients of the boundaries of hyperbolic groups, and relate this to some of the combinatorial results of Section 3. In Section 6, we set up some of the general machinery for analysing the topology of the boundaries of hyperbolic groups which split over two-ended subgroups. In Section 7, we look at some implications concerning connectedness properties of boundaries. In Section 8, we apply this specifically to global cut points. Finally, in Section 9, we discuss further the question of strong accessibility of groups over finite and two-ended subgroups.

Much of the material of the original version of this paper was worked out while visiting the University of Auckland. The first draft was written at the University of Melbourne. I would like to thank Gaven Martin as well as Craig Hodgson and Walter Neumann for their respective invitations. The paper was substantially revised in Southampton, with much of the material of Sections 1, 2, 3 and 5 added. I am also grateful to Martin Dunwoody for helpful conversations regarding the latter. Ultimately, as always, I am indebted to my ex-PhD supervisor David Epstein for first introducing me to matters hyperbolical.

1 Trees and splittings

In this section, we introduce some terminology and notation relating to simplicial trees and group splittings.

Let T be a simplicial tree, which we regard a 1–dimensional CW–complex. We write $V(T)$ and $E(T)$ respectively for the vertex set and edge set. Given $v, w \in V(T)$, we write $\mathrm{dist}(v, w)$ for the distance between v and w, in other words, the number of edges in the arc connecting v to w. If $\vec{e} \in \vec{E}(T)$ and $v \in V(T)$, we say that $\vec{e}$ "points towards" v if $\mathrm{dist}(v, \mathrm{tail}(\vec{e})) = \mathrm{dist}(v, \mathrm{head}(\vec{e})) + 1$.

If $S \subseteq T$ is a subgraph, we write $V(S) \subseteq V(T)$ and $E(S) \subseteq E(T)$ for the corresponding vertex and edge sets. A subtree of T is a connected subgraph. Of particular interest are "rays" and "biinfinite arcs" (properly embedded subsets homeomorphic to $[0, \infty)$ and $\mathbb{R}$ respectively.)

We may define the ideal boundary, ∂T, of T, as the set of cofinality classes of rays in Σ. We shall only be interested in ∂T as a set. (In fact, $T \cup \partial T$ can be given a natural compact topology as a dendron, as discussed in [4]. It can also be given a finer topology by viewing T has a Gromov hyperbolic space, and ∂T as its Gromov boundary.) If $S \subseteq T$ is a subgraph, we write $\partial S \subseteq \partial T$ for the subset arising from those rays which lie in S. Note that if β is a biinfinite arc, then $\partial\beta$ contains precisely two points, $x, y \in \partial T$. We say that β *connects* x to y.

Further discussion of general simplicial trees will be given in Sections 2 and 3. We now move on to consider group actions on trees.

Let G be a group. A *G–tree* is a simplicial tree, T, admitting a simplicial action of G without edge inversions. If $v \in V(T)$ and $e \in E(T)$, we write $G_T(v)$ and $G_T(e)$ for the corresponding vertex and edge stabilisers respectively. Where there can be no confusion, we shall abbreviate these to $G(v)$ and $G(e)$. Such a tree gives rise to a splitting of G as a graph of groups, G/T. We shall say that T is *cofinite* if T/G is finite. We shall usually assume that T is *minimal*, ie that there is no proper G–invariant subtree. This is the same as saying that T has no terminal vertex, or, on the level of the splitting, that no vertex group of degree one is equal to the incident edge groups. Such a vertex will be referred to as a *trivial vertex*. A subset (usually a subgroup) H, of G is *elliptic* with respect to T, if it lies inside some vertex stabiliser. If $\mathcal{H}$ is a set of subsets of G, we say that the splitting is *relative to* $\mathcal{H}$, if every element of $\mathcal{H}$ is an elliptic subset. We note that any finite subgroup of a group is elliptic with respect to every splitting. Thus any splitting of any group is necesarily relative to the set of all finite subgroups.

Suppose that F is a G–invariant subgraph of T, we can obtain a new G–tree, Σ, by collapsing each component of F to a point. We speak of the splitting T/G as being a *refinement* of the splitting Σ/G. Note that one may obtain a refinement of a given graph of groups, if one of the vertex groups splits relative to its incident edge groups.

We say that a G–tree, T', is a *subdivision* of T, if it is obtained by inserting degree–2 vertices into the edges of T in a G–equivariant fashion. Suppose that Σ is another G–tree. A *folding* of T onto Σ is a G–equivariant map of T onto Σ such that each edge of T either gets mapped homeomorphically onto an edge of Σ or gets collapsed to a vertex of Σ. A *morphism* of T onto Σ is a folding of some subdivision of T. Such maps are necessarily surjective provided that Σ is minimal. Clearly a composition of morphisms is a morphism.

We say that T *dominates* Σ (or that the splitting T/G *dominates* Σ/G) if there exists a morphism from T to Σ. It's not hard to see that this is equivalent to saying that every vertex stabiliser in T is elliptic with respect to Σ. We say that T and Σ are *equivalent* if each dominates the other. This is equivalent to saying that a subset of G is elliptic with respect to T if and only if it is elliptic with respect to Σ.

Suppose that T is cofinite. If T dominates Σ, then Σ is also cofinite. In this case, any morphism from T to Σ expands combinatorial distances by at most a bounded factor (namely the maximum number of edges into which we need to subdivide a given edge of T to get a folding.) Also, any two morphisms remain a bounded distance apart. In particular, any self-morphism of a cofinite tree is a bounded distance from the identity map, and is thus a quasiisometry. Suppose that T and Σ are equivalent, and that $\phi\colon T \longrightarrow \Sigma$ is a morphism. Let $\psi\colon \Sigma \longrightarrow T$ be any morphism. Now, since ψ expands distances by a bounded factor, and $\psi \circ \phi$ is a quasiisometry, it follows that ϕ is itself a quasiisometry. In summary, we have shown:

Lemma 1.1 *If T and Σ are equivalent cofinite G–trees, then any morphism from T to Σ is quasiisometry.* □

We see from the above discussion that there is a natural bijective correspondence between the boundaries, ∂T and $\partial\Sigma$, of T and Σ.

Lemma 1.2 *Suppose that T and Σ are cofinite G–trees with finite edge-stabilisers. If $\phi\colon T \longrightarrow \Sigma$ is a folding, then only finitely many edges of T get mapped homeomorhically under ϕ to any given edge of Σ.*

Proof If $\gamma \in \Gamma$ and $e, \gamma e \in E(T)$ both get mapped homeomorpically onto some edge $\epsilon \in E(\Sigma)$, then $\gamma \in \Gamma_\Sigma(\epsilon)$. There are thus only finitely many such edges in the Γ–orbit of e in $E(T)$. The result follows since $E(T)/\Gamma$ is finite. □

We shall need to elaborate a little on the notion of accessibility over finite groups. For the remainder of this section, all splittings will be assumed to be over finite groups, and the term "accessible" is assumed to mean "accessible over finite groups".

We shall say that a graph of groups is *reduced* if no vertex group of degree one or two is equal to an incident edge group. (Every graph of groups is a refinement of a reduced graph.) We say that a group G is "accessible" if there is a bound on the complexity (as measured by the number of edges) of a splitting of G as a reduced graph of groups (with finite edge groups). Among graphs of maximal complexity, one for which the sum of the orders of the edge stabilisers is minimal will be referred to as a "complete splitting". By Dunwoody's theorem [10], any finitely presented group is accessible. (This has been generalised to splittings over small subgroups by Bestvina and Feighn [1].)

This can be rephrased in terms of one-ended subgroups. For this purpose, we define a group to be *one-ended* if it is infinite and does not split non-trivially (over any finite subgroup). Thus, by Stallings's theorem, this coincides with the usual topological notion for finitely generated groups. Suppose that G is accessible, and we take a complete splitting of G. Now any splitting of a vertex group is necessarily relative to the incident edge groups, and so would give rise to a refined splitting. It is possible that this refined splitting may no longer be reduced, but in such a case, we can coalesce two vertex groups, to produce a reduced graph with one smaller edge stabiliser than the original, thereby contradicting completeness. In summary, we see that all the vertex groups of a complete splitting are either finite or one-ended. In fact, we see that the infinite vertex groups are precisely the maximal one-ended subgroups. It turns out that there is a converse to this statement: any group which can be represented as a finite graph of groups with finite edge groups and with all vertex groups finite or one-ended is necessarily accessible (see [9]).

Finally, suppose that G is accessible, and we represent it as a finite graph of groups over finite subgroups. Now each vertex group must be accessible. Taking complete splittings of each of the vertex groups, we can see that we can refine the original splitting in such a way that all the vertex groups are finite or one-ended. (It is possible that this refinement might not be reduced.)

Now, let G be an accessible group, and let T be a cofinite tree with finite edge stabilisers and with every vertex stabilisers either finite or one-ended. The infinite vertex groups are canonically determined. We have also observed that finite groups are always elliptic in any splitting. It follows that if T' is another such G–tree, then T and T' are equivalent, by Lemma 1.1. In particular ∂T and $\partial T'$ can be canonically (and hence G–equivariantly) identified. We can thus associate to any accessible group, G, a canonical G–set, $\partial_\infty G$, which we may identify with the boundary of any such G–tree.

Clearly in the case of a free group, we just recover the usual boundary. More generally, if G is (word) hyperbolic (and hence accessible) then we may identify $\partial_\infty G$ with the set of singleton components of the boundary, ∂G. In fact, as discussed in the introduction, we can write ∂G as a disjoint union $\partial_0 G \sqcup \partial_\infty G$, where each component of $\partial_0 G$ is the boundary of a maximal one-ended subgroup of G.

We shall make some further observations about accessible groups in connection with strong accessibility in Section 9.

2 Splittings over two-ended subgroups

The main aim of this section will be to give a proof of Theorem 2.3. We first introduce some terminology regarding "arc systems" which will be relevant to later sections.

Let T be a simplicial tree.

Definition An *arc system*, $\mathcal{B}$, on T consists of a set of biinfinite arcs in T.

We say that $\mathcal{B}$ is *edge-finite* if at most finitely many elements of $\mathcal{B}$ contain any given edge of T.

If G is a group, and T is a G–tree, then we shall assume that an arc system on T is G–invariant.

Recall that a subgroup, H, of G is "elliptic" if it fixes a vertex of T. If H is two-ended (ie virtually cyclic) then either H is elliptic, or else there is a biinfinite β in T which is H–invariant. In the latter case, we say that H is *hyperbolic* and that β is the *axis* of H. Clearly, the H–stabiliser of any edge of $\mathcal{B}$ is finite.

Suppose now that all edge stabilisers of T are finite. Then every hyperbolic two-ended subgroup of G lies in a unique maximal two-ended subgroup of G, namely the setwise stabiliser of the axis. Note also that there are only finitely many two-ended subgroups, H, with a given axis, $\mathcal{B}$, and with the number of edges of β/H bounded. In particular, we see that only finitely many G–conjugates of a given hyperbolic two-ended subgroup, H, can share the same axis.

Suppose, now, that $\mathcal{H}$ is a finite union of conjugacy classes of two-ended subgroups of G, and that $\mathcal{B}$ is the set of all axes of all hyperbolic elements of $\mathcal{H}$. (In other words, $\mathcal{B}$ is an arc-system with $\mathcal{B}/\Gamma$ finite, and such that the setwise stabiliser of each element of $\mathcal{B}$ is infinite, and hence two-ended.) We note:

Lemma 2.1 *The arc system $\mathcal{B}$ is edge-finite.*

Proof We want to show that any given edge lies in a finite number of elements of $\mathcal{B}$. Without loss of generality, we can suppose that $\mathcal{B}$ conists of the orbit of a single arc, β. Let H be the setwise stabiliser of β. Choose any edge $e \in T$. Let $K \leq G$ be the stabiliser of e. Without loss of generality, we may as well suppose that $e \in E(\beta)$. Note that $E(\beta)/H$ is finite. Now, the G–orbit, Ge, of e meets $E(\beta)$ in an H–invariant set consisting of finitely many H–orbits, say $Ge \cap E(\beta) = Hg_1e \cup Hg_2e \cup \cdots \cup Hg_ne$, where $g_i \in G$.

Suppose that $e \subseteq g\beta$, for some $g \in G$. Now $g^{-1}e \in E(\beta)$, so $g^{-1}e = hg_ie$ for some $h \in H$, and $i \in \{1, \ldots, n\}$. Thus $ghg_i \in K$, so $gH = kg_i^{-1}H$ for some $k \in K$. Since K is finite, there are finitely many possibilities for the right coset gH, and hence for the arc $g\beta$. □

Now, let $\mathcal{H}$ be any finite union of conjugacy classes of two ended subgroups of G, as above. Recall that to say that G splits over a finite subgroup relative to $\mathcal{H}$ means that there is a non-trivial G–tree with finite edge stabilisers, and with each element of $\mathcal{H}$ elliptic with respect to T. We can always take such a G–tree to be cofinite, and indeed to have only one orbit of edges. We say that $\mathcal{H}$ is *indecomposable* if G does not split over any finite group relative to $\mathcal{H}$.

In Section 3, we shall give a general criterion for indecomposability in terms of arc systems. For the moment, we note:

Lemma 2.2 *Suppose that G is a group and that T is a G–tree with finite edge stabilisers. Suppose that $\mathcal{H}$ is a finite union of conjugacy classes of two-ended subgroups of G. Let $\mathcal{B}$ be the arc system consisting of the set of axes of hyperbolic elements of G. If $\mathcal{H}$ is indecomposable, then each edge of T lies in at least two elements of $\mathcal{B}$.*

Proof Suppose that $T \neq \bigcup \mathcal{B}$. Then, collapsing each component of $\bigcup \mathcal{B}$ to a point, we obtain another G–tree, Σ, with finite edge stabilisers. Moreover, each element of $\mathcal{H}$ is elliptic with respect to Σ, contradicting indecomposability.

We thus have $T = \bigcup \mathcal{B}$. Suppose, for contradiction, that there is an edge of T which lies in precisely one element of $\mathcal{B}$. We may as well suppose that this is true of all edges of T. (For if not, let F be the union of all edges of T which lie in at least two elements of $\mathcal{B}$. Collapsing each component of F to a point, we obtain a new G–tree. We replace $\mathcal{B}$ by the set of axis of those elements of $\mathcal{H}$ which remain hyperbolic. Thus each element of the new arc system is the result of collapsing an element of the old arc system along a collection of disjoint compact subarcs.)

We now construct a bipartite graph, Σ, with vertex set an abstract disjoint union of $V(T)$ and $\mathcal{B}$, by deeming $x \in V(T)$ and $\beta \in \mathcal{B}$ to be adjacent in Σ if $x \in \beta$ in T. Now, it's easily verified that Σ is a simplicial tree, and that the stabiliser of each pair $(x, \mathcal{B})$ is finite. In other words, Σ is a G–tree with finite edge stabilisers. Finally, we note that each element of $\mathcal{H}$ is elliptic in Σ. This again contradicts the indecomposability of $\mathcal{H}$. $\square$

We now move on to considering splittings over two-ended subgroups. Suppose that Γ is a group, and that Σ is a cofinite Γ–tree (with no terminal vertex) and with two-ended edge-stabilisers. We can write $V(\Sigma)$ as a disjoint union, $V(\Sigma) = V_1(\Sigma) \sqcup V_2(\Sigma) \sqcup V_\infty(\Sigma)$, depending on whether the corresponding vertex stabiliser is one, two or infinite-ended. Note that $V_2(\Sigma)$ is precisely the set of vertices of finite degree.

We remark that if there is a bound on the order of finite subgroups of Γ, and there are no infinitely divisible elements, then each two-ended subgroup lies in a unique maximal two-ended subgroup. In this case, we can refine our splitting so that for each vertex $v \in V_1(\Sigma) \cup V_\infty(\Sigma)$, the incident edge groups are all maximal two-ended subgroups of $\Gamma(v)$. This is automatically true of the JSJ splitting of hyperbolic groups (as described in [5]), for example, though we shall have no need to assume this in this section.

It is fairly easy to see that the one-endedness or otherwise of Γ depends only on the infinite-ended vertex groups, $\Gamma(v)$ for $v \in V_\infty(\Sigma)$. In one direction, it easy to see that if one of these groups splits over a finite group relative to incident edge groups, then we can refine our splitting so that one of the new edge groups is finite. Hence Γ is not one-ended. In fact, we also have the converse. Recall that a "trivial vertex" of a splitting is a vertex of degree 1 such that the vertex group equals the adjacent edge group (ie it corresponds to a terminal vertex of the corresponding tree).

Theorem 2.3 *Suppose we represent a group, Γ, as finite graph of groups with two-ended vertex groups and no trivial vertices. Then, Γ is one-ended if and only if none of the infinite-ended vertex groups split intrinsically over a finite subgroup relative to the incident edge groups.*

Proof Let Σ be the Γ–tree corresponding to the splitting, and write $V(\Sigma) = V_1(\Sigma) \sqcup V_2(\Sigma) \sqcup V_\infty(\Sigma)$ as above. Given $v \in V(\Sigma)$ let $\Delta(v) \subseteq E(\Sigma)$ be the set of incident edges. We are supposing that for each $v \in V_\infty(\Sigma)$, the set of incident edge stabilisers, $\{\Gamma_\Sigma(e) \mid e \in \Delta(v)\}$, is indecomposable in the group $\Gamma_\Sigma(v)$. This is therefore true for all $v \in V(\Sigma)$. We aim to show that Γ is one-ended.

Suppose, for contradiction, that there exists a non-trivial minimal G–tree, T, with finite edge stabilisers. Let $\mathcal{B}$ be the arc system on T consisting of the axes of those Σ–edge stabilisers, $\Gamma_\Sigma(e)$, which are hyperbolic with respect to T. By Lemma 2.1, $\mathcal{B}$ is edge-finite.

Suppose, first, that $\mathcal{B} = \emptyset$, ie each group $\Gamma_\Sigma(e)$ for $e \in E(\Sigma)$ is elliptic in T. Suppose $v \in V(\Sigma)$. Since $\{\Gamma_\Sigma(e) \mid e \in \Delta(v)\}$ is indecomposable in $\Gamma_\Sigma(v)$, it follows that $\Gamma_\Sigma(v)$ must be elliptic in T. It therefore fixes a unique vertex of T. Suppose $w \in V(\Sigma)$ is adjacent to v. Since $\Gamma_\Sigma(v) \cap \Gamma_\Sigma(w)$ is infinite, it follows that $\Gamma_\Sigma(w)$ must also fix the same vertex of T. Continuing in this way, we conclude that this must be true of all Σ–vertex stabilisers. We therefore arrive at the contradiction that Γ fixes a vertex of T.

We deduce that $\mathcal{B} \neq \emptyset$. Now, choose any $\beta \in \mathcal{B}$ and any edge $\epsilon \in E(\beta)$. By construction, β is the axis of some edge stabiliser $\Gamma_\Sigma(e_0)$ for $e_0 \in E(\Sigma)$. Let $v \in V(\Sigma)$ be an endpoint of e_0. Now, $\Gamma_\Sigma(e_0) \subseteq \Gamma_\Sigma(v)$, so $\Gamma_\Sigma(v)$ is not elliptic in T. It follows that $v \notin V_1(\Sigma)$. If $v \in V_2(\Sigma)$, then β is the axis in T of $\Gamma_\Sigma(v)$, and hence of any edge $e_1 \in E(\Sigma)$ adjacent to e_0. In particular, ϵ lies in the axis of $\Gamma_\Sigma(e_1)$. If $v \in V_\infty(\Sigma)$, let $T(v)$ be the unique minimal $\Gamma_\Sigma(v)$–invariant subtree of T. Let $\mathcal{B}(v)$ be the set of axis of hyperbolic elements of $\{\Gamma_\Sigma(e) \mid e \in \Delta(v)\}$. Thus, $\mathcal{B}(v) \subseteq \mathcal{B}$ is an arc system on $T(v)$, and $\beta \in \mathcal{B}(v)$. By Lemma 2.2, there is some $\beta' \in \mathcal{B}(v) \setminus \{\beta\}$ with $\epsilon \in E(\beta')$. Now, β' is the axis of $\Gamma_\Sigma(e_1)$ for some edge $e_1 \in E(\Sigma)$ adjacent to e_0, as in the case where $v \in V_2(\Sigma)$. Now, in the same way, we can find some edge e_2 incident on the other endpoint of e_1, so that $\Gamma_\Sigma(e_2)$ is hyperbolic in T and contains ϵ in its axis. Continuing, we get an infinite sequence of edges, $(e_n)_{n\in\mathbb{N}}$, which form a ray in Σ, and which all have this property.

Now, since $\mathcal{B}$ is edge-finite, we can pass to a subsequence so that the axes of the groups $\Gamma_\Sigma(e_n)$ are constant. Since Σ is cofinite, we can find an edge $e \in E(\Sigma)$ and an element $\gamma \in \Gamma$ which is hyperbolic in Σ, and such that the axes of $\Gamma_\Sigma(e)$

and $\Gamma_\Sigma(\gamma e) = \gamma\Gamma_\Sigma(e)\gamma^{-1}$ in T are equal to α, say. In particular, $\gamma\alpha = \alpha$. Now, $\Gamma_\Sigma(e)$ has finite index in the setwise stabiliser of α, and so some power of γ lies in $\Gamma_\Sigma(e)$, contradicting the fact that γ is hyperbolic in Σ.

This finally contradicts the existence of the Γ–tree T. □

We note that Theorem 2.3 gives a means of describing the indecomposibility of a set of two-ended subgroups in terms of the "doubled" group, as follows.

Suppose that G is a group, and that $\mathcal{H}$ is a union of conjugacy classes of subgroups. We form a graph of groups with two vertices as follows. We take two copies of G as vertices, and connect them by a set of edges, one for each conjugacy class of subgroup in $\mathcal{H}$. We associate to each edge the corresponding group. We refer to the fundamental group of this graph of groups as the *double* of G in $\mathcal{H}$, and write it as $D(G,\mathcal{H})$. For example, if H is any subgroup of G and $\mathcal{H}$ is its conjugacy class, then we just get the amalgamated free product, $D(G,\mathcal{H}) \cong G *_H G$.

From Theorem 2.3, we deduce immediately:

Corollary 2.4 *Suppose that G is a group, and that $\mathcal{H}$ is a union of finitely many conjugacy classes of two-ended subgroups. Then, $\mathcal{H}$ is indecomposable in G if and only if the double, $D(G,\mathcal{H})$, is one-ended.* □

We note that Theorem 2.3 can be extended to allow for one-ended edge groups. The hypotheses remain unaltered. We simply demand that no vertex group splits over a finite group relative to the set of two-ended incident edge groups. The argument remains essentially unchanged. If, however, we allow for infinite-ended edge groups, then Theorem 2.3 and Corollary 2.4 may fail.

Consider, for example, a one-ended group, K, with an infinite order element $a \in K$. Let G be the free product $K * \mathbb{Z}$, and write $b \in G$ for the generator of the $\mathbb{Z}$ factor. Let $H \leq G$ be the subgroup generated by a and b. Thus, H is free of rank 2. Now, the conjugacy class of H is indecomposable in G. (For suppose that T is a G–tree with finite edge stabilisers and with H elliptic. Now, since K is one-ended, it is also elliptic. Since $K \cap H$ is infinite, and since $K \cup H$ generates G, we arrive at the contradiction that G is elliptic.) However, $G *_H G$ is not one-ended. In fact, $G *_H G \cong (K *_{\langle a\rangle} K) * \mathbb{Z}$. We remark that by taking $\langle a\rangle$ to be malnormal in K (for example taking K to be any torsion-free one-ended word hyperbolic group, and taking a to be any infinite order element which is not a proper power) we can arrange that H is malnormal in G.

3 Indecomposable arc systems

In this section, we look further at arc systems and give a combinatorial characterisation of indecomposability. First, we introduce some additional notation concerning trees.

Suppose $S \subseteq T$ is a subtree. We write $\pi_S\colon T \cup \partial T \longrightarrow S \cup \partial S$ for the natural retraction. Thus, $\pi_S((T \cup \partial T) \setminus (S \cup \partial S)) \subseteq V(S) \subseteq S$. If $R \subseteq S$ is another subtree, then $\pi_R \circ \pi_S = \pi_R$. Moreover, $\pi_R|(S \cup \partial S)$ is defined intrinsically to S.

If $v \in V(S)$, then $T \cap \pi_S^{-1}(v)$ is a subtree of T, which we denote by $F(S,v)$. Note that $F(s,v) \cap S = \{v\}$, and that $\partial F(S,v) = \partial T \cap \pi_S^{-1}(v)$. Also, $T = S \cup \bigcup_{v \in V(S)} F(S,v)$.

We begin by describing generalisations of Whitehead graphs. For the moment, we do not need to introduce group actions.

Let T be a simplicial tree. We write $\mathcal{S}(T)$ for the set of finite subtrees of T. We can think of $\mathcal{S}(T)$ as a directed set under inclusion. Given $S \in \mathcal{S}(T)$, we define an equivalence relation, $\approx_S$, on ∂T by writing $x \approx_S y$ if $\pi_S x = \pi_S y$. In other words, $x \approx_S y$ if and only if the arc connecting x to y meets S in at most one point. Clearly, if $S \subseteq R \in \mathcal{S}(T)$, then $\approx_R$ is finer than $\approx_S$. We therefore get a direct limit system of equivalence relations indexed by $\mathcal{S}(T)$. The direct limit (ie intersection) of these relations is just the equality relation on ∂T.

Suppose now that $\mathcal{B}$ is an arc system on T. We have another equivalence relation, $\approx_{\mathcal{B}}$, on ∂T defined as follows. We write $x \approx_{\mathcal{B}} y$ if $x = y$ or if there exists some $\beta \in \mathcal{B}$ such that $\partial\beta = \{x, y\}$. If the intersection of any two arcs of $\mathcal{B}$ is compact (as in most of the cases in which we shall be interested) then this is already an equivalence relation. If not, we take $\approx_{\mathcal{B}}$ to be the transitive closure of this relation.

Given $S \in \mathcal{S}(T)$, let $\sim_{S,\mathcal{B}}$ be the transitive closure of the union of the relations $\approx_S$ and $\approx_{\mathcal{B}}$. Thus, the relations $\sim_{S,\mathcal{B}}$ again form a direct limit system indexed by $\mathcal{S}(T)$. We write $\sim_{\mathcal{B}}$ for the direct limit.

Definition We say that the arc system $\mathcal{B}$ is *indecomposable* if there is just one equivalence class of $\sim_{\mathcal{B}}$ in ∂T.

We can give a more intuitive description of this construction which ties in with Whitehead graphs as follows. We fix our arc system $\mathcal{B}$. If $S \in \mathcal{S}(T)$, we abbreviate $\sim_{S,\mathcal{B}}$ to $\sim_S$. Note that, if $Q \subseteq \partial T$ is a $\sim_S$–equivalence class, then $Q = \partial T \cap \pi_S^{-1}\pi_S Q$. Let $\mathcal{W}(S)$ be the collection of all sets of the form $\pi_S Q$, as

Q runs over the set, $\partial T/\sim_S$, of $\sim_S$–classes. Thus, $\mathcal{W}(S)$ gives a partition of the subset $\bigcup \mathcal{W}(S)$ of $V(S)$. We refer to $\mathcal{W}(S)$ as a "subpartition" of $V(S)$ (ie a collection of disjoint subsets). There is a natural bijection between $\mathcal{W}(S)$ and the set $\partial T/\sim_S$.

Let us now suppose that $\bigcup \mathcal{B}$ is not contained in any proper subtree of T (for example if $\mathcal{B}$ is indecomposable). Let $\mathcal{B}(S) \subseteq \mathcal{B}$ be the set of arcs which meet S in a non-trivial interval (ie non-empty and not a point). If $\beta \in \mathcal{B}(S)$, we write $I(\beta)$ for the interval $\beta \cap S$, thought of abstractly, and write $\mathrm{fr}\, I(\beta)$ for the set consisting of its two endpoints. Let $Z(S)$ be the disjoint union $Z(S) = \bigsqcup_{\beta \in \mathcal{B}(S)} I(\beta)$, and let $\mathrm{fr}\, Z(S) = \bigsqcup_{\beta \in \mathcal{B}(S)} \mathrm{fr}\, I(\beta)$. There is a natural projection $p\colon Z(S) \longrightarrow S$ with $p(\mathrm{fr}\, Z(S)) \subseteq V(S)$. Now let $\mathcal{G}(S)$ be the quotient space $Z(S)/\cong$, where $\cong$ is the equivalence relation on $Z(S)$ defined by $x \cong y$ if and only if $x = y$ or $x, y \in \mathrm{fr}\, Z(S)$ and $px = py$. We see that $\mathcal{G}(S)$ is a 1–complex, with vertex set, $V(\mathcal{G}(S))$, arising from $\mathrm{fr}\, Z(S)$. The map p induces a natural map from $\mathcal{G}(S)$ to S, also denoted by p. Now, $p|V(\mathcal{G}(S))$ is injective, and $p(V(\mathcal{G}(S))) = \bigcup \mathcal{W}(S)$, where $\mathcal{W}(S)$ is the subpartition of $V(S)$ described earlier. Moreover, an element of $\mathcal{W}(S)$ is precisely the vertex set of connected component of $\mathcal{G}(S)$. If $\mathcal{B}$ is edge-finite, then $\mathcal{G}(S)$ will be a finite graph.

To relate this to the theory of Whitehead graphs, the following observation will be useful. Recall that a graph is 2–vertex connected if it is connected and has no cut vertex. (We consider a graph consisting of a single edge to be 2–vertex connected.)

Lemma 3.1 *Suppose that $S_1, S_2 \in \mathcal{S}(T)$ are such that $S_1 \cap S_2$ consists of a single edge $e \in E(S_1) \cap E(S_2)$. If $\mathcal{G}(S_1)$ and $\mathcal{G}(S_2)$ are 2–vertex connected, then so is $\mathcal{G}(S)$.*

Proof Let $S = S_1 \cup S_2 \in \mathcal{S}(T)$. Let v_1, v_2 be the endpoints of e which are extreme in S_1 and S_2 respectively. Let $V_1 = V(S_1) \setminus \{v_1\}$ and $V_2 = V(S_2) \setminus \{v_2\}$. Write $W_i = p^{-1}(V_i) \subseteq V(\mathcal{G}(S))$ so that $V(\mathcal{G}(S)) = W_1 \sqcup W_2$. Let $\mathcal{G}_i$ be the full subgraph spanned by W_i. Then $\mathcal{G}(S_i)$ is obtained by collapsing $\mathcal{G}_i$ to a single vertex. The result therefore follows from the following observation, of which we omit the proof. □

Lemma 3.2 *Suppose that $\mathcal{G}$ is a connected graph and that $\mathcal{G}_1$ and $\mathcal{G}_2$ are disjoint connected subgraphs. Write $\mathcal{G}_i'$ for the result of collapsing $\mathcal{G}_i$ to a single point in the graph $\mathcal{G}$. If $\mathcal{G}_1'$ and $\mathcal{G}_2'$ are both 2–vertex connected, then so is $\mathcal{G}$.* □

Suppose $v \in V(T)$. Write $S(v)$ for the subtree consisting of the union of all edges incident on v. If T is locally finite, then $S(v) \in \mathcal{S}(T)$. Applying Lemma 3.1 inductively we conclude:

Lemma 3.3 *Suppose that $\mathcal{B}$ is an arc system on the locally finite tree, T, such that $\bigcup \mathcal{B}$ is not contained in any proper subtree. If $\mathcal{G}(S(v))$ is 2–vertex connected for all $v \in V(T)$, then $\mathcal{B}$ is indecomposable.* □

The classical example of this, as discussed in the introduction, is that of Whitehead graphs. Suppose that G is a free group with free generators $a_1, \ldots, a_n$. Let T be the Cayley graph of G with respect to these generators. Thus, T is locally finite cofinite G–tree.

Let $\{\gamma_1, \ldots, \gamma_p\}$ be a finite set of non-trivial elements of G. It's easy to see that the indecomposability of the set of cyclic subgroups $\{\langle\gamma_1\rangle, \ldots \langle\gamma_p\rangle\}$ (as defined in Section 2) is equivalent to that of $\{H_1, \ldots, H_p\}$ where H_k is the maximal cyclic subgroup containing $\langle\gamma_k\rangle$. For this reason, we don't loose any generality by taking the elements γ_k to be indivisible, though this is not essential for what are going to say.

Now, let $\mathcal{B}$ be the arc system consisting of the set of axes of all conjugates of the elements γ_i. Now, the graph $\mathcal{G}(S(v))$ is independent of the choice of vertex $v \in V(T)$, so we may write it simply as $\mathcal{G}$. We can construct $\mathcal{G}$ abstractly as the graph with vertex set $\{a_1, \ldots, a_n, a_1^{-1}, \ldots, a_n^{-1}\}$ where the number of edges connecting $a_i^{\epsilon_i}$ to $a_j^{\epsilon_j}$ equals the total number of times the subword $a_i^{\epsilon_i} a_j^{-\epsilon_j}$ occurs in the (disjoint union of the) reduced cyclic words representing elements γ_k (where $\epsilon_i, \epsilon_j \in \{-1, 1\}$). Thus, the total number of edges in $\mathcal{G}$ equals the sum of the cyclically reduced word lengths of the elements γ_k. The fact that we are taking reduced cyclic words tells us immediately that there are no loops in $\mathcal{G}$. We call $\mathcal{G}$ the *Whitehead graph*. This agrees with the description in the introduction, except that we are now allowing for multiple edges. (To recover the description of the introduction, and that of the original paper [30], we can simply replace each multiple edge by a single edge. This has no consequence for what we are going to say.)

By Lemma 3.3, we see immediately that:

Proposition 3.4 *If $\mathcal{G}$ is 2–vertex connected, then $\mathcal{B}$ is indecomposable.* □

We shall see later, in a more general context, that the indecomposability of $\mathcal{B}$ is equivalent to the indecomposability of the set of subgroups $\{\langle\gamma_1\rangle, \ldots, \langle\gamma_p\rangle\}$.

By a "cut vertex" of $\mathcal{G}$ we mean a vertex of $\mathcal{G}$ which separates the component in which it lies. Now, if $\mathcal{G}$ contains a cut vertex, one can change the generators (in an explicit algorithmic fashion) so as to reduce the total length of $\mathcal{G}$ (allowing multiple edges) — cf [30]. Thus, after a linearly bounded number of steps, we arrive at a Whitehead graph with no cut vertex. (It follows that if we choose generators so as to minimise the sum of the cyclically reduced word lengths of the γ_k, then the Whitehead graph will have this property.) In this case, the Whitehead graph is either disconnected or 2–vertex connected. In the former case, $\mathcal{B}$ is clearly not indecomposable, whereas in the latter case it is (by Proposition 3.4). There is therefore a linear algorithm to decide indecomposability for a finite set of elements in a free group.

We remark that we can also recognise a free generating set by the same process. If $p = n$, then $\{\gamma_1, \ldots, \gamma_n\}$ forms a free generating set if and only if a minimal Whitehead graph (or any Whitehead graph without cut vertices) is a disjoint union of n bigons. (If the elements γ_i are all indivisible, then any component with 2 vertices must be a bigon.) The algorithm arising out of this procedure was one of the main motivations of the original paper [30].

We want to generalise some of this discussion of indecomposability to the context of groups accessible over finite groups, as alluded to in Section 2.

For the moment, suppose that G is any group, and that T and Σ are equivalent cofinite G–trees with finite edge stabilisers. There are morphisms $\phi\colon T \longrightarrow \Sigma$ and $\psi\colon \Sigma \longrightarrow T$. These morphisms are quasiisometries, and hence induce a canonical bijection between ∂T and $\partial\Sigma$. In this case, it is appropriate to deal with formal arc systems, ie (G–invariant) sets of unordered pairs of elements of $\partial T \equiv \partial\Sigma$. Such a formal arc system determines an arc system, $\mathcal{B}$, on T and one, $\mathcal{A}$, on Σ. There is a bijection between $\mathcal{B}$ and $\mathcal{A}$ such that corresponding arcs have the same ideal endpoints. Thus, if $\beta \in \mathcal{B}$, then $\phi(\beta)$ is a subtree of Σ, with $\partial\phi(\beta) \equiv \partial\beta$. We see that the corresponding arc, $\alpha \in \mathcal{A}$ is the unique biinfinite arc contained in $\phi(\beta)$. Note that we get relations $\sim_{\mathcal{B}}$ and $\sim_{\mathcal{A}}$ on $\partial T \equiv \partial\Sigma$, from the direct limit construction described earlier. Our first objective will be to check that these are equal. It follows that the indecomposability of $\mathcal{A}$ and $\mathcal{B}$ are equivalent (Lemma 3.5). We thus get a well-defined notion of indecomposability of formal arc systems for such trees.

Suppose that $S \in \mathcal{S}(T)$. For clarity, we write $\approx_{S,T}$ for the relation on ∂T abbreviated to $\approx_S$ in the previous discussion (ie $x \approx_{S,T} y$ if $\pi_S x = \pi_S y$). We thus have a direct limit system $(\approx_{S,T})_{S\in\mathcal{S}(T)}$. We similarly get another direct limit system $(\approx_{R,\Sigma})_{R\in\mathcal{S}(\Sigma)}$. We claim that these are cofinal. In other words, for each $S \in \mathcal{S}(T)$, there is some $R \in \mathcal{S}(\Sigma)$ such that the relation $\approx_{R,\Sigma}$ is finer than $\approx_{S,T}$, and conversely, swapping the roles of T and Σ.

To see this, let $\phi\colon T \longrightarrow \Sigma$ be a morphism, and let T' be an equivariant subdivision of T such that $\phi\colon T' \longrightarrow \Sigma$ is a folding. Suppose $R \in \mathcal{S}(\Sigma)$.

Applying Lemma 1.2, there is finite subtree, S, of T which contains every edge of T' that gets mapped homeomorphically to one of the edges of R. Suppose that $x, y \in \partial T \equiv \partial\Sigma$, and let α and β be the arcs in T and Σ respectively, connecting x to y. Thus $\beta \subseteq \phi\alpha$. Suppose that $x \approx_{S,T} y$. In other words, $\alpha \cap S$ is either empty or consists of a single vertex. We claim that the same is true of $\beta \cap R$. For any edge of $\beta \cap R$ is the image under ϕ of some edge ϵ of α in T'. By construction, ϵ is also an edge of S in T', giving a contradiction. This shows that $x \approx_{R,\Sigma} y$ as claimed. Swapping the roles of T and Σ, we deduce the cofinality of the direct limit systems as claimed.

Now, suppose that $\mathcal{B}$ and $\mathcal{A}$ are arc systems on T and Σ respectively, giving rise to the same formal arc system. We get identical relations $\approx_{\mathcal{B}} = \approx_{\mathcal{A}}$ on $\partial T = \partial\Sigma$, as defined earlier. Now, it follows that the direct limit systems $(\sim_{S,\mathcal{B}})_{S\in\mathcal{S}(T)}$ and $(\sim_{R,\mathcal{A}})_{R\in\mathcal{S}(\Sigma)}$ are cofinal, and so give rise to the same direct limit, namely $\sim_{\mathcal{B}} = \sim_{\mathcal{A}}$, as claimed earlier.

In particular, we see that $\mathcal{B}$ is indecomposable if and only if $\mathcal{A}$ is. In summary, reintroducing the group action, we have shown:

Lemma 3.5 *Suppose that T and Σ are equivalent cofinite G–trees with finite edge stabilisers. Suppose that $\mathcal{B}$ and $\mathcal{A}$ are arc systems on T and Σ respectively, corresponding to the same formal arc system on $\partial T \equiv \partial\Sigma$. Then, $\mathcal{B}$ is indecomposable if and only if $\mathcal{A}$ is indecomposable.* □

Suppose, now, that G is accessible over finite groups. As discussed in Section 1, we can associate to G a set $\partial_\infty G$, which we can identify with the boundary of any cofinite G–tree with finite edge stabilisers and finite and one-ended vertex stabilisers. We refer to such trees as *complete* G–trees. Any two complete G–trees are equivalent, so by Lemma 3.5, it makes sense to speak about a formal arc system on $\partial_\infty G$ as being indecomposable.

Suppose, now that $H \leq G$ is a two-ended subgroup. We say that H is *elliptic* if it lies inside some one-ended subgroup of G. Thus H is elliptic if and only it is elliptic with respect to some (and hence any) complete G–tree. Otherwise, we say that H is *hyperbolic*. In this case, there is a unique H–invariant unordered pair of points in $\partial_\infty G$ which we denote by ΛH. Thus, ΛH is the pair of endpoints of the axis of H in any complete G–tree. We refer to ΛH as the *limit set* of H. We note that if H' is another hyperbolic two-ended subgroup, and $\Lambda H \cap \Lambda H' \neq \emptyset$, then H and H' are commensurable, and hence lie in the same maximal two-ended subgroup.

Let $\mathcal{H}$ be a finite union of conjugacy classes of hyperbolic two-ended subgroups of G. Recall that $\mathcal{H}$ is "indecomposable" if we cannot write G as a non-trivial amalgamated free product or HNN–extension over a finite group with each element of H conjugate into a vertex group. It is easy to see that this property depends only on the commensurability classes of the elements of $\mathcal{H}$, so we may, if we wish, take all the elements of H to be maximal two-ended subgroups, in which case their limit sets are all disjoint. Note that we get a formal arc system, $\{\Lambda H \mid H \in \mathcal{H}\}$, on $\partial_\infty G$. We claim:

Proposition 3.6 *If the formal arc system $\{\Lambda H \mid H \in \mathcal{H}\}$ is indecomposable, then $\mathcal{H}$ is indecomposable.*

Proof Suppose not. Then there is a non-trivial cofinite G–tree, T, with finite edge stabilisers and with each element of $\mathcal{H}$ elliptic with respect to T. Now, as discussed in Section 1, we can refine the splitting T/G to a complete splitting, giving us a complete G–tree, Σ. We can recover T by collapsing T along a disjoint union of subtrees. Each element of H fixes setwise one of these subtrees.

Now, let $\mathcal{B}$ be the arc system on Σ given by the formal arc system, in other words, the set of axes of elements of $\mathcal{H}$. Thus each axis lies inside one of the collapsing subtrees. In particular, $\Sigma \neq \bigcup \mathcal{B}$, and so $\mathcal{B}$ is decomposable. □

We shall prove a converse to Proposition 3.6 in the case where G is finitely generated. For this we shall need a relative version of Stallings's theorem.

Let G be a finitely generated group, and let X be a Cayley graph of X (or, indeed, any graph on which G acts with finite vertex stabilisers and finite quotient). Given a subset $A \subseteq V(X)$ we write $E_A \subseteq E(X)$ for the set of edges with precisely one endpoint in A. Thus, to say that X has "more than one end" means that we can find an infinite subset, $A \subseteq V(X)$ such that its complement $B = V(X) \setminus A$ is also infinite, and such that $E_A = E_B$ is finite. Thus, Stallings's theorem [27] tells us that in such a case, G splits over a finite group.

Suppose, now that $H \leq G$ is a two ended subgroup, and that $C \subseteq V(X)$ is an H–orbit of vertices (or any H–invariant subset with C/H finite). Now, for all but finitely many G–images, gC, of C, we have either $gC \subseteq A$ or $gC \subseteq B$. For the remainder, we have three possibilities: either $gC \cap A$ is finite or $gC \cap B$ is finite, or else both of these subsets give us a neighbourhood of an end of H. We shall not say more about the last case, since it is precisely the case we wish to rule out. Note that this classification does not depend on the choice of H–orbit, C. A specific relative version of Stallings's theorem says the following:

Lemma 3.7 *Suppose G is a finitely group and $\mathcal{H}$ is a finite union of conjugacy classes of two-ended subgroups. Let X be a Cayley graph of G. Suppose we can find an infinite set, $A \subseteq V(X)$, such that E_A is finite and $B = V(X) \setminus A$ is infinite. Suppose that for any $H \in \mathcal{H}$ either $A \cap C$ or $B \cap C$ is finite for some (hence every) H–orbit of vertices, C. Then, $\mathcal{H}$ is decomposable (ie G splits over a finite group relative to $\mathcal{H}$).* □

In fact, a much stronger result follows immediately from the results of [9]. It may be stated as follows. Suppose G is any finitely generated group, and $A \subseteq G$ is an infinite subset, whose complement $B = G \setminus A$ is also infinite. Suppose that the symmetric difference of A and Ag is finite for all $g \in G$. Suppose that $H_1, \ldots, H_n$ are subgroups such that for all $g \in G$ and all $i \in \{1, \ldots, n\}$ either $gH_i \cap A$ or $gH_i \cap B$ is finite. Then G splits over a finite group relative to $\{H_1, \ldots, H_n\}$. (If fact, it's sufficient to rule out G being a non-finitely generated countable torsion group.)

Alternatively, one can deduce Lemma 3.7, as we have stated it, by applying Stallings's theorem to the double, $D(G, \mathcal{H})$, and using Corollary 2.4. We briefly sketch the argument. We may construct a Cayley graph, Y, for $D(G, \mathcal{H})$ by taking lots of copies of X, and stringing them together in a treelike fashion. Let's focus on a particular copy of X, which we take to be acted upon by G. Now each adjacent copy of X corresponds to an element $H \in \mathcal{H}$, and is connected ours by an H–orbit of edges. We refer to such edges as "amalgamating edges". The amalgamating edges corresponding to H are attached to X by an H–orbit, C_H, of vertices of X. By hypothesis, either $C_H \cap A$ is finite, in which case, we write E_H for the set of amalgamating edges which have an endpoint in $C_H \cap A$, or else, $C_H \cap B$ is finite, in which case, we write E_H for the set of amalgamating edges which have an endpoint in $C_H \cap B$. Now, for all but finitely many H, the set E_H is empty. Thus, the set $E_{\mathcal{H}} = \bigcup_{H \in \mathcal{H}} E_H$ is finite, and so $E_0 = E_A \cup E_{\mathcal{H}} \subseteq E(Y)$ is finite. Now, E_0 separates Y into two infinite components. Thus, by Stallings's theorem, $D(G, \mathcal{H})$ splits over a finite group, and so by Corollary 2.4, $\mathcal{H}$ is decomposable. With the details filled in, this gives another proof of Lemma 3.7.

We are now ready to prove a converse to Proposition 3.6:

Proposition 3.8 *Suppose that G is a finitely generated accessible group. Suppose that $\mathcal{H}$ is a finite union of conjugacy classes of hyperbolic two-ended subgroups. If $\mathcal{H}$ is indecomposable, then the formal arc system, $\{\Lambda H \mid H \in \mathcal{H}\}$, on $\partial_\infty G$, is indecomposable.*

Proof Let T be a complete G–tree, and let $\mathcal{B}$ be the corresponding arc system on T, ie the set of axes of elements of $\mathcal{H}$. Suppose, for contradiction, that $\mathcal{B}$ is decomposable. In other words, we can find $S \in \mathcal{S}(T)$ such that there is more than one $\sim_S$–class. By taking projections of $\sim_S$–classes as disussed in Section 1, we can write $V(S)$ as a disjoint union of non-empty subsets, $V(S) = W_1 \sqcup W_2$ with the property that if $\beta \in \mathcal{B}$, then β meets S, if at all, in compact interval (or point) with either both endpoints in W_1 or both endpoints in W_2. Let $F_i = \pi_S^{-1} W_i$. Thus, $T = S \cup F_1 \cup F_2$, and each component of each F_i is a subtree meeting S in a single point.

Now, let X be a Cayley graph of G. Let $f \colon V(X) \longrightarrow V(T)$ be any G–equivariant map. Let $A_i = f^{-1} F_i \subseteq V(X)$. Thus, $V(X) = A_1 \sqcup A_2$. Moreover, it is easily seen that $E_{A_1} = E_{A_2}$ is finite. (For example, extend f equivariantly to a map $f \colon X \longrightarrow T$ so that each edge of X gets mapped to a compact interval of T. Only finitely many G–orbits of such an interval can contain a given edge of T. Now, the image of an edge of E_{A_1} connects a vertex of F_1 to a vertex of F_2, and hence contains an edge of S. There are only finitely many such edges.)

Finally, suppose that $H \in \mathcal{H}$. Let $\beta \in \mathcal{B}$ be the axis of H. Without loss of generality, we can suppose that both ends of β are contained in F_1. Now suppose that C is any H–orbit of vertices of X. Then $f(C)$ remains within a bounded distance of β, from which we see easily that $f(C) \cap F_2$ is finite. Thus, $C \cap A_2$ is finite.

We have verified the hypotheses of Lemma 3.7, and so $\mathcal{H}$ is decomposable, contrary to our hypotheses. ◻

Note that Propositions 3.6 and 3.8 apply, in particular, to any finitely presented group, and even more specifically, to any hyperbolic group, G. In the latter case, $\partial_\infty G$ can be identified as a subset of the Gromov boundary, ∂G, as discussed in Section 2. If $H \leq G$ is a hyperbolic two-ended subgroup, then $\Lambda H \subseteq \partial G$ is the limit set of H by the standard definition. This ties in with the discussion of equivalence relations on ∂G in the introduction, and will be elaborated on in Section 5.

4 Quasiconvex splittings of hyperbolic groups

For most of the rest of this paper, we shall be confining our attention to hyperbolic groups. We shall consider how some of the general constructions of Sections 1–3 relate to the topology of the boundary in this case. Before we embark on this, we review some general facts about quasiconvex splittings of hyperbolic groups (ie splittings over quasiconvex subgroups). This elaborates on the account given in [5].

Throughout the rest of this paper, we shall use the notation $\mathrm{fr}\, A$ to denote the topological boundary (or "frontier") of a subset, A, of a larger topological space. We reserve the symbol "∂" for ideal boundaries.

Let Γ be any hyperbolic group. Let X be any locally finite connected graph on which Γ acts freely and cocompactly (for example a Cayley graph of Γ). We put a path metric, d, on X by assigning a positive length to each edge in a Γ–invariant fashion. Let $\partial\Gamma \equiv \partial X$ be the boundary of Γ. We may put a metric on $\partial\Gamma$ as described in [14]. This has the property that given a basepoint, $a \in V(X)$, there are constants, $A, B > 0$ and $\lambda \in (0,\infty)$ such that if $x, y \in \partial X$, then $A\lambda^\delta \leq \rho(x,y) \leq B\lambda^\delta$, where δ is the distance from a to some biinfinite geodesic connecting x to y. Although all the arguments of this paper can be expressed in purely topological terms, it will be convenient to have recourse to this metric.

Note that if $G \leq \Gamma$ is quasiconvex, then it is intrinsically hyperbolic, and we may identify its boundary, ∂G, with its limit set $\Lambda G \subseteq \partial\Gamma$. Note that G acts properly discontinuously on $\partial\Gamma \setminus \Lambda G$. The setwise stabiliser of ΛG in Γ is precisely the commensurator, $\mathrm{Comm}(G)$, of G in Γ (ie the set of all $g \in \Gamma$ such that $G \cap gGg^{-1}$ has finite index in G). In this case, G has finite index in $\mathrm{Comm}(G)$. In fact, $\mathrm{Comm}(G)$ is the unique maximal subgroup of Γ which contains G as finite index subgroup. We say that G is *full* if $G = \mathrm{Comm}(G)$.

We shall use the following notation. If $f\colon Z \longrightarrow [0,\infty)$ is a function from some set Z to the nonnegative reals, we write "$f(z) \to 0$ for $z \in Z$" to mean that $\{z \in Z \mid f(z) \geq \epsilon\}$ is finite for all $\epsilon > 0$. We similarly define "$f(z) \to \infty$ for $z \in Z$".

Lemma 4.1 *If $G \leq \Gamma$ is quasiconvex and $x \in \partial\Gamma$, then $\rho(gx, \Lambda G) \to 0$ for $g \in G$.*

Proof Since G acts properly discontinuously on $\partial\Gamma \setminus \Lambda G$, there can be no accumulation point of the G–orbit of x in this set. ◻

The following is also standard:

Lemma 4.2 *If $G \leq \Gamma$ is quasiconvex, then $\mathrm{diam}(\Lambda H) \to 0$ as H ranges over conjugates of G.* ◻

We want to go on to consider splittings of Γ. For this, we shall want to introduce some further notation regarding trees.

By a "directed edge" we mean an edge together with an orientation. We write $\vec{E}(T)$ for the set of directed edges. We shall always use the convention that $e \in E(T)$ represents the undirected edge underlying the directed edge $\vec{e} \in \vec{E}(T)$. We write $\mathrm{head}(\vec{e})$ and $\mathrm{tail}(\vec{e})$ respectively for the head and tail of $\vec{e}$. We use $-\vec{e}$ for the same edge oriented in the opposite direction, ie $\mathrm{head}(-\vec{e}) = \mathrm{tail}(\vec{e})$ and $\mathrm{tail}(-\vec{e}) = \mathrm{head}(\vec{e})$. If $\vec{e} \in \vec{E}(T)$ and $v \in V(T)$, we say that $\vec{e}$ "points towards" v if $\mathrm{dist}(v, \mathrm{tail}(\vec{e})) = \mathrm{dist}(v, \mathrm{head}(\vec{e})) + 1$.

If $v \in V(T)$, let $\Delta(v) \subseteq E(T)$ be the set of edges incident on v, and let $\vec{\Delta}(v) = \{\vec{e} \in \vec{E}(T) \mid \mathrm{head}(\vec{e}) = v\}$. Thus, the degree of v is $\mathrm{card}(\Delta(v)) = \mathrm{card}(\vec{\Delta}(v))$.

Given $\vec{e} \in \vec{E}(T)$, we write $\Phi(\vec{e}) = \Phi_T(\vec{e})$ for the connected component of T minus the interior of e which contains $\mathrm{tail}(\vec{e})$. Thus, $V(\Phi(\vec{e}))$ is the set of vertices, v, of T such that $\vec{e}$ points away from v.

Given $v \in V(T)$, we shall write $\vec{\Omega}(v) \subseteq \vec{E}(T)$ for the set of directed edges which point towards v. Thus, for each edge $e \in E(T)$, precisely one of the pair $\{\vec{e}, -\vec{e}\}$ lies in $\vec{\Omega}(v)$. Note that $\vec{e} \in \vec{\Omega}(v)$ if and only if $v \notin \Phi(\vec{e})$. Clearly $\vec{\Delta}(v) \subseteq \vec{\Omega}(v)$.

We now return to our hyperbolic group, Γ. Suppose that Γ acts without edge inversions on a simplicial tree, Σ, with Σ/Γ finite. We suppose that this action is minimal. Given $v \in V(\Sigma)$ and $e \in E(\Sigma)$, write $\Gamma(v)$ and $\Gamma(e)$ respectively for the corresponding vertex and edge stabilisers. Note that $\Gamma(v)$ is finite if and only if v has finite degree in Σ and finite incident edge stabilisers. If $v, w \in V(\Sigma)$ are the endpoints of an edge $e \in E(\Sigma)$, then $\Gamma(e) = \Gamma(v) \cap \Gamma(w)$.

As in [5], we may construct a Γ–equivariant map $\phi\colon X \longrightarrow \Sigma$ such that each edge of X either gets collapsed onto a vertex of Σ or mapped homeomorphically onto a closed arc in Σ. (Note that, after subdividing X if necessary, we can assume that, in the latter case, this closed arc is an edge of Σ.) Since the action of Γ is minimal, ϕ is surjective.

A proof of the following result can be found in [5], though it appears to be "folklore".

Proposition 4.3 *If $\Gamma(e)$ is quasiconvex for each $e \in E(\Sigma)$, then $\Gamma(v)$ is quasiconvex for each $v \in V(\Sigma)$.* □

We refer to such a splitting as a *quasiconvex splitting*.

We note that if a vertex group, $\Gamma(v)$, of a quasiconvex splitting has the property that all incident edge groups are of infinite index in $\Gamma(v)$, then $\Gamma(v)$ must be full in the sense described above. In other words, $\Gamma(v)$ is the setwise stabiliser of $\Lambda\Gamma(v)$. This will be the case in most situations of interest (in particular where all edge groups are finite or two-ended, but $\Gamma(v)$ is not).

Note that, if $v, w \in V(\Sigma)$, then $\Gamma(v) \cap \Gamma(w)$ is quasiconvex (since the intersection of any two quasiconvex subgroups is quasiconvex [26]). We see that $\Lambda\Gamma(v) \cap \Lambda\Gamma(w) = \Lambda(\Gamma(v) \cap \Gamma(w))$. In particular, if v, w are the endpoints of an edge $e \in E(\Sigma)$, then $\Lambda\Gamma(v) \cap \Lambda\Gamma(w) = \Lambda\Gamma(e)$.

As described in [5], there is a natural Γ–invariant partition of $\partial\Gamma$ as $\partial\Gamma = \partial_0\Gamma \sqcup \partial_\infty\Gamma$, where $\partial_0\Gamma = \bigcup_{v \in V(\Sigma)} \Lambda\Gamma(v)$, and $\partial_\infty\Gamma$ is naturally identified with $\partial\Sigma$. Note that $\partial_\infty\Gamma$ is dense in $\partial\Gamma$, provided that Σ is non-trivial. (In the case where the edge stabilisers are all finite, this agrees with the notion introduced for accessible groups in Section 2.)

Given $\vec{e} \in \vec{E}(\Sigma)$, we write

$$\Psi(\vec{e}) = \partial\Phi(\vec{e}) \cup \bigcup_{v \in V(\Phi(\vec{e}))} \Lambda\Gamma(v).$$

It's not hard to see that $\Psi(\vec{e})$ is a closed $\Gamma(e)$–invariant subset of $\partial\Gamma$. Moreover, $\Psi(\vec{e}) \cup \Psi(-\vec{e}) = \partial\Gamma$ and $\Psi(\vec{e}) \cap \Psi(-\vec{e}) = \mathrm{fr}\,\Psi(\vec{e}) = \Lambda\Gamma(e)$.

Now, $V(\Sigma) = \{v\} \sqcup \bigsqcup_{\vec{e} \in \vec{\Delta}(v)} V(\Phi(\vec{e}))$ and $\partial\Sigma = \bigsqcup_{\vec{e} \in \vec{\Delta}(v)} \partial\Phi(\vec{e})$. It follows that:

Lemma 4.4 $\partial\Gamma = \Lambda\Gamma(v) \cup \bigcup_{\vec{e} \in \vec{\Delta}(v)} \Psi(\vec{e})$. □

Moreover, for each $\vec{e} \in \vec{\Delta}(v)$, we have $\Lambda\Gamma(v) \cap \Psi(\vec{e}) = \Lambda\Gamma(e)$.

The above assertions become more transparent, given the following alternative description of $\Psi(\vec{e})$.

Let $m(e)$ be the midpoint of the edge e, and let $I(\vec{e})$ be the closed interval in Σ consisting of the segment of e lying between $m(e)$ and $\mathrm{tail}(\vec{e})$. Let $Q(e) = \phi^{-1}(m(e)) \subseteq X$ and $R(\vec{e}) = \phi^{-1}(\Phi(\vec{e}) \cup I(\vec{e})) \subseteq X$, where $\phi\colon X \longrightarrow \Sigma$ is the map described above. Note that $Q(e) = \mathrm{fr}\,R(\vec{e}) = R(\vec{e}) \cap R(-\vec{e})$. By the arguments given in [5], we see easily that $Q(e)$ and $R(\vec{e})$ are quasiconvex subsets of X. Moreover, $\Psi(\vec{e}) = \partial R(\vec{e})$.

Note that the collection $\{Q(e) \mid e \in E(\Sigma)\}$ is locally finite in X. It follows that, for any fixed $a \in X$, we have $d(a, Q(e)) \rightarrow \infty$ for $e \in E(\Sigma)$.

Now, fix some vertex, $v \in V(\Sigma)$. Recall that $\vec{\Omega}(v)$ is defined to be the set of all directed edges pointing towards v. Choose any $b \in \phi^{-1}(v) \subseteq X$. Now, if $\vec{e} \in \vec{\Omega}(v)$, we have $v \notin \Phi(\vec{e}) \cup I(\vec{e})$, and so $b \notin R(\vec{e})$. Since $Q(e) = \mathrm{fr}\,R(\vec{e})$, we have $d(b, R(\vec{e})) = d(b, Q(e))$. It follows that $d(b, R(\vec{e})) \rightarrow \infty$ for $\vec{e} \in \vec{\Omega}(v)$.

In fact, we see that $d(a, R(\vec{e})) \to \infty$ given any fixed basepoint, $a \in X$. Now, there are only finitely many Γ–orbits of directed edges, and so the sets $R(\vec{e})$ are uniformly quasiconvex. From the definition of the metric ρ on $\partial\Gamma$, it follows easily that $\operatorname{diam}(\Psi(\vec{e})) \to 0$, where diam denotes diameter with respect to ρ. In summary, we have shown:

Lemma 4.5 *For any* $v \in V(\Sigma)$, $\operatorname{diam}(\Psi(\vec{e})) \to 0$ *for* $\vec{e} \in \vec{\Omega}(v)$. □

We now add the hypothesis that $\Gamma(e)$ is infinite for all $e \in E(\Sigma)$.

Suppose $v \in V(\Sigma)$ and suppose K is any closed subset of $\Lambda\Gamma(v)$. Let $\vec{\Delta}_K(v) = \{\vec{e} \in \vec{\Delta}(v) \mid \Lambda\Gamma(e) \subseteq K\}$, and let $\Upsilon(v, K) = K \cup \bigcup_{\vec{e} \in \vec{\Delta}_K(v)} \Psi(\vec{e}) \subseteq \partial\Gamma$.

Lemma 4.6 *The set* $\Upsilon(v, K)$ *is closed in* $\partial\Gamma$.

Proof Suppose $x \notin \Upsilon(v, K)$. In particular, $x \notin K$, so $\epsilon = \rho(x, K) > 0$. Now, if $\vec{e} \in \vec{\Delta}_K(v)$ and $\rho(x, \Psi(\vec{e})) < \epsilon/2$, then $\operatorname{diam}(\Psi(\vec{e})) > \epsilon/2$ (since $K \cap \Psi(\vec{e}) \supseteq \Lambda\Gamma(e)$, which, by the hypothesis on edge stabilisers, is non-empty). By Lemma 4.5, this occurs for only finitely many such $\vec{e}$. Since each $\Psi(\vec{e})$ is closed, it follows that $\rho(x, \Upsilon(v, K))$ is attained, and hence positive. In other words, $x \notin \Upsilon(v, K)$ implies $\rho(x, \Upsilon(v, K)) > 0$. This shows that $\Upsilon(v, K)$ is closed. □

5 Quotients

In this section, we aim to consider quotients of boundaries of hyperbolic groups, and to relate this to indecomposability, thereby generalising some of the results of [23].

First, we recall a few elementary facts from point-set topology [17,16]. Let M be a hausdorff topological space. A subset of M is *clopen* if it is both open and closed. We may define an equivalence relation on M by deeming two points to be related if every clopen set containing one must also contain the other. The equivalence classes are called *quasicomponents*. A *component* of M is a maximal connected subset. Components and quasicomponents are always closed. Every component is contained in a quasicomponent, but not conversely in general. However, if M is compact, these notions coincide. Thus, if K and K' are distinct components of a compact hausdorff space, M, then there is a clopen subset of M containing K, but not meeting K'.

Suppose that M is a compact hausdorff space, and that $\approx$ is an equivalence relation on M. If the relation $\approx$ is closed (as a subset of $M \times M$), then the quotient space, $M/\!\approx$ is hausdorff.

The compact spaces of interest to us here will be the boundaries of hyperbolic groups. Suppose that G is a hyperbolic group, and that ∂G is its boundary. Now, any two ended subgroup, H, of G is necessarily quasiconvex, so its limit set, $\Lambda H \subseteq \partial G$, consists of pair of points. If H' is another two-ended subgroup, and $\Lambda H \cap \Lambda H' \neq \emptyset$, then H and H' are commensurable, and so lie in a common maximal two-ended subgroup. In particular, $\Lambda H = \Lambda H'$ (cf the discussion of accessible groups in Section 3).

Suppose that $\mathcal{H}$ is a union of finitely many conjugacy classes of two-ended subgroups of G. Let $\approx_{\mathcal{H}}$ be the equivalence relation defined on ∂G defined by $x \approx_{\mathcal{H}} y$ if and only if either $x = y$ or there exists $H \in \mathcal{H}$ such that $\Lambda H = \{x, y\}$. Now, it's a simple consequence of Lemma 4.2 that the relation $\approx_{\mathcal{H}}$ is closed. We write $M(G, \mathcal{H})$ for the quotient space $\partial G/\!\approx_{\mathcal{H}}$. Thus:

Lemma 5.1 *$M(G, \mathcal{H})$ is compact hausdorff.* □

We aim to describe when $M(G, \mathcal{H})$ is connected. Clearly, if G is one-ended so that ∂G is connected, this is necessarily the case. We can thus restrict attention to the case when G is infinite-ended.

Let T be a complete G–tree. As in Section 3, we can define $\partial_\infty G$ as ∂T. This also agrees with the notation introduced in Section 4, thinking of T as a quasiconvex splitting of G. In particular, we can identify $\partial_\infty G$ as a subset of ∂G. This set $\partial_0 G = \partial G \setminus \partial_\infty G$ is a disjoint union of the boundaries of the infinite vertex stabilisers of T, ie the maximal one-ended subgroups. In other words, the components of $\partial_0 G$ are precisely the boundaries of the maximal one-ended subgroups of G.

Let $\mathcal{H}$ be a set of two-ended subgroups as above. The subset, $\mathcal{H}_0$, of $\mathcal{H}$ consisting of those subgroups in $\mathcal{H}$ which are hyperbolic (ie with both limit points in $\partial_\infty G$), defines a formal arc system on $\partial_\infty G$. We aim to show that $M(G, \mathcal{H})$ is connected if and only if this arc system is indecomposable. This, in turn, we know to be equivalent to asserting that $\mathcal{H}_0$ is irreducible.

In fact, it's easy to see that the elliptic elements of $\mathcal{H}$ have no bearing on the connectivity or otherwise of $M(G, \mathcal{H})$. For this reason, we may as well suppose, for simplicity, that $\mathcal{H}$ consists entirely of hyperbolic two-ended subgroups. We therefore aim to show:

Theorem 5.2 *Let G be an infinite-ended hyperbolic group, and let $\mathcal{H}$ be a union of finitely many conjugacy classes of hyperbolic two-ended subgroups. Then, the quotient space $M(G,\mathcal{H})$ is connected if and only if $\mathcal{H}$ is indecomposable.*

First, we set about proving the "only if" bit. Let T be a complete G–tree. Thus, $\partial_\infty G$ is identified with ∂T, and $\mathcal{H}$ determines an arc system, $\mathcal{B}$, on T. We know (Propositions 3.6 and 3.8) that the indecomposability of $\mathcal{H}$ is equivalent to the indecomposability of $\mathcal{B}$.

We shall say that a subgraph, F, of T is *finitely separated* if there are only finitely many edges of T with precisely one endpoint in F. Now, it's not hard to see that F is finitely separated if and only if it's a finite union of finite intersections of subtrees of the form $\Phi(\vec{e})$ for $\vec{e} \in \vec{E}(T)$ (recalling the notation of Section 4).

Now, given a subgraph, $F \subseteq T$, we write

$$A(F) = \partial F \cup \bigcup_{v \in V(F)} \Lambda G(v)$$

(so that $A(T) = \partial G$). If F is finitely separated, then $A(F)$ is a finite union of finite intersections of sets of the form $\Psi(\vec{e})$, which are each closed by the remarks of Section 4. We conclude:

Lemma 5.3 *If $F \subseteq T$ is a finitely separated subgraph, then $A(F)$ is closed in ∂G.* □

We can now prove:

Lemma 5.4 *If $M(G,\mathcal{H})$ is connected, then the arc system $\mathcal{B}$ is indecomposable.*

Proof Suppose, to the contrary, that $\mathcal{B}$ is decomposable. Then, exactly as in the proof of Proposition 3.8, we can find two disjoint finitely separated subgraphs, F_1 and F_2 of T with $V(T) = V(F_1) \sqcup V(F_2)$ and $\partial T = \partial F_1 \sqcup \partial F_2$, and such that for each $\beta \in \mathcal{B}$, either $\partial\beta \subseteq \partial F_1$ or $\partial\beta \subseteq \partial F_2$. We see that $\partial G = A(F_1) \sqcup A(F_2)$.

Let $q\colon \partial G \longrightarrow \partial G/{\approx_{\mathcal{H}}} = M(G,\mathcal{H})$ be the quotient map. Now, from the construction, we see that if $x \approx_{\mathcal{H}} y$ then either $x, y \in \partial F_1 \subseteq A(F_1)$ or $x, y \in \partial F_2 \subseteq A(F_2)$. We therefore get that $M(G,\mathcal{H}) = q(A(F_1)) \sqcup q(A(F_2))$. But applying Lemma 5.3, the sets $q(A(F_i))$ are both closed in $M(G,\mathcal{H})$, contrary to the assumption that $M(G,\mathcal{H})$ is connected. □

Lemma 5.5 *If $\mathcal{H}$ is indecomposable, then $M(G,\mathcal{H})$ is connected.*

Proof Suppose, for contradiction, that we can write $M(G,\mathcal{H})$ as the disjoint union of two non-empty closed sets, K_1 and K_2. Let $L_i \subseteq \partial G$ be the preimage of K_i under the quotient map $\partial G \longrightarrow M(G,\mathcal{H})$. Thus, $\partial G = L_1 \sqcup L_2$. Let X be a Cayley graph of G. Now, we can give $X \cup \partial G$ a natural G–invariant topology as a compact metrisable space. Since $X \cup \partial G$ is normal, we can find disjoint open subsets, $U_i \subseteq X \cup \partial G$ with $L_i \subseteq U_i$. Now, $(X \cup \partial G) \setminus (U_1 \cup U_2) \subseteq X$ is compact, and so lies inside a finite subgraph, Y, of X. Let $A = U_1 \cap V(X)$ and let $B = V(X) \setminus A$. We need to verify that A satisfies the hypotheses of Lemma 3.7.

Note that $A \cup L_1$ and $B \cap L_2$ are both closed in $X \cup \partial G$. We see that A and B are both infinite. Recall that $E_A = E_B$ is the set of edges of X which have one endpoint in A and the other in B. Now, $E_A \subseteq E(Y)$, and so E_A is finite.

Finally, suppose that $H \in \mathcal{H}$ and that $C \subseteq V(X)$ is an H–orbit of vertices of X. Now, $C \cup \partial H$ is closed in $X \cup \partial G$. Without loss of generality we can suppose that $\Lambda H \subseteq L_1$. Since $B \cup L_2 \subseteq X \cup \partial G$ is closed, we see that $C \cap B$ is finite.

We have verified the hypotheses of Lemma 3.7, and so we arrive at the contradiction that $\mathcal{H}$ is decomposable. □

This concludes the proof of Theorem 5.2.

6 Splittings of hyperbolic groups over finite and two-ended subgroups

Suppose that a hyperbolic group splits over a collection of two-ended subgroups. We may in turn try to split each of the vertex groups over finite groups, thus giving us a two-step series of splittings. We want to study how the combinatorics of such splittings are reflected in the topology of the boundary. The combinatorics can be described in terms of the trees associated to each step of the splitting, together with arc systems on the trees of the second step which arise from the incident edge groups of the first step.

Suppose that Γ is a hyperbolic group, and that Σ is a cofinite Γ–tree with two-ended edge stabilisers. Note that this is necessarily a quasiconvex splitting (as described in Section 4), since a two-ended subgroup of a hyperbolic group is necessarily quasiconvex (see, for example, [14]). We shall fix some vertex,

$\omega \in V(\Sigma)$, and write $G = \Gamma(\omega)$. We suppose that G is not two-ended. By Proposition 4.3, G is quasiconvex, and hence intrinsically hyperbolic. We shall, in turn, want to consider splittings of G over finite groups, so to avoid any confusion later on, we shall alter our notation, so that it is specific to this situation.

Let Ξ be an indexing set which is in bijective correspondence with the set, $\vec{\Delta}(\omega)$, of directed edges of Σ with heads at ω. Thus, G permutes the elements of Ξ. There are finitely many G–orbits (since $\vec{\Delta}(\omega)/\Gamma(\omega)$ is finite). Given $\xi \in \Xi$, we write $H(\xi)$ for the stabiliser, in G, of ξ. Thus, if $\vec{e} \in \vec{\Delta}(\omega)$ is the edge corresponding to ξ, then $H(\xi) = \Gamma(e)$. In particular, $H(\xi)$ is two-ended. Let $J(\xi) = \Psi(\vec{e})$. Thus, $J(\xi)$ is a closed $H(\xi)$–invariant subset of ΛG. Moreover, $\operatorname{fr} J(\xi) = J(\xi) \cap \Lambda G = \Lambda H(\xi)$ consists of a pair of distinct points.

In this notation, we have:

Lemma 6.1 $\partial\Gamma = \Lambda G \cup \bigcup_{\xi\in\Xi} J(\xi)$. □

Lemma 6.2 $\operatorname{diam} J(\xi) \to 0$ *for* $\xi \in \Xi$. □

Here, Lemma 6.1 is a rewriting of Lemma 4.4, and Lemma 6.2 is a restriction of Lemma 4.5.

If $K \subseteq \Lambda G$ is closed, we write $\Xi(K) = \{\xi \in \Xi \mid \operatorname{fr} J(\xi) \subseteq K\}$, and write $\Upsilon(K) = K \cup \bigcup_{\xi\in\Xi(K)} J(\xi)$. Thus, Lemma 4.6 says that:

Lemma 6.3 $\Upsilon(K)$ *is a closed subset of* $\partial\Gamma$. □

These observations tell us all we need to know about the groups $H(\xi)$ and sets $J(\xi)$ for the rest of this section. Thus, for the moment, we can forget how they were constructed.

Now, G is intrinsically hyperbolic, with ∂G identified with ΛG. We write $\Lambda G = \Lambda_0 G \sqcup \Lambda_\infty G$, corresponding to the partition $\partial G = \partial_0 G \sqcup \partial_\infty G$, as described in Section 5. Let T be a complete G–tree, so that $\partial T \equiv \Lambda_\infty G$. We write $V_{\text{fin}}(T)$ and $V_{\text{inf}}(T)$ respectively, for the sets of vertices of T of finite and infinite degree. Thus, $\Lambda_0 G = \bigsqcup_{v\in V(T)} \Lambda G(v)$. We note that if T is non-trivial (ie not a point), then $\Lambda_\infty G$ is dense in ΛG.

Given $\xi \in \Xi$, the subgroup $H(\xi)$ is two-ended. It is either elliptic or hyperbolic with respect to the G–tree T. We write Ξ_{ell} and Ξ_{hyp}, respectively, for the sets of $\xi \in \Xi$ such that $H(\xi)$ is elliptic or hyperbolic.

If $\xi \in \Xi_{\rm ell}$, then $H(\xi)$ fixes a unique vertex $v(\xi) \in V_{\rm inf}(T)$, so that $H(\xi) \subseteq G(v(\xi))$ and $\operatorname{fr} J(\xi) \subseteq \Lambda G(v(\xi))$. Given $v \in V(T)$, we write $\Xi_{\rm ell}(v) = \{\xi \in \Xi \mid H(\xi) \subseteq G(v)\}$. Thus $\Xi_{\rm ell}(v) \subseteq \Xi_{\rm ell}$, and $\Xi_{\rm ell}(v) = \emptyset$ for all $v \in V_{\rm fin}(T)$. In fact, $\Xi_{\rm ell} = \bigsqcup_{v \in V(T)} \Xi_{\rm ell}(v)$.

Given $\xi \in \Xi_{\rm hyp}$, we write $\beta(\xi) \subseteq T$ for the unique biinfinite arc in T preserved setwise by $H(\xi)$. Note that, under the identification of ∂T and $\Lambda_0 G$, we have $\partial\beta(\xi) = \Lambda H(\xi)$.

Suppose that $F \subseteq T$ is a finitely separated subgraph. Recall from Section 5 that $A(F)$ is defined as $A(F) = \partial F \cup \bigcup_{v \in V(F)} \Lambda G(v)$. Thus, by Lemma 5.3, $A(F)$ is closed in ΛG and hence in $\partial\Gamma$. We abbreviate $A(\Phi(\vec{e}))$ to $A(\vec{e})$. (So that $A(\vec{e})$ has the form $\Psi(\vec{e})$ in the notation of Section 4.)

If $F \subseteq T$ is finitely separated, we write $\Xi(F) = \Xi(A(F)) = \{\xi \in \Xi \mid \operatorname{fr} J(\xi) \subseteq A(F)\}$. Thus, $\xi \in \Xi_{\rm ell} \cap \Xi(F)$ if and only if $v(\xi) \in V(F)$. Also, $\xi \in \Xi_{\rm hyp} \cap \Xi(F)$ if and only if $\partial\beta(\xi) \subseteq \partial F$.

If $\vec{e} \in \vec{E}(T)$, we shall abbreviate $\Xi(\vec{e}) = \Xi(\Phi(\vec{e}))$. Thus, $\xi \in \Xi(\vec{e})$ if and only if $\vec{e}$ points away from $v(\xi)$ or $\beta(\xi)$. Suppose $v_0 \in V(T)$. Let $\alpha \subseteq T$ be the arc joining v_0 to $v(\xi)$ or to the nearest point of $\beta(\xi)$. Then, $\{\vec{e} \in \vec{\Omega}(v_0) \mid \xi \in \Xi(\vec{e})\}$ consists of the directed edges in α which point towards v_0. In particular, this set is finite. Indeed, if $\Xi_0 \subseteq \Xi$ is finite, we see that $\{\vec{e} \in \vec{\Omega}(v_0) \mid \Xi_0 \cap \Xi(\vec{e}) \neq \emptyset\}$ is finite.

If $F \subseteq T$ is a finitely separated subgraph, we write

$$B(F) = A(F) \cup \bigcup_{\xi \in \Xi(F)} J(\xi).$$

In other words, $B(F) = \Upsilon(A(F))$, as defined earlier in this section. Thus, by Lemma 6.3, we have:

Lemma 6.4 *The set $B(F) \subseteq \partial\Gamma$ is closed, for any finitely separated subgraph, F, of T.* □

If $\vec{e} \in \vec{E}(T)$, we abbreviate $B(\vec{e}) = B(\Phi(\vec{e}))$.

Lemma 6.5 *If $v_0 \in V(T)$, then $\operatorname{diam} B(\vec{e}) \to 0$ for $\vec{e} \in \vec{\Omega}(v_0)$.*

Proof Suppose $\delta > 0$. By Lemma 6.2, there is a finite subset $\Xi_0 \subseteq \Xi$ such that if $\xi \in \Xi \setminus \Xi_0$ then $\operatorname{diam} J(\xi) \leq \delta/3$. Let $\vec{\Omega}_0 = \{\vec{e} \in \vec{\Omega}(v_0) \mid \Xi_0 \cap \Xi(\vec{e}) \neq \emptyset\}$. As observed above, $\vec{\Omega}_0$ is finite. Let $\vec{\Omega}_1 = \{\vec{e} \in \vec{\Omega}(v_0) \mid \operatorname{diam} A(\vec{e}) \geq \delta/3\}$. By Lemma 4.5, $\vec{\Omega}_1$ is also finite.

Suppose $\vec{e} \in \vec{\Omega}(v_0) \setminus (\vec{\Omega}_0 \cup \vec{\Omega}_1)$. Suppose $x \in B(\vec{e})$. If $x \notin A(\vec{e})$, then $x \in J(\xi)$ for some $\xi \in \Xi(\vec{e})$. Since $\vec{e} \notin \vec{\Omega}_0$, $\Xi_0 \cap \Xi(\vec{e}) = \emptyset$, so $\xi \notin \Xi_0$. Therefore, $\operatorname{diam} J(\xi) \leq \delta/3$. Now, $\operatorname{fr} J(\xi) \subseteq A(\vec{e})$, and so $\rho(x, A(\vec{e})) \leq \delta/3$. This shows that $B(\vec{e})$ lies in a $(\delta/3)$–neighbourhood of $A(\vec{e})$. Now, since $\vec{e} \notin \vec{\Omega}_1$, $\operatorname{diam} A(\vec{e}) < \delta/3$ and so $\operatorname{diam} B(\epsilon) < \delta$. □

Recall, from Section 3, that if $S \subseteq T$ is a subtree, then there is a natural projection $\pi_S \colon T \cup \partial T \longrightarrow S \cup \partial S$. If $v \in V(S)$, we write $F(S,v)$ for the subtree $T \cap \pi_S^{-1} v$. If $R \subseteq S$ is a subtree, then we see that $F(S,v) \subseteq F(R, \pi_R v)$. Recall that $\vec{\Delta}(S) = \{\vec{e} \in \vec{E}(T) \mid \operatorname{head}(\vec{e}) \in S, \operatorname{tail}(\vec{e}) \notin S\}$. If $v \in V(S)$, set $\vec{\Delta}(S,v) = \vec{\Delta}(S) \cap \vec{\Delta}(v)$. We write $\vec{\Omega}(S)$ for the set of all directed edges pointing towards S, ie $\vec{\Omega}(S) = \bigcap_{v \in V(S)} \vec{\Omega}(v)$. Clearly, $\vec{\Delta}(S) \subseteq \vec{\Omega}(S)$. Also if $R \subseteq S$ is a subtree, then $\vec{\Omega}(S) \subseteq \vec{\Omega}(R)$. If $v \in V(T) \setminus V(R)$, let $\vec{e}(R,v)$ be the directed edge with head at $\pi_R v$ which lies in the arc joining v to $\pi_R v$. In other words, $\vec{e}(R,v)$ is the unique edge in $\vec{\Delta}(R)$ such that $v \in \Phi(\vec{e}(R,v))$. Note that, if $v \in V(S) \setminus V(R)$, then $F(S,v) \subseteq \Phi(\vec{e}(R,v))$.

Let $\mathcal{T}$ be the set of all finite subtrees of T. Given $\delta > 0$, let

$$\mathcal{T}_1(\delta) = \{S \in \mathcal{T} \mid (\forall \vec{e} \in \vec{\Delta}(S))(\operatorname{diam} B(\vec{e}) < \delta)\}$$
$$\mathcal{T}_2(\delta) = \{S \in \mathcal{T} \mid (\forall v \in V(S) \cap V_{\mathrm{fin}}(T))(\operatorname{diam} B(F(S,v)) < \delta)\}$$
$$\mathcal{T}_3(\delta) = \{S \in \mathcal{T} \mid (\forall v \in V(S) \cap V_{\mathrm{inf}}(T))(\forall \vec{e} \in \vec{\Delta}(S,v))(\rho(\Lambda G(v), B(\vec{e})) < \delta)\}.$$

Let $\mathcal{T}(\delta) = \mathcal{T}_1(\delta) \cap \mathcal{T}_2(\delta) \cap \mathcal{T}_3(\delta)$.

It is really the collection $\mathcal{T}(\delta)$ in which we shall ultimately be interested. It can be described a little more directly as follows. A finite tree, S, lies in $\mathcal{T}(\delta)$ if and only if for each $v \in V(S)$, we have either $v \in V_{\mathrm{fin}}(T)$ and $\operatorname{diam} B(F(S,v)) < \delta$ or else $v \in V_{\mathrm{inf}}(T)$ and for all $\vec{e} \in \vec{\Delta}(S,v)$ we have $\operatorname{diam} B(\vec{e}) < \delta$ and $\rho(\Lambda G(v), B(\vec{e})) < \delta$. It is this formulation we shall use in applications.

Note that if $R \in \mathcal{T}_1(\delta)$, then, in fact, $\operatorname{diam} B(\vec{e}) < \delta$ for all $\vec{e} \in \vec{\Omega}(R)$. We see that if $R \in \mathcal{T}_1(\delta)$, $S \in \mathcal{T}$ and $R \subseteq S$, then $S \in \mathcal{T}_1(\delta)$. More to the point, we have:

Lemma 6.6 *If $R \in \mathcal{T}(\delta)$, $S \in \mathcal{T}$ and $R \subseteq S$, then $S \in \mathcal{T}(\delta)$.*

Proof As observed above, $S \in \mathcal{T}_1(\delta)$.

Suppose that $v \in V(S) \cap V_{\rm fin}(T)$. If $v \in V(R)$, then $F(S,v) \subseteq F(R,v)$, and so $B(F(S,v)) \subseteq B(F(R,v))$. Therefore, $\operatorname{diam} B(F(S,v)) \leq \operatorname{diam} B(F(R,v)) < \delta$, since $R \in \mathcal{T}_2(\delta)$. On the other hand, if $v \notin V(R)$, then $F(S,v) \subseteq \Phi(\vec{e}(R,v))$, so $\operatorname{diam} B(F(S,v)) \leq \operatorname{diam} B(\vec{e}(R,v)) < \delta$, since $R \in \mathcal{T}_1(\delta)$. This shows that $S \in \mathcal{T}_2(\delta)$.

Finally, suppose $v \in V(S) \cap V_{\rm inf}(T)$ and $\vec{e} \in \vec{\Delta}(S,v)$. If $v \in V(R)$, then $\vec{e} \in \vec{\Delta}(R,v)$, so $\rho(\Lambda G(v), B(\vec{e}))$, since $R \in \mathcal{T}_3(\delta)$. On the other hand, if $v \notin V(R)$, then $\{v\} \cup \Phi(\vec{e}) \subseteq F(R, \vec{e}(R,v))$, and so $\Lambda G(v) \cup B(\vec{e}) \subseteq B(F(R,\vec{e}(R,v)))$. But $\operatorname{diam} B(F(R,\vec{e}(R,v))) < \delta$, since $R \in \mathcal{T}_1(\delta)$. In particular, $\rho(\Lambda G(v), B(\vec{e})) < \delta$. This shows that $S \in \mathcal{T}_3(\delta)$. □

Lemma 6.7 $\mathcal{T}(\delta) \neq \emptyset$.

Proof Using Lemma 6.5, we can certainly find some $R \in \mathcal{T}_1(\delta)$. We form another finite tree, $S \supseteq R$, by adjoining a finite number of adjacent edges as follows. If $v \in V(R) \cap V_{\rm fin}(T)$, we add all edges which are incident on v. If $v \in V(R) \cap V_{\rm inf}(T)$, we add all those incident edges, e, which correspond to $\vec{e} \in \vec{\Delta}(R,v)$ for which $\rho(\Lambda G(v), B(\vec{e})) \geq \delta$. By Lemma 4.1, and the fact that $\vec{\Delta}(v)/G(v)$ is finite, there are only finitely many such $\vec{e}$. We thus see that S is finite. The fact that $S \in \mathcal{T}(\delta)$ follows by essentially the same arguments as were used in the proof of Lemma 6.6. □

7 Connectedness properties of boundaries of hyperbolic groups

In this section, we continue the analysis of Section 6, bringing connectedness assumptions into play.

Suppose, as before, that Γ is a hyperbolic group, and that Σ is a cofinite $\Gamma-$tree with two-ended edge stabilisers. We now add the assumption that Γ is one ended, so that $\partial\Gamma$ is a continuum. In this case, we note:

Lemma 7.1 *For each $\vec{e} \in \vec{E}(\Sigma)$, the set $\Psi(\vec{e})$ is connected.*

Proof Since $\Gamma(e)$ is two-ended, we have $\mathrm{fr}\,\Psi(\vec{e}) = \Lambda\Gamma(e) = \{a, b\}$, where $a, b \in \Psi(\vec{e})$ are distinct. Moreover, $\Psi(\vec{e})$ is closed and $\Gamma(e)$–invariant. Also $\Psi(\vec{e}) \neq \{a, b\}$, since it must, for example, contain all points of $\partial\Phi(\vec{e})$.

Let K be a connected component of $\Psi(\vec{e})$. We claim that $K \cap \{a, b\} \neq \emptyset$. To see this, suppose $a, b \notin K$. There are subsets $K_1, K_2 \subseteq \Psi(\vec{e})$, containing K, with $a \notin K_1$, $b \notin K_2$, and which are clopen in $\Psi(\vec{e})$. Let $L = K_1 \cap K_2$. Thus, $K \subseteq L \subseteq \Psi(\vec{e}) \setminus \mathrm{fr}\,\Psi(\vec{e})$. Since $\Psi(\vec{e})$ is closed in $\partial\Gamma$, so is L, and since $\Psi(\vec{e}) \setminus \partial\Psi(\vec{e})$ is open in $\partial\Gamma$, so also is L. In other words, L is clopen in M, contradicting the hypothesis that $\partial\Gamma$ is connected.

Suppose, then, that $a \in K$. Let $H \leq \Gamma(e)$ be the subgroup (of index at most 2) fixing a (and hence b). We see that K is H–invariant. Now $\Lambda H = \{a, b\}$ so either $b \in K$, or $K = \{a\}$. In the former case, we see that $K = \Psi(\vec{e})$, showing that $\Psi(\vec{e})$ is connected. In the latter case, we see, by a similar argument, that the component of K containing b equals $\{b\}$, giving the contradiction that $\Psi(\vec{e}) = \{a, b\}$. □

Now, as in Section 6, we focus on one vertex $\omega \in V(\Sigma)$, and write $G = \Gamma(\omega)$. Let T be a complete G–tree. Now, $\Lambda G = \Lambda_0 G \sqcup \Lambda_\infty G$, where $\Lambda_0 G = \bigsqcup_{v \in V(T)} \Lambda G(v)$ and $\Lambda_\infty G$ is identified with ∂T. It is possible that T may be trivial, but most of the following discussion will be vacuous in that case. If not, then $\Lambda_\infty G$ is dense in ΛG.

We now reintroduce the notation used in Section 6, namely Ξ, $J(\xi)$, $H(\xi)$, $B(\vec{e})$, etc. Note that if $\xi \in \Xi$ corresponds to the directed edge $\vec{\epsilon}$ of Σ, then $J(\xi)$ equals $\Psi(\vec{\epsilon})$ and the closure of $\partial\Gamma \setminus J(\xi)$ in $\partial\Gamma$ equals $\Psi(-\vec{\epsilon})$ (in the notation of Section 4). Thus, rephrasing Lemma 7.1, we get:

Lemma 7.2 *For each $\xi \in \Xi$, the set $J(\xi)$ is connected. Moreover, the closure of $\partial\Gamma \setminus J(\xi)$ in $\partial\Gamma$ is also connected.* □

Let $\mathcal{B} = \{\beta(\xi) \mid \xi \in \Xi_{\mathrm{hyp}}\}$. Now, Ξ_{hyp}/G is finite, so Lemma 2.1 tells us that:

Lemma 7.3 *The arc system $\mathcal{B}$ is edge-finite.* □

Now, since Γ is one-ended, the set of two-ended subgroups $\mathcal{H} = \{H(\xi) \mid \xi \in \Xi_{\mathrm{hyp}}\}$ is indecomposable. Since $\mathcal{B}$ is the set of axes of elements of $\mathcal{H}$, we see by Proposition 3.8 that:

Lemma 7.4 *$\mathcal{B}$ is indecomposable.* □

Alternatively, one can give a direct proof of Lemma 7.4 along the lines of Lemma 5.4. Thus, if $\mathcal{B}$ is decomposable, we can find two finitely separated subgraphs, F_1 and F_2, of T, so that $\partial G = A(F_1) \sqcup A(F_2)$, and such that for all $\xi \in \Xi_{\text{hyp}}$, either $\partial\beta(\xi) \in \partial F_1$, or $\partial\beta(\xi) \in \partial F_2$. It follows that $\partial\Gamma = B(F_1) \sqcup B(F_2)$ are closed in $\partial\Gamma$, contradicting the assumption that $\partial\Gamma$ is connected.

To go further, we shall want some more general observations and notation regarding simplicial trees. For the moment, T can be any simplicial tree, and $\mathcal{B}$ any arc system on T.

In Section 3, we associated to any finite subtree, $S \subseteq T$, an equivalence relation, $\sim_S \,=\, \sim_{S,\mathcal{B}}$, on ∂T. This, in turn, gives us a subpartition, $\mathcal{W}(S)$, of the set $V(S)$ of vertices of S. The elements of $\mathcal{W}(S)$ are the vertex sets of the connected components of the Whitehead graph, $\mathcal{G}(S)$.

More generally, we shall say that a subtree, S, of T is *bounded* if it has finite diameter in the combinatorial metric on T. In particular, every arc of $\mathcal{B}$ meets S, if at all, in a compact interval (or point). We define the equivalence relation, $\sim_S \,=\, \sim_{S,\mathcal{B}}$ on ∂T in exactly the same way as for finite trees. We also get a graph $\mathcal{G}(S)$, and a subpartition, $\mathcal{W}(S)$ of $V(S)$ as before. Note that if $\mathcal{B}$ is edge-finite, then $\mathcal{G}(S)$ is locally finite.

We have already observed that if $R \subseteq S$ is a subtree of S, then the relation $\sim_R$ is coarser than the relation $\sim_S$ (ie $x \sim_S y$ implies $x \sim_R y$). Moreover, the subpartition, $\mathcal{W}(R)$ of $V(R)$ can be described explicitly in terms of the subpartition $\mathcal{W}(S)$ and the map $\pi_R | V(S)\colon V(S) \longrightarrow V(R)$. To do this, define $\cong$ to be the equivalence relation on $\mathcal{W}(S)$ generated by relations of the form $W \cong W'$ whenever $\pi_R W \cap \pi_R W' \neq \emptyset$. An element of $\mathcal{W}(R)$ is then a union of sets of the form $\pi_R W$ as W ranges over some $\cong$–class in $\mathcal{W}(S)$. For future reference, we note:

Lemma 7.5 *Suppose $R \subseteq S$ are bounded subtrees of T. If $W \in \mathcal{W}(S)$, $W \subseteq V(R)$, and $W \cap \pi_R(V(S) \setminus V(R)) = \emptyset$, then $W \in \mathcal{W}(R)$.*

Proof If $W' \in \mathcal{W}(S)$ and $W \cap \pi_R W' \neq \emptyset$, then $W \cap W' \neq \emptyset$. (To see this, choose $v \in W'$ with $\pi_R v \in W \subseteq V(R)$. Since $W \cap \pi_R(V(S) \setminus V(R)) = \emptyset$, it follows that $v \in V(R)$, so $\pi_R v = v$. Thus $v \in W \cap W'$.) Since $W, W' \in \mathcal{W}(S)$ we thus have $W = W'$, so $W' = \pi_R W'$. This shows that any set of the form $\pi_R W'$ for $W' \in \mathcal{W}(S)$ which meets W must, in fact, be equal to W. From the description of $\mathcal{W}(R)$ given above, we see that $W \in \mathcal{W}(R)$. □

Given a directed edge $\vec{e} \in \vec{E}(T)$, let $\mathcal{S}(\vec{e})$ be the set of finite subtrees, S, of T with the property that $\vec{\Delta}(\mathrm{head}(\vec{e})) \cap \vec{E}(S) = \{\vec{e}\}$ (ie $e \subseteq S$, and $\mathrm{head}(\vec{e})$ is a terminal vertex of S). Given $S \in \mathcal{S}(\vec{e})$, we define the equivalence relation $\simeq_S$ on $\partial\Phi(\vec{e})$ to be the transitive closure of relations of the form $x \simeq_S y$ whenever $\pi_S x = \pi_S y$ or $\partial\beta = \{x, y\}$ for some $\beta \in \mathcal{B}$, with $\beta \subseteq \Phi(\vec{e})$. Clearly, if $x \simeq_S y$ then $x \sim_S y$. Also, if $R, S \in \mathcal{S}(\vec{e})$ with $R \subseteq S$, then $x \simeq_S y$ implies $x \simeq_R y$. We can also define a subpartition, $\mathcal{W}(S, \vec{e})$, of $V(S) \setminus \{\mathrm{head}(\vec{e})\}$, in a similar manner to $\mathcal{W}(S)$, as described in Section 3.

Suppose now that $\mathcal{B}$ is edge-finite and indecomposable, and suppose $S \in \mathcal{S}(\vec{e})$. Suppose $Q \subseteq \partial\Phi(\vec{e})$ is a $\simeq_S$–class. Since there is only one $\sim_S$–class, there must be some $\beta \in \mathcal{B}$ with one endpoint in Q and one endpoint in $\partial\Phi(-\vec{e})$. Thus, $e \subseteq \beta$. It follows that the number of $\simeq_S$–classes is bounded by the number of arcs in $\mathcal{B}$ containing the edge e. By the edge-finiteness assumption, this number is finite. It follows that, as the trees $S \in \mathcal{S}(\vec{e})$ get bigger, the relations $\simeq_S$ must stabilise. More precisely, there is a (unique) equivalence relation, $\simeq$, on $\partial\Phi(\vec{e})$ such that the set $\mathcal{S}_0(\vec{e}) = \{S \in \mathcal{S}(\vec{e}) \mid \simeq_S = \simeq\}$ contains all but finitely many elements of $\mathcal{S}(\vec{e})$. Note that if $R \in \mathcal{S}_0(\vec{e})$, $S \in \mathcal{S}(\vec{e})$, and $R \subseteq S$, then $S \in \mathcal{S}_0(\vec{e})$. Note also that there are finitely many $\simeq$–classes.

We now return to the set-up described earlier, with T a complete G–tree, and with $\mathcal{B} = \{\beta(\xi) \mid \xi \in \Xi_{\mathrm{hyp}}\}$. We have seen that $\mathcal{B}$ is edge-finite and indecomposable. We note:

Lemma 7.6 *Suppose $\vec{e} \in \vec{E}(T)$ and $x, y \in \partial\Phi(\vec{e})$. If $x \simeq y$, then x and y lie in the same connected component of $B(\vec{e})$.*

Proof Suppose, for contradiction that x and y lie in different components of $B(\vec{e})$. We can partition $B(\vec{e})$ into two closed subsets, $B(\vec{e}) = K \sqcup L$, with $x \in K$ and $y \in L$.

Let $\delta = \frac{1}{2}\rho(K, L) > 0$. By Lemma 6.7, we can find some $R \in \mathcal{T}(\delta)$. By Lemma 6.6, we can suppose that $S = R \cap (e \cup \Phi(\vec{e})) \in \mathcal{S}_0(\vec{e})$. (For example, take R to be the smallest tree containing a given element of $\mathcal{T}(S)$ and a given element of $\mathcal{S}_0(\vec{e})$.) Thus, $\simeq_S = \simeq$, so in particular, $x \simeq_S y$. Note that, if $v \in V(S) \setminus \{\mathrm{head}(\vec{e})\}$, then $F(R, v) = F(S, v)$ (in the notation of Section 2).

Now, from the definition of the relation $\simeq_S$, we have a finite sequence, $x = x_0, x_1, \ldots, x_n = y$ of points of $\partial\Phi(\vec{e})$, such that for each i, either $\pi_S x_i = \pi_S x_{i+1}$, or there is some $\xi \in \Xi_{\mathrm{hyp}}$, with $\partial\beta(\xi) = \{x_i, x_{i+1}\}$. Now, $\partial\Phi(\vec{e}) \subseteq B(\vec{e}) = K \sqcup L$, so for each i, either $x_i \in K$ or $x_i \in L$. We claim, by induction on i, that $x_i \in K$ for all i.

Suppose, then, that $x_i \in K$. Suppose first, that $\{x_i, x_{i+1}\} = \partial\beta(\xi)$ for some $\xi \in \Xi_{\rm hyp}$. We have that $x_i, x_{i+1} \in J(\xi) \subseteq B(\vec{e})$. Moreover, by Lemma 6.1, $J(\xi)$ is connected. It follows that $x_{i+1} \in K$.

We can thus suppose that $\pi_S x_i = \pi_S x_{i+1} = v \in V(S) \setminus \{{\rm head}(\vec{e})\}$. Thus, $x_i, x_{i+1} \in \partial F(S,v) = \partial F(R,v) \subseteq B(F(R,v))$. Now, if $v \in V_{\rm fin}(T)$, then, since $R \in \mathcal{T}(\delta)$, we have ${\rm diam}\, B(F(R,v)) < \delta$. Therefore, $\rho(x_i, x_{i+1}) < \delta$ and so $x_{i+1} \in K$. Thus, we can assume that $v \in V_{\rm inf}(T)$. Since $x_i \in \partial F(R,v)$, we have $x_i \in \partial\Phi(\vec{\epsilon})$ for some $\vec{\epsilon} \in \vec{\Delta}(R,v)$. Again, since $R \in \mathcal{T}(\delta)$, we have ${\rm diam}\, B(\vec{\epsilon}) < \delta$ and $\rho(B(\vec{\epsilon}), \Lambda G(v)) < \delta$. Thus, $\rho(x_i, \Lambda G(v)) < 2\delta$. Similarly, $\rho(x_{i+1}, \Lambda G(v)) < 2\delta$. Now, $\Lambda G(v)$ is connected, and so it again follows that $x_{i+1} \in K$.

Thus, by induction on i, we arrive at the contradiction that $y = x_n \in K$. This shows that x and y lie in the same component of $B(\vec{e})$ as required. $\square$

Now, fix some $v \in V_{\rm inf}(T)$, so that $G(v)$ is one-ended, and $\Lambda G(v)$ is a subcontinuum of $\partial\Gamma$.

We say that a $G(v)$–invariant subtree, S, of T is *stable about* v if $S \cap \Phi(\vec{e}) \in \mathcal{S}_0(\vec{e})$ for all $\vec{e} \in \vec{\Delta}(v)$. Note that, since $\vec{\Delta}(v)/G(v)$ is finite, $S/G(v)$ is finite. In particular, we see that S is bounded (ie has finite diameter). Note that, since S contains every edge of T incident on v, we have $\pi_S \partial T \subseteq V(S) \setminus \{v\}$. Let $\sim_S = \sim_{S,\mathcal{B}}$ be the equivalence relation on ∂T as defined in Section 3 (in the case of finite trees). We remark that $\sim_S$ is independent of the choice of stable tree, S, since it is easily seen to be definable purely in terms of the arc system $\mathcal{B}$, and the relations, $\simeq$ for $\vec{e} \in \vec{\Delta}(v)$. We shall thus write $\sim_S$ simply as $\sim$. Clearly, $\sim$ is $G(v)$–invariant. (It need not be trivial, since we are only assuming that S is bounded.)

We can certainly construct a stable tree about v by taking $S = \bigcup_{\vec{e} \in \vec{\Delta}(v)} S(\vec{e})$. In this case, $S \cap \Phi(\vec{e}) = S(\vec{e}) \in \mathcal{S}_0(\vec{e})$.

Note that we get a subpartition, $\mathcal{W}(S)$, of $V(S)$, as described in Section 3. Note that $\bigcup \mathcal{W}(S) \subseteq \pi_S \partial T$. In particular, $v \notin \bigcup \mathcal{W}(S)$.

Lemma 7.7 *The setwise stabiliser, in $G(v)$, of every $\sim$–class is infinite.*

Proof As described in Section 3, each $\sim$–class corresponds to an element of $\mathcal{W}(S)$. Moreover, $(\bigcup \mathcal{W}(S))/G(v) \subseteq V(S)/G(v)$ is finite. Thus, the lemma is equivalent to asserting that each element of $\mathcal{W}(S)$ is infinite.

Suppose, to the contrary, that $W \in \mathcal{W}(S)$ is finite. Let $\vec{\Delta}_0 = \{\vec{e} \in \vec{\Delta}(v) \mid W \cap S(\vec{e}) \neq \emptyset\}$, and let $R = \bigcup_{\vec{e} \in \vec{\Delta}_0} S(\vec{e})$. Thus, R is a finite subtree of

S, and $W \subseteq V(R)$. Moreover, $\pi_R(V(S) \setminus V(R)) = \{v\}$, so, in particular, $W \cap \pi_R(V(S) \setminus V(R)) = \emptyset$. Thus, by Lemma 7.5, $W \in \mathcal{W}(R)$. But $v \in \bigcup \mathcal{W}(R)$ (since any element of $\partial\Phi(\vec{e})$ for $\vec{e} \in \vec{\Delta}(v) \setminus \vec{\Delta}_0$ projects to v under π_R). Thus, $\mathcal{W}(R) \neq \{W\}$. This shows that there is more than one $\sim_R$–class, contradicting the fact that $\mathcal{B}$ is indecomposable. □

Finally, we note:

Lemma 7.8 *If $x, y \in \partial T$ with $x \sim y$, then x and y lie in the same quasi-component of $\partial\Gamma \setminus \Lambda G(v)$.*

Proof In fact, we shall show that x and y both lie in a compact connected subset, K, of $\partial\Gamma \setminus \Lambda G(v)$.

By the definition of the relation $\sim = \sim_S$, we can assume that either $\pi_S x = \pi_S y$ or there is some $\xi \in \Xi_{\rm hyp}$ with $\partial\beta(\xi) = \{x, y\}$.

In the former case, let $w = \pi_S x = \pi_S y$. Thus, $w \in V(S(\vec{e}))$ for some $\vec{e} \in \vec{\Delta}(v)$. Since $S(\vec{e}) \in \mathcal{S}_0(\vec{e})$, we have $x \simeq y$, and so, by Lemma 7.6, x and y lie in the same component of $B(\vec{e})$. Call this component K. Thus, K is closed in $B(\vec{e})$ and hence in $\partial\Gamma$. Note that, from the definition of $B(\vec{e})$, we have $B(\vec{e}) \cap \Lambda G(v) = \emptyset$ and so $K \cap \Lambda G(v) = \emptyset$.

In the latter case, set $K = J(\xi)$. Thus, by Lemma 6.1, K is connected. Also $K \cap \Lambda G = \{x, y\} \subseteq \partial T$, and so, again, $K \cap \Lambda G(v) = \emptyset$. □

8 Global cut points

In this section, we set out the "inductive step" of the proof that a strongly accessible hyperbolic group has no global cut points in its boundary. In the light of the result announced in [8], we see that this, in fact, applies to all one-ended hyperbolic groups. A more direct proof of the general case was given in [28] using the results of [4,6,19]. (See also [7].)

Specifically, we shall show:

Theorem 8.1 *Suppose that Γ is a one-ended hyperbolic group. Suppose that we represent Γ as a finite graph of groups over two-ended subgroups. Suppose that each maximal one-ended subgroup of each vertex group has no global cut point in its boundary (as an intrinsic hyperbolic group). Then, $\partial\Gamma$ has no global cut point.*

Before we start on the proof, we give a few general definitions and observations relating to global cut points.

Suppose that M is any continuum, ie a compact connected hausdorff space. (For the moment, the compactness assumption is irrelevant.) If $p \in M$, and $O, U \subseteq M$, we write OpU to mean that O and U are non-empty open subsets and that M is (set theoretically) a disjoint union $M = O \sqcup \{p\} \sqcup U$. Note that $\mathrm{fr}\, O = \mathrm{fr}\, U = \{p\}$. Also, it's not hard to see that $O \cup \{p\}$ and $U \cup \{p\}$ are connected. (More discussion of this is given in [4].) We say that a point $p \in M$ is a *global cut point* if there exist $O, U \subseteq M$ with OpU.

Definition If $Q \subseteq M$ is any subset, and $p \in M$, we say that Q is *indivisible in M at p* if whenever we have $O, U \subseteq M$ with OpU, then either $Q \cap O = \emptyset$ or $Q \cap U = \emptyset$.

If $R \subseteq M$ is another subset, we say that Q is *indivisible in M over R*, if it is indivisible in M at every point of R.

We say that Q is *(globally) indivisible* in M if it is indivisible at every point of M.

Thus, M is indivisible in itself if and only if it does not contain a global cut point.

Obviously, if $P \subseteq Q \subseteq M$ and Q is indivisible in M, then so is P. Also any subcontinuum of M with no global cut point is indivisible in M. We shall need the following simple observations:

Lemma 8.2 *If $P, Q \subseteq M$ are indivisible in M, and $\mathrm{card}(P \cap Q) \geq 2$, then $P \cup Q$ is indivisible in M.*

Proof Suppose OpU. Choose any $x \in P \cap Q \setminus \{p\}$. We can assume that $x \in O$, so that $P \cap U = Q \cap U = \emptyset$. Thus $(P \cup Q) \cap U = \emptyset$. □

Lemma 8.3 *Suppose that $\mathcal{Q}$ is a chain of indivisible subsets of M (ie if $P, Q \in \mathcal{Q}$, then $P \subseteq Q$ or $Q \subseteq P$). Then $\bigcup \mathcal{Q}$ is indivisible.*

Proof Suppose OpU, and $x \in O \cap (\bigcup \mathcal{Q})$ and $y \in U \cap (\bigcup \mathcal{Q})$. Then $x, y \in Q$ for some $Q \in \mathcal{Q}$, contradicting the indivisibility of Q. □

Lemma 8.4 *If Q is indivisible in M, then so is its closure, $\bar{Q}$.*

Proof If OpU, then we can assume that $O \cap Q = \emptyset$, so $O \cap \bar{Q} = \emptyset$. □

Now, let Γ be a one-ended hyperbolic group, and let Σ be a cofinite Γ–tree with two-ended edge stabilisers. We begin with the following observation:

Lemma 8.5 *If $\Lambda\Gamma(v)$ is indivisible in $\partial\Gamma$ for all $v \in V(\Sigma)$, then $\partial\Gamma$ is indivisible.*

Proof Note that if $v, w \in V(\Sigma)$ are adjacent, then $\Gamma(v) \cap \Gamma(w)$ is two-ended, so $\Lambda\Gamma(v) \cap \Lambda\Gamma(w) = \Lambda(\Gamma(v) \cap \Gamma(w))$ consists of a pair of points. Thus, by Lemma 8.2, $\Lambda\Gamma(v) \cap \Lambda\Gamma(w)$ is indivisible in $\partial\Gamma$. By an induction argument, we see that $\bigcup_{v \in V(S)} \Lambda\Gamma(v)$ is indivisible for any finite subtree, $S \subseteq \Sigma$. Taking an exhaustion of Σ by an increasing sequence of finite subtrees, and applying Lemma 8.3, we see that $\bigcup_{v \in V(\Sigma)} \Lambda\Gamma(v)$ is indivisible. But this set is dense in $\partial\Gamma$ (since it is non-empty and Γ–invariant). The result follows by Lemma 8.4. □

In fact, it's enough to verify the hypotheses of Lemma 8.5 for those $v \in V(\Sigma)$ for which $\Gamma(v)$ is not two-ended. To see this, first note that if α is a finite arc connecting two points $v_0, v_1 \in V(\Sigma)$ such that $\Gamma(v)$ is two ended for all $v \in V(\alpha) \setminus \{v_0, v_1\}$, then the groups $\Gamma(e)$ and $\Gamma(v)$ are all commensurable for all $e \in E(\alpha)$ and $v \in V(\alpha) \setminus \{v_0, v_1\}$. Now, since Γ is hyperbolic and not two-ended, there must be some $v_0 \in V(\Sigma)$ such that $\Gamma(v_0)$ is not two-ended. Suppose that $v \in V(\Sigma)$ is some other vertex. Connect v to v_0 by an arc in Σ, and let w be the first vertex of this arc for which $\Gamma(w)$ is not two-ended. Thus, $\Gamma(v) \cap \Gamma(w)$ has finite index $\Gamma(v)$, and so $\Lambda\Gamma(v) \subseteq \Lambda\Gamma(w)$. Clearly, if $\Lambda\Gamma(w)$ is indivisible in $\partial\Gamma$, then so is $\Lambda\Gamma(v)$.

As in Section 7, we now fix $\omega \in V_{\mathrm{inf}}(\Sigma)$ and set $G = \Gamma(\omega)$. We are interested in the indivisibility properties of ΛG as a subset of $\partial\Gamma$. We aim to show that if ΛG is indivisible in $\partial\Gamma$ at each point of $\Lambda_0 G$, then it is (globally) indivisible in $\partial\Gamma$ (Corollary 8.8). Moreover, if $\Lambda G(v)$ is indivisible in $\partial\Gamma$ at some point $p \in \Lambda G(v)$, then ΛG is also indivisible in $\partial\Gamma$ at p (Proposition 8.9). As a corollary, we deduce (Corollary 8.10) that if $\Lambda G(v)$ is indivisible in $\partial\Gamma$ for all $v \in V(T)$, then ΛG is indivisible in $\partial\Gamma$. (Note that this is the essential ingredient in showing that $\partial\Gamma$ has no global cut point, as in Lemma 8.5.)

Recall the notation Ξ, $J(\xi)$, $H(\xi)$, $B(\vec{e})$ etc from Section 6. We begin with the following observation:

Lemma 8.6 *ΛG is indivisible in $\partial\Gamma$ over $\partial\Gamma \setminus \Lambda G$.*

Proof Suppose $p \in \partial\Gamma \setminus \Lambda G$. Then, by Lemma 6.1, $p \in J(\xi) \setminus \mathrm{fr}\, J(\xi)$ for some $\xi \in \Xi$. Let K be the closure of $\partial\Gamma \setminus J(\xi)$ in $\partial\Gamma$. By Lemma 7.2, K is connected. Moreover $\Lambda G \subseteq K$. Suppose $O, U \subseteq M$ with OpU. Without loss of generality, we can suppose that $K \cap U = \emptyset$. (Otherwise $O \cap K$ and $U \cap K$ would partition K.) But $\Lambda G \subseteq K$, and so $\Lambda G \cap U = \emptyset$. □

Recall the notation $\mathcal{S}_0(\vec{e})$, $\simeq_S$ etc from Section 7.

For each $\vec{e} \in \vec{E}(T)$, we shall choose $S(\vec{e}) \in \mathcal{S}_0(\vec{e})$. We do this equivariantly with respect to the action of G. Thus, $N = \max\{\mathrm{diam}\, S(\vec{e}) \mid \vec{e} \in \vec{E}(T)\} < \infty$ (where diam denotes diameter with respect to combinatorial distance in T).

Lemma 8.7 *ΛG is indivisible in $\partial\Gamma$ over $\Lambda_\infty G$.*

Proof Clearly, we can assume that $\Lambda_\infty G$ is non-empty, and hence dense in ΛG. Suppose that $p \in \Lambda_\infty G$, and $O, U \subseteq \partial\Gamma$ with OpU. If $O \cap \Lambda G \neq \emptyset$, then $O \cap \Lambda_\infty G \neq \emptyset$, and similarly for U. Thus, suppose, for contradiction, that there exist $x \in O \cap \Lambda_\infty G$ and $y \in U \cap \Lambda_\infty G$. Clearly x, y and p are all distinct.

Now, let $v \in V(T)$ be the median of the points $x, y, p \in \partial T$. In other words, v is the unique intersection point of the three arcs connecting the points x, y and p pairwise. Let α be the ray from v to p, and let $w \in V(T)$ be that vertex at distance $N+1$ from v along α. Let $\vec{e}$ be the directed edge of α pointing towards p with $\mathrm{head}(\vec{e}) = w$ (so that $\mathrm{dist}(v, \mathrm{tail}(\vec{e})) = N$. Thus $x, y \in \partial\Phi(\vec{e})$ and $p \in \partial\Phi(-\vec{e})$.

Write $S = S(\vec{e})$, so that $\mathrm{diam}\, S \leq N < \mathrm{dist}(v, w)$. Now v is the nearest point to w in the biinfinite arc connecting x to y. We see that this arc does not meet S, and so $\pi_S x = \pi_S y$. In particular, $x \simeq_S y$, and so, since $S \in \mathcal{S}_0(\vec{e})$, we have $x \simeq y$. By Lemma 7.6, x and y lie in the same component of $B(\vec{e})$. But, $\partial\Phi(-\vec{e}) \cap B(\vec{e}) = \emptyset$, and so $p \notin B(\vec{e})$. But this contradicts the fact that p separates x from y. (More formally, $O \cap B(\vec{e})$ and $U \cap B(\vec{e})$ partition $B(\vec{e})$ into two non-empty open sets.) □

Putting Lemma 8.7 together with Lemma 8.6, we obtain:

Corollary 8.8 *If ΛG is indivisible in $\partial\Gamma$ over $\Lambda_0 G$, then ΛG is (globally) indivisible in $\partial\Gamma$.* □

Next, we show:

Proposition 8.9 *If $\Lambda G(v)$ is indivisible in $\partial\Gamma$ at the point $p \in \Lambda G(v)$, then ΛG is indivisible in $\partial\Gamma$ at p.*

Proof First, note that if T is trivial, then $G = G(v)$, so there is nothing to prove. We can thus assume that T is non-trivial.

Suppose that $O, U \subseteq \partial\Gamma$ with OpU. Since $\Lambda G(v)$ is indivisible in $\partial\Gamma$ at p, we can assume that $U \cap \Lambda G(v) = \emptyset$. We claim that $U \cap \Lambda G = \emptyset$. Since $\Lambda_\infty G$ is dense in ΛG, it's enough to show that $U \cap \Lambda_\infty G = \emptyset$.

Suppose, to the contrary, that there is some $x \in U \cap \Lambda_\infty G$. Let $G_0 \subseteq G(v)$ be the setwise stabiliser of the $\sim$–class of x. By Lemma 7.7, G_0 is infinite. Now a hyperbolic group cannot contain an infinite torsion subgroup (see for example [14]) and so we can find some $g \in G_0$ of infinite order.

Now, for each $i \in \mathbb{Z}$, $g^i x \sim x$, so, by Lemma 7.8, there is a connected subset (in fact a subcontinuum), K, containing x and $g^i x$, with $K \cap \Lambda G(v) = \emptyset$. Since $p \in \Lambda G(v)$, we have $K \subseteq \partial\Gamma \setminus \{p\}$. Thus, $K \subseteq U$. (Otherwise $O \cap K$ and $U \cap K$ would partition K.) In particular, $g^i x \in U$. Now, as $i \to \infty$, the sequences $g^i x$ and $g^{-i} x$ converge on distinct points, $a, b \in \Lambda G_0 \subseteq \Lambda G(v)$. Since $U \cup \{p\}$ is closed, we have $a, b \in U \cup \{p\}$, and so, without loss of generality, $a \in U$. But now, $a \in U \cap \Lambda G(v)$, contradicting the assumption that $U \cap \Lambda G(v) = \emptyset$. $\square$

Putting Proposition 8.9 together with Corollary 8.8, we get:

Corollary 8.10 *Suppose that, for all $v \in V_{\mathrm{inf}}(T)$, the continuum $\Lambda G(v)$ is indivisible in $\partial\Gamma$ over $\Lambda G(v)$. Then, ΛG is (globally) indivisible in $\partial\Gamma$.* $\square$

Of course, it's enough to suppose that each continuum $\Lambda G(v)$ has no global cut point.

Finally, putting Corollary 8.10 together with Lemma 8.5, we get the main result of this section, namely Theorem 8.1.

9 Strongly accessible groups

In this final section, we look once more at the property of strong accessibility over finite and two-ended subgroups. We begin with general groups, and specialise to finitely presented groups. We finish by showing how Theorem 8.1, together with the results of [4,6] imply that the boundary of a one-ended strongly accessible hyperbolic group has no global cut point (Theorem 9.3).

As discussed in the introduction, the issue of strong accessibility is concerned with sequences of splittings over a class of subgroups (in particular, the class of finite and two-ended subgroups), and when such sequences must terminate.

In general, this may depend on the choices of splittings that we make at each stage of the process. We first describe a few general results which imply, at least for finitely presented groups, that we can assume that at any given stage, we can split over finite groups whenever this is possible.

Suppose, for the moment, that Γ is any group, and that G_1 and G_2 are one-ended subgroups with $G_1 \cap G_2$ infinite. Then the group, $\langle G_1 \cup G_2 \rangle$, generated by G_1 and G_2 is also one-ended. (For if not, there is a non-trivial action of $\langle G_1 \cup G_2 \rangle$ on a tree, T, with finite edge stabilisers. Now, since the groups, G_i are one-ended, they each fix a unique vertex of T. Since $G_1 \cap G_2$ is infinite, this must be the same vertex, contradicting the non-triviality of the action.) Note that essentially the same argument works if G_1 is one-ended and G_2 is two-ended.

Similarly, suppose that $G \leq \Gamma$ is one-ended, and $g \in \Gamma$ with $G \cap gGg^{-1}$ infinite. Then $\langle G, g \rangle$ is one-ended. (Since if $\langle G, g \rangle$ acts on a tree, T, with finite edge stabilisers, then G and gGg^{-1} must fix the same unique vertex of T. Thus, g must also fix this vertex, again showing that the action is trivial.) Recall that the commensurator, $\mathrm{Comm}(G)$, of G is the set of elements $g \in \Gamma$ such that $G \cap gGg^{-1}$ has finite index in G. Thus, $\mathrm{Comm}(G)$ is a subgroup of Γ containing G. We see that if G is one-ended, then so is $\mathrm{Comm}(G)$.

Now, suppose that Γ is accessible over finite groups. Then every one-ended subgroup of Γ is contained in a unique maximal one-ended subgroup of Γ. Each maximal one-ended subgroup is equal to its commensurator, and there are only finitely many conjugacy classes of such subgroups. If G is a maximal one-ended subgroup, and $H \leq G$ is two-ended, then either $H \leq G$ or else $H \cap G$ is finite. Moreover, H can lie in at most one maximal one-ended subgroup. These observations follow from the remarks of the previous two paragraphs. They can also be deduced by considering the action of H on a complete Γ–tree.

Now, suppose that Γ splits as an amalgamated free product or HNN–extension over a two-ended subgroup. This corresponds to a Γ–tree, Σ, with just one orbit of edges, and with two-ended edge stabiliser. We consider two cases, depending on whether or not the edge group is elliptic or hyperbolic, ie whether or not it lies in a one-ended subgroup of Γ.

Consider, first, the case where the edge stabiliser of Σ does not lie in a one-ended subgroup, and hence intersects every one-ended subgroup in a finite group. In this case, we have:

Lemma 9.1 *Suppose $v \in V(\Sigma)$. Then, each maximal one-ended subgroup of $\Gamma(v) = \Gamma_\Sigma(v)$ is a maximal one-ended subgroup of Γ. Moreover, every maximal one-ended subgroup of Γ arises in this way (for some $v \in V(\Sigma)$).*

Proof Suppose, first, that G is any one-ended subgroup of Γ. Then, G must lie inside some (unique) vertex stabiliser $\Gamma(v)$. (Otherwise, G would split over a group of the form $G \cap H$, where H is an edge-stabiliser. But $G \cap H$ is finite, contradicting the fact that G is one-ended.) If G is maximal in Γ, then clearly it is also maximal in $\Gamma(v)$.

Conversely, suppose that G is a maximal one-ended subgroup of a vertex stabiliser, $\Gamma(v)$. Let G' be the unique maximal one-ended subgroup of Γ containing G. By the first paragraph, G' lies inside some vertex group, which must, in this case, be $\Gamma(v)$. By maximality in $\Gamma(v)$, we must therefore have $G = G'$. □

The second case is when an edge group lies inside some one-ended subgroup. To consider this case, fix and edge e of Σ, with endpoints $v, w \in V(\Sigma)$. Now, $\Gamma(e)$ lies inside a unique maximal one-ended subgroup, Γ_0, of Γ. Any other maximal one-ended subgroup of Γ must intersect $\Gamma(e)$ in a finite subgroup. In this case, we have:

Lemma 9.2 *Γ_0 splits as an amalgamated free product or HNN extension over $\Gamma(e)$, with incident vertex groups equal to $\Gamma_0 \cap \Gamma(v)$ and $\Gamma_0 \cap \Gamma(w)$. Each maximal one-ended subgroup of $\Gamma(v)$ is a maximal one-ended subgroup of $\Gamma_0 \cap \Gamma(v)$ or of Γ (and similarly for w). Every maximal one-ended subgroup of $\Gamma_0 \cap \Gamma(v)$ arises in this way. Each maximal one-ended subgroup of Γ is conjugate, in Γ, to Γ_0 or to a maximal one-ended subgroup of $\Gamma(v)$ or $\Gamma(w)$.*

Proof Suppose G is a maximal one-ended subgroup of Γ. Either G contains some edge-stabiliser, so that some conjugate of G contains $\Gamma(e)$ and hence equals Γ_0, or else G meets each edge stabiliser in a finite group. In the latter case, we see, as in Lemma 9.1, that G is a maximal one-ended subgroup of a vertex group.

Now suppose that G is a maximal one-ended subgroup of $\Gamma(v)$. Let G' be the maximal one-ended subgroup of Γ containing G. From the first paragraph, we see that either $G' = \Gamma_0$, or G' is a maximal one-ended subgroup of $\Gamma(v)$. In the former case, we see that $G \subseteq \Gamma_0 \cap \Gamma(v)$, and must therefore be maximal one-ended in $\Gamma_0 \cap \Gamma(v)$. The latter case, we obtain $G = G'$.

Finally suppose that G is a maximal one-ended subgroup of $\Gamma_0 \cap \Gamma(v)$. Let G' be the maximal one-ended subgroup of $\Gamma(v)$ containing G. From the previous paragraph, we see that $G' \subseteq \Gamma_0 \cap \Gamma(v)$, so $G = G'$.

It remains to show that Γ_0 splits over $\Gamma(e)$ in the manner described. This amounts to showing that if H is an edge stabiliser and a subgroup of $\Gamma_0 \cap \Gamma(v)$, then H is conjugate in $\Gamma_0 \cap \Gamma(v)$ to $\Gamma(e)$, (and similarly for w).

We know that there must be some $g \in \Gamma(v)$ such that $H = g\Gamma(e)g^{-1}$ Now, $H \subseteq \Gamma_0 \cap g\Gamma_0 g^{-1}$. Since H is infinite, it follows that the group generated by Γ_0 and $g\Gamma_0 g^{-1}$ must be one-ended, and so, by maximality, must equal Γ_0. Hence, $g\Gamma_0 g^{-1} = \Gamma_0$. In particular, $g \in \mathrm{Comm}(\Gamma_0)$. But, from the earlier discussion, $\mathrm{Comm}(\Gamma_0) = \Gamma_0$, and so $g \in \Gamma_0 \cap \Gamma(v)$ as required. □

We now go on to describe the notion of strong accessibility. To set up the notation, let Γ be any group, and let $\mathcal{C}$ be any conjugacy–invariant set of subgroups of Γ. (In the case of interest, $\mathcal{C}$ will be the set of all finite and two-ended subgroups of Γ.) We want to look at sequences of splittings of Γ over $\mathcal{C}$, where the only information retained at each stage will be the vertex groups of the previous splittings. In other words, we get a sequence of conjugacy invariant sets of subgroups of Γ. (In fact, if $\mathcal{C}$ is closed under isomorphism, we can just view these as isomorphism classes of groups.) Note that finite groups can never split non-trivially, and so for our purposes, we can throw away finite subgroups whenever they arise.

To be more formal, suppose that $\mathcal{J}$ and $\mathcal{J}'$ are both conjugacy invariant sets of subgroups of Γ. We say that $\mathcal{J}'$ is obtained by splitting $\mathcal{J}$ over $\mathcal{C}$ if it has the form $\mathcal{J}' = \bigcup_J \mathcal{J}(J)$, where $\mathcal{J}(J)$ is the set of (Γ–conjugacy classes of) infinite vertex groups of some splitting of J as a finite graph of groups over $\mathcal{C}$, and where J ranges over a conjugacy transversal in $\mathcal{J}$. Thus, a sequence of splittings of Γ over $\mathcal{C}$ consists of a sequence, $\mathcal{J}_0, \mathcal{J}_1, \mathcal{J}_2, \ldots$, where $\mathcal{J}_0 = \{\Gamma\}$, and each $\mathcal{J}_{i+1}$ is obtained as a splitting of $\mathcal{J}_i$ over $\mathcal{C}$ in the manner just described. Note that, by induction, each of the sets $\mathcal{J}_i$ is a finite union of conjugacy classes in Γ. Note also that we can assume, if we wish, by introducing some intermediate steps, that each $\mathcal{J}_{i+1}$ is obtained from $\mathcal{J}_i$ by splitting one of the conjugacy classes of $\mathcal{J}_i$ as an amalgamated free product or HNN extension, while leaving the remaining groups unchanged. We say that the sequence terminates, if for some n, none of the elements of $\mathcal{J}_n$ split non-trivially over $\mathcal{C}$. We say that Γ is *strongly accessible* over $\mathcal{C}$ if there exists such a sequence which terminates.

Suppose that $\mathcal{J}$ is a union of conjugacy classes of subgroups of Γ, each accessible over finite groups. Let $\mathcal{F}(\mathcal{J}) = \bigcup_{J \in \mathcal{J}} \mathcal{F}(J)$, where $\mathcal{F}(J)$ is the set of maximal one-ended subgroups of J. Thus $\mathcal{F}(\mathcal{J})$ is obtained by $\mathcal{J}$ by splitting over the class of finite subgroups of Γ, in the sense defined above.

Let us now suppose that Γ is finitely presented, and that $\mathcal{C}$ is the set of all finite and one-ended subgroups of Γ. Suppose that $(\mathcal{J}_i)_i$ is a sequence of splitting of Γ over $\mathcal{C}$. By induction, each element of each $\mathcal{J}_i$ is finitely presented and hence accessible over finite groups. We can thus form a sequence $(\mathcal{F}_i)_i$ where $\mathcal{F}_i = \mathcal{F}(\mathcal{J}_i)$. Now, we can assume that $\mathcal{J}_{i+1}$ is obtained from $\mathcal{J}_i$ by splitting an element of $\mathcal{J}_i$ as an amalgamated free product or HNN extension either over a finite group or over a two-ended group. In the former case, we see that $\mathcal{F}_{i+1} = \mathcal{F}_i$. In the latter case, we see, from Lemmas 9.1 and 9.2, that $\mathcal{F}_{i+1}$ is obtained from $\mathcal{F}_i$ by first splitting some element over a two-ended subgroup, and then, if necessary splitting over some finite subgroups to reduce ourselves again to one-ended groups. Thus, after inserting some intermediate steps if necessary, we can suppose that the sequence $(\mathcal{F}_i)_i$ is also a sequence of splittings of Γ over $\mathcal{C}$. If the sequence $(\mathcal{J}_i)_i$ terminates at $\mathcal{J}_n$, then $\mathcal{F}_n = \mathcal{F}(\mathcal{J}_n) = \mathcal{J}_n$, so $(\mathcal{F}_i)_i$ also terminates (and in the same set of subgroups).

In summary, we see that if Γ is finitely presented, and strongly accessible over $\mathcal{C}$, then we can find a terminating sequence of splittings over $\mathcal{C}$ where we split over finite groups wherever possible (in priority to splitting over two-ended subgroups). In other words, we only ever need to split one-ended groups over two-ended subgroups and to split infinite-ended and two-ended groups over finite subgroups.

Finally, suppose that Γ is a strongly accessible one-ended hyperbolic group, and that $\mathcal{J}_0, \mathcal{J}_1, \ldots, \mathcal{J}_n$ is a sequence of splitting of Γ over finite and one-ended subgroups, which terminates in $\mathcal{J}_n$. In this case, each elements of each $\mathcal{J}_i$ is quasiconvex, and hence intrinsically hyperbolic. Moreover, we can suppose, as above, that the only groups we ever split over two-ended groups are one-ended.

Now, each element of $\mathcal{J}_n$ is one-ended and does not split over any two-ended subgroup. From the results of [4,6], we see that each element of $\mathcal{J}_n$ has no global cut point in its boundary. Now, applying Theorem 8.1 inductively, we conclude that this is also true of Γ.

We have shown:

Theorem 9.3 *Suppose that Γ is a one-ended hyperbolic group which is strongly accessible over finite and two-ended subgroups. Then, $\partial\Gamma$ has no global cut point.* □

As mentioned in the introduction, Delzant and Potyagailo have shown that every finitely presented group, Γ, is strongly accessible over any "elementary" class of subgroups, $\mathcal{C}$. In particular, this deals with the case where Γ is hyperbolic, and where $\mathcal{C}$ is the set of finite and two-ended subgroups of Γ. We thus conclude that the boundary of any one-ended hyperbolic group has no global cut point, and is thus locally connected by the result of [3].

References

[1] **M Bestvina**, **M Feighn**, *Bounding the complexity of simplicial actions on trees*, Invent. Math. 103 (1991) 449–469

[2] **M Bestvina**, **M Feighn**, *A combination theorem for negatively curved groups*, J. Differential Geometry 35 (1992) 85–101

[3] **M Bestvina**, **G Mess**, *The boundary of negatively curved groups*, Journal Amer. Math. Soc. 4 (1991) 469–481

[4] **B H Bowditch**, *Treelike structures arising from continua and convergence groups*, Memoirs Amer. Math. Soc. (to appear)

[5] **B H Bowditch**, *Cut points and canonical splittings of hyperbolic groups*, Acta Math. 180 (1998) 145–186

[6] **B H Bowditch**, *Group actions on trees and dendrons*, Topology 37 (1998) 1275–1298

[7] **B H Bowditch**, *Connectedness properties of limit sets*, Transactions Amer. Math. Soc. (to appear)

[8] **T Delzant**, **L Potyagailo**, *Accessibilité hiérarchique des groupes de présentation finie*, Strasbourg/Lille preprint (1998)

[9] **W Dicks**, **M J Dunwoody**, *Groups acting on graphs*, Cambridge Studies in Advanced Mathematics No. 17, Cambridge University Press (1989)

[10] **M J Dunwoody**, *The accessibility of finitely presented groups*, Invent. Math. 81 (1985) 449–457

[11] **M J Dunwoody**, **M E Sageev**, *JSJ–splittings for finitely presented groups over slender subgroups*, Invent. Math. (to appear)

[12] **S Ferry**, *A stable converse to the Vietoris–Smale theorem with applications to shape theory*, Transactions Amer. Math. Soc. 261 (1980) 369–386

[13] **K Fujiwara**, **P Papasoglu**, *JSJ decompositions of finitely presented groups and complexes of groups*, preprint (1997)

[14] **E Ghys**, **P de la Harpe**, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math. 83, Birkhäuser (1990)

[15] **M Gromov**, *Hyperbolic groups*, from: "Essays in Group Theory", S M Gersten (editor) MSRI Publications No. 8, Springer–Verlag (1987) 75–263

[16] **J G Hocking**, **G S Young**, *Topology*, Addison–Wesley (1961)

[17] **J L Kelley**, *General topology*, Graduate Texts in Mathematics 21, Springer–Verlag (reprint of Van Nostrand edition 1955)

[18] **J Krasinkiewicz**, *Local connectedness and pointed 1–movability*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Atronom. Phys. 25 (1977) 1265–1269

[19] **G Levitt**, *Non-nesting actions on real trees*, Bull. London Math. Soc. 30 (1998) 46–54

[20] **I Martinez**, *Bord d'un produit amalgamé sur* **Z** *de deux groupes libres ou de surface*, Orsay preprint (1992)

[21] **M Mihalik**, *Semistability at* ∞ *of finitely generated groups, and solvable groups*, Topology and its Appl. 24 (1986) 259–264

[22] **M L Mihalik**, **S T Tschantz**, *Semistability of amalgamated products and HNN–extensions*, Memoirs Amer. Math. Soc. No. 471, Providence, Rhode Island (1992)

[23] **J P Otal**, *Certaines relations d'équivalence sur l'ensemble des bouts d'un groupe libre*, Journal London Math. Soc. 46 (1992) 123–139

[24] **E Rips**, **Z Sela**, *Cyclic splittings of finitely presented groups and the canonical JSJ decomposition*, Annals of Math. 146 (1997) 53–109

[25] **Z Sela**, *Structure and rigidity in (Gromov) hyperbolic groups and discrete groups in rank* 1 *Lie groups II*, Geom. Funct. Anal. 7 (1997) 561–593

[26] **H Short**, *Quasiconvexity and a theorem of Howson's*, from: "Group theory from a geometrical viewpoint", E Ghys, A Haefliger and A Verjovsky (editors), World Scientific (1991) 168–176

[27] **J R Stallings**, *Group theory and three–dimensional manifolds*, Yale Math. Monographs No. 4, Yale University Press, New Haven (1971)

[28] **G A Swarup**, *On the cut point conjecture*, Electron. Res. Announc. Amer. Math. Soc. 2 (1996) 98–100

[29] **E L Swenson**, *A cut point tree for a continuum*, preprint (1997)

[30] **J H C Whitehead**, *On certain sets of elements in a free group* Proc. London Math. Soc. 41 (1936) 48–56

Faculty of Mathematical Studies, University of Southampton
Highfield, Southampton SO17 1BJ, Great Britain

Email: `bhb@maths.soton.ac.uk`

Received: 15 November 1997 Revised: 10 August 1998

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 99–116

Controlled embeddings into groups that have no non-trivial finite quotients

Martin R Bridson

Abstract If a class of finitely generated groups $\mathcal{G}$ is closed under isometric amalgamations along free subgroups, then every $G \in \mathcal{G}$ can be quasi-isometrically embedded in a group $\widehat{G} \in \mathcal{G}$ that has no proper subgroups of finite index.

Every compact, connected, non-positively curved space X admits an isometric embedding into a compact, connected, non-positively curved space $\overline{X}$ such that $\overline{X}$ has no non-trivial finite-sheeted coverings.

AMS Classification 20E26, 20E06, 53C70; 20F32, 20F06

Keywords Finite quotients, embeddings, non-positive curvature

David Epstein's lucid writings, particularly those on automatic groups, had a strong influence on me when I was a graduate student. Since then, during many hours of enjoyable conversation, I have continued to benefit from his great insight into mathematics. It was therefore a great pleasure to speak at his birthday celebration and it is an equal pleasure to write an article for this volume.

0 Introduction

In this article I shall address the following general question: given a finitely generated group G that satisfies certain desirable properties, when can one embed G into a group which retains these desirable properties but does not have any non-trivial finite quotients? My interest in this question arises from a geometric problem that is the subject of Theorem C.

Our discussion begins with a general embedding theorem which is similar to results that were proved in the wake of the landmark paper by Higman, Neumann and Neumann [11]. The novel element in the result presented here is that we control the geometry of the embedding.

Theorem A *Let $\mathcal{G}$ be a class of finitely generated groups. If $\mathcal{G}$ is closed under the operation of isometric amalgamation along finitely generated free groups, then every $G \in \mathcal{G}$ can be quasi-isometrically embedded in a group $\widehat{G} \in \mathcal{G}$ that has no proper subgroups of finite index.*

The definition of isometric amalgamation is given in Section 1. There are various interesting classes of groups that are closed under amalgamations along arbitrary finitely generated free groups, for example the class of all finitely presented groups, groups of type F_n, and groups of a given (cohomological or geometric) dimension $n \geq 2$. The benefit of restricting the geometry of the amalgamation becomes apparent when the defining properties of $\mathcal{G}$ are more geometric in nature. For example, the class of groups which satisfy a polynomial isoperimetric inequality is not closed under the operation of amalgamation along arbitrary finitely generated free groups (or indeed along quasi-isometrically embedded free groups), but it is closed under amalgamation along isometrically embedded subgroups (Corollary 4.2).

A refinement of the proof of Theorem A yields:

Theorem B *Every finitely presented group G can be embedded in a finitely presented group $\widehat{G}$ that has no non-trivial finite quotients and whose Dehn function $f_{\widehat{G}}$ satisfies:*

$$f_{\widehat{G}}(n) \leq n f_G(n).$$

One can (simultaneously) arrange for the isodiametric function of $\widehat{G}$ to be no greater than that of G.

Theorem A does not apply directly to the class of groups that arise as fundamental groups of compact non-positively curved spaces.[1] Nevertheless, using a more subtle argument based on the same blueprint of proof, in Section 3 we shall prove the following theorem. (We say that a covering $\widehat{Z} \to Z$ is 'non-trivial' if $\widehat{Z}$ is connected and $\widehat{Z} \to Z$ is not a homeomorphism.)

Theorem C *Every compact, connected, non-positively curved space X admits an isometric embedding into a compact, connected, non-positively curved space $\overline{X}$ such that $\overline{X}$ has no non-trivial finite-sheeted coverings. If X is a polyhedral complex of dimension $n \geq 2$, then one can arrange for $\overline{X}$ to be a complex of the same dimension.*

[1] Throughout this article we use the term 'non-positive curvature' in the sense of A.D. Alexandrov [3].

Any local isometry between compact non-positively curved spaces induces an injection on fundamental groups [3, II.4], so in the notation of Theorem C we have $\pi_1 X \hookrightarrow \pi_1 \overline{X}$. Since $\overline{X}$ has no non-trivial finite-sheeted coverings, $\pi_1 \overline{X}$ has no proper subgroups of finite index. Thus Theorem C gives a solution to our general embedding problem for the class of groups that arise as fundamental groups of compact non-positively curved spaces. An extension of Theorem C yields the corresponding result for groups that act properly and cocompactly on CAT(0) spaces (3.6).

The fundamental groups of the most classical examples of non-positively curved spaces, quotients of symmetric spaces of non-compact type, are residually finite. In 1995 Dani Wise produced the first examples of compact non-positively curved spaces whose fundamental groups have no non-trivial finite quotients [21]. He also constructed semihyperbolic groups that are not virtually torison free, cf (3.7). Subsequently, Burger and Mozes [5] constructed compact non-positively curved 2-complexes whose fundamental groups are simple. Fundamental groups of compact negatively curved spaces, on the other hand, are never simple [8], [16].

One might hope to prove an analogue of Theorem A in which the enveloping group $\widehat{G}$ is simple. However the techniques described in this article are clearly inadequate in this regard. Indeed, finitely presented simple groups have solvable word problems and hence so do their finitely presented subgroups. Thus if one wishes to embed a given finitely presented group G into a finitely presented simple group, then one must make essential use of the fact that G has a solvable word problem. Higman conjectures that the solvability of the word problem is the only obstruction to the existence of such an embedding [10] (cf [4], [17]).

This article is organized as follows. In Section 1 we describe some examples of groups that are not residually finite and define isometric amalgamation. In Section 2 we prove Theorem A. In Section 3 we discuss spaces of non-positive curvature and prove Theorem C. In Section 4 we examine the effect of isometric amalgamations on isoperimetric and isodiametric inequalities and prove Theorem B.

This article grew out of a lecture which I gave at the conference on Geometric Group Theory at Canberra in July 1996. I would like to thank the organizers of that conference. I would particularly like to thank Chuck Miller for arranging my visit and for welcoming me so warmly.

1 Residual finiteness and isometric amalgamation

A group G is said to be *residually finite* if for every non-trivial element $g \in G$ there is a finite group Q and an epimorphism $\phi\colon G \twoheadrightarrow Q$ such that $\phi(g) \neq 1$. As a first step towards producing groups with no finite quotients, we must gather a supply of groups that are not residually finite. The Hopf property provides a useful tool in this regard. A group H is said to be *Hopfian* if every epimorphism $H \twoheadrightarrow H$ is an isomorphism — in other words, if $N \subset H$ is normal and $H/N \cong H$ then $N = \{1\}$.

The following result was first proved by Malcev [14].

1.1 Proposition *If a finitely generated group is residually finite then it is Hopfian.*

Proof Let G be a finitely generated group and suppose that there is an epimorphism $\phi\colon G \to G$ with non-trivial kernel. We fix $g_0 \in \ker\phi \smallsetminus \{1\}$ and for every $n > 0$ we choose $g_n \in G$ such that $\phi^n(g_n) = g_0$.

If there were a finite group Q and a homomorphism $p\colon G \to Q$ such that $p(g_0) \neq 1$, then all of the maps $\phi_n := p\phi^n$ would be distinct, because $\phi_n(g_n) \neq 1$ whereas $\phi_m(g_n) = 1$ if $m > n$. But there are only finitely many homomorphisms from any finitely generated group to any finite group (because the images of the generators determine the map). □

1.2 Examples The following group was discovered by Baumslag and Solitar [6]:

$$\mathrm{BS}(2,3) = \langle a, t \mid t^{-1}a^2t = a^3\rangle.$$

The map $a \mapsto a^2, t \mapsto t$ is onto: a is in the image because $a = a^3a^{-2} = (t^{-1}a^2t)a^{-2}$. However this map is not an isomorphism: $[a, t^{-1}at]$ is a non-trivial element of the kernel. Meier [15] noticed that the salient features of this example are present in many other HNN extensions of abelian groups. Some of these groups were later studied by Wise [19], among them

$$T(n) = \langle a, b, t_a, t_b \mid [a,b] = 1,\ t_a^{-1}at_a = (ab)^n,\ t_b^{-1}bt_b = (ab)^n\rangle,$$

which is the fundamental group of a compact non-positively curved 2-complex (see (3.1)). If $n \geq 2$ then certain non-trivial commutators, for example $g_0 = [t_a(ab)t_a^{-1}, b]$, lie in the kernel of the epimorphism $T(n) \twoheadrightarrow T(n)$ given by $a \mapsto a^n, b \mapsto b^n, t_a \mapsto t_a, t_b \mapsto t_b$. The proof of (1.1) shows that g_0 has trivial image in every finite quotient of $T(n)$.

1.3 Definition of Isometric Amalgamation Let $H \subset G$ be a pair of groups with fixed finite generating sets. If, in the corresponding word metrics, $d_G(h,h') = d_H(h,h')$ for all $h,h' \in H$, then we say that H is *isometrically embedded* in G.

Consider a finite graph of groups (in the sense of Serre [18]). If one can choose finite generating sets for the vertex groups G_i and the edge groups $H_{i,j}$ such that the inclusions of the edge groups are all isometric embeddings, then we say that the fundamental group Γ of the graph of groups is obtained by an *isometric amalgamation of the* G_i *along the* $H_{i,j}$ or, more briefly, Γ *is an isometric amalgam of the* G_i.

Note that, with respect to the natural choice of generators, all of the vertex and edge groups are isometrically embedded in the amalgam. Note also that, even in the basic cases of HNN extensions and amalgamated free products, the above definition is more stringent than simply requiring that for each i,j there exist choices of generators (depending on i,j) with respect to which $H_{i,j} \hookrightarrow G_i$ is an isometric embedding.

Free products of finitely generated groups are (trivial) examples of isometric amalgams. One can also obtain both $G \times \mathbb{Z}$ and $G * \mathbb{Z}$ from G by isometric amalgamations: each is the fundamental group of a graph of groups with one vertex group G and one edge group; to obtain $G \times \mathbb{Z}$ one takes G as edge group and uses the identity map as the inclusions; to obtain $G * \mathbb{Z}$ one takes the edge group to be trivial.

1.4 Lemma *Let $\mathcal{G}$ be as in Theorem A and let $T(n)$ be as in (1.2). If $G \in \mathcal{G}$ then $G * T(n) \in \mathcal{G}$.*

Proof Fix a finite generating set $\mathcal{S}$ for G. As above $G * \mathbb{Z} \in \mathcal{G}$; let a be a generator of the $\mathbb{Z}$ free factor. The cyclic subgroup generated by a is isometrically embedded with respect to the generating system $\mathcal{S} \cup \{a\}$. We add a further stable letter b that commutes with a, thus obtaining $G * \mathbb{Z}^2 \in \mathcal{G}$.

With respect to $\mathcal{S} \cup \{a, b, (ab)^n\}$, the cyclic subgroups generated by a, b and $(ab)^n$ are all isometrically embedded. Thus $G * T(n)$ can be obtained from $G * \mathbb{Z}^2$ by an isometric amalgamation: the underlying graph of groups has one vertex group, $G * \mathbb{Z}^2$, there are two edges in the graph and both edge groups are cyclic; the homomorphism at one end of each edge sends the generator to $(ab)^n$, and the maps at the other ends are onto $\langle a \rangle$ and $\langle b \rangle$ respectively. □

2 The proof of Theorem A

In order to clarify the exposition, we shall first prove a simplified version of Theorem A in which we do not examine the geometry of the amalgamations involved.

2.1 Lemma *Let $\mathcal{G}$ be a class of groups that is closed under the operation of amalgamation along finitely generated free groups. If $G \in \mathcal{G}$ is finitely generated, then it can be embedded in a finitely generated group $\widehat{G} \in \mathcal{G}$ that has no proper subgroups of finite index.*

Proof The following proof is chosen with Theorem A in mind (shorter proofs exist). A similar construction was used in [21].

Step 0 Replacing G by $G_0 = G * T(n)$ if necessary, we may assume that G contains an element of infinite order $g_0 \in G$ whose image in every finite quotient of G_0 is trivial (see (1.2)). Let $\{b_1, \dots, b_n\}$ be a generating set for G_0. We replace G_0 by $G_1 = G_0 * \mathbb{Z}$, and take as generators $\mathcal{A}' := \{t, b_1 t, \dots, b_n t\}$, where t generates the free factor $\mathbb{Z}$. We relabel the generators $\mathcal{A}' = \{a_0, \dots, a_n\}$.

Step 1 We take an HNN extension of G_1 with n stable letters:

$$E_1 = \langle G_1, s_0, \dots, s_n \mid s_i^{-1} a_i s_i = g_0^{p_i},\ i = 0, \dots, n \rangle.$$

where the p_i are any non-zero integers. Now, since each a_i is conjugate to a power of g_0 in E_1, the only generators of E_1 that can survive in any finite quotient are the s_i. However, since there is an obvious retraction of E_1 onto the free subgroup generated by the s_i, the group E_1 still has plenty of finite quotients.

Step 2 We repeat the extension process, this time introducing stable letters τ_i to make the generators s_i conjugate to g_0:

$$E_2 = \langle E_1, \tau_0, \dots, \tau_n \mid \tau_i^{-1} s_i \tau_i = g_0,\ i = 0, \dots, n \rangle.$$

Step 3 Add a single stable letter σ that conjugates the free subgroup of E_2 generated by the s_i to the free subgroup of E_2 generated by the τ_i:

$$E_3 = \langle E_2, \sigma \mid \sigma^{-1} s_i \sigma = \tau_i,\ i = 0, \dots, n \rangle.$$

At this stage we have a group in which all of the generators except σ are conjugate to g_0. In particular, every finite quotient of E_3 is cyclic.

Step 4 Because no power of a_0 lies in either of the subgroups of E_2 generated by the s_i or the τ_i, the normal form theorem for HNN extensions implies that $\{a_0, \sigma\}$ freely generates a free subgroup of E_3.

We define $\widehat{G}$ to be an amalgamated free product of two copies of E_3,

$$\widehat{G} = E_3 *_F \overline{E_3},$$

where $F = F(x, y)$ is a free group of rank two; the inclusion into E_3 is $x \mapsto a_0$ and $y \mapsto \sigma$, and the inclusion into $\overline{E}_3$ is $x \mapsto \overline{\sigma}$ and $y \mapsto \overline{a}_0$. All of the generators of $\widehat{G}$ are conjugate to a power of either g_0 or $\overline{g_0}$, and therefore cannot survive in any finite quotient. In other words, $\widehat{G}$ has no finite quotients. □

The following lemma enables us to gauge the geometry of the embeddings in the preceding construction.

2.2 Lemma *Let G be a group with finite generating set $\mathcal{A}$, where no $a \in \mathcal{A}$ represents $1 \in G$.*

(1) *In any HNN extension of G with finitely many stable letters $s_0, \dots, s_n$, the free subgroup generated by $S = \{s_0, \dots, s_n\}$ is isometrically embedded with respect to $\mathcal{A} \cup S$. If $\langle a \rangle \subset G$ is isometrically embedded and has trivial intersection with the amalgamated subgroups of s_i then $\mathrm{gp}\{a, s_i\}$ is isometrically embedded in the HNN extension.*

(2) *If $H \subset G$ is isometrically embedded with respect to $\mathcal{A}$, then H is also isometrically embedded in any isometric amalgamation involving G as a vertex group (provided the amalgamation is isometric with respect to the same generating set $\mathcal{A}$).*

(3) *Let $g \in G \smallsetminus \{1\}$. The cyclic subgroups of $G * \langle t \rangle$ generated by t, by $[g, t]$, and by each (at) with $a \in \mathcal{A}$, are all isometrically embedded with respect to the choice of generators $\mathcal{A}^* = \{at, [g, t], t \mid a \in \mathcal{A}\}$.*

Proof (1) and (2) follow from the normal form theorem for graphs of groups [18].

The normal form theorem for free products tells us that if we write $[g, t]^n$ as a word in the generators $\mathcal{A} \cup \{t\}$, then that word must contain at least $2n$ occurences of $t^{\pm 1}$. Each of the elements of $\mathcal{A}^*$ contains at most two occurences of $t^{\pm 1}$, therefore $d_{\mathcal{A}^*}(1, [g, t]^n) = n$.

If a word over $\mathcal{A} \cup \{t\}$ equals $(at)^n$ in $G * \langle t \rangle$, then its exponent sum in t must be n. Therefore, since each of the generators in $\mathcal{A}^*$ has t-exponent sum 1 or 0, we have $d_{\mathcal{A}^*}(1, (at)^n) = n$. □

2.3 The Proof of Theorem A We follow the proof of (2.1). What we must ensure is that at each stage the embedding which we described can be performed by means of an *isometric* amalgamation.

First we choose a finite generating set $\mathcal{A}$ for $G_0 = G * T(n)$ so that $G \hookrightarrow G_0$ is an isometric embedding, and we fix an element $g \in G_0$ whose image is trivial in every finite quotient of G_0. Then as generators for $G_1 = G_0 * \langle t \rangle$ we take $\mathcal{A}^* := \{at, [g,t], t \mid a \in \mathcal{A}\}$. Note the difference with (2.1) — we have included $[g,t]$. Define $g_0 = [g,t]$.

Lemma 2.2(3) assures us that the amalgamations carried out in Step 1 of the proof of (2.1) are along isometrically embedded subgroups provided that we take all $p_i = 1$. And parts (1) and (2) of Lemma 2.2 imply that the amalgamations carried out in Steps 2, 3 and 4 of (2.1) are also along isometrically embedded subgroups. Thus we obtain the desired group $\widehat{G} \in \mathcal{G}$ that has no finite quotients.

We have the inclusions $G \subset G_0 \subset G_1 \subset \widehat{G}$. The third inclusion was constructed to be an isometric embedding. The first and second inclusions are obviously isometric embeddings with respect to natural choices of generators. But it does not follow that $G \hookrightarrow \widehat{G}$ is an isometric embedding, because at the end of Step 0 of the proof we switched from the obvious set of generators for G_1 to a less natural set that was suited to our purpose. On the other hand, for any finitely generated group H, the identity map between the metric spaces obtained by endowing H with different word metrics is bi-Lipschitz. Thus, $G \subset \widehat{G}_0$ is a quasi-isometric embedding (with respect to any choice of word metrics). $\square$

For future reference we note:

2.4 Lemma *The cyclic subgroups generated by all of the stable letters introduced in the above construction are isometrically embedded in* $\widehat{G}$.

3 The non-positively curved case

The proof that we shall give of Theorem C is entirely self-contained except that we do not prove the basic facts about non-positively curved spaces that are listed (3.2). One could shorten the proof of Theorem C considerably by using the complexes constructed in [21] or [5] in place of Lemmas 3.3 and 3.5. However those constructions are rather complicated, so we feel that there is benefit in presenting a more direct account.

The example given in (4.3(2)) shows that the class of groups which act properly and cocompactly on spaces of non-positive curvature does not satisfy the conditions of Theorem A. Nevertheless, with appropriate attention to detail, one can use the blueprint of our proof of Theorem A to prove Theorem C, and this is what we shall do. First we need to know that there exists a compact non-positively curved 2-complex whose fundamental group is not residually finite.

3.1 Wise's Examples [19] Let

$$T(n) = \langle a, b, t_a, t_b \mid [a,b] = 1,\ t_a^{-1} a t_a = (ab)^n,\ t_b^{-1} b t_b = (ab)^n \rangle.$$

In Section 1 we saw that if $n \geq 2$ then this group is not Hopfian and therefore not residually finite. $T(n)$ is the fundamental group of the non-positively curved 2-complex $X(n)$ that one constructs as follows: take the (skew) torus formed by identifying opposite sides of a rhombus with sides of length n and small diagonal of length 1; the loops formed by the images of the sides of the rhombus are labelled a and b respectively; to this torus attach two tubes $S \times [0,1]$, where S is a circle of length n; one end of the first tube is attached to the loop labelled a and one end of the second tube is attached to the loop labelled b; in each case the other end of the tube wraps n times around the image of the small diagonal of the rhombus.

Any complex obtained by attaching tubes along local geodesics in the above manner is non-positively curved in the natural length metric (see [3, II.11]). We shall need the following additional facts concerning metric spaces of non-positive curvature; see [3] for details.

3.2 Proposition *Let X be a compact, connected, geodesic space of non-positive curvature. Fix $x \in X$.*

(1) *Each homotopy class in $\pi_1(X,x)$ contains a unique shortest loop based at x. This based loop is the unique local geodesic in the given homotopy class.*

(2) *Each conjugacy class in $\pi_1(X,x)$ is represented by a closed geodesic in X (ie a locally isometric embedding of a circle). In other words, every loop in X is freely homotopic to a closed geodesic (which need not pass through x). If two closed geodesics are freely homotopic then they have the same length.*

(3) *$\pi_1(X,x)$ is torsion-free.*

(4) *Metric graphs are non-positively curved.*

(5) *The induced path metric on the 1-point union of two non-positively curved spaces is again non-positively curved.*

(6) *If X is a compact non-positively curved space, Z is a compact length space and $i_1, i_2\colon Z \to X$ are locally isometric embeddings, then, when*

endowed with the induced path metric, the quotient of $X \cup (Z \times [0, L])$ *by the equivalence relation generated by* $i_1(z) \sim (z, 0)$ *and* $i_2(z) \sim (z, L)$ *is non-positively curved. Moreover, if* L *is greater than the diameter of* X*, then* X *is isometrically embedded in the quotient.*

A particular case of (6) that we shall need is where X is the disjoint union of spaces X_1 and X_2, and Z is a circle. In this case the quotient is obtained by joining X_1 to X_2 with a cylinder whose ends are attached along closed geodesics.

3.3 Lemma *There exists a compact, connected, non-positively curved 2-complex* K *with basepoint* $x_0 \in K$ *such that:*

(1) *there is an element* $g_0 \in \pi_1(K, x_0)$ *whose image in every finite quotient of* $\pi_1(K, x_0)$ *is trivial;*

(2) $\pi_1(K, x_0)$ *is generated by a finite set of elements each of which is represented by a closed geodesic that passes through* x_0 *and has integer length;*

(3) g_0 *is represented by a closed geodesic of length 1 that passes through* x_0.

Proof Let X be a compact, connected, 2-complex of non-positive curvature and let $g_0 \in \pi_1 X$ be a non-trivial element whose image in every finite quotient of $\pi_1 X$ is trivial (the spaces $X(n)$ of (3.1) give such examples). We choose a point x_0 on a closed geodesic that represents the conjugacy class of g_0. Suppose that $\pi_1(X, x_0)$ is generated by $\{b_1, \dots, b_n\}$, let β_i be the shortest loop based at x_0 in the homotopy class b_i, and let l_i be the length of β_i. Let l_0 be the length of the closed geodesic representing g_0. Replacing g_0 by a proper power if necessary, we may assume that $l_0 > l_i$ for $i = 1, \dots, n$.

Consider the following metric graph Λ: there are $(n+1)$ vertices $\{v_0, \dots, v_n\}$ and $2n$ edges $\{e_1, \varepsilon_1, \dots, e_n, \varepsilon_n\}$; the edge e_i connects v_0 to v_i and has length $(l_0 - l_i)/2$; the edge ε_i is a loop of length l_0 based at v_i. We obtain the desired complex K by gluing Λ to X, identifying v_0 with x_0, and then scaling the metric by a factor of l_0 so that the closed geodesic representing $g_0 \in \pi_1(K, x_0)$ has length 1.

Let $\gamma_i \in \pi_1(K, x_0)$ be the element given by the geodesic c_i that traverses e_i, crosses ε_i, and then returns along e_i, that is $c_i = e_i \varepsilon_i \overline{e}_i$, where the overline denotes reversed orientation. Note that $\pi_1(K, x_0)$ is the free product of $\pi_1(X, x_0)$ and the free group generated by $\{\gamma_1, \dots, \gamma_n\}$. As generating set for $\pi_1(K, x_0)$ we choose $\{b_i\gamma_i, b_i\gamma_i^2 \mid i = 1, \dots, n\}$.

According to parts (4) and (5) of the preceding proposition, K has non-positive curvature. Moreover, the concatenation of any non-trivial locally geodesic loop in X, based at x_0, and any non-trivial locally geodesic loop in Λ based at v_0 is a closed geodesic in K. Thus $\beta_i c_i$ and $\beta_i e_i \varepsilon_i^2 \overline{e}_i$ are closed geodesics in K; the former has length 2 and the latter has length 3; the former represents $b_i \gamma_i$ and the latter represents $b_i \gamma_i^2$. □

3.4 The proof of Theorem C Given a compact, connected, non-positively curved space X we must isometrically embed it in a compact, connected, non-positively curved space $\overline{X}$ whose fundamental group has no non-trivial finite quotients. Moreover the embedding must be such that if X is a complex of dimension at most $n \geq 2$ then so is $\overline{X}$. We give two constructions, the first in outline and the second in detail.

First Proof We form the 1-point union of X with one of the complexes $X(n)$ described in (3.1) thus ensuring that some element g_0 of the fundamental group has trivial image in every finite quotient. We then apply the construction of (3.3), gluing a metric graph to our space to obtain a space X' whose fundamental group is generated by elements represented by closed geodesics that pass through a basepoint on a closed geodesic representing g_0. To complete the proof one follows the argument of Lemma 3.5 with X' in place of K (taking the cylinders attached to be sufficiently long so that X is isometrically embedded in the resulting space, 3.2(6)).

Second Proof Choose a finite set of generators for $\pi_1 X$, and let $c_1, \ldots, c_N$ be closed geodesics in X representing the conjugacy classes of these elements. Lemma 3.5 gives a compact non-positively curved 2-complex K_4 whose fundamental group has no finite quotients; fix a closed geodesic c_0 in K_4. Take N copies of K_4 and scale the metric on the i-th copy so that the length of c_0 in the scaled metric is equal to the length $l(c_i)$ of c_i. Then glue the N copies of K_4 to X using cylinders $S_i \times [0, L]$ where S_i is a circle of length $l(c_i)$; the ends of $S_i \times [0, L]$ are attached by arc length parametrizations of c_0 and c_i respectively. Call the resulting space $\overline{X}$.

Part (6) of (3.2) assures us that $\overline{X}$ is non-positively curved, and if the length L of the gluing tubes is sufficiently large then the natural embedding $X \hookrightarrow \overline{X}$ will be an isometry.

It remains to construct K_4.

3.5 Lemma *There exists a compact non-positively curved 2-complex K_4 whose fundamental group has no finite quotients.*

Proof Let K be as in (3.3). We mimic the argument of (2.1), with $\pi_1(K, x_0)$ in the rôle of G_1. At each stage we shall state what the fundamental group of the complex being constructed is; in each case this is a simple application of the Seifert-van Kampen theorem.

Let c_0 be the closed geodesic of length 1 representing g_0. Let $\{a_0, \ldots, a_n\}$ be the generators given by 3.3(2), let α_i be the closed geodesic through x_0 that represents a_i, and suppose that α_i has length p_i. For each i, we glue to K a cylinder $S_{p_i} \times [0,1]$, where S_{p_i} is a circle of length p_i, with basepoint v_i; one end of the cylinder is attached to α_i while the other end wraps p_i-times around c_0, and $v_i \times \{0,1\}$ is attached to x_0. Let K_1 be the resulting complex. By the Seifert-van Kampen theorem, $\pi_1(K_1, x_0) = E_1$, in the notation of (2.1). Part (6) of (3.2) implies that K_1 is non-positively curved.

The images in K_1 of the paths $v_i \times [0,1]$ give an isometric embedding into K_1 of the metric graph Y that has one vertex and n edges of length 1; call the corresponding free subgroup $F_1 \subset E_1$ (it is the subgroup generated by the s_i in (2.1)).

Step 2 of (2.1) is achieved by attaching n cylinders of unit circumference $S_1 \times [0,1]$ to K_1, the ends of the i-th cylinder being attached to c_0 and to the image of $v_i \times [0,1]$. The resulting complex K_2 has $\pi_1(K_2, x_0) = E_2$. As in the previous step, the free subgroup $F_2 \subset E_2$ generated by the basic loops that run along the new cylinders is the π_1-image of an isometric embedding $Y \to K_2$. (This F_2 is the subgroup generated by the τ_i in (2.1).)

To achieve Step 3 of (2.1), we now glue $Y \times [0, L]$ to K_2 by attaching the ends according to the isometric embeddings that realize the embeddings $F_1, F_2 \subset \pi_1(K_2, x_0)$. This gives us a compact non-positively curved complex K_3 with fundamental group E_3 (in the notation of (2.1)). Let v be the vertex of Y, observe that $v \times \{0, L\}$ is attached to $x_0 \in K_3$, and let $\sigma \in \pi_1(K_3, x_0)$ be the homotopy class of the loop $[0, L] \to K_3$ given by $t \mapsto (v, t)$.

We left open the choice of L, the length of the mapping cylinder in Step 3, we now specify that it should be p_0, the length of the geodesic representing the generator a_0. An important point to observe is that the angle at x_0 between the image of $v \times [0, L]$ and any path in $K_1 \subset K_3$ is π. Thus the free subgroup $\mathrm{gp}\{a_0, \sigma\}$ is the π_1-image in $\pi_1(K_3, x_0)$ of an isometry from the metric graph Z with one vertex (sent to x_0) and two edges of length $L = p_0$. In fact, we have two such isometries $Z \to K_3$, corresponding to the free choice we have of which edge of Z to send to the image of $v \times [0, L]$. We use these two maps to realize Step 4 of the construction on (2.1): we apply part (6) of (3.2) with X equal to the disjoint union of two copies of K_3 and with the two maps $Z \to K_3$

employed as the local isometries i_1, i_2, the image of one of the maps being in each component of X. The resulting space is the desired complex K_4. □

By gluing non-positively curved orbi-spaces (in the sense of Haefliger [9]), or by performing equivariant gluing, one can extend Theorem C to include groups with torsion. We refer the reader to [3, II.11] for the technical tools that make this adaptation straightforward.

3.6 Theorem *If a group G acts properly and cocompactly by isometries on a $CAT(0)$ space Y then one can embed G in a group $\widehat{G}$ that acts properly and cocompactly by isometries on a $CAT(0)$ space $\overline{Y}$ and has no proper subgroups of finite index. If Y is a polyhedral complex of dimension $n \geq 2$ then so is $\overline{Y}$.*

Since the group G need not be torsion-free, (3.6) shows in particular that there exist compact non-positively curved orbihedra, with finite local groups, that are not finitely covered by any polyhedron (where 'covered' refers to covering in the sense of orbispaces and 'polyhedron' means an orbihedron whose local groups are trivial). We close our discussion of non-positively curved spaces with an explicit example to illustrate this point. The first examples of this type were discovered by my student Wise [20], and the following example is essentially contained in his work.

3.7 A semihyperbolic group that is not virtually torison-free

In the hyperbolic plane $\mathbb{H}^2$ we consider a regular quadrilateral Q with vertex angles $\pi/4$. Let α and β be hyperbolic translations that identify the opposite sides of Q. Then Q is a fundamental domain for the action of $G = \mathrm{gp}\{\alpha, \beta\}$; the commutator $[\alpha, \beta]$ acts as a rotation through π at one vertex of Q, and away from the orbit of this vertex the action of G is free. Thus the quotient orbifold $V = \mathbb{H}^2/G$ is a torus with one singular point, and at that singular point the local group is $\mathbb{Z}_2$.

Let $X(n)$ and $T(n)$ be as in (3.1) and fix a closed geodesic c in the homotopy class of a non-trivial element g_0 in the kernel of a self-surjection $T(n) \twoheadrightarrow T(n)$. We scale the metric on $X(n)$ so that this geodesic has length $l = |\alpha| = |\beta|$. Then we take a copy of $X(n)$ and consider the orbispace $\overline{V}$ obtained by gluing it to V using a tube $S_l \times [0,1]$ one end of which is glued to c and the other end of which is glued to the image in V of the axis of α.

$\overline{V}$ inherits the structure as a (non-positively curved) orbihedron in which the only singular point is the original one; at this singular point the local structure is as it was in V. The fundamental group $\widehat{G}$ of $\overline{V}$ is $G *_{\mathbb{Z}} T(n)$, where the amalgamation identifies $g_0 \in T(n)$ with $\alpha \in G$. Now, g_0 has trivial image in every finite quotient of $T(n)$, therefore $[\alpha, \beta] = [g_0, \beta]$ has trivial image in every finite quotient of $\widehat{G}$. It follows that $[\alpha, \beta]$, which has order two, lies in every subgroup of $\widehat{G}$ that has finite index.

In the case $n = 2$, the group $\widehat{G}$ has the following presentation:

$$\langle a, b, s, t, \alpha, \beta \mid \alpha = [s^{-1}(ab)s, b],\ [a, b] = [\alpha, \beta]^2 = 1,\ t^{-1}bt = s^{-1}as = (ab)^2\rangle.$$

4 Isoperimetric inequalities

Isoperimetric inequalities for finitely presented groups $G = \langle \mathcal{A} \mid \mathcal{R}\rangle$ measure the complexity of the word problem. If a word w in the free group $F(\mathcal{A})$ represents the identity in G, then there is an equality

$$w = \prod_{i=1}^{N} x_i^{-1} r_i x_i$$

in $F(\mathcal{A})$, where $r_i \in \mathcal{R}^{\pm 1}$. Isoperimetric inequalities give upper bounds on the integer N in a minimal such expression. The bounds are given as a function of the length of w, and the function $f_G\colon \mathbb{N} \to \mathbb{N}$ giving the optimal bound is called the *Dehn function* of the presentation. If there is a constant $K > 0$ such that the functions $g, h\colon \mathbb{N} \to \mathbb{N}$ satisfy $g(n) \leq K\, h(Kn) + Kn$, then one writes $g \preceq h$. It is not difficult to show (see [1] for example) that the Dehn functions of different finite presentations of a fixed group are $\simeq$ equivalent, where $f \simeq g$ means that $f \preceq g$ and $g \preceq f$.

As an alternative measure of complexity for the word problem, instead of trying to bound the integer N in the above equality one might seek to bound the length of the conjugating elements x_i. In this case the function giving the optimal bound is called the *isodiametric function* of the group, which we write $\Phi_G(n)$. Again, this function is $\simeq$ independent of the chosen presentation (see [7]).

We refer the reader to [7] for more information and references concerning Dehn functions and isodiametric functions and their (useful) interpretation in terms of the geometry of van Kampen diagrams.

4.1 Proposition *If G is an isometric amalgam of a finite collection $\{G_i \mid i \in I\}$ of finitely presented groups, then the Dehn function $f_G(n)$ of G is $\preceq n^2 + n\, \max_i f_{G_i}(n)$.*

Proof A diagrammatic version of the proof is given in (4.3(3)), here we present a more algebraic proof.

By definition, G is the fundamental group of a finite graph of groups. For the sake of notational convenience we shall assume that there are no loops in the graph of groups under consideration. The proof in the general case is entirely similar but notationally cumbersome.

Thus we have a finite tree with vertex set I and a set of edges $\mathcal{E} \subset I \times I$. At the vertex indexed i the vertex group is G_i. Let $H_{i,j}$ be the edge group associated to $(i,j) \in \mathcal{E}$. By definition, (1.3), there are finite generating sets $\mathcal{A}_i$ for the G_i and subsets $\mathcal{B}_{i,j} \subset \mathcal{A}_i$ with specified bijections $\phi_{i,j}\colon \mathcal{B}_{i,j} \to \mathcal{B}_{j,i}$ for each $(i,j) \in \mathcal{E}$; the set $\mathcal{B}_{i,j}$ generates $H_{i,j}$, each of the inclusions $H_{i,j} \hookrightarrow G_i$ is isometric with respect to these choices of generators, and $\phi_{i,j} = \phi_{j,i}^{-1}$.

We fix finite presentations $\langle \mathcal{A}_i \mid \mathcal{R}_i \rangle$ for the G_i. Then,

$$G \cong \langle \mathcal{A} \mid \mathcal{R},\ \phi_{i,j}(b) = b, \forall b \in \mathcal{B}_{i,j} \rangle,$$

where $\mathcal{A} = \coprod_i \mathcal{A}_i$, $\mathcal{R} = \coprod_i \mathcal{R}_i$, and (i,j) runs over $\mathcal{E}$

Let W be a word in the generators $\mathcal{A}$. Suppose that W is identically equal to a product $u_1 \dots u_m$, where each u_k is a word over one of the alphabets $\mathcal{A}_{i(k)}$ and each $\mathcal{A}_{i(k)} \neq \mathcal{A}_{i(k+1)}$. Under these circumstances W is said to have *alternating length* m. The normal form theorem for amalgamated free products [13] (or more generally graph products [18]) ensures that this notion of length is well-defined. It also tells us that if $W = 1$ in G then at least one of the subwords u_k is equal in $G_{i(k)}$ to a word ω in the generators $\mathcal{B}_{i(k),i(k\pm1)}$. Because $H_{i(k),i(k\pm1)}$ is isometrically embedded in $G_{i(k)}$, we can replace u_k by ω without increasing the length of W. This can be done at the cost of applying at most $f_{G_{i(k)}}(2|u_k|)$ relations. We apply $|\omega|$ relations to replace each letter b of ω with $\phi_{i(k),i(k\pm1)}(b)$. Then, without applying any more relations, we group ω together with the neighbouring word $u_{k\pm1}$. The net effect of this operation is to reduce the alternating length of W without increasing its actual length. By repeating this operation fewer than $|W|$ times we can replace W by a word W' with $|W'| \leq |W|$ that involves letters from only one of the alphabets $\mathcal{A}_i$. Since W' represents the identity in G_i, we can then reduce W' to the empty word by applying at most $f_{G_i}(|W'|)$ relators from $\mathcal{R}_i$.

The total number of relators applied in the reduction of W to W' is fewer than $m|W| + m \max_i f_{G_i}(|W|)$, where m is the alternating length of W. Therefore the total number of relators that we had to apply in reducing W to the empty word was less than $|W|^2 + |W| \max_i f_{G_i}(|W|)$. □

4.2 Corollary *The class of groups that satisfy a polynomial isoperimetric inequality is closed under the formation of isometric amalgamations along finitely generated subgroups.*

4.3 Remarks

(1) If instead of considering isometric amalgamations we considered the fundamental groups of graphs of groups in which the edge groups were only quasi-isometrically embedded, then the above proof would break down at the point where we noted that $|W'| \leq |W|$. In fact Proposition 4.1 would be false under this weaker hypothesis: consider the Baumslag-Solitar groups for example.

(2) Let D be the direct product of the free group on $\{a, b\}$ and the free group on $\{c, d\}$. Let $L = \text{gp}\{ac, bc\}$. For a suitable choice of generators, L is isometrically embedded in D. It is shown in [2] and [3] that $D *_L D$ has a cubic Dehn function, whereas D has a quadratic Dehn function. Thus, in general, isometric amalgamations may increase the polynomial degree of Dehn functions.

(3) The proof of (4.1) can be recast as an induction argument in which one proves that the area of a minimal van Kampen diagram for W is $m(\max_i f_{G_i}(|W|) + |W|)$, where m is the alternating length of W. This admits a simple geometric proof which we shall now sketch.

Draw a circle labelled by W, divide it into m subarcs according to the decomposition of W as an alternating word. Maintaining the notation established in the proof of (4.1), we draw a chord in the disc connecting the endpoints of the circular arc labelled by u_k. We label the chord by a geodesic word $\omega \in \mathcal{B}^*_{i(k),i(k\pm1)}$ that is equal to u_k in G. We fill the subdisc with boundary labelled $u_k\omega^{-1}$ using a minimal-area van Kampen diagram over the given presentation of $G_{i(k)}$. We then attach to the chord labelled ω faces corresponding to relators of the type $\phi_{i(k),i(k\pm1)}(b)$; the effect of this is to replace ω by the corresponding word in the generators $\mathcal{B}_{i(k\pm1),i(k)}$. By induction, we may fill the remaining subdisc with a van Kampen diagram of area no greater than $(m-1)(\max_i f_{G_i}(|W|) + |W|)$. We may choose u_k so that $2|u_k| \leq |W|$, and hence $|u_k| + |\omega| \leq |W|$. Therefore the area of the whole diagram is no greater than $m\,(\max_i f_{G_i}(|W|) + |W|)$, completing the induction.

A simple induction on alternating length, in the manner of (4.3(3)), allows one to show that (with respect to the finite presentations considered in (4.1)) every null-homotopic word W of alternating length m bounds a van Kampen diagram in which every vertex can be joined to the basepoint of the diagram by a path in the 1-skeleton that has length at most $|W| + \max_i \Phi_{G_i}(|W|)$. Thus:

4.4 Proposition *If G is an isometric amalgam of a finite collection $\{G_i \mid i \in I\}$ of finitely presented groups, then the isodiametric function $\Phi_G(n)$ of G is $\preceq \max_i \Phi_{G_i}(n)$.*

4.5 The Proof of Theorem B Given an infinite finitely presented group G, we replace it by $G * \mathbb{Z}$. This does not change the Dehn function or the isodiametric function of G but it allows us to assume that G is generated by a finite set of elements $\{a_i, \dots, a_r\}$ such that each $\langle a_i \rangle$ is isometrically embedded in G (see 2.2(3)).

The fundamental group S of any of the spaces $\overline{X}$ yielded by Theorem C will satisfy a quadratic isoperimetric inequality and a linear isodiametric inequality [3, III]. At the level of π_1, the proof of Theorem C was exactly parallel to that of (2.1), so Lemma 2.4 implies that S contains an isometrically embedded infinite cyclic subgroup $\langle s \rangle$.

The group $\widehat{G}$ whose existence is asserted in Theorem B is obtained by taking an amalgamated free product of G and m copies of S: the cyclic subgroup $\langle s \rangle$ in the i-th copy of S is identified with $\langle a_i \rangle \subset G$. In other words, $\widehat{G}$ is the fundamental group of a tree of groups in which there is one vertex of valence m, with vertex group G, and m vertices of valence 1, each with vertex group S; each edge group is infinite cyclic and the generator of the i-th edge group is mapped to $s \in S$ and $a_i \in G$.

Proposition 4.1 tells us that the Dehn function of $\widehat{G}$ is $\preceq n f_G(n)$, and Proposition 4.4 tells us that the isodiametric function of $\widehat{G}$ is no worse than that of G. ◻

References

[1] **J M Alonso**, *Inégalités isopérimétriques et quasi-isométries*, C.R.A.S. Paris Série 1, 311 (1990) 761–764

[2] **M R Bridson**, *Doubles, finiteness properties of groups, and quadratic isoperimetric inequalities*, J. Alg. to appear

[3] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, book preprint

[4] **W W Boone**, **G Higman**, *An algebraic characterization of the solvability of the word problem*, J. Austral. Math. Soc. 18 (1974) 41–53

[5] **M Burger**, **S Mozes**, *Finitely presented simple groups and products of trees*, C.R.A.S. Paris (1) 324.I (1997) 747–752

[6] **G Baumslag**, **D Solitar**, *Some two-generator one-relator non-Hopfian groups*, Bull. Amer. Math. Soc. 68 (1962) 199-201

[7] **S M Gersten**, *Isoperimetric and isodiametric functions of finite presentations*, from: “Geometric group theory, vol. 1”, LMS lecture notes 181 (G Niblo and M Roller, editors) Camb. Univ. Press (1993)

[8] **M Gromov**, *Hyperbolic groups*, from: “Essays in group theory”, (S M Gersten, editor) MSRI Publication 8, Springer–Verlag (1988) 75–263

[9] **A Haefliger**, *Complexes of groups and orbihedra* , from: “Group Theory From a Geometrical Viewpoint”, (E Ghys, A Haefliger, A Verjovsky, editors) World Scientific (1991) 504–540

[10] **G Higman**, *Finitely presented infinite simple groups*, Notes on Pure Maths. 8, Australian National University, Canberra (1974)

[11] **G Higman**, **B H Neumann**, **Hanna Neumann**, *Embedding theorems for groups*, J. London. Math. Soc. 24 (1949) 247–254

[12] **T Hsu**, **D Wise**, *Embedding theorems for non-positively curved polygons of finite groups* , J. Pure Appl. Alg. to appear

[13] **R C Lyndon**, **P E Schupp**, *Combinatorial group theory*, Springer–Verlag, Berlin (1977)

[14] **A I Malcev**, *On isomorphic matrix representations of infinite groups*, Mat. Sb. 8 (1940) 405–422

[15] **D Meier**, *Non-Hopfian groups*, J. London. Math. Soc. (2) 26 (1982) 265–270

[16] **A Yu Ol’shanskii**, *SQ universality of hyperbolic groups*, Mat. Sborn. 186 (1995) 119–132

[17] **P E Schupp**, *Embeddings into simple groups*, J. London. Math. Soc. 13 (1976) 90–94

[18] **J-P Serre**, *Trees*, Springer–Verlag, Berlin, Heidelberg, New York, 1980 Translation of “Arbres, Amalgames, SL_2”, Astérisque 46 (1977)

[19] **D T Wise**, *An automatic group that is not Hopfian*, J. Alg. 180 (1996) 845–847

[20] **D T Wise**, *Non-positively curved squared complexes, aperiodic tilings, and non-residually finite groups*, PhD Thesis, Princeton Univ. (1996)

[21] **D T Wise**, *A non-positively curved squared complex with no finite covers*, preprint (1995)

Mathematical Institute, 24–29 St Giles’, Oxford, OX1 3LB

Email: `bridson@maths.ox.ac.uk`

Received: 16 November 1997

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 117–125

All Fuchsian Schottky groups are classical Schottky groups

Jack Button

Abstract Not all Schottky groups of Möbius transformations are classical Schottky groups. In this paper we show that all Fuchsian Schottky groups are classical Schottky groups, but not necessarily on the same set of generators.

AMS Classification 20H10; 30F35, 30F40

Keywords Möbius transformation, Fuchsian group, Schottky group

1 Introduction

A Schottky group of genus g is a group of Möbius transformations acting on the Riemann sphere $\overline{\mathbb{C}}$ generated by g elements $A_i, 1 \leq i \leq g$, each of which possesses a pair of Jordan curves $C_i, C_i' \subseteq \overline{\mathbb{C}}$, with the property that the $2g$ curves are mutually disjoint and that A_i maps C_i onto C_i' where the outside of C_i is sent onto the inside of C_i'. Direct use of combination theorems tells us that the resulting group is free on g generators, is discrete with a fundamental domain the region exterior to the $2g$ curves, and consists entirely of loxodromic and hyperbolic elements.

If in addition we can take all the Jordan curves to be geometric circles then the resulting group is called a classical Schottky group (or sometimes in order to be more specific we say it is classical on the generators $A_1, \ldots, A_g$). Marden [2] showed that not all Schottky groups are classical Schottky groups. Put very briefly, he argued that the algebraic limit of classical Schottky groups must be geometrically finite and so his isomorphism theorem implies that the ordinary set Ω of this limit cannot be empty. But most groups on the boundary of Schottky space have an empty ordinary set, so Schottky space strictly contains classical Schottky space. However, this argument is certainly non-constructive, raising the question of finding an explicit nonclassical Schottky group. Zarrow [7] claimed to have found such an example, but the paper of Sato [5] shows that it is in fact a classical Schottky group. A little later Yamamoto [6] did construct a nonclassical Schottky group.

The purpose of this paper is to show that if we examine the most straightforward cases where we might expect to find a counterexample, namely Fuchsian Schottky groups, then this approach is doomed to failure as all such groups are classical Schottky groups. Specifically we show that:

(**1**) Given a Fuchsian Schottky group G of any genus g then there exists a generating set for G of g hyperbolic Möbius transformations on which G is classical.

(**2**) The Fuchsian Schottky group G is classical on all possible generating sets if and only if $g = 2$ and G is generated by a pair of hyperbolic elements with intersecting axes.

(**3**) There exists a Fuchsian group which is Schottky on a particular generating set, but which cannot be classical on those generators.

The author would like to thank the referee for comments on an earlier draft of this paper.

2 Proof of Main Theorem

Given any finitely generated Fuchsian group G (namely a discrete subgroup of $PSL(2,\mathbb{R})$) containing no elliptic elements, we form the quotient surface $S = U/G$ where U is the upper half plane. The complete hyperbolic surface S has ideal boundary $\partial S = (\overline{\mathbb{R}} \cap \Omega_G)/G$, where $\overline{\mathbb{R}}$ is the boundary of U in the Riemann sphere $\overline{\mathbb{C}}$ and Ω_G is the ordinary set of G. Note that G is Schottky if and only if S is a closed surface minus at least one hole (although S cannot be a one-holed sphere). This is because a Fuchsian group G with a quotient surface S as above must be free and purely hyperbolic, and this implies (see, say [3]) that G is indeed Schottky.

If S is a surface of genus n with h holes then G will be a free group of some rank r. The process of doubling S along its boundary corresponds to considering the quotient of the whole ordinary set Ω_G by G. As G is a Schottky group, Ω_G/G is topologically a closed surface of genus r. Therefore we conclude that $r = 2n + h - 1$ (with $n \geq 0, h \geq 1$ and $r \geq 1$).

The idea of the proof of theorem 1 is that given any such surface $S = U/G$, we find a particular reference surface, homeomorphic to S, which has a system of simple closed geodesics $\gamma_1, \ldots, \gamma_r$ corresponding to a generating set for G. We also find disjoint complete simple geodesics $l_1, \ldots, l_r$ on this reference surface which are properly embedded (they can be thought of as having their endpoints up the "spouts"), where l_i intersects γ_i once and is disjoint from γ_j $(j \neq i)$. We will find that if we cut along these geodesics $l_1, \ldots, l_r$, a disc is obtained. We are then able to transfer these curves across to S. By viewing the process upstairs in the upper half plane U we get a fundamental domain for G, and then we can see directly that G is classical Schottky on our generating set.

Theorem 1 *Given a Fuchsian Schottky group G of any genus g then there exists a generating set for G of g hyperbolic Möbius transformations on which G is classical.*

Proof We prove the result by taking a standard Fuchsian classical Schottky group $G_{n,h}$ for each possible topological surface of genus n and h holes, and transfer the two sets of geodesics to curves on any other surface homeomorphic to $U/G_{n,h}$. These can be replaced by geodesics with all necessary properties preserved.

First consider $h = 1$. We choose $2n$ hyperbolic elements $A_1, \ldots, A_{2n}$ so that their axes all intersect at the same point, and ensure that $G_{n,1} = \langle A_1, \ldots, A_{2n} \rangle$ is classical Schottky by choosing the multipliers of the A_i in order to obtain for each group $\langle A_i \rangle$ a fundamental domain Δ_i consisting of the intersection of the exteriors of two geodesics L_i and $L_i' = A_i(L_i)$ so that all conditions of the free product combination theorem are satisfied; namely that

$$\Delta_i \cup \Delta_j = U \text{ for } i \neq j \text{ and } \bigcap_i \Delta_i \neq \emptyset.$$

Then we have a fundamental domain $\Delta_{n,1}$ (homeomorphic to a disc) for the discrete group $G_{n,1}$. There is one cycle of boundary intervals and so by the discussion above, the surface $S_{n,1} = U/G_{n,1}$ is indeed of genus n with boundary a circle.

We can project the axes of A_i down onto the surface to obtain our simple closed geodesics γ_i, and do the same with each L_i, which gives us the complete simple geodesic l_i right up to its two endpoints on the boundary. These have the appropriate properties mentioned earlier, and we see that the surface becomes a disc after cutting along all the geodesics $l_1, \ldots l_{2n}$.

The group $G_{2,1}$ and the projection of these geodesics are illustrated in figures 1 and 2.

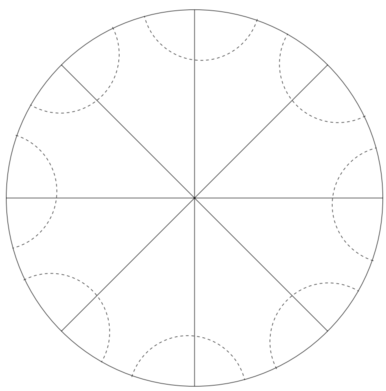


Figure 1

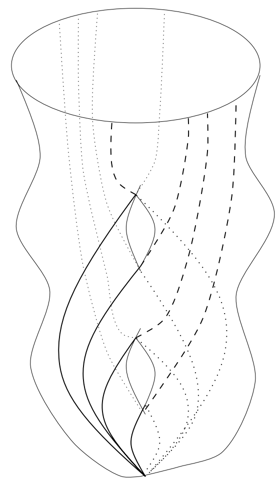


Figure 2

In order to construct $G_{n,h}$ when $h \geq 2$, take $G_{n,1}$ and choose an open interval I between one endpoint of some L_i and the nearest endpoint of a neighbouring geodesic L_j. This interval lies inside the ordinary set of $G_{n,1}$. Then inductively nest $h-1$ geodesics inside the previous one, so that each geodesic has endpoints in I. We then find hyperbolic transformations $A_{2n+1}, \ldots, A_{2n+h-1}$ with axes these geodesics and with each transformation having two geodesics L_i and $L'_i = A_i(L_i)$, where $2n+1 \leq i \leq 2n+h-1$, which it pairs. If these fundamental domains are correctly placed then $G_{n,h} = \langle A_1, \ldots A_{2n+h-1} \rangle$ is a discrete group having the correct quotient surface $S_{n,h} = U/G_{n,h}$ with a disc for a fundamental domain $\Delta_{n,h}$, where $\partial\Delta_{n,h}$ consists of $4n+2h-2$ geodesics L_i and L'_i, along with the same number of intervals of $\overline{\mathbb{R}}$. The geodesics and intervals alternate as we go round the boundary of the disc. Also the projections of these axes and of these paired geodesics which define γ_i and l_i have all the same properties as mentioned before. The case $n = 1$, $h = 5$ is pictured in figures 3 and 4.

Now given any Fuchsian Schottky group G with quotient surface S and boundary ∂S, there exists a homeomorphism

$$h\colon S_{n,h} \cup \partial S_{n,h} \mapsto S \cup \partial S$$

for some n and h. We also have natural continuous projections

$$\begin{array}{llll} p\colon & U \cup (\Omega_{G_{n,h}} \cap \overline{\mathbb{R}}) & \mapsto & S_{n,h} \cup \partial S_{n,h} \\ q\colon & U \cup (\Omega_G \cap \overline{\mathbb{R}}) & \mapsto & S \cup \partial S \end{array}$$

where p and q are both covering maps, and both domains are simply connected

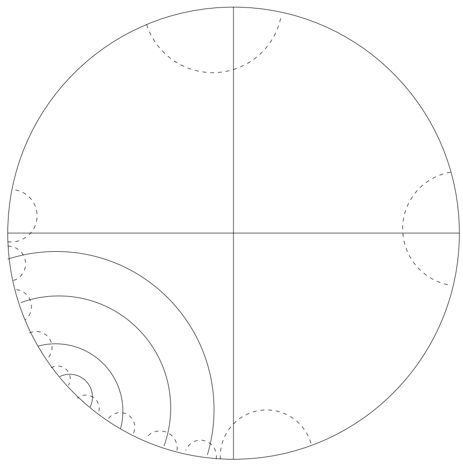


Figure 3

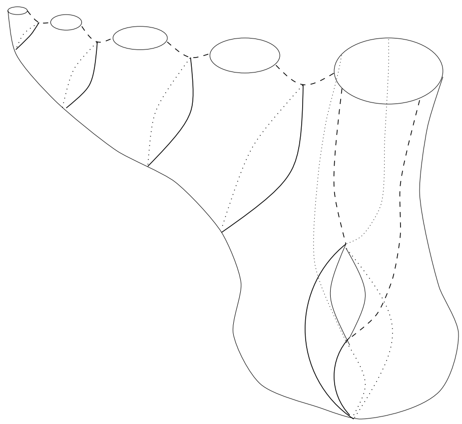


Figure 4

covering spaces of their images (where the elementary neighbourhoods of points downstairs are open discs, or half discs for points on the boundary).

By the lifting theorem, we have a continuous map

$$H\colon U \cup (\Omega_{G_{n,h}} \cap \overline{\mathbb{R}}) \mapsto U \cup (\Omega_G \cap \overline{\mathbb{R}})$$

which is a lift of hp, so that $hp = qH$. By reversing p and q, we see that H is a homeomorphism.

Take any element $g \in G_{n,h}$. This is a deck transformation of p and so $pg = p$. Conjugating g by H, we have $q(HgH^{-1}) = q$, thus HgH^{-1} is a deck transformation of q and therefore H defines an isomorphism of $G_{n,h}$ onto G by conjugation.

Note that H maps U to U and $\Omega_{G_{n,h}} \cap \overline{\mathbb{R}}$ to $\Omega_G \cap \overline{\mathbb{R}}$, because it is a lift of h which sends boundary points to and from boundary points. Therefore the image under H of the fundamental domain $\Delta_{n,h}$ is a disc in U. But $H(\partial\Delta_{n,h})$ will consist of $4n+2h-2$ disjoint closed intervals of $\overline{\mathbb{R}}$, along with curves $H(L_i)$ and $H(L_i')$ lying entirely in U apart from their endpoints which are also endpoints of these intervals of $\overline{\mathbb{R}}$. We find that the order in which the images under H of the L_i, L_i' and the intervals appear around $\partial H(\Delta_{n,h}) = H(\partial\Delta_{n,h}) \subseteq U \cup (\Omega_G \cap \overline{\mathbb{R}})$ is the same as the original order around $\partial\Delta_{n,h}$ (or the opposite order if H is orientation reversing).

By setting $B_i = HA_iH^{-1}$ we obtain a generating set for G, and because A_i sends the geodesic L_i to L_i', we see that B_i sends the curve $H(L_i)$ to the curve

$H(L_i')$. Also it is easy to check that the disc $H(\Delta_{n,h})$ is a fundamental domain for the action of G on U. In particular, the intersection of the exteriors in U of $H(L_i)$ and $H(L_i')$ is a fundamental domain for $\langle B_i\rangle$. We replace these two curves by geodesics M_i and $M_i' = B_i(M_i)$ which have the same endpoints. Just as in [1], this gives us $2n+h-1$ pairs of geodesics freely homotopic to the curves they replaced, and paired by a generating set B_i with another fundamental domain D_i for each group $\langle B_i\rangle$ that lies between these two geodesics. The free product combination theorem can be applied to $\langle B_1\rangle, \dots \langle B_{2n+h-1}\rangle$, as $D_i \cup D_j = U$ for $i \neq j$ and $\bigcap_i D_i \neq \emptyset$. We can see this by looking at the endpoints of the geodesics which have not been changed when passing from curves. Therefore, by reflecting this picture in the real axis, the group G is generated by elements B_i, each of which possesses a pair of mutually disjoint geometric circles C_i and C_i', with the outside of C_i being sent by B_i onto the inside of C_i'. By definition, G is a classical Schottky group. □

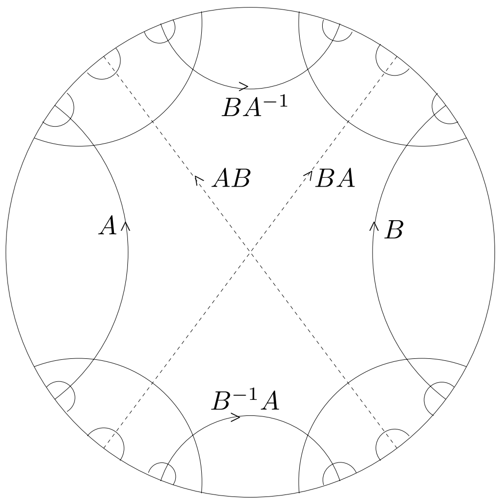


Figure 5

3 Proof of other Theorems

Suppose we are given any two hyperbolic elements A and B with different axes. We want to know when $G = \langle A, B\rangle$ is free, discrete and purely hyperbolic (hence Schottky). This problem falls naturally into two cases.

(**A**) The two hyperbolic elements have intersecting axes. Then it is well known that G is free, discrete and purely hyperbolic if and only if the commutator

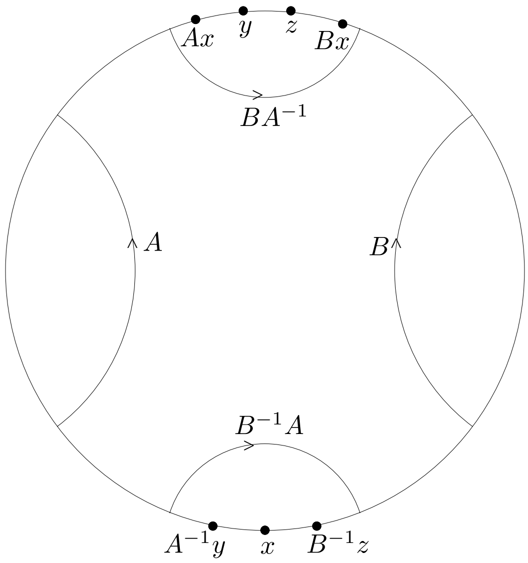


Figure 6

$ABA^{-1}B^{-1}$ is hyperbolic. See for instance [4] where this is shown by explicitly exhibiting two pairs of geometric circles, one paired by A and one by B. In this case the quotient surface is a one holed torus and, as any generating pair will have intersecting axes, we see that G is classical on every possible generating pair.

Alternatively we can see this directly from section 1 by using the fact that there will exist a homeomorphism from our standard surface to the quotient surface of G that takes the two simple closed geodesics γ_1, γ_2 onto two curves freely homotopic to the simple closed geodesics corresponding to any generating pair of G.

(**B**) The hyperbolic elements have non-intersecting axes. If so then all generating pairs of G must have non-intersecting axes, or else we are back in case (A).

First suppose G is a classical Schottky group on these two generators A and B. Without loss of generality we can replace any generator by its inverse so that we get a picture such as the one in figure 5, with the arrows on the two generators in the same direction. The quotient surface is a three holed sphere. Note that the axis of AB projects down onto a "figure of eight" geodesic, and so this group cannot be classical on the generating pair $\langle A, AB\rangle$.

Theorem 2 *A group G that has a quotient surface which is not a one holed* *torus cannot be classical on all generating sets.*

Proof We have already considered any G generated by two elements. Given any G generated by three or more elements, we can find a pair of generators with non-intersecting axes, and use the above argument on the subgroup generated by this pair. As the subgroup is not classical on all generating sets, nor is G. □

Finally we show the existence of a Fuchsian group generated by two elements which is Schottky, but not classical, on this generating pair.

Lemma 1 *A group $G = \langle A, B\rangle$ (where A and B are hyperbolic elements with non-intersecting axes, oriented as in figure 5) is classical on $\langle A, B\rangle$ if and only if both fixed points of $B^{-1}A$ lie in the interval between the repelling fixed points of A and B.*

Proof If we know G is classical on $\langle A, B\rangle$ then we can build up a pattern of nested circles as in figure 5, and see the location of the fixed points of the axes directly. Conversely if we only have information as in figure 6 then we consider the image of a suitable point x under the generators.

The axis of $B^{-1}A$ is sent to the axis of BA^{-1} by both generators, and also note that the arrows on BA^{-1} and $B^{-1}A$ are as in the picture (for instance consider the image of a fixed point of A). Then we choose any x inside the interval enclosed by the axis of $B^{-1}A$, and mark it and its images under A and B. We can take any two points y and z in the interval between Ax and Bx, and use these as endpoints for the geometric circles we require.

We can see that $A^{-1}y$ will be closer than x to the repelling fixed point of A, and similarly with $B^{-1}z$ and B. This gives us four endpoints $y, z, A^{-1}y$ and $B^{-1}z$, one for each circle. We have four more endpoints to mark but this choice is totally arbitrary: merely pick any point in the interval between A's fixed points, along with its image under A, and do the same for B too. This provides us with our two pairs of circles which show that G is discrete, and classical on $\langle A, B\rangle$. □

Theorem 3 *The Fuchsian group in figure 7, which is Schottky on the generators A and B, is not classical on them.*

Proof The exterior F of the two pairs of curves C_A, C'_A (paired by A) and C_B, C'_B (paired by B) is a fundamental domain, and is sent by the element BA^{-1} inside the circle $C(= B(C_A))$. The attracting fixed point of BA^{-1} must lie inside C and therefore it separates the fixed points of A. □

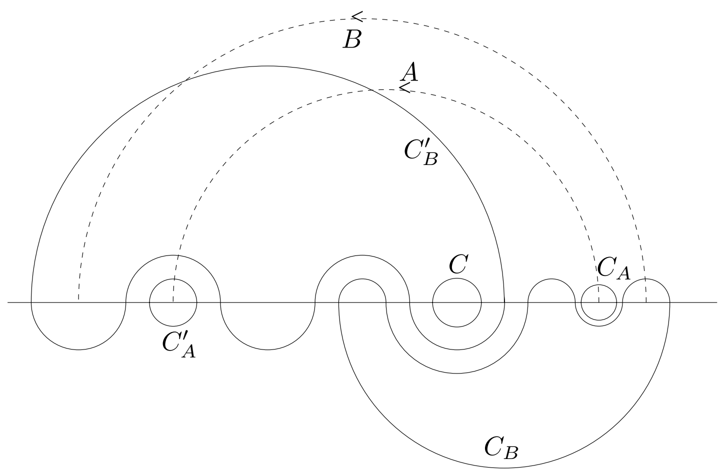


Figure 7

References

[1] **M Dehn**, **J Stillwell**, *Appendix: The Dehn–Nielsen Theorem*, from: "Papers on Group Theory and Topology", Springer–Verlag (1987) 363–395

[2] **A Marden**, *The geometry of finitely generated Kleinian groups*, Ann. of Math. 99 (1974) 383–462

[3] **B Maskit**, *A characterisation of Schottky groups*, J. Analyse Math. 19 (1967) 227–230

[4] **N Purzitsky**, *Two–Generator Discrete Free Products*, Math. Z. 126 (1972) 209–223

[5] **H Sato**, *On a paper of Zarrow*, Duke Math. J. 57 (1988) 205–209

[6] **H Yamamoto**, *An example of a nonclassical Schottky group*, Duke Math. J. 63 (1991) 193–197

[7] **R Zarrow**, *Classical and nonclassical Schottky groups*, Duke Math. J. 42 (1975) 717–724

Wadham College, Oxford, OX1 3PN, UK

Email: `button@maths.ox.ac.uk`

Received: 13 May 1998 Revised: 15 October 1998

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997 **127**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 127–138

On the Burau representation modulo a small prime

D Cooper
D D Long

Abstract We discuss techniques for analysing the structure of the group obtained by reducing the image of the Burau representation of the braid group modulo a prime. The main tools are a certain sesquilinear form first introduced by Squier and consideration of the action of the group on a Euclidean building.

AMS Classification 20F36; 57M07 57M25

Keywords Burau representation, braid group, Euclidean building, Squier form

1 Introduction

Despite the work of many authors, the group theoretic image of linear representations of the braid groups remains mysterious in most cases. The first nontrivial example, the *Burau representation* is not at all well understood. This representation

$$\beta_n\colon B_n \to GL(n-1, \mathbf{Z}[t, t^{-1}])$$

is known not to be faithful for $n \geq 6$ ([5] and [6]) but the nature of the image group and in particular, a presentation for the image group has not been found. In [3], we simplified the problem by composing β_n with the map which reduces coefficients modulo 2. In this way, we were able to give a presentation for the image of the simplified representation $\beta_4 \otimes \mathbf{Z}_2$. (Throughout this paper we use the notation $\mathbf{Z}_p$ for the finite field with p elements.) Of course, the motivation for this approach comes from the classical problem of whether the representation β_4 is faithful and to this end we pose the question:

Question 1.1 Is there any prime p for which the representation

$$\beta_4 \otimes \mathbf{Z}_p\colon B_4 \to GL(3, \mathbf{Z}_p[t, t^{-1}])$$

is faithful?

It is a consequence of some results of this note that the representation is not faithful in the case $p=3$, (below we exhibit a braid word in the kernel) however the program for attacking the problem runs into difficulty at the final stage when $p=5$. This case remains open and has some features which suggest it may be different to the first two primes.

In order to describe our approach, we recall that the group $GL(3,\mathbf{Z}_p[t,t^{-1}])$ acts on a certain contractible two dimensional simplicial complex, $\Delta=\Delta(p)$ a so-called *Euclidean building* (see [2]). This is defined by embedding

$$GL(3,\mathbf{Z}_p[t,t^{-1}]) \longrightarrow GL(3,\mathbf{Z}_p(t))$$

where $\mathbf{Z}_p(t)$ is the field of fractions of the ring $\mathbf{Z}_p[t,t^{-1}]$. This target group admits a discrete rank one valuation defined by $\nu(p/q)=degree(q)-degree(p)$. A standard construction now yields the complex Δ. We briefly outline how this building and action are defined, restricting our attention to the case $n=4$, since this is the only case in which we shall subsequently be interested. This will serve the additional purpose of establishing notation. Standard properties of ν imply that

$$\mathcal{O}=\{\, x\in\mathbf{Z}_p(t) \;\mid\; \nu(x)\geq 0\,\}$$

is a subring of $\mathbf{Z}_p(t)$, the *valuation ring* associated to ν. This is a local ring and the unique maximal ideal is easily seen to be $\mathcal{M}=\{\, x\in\mathbf{Z}_p(t) \;\mid\; \nu(x)>0\,\}$, a principal ideal. Choose some generator π for this ideal. This element is called a *uniformizing parameter* and by construction we have that $\nu(\pi)=1$. Since $\mathcal{M}$ is maximal, the quotient $k=\mathbf{Z}_p(t)/\mathcal{M}$ is a field, the *residue class field*. One sees easily that in this case, the residue class field is $\mathbf{Z}_p$.

Now let V be the vector space $\mathbf{Z}_p(t)^3$. By a *lattice* in V we shall mean an $\mathcal{O}$–submodule, L, of the form $L=\mathcal{O}x_1\oplus\mathcal{O}x_2\oplus\mathcal{O}x_3$ where $\{x_1,x_2,x_3\}$ is some basis for V. Thus the columns of a non-singular 3×3 matrix with entries in $\mathbf{Z}_p(t)$ defines a lattice. The standard lattice is the one corresponding to the identity matrix. We define two lattices L and L' to be equivalent, if for some $\lambda\in\mathbf{Z}_p(t)^*$ we have $L=\lambda L'$. We denote equivalence class by $[L]$. The building Δ is defined as a flag complex in the following way. The points are equivalence classes of lattices, and $[L_0],....,[L_k]$ span a k–simplex (in our situation $k=0,1,2$ are the only possibilities) if and only if one can find representatives so that $\pi L_0\subset L_1\subset ...\subset L_k\subset L_0$.

All 2–simplices are of the form $\{[x_1,x_2,x_3],[x_1,x_2,\pi x_3],[x_1,\pi x_2,\pi x_3]\}$; this is usually referred to as a *chamber* and denoted by C. Clearly the group $GL_3(\mathbf{Z}_p(t))$ acts on lattices and one sees easily that incidence is preserved, so that the group acts simplicially on Δ. It is shown in [2] that this building is a so-called *Euclidean building*, in particular, it is contractible and can be equipped with a metric which makes it into a $CAT(0)$ space and for which $GL_3(\mathbf{Z}_p(t))$ acts as a group of isometries. The metric is such that each 2 dimensional simplex is isometric to a unit Euclidean equilateral triangle.

We now return to our situation. One of the difficulties of dealing with representations of braid groups is that it is extremely difficult to determine exactly which matrices are in the image. We bypass this by dealing with a group which contains $im(\beta_4 \otimes \mathbf{Z}_p)$. To define this group, we recall that it was shown by Squier [7] that the Burau representation is unitary in the sense that there is a matrix

$$J = \begin{pmatrix} -(s+1/s) & 1/s & 0 \\ s & -(s+1/s) & 1/s \\ 0 & s & -(s+1/s) \end{pmatrix}$$

with the property that $A^*JA = J$ for all $A \in im(\beta_n)$. Here the involution $*$ comes from extending the involution of $\mathbf{Z}_p[t,t^{-1}]$ generated by $t \to 1/t$ to the matrix group by $(a_{i,j})^* = (a^*_{j,i})$, where $s^2 = t$.

We define the subgroup $Isom_J(\Delta)$ of $GL(3,\mathbf{Z}_p(t))$ to be those matrices with Laurent polynomial entries which are unitary for the form J. The advantage of dealing with this subgroup is that the condition that a matrix lies inside $Isom_J(\Delta)$ is easily used.

The strategy now is to examine the action of $Isom_J(\Delta)$ on Δ. This is interesting in its own right. Moreover, the greater ease of dealing with this subgroup means that we are able to compute the complex $\Delta/Isom_J(\Delta)$ together with all vertex, edge and 2–simplex stabilisers. We then appeal to results of Haefliger [4] to compute a presentation for the group $Isom_J(\Delta)$.

Now recall that homotheties act trivially on Δ so that the presentation for $Isom_J(\Delta)$ is to be compared with the following presentation of $B_4/centre(B_4)$:

Lemma 1.2 *The group $B_4/centre(B_4)$ is presented as*

$$\langle\, x, y \ \mid\ x^4 = y^3 = 1 \quad [x^2, yxy] = 1 \,\rangle$$

where $x = \sigma_1\sigma_2\sigma_3$ and $y = x\sigma_1$.

This is presumably well known to the experts—it is derived in [3]. The starting point for this work is:

Lemma 1.3 *The group $stab_J(I)$ acts on Δ as a finite group.*

Sketch of proof If $A \in stab_J(I)$, then its action on Δ is unchanged by homothety and it's easily seen that we can adjust any such A by applying $\pm t^k$ so that $A \in SL(3, \mathcal{O})$. Rewriting the unitary condition as $A^* = JA^{-1}J^{-1}$ and noting that $J \in GL(3, \mathcal{O})$, we see that $A^* \in SL(3, \mathcal{O})$. However the only matrices with Laurent polynomial entries for which A and A^* have all entries valuing positively are the constant matrices.

Thus we have shown that the only such A have constant entries up to homothety. In particular, they are unchanged by setting $t = 1$, so that $stab_J(I)$ can be regarded as a subgroup of the finite group $GL(3, \mathbf{Z}_p)$, completing the proof. □

This has the immediate corollary:

Corollary 1.4 *For every vertex $v \in \Delta$, $stab_J(v)$ is a finite group.*

Proof The building Δ is locally finite, in fact the link of every vertex is the flag manifold in the vector space $\mathbf{Z}_p^3$. The stabiliser of any vertex acts on this set as a group of permutations, so by passing to a subgroup of finite index in $stab_J(v)$ we obtain a subgroup which acts as the identity on all vertices in the link. Since every vertex is connected to I by some chain of vertices, we see that for every v, there is a subgroup of finite index which lies inside $stab_J(I)$, a finite group. □

We now focus on the case $p = 3$. In this case one finds by calculation:

Theorem 1.5 *At the prime 3, group $stab_J(I)$ acts on Δ as $\mathbf{Z}_4 \cong \langle x \rangle$.*

Remark 1.6 For $p = 2, 3, 5$, the group $stab_J(I)$ acts as the cyclic group $\mathbf{Z}_4$. For $p = 7$ it is cyclic of order 8 and for $p = 11$, cyclic of order 12.

One important difference between the case $p = 2$ and that of the larger primes is that it is one of the consequences of the results of [3] that $Isom_J(\Delta(2)) \cong im(\beta_4 \otimes \mathbf{Z}_2)$, this is not so for (at least some and conjecturally all) primes $p \geq 3$. In particular, for $p = 3$, we are able to construct (see below) an element $u \in Isom_J(\Delta(3))$ which has order 6; it is easy to see that this element does not lie in the subgroup $im(\beta_4 \otimes \mathbf{Z}_3)$. Its matrix is given by:

$$u = \begin{pmatrix} 2+t+t^2 & 2+t^2 & 2+2t+2t^2 \\ 2+2t^2 & 2+t+2t^2 & 2+t+t^2 \\ 2+t & 2+t & 2+2t \end{pmatrix}$$

However, having noted this difference, the qualitative picture of the quotient complex is very similar to the case $p = 2$; the complex consists of a compact piece coming from behaviour of groups close to the identity lattice, together with a single annular end. Application of Haefliger's methods yields the following group theoretic result:

Theorem 1.7 *When $p = 3$, the group $Isom_J(\Delta)$ is presented as:*

Generators: x, y, u

Relations:

(1) $x^4 = y^3 = u^6 = 1$

(2) $[x^2, yxy] = 1$

(3) $[x, u^{-1}x^{-1}y^{-1}xyxy] = 1$

(4) $[yxy, u^{-1}x^{-1}y^{-1}xyxy] = 1$

(5) $[xyx, u^2] = 1$

(6) $[x^2yx, u^3] = 1$

(7) $(u^2x^2yx)^2 = (x^2yxu^2)^2$

(8) *Infinitely many other relations to do with nilpotence.*

Of course the verification that these relations hold is a trivial matter of multiplying matrices modulo 3. We remark that the relations contained in (8) are explicitally known.

We claim that a computer application of the Reidemeister–Schreier algorithm contained in the computer program *GAP* applied to the presentation involving the first seven relations proves:

Corollary 1.8 *The index $[Isom_J(\Delta) : \langle x, y\rangle]$ is finite.*

This index is a divisor of 162. The corollary already implies that $im(\beta_4 \otimes \mathbf{Z}_3)$ is not faithful. One way to see this is that one sees easily (for example from the matrix representation) that the element $w = u^{-1}x^{-1}y^{-1}xyxy$ has infinite order. The presentation implies that it commutes with x. However, since $[Isom_J(\Delta) : \langle x, y\rangle]$ is finite, some power of w lies in the subgroup generated by x and y and this gives an unexpected element commuting with x. Alternatively, in the course of the proof, one discovers that $Isom_J(\Delta)$ contains arbitrarily large soluble subgroups and this can also be used to show that the representation is not faithful. In fact, one can be more specific; the computer can be used to give a presentation for the subgroup generated by $\langle x, y\rangle$; one finds for example, that there is a relation (where $\overline{x} = x^{-1}$ and $\overline{y} = y^{-1}$):

$$\overline{x}.\overline{y}.\overline{x}.y.x.\overline{y}.x.y.x.\overline{y}.x.\overline{y}.\overline{x}.y.x.y.\overline{x}.\overline{y}.\overline{x}.\overline{y}.\overline{x}.y.x.\overline{y}.\overline{x}.\overline{y}.\overline{x}.y.x.\overline{y}.x.y.x.\overline{y}.x.\overline{y}.$$
$$\overline{x}.y.x.y.x.\overline{y}.\overline{x}.\overline{y}.x.y.x.\overline{y}.\overline{x}.\overline{y}.\overline{x}.y.x.\overline{y}.x.y.x.\overline{y}.x.\overline{y}.\overline{x}.y.x.y.x.\overline{y}.\overline{x}.\overline{y}.x.y.x.\overline{y}$$

That this relation does not hold in the braid group is easily checked by computing the integral Burau matrix.

2 Outline of the proof for $p = 3$

In spirit, if not in detail, the proof follows the ideas introduced in [3], to which we refer the reader. We work outwards from the identity lattice, successively identifying point stabilisers. This enables us to find representatives for each orbit and hence build the quotient complex. The compact part alluded to above comes from the action of the group on vertices fairly close to the orbit of the identity; as one moves farther away there is a certain amount of stabilisation and it is this which gives rise to the single annular end.

We refer to the orbit of the lattice I as the *group points*. The result Lemma 1.3 shows that every group point has stabiliser $\mathbf{Z}_4$. We recall the link of any vertex may be considered as the flag geometry of the vector space $\mathbf{Z}_3^3$, so that every vertex has 26 points in its link, and each vertex in the link is adjacent to four other vertices in the link.

We need to recall the notation introduced in [3]. We make a (noncanonical) choice of representative lattices for each of the 26 vertices by writing down matrices whose columns define the lattice. Subsequent vertices are coded by using these matrices, regarded as elements of $GL(3, \mathbf{Z}_3[t, t^{-1}])$ as acting on Δ. As an example, denoting the matrix representative chosen for the thirteenth vertex by M_{13}, then the representative elements in the link of the the thirteenth vertex are chosen to be $M_{13}.M_j$ for $1 \leq j \leq 26$. Of course, one vertex has several names in this notation, for example the identity vertex appears in the link of each of its vertices.

The first task is to examine how many group points lie in the link of the identity.

Lemma 2.1 *$Link(I)$ contains precisely 18 group points:*

$$y, \quad y^2, \quad x.y, \quad x.y^2, \quad x^2.y, \quad x^2.y^2,$$
$$x^3.y, \quad x^3.y^2, \quad y.x.y, \quad (y.x.y)^{-1}, \quad x.y.x.y, \quad x.(y.x.y)^{-1},$$
$$w, \quad w^{-1}, \quad (y.x.y)^{-1}.w, \quad x.(y.x.y)^{-1}.w, \quad y.x.y.w^{-1}, \quad x.y.x.y.w^{-1},$$

where w is the element introduced at the end of section 1.

Of course the fact that these are all group points is immediate and the fact that they are distance one from I is a calculation. The content of the lemma is that there are no more group points. This proved by noting that the lattice

$$M_{19} = \begin{pmatrix} 1 & 0 & t \\ 0 & t & 0 \\ 0 & 0 & t \end{pmatrix}$$

is in the link of the identity and is stabilised by the element u. Thus it cannot be a group point as its stabiliser contains an element of order 6. The action of known group elements now accounts for all the other elements in $Link(I)$.

We indicate briefly how one can construct any isometries which may exist in the stabiliser of M_{19}, in particular, how one can find the element u. This involves an elaboration of the method used in Lemma 1.3.

Suppose that $g \in Isom_J(\Delta)$ has $g[M_{19}] = [M_{19}]$. The definition shows that this is the same as the existence of an element $\alpha \in GL_3(\mathcal{O})$ with $g.M_{19} = M_{19}.\alpha$. The form of the elements M_{19} and g means that α has Laurent polynomial entries. Then

$$\alpha^*(M_{19}^* J M_{19})\alpha = (M_{19}\alpha)^*.J.(M_{19}\alpha) = (g.M_{19})^*.J.(g.M_{19}) = M_{19}^* J M_{19}$$

since g is an isometry. It follows that α is an isometry of the form $M_{19}^* J M_{19}$ and although unlike Lemma 1.3, this form does not have its matrix lying in $GL_3(\mathcal{O})$, we have a bound on the valuations of its entries, so that exactly as in the lemma, we have a bound on the valuations possible for the entries of α. Since we are dealing with a fixed finite field, it follows that there are only a finite number of possibilities for the entries of α and one can check by direct enumeration which of these make $M_{19}\alpha M_{19}^{-1}$ into a J isometry. (In fact sharper, more practical methods exist, but this would take us too far afield.)

We now give some indication of how one can give complete descriptions of all vertex stabilisers. The idea is to work outwards from the identity; it turns out that we need no more elements than those we have already introduced.

Recalling the notation defined above, a calculation shows that that action of u on its link is given by the permutation

$$(7^*)(11^*)(18^*)(23^*)(3^*13^*)(6^*8^*)(14^*24^*26^*)(17^*21^*20^*)$$
$$(1^*5^*12^*22^*19^*10^*)(4^*16^*25^*15^*2^*9^*)$$

where x^* is shorthand for $M_{19}.x$. The two six cycles consist of 12 group points, ($I = 2^*$), there are 14 points in the orbit of M_{19} and two remaining, as yet unidentified points, 7^* and 11^*. Points in the orbit of M_{19} we refer to as $n-$points. Observe that neither of the unidentified points can be group points as they contain an element of order 6 in their stabiliser.

Using this information we now show:

Lemma 2.2 *The group $stab_J(M_{19})$ acts on Δ as a finite group $\mathbf{Z}_6 \cong \langle u\rangle$.*

Sketch of Proof First consider the map $i_0\colon stab_J(M_{19}) \to Aut(Link(M_{19}))$. We begin by noting that this map is injective, for any element of the kernel must fix every vertex in $Link(M_{19})$, in particular the vertex I, so that the kernel can only consist of powers of the element x. However, one checks that no element of the group $\langle x\rangle$ other than the identity fixes M_{19} proving the assertion.

We refer to the above permutation, where we recall the vertex 2^* is the identity vertex. Pick an element $\gamma \in stab_J(M_{19})$; it is type-preserving so that it must map the group points in $Link(M_{19})$ which correspond to lines back to lines, and those which correspond to planes to planes. Since u acts transitively on this orbit, we can find some power of u so that $u^k.\gamma$ fixes the vertex 2^*. Now exactly as in the previous paragraph, we deduce that $u^k.\gamma = I$, so that γ is a power of u as required. □

We now analyse the two new points 7^* and 11^*. We have already shown that these are not group points; we now show that they are not n–points.

Firstly, one finds that $xyx(7^*) = 11^*$, so that this is only one orbit of point and moreover that u acts as an element of order two on $Link(7^*)$. Moreover, we can construct a potentially new element in $stab_J(7^*)$ namely $u_1 = (xyx)^{-1}.u.xyx$. A calculation reveals that the action of the group $\langle u, u_1\rangle$ on $Link(7^*)$ is the dihedral group D_3. It now follows from 1.5 and 2.2 that the orbit of 7^* is distinct from that of the group and n–points.

In fact, the stabiliser is larger than this and one finds that there is an element $h \in \langle x, y, u\rangle$ of order 3 which commutes with this dihedral group.

$$h = \begin{pmatrix} 1+t^4 & 1+t^2+t^4 & 1+t+2t^2+2t^3 \\ 2t+2t^2+2t^4 & 2+t^2+2t^4 & 2+2t+t^2+t^3 \\ 0 & 0 & 2t^2 \end{pmatrix}$$

We omit the arguments which identify the stabilisers of these two points, as this is slightly special, however the results are that one shows successively:

Lemma 2.3 *The map $i_1\colon stab_J(7^*) \to Aut(Link(7^*))$ has $im(i_1) \cong \mathbf{Z}_3 \times D_3$.*

Corollary 2.4 *The group* $stab_J(7^*)$ *has order* 54 *with structure given by the nonsplit central extension:*

$$1 \to \langle u^2 \rangle \cong \mathbf{Z}_3 \to stab_J(7^*) \to \mathbf{Z}_3 \times D_3 \to 1$$

The orbit type for the action of $stab_J(7^*)$ acting on its stabiliser is $\{9, 9, 3, 3, 1, 1\}$ where the orbits of size 9 are n–points, the orbits of size 3 are of type 7^* and there are two points as yet unaccounted for, namely $M_{19}.M_7.M_7$ and $M_{19}.M_7.M_{11}$ for which we adopt the notational shorthand $7^{(2)}$ and $11^{(2)}$. As above, $xyx(7^{(2)}) = 11^{(2)}$.

This is the point at which the behaviour stabilises. For later use, it is more convenient to define for $i \geq 0$, a sequence of elements $\alpha_{i+1} = (xyx)^{-i}u.u_1(xyx)^i$. Then we have:

Theorem 2.5 *For* $k \geq 2$, *the map* $i_k\colon stab_J(7^{(k)}) \to Aut(Link(7^{(k)}))$ *has image of order* 54.

Moreover, $stab_J(7^{(k)})$ *is generated by the elements* u, h, α_1, ... , α_k.

Sketch Proof The argument is inductive; we explain the step $k = 2$ which contains all the essential ingredients. We set $H(2) = \langle u, \alpha_1, \alpha_2 \rangle \leq stab_J(7^{(2)})$. Note that every element of $H(2)$ stabilises $7^{(3)}$ and $11^{(3)}$. We refer to Figure 1, which shows the hexagon $Link(7^{(2)})/H(2)$. Our claim is that no element of $\eta \in stab_J(7^{(2)})$ can move $7^{(3)}$.

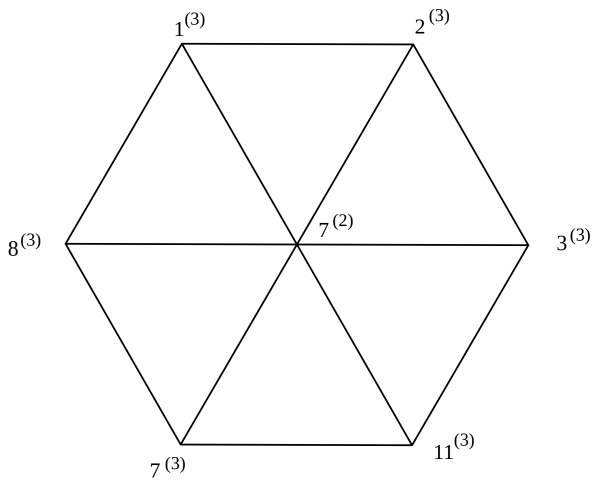


Figure 1

We argue as follows. Note that since elements in vertex stablisers are type preserving, the only possibilities for $\eta(7^{(3)})$ (assuming that it is moved) are the $H(2)$ orbit of $1^{(3)}$ or the $H(2)$ orbit of $3^{(3)}$.

However, the former orbit contains 9 elements and the latter 3, so that in any case, η must move some element in the $H(2)$ orbit of $1^{(3)}$ back into this orbit. By composing with an element of $H(2)$, we see that this implies the existence of an element moving $7^{(3)}$ lying in $stab_J(k^{(3)}) \cap im(i_2)$ where $k^{(3)}$ lies in the $H(2)$ orbit of $1^{(3)}$. After conjugating by an element of $H(2)$, we may assume that this element lies in $stab_J(1^{(3)}) = (xyx)^{-1}stab_J(2^{(3)})xyx$. But $2^{(3)} = 7^*$, so that $stab_J(1^{(3)}) = (xyx)^{-1}stab_J(7^*)xyx$. An examination of the generating elements shows that no element of this latter group moves $7^{(3)}$, a contradiction.

A similar argument establishes that $stab_J(7^{(2)})$ stabilises $11^{(3)}$.

We now show that $im(i_2)$ is a group of order at most 54. The reason is this: All of $im(i_2)$ stabilises $7^{(3)}$ hence permutes the four points in the link adjacent to it, however one of these points is $11^{(3)}$, which is also fixed by the whole group. Therefore by passing to a subgroup of $im(i_2)$ of index at most 3 we stabilise the point $3^{(3)}$. Arguing similarly for $3^{(3)}$, we deduce that $im(i_2)$ contains a subgroup of index at most 9 which stabilises $2^{(3)} = 7^*$. This is a group whose structure is already completely determined and one finds that $stab_J(7^*)$ acts on $Link(7^{(2)})$ as a group of order 6, proving the claim.

Now the group $H(2)$ is easily analysed; in particular, one shows easily that it acts on the link as a group of order 54. This establishes that $i_2(H(2)) = im(i_2)$ as required.

The kernel of the map $i_2\colon stab_J(7^{(2)}) \to Aut(Link(7^{(2)}))$ is a subgroup of $stab_J(7^*)$. Recalling that $H(2)$ is generated by u, α_1 and α_2, it follows that u, h, α_1 and α_2 generate $stab(7^{**})$, completing the first step of the induction. □

Given this theorem, one can now give a complete description of the groups $stab_J(7^{(k)})$ by analysing how $ker(i_k) \leq stab_J(7^{(k-1)})$ acts on $Link(7^{(k)})$. As a consequence, one proves that the group $stab_J(7^{(k)})$ has order 2.3^{2k+1}. It follows immediately that the orbits $7^{(k)}$ are all distinct.

We recap our progress so far. From the information that the stabiliser of the lattice I is a cyclic group of order four, we have identified the stabiliser of every vertex in the building; this information suffices to deduce that the orbits for the action of $Isom_J(\Delta)$ on Δ are precisely $I, M_{19}, 7^{(k)}$ for $k \geq 1$. Moreover, this already shows:

Corollary 2.6 *The group $Isom_J(\Delta)$ is generated by x, y and u.*

The construction of the entire complex $\Delta/Isom_J(\Delta)$ rests largely on the work set forth above and we shall not go into it in detail. Broadly it involves two steps: The identification of a candidate set of orbits of edges and triangles coming from the action of fairly short elements, followed by the proof that no further identifications are possible. This latter step is accomplished by the detailed understanding we have developed of the vertex stabilisers. This task gets easier as one moves further away from the group points, as stabilisers get larger and there are less orbits to be considered; eventually the action of stabilisers on links becomes constant. As a result, the complex has a fairly natural decomposition into two pieces; a compact part and some "tubes". We refer the reader to [3] for details in the case $p = 2$. For example, a picture of the tube comes from the concatenation of hexagons shown in Figure 2.

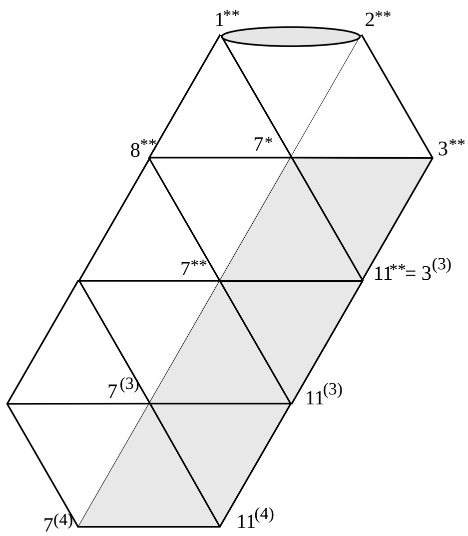


Figure 2

3 The case $p = 5$

The analysis in this case follows the same outline as indicated above, though of course the details become much more complicated. Nonetheless, one obtains a presentation of the group $Isom_J(\Delta(5))$. The quotient complex has interesting features not present in the first two cases; for example in contrast to the cases $p = 2$ and $p = 3$, the complex which emerges has three annular ends.

Once again one finds extra elements in $Isom_J(\Delta(5))$ which it turns out do not lie in the group generated by x and y. The simplest of these is the element β_2 shown below:

$$\beta_2 = \begin{pmatrix} 4 & 1+2t+2t^2 & 3+t \\ 1+t & 4+2t & 2+2t \\ 1 & 4+3t+4t^2 & 2+2t \end{pmatrix}$$

This is an element of order 4 and one finds that:

Theorem 3.1 *The group* $Isom_J(\Delta(5))$ *is generated by* x, y *and* β_2

In fact, we are able to complete all the analysis up until the very last step and in particular, we are able to find a presentation of the group $Isom_J(\Delta(5))$. It is rather complicated and *GAP* was unable to show that the index $[Isom_J(\Delta(5)) : \langle x, y\rangle]$ was finite. We have been unable to prove that it is infinite and unable to analyse the situation sufficiently to prove or disprove that $\langle x, y\rangle$ contains no extra relations.

Acknowledgment Both authors are supported in part by the NSF.

References

[1] **J Birman**, *Braids, Links and Mapping Class groups*, Annals of Math. Studies 82, Princeton University Press (1975)

[2] **K Brown**, *Buildings*, Springer–Verlag (1989)

[3] **D Cooper**, **D D Long**, *A presentation for the image of Burau(4)*$\otimes\mathbf{Z}_2$, Inventiones Math. 127 (1997) 535–570

[4] **A Haefliger**, *Complexes of groups and orbihedra*, from: "Group Theory from a Geometrical Viewpoint", E Ghys, A Haefliger and A Verjovsky (editors), World Scientific (1991) 504–540

[5] **J Moody**, *The faithfulness question for the Burau representation*, Proc. Amer. Math. Soc. 119 (1993) 671–679

[6] **D D Long**, **M Paton**, *The Burau representation is not faithful for* $n \geq 6$, Topology 32 (1993) 439–447

[7] **C Squier**, *The Burau representation is unitary*, Proc. Amer. Math. Soc. 90 (1984) 199–202

Department of Mathematics
University of California
Santa Barbara, CA 93106, USA

Email: `cooper@math.ucsb.edu, long@math.ucsb.edu`

Received: 26 August 1997

ISSN 1464-8997 **139**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 139–158

Folding sequences

M J Dunwoody

Abstract Bestvina and Feighn showed that a morphism $S \to T$ between two simplicial trees that commutes with the action of a group G can be written as a product of elementary folding operations. Here a more general morphism between simplicial trees is considered, which allow different groups to act on S and T. It is shown that these morphisms can again be written as a product of elementary operations: the Bestvina–Feighn folds plus the so-called "vertex morphisms". Applications of this theory are presented. Limits of infinite folding sequences are considered. One application is that a finitely generated inaccessible group must contain an infinite torsion subgroup.

AMS Classification 20E08; 57M07

Keywords Groups acting on trees, free groups

Dedicated to David Epstein on the occasion of his 60th birthday

1 Introduction

A morphism $\phi\colon S \to T$ of finite trees can be written as a product of elementary *folds*, in which two edges with a common vertex are folded together, and an isomorphism. Bestvina and Feighn [1] have given a generalization of this result. The case they consider is when S and T are (generally infinite) simplicial G–trees for which $G\backslash S$ and $G\backslash T$ are finite graphs T is minimal, and G and the edge stabilizers of T in G are finitely generated. The morphism now becomes a product of equivariant folds and an isomorphism. In each such fold a whole orbit of pairs of edges are folded together. Such an operation is easy to describe in terms of its effect on the quotient graph $G\backslash S$ and the edge and vertex stabilizers of S. These are specified in a *graph of groups* determined by a labelling of the edges and vertices of $G\backslash S$. In this paper a further generalization is given. We now allow different groups to act on S and T. Thus S is a G–tree and T is an H–tree and a *morphism* $\phi\colon S \to T$ incorporates a homomorphism $\tilde{\phi}\colon G \to H$, so that if we regard T as a G–tree via $\tilde{\phi}$ then ϕ is a morphism of G–trees. As well as the basic folding operations of [1] it is also necessary to include *vertex morphisms* each of which changes just one vertex label of the corresponding graph of groups. It is possible to generalize the Bestvina–Feighn result for the case when $\tilde{\phi}$ restricts to an injective homomorphism on point stabilizers of S. Under similar restrictions to those specified for a G–morphism, ϕ is a product of elementary folds, vertex morphisms and an isomorphism. A sequence of such operations is called a *folding sequence.* We can think of each tree in the sequence as the realization of a combinatorial tree. The folding and vertex morphisms correspond to morphisms of the combinatorial trees. If we interpret our folding sequence as a folding sequence of combinatorial trees then we also have to allow subdivision operations. This is because two different combinatorial trees may have isomorphic realizations as $\mathbf{R}$–trees. However if this does happen, then the two trees have isomorphic subdivisions.

Folding sequences are surprisingly useful. They yield theoretical results on decompositions of groups and also provide a way of constructing groups with strange properties.

A G–tree S is called reduced if for every edge $e \in ES, G_e = G_{\iota e}$ implies $\iota e, \tau e$ are in the same orbit. Let S be a reduced G–tree in which every edge group is finite. Let $\overline{S} = G\backslash S$ and let $(G(-), \overline{S})$ be the corresponding graph of groups. Put

$$\eta(S) = \sum_{e \in E\overline{S}} 1/|G(e)|.$$

Linnell [12] proved that $\eta(S) \leq 2d_G(\omega\mathbf{Q}G) - 1$ where $d_G(\omega\mathbf{Q}G)$ is the minimal number of generators of the augmentation ideal $\omega\mathbf{Q}G$ as a $\mathbf{Q}G$–module. Linnell's argument uses norms in W^*–algebras. Using a folding sequence argument we show that $\eta(S) \leq d(G)$, the minimal number of generators of G. If all the edge stabilizers of S are trivial, then $\eta(S) = |E\overline{S}|$ and so $|E\overline{S}| \leq d(G)$. This is a weak version of the Grushko–Neumann Theorem (see [4] or [16]). A stronger version of the Grushko–Neumann Theorem is obtained by a closer examination of the folding sequence. Stallings [16] has given a proof of this result using this approach.

Let G be a group. In [8] and [9] I introduced the idea of a *G–protree.* A *splitting sequence* of G–trees $T_1, T_2, \ldots$ is a sequence such that for each n there is a surjective G–map $T_n \to T_{n-1}$ obtained by contracting finitely many orbits of edges. A G–protree P arises as the inverse limit of this sequence. As shown in [9], if ET_n is countable for all n, then P has a realization as an $\mathbf{R}$–tree, on which G acts by isometries. In this $\mathbf{R}$–tree the set of branch points intersects each segment in a nowhere dense subset. A finitely generated group G is said to be *inaccessible* if there is a splitting sequence of reduced G–trees as above, for which all edge groups are finite and the number of G–orbits of VT_n (or ET_n) tends to infinity. In this case we obtain a G–protree P with infinitely many orbits of edges.

We prove in Section 3 that if G is finitely generated and P is a G–protree with countably many edges then the realization of P is a direct limit of a folding sequence of simplicial $\mathbf{R}$–trees. If the G_n–tree S_n is the n–th term of the sequence, then there is a surjective homomorphism $\tilde{\rho}_n\colon G_n \to G_{n+1}$ and G is the direct limit of this system of homomorphisms in the category of groups. This description of G gives information as to the subgroup structure of G. In particular either $G \cong G_n$ for all sufficiently large n or G must contain a subgroup which is the union of a properly ascending chain of finitely generated subgroups each of which is contained in an edge stabilizer of P. It follows that an inaccessible group must contain an infinite locally finite subgroup. If every edge stabilizer of S_n in G is cyclic (not necessarily finite), then G must contain a non-cyclic subgroup that is locally cyclic. It also follows that if G has an infinite splitting sequence then for any integer k there is an integer n such that G contains a nontrivial element which fixes an edge path in T_n of length at least k. This is also implied by Sela's results on acylindrical accessibility [14].

Infinite folding sequences were used first by Bestvina and Feighn [2] to give an example of a finitely generated group which had an infinite splitting sequence in which all edge groups are free abelian of rank 2. Subsequently [7], [8], [9] I gave a number of examples of inaccessible groups all of which were constructed (essentially) by means of folding sequences.

Martin and Skora [13] have obtained some results on accessible convergence groups acting on S^2. It is not hard to show that an infinite locally finite group cannot act as a convergence group on S^2. Hence by Theorem 4.5 a finitely generated convergence group acting on S^2 must be accessible. Thus the accessibility condition in the results of Martin and Skora can be removed (or replaced by a finite generation condition). In particular it follows that if $G \subset \mathrm{Hom}(S^2)$ is an orientation preserving convergence group, then there is a simplicial G–tree T such that $G\backslash T$ is a finite graph, all edge stabilizers are finite, and if $v \in VT$, then the ordinary set $O(G_v)$ is simply connected.

2 Folding

We recall and modify some of the terminology of [6] or [15].

Let G be a group. A *G–tree* T is an $\mathbf{R}$–tree with G acting on the left by isometries. A G–tree is *minimal* if it has no proper G–subtree.

Given an $\mathbf{R}$–tree T and $x \in T$, define $B_x = \{[x,y] | y \in T - x\}$. Define an equivalence relation on B_x by

$[x, y] \sim [x, z]$ if $[x, y] \cap [x, z] = [x, w]$ for some $w \in T - x$.

A *direction at* x is an element of $B_x/\sim$. There is a bijection between directions at x and the components of $T - x$. A *point of reflection* x of a G–tree T is a point with two directions for which there exists $g \in G$ which fixes x and transposes the two directions at x. We say that $x \in T$ is an *ordinary point* if there are exactly two directions at x but x is not a point of reflection. A *branch point* is a point x with more than two directions or equivalently for which $T - x$ has more than two components. A *vertex* is a point which is not an ordinary point.

An $\mathbf{R}$–tree is *simplicial* if the set of vertices is discrete. For each $x \in T$, let $d(v)$ denote the number of directions at x.

A *morphism* from a G–tree S to a G–tree T is a G–map $\phi\colon S \to T$ such that for each segment $[x, y]$ of S there is a segment $[x, w] \subset [x, y]$ such that $\phi|_{[x,w]}$ is an isometry.

Alternatively ([6]) $\phi\colon S \to T$ is a morphism if every segment has a finite subdivision such that ϕ restricts to an isometry on each segment of the subdivision.

We generalize the notion of morphism to allow different groups to act on domain and codomain. Thus if S is a G–tree and T is an H–tree, a *morphism* $\phi\colon S \to T$ is a homomorphism $\tilde{\phi}\colon G \to H$, and a map $\phi\colon S \to T$ such that if we regard T as a G–tree via $\tilde{\phi}$ then ϕ is a morphism when regarded as a morphism of G–trees. Such morphisms are discussed in unpublished work of Skora.

A simplicial $\mathbf{R}$–tree T can be regarded as the *realization* of a simplicial complex, which is a (combinatorial) tree. This will also be denoted T. Thus VT will correspond to a non-empty closed discrete subset of the $\mathbf{R}$–tree containing all branch points and ET will be the set of closures of components of $T - VT$, where VT is such that each element of ET is a closed segment the endpoints of which are elements of VT. As a combinatorial tree the vertices of the edge e are denoted $\iota e, \tau e$. When regarded as a protree the edges of a tree are regarded as directed pairs. Usually an edge of a tree is not directed.

Bestvina and Feighn [1] have shown that any morphism of simplicial G–trees is a product of subdivisions and folds (which are described as operations on the corresponding combinatorial G–trees). Folds are classified according to their effect on the quotient graph. The quotient graph $X = G\backslash T$, together with a labelling by subgroups of G which are the stabilizers of a lift of a maximal subtree X_0 of X, is known as a graph of groups $(G(-), X)$. See [4] for an account of this theory. The basic folds of Type I, II and III are shown below in Figure 1. These are denoted Type IA, IIA and IIIA in [1]. Bestvina and Feighn

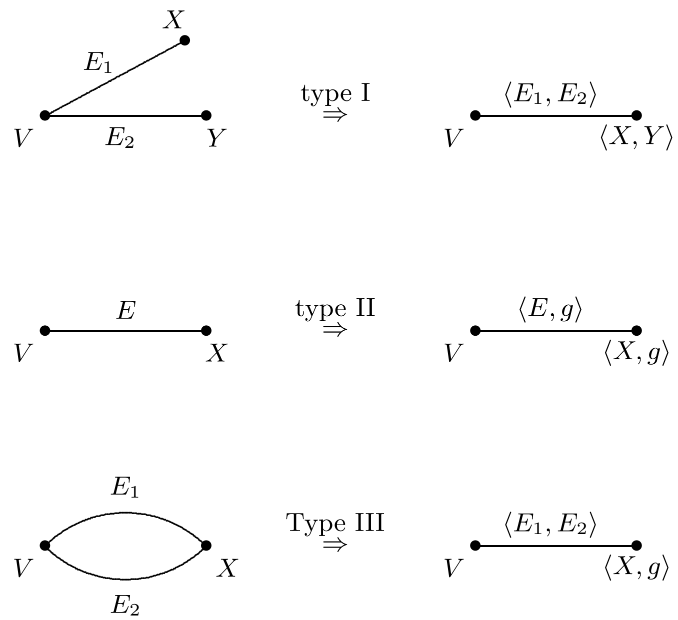


Figure 1

list other basic folds (Type IB,IIB, IIIB and IIIC). But as they remark, each of these is equivalent to a combination of Type A folds and subdivisions.

In [9] I introduced vertex morphisms. A *vertex morphism* is a morphism $\theta\colon S \to T$ of simplicial $\mathbf{R}$–trees for which the only change in the corresponding graph of groups is a change in the label of one of the vertices. Thus if the label U is changed to V then there is a surjective homomorphism $\theta_U\colon U \to V$ which restricts to the identity map on subgroups which label incident edges. For vertex morphisms the group G acting on S is different from the group H acting on T. We now generalize the Bestvina–Feighn result to allow different groups to act on domain and codomain.

Theorem 2.1 *Let S, T be simplicial $\mathbf{R}$–trees. Let G act by isometries on S and let H act by isometries on T so that $G\backslash S$ is finite, and all edge stabilizers of T in H are finitely generated. Also T is a minimal H–tree. Let $\phi\colon S \to T$ be a morphism, such that the corresponding homomorphism $\tilde{\phi}\colon G \to H$ is surjective, and restricts to an injective map on each point stabililizer, then ϕ can be written as a product of basic folds and vertex morphisms.*

Proof We adapt the proof of the Proposition in Section 2 of [1].

Step 0 We show that if K is a finite simplicial subtree of S, then we can factor ϕ as $\gamma\beta$ where β is a product of folds and vertex morphisms and γ restricted to $\beta(K)$ is an embedding. Also $\tilde{\gamma}$ is injective on all point stabilizers. If $\phi|K$ is not already an embedding then we can perform a basic fold $\phi_1\colon S \to S_1$ folding together edges e_1, e_2 of S so that $\phi(e_1) = \phi(e_2)$ and e_1, e_2 are distinct edges of X. The basic fold ϕ_1 produces at most one new edge group and one new vertex group. The new edge group is a subgroup of an existing vertex group. It follows that $\tilde{\phi}_1$ restricts to an injective homomorphism on the stabilizers of all except one orbit of vertices of S_1 and on the stabilizers of all edges. Thus there is a vertex morphism $\nu_1\colon S_1 \to T_1$ such that $\phi\colon S \to T$ factors $\phi = \phi^{(1)}\nu_1\phi_1$ as a morphism of $\mathbf{R}$–trees (regarding T as an H–tree), and also $\tilde{\phi}^{(1)}\colon G_1 \to H$, the homomorphism corresponding to $\phi^{(1)}$, retricts to an injective homomophism on all point stabilizers. Note that $\nu_1\phi_1(K)$ has fewer edges than K. We can therefore proceed by induction on the number of edges of K.

Step 1 We now claim that we can factor ϕ as $\gamma\beta$ so that γ induces a homeomorphism of quotient graphs, $\tilde{\gamma}$ is injective on point stabilizers and β is a product of basic folds and vertex morphisms. This follows exactly as in the corresponding argument in [1]. The fact that T is a minimal H–tree and $\tilde{\phi}$ is surjective, together imply that the induced morphism $G\backslash S \to H\backslash T$ is a surjective simplicial map. One then uses an induction argument based on the number of edges of $G\backslash S$, using Step 0.

Step 2 Since edge stabilizers in T are finitely generated, we can use the argument of [1] to show that ϕ can be factored $\phi = \gamma\beta$ as in Step 1 and in addition $\tilde{\gamma}$ induces an isomorphism on all edge stabilizers.

Step 3 It follows as in [1] that the γ obtained in Step 2 is an isomorphism. □

We say that in the G–tree S an edge $e \in ES$ is *compressible* if $G_{\iota e} = G_e$ and ιe and τe lie in different G–orbits. We say that S is *reduced* if it has no compressible edges. For any G–finite G–tree S there is a reduced G–tree S^* for which VS^* is a G–retract of S: S^* is obtained from S by compressing compressible edges. The retraction is not, in general, uniquely determined. The retraction is determined by a *compressing forest* F defined as follows:

(1) F is a subgraph of $G\backslash S = \overline{S}$.

(2) The edges of F are oriented (given arrows) so that each vertex $v \in VF$ has at most one arrow pointing away from it.

(3) If $e \in EF$ then $G(e) = G(\iota e)$, where the arrow on e points from ιe to τe.

(4) F is maximal with respect to properties (1), (2) and (3). In particular $VF = V\overline{S}$.

In each component c of F there is exactly one vertex v_c which has no arrow pointing away from it. The retraction $S \to S^*$ corresponding to F induces a retraction $\rho\colon \overline{S} \to \overline{S}^*, \rho(v) = v_c, v \in c$.

It is often convenient to work with reduced trees. We know that it is possible to factorize a morphism of reduced trees as a product of subdivisions, folds and vertex morphisms. Unfortunately subdividing a tree always produces compressible edges. We introduce some modified folding operations which allow us to factorize a morphism of reduced trees so that the intermediate trees obtained are also reduced. These modified folds are shown in Figures 2 ,3 and 4.

Every morphism of G–trees is a product of subdivisions and folds of types I, II and III. Let $\phi\colon S \to T$ be such a fold. Given a compressing forest F in $\overline{S}$, we will describe how to construct a compressing forest F' in $\overline{T}$ and describe the corresponding induced morphism $\phi^*\colon S^* \to T^*$. Again these are best described by their effect on the labelled quotient graphs.

Subdivision induces an isomorphism on the corresponding reduced trees, since one enlarges the compressing forest to include half the subdivided edge. Thus a morphism of reduced trees can always be written as a product of isomorphisms and the morphisms $\phi^*\colon S^* \to T^*$ induced by type I, II and III folds. These are discussed in detail below.

We consider the effect of folds on the quotient graph $\overline{S}$ and the quotient reduced graph $\overline{S}^*$. In the subsequent discussion, and in the diagrams of graphs of groups, the group corresponding to a given edge or vertex is denoted with the corresponding capital letter, eg the group corresponding to vertex v is V and the group of e_1 is E_1. For any vertex w, put $\rho(w) = w^*$, which therefore has the group W^*. Note that if $W = W^*$ then we can change the arrows on F so that w has no arrows pointing away from it (by reversing all the arrows on the geodesic from w to w^*). A change like this has no effect on $\overline{S}^*$.

We now list the different possibilities for the fold ϕ and the resulting induced fold ϕ^*

Type I

$e_1, e_2 \in F$

We choose the new compressing forest F' to contain all $x \in F, x \neq e_1, e_2$. Also e_1, e_2 fold to form the edge $\langle e_1, e_2\rangle$, which is included in F' with an arrow pointing away from pivot vertex v if and only if one of the edges e_1, e_2 has arrow pointing away from v. It is easy to check that F' is a compressing forest and ϕ induces an isomorphism on S^*, since the folding takes place in a part of the tree that is compressed both before and after the fold.

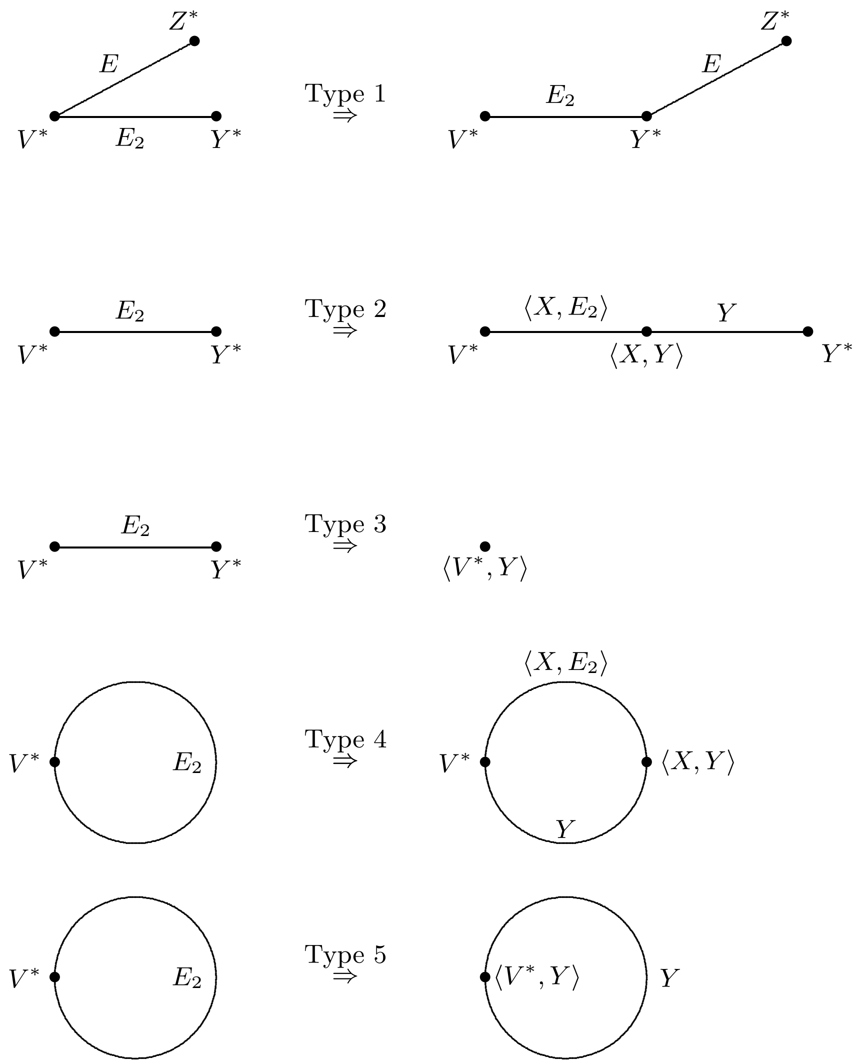


Figure 2

$e_1 \in F, e_2 \notin F$ and v, y in different components of F

Suppose first that the arrow on e_1 goes from x to v. Then $X = E_1$. After the fold F' is obtained from F by deleting e_1. If $X \leq E_2$, then ϕ^* consists of a composite of Type 1 folds for each edge e which has a vertex w in the same component of F as v but for which the arrowed path from w to v^* passes through x. It is important to note that in each such Type 1 fold $E \leq E_2$. Assume then that $X \not\leq E_2$. If $\langle X, E_2\rangle \neq V^*$ then after folding the new compressing forest is obtained by omitting the folded edge and also the edge

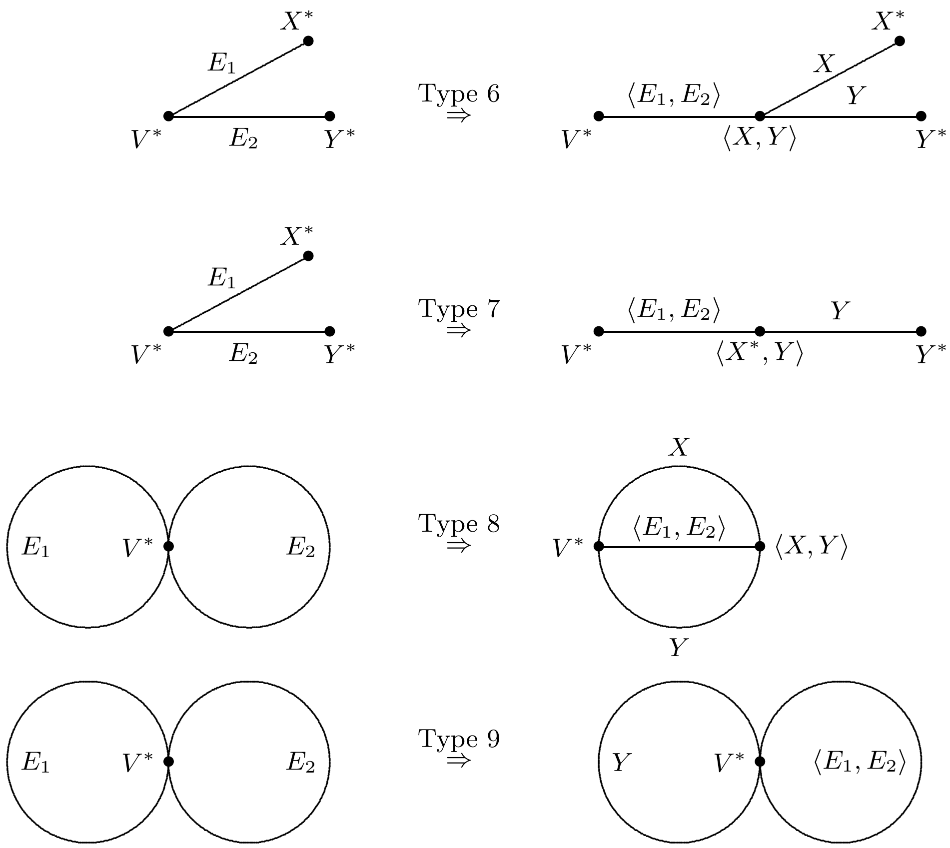


Figure 3

originally pointing away from y if $Y \neq Y^*$. Note that $\langle X, E_2\rangle \neq \langle X, Y\rangle$, since $\langle X, E_2\rangle$ is a subgroup of V^* but Y is not contained in V^*. Such a fold is called a Type 2 fold. Note that we can assume $E_2 \neq Y$ in a Type 2 fold, since if $E_2 = Y$, then because v, y are in different components of F we could get a bigger compressing forest by adding e_2. If $Y = Y^*$, then the induced fold is a combination of Type II folds. Similarly if $\langle X, E_2\rangle = V^*$ (so that the folded edge must be added to F) and $Y \neq Y^*$, then the induced fold is a combination of Type II folds. If $\langle X, E_2\rangle = V^*$ and $Y = Y^*$ then the induced fold is a Type 3 fold.

If the arrow on e_1 goes from v to x, then the fold produces a compressible edge which can be included in the the new compressing forest with arrow going from v to $\langle x, y\rangle$. If there are arrows in F pointing away from x and y then these edges must be omitted from the new compressing forest. If $X \neq X^*(= V^*)$ and $Y \neq Y^*$, the effect on $\overline{S}^*$ is a Type 2 fold (with $\langle X, E_2\rangle = X$). Note that

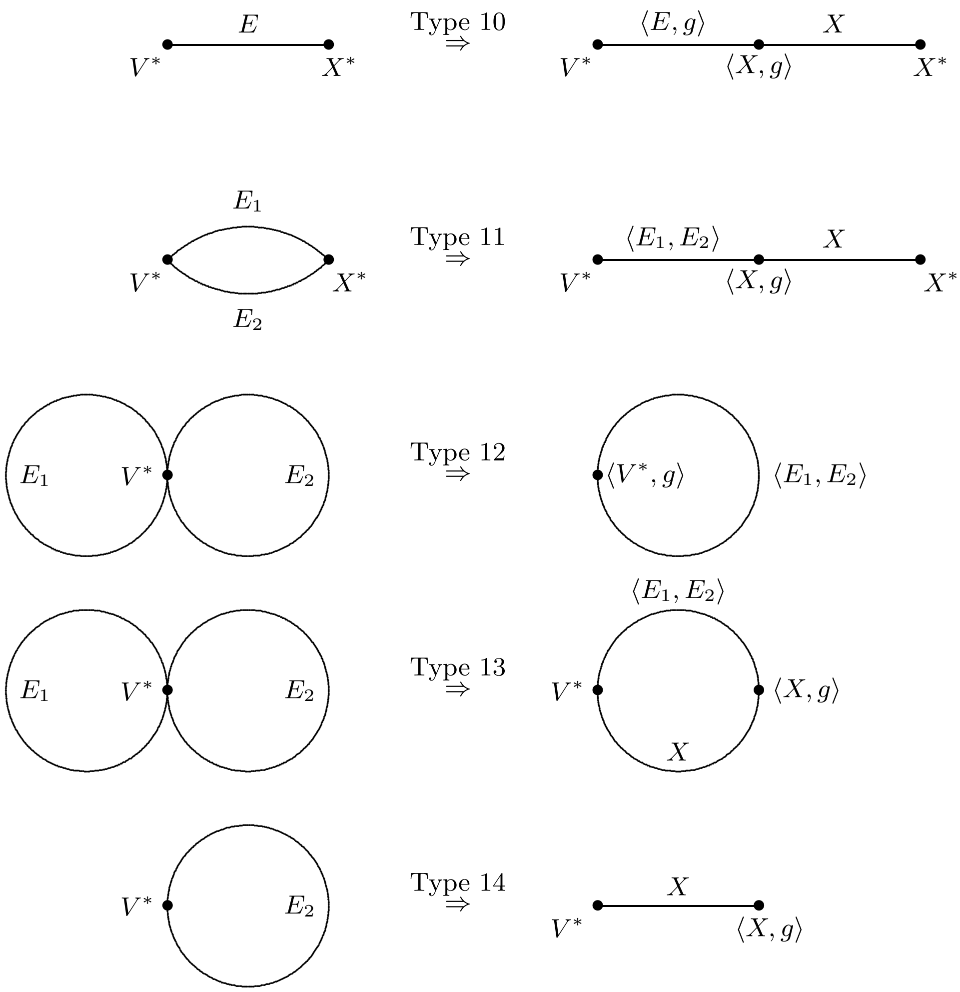


Figure 4

E_2 is a proper subgroup of X, since otherwise we could add e_2 to F and get a larger compressing forest in S^*. The induced fold for $X = X^*$ and $Y \neq Y^*$ is a combination of Type II folds (with y as the pivot vertex instead of v). The vertex which is initially labelled V^* finishes with label $\langle V^*, Y\rangle$ and the vertex with label Y^* is unchanged. The folded edge becomes a vertex if $X = X^*$ and $Y = Y^*$. Thus we have a Type 3 fold.

$e_1 \in F, e_2 \notin F$ and v, y in the same component of F

We can assume $E_2 \neq Y$, since if $E_2 = Y$ we could change F so that it included both e_1 and e_2 which is a case already considered. To see this note that

$v^* = y^*$. If there is no edge of F pointing away from y then $v^* = y$ and $V = Y$ and we can change arrows so that there is an edge in F pointing away from y. Now change F so that it includes e_2 and omits this edge. Thus we can assume $E_2 \neq Y$. The analysis for this case is now similar to that when v, y are in different components. The induced fold is of Type 4 if $\langle X, E_2\rangle \neq V^*$ and of Type 5 if $\langle X, E_2\rangle = V^*$. Note that, since the part of the graph of groups we are concerned with in this case is not a tree, it cannot be assumed that all the edge labels are subgroups of the incident vertex labels. Thus in a Type 4 fold, Y is not assumed to be a subgroup of V^*—it is conjugate to a subgroup of V^*. There is no analogous case to Type 3.

$e_1 \notin F, e_2 \notin F, v, x, y$ in distinct components of F

If either $X = X^*$ or $Y = Y^*$, then we can change the arrows on F so that either x or y has no edges pointing away from it. Thus if F contains edges pointing away from both x and y, then we can assume $X \neq X^*$ and $Y \neq Y^*$. In this case we must omit at least one of these edges from F after the fold. If $\langle X, Y\rangle \neq X$ then we must omit the edge of F with initial vertex x. Similarly if $\langle X, Y\rangle \neq Y$ then we must omit the edge of F with initial vertex y. If $\langle X, Y\rangle = X = Y$ then we need only omit one of the two edges, and we can choose either. First consider the case when both edges are omitted. The fold in this case is a Type 6 fold if $V^* \neq \langle E_1, E_2\rangle$. Note that $E_1 \neq X$ and $E_2 \neq Y$, since otherwise we could add e_1 or e_2 to F, contradicting its maximality. If $V^* = V = \langle E_1, E_2\rangle$ then the folded edge is compressible and can be added to F. The induced fold in this case is a combination of Type II folds: first operating on the edge e_1 by increasing E_1 to X and V^* to $\langle V^*, X\rangle$ and then operating on the edge e_2 by increasing E_2 to Y and $\langle V^*, X\rangle$ to $\langle X, Y\rangle$. For any edge of $\overline{S}$ that is not in F which has a vertex w for which the path from w to w^* passes through x or y it is necessary to carry out a Type 1 fold in the reduced graph. Such an edge, which initially is incident with x^* in $\overline{S}^*$ becomes incident with $\langle x, y\rangle$ in $\overline{T}^*$.

Consider now the case when only one edge is omitted. This happens for example if $X = X^*$ and $Y \neq Y^*$ then the induced fold is of Type 7. If $X = X^*$ and $Y = Y^*$ then the induced fold is just a Type I fold. If v, x, y are in different components of F then both $\langle X, Y\rangle \neq X$ and $\langle X, Y\rangle \neq Y$, since $X \leq Y$ implies x, y are in the same component of F. It follows that the edges after the fold cannot be added to F.

$e_1 \notin F, e_2 \notin F, v, x, y$ not in distinct components of F

This case is similar to the previous case. We can still assume that $E_1 \neq X$ and $E_2 \neq Y$. For if say $E_1 = X$, and v, x are in the same component of F, then either there is an edge in F pointing away from x or $X = V = V^*$ and there is an edge in F pointing away from v. We can then change F by removing this edge and replacing it by e_1. Such a change induces an isomorphism on the reduced graph. The fold will now involve an edge of F and has been considered previously.

Suppose v, x, y are all in the same component of F so that $V^* = X^* = Y^*$ and $\langle X, Y\rangle \neq V^*, \langle X, Y\rangle \neq X, \langle X, Y\rangle \neq Y$. The induced fold is of Type 8. Again it may be necessary to alter by Type 1 folds the incidence of edges to vertices in $\overline{S}^*$. The similarity with the case when v, x, y are in different components of F is because in both cases F is altered in the same way; by omitting the edges pointing away from the identified vertex $\langle x, y\rangle$. It may now be the case that $\langle X, Y\rangle = X$ say. In this case there would be a compressible edge produced and so we can add an extra edge to F and the induced fold is of Type 9.

Type II

$e \in F$

In such a fold $V \neq E$ and so the arrow on e must point from x to v. We can include the folded edge $\langle e, g\rangle$ in F', with arrow pointing from $\langle x, g\rangle$ to v.

$e \notin F, v, x$ **in different components of** F

We obtain a Type 10 fold for the case when $X \neq X^*$. Type 1 folds in $\overline{S}^*$ are necessary corresponding to any edge of $\overline{S} - F$ joined to a vertex w for which the path from w to w^* passes through x. If $X = X^*$ then the induced fold is just a Type II fold.

$e \notin F, v, x$ **in the same component of** F

This is the same as the previous case except that the vertices v^* and x^* are identified before and after the folds. This gives rise to folds of Type 4 and 5.

Type III

$e_1, e_2 \notin F, v, x$ **in different components.**

We obtain a Type 11 fold when $X \neq X^*$. Again Type 1 folds may be neccessary corresponding to any edge of $\overline{S} - F$ joined to a vertex w for which the path from w to w^* passes through x. If $X = X^*$ then the induced fold is just a Type III fold.

$e_1, e_2 \notin F, v, x$ **in the same component of** F

This produces a Type 12 fold if $X = X^*(= V^*)$, and a Type 13 fold if $X \neq X^*$.

$e_1 \in F, e_2 \notin F$

In this case, since e_2 has both its vertices in the same component of F it may be the case that $E_2 = X$. We obtain a Type 14 fold.

We see then that the induced folds in reduced trees may just be a Type I, II or III fold, but it may be of a type which creates a new vertex. For example a Type 6 fold creates a new vertex.

Theorem 2.1 can now be adapted for morphisms between reduced trees.

Theorem 2.2 *Let S, T be simplicial reduced $\mathbf{R}$–trees. Let G act by isometries on S and let H act by isometries on T so that $G \backslash S$ is finite, and all edge stabilizers of T in H are finitely generated. Also T is a minimal H–tree. Let $\phi\colon S \to T$ be a morphism, such that the corresponding homomorphism $\tilde{\phi}\colon G \to H$ is surjective, and restricts to an injective map on each point stabililizer, then ϕ can be written as a product of folds of Type I, II and III or of Types 1 – 14 and vertex morphisms and all the intermediate trees are reduced.*

This result enables us to deduce certain bounds on the complexity of decompositions of finitely generated groups.

Let S be a G–tree with finite edge stabilizers. Define

$$\eta(S) = \sum_{e \in E\overline{S}^*} 1/|G(e)|.$$

Theorem 2.3 *Let G be a finitely generated group for which $d(G)$ is the minimal number of generators, then $\eta(S) \leq d(G)$.*

Proof Let W be a free group of rank $d(G)$ and let X be the W–tree with one orbit of vertices on which W acts freely, and for which $\eta(\overline{X}^*) = d(G)$. We regard both X and S as simplicial $\mathbf{R}$–trees. A surjective homomorphism $\tilde{\alpha}\colon W \to G$ induces a morphism $\alpha\colon X \to S$. By Theorem 2.1 α is a product of basic folds and vertex morphisms. We consider the induced folds on the reduced trees. One can check without too much difficulty that $\eta(S)$ does not increase for each of the induced folds described above. For example, for a fold of Type 6

$$\eta(S) - \eta(T) = \frac{1}{|E_1|} + \frac{1}{|E_2|} - \frac{1}{|\langle E_1, E_2\rangle|} - \frac{1}{|X|} - \frac{1}{|Y|}.$$

We can assume $|E_1| \leq |E_2|$. Also we know that $E_1 < X$ and $E_2 < Y$. Thus $\frac{1}{|X|} \leq \frac{1}{2|E_1|}$ and $\frac{1}{|Y|} \leq \frac{1}{2|E_2|} \leq \frac{1}{2|E_1|}$, so that $\frac{1}{|X|} + \frac{1}{|Y|} \leq \frac{1}{|E_1|}$. Also

$\frac{1}{|\langle E_1, E_2\rangle|} \leq \frac{1}{|E_2|}$. It is clear in this case that $\eta(S) - \eta(T) \geq 0$. Similar arguments show that $\eta(S)$ does not increase in each of the other cases. A vertex morphism will leave edge groups unchanged and cannot increase $\eta(S)$. The theorem is proved. □

Let us consider the case when G is a finitely generated group and S is a G–tree with trivial edge stabilizers. In this case $\eta(S) = |E\overline{S}^*|)$, and so we see that the number of edge orbits in a minimal reduced G–tree is bounded by $d(G)$. In fact we obtain stronger versions of the Grushko–Neumann Theorem by examining the folding sequence in this case. Thus we obtain the following theorem, first obtained in [4, I, 10.3].

Theorem 2.4 *Let S be a G–tree and let T be a reduced minimal H–tree for which G acts freely on ES and H acts freely on ET. Also suppose H is finitely generated. Let α: $S \to T$ be a morphism. If $\tilde{\alpha}$: $G \to H$ is surjective then there is a G–tree S' and a morphism α': $S' \to T$ that induces an isomorphism $G\backslash S' \to H\backslash T$ and $\tilde{\alpha}'$ induces a surjective homomorphism $G_v \to H_{\alpha'(v)}$ for each vertex $v \in VS'$.*

Proof We can carry out vertex morphisms on S and replace each vertex stabilizer by its image under $\tilde{\alpha}$. We will then have a $\hat{G}$–tree $\hat{S}$ for which there is a morphism $\hat{\phi}$: $\hat{S} \to T$ for which the corresponding homomorphism $\hat{G} \to H$ is injective on all point stabilizers. By Theorem 2.1 $\hat{\phi}$ is a product of basic folds, subdivisions and vertex morphisms. We consider the induced operations on the corresponding reduced trees. Since all edge groups are trivial, the only possible induced folds that can occur are Type I, III, 1, 3 and 5 (with $E_2 = X = \{1\}$). If we carry out the same sequence of induced folds on S^* (leaving out all the vertex morphisms), we will obtain the G–tree T' with the required properties. □

3 Folding sequences

A folding sequence (T_n), is a sequence of combinatorial trees T_n, satisfying the following conditions:

(a) T_n is a minimal G_n–tree, where G_n is finitely generated.

(b) T_{n+1} can be obtained from T_n either by subdivision, or by a I, II or III fold followed by a vertex morphism.

It is often the case that corresponding to a folding sequence (T_n) is a *folding sequence of simplicial* $\mathbf{R}$*–trees*, in which we replace each tree by a realization and the folding operations induce morphisms of $\mathbf{R}$–trees. In this case we will risk confusion by using T_n to denote both the tree and its realization as an $\mathbf{R}$–tree. There are examples of folding seqences which cannot be realized in the above way. For example if for each n, $G_{2n-1}\backslash T_{2n-1}$ is a tree with two edges e_{2n-1}, f_{2n-1}, and T_{2n} is obtained from T_{2n-1} by subdividing e_{2n-1} into two edges e_{2n} and e_{2n+1}. Then T_{2n+1} is obtained from T_{2n} by a Type I fold, in which e_{2n} and f_{2n-1} are folded together to form f_{2n+1}. We call such a folding sequence *reducible*. Thus a folding sequence is reducible if it satisfies the following condition:

There exists n, such that for each $m \geq n$ there is a proper subset $E_m \subset ET_m$ which is invariant under G_m and such that if the folding operation involves an edge of E_m then the resulting edges are in E_{m+1}.

Thus if the folding operation is subdivision of an edge of E_m, then the resulting edges are all in E_{m+1}; and if the operation is a Type I fold in which one of the edges is in E_m, then the resulting edge is in E_{m+1}. In the the above example the folding sequence is reducible since the sets $E_{2m} = E_{2m-1} = \{f_{2m-1}\}$, satisfy the above condition. A folding sequence is *irreducible* if it is not reducible.

Theorem 3.1 *Let (T_n) be an irreducible folding sequence of combinatorial trees. The sequence can be realized as a folding sequence of morphisms of simplicial $\mathbf{R}$–trees in which group actions are by isometries.*

Proof For each n it is possible to realize the finite folding sequence $T_1, T_2, \ldots,$ T_n as a folding sequence of morphisms of simplicial $\mathbf{R}$–trees in which the group actions are by isometries. To produce such a realization one has to assign a common length to the edges in each orbit of edges in such a way that the lengths are compatible with subdivision and so that Type I and Type III folds take place between edges of equal length. To achieve such a realization assign lengths to the edges of T_n and work backwards, noting that the lengths of edges of T_i are determined by the lengths of edges of T_{i+1}. For each $n = 1, 2, \ldots,$ let $z_n = (\xi_{n1}, \xi_{n2}, \xi_{n3}, \ldots, \xi_{nk})$ be the length of the edges $e_1, e_2, \ldots e_k$ of $G_1\backslash T_1$ in such a solution. We may assume that for each $n, |z_n| = \sum_{i=1}^{k} \xi_{ni} = 1$. By compactness for the standard $n-1$–simplex $|\sigma_{n-1}|$, the sequence z_n has a convergent subsequence. Let $w_1 = (\xi_1, \xi_2, \ldots \xi_k)$ be the limit point of a convergent subsequence. Note that some of values ξ_i may be zero, but not all. We now repeat the above process. For each term of the convergent subsequence for w_1, we can find a vector corresponding to a solution for the edges of $G_2\backslash T_2$. The lengths of these vectors is bounded, since $|w_1| = 1$. Again by compactness there is a convergent subsequence converging to w_2 and assigning the coefficients of w_2 to $G_2\backslash T_2$ will be consistent with assigning the coefficients of w_1 to the lengths of the edges of $G_1\backslash T_1$. Note that if an edge has been assigned zero length then when subdivided the parts have zero length and it can be part of a Type I fold with another edge of zero length. Again repeating this process we can eventually assign lengths to all the edges of $G_n\backslash T_n$ for every n which are consistent with the folding process. If all these lengths are non-zero then we have realized the folding sequence as a folding sequence of simplicial $\mathbf{R}$–trees. If some of the edges have zero length assigned to them, then it is easy to see that the folding sequence is reducible. Thus we take $E_m \subset ET_m$ to be the set of edges assigned zero length. □

It is easy to construct the *limit* of such a folding sequence of $\mathbf{R}$–trees. Let $\theta_n = \rho_n\rho_{n-1}\dots\rho_1\colon T_1 \to T_{n+1}$. Let d_n be the $\mathbf{R}$–tree metric in T_n. We define a pseudometric d in T_1 by $d(x,y) = \lim_{n\to\infty}(d_n(\theta_n(x)), d_n(\theta_n(y)))$. We put $T = T_1/\sim$, where $x \sim y$ if $d(x,y) = 0$. Clearly d induces a metric on T and this metric space is called the limit of the folding sequence.

I am grateful to Brian Bowditch for supplying the proof of the following theorem.

Theorem 3.2 *The limit T of the folding sequence T_n is an $\mathbf{R}$–tree.*

Proof Let (S,d) be a metric space. In the terminology of [3], d is a path metric if given any two points $X, Y \in S$ and $\epsilon > 0$ there is a rectifiable path joining X and Y of length at most $d(X,Y) + \epsilon$. Each (T_n, d_n) satisfies the stronger condition that any two points $X, Y \in T_n$ are joined by a path of length $d(X,Y)$. Since for any $x, y \in T_1, (d_n(\theta_n(x)), d_n(\theta_n(y)))$ is a decreasing sequence, it follows easily that d as defined above is a path metric on T. It now follows from [3] Proposition 3.4.2 that T is an $\mathbf{R}$–tree if given any four points X, Y, Z, W they can be partitioned into two sets of two elements, without loss of generality, $\{\{X, y\}, \{Z, W\}\}$, so that

$$d(X,Y) + d(Z,W) \le d(X,Z) + d(Y,W) = d(Y,Z) + d(X,W).$$

Since this condition is satisfied in each T_n, it must also be satisfied in T. Thus T is an $\mathbf{R}$–tree. □

If G is the direct limit in the category of groups of the sequence of homomorphisms $\rho_n\colon G_n \to G_{n+1}$ then there is an action of G on T via isometries. Suppose in addition the folding sequence satisfies the following condition

(c) Two edges of T_n cannot be folded together if they arose as subdivided parts of the same edge of T_m for some $m < n$.

In this case the natural map $\phi_n\colon T_n \to T$ restricts to an isometry on each edge of T_n and it is therefore a morphism of $\mathbf{R}$–trees. It is easy to check that T

is the direct limit of the sequence of folding morphisms in the category $\mathcal{T}$ of $\mathbf{R}$–trees and morphisms.

As noted above, it is best to describe folding operations in terms of their effect on the quotient graphs $G_n\backslash T_n$. Note that (c) applies to T_n and not to $G_n\backslash T_n$. Thus it is possible for the n-th fold in the folding sequence to fold together edges that arose as subdivided edges of $G_m\backslash T_m$ for some $m < n$. An example of this is given in [8]. What happens is that, in T_n, the edges folded together occur as subdivided parts of different edges in the same G_m–orbit in T_m.

Let G be a finitely generated group. Suppose we have an infinite folding sequence with limit T and suppose that $\tilde{\phi}_n\colon G_n \to G$ is not an isomorphism for any n. This means that the folding sequence must have infinitely many vertex morphisms. There is then an induced folding sequence of reduced trees. We examine the induced folds listed above. For induced folds of type I, III and 3 there is a decrease in the number of orbits of edges. For a fold of type 12, 13 or 14 there is a decrease in the rank of $H_1(\overline{S}^*)$ and for a fold of type 1 there is no change in vertex groups. Thus the sequence must contain infinitely many induced folds of types other than I, III, 1, 3, 12, 13 or 14. However each such induced fold, which is not an isomorphism, produces a new edge group that properly contains one of the old edge groups. In the situation when the maps $\phi_n\colon T_n \to T$ are morphisms of $\mathbf{R}$–trees, for example if condition (c) is satisfied, each edge stabilizer of T_n fixes an arc of T. Since each T_n has finitely many orbits of edges, using a graph theoretic argument (König's Lemma) it is possible to find a sequence of edge stabilizers from a subsequence of the T_n's for which the inclusions are proper. It follows that G contains a subgroup H that is not finitely generated but every finitely generated subgroup of H fixes an arc of T. Thus we have the following result.

Theorem 3.3 *Let the G–tree T be the direct limit in $\mathcal{T}$ of the folding sequence T_n of simplicial trees, where T is a G_n–tree. Then either there exists m such that $G = G_n$ for all $n \geq m$ or G contains a subgroup H that is not finitely generated but every finitely generated subgroup of H fixes an arc of T.*

In [8] I introduced the concept of a G–protree. Protrees arise naturally in studying inaccessible groups. Let G be a finitely generated group. Let $\mathcal{B}(G)$ denote the Boolean ring consisting of all subsets $a \subset G$ of almost invariant sets. Thus $a \in \mathcal{B}(G)$ if and only if the sets a and ag are almost equal for every $g \in G$. In [4] it is shown that there is a *nested* G–set E which generates $\mathcal{B}(G)$ as a Boolean ring. The group G is accessible if and only if E can be chosen to be G–finite, in which case E can be regarded as the edge set of a simplicial G–tree. If G is inaccessible then E is not G–finite. In this case E is a combinatorial object called a nice G–protree, which has a realization (also called a G–protree) as an $\mathbf{R}$–tree in which the set of branch points intersects each segment in a nowhere dense subset.

If G is finitely generated, then any G–tree T is a strong limit of a sequence T_n of $\mathbf{R}$–trees, where T_n is a G_n–tree and the action is geometric, ie it arises from a foliation on a finite 2–complex. See [11] for a precise definition and a proof of the above statement. However in a geometric action an orbit which is nowhere dense must be discrete (see [11]). Thus if G is finitely generated and T is a G–protree, then T is a strong limit of a folding sequence of simplicial trees. This gives the following result.

Theorem 3.4 *Let G be a finitely generated group and let P be a nice G–protree. Then either*

(i) *there is a reduced G–tree T such that for every $v \in VT, G_v$ is finitely generated and fixes a vertex of P and for every $e \in ET, G_e$ is finitely generated and fixes an edge of P,*

or

(ii) *the group G contains a subgroup H that is not finitely generated but every finitely generated subgroup of H fixes an edge of P.*

Note that if G is finitely presented then $\tilde{\phi}_n$ must be an isomorphism for n large and so (i) must hold. This can be used to give a proof that finitely presented groups are accessible. This was first proved in [5]. We have seen that if G is finitely generated then we can construct a G–protree P corresponding to a nested set of generators of $\mathcal{B}(G)$. There is then a folding sequence which has limit P. If the situation (i) of Theorem 3.4 prevails then for each $v \in VT, G_v$ will have at most one end and so G will be accessible. Thus if G is inaccessible then condition (ii) must be satisfied giving the following result.

Theorem 3.5 *Let G be a finitely generated inaccessible group. Then G contains an infinite locally finite subgroup.*

Proof This follows immediately from Theorem 3.4. □

Corollary 3.6 *Let G be a finitely generated discrete convergence group acting on S^2. Then G is accessible.*

Proof By Theorem 3.5 it suffices to show that a locally finite discrete convergence group must be finite. Suppose that H is an infinite locally finite discrete convergence group acting on S^2. By [10] Theorem 5.11, $L(H)$ (the set of limit points of H) consists of exactly one point x_0, which is fixed by H. A finite group of homeomorphisms with a fixed point is conjugate in $Hom(S^2)$ to a cyclic or dihedral group acting linearly on S^2 . An increasing chain of such groups would have to have two fixed points, contradicting the statement above that there is a unique fixed point.

References

[1] **M Bestvina**, **M Feighn**, *Bounding the complexity of simplicial group actions on trees*, Invent. Math. 103 (1991) 449–469

[2] **M Bestvina**, **M Feighn**, *A counterexample to generalized accessibility*, from: "Arboreal Group Theory", R C Alperin (editor), MSRI Publications 19, Springer–Verlag (1991) 133–142

[3] **B H Bowditch**, *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, from: "Group Theory from a Geometrical Viewpoint", E Ghys, A Haefliger and A Verjovsky (editors), World Scientific (1991) 64–167

[4] **W Dicks**, **M J Dunwoody**, *Groups acting on graphs*, Cambridge University Press, Cambridge, UK (1989)

[5] **M J Dunwoody**, *The accessibility of finitely presented groups*, Invent. Math. 81 (1985) 449–57

[6] **M J Dunwoody**, *Groups acting on* **R***–trees*, Commun. Algebra 19 (1991) 2125–2136 *Erratum*, Commun. Algebra 19 (1991) 3191

[7] **M J Dunwoody**, *An inaccessible group*, from: "The Proceedings of Geometric Group Theory 1991", G A Niblo and M Roller (editors), LMS Lecture Note Series 181, Cambridge University Press (1993) 75–78

[8] **M J Dunwoody**, *Inaccessible groups and protrees*, J. Pure Appl. Algebra 88 (1993) 63–78

[9] **M J Dunwoody**, *Groups acting on protrees*, J. London Math. Soc. 56 (1997) 125–136

[10] **F W Gehring**, **G J Martin**, *Discrete quasiconformal groups I*, Proc. London Math. Soc. 55 (1987) 331–358

[11] **G Levitt**, **F Paulin**, *Geometric group actions on trees*, Amer. J. Math. 119 (1997) 83–102

[12] **P A Linnell**, *On accessibility of groups*, J. Pure Appl. Algebra 30 (1983) 39–46

[13] **G J Martin**, **R K Skora**, *Group actions on the* 2*–sphere*, Amer. J. Math. 111 (1989) 387–402

[14] **Z Sela**, *Acylindrical accessibility for groups*, Invent. Math. 129 (1997) 527–565

[15] **R K Skora**, *Splittings of surfaces*, Bull. Amer. Math. Soc. 23 (1990) 85–90

[16] **J R Stallings**, *Foldings of G–trees*, from: "Arboreal Group Theory", R C Alperin (editor), MSRI Publications 19, Springer–Verlag (1991) 355–368

Faculty of Math.Studies
University of Southampton,
Southampton, SO9 5NH, UK

Email: `mjd@maths.soton.ac.uk`

Received: 27 October 1997

ISSN 1464-8997 **159**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 159–166

Characterisation of a class of equations with solutions over torsion-free groups

Roger Fenn
Colin Rourke

Abstract We study equations over torsion-free groups in terms of their "t–shape" (the occurences of the variable t in the equation). A t–shape is *good* if any equation with that shape has a solution. It is an outstanding conjecture [5] that all t–shapes are good. In [2] we proved the conjecture for a large class of t–shapes called *amenable*. In [1] Clifford and Goldstein characterised a class of good t–shapes using a transformation on t–shapes called the *Magnus derivative*. In this note we introduce an inverse transformation called *blowing up*. Amenability can be defined using blowing up; moreover the connection with differentiation gives a useful characterisation and implies that the class of amenable t–shapes is strictly larger than the class considered by Clifford and Goldstein.

AMS Classification 20E34, 20E22; 20E06, 20F05

Keywords Groups, adjunction problem, equations over groups, shapes, Magnus derivative, blowing up, amenability

1 Introduction

Let G be a group. An expression of the form

$$r = g_1 t^{\varepsilon_1} g_2 t^{\varepsilon_2} g_3 \cdots t^{\varepsilon_k} = 1, \tag{1}$$

where $k \geq 1$, $g_i \in G$ and $\varepsilon = \pm 1$, is called an *equation* over G in the *variable* t with *coefficients* $g_1, g_2, \ldots, g_k$. The equation is said to have a *solution* if G embeds in a group H containing an element t for which (1) holds. This is equivalent to saying that the natural map

$$G \longrightarrow \frac{G * \langle t \rangle}{\langle r = 1 \rangle}$$

is injective.

The equation is said to be *reduced* if it contains no subword tt^{-1} or $t^{-1}t$ (ie each coefficient which separates a pair t, t^{-1} is non-trivial). The equation is said to be *cyclically reduced* if all cyclic permutations are reduced and, unless explicitly stated otherwise, all equations are assumed to be cyclically reduced.

The *t–shape* of the word r is the sequence $t^{\varepsilon_1}t^{\varepsilon_2}\cdots t^{\varepsilon_k}$.

We use the abbreviated notation t^m for the sequence $tt\cdots t$ (m times) and t^{-m} for the sequence $t^{-1}t^{-1}\cdots t^{-1}$ (m times). We call the t–shape t^m ($m\in\mathbb{Z}$, $m\neq 0$) a *power* shape. If a t–shape is not a power then after cyclic permutation it can be written in the form

$$t^{r_1}t^{-r_2}t^{r_3}\cdots t^{-r_u},\ u>1$$

where each r_i is positive.

The sum $\varepsilon = r_1 - r_2 + \ldots - r_u$ is called the *degree* of the t–shape. The sum $w = r_1 + r_2 + \ldots + r_u$ is called the *width* of the t–shape. Note that the width is the length of the corresponding equation.

We call a cyclic t–shape *good* if any corresponding equation with torsion-free coefficients has a solution.

Conjecture [5] *All t–shapes are good.*

The conjecture is a special case of the adjunction problem [6] and for a brief history, see the introduction to [2]. The torsion-free condition is necessary because the t–shape tt^{-1} is good [3] but for example the equation $ata^2t^{-1}=1$ has no solution over a group in which a has order 4.

The conjecture is known to be true in many cases. Levin [5] has proved that power shapes are good (without the torsion-free hypothesis). Klyachko [4] has proved that t–shapes of degree ±1 are good. Furthermore both Clifford and Goldstein [1] and ourselves [2] have extended Klyachko's results to larger classes of t–shapes. The class of good t–shapes in [1] are characterised in terms of the *Magnus derivative* and for definitiveness we will call them *CG–good*. The class of good t–shapes in [2] are called *amenable*. No usable characterisation of amenability was given in [2] and it is the purpose of this note to supply such a characterisation and to compare the two classes.

The rest of the paper is organised as follows. In the next section (section 2) we review the Magnus derivative (an operation on t–shapes which we refer to simply as *differentiation*) and define the class of CG–good shapes. In section 3 we define another operation on t–shapes called *blowing up* and prove that it is the inverse of differentiation. Finally in section 4 we give two simple characterisations of amenable shapes. The first in terms of blowing up and the second, similar to the characterisation of CG–good shapes, in terms of differentiation. We conclude that the class of amenable shapes is strictly larger than the class of CG–good shapes.

Acknowledgements We are grateful to Martin Edjvet for suggesting that there might be a connection between the results of the Clifford–Goldstein paper and ours. We thank the referee for helpful comments.

2 The Magnus derivative

Let $T = t^{\varepsilon_1} t^{\varepsilon_2} \cdots t^{\varepsilon_w}$, where $\varepsilon_i = \pm 1$, be a t–shape. We regard T as a cyclic t–shape and we define the cyclic t–shape $D(T)$, the *Magnus derivative* or simply *derivative* of T, as follows.

Arrange the signs of the exponent powers around a circle. The t–shape is well defined by this up to cyclic symmetry. Between each occurence of $+,+$ insert a new $+$, between each occurence of $-,-$ insert a new $-$ and in all other cases do nothing. Now delete the original signs. The remaining cyclic sequence of signs defines a new t–shape, $D(T)$.

For example $\quad tttt^{-1}tt^{-1}t^{-1}t \ \stackrel{D}{\to}\ ttt^{-1}t \ \stackrel{D}{\to}\ tt.$

The following is easy to prove.

Lemma *Let the cyclic t–shape T have degree $\varepsilon(T)$ and width $w(T)$ then:*

1) $\varepsilon(DT) = \varepsilon(T)$.

2) $w(DT) \le w(T)$ *with equality if and only if T is empty or a power shape.*

3) $D(T) = T$ *if and only if T is empty or a power shape.*

4) $D^\alpha(T)$ *is empty or a power shape if $\alpha > w(T)/2$.*

5) *If $T = t^{r_1} t^{-r_2} t^{r_3} \cdots t^{-r_k}$, where $r_i \ge 1$, is not a power shape then $DT = t^{r_1 - 1} t^{-r_2+1} \cdots t^{-r_k+1}$.* □

We can illustrate the effect of differentiation by looking at the *graph* of the t–shape $T = t^{\varepsilon_1} t^{\varepsilon_2} \cdots t^{\varepsilon_w}$.

This is a function $f = f_T \colon [0, w] \to \mathbb{R}$ defined as follows. Define $f(0) = 0$ and for integers i in the range $0 < i \le w$ $f(i) = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_i$. Extend f over the whole interval by piecewise-linear interpolation. Notice that the graph of the t–shape starts at $(0, 0)$ and finishes at (w, ε).

Figure 1 shows the graph of the example above and the effect of differentiation which 'smooths off' the peaks and troughs until a straight line graph is left.

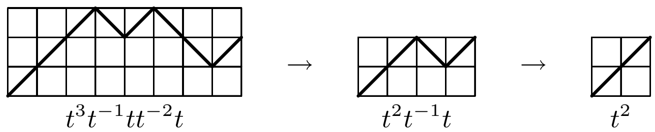


Figure 1: Differentiation

A *clump* in a cyclic t–shape is defined to be a maximal connected subsequence of the form t^m where $|m| > 1$. A *one-clump shape* is a shape with just one clump, which is not the whole sequence, ie, after possible cyclic permutation and inversion, a shape of the form $t^m t^{-1}(tt^{-1})^r$ where $m > 1$ and $r \geq 0$. We can now define CG–good. A t–shape is *CG–good* if, after a (possibly empty) sequence of differentiations it becomes a one-clump shape.

Theorem (Clifford–Goldstein [1]) *All CG–good shapes are good.* □

3 Blowing up

We shall now introduce the notion of *blowing up* of a t–shape which was implicit in [2].

We consider non-cyclic t–shapes whose graphs start and end at level 0 and which lie between levels $-m$ and 0. Such a t–shape will be called an *m–block*. An m–block whose graph reaches level $-m$ at some point will be called a *full m–block*.

Definition *m–blow up* Start with a given cyclic t–shape. Between each pair $t^{-1}t$ (ie at local minima of the graph) insert a full m–block. Between other pairs insert a general m–block (see figure 2).

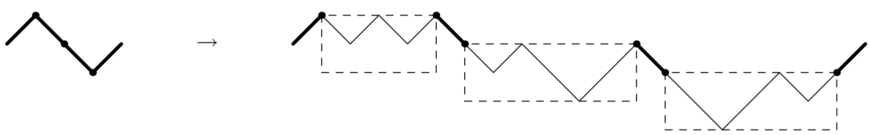


Figure 2: An example of a 2–blow-up

The definition of blow up is not explicit in [2]. However we shall see later that it coincides with the concept of normal form given on page 69 of [2].

Notice that a 0–blow up of a shape T is the original shape T but that, in general, the result of blowing up depends on the choices of the blocks. We use the notation $B^m(T)$ for the set of m–blow ups of T and we abbreviate B^1 to B.

We now prove that blowing up is anti-differentiation.

Lemma 3.1 *$U \in B(T)$ if and only if $D(U) = T$.*

Proof We give a graphical description of D. Start with the graph of a t–shape T. Introduce a new vertex halfway along each edge of the graph. At each local maximum (respectively minimum) join the new vertices just below (respectively above) and truncate. Now contract the horizontal edges and discard the old vertices. The result is the graph of $D(T)$.

This process is illustrated in figure 3, where the new vertices are open dots and the old vertices are black dots.

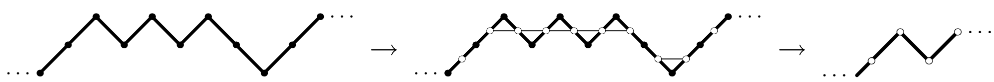


Figure 3: Graphical differentiation

To see the connection with 1–blow ups consider the following alternative description. Introduce the new vertices as before but slide them up to the top of the edges. Discard all the locally minimal vertices of the graph of T and again reduce the resulting graph by contracting horizontal edges (see figure 4). In this description it is clear that the discarded pieces are precisely 1–blocks and the lemma follows. □

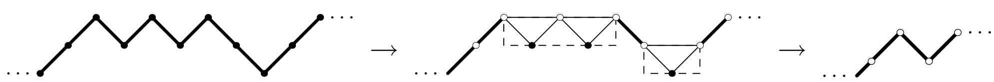


Figure 4: Differentiation and 1–blow up

For the next lemma we need to extend differentiation and blowing up to m–blocks. If T is an m–block then we define an n–blow up by inserting full n–blocks at local minima and general n–blocks at all other vertices, including the first and last vertex (in other words we prefix and append a general n–block). It can then be seen that the n–blow up of an m–block is an $(m+n)$–block and if the original block is full, then the blow up is also full.

We extend differentiation by using the same rule as for cyclic t–shapes. In graphical terms it has the same meaning as in the last proof: Discard all the locally minimal vertices of the graph and reduce by contracting horizontal edges. The proof of the previous lemma then shows that B and D are inverse operations on m–blocks.

Lemma 3.2 (a) $B \circ B^m \subset B^{m+1}$ (b) $DB^{m+1} \subset B^m$.

Proof A 1–blow up of an m–blow up can be obtained by 1–blowing up the inserted m–blocks. Part (a) now follows from the remarks above. To see part (b) observe that D of a $(m+1)$–blow up is obtained by differentiating the inserted pieces and thus results in an m–blow up. □

Corollary 3.3 (a) $B \circ B^m = B^{m+1}$ (b) $B^n = B \circ \ldots \circ B$ *(n factors)* (c) $B^n \circ B^m = B^{n+m}$.

Proof (a) By part (a) of lemma 3.2 we just have to show that if $U \in B^{m+1}(T)$ then $U \in B \circ B^m(T)$. But $D(U) \in B^m(T)$ by part (b), and $U \in B(D(U))$ by lemma 3.1 and hence $U \in B(D(U)) \subset B \circ B^m(T)$.

Parts (b) and (c) follow by induction. □

Corollary 3.4 $U \in B^n(T)$ *if and only if* $D^n(U) = T$.

Proof Repeat lemma 3.1 n times. □

We now turn to the connection of blowing up with the concept of normal form defined in [2].

On page 69 of [2] we define a word in *normal form* based on a particular cyclic t–shape T as a word obtained from T by inserting elements of certain subsets (X, J and Y defined on page 65) of the kernel of the exponential map $\varepsilon\colon G*\langle t\rangle \to \mathbb{Z}$ at *top* (between t and t^{-1}), *middle* (between t and t or t^{-1} and t^{-1}) and *bottom* (between t^{-1} and t) positions respectively. Inspecting the definitions of X, J and Y, it can be seen that this corresponds to inserting m–blocks and then allowing a controlled amount of cancellation. To be precise, define a *leading string* of an m–block to be an initial string $t^{-1}t^{-1}\ldots t^{-1}$ and a *trailing string* to be a final string $tt\ldots t$. Cancellation is allowed for specified leading and trailing strings of all blocks. The defining condition on X is that the graph of the corresponding block must meet level 0 after deletion of leading and trailing strings and the defining condition for Y is that the block must be full. There is no condition on J. We call the blocks corresponding to elements of X, J and Y, *top*, *middle* and *bottom* blocks, respectively and we denote the set of words in normal form based on the cyclic t–shape T by $NF(T)$.

Lemma 3.5 $NF(T) = B^m(T)$.

Proof Blowing up corresponds to normal form with no cancellation allowed and hence $NF(T) \supset B^m(T)$. For the converse suppose that U is in normal form based on T and that for a particular top block D the leading t^{-1} is allowed to cancel. Define the $(m-1)$–block B by $D = t^{-1}BtC$ (see figure 5). Then figure 5 makes clear that U can also be obtained by appending B to the block inserted in the previous place and replacing D by C. After these substitutions there are fewer allowed cancellations.

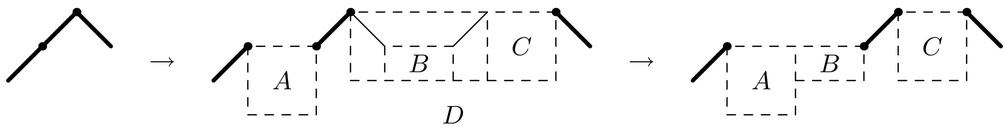


Figure 5: The simplification move

Similar arguments simplify the situation if cancellation takes place at the end of a top block or at either end of a middle block. (Notice that no cancellation can take place at bottom blocks.) Thus by repeating simplifications of this type a finite number of times, we see that U is an m–blow up of T. □

4 Amenability

We now recall the definition of *amenable* t–shapes from [2].

Recall that a *clump* in a cyclic t–shape is a maximal connected subsequence of the form t^m or t^{-m} where $m > 1$. These are said to have *order* m and $-m$ respectively. We call a clump of positive order an *up* clump and a clump of negative order a *down* clump. A t–shape is said to be *suitable* if it has exactly one up clump which is not the whole sequence and possibly some down clumps, or if it has exactly one down clump which is not the whole sequence and possibly some up clumps. It follows that, after a possible cyclic rotation or inversion, a suitable t–shape has the form

$$t^s t^{-r_0} t t^{-r_1} t \dots t t^{-r_k}$$

where $s > 1$, $k \geq 0$ and $r_i \geq 1$ for $i = 0, \dots, k$.

We now define amenable t–shapes. Using lemma 3.5 above we can rephrase the definition on page 69 of [2] as follows.

Definition *Amenable t–shapes* A t–shape which is the m–blow up of a suitable t–shape is called *amenable*.

Theorem (Fenn–Rourke [2]) *Amenable shapes are good.* □

We now turn to the characterisation of amenability. Using corollary 3.4, the definition of amenability says that a shape is amenable if and only if it eventually differentiates to a suitable shape. But now a suitable t–shape is either a one clump shape or differentiates to $t^s t^{-r}$ for some $r, s \geq 1$. This in turn either eventually differentiates to tt^{-1} or to $t^s t^{-1}$ or to tt^{-r} for some $r, s \geq 2$. Now the last two are one clump shapes and so we can see that a suitable shape either eventually differentiates to a one clump shape or to tt^{-1}. To make the final characterisation of amenability as simple as possible, we make the shape tt^{-1} an honorary amenable shape (it is good [3]) and then we have the following simple characterisation.

Theorem 4.1 (Characterisation of amenability) *A shape is amenable if and only if, after a (possibly empty) sequence of differentiations, it becomes either a one-clump shape or the shape tt^{-1}.* □

Corollary 4.2 *Amenable shapes are a strictly larger class than CG–good shapes.* □

Final remarks (1) The class of amenable shapes which are not CG–good are precisely those which eventually differentiate to tt^{-1}: an example would be $tt^{-1}t^2t^{-2}$. It seems that the methods of Clifford and Goldstein can be extended with little extra work to the smaller class of shapes which eventually differentiate to the shape t^2t^{-2}. However we cannot see how to extend their methods to cover all amenable shapes.

(2) The remark at the top of page 70 of [2], which was left unproven, can be quickly proved using theorem 4.1.

References

[1] **A Clifford**, **R Z Goldstein**, *Tesselations of S^2 and equations over torsion-free groups*, Proc. Edinburgh Maths. Soc. 38 (1995) 485–493

[2] **Roger Fenn**, **Colin Rourke**, *Klyachko's methods and the solution of equations over torsion-free groups*, l'Enseign. Math. 42 (1996) 49–74

[3] **G Higman**, **B H Neumann**, **Hanna Neumann**, *Embedding theorems for groups*, J. London Maths. Soc. 24 (1949) 247–254

[4] **A Klyachko**, *Funny property of sphere and equations over groups*, Comm. in Alg. 21 (7) (1993) 2555–2575

[5] **F Levin**, *Solutions of equations over groups*, Bull. Amer. Math. Soc. 68 (1962) 603–604

[6] **B H Neumann**, *Adjunction of elements to groups*, J. London Math. Soc. 18 (1943) 4–11

School of Mathematical Sciences, Sussex University
Brighton, BN1 9QH, UK
and
Mathematics Institute, University of Warwick
Coventry, CV4 7AL, UK

Email: `R.A.Fenn@sussex.ac.uk, cpr@maths.warwick.ac.uk`

Received: 15 November 1997

ISSN 1464-8997 **167**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 167–180

At most 27 length inequalities define Maskit's fundamental domain for the modular group in genus 2

David Griffiths

Abstract

In recently published work Maskit constructs a fundamental domain $\mathcal{D}_g$ for the Teichmüller modular group of a closed surface $\mathcal{S}$ of genus $g \geq 2$. Maskit's technique is to demand that a certain set of $2g$ non-dividing geodesics $\mathcal{C}_{2g}$ on $\mathcal{S}$ satisfies certain shortness criteria. This gives an a priori infinite set of length inequalities that the geodesics in $\mathcal{C}_{2g}$ must satisfy. Maskit shows that this set of inequalities is finite and that for genus $g = 2$ there are at most 45. In this paper we improve this number to 27. Each of these inequalities: compares distances between Weierstrass points in the fundamental domain $\mathcal{S} \setminus \mathcal{C}_4$ for $\mathcal{S}$; and is realised (as an equality) on one or other of two special surfaces.

AMS Classification 57M50; 14H55, 30F60

Keywords Fundamental domain, non-dividing geodesic, Teichmüller modular group, hyperelliptic involution, Weierstrass point

0 Introduction and preliminaries

In this paper we consider a fundamental domain defined by Maskit in [8] for the action of the Teichmüller modular group on the Teichmüller space of a closed surface of genus $g \geq 2$ in the special case of genus $g = 2$. McCarthy and Papadopoulos [9] have also defined such a fundamental domain, modelled on a Dirichlet region; for punctured surfaces there is the celebrated cell decomposition and associated fundamental domain due to Penner [10]. For genus $g = 2$ Semmler [11] has defined a fundamental domain based on locating the shortest dividing geodesic. Also for low signature surfaces the reader is referred to the papers of Keen [3] and of Maskit [7], [8].

Throughout $\mathcal{S}$ will denote a closed orientable surface of genus $g = 2$, with some fixed hyperbolic metric. We say that a simple closed geodesic γ on $\mathcal{S}$

is: *dividing* if $\mathcal{S} \setminus \gamma$ has two components; or *non-dividing* if $\mathcal{S} \setminus \gamma$ has one component. By *non-dividing geodesic* we shall always mean simple closed non-dividing geodesic. We denote the length of γ with respect to the hyperbolic metric on $\mathcal{S}$ by $l(\gamma)$. Let $|\alpha \cap \beta|$ denote the number of intersection points of two distinct geodesics α, β.

We define a *chain* $\mathcal{C}_n = \gamma_1, \dots, \gamma_n$ to be an ordered set of non-dividing geodesics such that: $|\gamma_i \cap \gamma_{i+1}| = 1$ for $1 \leq i \leq n-1$ and $\gamma_i \cap \gamma_{i'} = \emptyset$ otherwise. We say that a chain $\mathcal{C}_n$ has *length* n, where $1 \leq n \leq 5$. Likewise we define a *bracelet* $\mathcal{B}_n = \gamma_1, \dots, \gamma_n$ to be an ordered set of non-dividing geodesics such that: $|\gamma_i \cap \gamma_{i+1}| = 1$ for $1 \leq i \leq n-1$, $|\gamma_n \cap \gamma_1| = 1$ and $\gamma_i \cap \gamma_{i'} = \emptyset$ otherwise. Again we say that $\mathcal{B}_n$ has *length* n, where $3 \leq n \leq 6$. Following Maskit, we call a bracelet of length 6 a *necklace*.

For $n \leq 4$ a chain of length n can be always be extended to a chain of length $n+1$. For $n = 4$ this extension is unique. Likewise a chain of length 5 extends uniquely to a necklace. So chains of length 4 or 5 and necklaces can be considered equivalent. We shall usually work with length 4 chains, which we call *standard*. (Maskit, for genus g, usually works with chains of length 2g+1, which he calls standard.)

As Maskit shows in [8] each surface, standard chain pair $\mathcal{S}, \mathcal{C}_4$ gives a canonical choice of generators for the Fuchsian group F such that $\mathbb{H}^2/F = \mathcal{S}$ and hence a point in $\mathcal{DF}(\pi_1(\mathcal{S}), PSL(2,\mathbb{R}))$, the set of discrete faithful representations of $\pi_1(\mathcal{S})$ into $PSL(2,\mathbb{R})$. Essentially this representation corresponds to the fundamental domain $\mathcal{S} \setminus \mathcal{C}_4$ together with orientations for its side pairing elements. As Maskit observes, it is well known that $\mathcal{DF}(\pi_1(\mathcal{S}), PSL(2,\mathbb{R}))$ is real analytically equivalent to Teichmüller space. So, we define the *Teichmüller space* of closed orientable genus $g = 2$ surfaces $\mathcal{T}_2$ to be the set of pairs $\mathcal{S}, \mathcal{C}_4$.

We say that a standard chain $\mathcal{C}_4 = \gamma_1, \dots, \gamma_4$ is *minimal* if for any chain $\mathcal{C}'_m = \gamma_1, \dots, \gamma_{m-1}, \alpha_m$ we have $l(\gamma_m) \leq l(\alpha_m)$ for $1 \leq m \leq 4$. We then define the *Maskit domain* $\mathcal{D}_2 \subset \mathcal{T}_2$ to be the set of surface, standard chain pairs $\mathcal{S}, \mathcal{C}_4$ with $\mathcal{C}_4$ minimal.

For $\mathcal{C}_4$ to be minimal the geodesics $\gamma_1, \dots, \gamma_4$ must satisfy an a priori infinite set of length inequalities. For genus g, Maskit gives an algorithm using cut-and-paste to show that only a finite number N_g of length inequalities need to be satisfied. Applying his algorithm to genus $g = 2$, Maskit showed that $N_2 \leq 45$. We establish an independent proof that $N_2 \leq 27$. We could have shown that 18 of Maskit's 45 inequalities follow from the other 27. However, by tayloring all our techniques to the special case of genus 2, we are able to produce a much shorter proof.

The fact that 18 of Maskit's 45 inequalities follow from the other 27 follows from applications of Theorem 2.2 (which appeared as Theorem 1.1 in [4]) and of Corollary 2.5. The latter follows immediately from Theorem 2.4, for which we give a proof in this paper. This is a characterisation of the octahedral surface $\mathcal{O}ct$ (the well known genus two surface of maximal symmetry group) in terms of a finite set of length inequalities.

The 27 length inequalities have the properties that: each is realised on one or other of two special surfaces (for all but 2 this special surface is $\mathcal{O}ct$); and each compares distances between Weierstrass points in the fundamental domain $\mathcal{S} \setminus \mathcal{C}_4$ for $\mathcal{S}$.

The author would like to thank Bill Harvey, Bernie Maskit, Peter Buser, Klaus-Dieter Semmler and Christophe Bavard for hospitality and helpful discussions. The author was supported for this work by the Swiss National Science Foundation on a Royal Society Exchange Fellowship at EPFL, Lausanne, Switzerland and is currently supported by the French Government as a boursier on a Sejour Scientifique.

1 The hyperelliptic involution and the main result

It is well known that every closed genus two surface without boundary $\mathcal{S}$ admits a uniquely determined *hyperelliptic involution*, an isometry of order two with six fixed points, which we denote by $\mathcal{J}$. The fixed points of $\mathcal{J}$ are known as *Weierstrass points*. Every simple closed geodesic $\gamma \subset \mathcal{S}$ is setwise fixed by $\mathcal{J}$, and the restriction of $\mathcal{J}$ to γ has no fixed points if γ is dividing and two fixed points if γ is non-dividing (see Haas–Susskind [2]). So every non-dividing geodesic on $\mathcal{S}$ passes through two Weierstrass points. It is a simple consequence that sequential geodesics in a chain intersect at Weierstrass points. We say that two non-dividing geodesics α, β *cross* if $\alpha \neq \beta$ and $\alpha \cap \beta$ contains a point that is not a Weierstrass point.

The quotient *orbifold* $\mathcal{O} \cong \mathcal{S}/\mathcal{J}$ is a sphere with six order two cone points, endowed with a fixed hyperbolic metric. Each cone point on $\mathcal{O}$ is the image of a Weierstrass point under the projection $\mathcal{J}\colon \mathcal{S} \to \mathcal{O}$ and each non-dividing geodesic on $\mathcal{S}$ projects to a simple geodesic between distinct cone points on $\mathcal{O}$ – what we shall call an *arc*. Definitions of chains, bracelets and crossing all pass naturally to the quotient.

Let $\mathcal{C}_4$ be a standard chain on $\mathcal{S}$, which extends to a necklace $\mathcal{N}$. We number Weierstrass points on $\mathcal{N}$ so that $\omega_i = \gamma_{i-1} \cap \gamma_i$ for $2 \leq i \leq 6$ and $\omega_1 = \gamma_6 \cap \gamma_1$.

Choose an orientation upon $\mathcal{S}$ and project to the quotient orbifold $\mathcal{O} \cong \mathcal{S}/\mathcal{J}$ – for the rest of the paper we shall work on the quotient orbifold $\mathcal{O}$. We label the components of $\mathcal{O} \setminus \mathcal{N}$ by $\mathcal{H}, \overline{\mathcal{H}}$ so that $\gamma_1, \dots, \gamma_6$ lie anticlockwise around $\mathcal{H}$. Label by $\beta_{j,k}^{i_1,i_2,\dots,i_n}$ (respectively $\overline{\beta_{j,k}^{i_1,i_2,\dots,i_n}}$) the arc between the cone points ω_j, ω_k ($j < k$) crossing the sequence of arcs $\gamma_{i_1}, \gamma_{i_2}, \dots, \gamma_{i_n}$ and having the subarc between ω_j, γ_{i_1} lying in $\mathcal{H}$ (respectively $\overline{\mathcal{H}}$).

Our main result is then the following. (We abuse notation so that $\beta_{1,6} = \overline{\beta_{1,6}} = \gamma_6$ and $\beta_{2,3} = \overline{\beta_{2,3}} = \gamma_2$. We then have repetitions, $l(\gamma_2) \leq l(\gamma_6)$ twice, and redundancies, $l(\gamma_2) \leq l(\gamma_2)$ also twice.)

Theorem 1.1 *The standard chain $\mathcal{C}_4$ is minimal if the following are satisfied:*

(1) $l(\gamma_1) \leq l(\gamma_i), i \in \{2,3,4,5\}$

(2) $l(\gamma_2) \leq l(\beta_{i,j}), l(\overline{\beta_{i,j}}), l(\beta_{2,5}^6), l(\overline{\beta_{2,5}^6}),\ i \in \{1,2\},\ j \in \{3,4,5,6\}$

(3) $l(\gamma_3) \leq l(\beta_{3,j}), l(\overline{\beta_{3,j}}), l(\beta_{3,4}^6), l(\overline{\beta_{3,4}^6}),\ j \in \{5,6\}$

(4) $l(\gamma_4) \leq l(\beta_{4,6}), l(\overline{\beta_{4,6}})$.

Each length $l(\gamma_i)$ or $l(\beta_{j,k})$ (respectively $l(\overline{\beta_{j,k}})$) is a distance between cone points in $\mathcal{H}$ (respectively $\overline{\mathcal{H}}$). Likewise each length $l(\beta_{j,k}^6)$, $l(\overline{\beta_{j,k}^6})$ is a distance between cone points in $\mathcal{O} \setminus \mathcal{C}_5$. So each length inequality in Theorem 1.1 compares distances between cone points in $\mathcal{O} \setminus \mathcal{C}_5$ (and hence distances between Weierstrass points in $\mathcal{S} \setminus \mathcal{C}_4$).

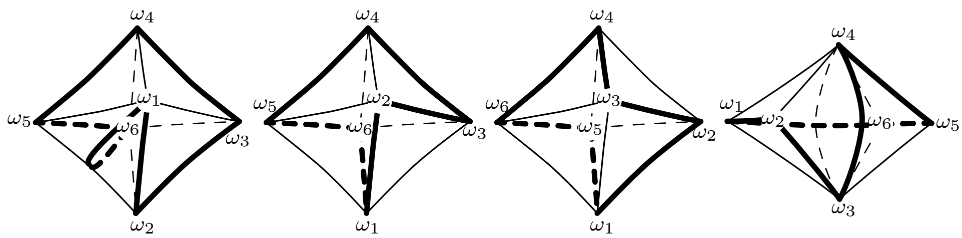


Figure 1: How the length inequalities in Theorem 1.1 are realized on $\mathcal{O}ct$ and $\mathcal{E}$

Theorem 1.1 gives a sufficient list of inequalities. As to the necessity each inequality, we make the following observation. Each inequality is realised (as an equality) on either $\mathcal{O}ct$ or $\mathcal{E}$ – cf Theorem 1.1 in [5]. The octahedral orbifold $\mathcal{O}ct$ is the well known orbifold of maximal conformal symmetry group. Any minimal standard chain on $\mathcal{O}ct$ lies in its set of shortest arcs. This arc set has the combinatorial edge pattern of the Platonic solid. The exceptional orbifold $\mathcal{E}$, which was constructed in [5], has conformal symmetry group $\mathbb{Z}_2 \times \mathbb{Z}_2$. However it is not defined by the action of its symmetry group alone, it also requires a certain length inequality to be satisfied. Any minimal standard chain on $\mathcal{E}$ lies in its set of shortest and second shortest arcs.

In Figure 1 we have illustrated necklaces on $\mathcal{O}ct$ and $\mathcal{E}$ that are the extentions of minimal standard chains. As with other figures in this paper, we use wire frame diagrams to illustrate the orbifolds. Solid (respectively dashed) lines represent arcs in front (respectively behind) the figure. Thick lines represent arcs in the necklace $\mathcal{N}$. The minimal standard chain on $\mathcal{E}$ in Figure 1 has: $l(\gamma_1) = l(\gamma_5)$; $l(\gamma_2) = l(\overline{\beta_{1,3}}) = l(\beta_{1,4}) = l(\beta_{2,4})$; $l(\gamma_3) = l(\overline{\beta^6_{3,4}}) = l(\overline{\beta_{3,5}}) = l(\overline{\beta_{3,6}})$; $l(\gamma_4) = l(\beta_{4,6})$. Making such a list for all the orbifolds in Figure 1, together with their mirror images, we see that all the inequalities in Therem 1.1 are realised as equalities on either $\mathcal{O}ct$ or $\mathcal{E}$.

2 Length inequalities for systems of arcs

In order to prove Theorem 1.1 we need a number of length inequality results for systems of arcs. Let $\mathcal{K}_4 = \kappa_{0,1}, \ldots, \kappa_{3,0}$ denote a length 4 bracelet such that each component of $\mathcal{O} \setminus \mathcal{K}_4$ contains an interior cone point. Using mod 4 addition throughout, label cone points: on $\mathcal{K}_4$ by $c_k = \kappa_{k-1,k} \cap \kappa_{k,k+1}$ for $k \in \{0, \ldots, 3\}$; and off $\mathcal{K}_4$ by c_l for $l \in \{4, 5\}$. Label by $\mathcal{O}_l$ the component of $\mathcal{O} \setminus \Upsilon$ containing c_l and label arcs in $\mathcal{O}_l$ so that $\kappa_{k,l}$ is between c_k, c_l. Let λ_k denote the arc between c_4, c_5 crossing only $\kappa_{k,k+1} \subset \mathcal{K}_4$.

The following two results appeared as Lemma 2.3 in [5] (in Maskit's terminology this is a cut-and-paste) and as Theorem 1.1 in [4] respectively.

Lemma 2.1 (i) $2l(\kappa_{0,4}) < l(\lambda_0) + l(\lambda_3)$ (ii) $2l(\kappa_{3,0}) < l(\lambda_0) + l(\lambda_2)$.

Theorem 2.2 *If* $l(\kappa_{3,4}) \leq l(\kappa_{0,4})$, $l(\kappa_{3,5}) \leq l(\kappa_{0,5})$, $l(\lambda_0) \leq l(\lambda_2)$ *then* $l(\kappa_{3,4}) = l(\kappa_{0,4})$, $l(\kappa_{3,5}) = l(\kappa_{0,5})$, $l(\lambda_0) = l(\lambda_2)$.

Corollary 2.3 *If* $l(\kappa_{3,4}) \leq l(\kappa_{0,4})$, $l(\kappa_{3,5}) \leq l(\kappa_{0,5})$, $l(\kappa_{1,4}) \leq l(\kappa_{2,4})$ *then* $l(\kappa_{1,5}) \geq l(\kappa_{2,5})$.

Proof of Corollary 2.3 Since $l(\kappa_{3,4}) \leq l(\kappa_{0,4})$, $l(\kappa_{3,5}) \leq l(\kappa_{0,5})$ Theorem 2.2 implies that $l(\lambda_0) \geq l(\lambda_2)$. Moreover $l(\kappa_{1,4}) \leq l(\kappa_{2,4})$ and so again, by Theorem 2.2, $l(\kappa_{1,5}) \geq l(\kappa_{2,5})$. □

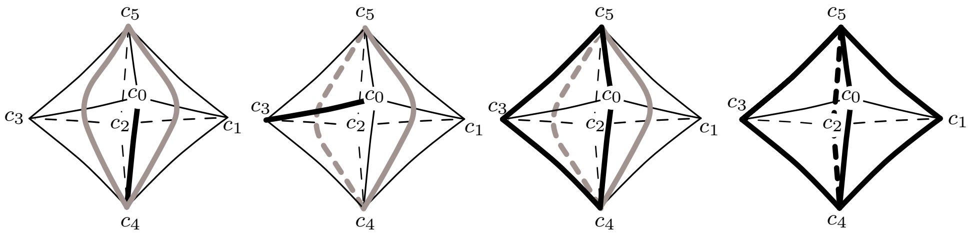


Figure 2: Arc sets for Lemma 2.1, for Theorem 2.2 and for Corollary 2.3

Theorem 2.4 *Suppose* $l(\kappa_{2,3}) \leq l(\kappa_{2,l})$, $l(\kappa_{3,0}) \leq l(\kappa_{1,2}) \leq \{l(\kappa_{0,l}), l(\kappa_{1,l})\}$ *and* $l(\kappa_{0,1}) \leq \{l(\kappa_{0,l}), l(\kappa_{3,l})\}$ *then* $l(\kappa_{k,l}) = l(\kappa_{k,k+1})$ *for each* k, l *and* $\mathcal{O}$ *is the octahedral orbifold.*

Proof of Theorem 2.4 We postpone this until Section 3. □

Corollary 2.5 *Suppose* $l(\kappa_{2,3}) \leq l(\kappa_{2,l})$, $l(\kappa_{1,2}) \leq \{l(\kappa_{0,l}), l(\kappa_{1,l})\}$ *and* $l(\kappa_{0,1}) \leq \{l(\kappa_{0,l}), l(\kappa_{3,l})\}$ *then* $l(\kappa_{3,0}) \geq l(\kappa_{1,2})$.

Proof of Corollary 2.5 If $l(\kappa_{3,0}) \leq l(\kappa_{1,2})$ then by Theorem 2.4 $l(\kappa_{k,l}) = l(\kappa_{k,k+1})$ for each k, l. In particular $l(\kappa_{3,0}) = l(\kappa_{1,2})$. So $l(\kappa_{3,0}) \geq l(\kappa_{1,2})$. □

3 The proofs

Proof of Theorem 1.1 Let α_m denote an arc such that $\mathcal{C}'_m = \gamma_1, \ldots, \gamma_{m-1}, \alpha_m$ is a chain, for $1 \leq m \leq 4$, $\alpha_m \neq \gamma_m$. We will show that $l(\gamma_m) \leq l(\alpha_m)$ for arcs of the form $\beta_{j,k}^{i_1,i_2,\ldots,i_n}$. The same arguments work for arcs of the form $\overline{\beta_{j,k}^{i_1,i_2,\ldots,i_n}}$. Let $X(\alpha, \beta)$ denote the number of crossing points of a distinct pair of arcs α, β – ie the number of intersection points of α, β that are not cone points. Let $n = \infty$, if $X(\gamma_m, \alpha_i) = 0$ for $i \in \{1, \ldots, 6\}$; otherwise, let $n =$ min $i \in \{1, \ldots, 6\}$ such that $X(\gamma_n, \alpha_i) > 0$. We note that $n \geq m$.

Let $P_{m,n,p}$ be the proposition that $l(\gamma_m) \leq l(\alpha_m)$ for $X(\alpha_m, \gamma_n) = p$. Clearly, if $n = \infty$ then $p = 0$. For $n \in \{5, 6\}$ it is not hard to show that $p = 1$. For $n \in \{1, \ldots, 4\}$ we consider $p = 1$ and $p > 1$. We order the propositions as follows: $P_{4,\infty,0}, \ldots, P_{1,\infty,0}$ which is followed by $P_{4,6,1}, P_{4,5,1}, \ldots, P_{1,6,1}, P_{1,5,1}$ followed by $P_{4,4,1}, P_{4,4,p>1}$ which is followed by $P_{3,4,1}, P_{3,4,p>1}, P_{3,3,1}, P_{3,3,p>1}$ followed by $P_{2,4,1}, P_{2,4,p>1}, \ldots, P_{2,2,1}, P_{2,2,p>1}$ followed by $P_{1,4,1}, P_{1,4,p>1}, \ldots, P_{1,1,1}, P_{1,1,p>1}$.

Suppose $n = \infty$, α_m does not cross $\mathcal{N}$. If $m > 1$ then $P_{m,\infty,0}$ is a hypothesis. If $m = 1$ then either $P_{1,\infty,0}$ is a hypothesis, $\alpha_1 = \gamma_i$ for some $i \in \{2, \ldots, 5\}$, or $P_{1,\infty,0}$ follows from the hypotheses, $l(\gamma_1) \le l(\gamma_i), l(\gamma_i) \le l(\alpha_1)$ for some $i \in \{2,3,4\}$.

Suppose $n \in \{5,6\}$, α_m crosses $\mathcal{N}$ but does not cross $\mathcal{C}_4$.

For $m = 4$, by inspection, $\alpha_4 = \beta^6_{4,5}$. So α_m, γ_m share endpoints, $n > m+1$ and we can apply the argument (i) below. So we have $P_{4,n,1}$ for $n \in \{5,6\}$.

In Figures 3,4,5 we illustrate applications of length inequalities results to the proof. As above we use wire frame figures of the octahedral orbifold, with the necklace $\mathcal{N}$ in thick black. Other arcs are in thick grey. Figures have been drawn so arcs in the application correspond to arcs in the length inequality result.

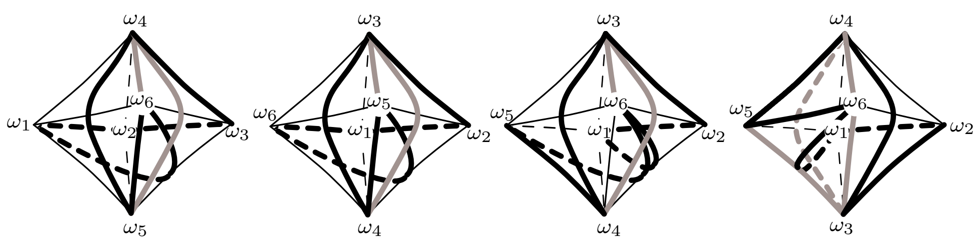


Figure 3: Application (i) for $\alpha_4 = \beta^6_{4,5}, \alpha_3 = \beta^5_{3,4}$ and $\beta^{6,5}_{3,4}$ and of Theorem 2.2, (ii) for $\alpha_3 = \beta^6_{3,5}$

For $m = 3$. By inspection, α_3 is one of $\beta^5_{3,4}, \beta^6_{3,4}, \beta^{6,5}_{3,4}, \beta^6_{3,5}$. For $\beta^5_{3,4}, \beta^{6,5}_{3,4}, \beta^6_{3,4}$: γ_m, α_m share endpoints, $n > m+1$ and so we can apply either argument (i) or (ii) below. For $\beta^6_{3,5}$ we can apply Theorem 2.2 in conjunction with argument (ii): by hypothesis $l(\gamma_4) \le l(\beta_{4,6})$ and by argument (ii) $l(\gamma_3) \le l(\beta^6_{3,4})$ and so $l(\beta^6_{3,5}) \ge l(\beta_{3,6})$. Again by hypothesis $l(\gamma_3) \le l(\beta_{3,6})$ and so $l(\gamma_3) \le l(\beta_{3,6}) \le l(\beta^6_{3,5})$. This gives $P_{3,n,1}$ for $n \in \{5,6\}$.

For $m = 2$, α_2 is one of $\beta^5_{2,3}, \beta^6_{2,3}, \beta^{6,5}_{2,3}$ or one of $\beta^5_{2,4}, \beta^6_{2,4}, \beta^{6,5}_{2,4}, \beta^6_{2,5}, \beta^5_{1,3}, \beta^5_{1,4}$. By hypothesis $l(\gamma_2) \le l(\beta^6_{2,5})$. For $\beta^6_{2,3}, \beta^{6,5}_{2,3}, \beta^5_{2,3}$ we can again apply either argument (i) or (ii). For $\beta^5_{2,4}, \beta^6_{2,4}, \beta^{6,5}_{2,4}, \beta^5_{1,3}$ we apply Theorem 2.2 in conjunction with argument (ii). We give the argument for $\beta^5_{2,4}$. By argument (ii), we have $l(\gamma_2) < l(\beta^5_{2,3})$. Also, by hypothesis, $l(\gamma_3) \le l(\beta_{3,5})$ and so by Theorem 2.2 $l(\beta_{2,5}) < l(\beta^5_{2,4})$. Again, by hypothesis, $l(\gamma_2) \le l(\beta_{2,5})$ and so $l(\gamma_2) \le l(\beta_{2,5}) < l(\beta^5_{2,4})$.

For $\alpha_2 = \beta^5_{1,4}$ we argue as follows. By hypothesis we have $l(\gamma_3) \le l(\beta_{3,5}), l(\overline{\beta_{3,6}})$ and $l(\gamma_2) \le l(\beta_{1,5}), l(\gamma_6), l(\beta_{2,5}), l(\overline{\beta_{2,6}})$ and $l(\gamma_1) \le l(\beta_{1,5}), l(\gamma_6), l(\gamma_4), l(\overline{\beta_{4,6}})$. By Corollary 2.5: $l(\beta^5_{1,4}) \ge l(\gamma_2)$. Hence $P_{2,n,1}$ for $n \in \{5,6\}$.

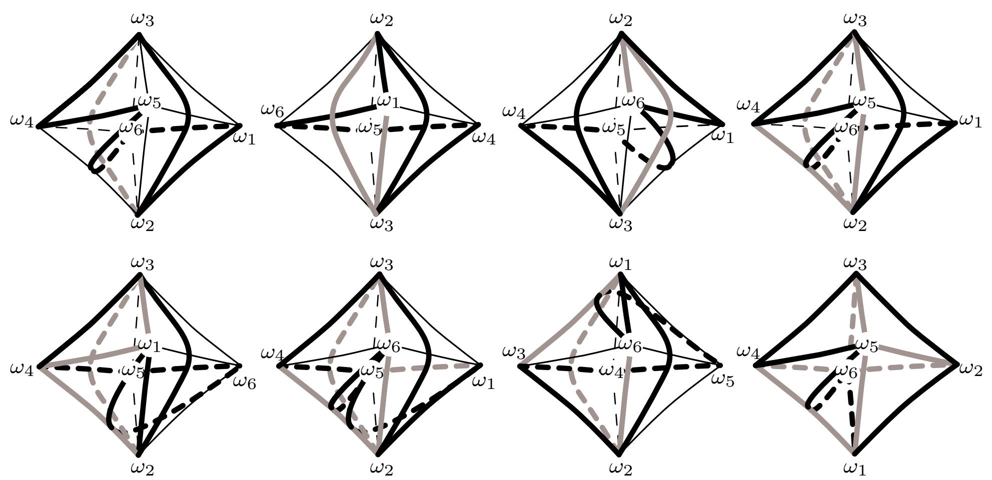


Figure 4: Applications of (i) or (ii) for $\alpha_2 = \beta^5_{2,3}, \beta^6_{2,3}$ and $\beta^{6,5}_{2,3}$; of Theorem 2.2, (ii) for $\alpha_2 = \beta^5_{2,4}, \beta^6_{2,4}, \beta^{6,5}_{2,4}$ and $\beta^5_{1,3}$; and of Corollary 2.5 for $\alpha_2 = \beta^5_{1,4}$

For $m = 1$. If $\{j,k\} \neq \{1,2\}$ or $\{j,k\} \neq \{5,6\}$ then $l(\gamma_1) \leq l(\gamma_i)$, $l(\gamma_i) \leq l(\alpha_1)$ are hypotheses, or preceding propositions, for some $i \in \{2,3,4\}$. If $\{j,k\} = \{1,2\}$ then, by inspection, $\alpha_1 = \beta^5_{1,2}$ we can again apply argument (i). By inspection there is no such α_1 for $\{j,k\} = \{5,6\}$. This completes $P_{m,n,1}$ for $n \in \{5,6\}$.

We now give the arguments for: α_m, γ_m share endpoints and $n > m+1$. The arc set $\Gamma := \alpha_m \cup \gamma_m$ divides $\mathcal{O}$ into two components. Either: (i) Γ divides one cone point (c) from three; or (ii) Γ divides two cone points from two. For (i) we let $\mathcal{O}_c$, $\mathcal{O}'_c$ denote the components of $\mathcal{O} \setminus \Gamma$ so that $c \in \mathcal{O}_c$ and we let α'_m (respectively α''_m) denote the arc between ω_m, c (respectively between ω_{m+1}, c) in $\mathcal{O}_c$.

First $m = 4$, (i), $n = 6$. None of $\gamma_1, \gamma_2, \gamma_3$ crosses $\Gamma = \alpha_4 \cup \gamma_4$, so $\mathcal{C}_3 = \gamma_1, \gamma_2, \gamma_3$ lies in one or other component of $\mathcal{O} \setminus \Gamma$. Now $\mathcal{C}_3$ contains three cone points disjoint from Γ, so $\mathcal{C}_3 \subset \mathcal{O}'_c$. So $c = \omega_6$ and $\mathcal{C}'_4 = \gamma_1, \gamma_2, \gamma_3, \alpha'_4$ is a chain. We observe – see Figure 3 – that $\alpha'_4 = \beta_{4,6}$ and hence $l(\gamma_4) \leq l(\alpha'_4)$ is a hypothesis. By Lemma 2.1(i): $2l(\alpha'_4) < l(\gamma_4) + l(\alpha_4)$ and so $l(\gamma_4) \leq l(\alpha'_4) < l(\alpha_4)$.

Second $m = 3$, (i), $n \in \{5,6\}$. Neither γ_1 nor γ_2 crosses $\Gamma = \alpha_3 \cup \gamma_3$, so $\mathcal{C}_2 = \gamma_1, \gamma_2$ lies in one or other component of $\mathcal{O} \setminus \Gamma$. Now $\mathcal{C}_2$ contains two cone points disjoint from Γ, so $\mathcal{C}_2 \subset \mathcal{O}'_c$, $c = \omega_5$ or ω_6 and $\mathcal{C}'_3 = \gamma_1, \gamma_2, \alpha'_3$ is a chain. We observe – see Figure 3 – that $\alpha'_3 = \beta_{3,5}$ or $\alpha'_3 = \beta_{3,6}$ and hence $l(\gamma_3) \leq l(\alpha'_3)$ is hypothesis. Again, by Lemma 2.1(i): $2l(\alpha'_3) < l(\gamma_3) + l(\alpha_3)$ and so $l(\gamma_3) \le l(\alpha_3') < l(\alpha_3)$. For (ii) we have that $\alpha_3 = \beta_{3,4}^6$ and $l(\gamma_3) \le l(\beta_{3,4}^6)$ is a hypothesis.

Next $m = 2$, (i), $n \in \{4,5,6\}$. The arc γ_1 does not cross $\Gamma = \alpha_2 \cup \gamma_2$, so $\gamma_1 \subset \mathcal{O}_c'$ and $c \in \{\omega_4, \omega_5, \omega_6\}$ (respectively $\gamma_1 \subset \mathcal{O}_c$ and $c = \omega_1$). For $n \in \{5,6\}$ – see Figure 4 – we have that $\alpha_2' = \beta_{2,6}$ (respectively $\alpha_2'' = \beta_{1,3}$). For $n = 4$ – see Figure 5 – we have that $\alpha_2' = \beta_{2,4}$ or $\beta_{2,5}$ (respectively there is no such α_2). So $l(\gamma_2) \le l(\alpha_2')$ (respectively $l(\gamma_2) \le l(\alpha_2'')$) is a hypothesis. By Lemma 2.1(i): $2l(\alpha_2')$ or $2l(\alpha_2'') < l(\gamma_2) + l(\alpha_2)$ and so $l(\gamma_2) \le l(\alpha_2') < l(\alpha_2)$ (respectively $l(\gamma_2) \le l(\alpha_2'') < l(\alpha_2)$).

For (ii), again, γ_1 lies in one component of $\mathcal{O} \setminus \Gamma$. Let α_2''' denote the unique arc disjoint from Γ in this component of $\mathcal{O} \setminus \Gamma$. For $n \in \{5,6\}$ – again see Figure 4– we have that $\alpha_2''' = \gamma_6$. For $n = 4$ – again see Figure 5 – we have $\alpha_2''' = \beta_{1,4}$ or $\beta_{1,5}$. So $l(\gamma_2) \le l(\alpha_2''')$ is a hypothesis. By Lemma 2.1(ii): $2l(\alpha_2''') < l(\gamma_2) + l(\alpha_2)$ and so $l(\gamma_2) \le l(\alpha_2''') < l(\alpha_2)$.

Finally, $m = 1$, (i), $n \in \{3, \ldots, 6\}$. For $n \in \{5,6\}$: $\alpha_1' = \overline{\beta_{2,6}}$ and $l(\gamma_2) \le l(\alpha_1')$ is a hypothesis. For $n \in \{3,4\}$: $l(\gamma_2) \le l(\alpha_1')$ is a proceeding proposition. Since $l(\gamma_1) \le l(\gamma_2)$ is a hypothesis, we have that $l(\gamma_1) \le l(\gamma_2) \le l(\alpha_1')$. By Lemma 2.1(i): $2l(\alpha_1') < l(\gamma_1) + l(\alpha_1)$ and so $l(\gamma_1) \le l(\alpha_1') < l(\alpha_1)$.

For (ii), $n \in \{5,6\}$, there is no such α_1. For $n \in \{3,4\}$, we let α_3' denote the unique arc disjoint from Γ in the same component of $\mathcal{O} \setminus \Gamma$ as γ_2. Here $\mathcal{C}_3' = \gamma_1, \gamma_2, \alpha_3'$ is a chain and so $l(\gamma_3) \le l(\alpha_3')$ is a proceeding proposition. Since $l(\gamma_1) \le l(\gamma_3)$ is a hypothesis, we have that $l(\gamma_1) \le l(\gamma_3) \le l(\alpha_3')$. By Lemma 2.1(ii): $2l(\alpha_3') < l(\gamma_1) + l(\alpha_1)$ and so $l(\gamma_1) \le l(\alpha_3') < l(\alpha_1)$.

Now suppose $n \in \{1, \ldots, 4\}$, α_m crosses $\mathcal{C}_4$.

Lemma 3.1 *Suppose that either $X(\alpha_m, \gamma_n) > 1$ or α_m, γ_n share an endpoint. Then there exist arcs α_m', γ_n' between the same respective endpoints as α_m, γ_n such that $l(\alpha_m') < l(\alpha_m)$ or $l(\gamma_n') < l(\gamma_n)$; $X(\alpha_m', \gamma_n), X(\gamma_n', \gamma_n) < X(\alpha_m, \gamma_n)$; and $X(\alpha_m', \gamma_i) = X(\gamma_n', \gamma_i) = 0$ for $i \le n-1$. In particular $\mathcal{C}_m' = \gamma_1, \ldots, \gamma_{m-1}, \alpha_m', \mathcal{C}_n'' = \gamma_1, \ldots, \gamma_{n-1}, \gamma_n'$ are both chains.*

Proof This result is essentially Proposition 3.1 in [5], with additional observations upon the number of crossing points. However, upon going through the proof, these observations become clear. □

The following argument gives $P_{m,n,p>1}$: it uses induction on p, the first induction step being the set of propositions that precede $P_{m,n,p>1}$.

Let $X(\alpha_m, \gamma_n) = p > 1$ and so by Lemma 3.1 there exist arcs α'_m, γ'_n as stated. Let $p' = X(\alpha'_m, \gamma_n) < p$, $p'' = X(\gamma'_m, \gamma_n) < p$. We note that $l(\gamma_m) \leq l(\alpha'_m)$ is either: $P_{m,n,p'>1}$ if $p' > 1$; or a preceding proposition if $p' \leq 1$. Likewise, $l(\gamma_n) \leq l(\gamma'_n)$ is either: $P_{m,n,p''>1}$ if $n = m$ and $p'' > 1$; or a preceding proposition if $n > m$ or $p'' \leq 1$. Since $l(\alpha'_m) < l(\alpha_m)$ or $l(\gamma'_n) < l(\gamma_n)$ it follows, by induction on p, that $l(\gamma_m) \leq l(\alpha'_m) < l(\alpha_m)$.

So, for the rest of the proof, we may suppose that $X(\alpha_m, \gamma_n) = 1$.

Lemma 3.2 *Suppose that α_m, γ_n have distinct endpoints and that $k > n+1$. Then there exist arcs α'_m, γ'_n between ω_j, ω_{n+1} and ω_n, ω_k such that $l(\alpha'_m) < l(\alpha_m)$ or $l(\gamma'_n) < l(\gamma_n)$ and $X(\alpha'_m, \gamma_i) = X(\gamma'_n, \gamma_i) = 0$ for $i \leq n$. In particular $\mathcal{C}'_m = \gamma_1, \ldots, \gamma_{m-1}, \alpha'_m, \mathcal{C}''_n = \gamma_1, \ldots, \gamma_{n-1}, \gamma'_n$ are both chains.*

Proof This is essentially Lemma 3.3 in [5], again with additional observations upon the number of crossing points. Again, these observations are clear. □

We now give two general arguments using these two lemmas.

Suppose: (1) α_m, γ_n share an endpoint. Again we can apply Lemma 3.1: there exist arcs α'_m, γ'_n as stated. In particular $X(\alpha'_m, \gamma_i) = X(\gamma'_n, \gamma_i) = 0$ for $i \leq n$. So $l(\gamma_m) \leq l(\alpha'_m)$, $l(\gamma_n) \leq l(\gamma'_n)$ are both preceding propositions. Since $l(\alpha'_m) < l(\alpha_m)$ or $l(\gamma'_n) < l(\gamma_n)$, it follows that $l(\gamma_m) \leq l(\alpha'_m) < l(\alpha_m)$.

Suppose: (2) α_m, γ_n have distinct endpoints and $k > n + 1$. By Lemma 3.2 there exist arcs α'_m, γ'_n as stated. Again $l(\gamma_m) \leq l(\alpha'_m), l(\gamma_n) \leq l(\gamma'_n)$ are both preceding propositions. As $l(\alpha'_m) < l(\alpha_m)$ or $l(\gamma'_n) < l(\gamma_n)$, we have that $l(\gamma_m) \leq l(\alpha'_m) < l(\alpha_m)$.

For $m = 4 : j = 4, k \in \{5, 6\}$ and $n = 4 : \alpha_4, \gamma_4$ share the endpoint ω_4 (1).

For $m = 3 : j = 3, k \in \{4, 5, 6\}$. For $n = 4$ if $k \in \{4, 5\}$ then α_3, γ_4 share the endpoint ω_k (1); if $k = 6$ then α_3, γ_4 have distinct endpoints and $k > n + 1$ (2). For $n = 3 : \alpha_3, \gamma_3$ share the endpoint ω_3 (1).

For $m = 2 : j \in \{1, 2\}, k \in \{3, \ldots, 6\}$. For $n = 4$ if $k = 3$ then, by inspection, α_2 is one of $\beta^4_{2,3}, \beta^{4,5,6}_{2,3}, \beta^{5,4}_{2,3}, \beta^{6,4}_{2,3}, \beta^{6,5,4}_{2,3}$, and we can apply argument (i) or (ii), or is one of $\beta^4_{1,3}, \beta^{4,5,6}_{1,3}, \beta^{5,4}_{1,3}$, and we apply Theorem 2.2 in conjunction with argument (ii) – see Figure 5. If $k \in \{4, 5\}$ (1); if $k = 6$ (2). For $n = 3$ if $k \in \{3, 4\}$ (1); if $k \in \{5, 6\}$ (2). For $n = 2$ if $k = 3$ (1); if $k \in \{4, 5, 6\}$ (2).

Finally $m = 1$. Suppose $n = 4$. If $\{j, k\} \neq \{1, 2\}$ or $\{j, k\} \neq \{5, 6\}$ then $l(\gamma_1) \leq l(\gamma_i), l(\gamma_i) \leq l(\alpha_1)$ are both preceding propositions for some

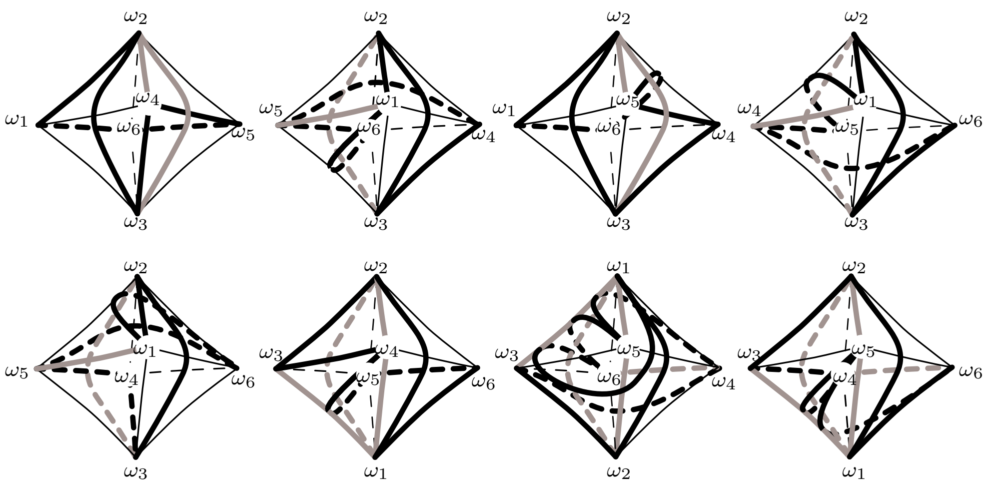


Figure 5: For $\alpha_2 = \beta^4_{2,3}, \beta^{4,5,6}_{2,3}, \beta^{5,4}_{2,3}, \beta^{6,4}_{2,3}$ and $\beta^{6,5,4}_{2,3}$ applications of (i) or (ii); and for $\alpha_2 = \beta^4_{1,3}, \beta^{4,5,6}_{1,3}$ and $\beta^{5,4}_{1,3}$ applications of Theorem 2.2, (ii)

$i \in \{2,3,4\}$. If $\{j,k\} = \{1,2\}$ we can apply (i) or (ii). There is no such α_1 for $\{j,k\} = \{5,6\}$.

Now suppose $n = 3$. If $\{j,k\} \neq \{1,2\}$ or $\{j,k\} \not\subset \{4,5,6\}$ then $l(\gamma_1) \leq l(\gamma_i), l(\gamma_i) \leq l(\alpha_1)$ are both preceding propositions for some $i \in \{2,3\}$. Again, if $\{j,k\} = \{1,2\}$ we can apply (i) or (ii). For $\{j,k\} \subset \{4,5,6\}$ either $j = 4$ (1) or $j = 5$ (2).

Now suppose $n = 2$. If $\{j,k\} \neq \{1,2\}$ or $\{j,k\} \not\subset \{3,\dots,6\}$ (ie $j \in \{1,2\}, k \in \{3,\dots,6\}$) then $l(\gamma_1) \leq l(\gamma_2), l(\gamma_2) \leq l(\alpha_1)$ are both preceding propositions. For $\{j,k\} = \{1,2\}$ (1). For $\{j,k\} \subset \{3,\dots,6\}$ either $j = 3$ (1); or $j \in \{4,5,6\}$ (2).

Finally $n = 1$. Either j or $k \in \{1,2\}$ (1); or $\{j,k\} \subset \{3,\dots,6\}$ (2). □

Proof of Theorem 2.4 As $l(\kappa_{3,0}) \leq l(\kappa_{0,5}), l(\kappa_{2,3}) \leq l(\kappa_{2,5}), l(\kappa_{0,1}) \leq l(\kappa_{0,4})$, by Corollary 2.3, we have that $l(\kappa_{1,2}) \geq l(\kappa_{2,4})$. Likewise, since $l(\kappa_{3,0}) \leq l(\kappa_{0,4}), l(\kappa_{2,3}) \leq l(\kappa_{2,4}), l(\kappa_{0,1}) \leq l(\kappa_{0,5})$ we have that $l(\kappa_{1,2}) \geq l(\kappa_{2,5})$. That is $l(\kappa_{1,2}) \geq l(\kappa_{2,l})$.

The arc set K divides $\mathcal{O}$ into eight triangles. We label these as follows: let t_k (respectively T_k) denote the triangle with one edge $\kappa_{k,k+1}$ and one vertex c_4 (respectively c_5). We shall use $\angle c_l t_k$ to denote the angle at the c_l–vertex of t_k, et cetera. Cut $\mathcal{O}$ open along $\kappa_{3,0} \cup \kappa_{0,1} \cup \kappa_{1,4} \cup \kappa_{1,2} \cup \kappa_{1,5}$ to obtain a domain Ω.

We show that $l(\kappa_{2,3}) \leq l(\kappa_{2,l}), l(\kappa_{3,0}) \leq l(\kappa_{1,2}) \leq \{l(\kappa_{0,l}), l(\kappa_{1,l})\}, l(\kappa_{0,1}) \leq l(\kappa_{0,l})$ implies that $\min_l l(\kappa_{3,l}) \leq l(\kappa_{0,1})$ with equality *if and only if* $\mathcal{O}$ is the octahedral orbifold. First we show that: $\angle c_2 t_2 \leq \angle c_4 t_0$ *or* $\angle c_2 T_2 \leq \angle c_5 T_0$.

Now $l(\kappa_{1,2}) \leq l(\kappa_{1,l})$, $l(\kappa_{3,0}) \leq l(\kappa_{0,l}$ so $\angle c_2 t_1 \geq \angle c_4 t_1$, $\angle c_2 T_1 \geq \angle c_5 T_1$, $\angle c_3 t_3 \geq \angle c_4 t_3$, $\angle c_3 T_3 \geq \angle c_5 T_3$, which imply

$$\angle c_2 t_1 + \angle c_2 T_1 + \angle c_3 t_3 + \angle c_3 T_3 \geq \angle c_4 t_1 + \angle c_5 T_1 + \angle c_4 t_3 + \angle c_5 T_3$$
$$\Leftrightarrow (\pi - \angle c_2 t_1 - \angle c_2 T_1) + (\pi - \angle c_3 t_3 - \angle c_3 T_3) \leq (\pi - \angle c_4 t_1 - \angle c_4 t_3) + (\pi - \angle c_5 T_1 - \angle c_5 T_3)$$
$$\Leftrightarrow (\angle c_2 t_2 + \angle c_2 T_2) + (\angle c_3 t_2 + \angle c_3 T_2) \leq (\angle c_4 t_2 + \angle c_4 t_0) + (\angle c_5 T_2 + \angle c_5 T_0)$$

and $l(\kappa_{2,3}) \leq l(\kappa_{2,l})$ so $\angle c_3 t_2 \geq \angle c_4 t_2, \angle c_3 T_2 \geq \angle c_5 T_2 \Rightarrow \angle c_2 t_2 + \angle c_2 T_2 \leq \angle c_4 t_0 + \angle c_5 T_0 \Rightarrow \angle c_2 t_2 \leq \angle c_4 t_0$ or $\angle c_2 T_2 \leq \angle c_5 T_0$.

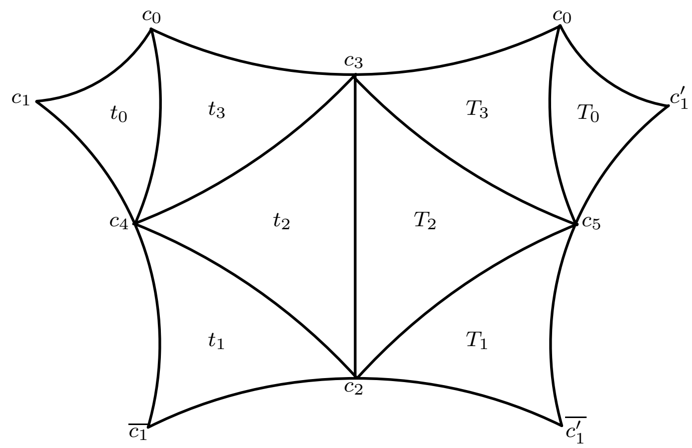


Figure 6: The triangles t_k, T_k in the domain Ω

Up to relabelling, we may suppose that $\angle c_2 t_2 \leq \angle c_4 t_0$. We now show that $l(\kappa_{3,4}) \leq l(\kappa_{0,1})$. There are two arguments. Firstly we show that if $\angle c_3 t_2 \geq \pi - \theta$ then $l(\kappa_{0,4}) < l(\kappa_{3,0})$ – contradicting a hypothesis. So $\angle c_3 t_2 < \pi - \theta$ and we then show that $l(\kappa_{3,4}) \leq l(\kappa_{0,1})$. The angle θ is given as follows. Let $\mathcal{I}_2$ be an isoceles triangle with vertices v_2, v_3, v_4 and edges $\varepsilon_{2,3}, \varepsilon_{2,4}, \varepsilon_{3,4}$ such that $l(\varepsilon_{2,3}) = l(\varepsilon_{2,4}) = l(\kappa_{2,4})$ and $\angle v_2 \mathcal{I}_2 = \angle c_2 t_2$. Then $\theta = \angle v_3 \mathcal{I}_2 = \angle v_4 \mathcal{I}_2$.

Let C_2, C_4 denote circles of radius $l(\kappa_{2,4})$ about c_2, c_4 respectively. As in Figure 7 c_3 must lie inside C_2 since $l(\kappa_{2,3}) \leq l(\kappa_{2,4})$. Likewise c_0 must lie outside C_4 since $l(\kappa_{0,4}) \geq l(\kappa_{1,2}) \geq l(\kappa_{2,4})$. Similarly c_1 must lie outside C_4 since $l(\kappa_{1,4}) \geq l(\kappa_{1,2}) \geq l(\kappa_{2,4})$. Moreover since the angle sum at any cone point is π: $\angle c_3 t_2 + \angle c_3 t_3 < \pi$. In Figure 6 we have also constructed the point x as the intersection of the radius through $\kappa_{2,3}$ and $\mathcal{C}_4$. Let t_x denote the triangle spanning x, c_3, c_4.

Now $\angle c_3 t_2 \geq \pi - \theta$ is equivalent to $\angle c_3 t_x \leq \theta$. It follows that $\angle c_4 t_x \geq \angle c_3 t_x$. By inspection $\angle c_4 t_3 > \angle c_4 t_x$ and $\angle c_3 t_x > \angle c_3 t_3$. So $\angle c_4 t_3 > \angle c_4 t_x \geq \angle c_3 t_x > \angle c_3 t_3$ or equivalently $l(\kappa_{0,4}) < l(\kappa_{0,3})$.

So $\angle c_3 t_2 < \pi - \theta$ and we will compare t_2, t_0. Firstly, $\angle c_3 t_2 < \pi - \theta$ implies that $l(\kappa_{3,4}) \leq l(\varepsilon_{3,4})$. (Recall that $\varepsilon_{3,4}$ is an edge of $\mathcal{I}_2$.) Let $\mathcal{I}_0$ be an isoceles triangle with vertices v_0, v_1, v_4 and edges $\varepsilon_{0,1}, \varepsilon_{1,4}, \varepsilon_{0,4}$ such that $l(\varepsilon_{1,4}) = l(\varepsilon_{0,4}) = l(\kappa_{2,4})$ and $\angle v_4 \mathcal{I}_0 = \angle c_4 t_0$. Since $l(\kappa_{0,4}), l(\kappa_{1,4}) \geq l(\kappa_{1,2}) \geq l(\kappa_{2,4})$ we then observe that $l(\kappa_{0,1}) \geq l(\varepsilon_{0,1})$. As $\angle c_2 t_2 \leq \angle c_4 t_0$ we have that $l(\varepsilon_{3,4}) \leq l(\varepsilon_{0,1})$. Therefore $l(\kappa_{0,1}) \geq l(\varepsilon_{0,1}) \geq l(\varepsilon_{3,4}) \geq l(\kappa_{3,4})$.

We have equality *if and only if* $\angle c_2 t_2 = \angle c_4 t_0$ and $l(\kappa_{2,3}) = l(\kappa_{2,4}) = l(\kappa_{0,4}) = l(\kappa_{1,4})$. From above $\angle c_2 t_2 = \angle c_4 t_0$ *if and only if* $l(\kappa_{1,2}) = l(\kappa_{1,l}), l(\kappa_{3,0}) = l(\kappa_{0,l})$ and $l(\kappa_{2,3}) = l(\kappa_{2,l})$. So we have that $l(\kappa_{0,1}) = l(\kappa_{3,4})$ and $l(\kappa_{1,2}) = l(\kappa_{2,3}) = l(\kappa_{3,0}) = l(\kappa_{0,l}) = l(\kappa_{1,l}) = l(\kappa_{2,l})$.

That is: t_1, T_1 are isometric equilateral triangles and t_0, T_0, t_2, t_3 (respectively T_2, T_3) are isometric isoceles triangles. By considering angle sums at c_4, c_5 : $\angle c_4 t_2 = \angle c_4 t_3 = \angle c_5 T_2 = \angle c_5 T_3$. So: t_1, T_1 are isometric equilateral triangles and $t_0, T_0, t_2, t_3, T_2, T_3$ are isometric isoceles triangles. By the angle sum at $c_3 : \angle c_3 t_2 = \angle c_3 t_3 = \angle c_3 T_2 = \angle c_3 T_3 = \pi/4$ and so $\angle c_0 t_0 = \angle c_1 t_0 = \angle c_0 T_0 = \angle c_1 T_0 = \pi/4$. Again, by considering angle sums at c_0, c_1 all the angles are $\pi/4$, all of the edges are of equal length. So $\mathcal{O}$ is the octahedral orbifold. □

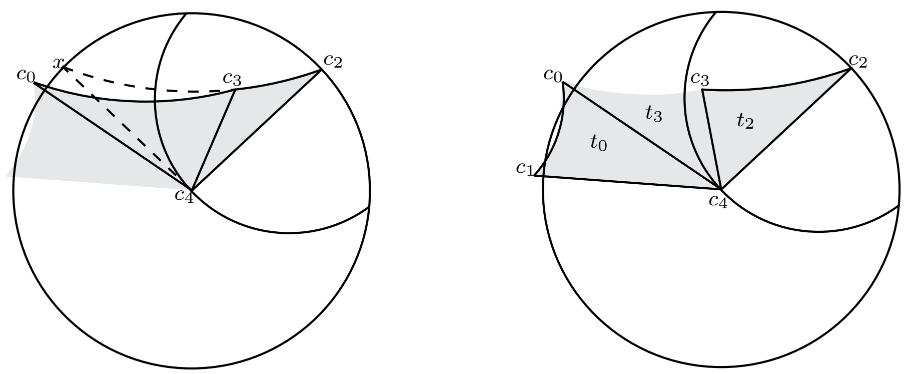


Figure 7: Arguments for $\angle c_3 t_2 \geq \pi - \theta$ and for $\angle c_3 t_2 < \pi - \theta$

References

[1] **P Buser**, *Geometry and Spectra of Compact Riemann Surfaces*, Birkhäuser (1992) 459–274

[2] **A Haas**, **P Susskind**, *The geometry of the hyperelliptic involution in genus two*, Proc. Amer. Math. Soc. 105 (1989) 159–165

[3] **L Keen**, *On fundamental domains and the Teichmüller modular group*, from: "Contributions to Analysis", Academic Press, New York and London (1974) 185–194

[4] **D Griffiths**, *Length inequalities for systems of geodesic loops on a surface of genus two: 1*, Bulletin of the LMS, 28 (1996) 505–508

[5] **D Griffiths**, *The side-pairing elements of Maskit's fundamental domain for the modular group in genus two*, Annal. Acad. Sci. Fenn. (to appear)

[6] **B Maskit**, *Parameters for Fuchsian groups I: Signature (0,4)*, from; "Holomorphic Functions and Moduli I", Math. Sci Res. Inst. Pub. 11, Springer–Verlag, New York (1988) 251–265

[7] **B Maskit**, *Parameters for Fuchsian groups II: topological type (1,1)*, Annal. Acad. Sci. Fenn. Ser. A.I. 14 (1990) 265–275

[8] **B Maskit**, *A Picture of Moduli Space*, Invent. Math. 126 (1996) 341–390

[9] **J McCarthy**, **A Papadopoulos**, *Fundamental Domains in Teichmüller Space*, Annal. Acad. Sci. Fenn. Ser. A.I. 21 (1996) 151–166

[10] **R C Penner**, *Weil–Petersson Volumes*, J. Diff. Geo. 35 (1992) 559–608

[11] **K-D Semmler**, *An explicit Fundamental Domain for the Teichmüller*, Doctoral Thesis, EPFL (1988)

Laboratoire de Mathematiques Pures de Bordeaux
Universite Bordeaux 1, 351 cours de la liberation
Talence 33405, Cedex, France

Email: `griffith@math.u-bordeaux.fr`

Received: 18 November 1997

ISSN 1464-8997 **181**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 181–248

Simplicité de groupes d'automorphismes d'espaces à courbure négative

Frédéric Haglund
Frédéric Paulin

Abstract We prove that numerous negatively curved simply connected locally compact polyhedral complexes, admitting a discrete cocompact group of automorphisms, have automorphism groups which are locally compact, uncountable, non linear and virtually **simple**. Examples include hyperbolic buildings, Cayley graphs of word hyperbolic Coxeter systems, and generalizations of cubical complexes, that we call *even* polyhedral complexes. We use tools introduced by Tits in the case of automorphism groups of trees, and Davis–Moussong's geometric realisation of Coxeter systems.

Résumé Nous montrons que de nombreux complexes polyédraux simplement connexes, localement compacts, à courbure négative, admettant un groupe discret cocompact d'automorphismes, ont leur groupe d'automorphismes localement compact, non dénombrable, non linéaire et virtuellement **simple**. Parmi les exemples, certains sont des immeubles hyperboliques, des graphes de Cayley de systèmes de Coxeter hyperboliques au sens de Gromov, et des généralisations de complexes cubiques, que nous appelons des complexes polyédraux *pairs*. Nous utilisons des outils dus à Tits dans le cas des groupes d'automorphismes d'arbres, et la réalisation géométrique de Davis–Moussong des systèmes de Coxeter.

AMS Classification 20E32, 51E24, 20F55; 20B27, 51M20

Keywords Simple group, polyhedral complex, even polyhedron, word hyperbolic group, hyperbolic building, Coxeter group

1 Introduction

J. Tits a démontré dans [31] que le groupe des automorphismes (sans inversion) d'un arbre (différent de la droite) homogène ou semi-homogène localement fini, est localement compact, non dénombrable et simple. Le but de cet article est de démontrer la simplicité de groupes d'automorphismes de nombreux complexes polyédraux localement finis, ayant des propriétés de courbure négative, comme par exemple des immeubles hyperboliques ou des complexes cubiques.

Un *immeuble hyperbolique* (voir [19]) est un immeuble de type un système de Coxeter $(W(P), S(P))$ de la forme suivante. Soit P un polyèdre (compact convexe, pas forcément un simplexe) de l'espace hyperbolique réel $\mathbb{H}^n$ de dimension n, avec P *de Coxeter* (i.e. ses angles dièdres sont de la forme $\frac{\pi}{k}$ avec k un entier au moins 2). Alors $S(P)$ est l'ensemble des réflexions (orthogonales) sur les faces de codimension 1 de P, et $W(P)$ le groupe d'isométries de $\mathbb{H}^n$ engendré par $S(P)$.

Un premier exemple est l'immeuble de Bourdon $I_{p,q}$ avec $p \geq 5, q \geq 3$, qui est l'unique complexe polyédral de dimension 2, dont les polygones sont des copies du p–gone hyperbolique régulier à angles droits P_p, et le link de chaque sommet est isomorphe au graphe biparti complet à $q+q$ sommets (voir [7]). Il existe une numérotation des arêtes de $I_{p,q}$ (unique une fois numérotées les arêtes d'un polygone fixé) par $I = \{1, \cdots p\}$ de sorte que le long du bord de chaque polygone les arêtes apparaissent avec l'ordre cyclique ou l'ordre inverse. L'ensemble des polygones de $I_{p,q}$ est alors un système de chambres sur I, deux chambres étant i–adjacentes si et seulement si les polygones correspondants se rencontrent le long d'une arête numérotée i. Il est facile (voir [19]) de montrer que $I_{p,q}$ est un immeuble de type $(W(P_p), S(P_p))$.

Théorème 1.1 *Le groupe des automorphismes préservant le type de l'immeuble de Bourdon $I_{p,q}$ est un groupe localement compact, non dénombrable, non linéaire au moins si p est multiple de 4, et simple.*

Dans [23] sont construits de nombreux autres exemples. Soit L un m–gone généralisé fini épais classique (i.e. un graphe biparti complet à $p+q$ sommets avec $p, q \geq 3$ si $m = 2$, ou si $m \geq 3$, l'immeuble sphérique de rang 2 d'un groupe de Chevalley fini $\underline{G}(\mathbb{F}_q)$, avec $\underline{G}$ un groupe algébrique simple, de groupe de Weyl le groupe diédral D_{2m} d'ordre $2m$). Par exemple, L peut être l'immeuble des drapeaux du plan projectif sur le corps fini $\mathbb{F}_q$, avec $m = 3$. Soit k un entier pair au moins 6. Alors dans [23] est construit un 2–complexe polyédral $A_{k,L}$, dont les polygones sont des copies du k–gone hyperbolique $P_{k,m}$ régulier à angles $\frac{\pi}{m}$, et le link de chaque sommet est isomorphe au graphe biparti L. L'ensemble de ses polygones possède aussi une structure naturelle d'immeuble de type $(W(P_{k,m}), S(P_{k,m}))$ (voir [19]).

Théorème 1.2 *Le groupe des automorphismes de l'immeuble hyperbolique $A_{k,L}$ est un groupe localement compact, non dénombrable, non linéaire au moins si k est multiple de 4, et virtuellement simple.*

En fait, $A_{k,L}$ est la réalisation géométrique au sens de Davis–Moussong (voir [25]) du système de Coxeter $W(k,L)$, dont la matrice de Coxeter est la matrice d'adjacence du graphe L, où les 1 et 0 ont été remplacés par des $\frac{k}{2}$ et ∞ respectivement. Le 1–squelette de la réalisation géométrique de Davis–Moussong d'un système de Coxeter (W,S) s'identifie au graphe de Cayley de (W,S), et nous montrons que tout automorphisme du 1–squelette s'étend à cette réalisation géométrique (voir section 5.1). Appelons *mur* du graphe de Cayley l'ensemble des points fixes d'un conjugué d'un élément de S. Un mur est *propre* si aucune des deux composantes du complémentaire du mur ne reste à distance bornée du mur. Un automorphisme du graphe de Cayley fixe *strictement* un mur s'il fixe le mur et n'échange pas les deux composantes de son complémentaire.

Un système de Coxeter est dit *rigide* s'il n'existe pas d'automorphisme non trivial de son diagramme qui fixe les arêtes de poids fini issues d'un de ses sommets. Généralisant [23], nous montrons que le groupe des automorphismes (de graphe) du graphe de Cayley de (W,S) est non dénombrable si et seulement si (W,S) est non rigide, et nous le calculons exactement dans le cas rigide (voir théorème 5.12).

Théorème 1.3 *Si (W,S) est un système de Coxeter, avec W ne contenant pas de sous-groupe isomorphe à $\mathbb{Z}+\mathbb{Z}$, alors le quotient, par son sous-groupe distingué des éléments fixant l'infini, du sous-groupe G^+ des automorphismes du graphe de Cayley de (W,S) engendré par les fixateurs stricts de murs propres, est simple. Il est non trivial, donc non dénombrable, si et seulement si (W,S) n'est pas rigide.*

Un *complexe cubique* de dimension n est un complexe polyédral P, dont les polyèdres sont des cubes euclidiens $[-\frac{1}{2},\frac{1}{2}]^k$, tout cube de P étant contenu dans un cube de dimension (maximale) n. Il est dit CAT(0) s'il est simplement connexe, et si pour tout cube c de P, le link $lk(c)$ de c vérifie la condition suivante: tout cycle d'arêtes dans $lk(c)$ est de longueur au moins 3, et si de longueur 3, borde un simplexe de $lk(c)$. Pour toute arête a de P, il existe un unique sous-complexe (de la subdivision barycentrique) de P, appelé *mur* ("geometric hyperplane" par M. Sageev [28]), rencontrant a en son milieu, et dont toute intersection non triviale avec un cube de dimension n de P est un hyperplan $[-\frac{1}{2},\frac{1}{2}]^k\times\{0\}\times[-\frac{1}{2},\frac{1}{2}]^{n-k-1}$ de ce cube. Par exemple, si $n=1$, alors P est un arbre, et un mur est le milieu d'une arête.

Nous introduisons une notion de *polyèdre pair* (section 4.1) et donc de *complexe polyédral pair* (i.e. dont tous les polyèdres sont pairs), généralisant strictement celle de cube et complexe cubique, avec ses murs. Un polyèdre d'un espace à courbure constante est pair s'il est symétrique par rapport à l'hyperplan médiateur de chacune de ses arêtes, et si un tel hyperplan ne passe pas par un de ses sommets. Nous donnons en section 4.1 la construction explicite de tous les polyèdres pairs euclidiens ou hyperboliques, à partir des systèmes de Coxeter finis, ainsi que la liste complète des polyèdres hyperboliques pairs de dimension 2 et 3 qui sont eux-mêmes des polyèdres de Coxeter. M. Davis nous a signalé que nos polyèdres pairs sont, du point de vue combinatoire, exactement les *zonotopes de Coxeter* (aussi appelés "Coxeter cell" dans [16]), i.e. les polyèdres duaux de l'arrangement d'hyperplans formé par les hyperplans fixes des conjugués des réflexions d'un système de Coxeter fini. Nos complexes polyédraux pairs sont donc, du point de vue combinatoire, des cas particuliers de "zonotopal cell complex" au sens de [17]. Notons qu'il existe des polyèdres pairs non isométriques ayant même combinatoire.

Théorème 1.4 *Soit P un complexe polyédral pair (par exemple cubique), localement fini,* $\mathrm{CAT}(0)$*, admettant un groupe discret cocompact d'automorphismes qui est hyperbolique au sens de Gromov. Alors le groupe d'automorphismes G^+ de P engendré par les fixateurs stricts de murs propres est presque simple (au sens que tout éventuel sous-groupe distingué propre est relativement compact). Si P est* $\mathrm{CAT}(-1)$ *et tout point de P appartient à une droite géodésique, alors G^+ est simple, et non dénombrable si non trivial.*

Bien sûr, G^+ peut être trivial. Pour tout type de polyèdre euclidien pair possible, nous construisons (section 5.4) un complexe polyédral pair $\mathrm{CAT}(-1)$, dont les cellules maximales sont de ce type, et dont le groupe G^+ est non dénombrable. Un arbre homogène ou semi-homogène localement fini admet un groupe discret cocompact d'automorphismes qui est libre, donc hyperbolique au sens de Gromov (voir section 2 pour des rappels sur cette notion.) Nous retrouvons ainsi le résultat de J.Tits. La condition de locale finitude n'est pas vraiment nécessaire (voir section 7). La condition d'hyperbolicité n'est sans doute pas optimale. Mais comme le montre le cas du produit de deux arbres homogènes, il faut une hypothèse d'irréductibilité sur P. Nous renvoyons à [11] pour un critère ingénieux de simplicité sur les groupes discrets d'automorphismes du produit de deux arbres.

Une généralisation immédiate du théorème B de Niblo–Reeves [26] est la suivante.

Théorème 1.5 *Soit P un complexe polyédral pair* $\mathrm{CAT}(0)$ *de dimension finie. Toute action polyédrale sur P d'un groupe ayant la propriété* (T) *de Kazhdan a un point fixe global.*

Pour généraliser la situation des exemples ci-dessus, nous introduisons (section 3) une notion abstraite d'ensemble discret X muni d'un système de *murs*, modélisant les propriétés de l'ensemble des sommets d'un complexe polyédral cubique (ou pair) CAT(0) et de la famille de ses hyperplans médiateurs des arêtes, ou d'un groupe de Coxeter W muni de sa famille de murs (voir [27, page 14]). Dans les sections 4.2 à 4.4, nous étudions l'espace à murs canoniquement associé à un complexe polyédral pair.

Sous des hypothèses d'hyperbolicité au sens de Gromov (voir section 2.2 pour les propriétés que nous utiliserons) du graphe d'incidence de cette famille de murs, nous montrons (section 6) un théorème de simplicité sur des groupes de bijections de X préservant le système de murs, vérifiant une condition (P) analogue à celle introduite par J. Tits [31] dans le cas des arbres. Le lemme clef 6.4 sur les commutateurs est analogue au lemme 4.3 de [31]. Enfin, en section 7, nous appliquons ce théorème de simplicité à nos exemples.

Nous remercions F. Choucroun, pour son exposé sur l'article de J. Tits, qui a servi de point de départ à ce travail, ainsi que S. Mozes et M. Davis.

2 Rappels sur les espaces métriques hyperboliques

Nous renvoyons à [21, 20] pour les définitions, références, historiques et preuves des propriétés rappelées ci-dessous des espaces métriques hyperboliques au sens de Gromov, à [9, 6] pour celles des espaces métriques CAT(χ) au sens d'Alexandroff–Topogonov et à [8] pour celles des complexes polyédraux. Le lecteur connaisseur peut se ramener directement à la proposition 2.1.

2.1 Définitions diverses

Une *géodésique* d'un espace métrique X est une isométrie d'un intervalle I de $\mathbb{R}$ dans X. On parle de *segment, rayon* ou *droite géodésique* si I est de la forme $[a, b], [a, +\infty[$ ou $\mathbb{R}$. Un espace métrique est *géodésique* si par deux de ses points passe un segment géodésique.

Un espace géodésique est *hyperbolique* (au sens de Gromov) s'il existe une constante $\delta \geq 0$ (dite *constante d'hyperbolicité*) telle que tout point de tout côté de tout triangle géodésique est à distance au plus δ d'un point de l'un des deux autres côtés. Un groupe de type fini G, muni d'une partie génératrice S, est *hyperbolique* (au sens de Gromov) si le graphe de Cayley de G pour S, muni de sa métrique naturelle, est hyperbolique. Une application $f\colon X \to Y$ entre deux espaces métriques est une *quasi-isométrie* s'il existe des constantes $\lambda \geq 1, c, c' \geq 0$ telles que pour tous x, y dans X et z dans Y :

$$\frac{1}{\lambda}d(x,y) - c \leq d(f(x), f(y)) \leq \lambda d(x,y) + c \quad \text{et} \quad d(z, f(X)) \leq c'.$$

Un espace géodésique quasi-isométrique à un espace hyperbolique est encore hyperbolique, donc l'hyperbolicité d'un groupe ne dépend pas de la partie génératrice fixée.

Deux rayons géodésiques sont *asymptotes* si leur distance de Hausdorff est finie. Ceci définit une relation d'équivalence sur l'ensemble des rayons géodésiques dans X. L'ensemble des classes d'équivalence est appelé le *bord* (ou espace à l'infini) de X, et noté ∂X. Il existe une topologie naturelle sur $\overline{X} = X \cup \partial X$, métrisable compacte lorsque X est hyperbolique, localement compact, complet. Toute quasi-isométrie entre deux espaces hyperboliques s'étend continûment en un homéomorphisme de ∂X sur ∂Y.

Soit X un espace géodésique et $\chi \in \mathbb{R}$. Soit $\mathbb{X}^2_\chi$ le plan riemannien complet simplement connexe à courbure constante χ ($\mathbb{X}^2_\chi$ est le plan hyperbolique, le plan euclidien, la sphère de dimension 2 si $\chi = -1, 0, 1$). Soit $\triangle = [xy] \cup [yz] \cup [zx]$ un triangle géodésique dans X. Soit $\overline{\triangle} = [\overline{x}\,\overline{y}] \cup [\overline{y}\,\overline{z}] \cup [\overline{z}\,\overline{x}]$ un triangle géodésique dans $\mathbb{X}^2_\chi$ ayant mêmes longueurs des côtés que $\triangle$. Si $s \in \triangle$, le point sur le côté correspondant de $\overline{\Delta}$, à la même distance des extrémités que s, est noté $\overline{s}$. Un triangle géodésique $\triangle$ dans X est CAT(χ) s'il est plus “pincé” que le triangle correspondant de l'espace modèle, i.e. si, pour tous points $s, t \in \triangle$, on a

$$d_X(s,t) \leq d_{\mathbb{X}^2_\chi}(\overline{s}, \overline{t}).$$

Un espace géodésique est CAT(χ) si tout triangle géodésique de X est CAT(χ). Si $\chi < 0$, un espace CAT(χ) est hyperbolique au sens de Gromov.

Un *complexe polyédral* P est un complexe cellulaire (voir par exemple [30]) dont les cellules sont des polyèdres (compacts convexes) d'un espace à courbure constante, et dont les applications d'attachements sont cellulaires et localement isométriques sur chaque cellule ouverte. Un *complexe polygonal* est un complexe polyédral de dimension 2. Un complexe polyédral, dont les polyèdres sont des simplexes ne se rencontrant qu'au plus en une face, est précisément (la réalisation géométrique d') un complexe simplicial.

Un *automorphisme* de complexe polyédral de P est un automorphisme du complexe cellulaire P. Nous identifierons deux automorphismes qui envoient chaque cellule ouverte sur une même cellule ouverte. Un automorphisme est dit *isométrique* (ou une *isométrie polyédrale*) si sa restriction à chaque polyèdre est isométrique. Par exemple, si P est un rectangle euclidien non carré, alors P admet 4 isométries polyédrales, et 8 automorphismes. Si P est muni de la topologie faible usuelle, le groupe des automorphismes de P sera muni de la topologie compacte–ouverte. Si P est localement fini, alors Aut G est localement compact, et le fixateur de tout polyèdre de P est un groupe compact profini.

Si P n'a qu'un nombre fini de classe d'isométrie de polyèdres, alors (voir [8]) il existe une métrique d (naturelle pour les automorphismes de P) géodésique et complète, ainsi définie. Une *géodésique brisée* γ de P est une courbe qui, par morceaux, est contenue et géodésique dans un polyèdre de P. Sa longueur $\ell(\gamma)$ est la somme des longueurs des morceaux géodésiques précédents. Alors $d(x,y)$ est la borne inférieure des longueurs des géodésiques brisées entre x et y.

Sauf mention explicite du contraire, tout complexe polyédral sera muni de cette distance. Toute isométrie polyédrale est une isométrie pour cette distance. La topologie faible et la topologie induite par cette distance coïncident si et seulement si P est localement fini. Voir [21] pour l'équivalence, dans le cas des complexes cubiques, entre la définition ci-dessus de CAT(0) et celle donnée en introduction.

Si C est un complexe polyédral n'ayant qu'un nombre fini de types d'isométrie de cellules, et $x \in C$, nous noterons $lk(x,C)$ l'espace des germes de segments géodésiques issus de x. Il possède une structure naturelle de complexe polyédral, dont les cellules sont des polyèdres sphériques.

Un *graphe* est un 1–complexe simplicial connexe. En identifiant chaque arête à $[-\frac{1}{2},\frac{1}{2}]$, on obtient un complexe polyédral. Sa métrique est l'unique métrique géodésique rendant chaque arête isométrique à $[0,1]$. Un *arbre* est un graphe simplement connexe. Un arbre est CAT$(-\infty)$, i.e. CAT(χ) pour tout $\chi \in \mathbb{R}$.

2.2 Groupes d'isométries non élémentaires

Soit Y un espace métrique complet, géodésique et hyperbolique, tel que par deux points de $Y \cup \partial Y$ passe un segment, rayon ou droite géodésique (cette dernière condition est toujours remplie si Y est localement compact). On note $\partial^2 Y$ l'espace des couples de points distincts de ∂Y. On note $\overline{Z}$ l'adhérence dans $Y \cup \partial Y$ d'une partie Z de Y, et $\partial Z = \overline{Z} \cap \partial Y$.

Une isométrie g de Y est dite *hyperbolique* si pour un (donc pour tout) point x dans Y, l'application de $\mathbb{Z}$ dans Y qui à k associe $g^k x$ est une quasi-isométrie sur son image. En particulier, g admet alors exactement deux points fixes dans ∂Y.

Soit G un sous-groupe du groupe des isométries de Y (n'agissant peut-être pas proprement discontinument). Définissons l'*ensemble limite* ΛG de G comme l'adhérence dans ∂Y de l'ensemble des points fixes dans ∂Y des éléments hyperboliques de G. Le groupe G est dit *non élémentaire* si son ensemble limite contient au moins trois points et ne contient pas de point fixe global (cette dernière condition est toujours remplie si Y est localement compact et G discret). Si G est non élémentaire, ΛG est non dénombrable et sans point isolé; c'est l'ensemble d'accumulation dans ∂Y de l'orbite par G de tout point de Y; c'est le plus petit fermé non vide invariant par G dans ∂X; l'orbite par G de tout point de ΛG est dense dans ΛG. On note $\Lambda^2 G$ l'ensemble des couples de points distincts de ΛG.

Remarque Par exemple, si Y est localement compact, si G contient un sous-groupe agissant proprement discontinûment avec quotient compact sur Y, alors G est non élémentaire et $\Lambda G = \partial Y$.

Proposition 2.1 *Si G est non élémentaire, alors l'ensemble des couples des point fixes des éléments hyperboliques de G est dense dans $\Lambda^2 G$.*

Soit H un sous-groupe distingué non trivial de G. Si G est non élémentaire, alors ou bien H est contenu dans le noyau de l'action de G sur ΛG, ou bien H est non élémentaire, d'ensemble limite égal à celui de G.

Preuve La première assertion est due à [21, Corollaire 8.2.G].

Pour la seconde assertion, supposons que $h \in H$ n'agisse pas trivialement sur l'ensemble limite de G. Montrons tout d'abord que H contient au moins un élément hyperbolique.

Soit $a \in \Lambda G$ tel que $ha \neq a$. Par invariance, ha est dans ΛG. Soit δ une constante d'hyperbolicité de X. Soit U un voisinage ouvert suffisamment petit de a dans $Y \cup \partial Y$, de sorte que U et hU soient disjoints, et séparés d'une distance grande devant δ. Soit g un élément hyperbolique de G, dont les points fixes répulsif $g^- \in \Lambda G$ et attractifs $g^+ \in \Lambda G$ sont dans U et hU respectivement. Soit γ une géodésique entre g^- et g^+. Soit y un point de $\gamma \cap U$. En particulier, hy appartient à hU.

Si n est assez grand, alors $g^n y$ est proche de g^+, donc appartient à hU. Donc $h^{-1} g^n y$ appartient à U, et si n est assez grand, $g^{-n} h^{-1} g^n y$ est beaucoup plus

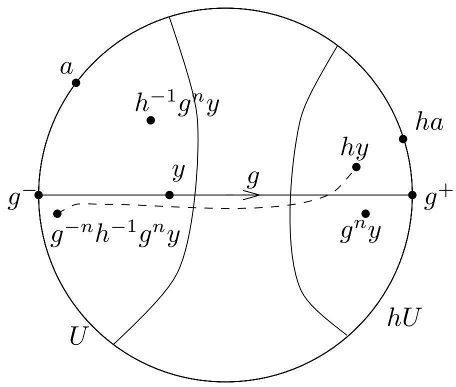


Figure 1: Construction d'un élément hyperbolique dans H

proche de g^- que y. Donc il existe une constante K (ne dépendant que de δ) telle que y est à distance au plus K d'un segment géodésique entre $g^{-n}h^{-1}g^n y$ et hy. Quitte à avoir pris U suffisamment petit, on a

$$\inf\{d(y, hy), d(y, g^{-n}h^{-1}g^n y)\} > 2K + 1000\delta.$$

Par [21, Lemma 8.1.A], on en déduit que $h(g^{-n}h^{-1}g^n)^{-1}$ est hyperbolique. Comme H est distingué, ceci montre notre affirmation préliminaire.

Maintenant, comme les conjugués d'un élément hyperbolique h de H sont encore dans H, que l'orbite par G d'un point fixe de h est contenue et dense dans ΛG, on en déduit que $\Lambda H = \Lambda G$. En particulier ΛH contient au moins trois point. Si H fixait un point a de ΛH, celui-ci serait unique [21, 8.2.D]. Comme H est distingué dans G, le point a serait fixe par G, ce qui est impossible. □

Lemme 2.2 *Supposons ∂Y non vide sans point isolé. Si Y est localement compact, le noyau de l'action de G sur le bord de Y est relativement compact dans le groupe des isométries de Y (donc compact si G est fermé dans le groupe des isométries de Y). Si Y est* CAT(-1) *et tout point de Y appartient à une droite géodésique, alors G agit fidèlement sur le bord.*

Preuve Pour la première assertion, soient x, y, z trois points distincts de ∂Y et p une quasi-projection de x sur une géodésique entre y et z. Une isométrie de Y qui fixe (point par point) le bord de Y bouge p d'une distance inférieure à une constante. Le résultat découle alors du théorème d'Ascoli.

Pour la seconde assertion, soit $g \in G$ fixant le bord de Y. Soit $x \in Y$ et $a, b \in \partial Y$ les extrémités d'une droite géodésique D passant par x. Soient

a', b' deux points proches et distincts de a, b respectivement. Soit p, p' l'unique projection de a', b' sur D. Alors par unicité, p et p' sont fixes par l'isométrie g, et $x \in [p, p']$ aussi, par unicité du segment géodésique entre deux points. □

3 Espaces à murs

Soit X un ensemble. Un *mur* de X est une partition de X en deux sous-ensembles, appelés les *demi-espaces* définis par le mur. Un mur *sépare* deux points x et y de X si et seulement si x appartient à l'un des demi-espaces définis par le mur et y appartient à l'autre. Un *système de murs sur* X est un ensemble $\mathcal{M}$ de murs de X tel que:

(M) Pour tous x et y distincts dans X, l'ensemble $\mathcal{M}(x, y)$ des murs de $\mathcal{M}$ séparant x et y est fini non vide.

Un *espace à murs* est un couple $(X, \mathcal{M})$, où X est un ensemble et $\mathcal{M}$ un système de murs sur X. Tout singleton de X est alors l'intersection des demi-espaces qui le contiennent.

Dans un espace à murs $(X, \mathcal{M})$, on dit qu'un point z est *entre* deux points x et y si $\mathcal{M}(x, y)$ est la réunion (nécessairement disjointe) de $\mathcal{M}(x, z)$ et $\mathcal{M}(z, y)$. Le *graphe associé* à $(X, \mathcal{M})$ est le graphe ayant X pour ensemble de sommets, et une arête entre deux sommets x et y si et seulement si les seuls points de X entre x et y sont x et y. On note $\mathcal{G} = \mathcal{G}(X, \mathcal{M})$ ce graphe, qui est connexe d'après l'axiome (M). Un mur M de X est dit *transverse* à une arête de $\mathcal{G}(X, \mathcal{M})$ lorsqu'il sépare ses extrémités.

Un espace à murs $(X, \mathcal{M})$ est dit *hyperbolique* si son graphe associé est un espace métrique hyperbolique au sens de Gromov, et s'il vérifie la condition (H) suivante de non trivialité et de compatibilité entre la structure métrique de $\mathcal{G}$ et le système de demi-espaces défini par $\mathcal{M}$:

(H) Pour tout $\xi \in \partial\mathcal{G}$, l'ensemble des parties de $\mathcal{G} \cup \partial\mathcal{G}$ de la forme $\overline{A}$, où A est un demi-espace de $(X, \mathcal{M})$ tel que $\overline{A}$ contient ξ dans son intérieur, est une base de voisinages de ξ dans $\mathcal{G} \cup \partial\mathcal{G}$.

3.1 Automorphismes d'espaces à murs et propriété (P) de Tits

Soit $(X, \mathcal{M})$ un espace à murs. Un *automorphisme* ϕ de $(X, \mathcal{M})$ est une bijection de X préservant $\mathcal{M}$. Il induit un automorphisme du graphe $\mathcal{G}$, encore noté ϕ. Si $(X,\mathcal{M})$ est hyperbolique, alors ϕ induit un homéomorphisme du bord hyperbolique $\partial\mathcal{G}$ de $\mathcal{G}$, toujours noté ϕ.

Si $\mathrm{Aut}(X,\mathcal{M})$ est le groupe des automorphismes de (X,M), et $\mathrm{Aut}(\mathcal{G})$ le groupe des automorphismes de graphe de $\mathcal{G}$, alors l'application $\phi \mapsto \phi$ est une injection de $\mathrm{Aut}(X,\mathcal{M})$ dans $\mathrm{Aut}(\mathcal{G})$, en général non surjective (voir toutefois la preuve du théorème 5.1). Nous identifierons $\mathrm{Aut}(X,\mathcal{M})$ avec son image dans $\mathrm{Aut}(\mathcal{G})$. Lorsque $\mathcal{G}$ est localement fini, nous munirons $\mathrm{Aut}(\mathcal{G})$ de la topologie compacte–ouverte et $\mathrm{Aut}(X,\mathcal{M})$ de la topologie induite.

Un automorphisme *fixe strictement* un mur M s'il fixe les sommets de toute arête transverse à M. Un automorphisme d'un espace à murs *fixe strictement* un demi-espace A s'il fixe A et fixe strictement le mur $M=\{\mathcal{A},X\setminus A\}$.

Lemme 3.1 *Un automorphisme fixant strictement un mur M préserve chacun des demi-espaces de X définis par M.*

Preuve Remarquons d'abord que si M sépare deux points x,y, alors tout chemin entre x et y dans $\mathcal{G}$ contient une arête de $\mathcal{G}$ transverse à M.

Notons $M=\{A,X\setminus A\}$ et $V(M)$ l'ensemble des sommets d'arêtes de $\mathcal{G}$ transverses à M. Si x appartient au demi-espace A, soit p un point de $V(M)$ à distance minimale de x. Par minimalité, p est dans A. Si ϕ fixe strictement M, alors il fixe point par point $V(M)$. Il envoie un chemin γ de longueur minimale entre x et p sur un chemin de même longueur entre $\phi(p)=p$ et $\phi(x)$. Si $\phi(x)$ n'est pas dans A, alors le chemin $\phi(\gamma)$ doit contenir une arête transverse à M, ce qui contredit le fait que ϕ préserve la distance combinatoire à $V(M)$. □

On appelle *chaîne* une suite $(A_i)_{i\in\mathbb{Z}}$ de demi-espaces qui est strictement décroissante pour l'inclusion. Un automorphisme *fixe strictement* cette chaîne s'il fixe strictement chaque mur $M_i=\{A_i,X\setminus A_i\}$. Par le lemme précédent, il préserve alors chaque demi-espace A_i.

Soit G un groupe d'automorphismes de $(X,\mathcal{M})$. Si $M=\{A,X\setminus A\}$ est un mur de $\mathcal{M}$, soit G_M le sous-groupe de G fixant strictement M. Par le lemme précédent, le groupe G_M préserve les ensembles $X\setminus A$ et A. Nous notons G_A (resp. $G_{X\setminus A}$) le groupe des permutations de A (resp. $X\setminus A$) induit par G_M. Le produit des restrictions donne un morphisme injectif

$$G_M \to G_A \times G_{X\setminus A}.$$

Soit $C=(A_i)_{i\in\mathbb{Z}}$ une chaîne. Soit G_C le sous-groupe de G fixant strictement C. Pour tout i, le groupe G_C préserve l'ensemble $A_i\setminus A_{i+1}$, et nous notons

$G_{C,i}$ le groupe des permutations de cet ensemble induit par G_C. Le produit direct des restrictions $G_C \to G_{C,i}$ est un morphisme

$$G_C \to \prod_{i \in \mathbb{Z}} G_{C,i}.$$

Lemme 3.2 *Ce morphisme est injectif.*

Preuve Il suffit de montrer que pour toute chaîne $C = (A_i)_{i\in\mathbb{Z}}$ de $(X, \mathcal{M})$, la réunion des $A_i \setminus A_{i+1}$ vaut tout X. Supposons par l'absurde qu'il existe un point x n'appartenant pas à cette réunion. Supposons que x appartient à A_0 (si $x \in X \setminus A_0$, le raisonnement est le même, quitte à renverser l'ordre de $\mathbb{Z}$). Soit $x_0 \in A_0 \setminus A_1$. Alors x_0 appartient à $X \setminus A_i$ et x appartient à A_i pour tout $i \geq 1$. Donc le mur $M_i = \{X \setminus A_i, A_i\}$ sépare x_0 et x pour tout $i \geq 1$, ce qui contredit la finitude de $\mathcal{M}(x, x_0)$. □

La définition suivante est alors analogue à la propriété homonyme de [31].

Définition 3.3 On dit qu'un groupe G d'automorphismes de $(X, \mathcal{M})$ *vérifie la propriété* (P) si pour tout mur M et toute chaîne C, les morphismes précédents sont surjectifs, i.e. des isomorphismes.

Lemme 3.4 *Soit G un groupe d'automorphismes d'un espace à murs, ayant la propriété* (P)*. Alors le sous-groupe de G engendré par les fixateurs stricts de murs coïncide avec le sous-groupe de G engendré par les fixateurs stricts de demi-espaces.*

Preuve Le second groupe est contenu dans le premier, par définition. Il suffit donc de montrer que tout élément g de G fixant strictement un mur $M = \{A^-, A^+\}$ est produit de deux éléments g^-, g^+ fixant strictement les demi-espaces A^-, A^+ respectivement. Par la propriété (P), le morphisme $G_M \to G_{A^-} \times G_{A^+}$ est surjectif. Il suffit de prendre pour g^-, g^+ des préimages de $(g|_{A^-}, id)$ et $(id, g|_{A^+})$ respectivement. □

Considérons la propriété suivante d'un espace à murs $(X, \mathcal{M})$.

> (M′) Pour tous demi-espaces A, B de $(X, \mathcal{M})$, avec B rencontrant A et son complémentaire, tout automorphisme fixant strictement le mur $M = \{A, X \setminus A\}$ préserve B.

Dans le cas d'un arbre, cette condition est vide (donc n'apparaît pas dans [31]).

Lemme 3.5 *Si un espace à murs* $(X,\mathcal{M})$ *vérifie la condition* (M′), *alors le groupe de tous ses automorphismes vérifie la propriété* (P).

Preuve Soit $M=\{A^-,A^+\}$ un mur de $(X,\mathcal{M})$. Soit $h^\pm$ la restriction à $A^\pm$ d'un automorphisme $\overline{h}^\pm$ de $(X,\mathcal{M})$ fixant strictement M. Comme $A^-\cup A^+=X$, soit g la bijection de X valant $h^\pm$ sur $A^\pm$. Montrons que g préserve $\mathcal{M}$, ce qui impliquera la surjectivité de $\mathrm{Aut}(X,\mathcal{M})_M\to\mathrm{Aut}(X,\mathcal{M})_{A^-}\times\mathrm{Aut}(X,\mathcal{M})_{A^+}$. Soit $N=\{B,X\setminus B\}$ un mur de $(X,\mathcal{M})$. Si B est contenu dans $A^\pm$, alors $g(B)=h^\pm(B)\subset A^\pm$, donc $g(B)=\overline{h}^\pm(B)$ est un demi-espace de $(X,\mathcal{M})$. D'où $g(N)$ est encore un mur de $(X,\mathcal{M})$. Si B rencontre à la fois A^- et A^+, alors les deux automorphismes $\overline{h}^-$ et $\overline{h}^+$ préservent B par la propriété (M′). Donc $\overline{h}^\pm$ préserve $B\cap A^\pm$. D'où g préserve B, et $g(N)=N$ est encore un mur de $(X,\mathcal{M})$.

Soit $C=(A_i)_{i\in\mathbb{Z}}$ une chaîne de $(X,\mathcal{M})$, et soit h_i la restriction à $A_i\setminus A_{i+1}$ d'un automorphisme $\overline{h}_i$ de $(X,\mathcal{M})$ fixant strictement M_i. Comme $X=\bigcup_{i\in\mathbb{Z}}A_i\setminus A_{i+1}$ (voir la preuve du lemme 3.2), il existe une bijection g de X valant h_i sur $A_i\setminus A_{i+1}$. Soit B un demi-espace de $(X,\mathcal{M})$. On montre comme précédemment que si B est contenu dans un $A_i\setminus A_{i+1}$, alors $g(B)$ est encore un demi-espace, et que, par la propriété (M′), si B rencontre au moins deux $A_i\setminus A_{i+1}$, alors $g(B)=B$. Donc g est un automorphisme de $(X,\mathcal{M})$. Ceci montre la surjectivité de $\mathrm{Aut}\,(X,\mathcal{M})_C\to\prod_{i\in\mathbb{Z}}\mathrm{Aut}\,(X,\mathcal{M})_{C,i}$. □

Un mur d'un espace à murs hyperbolique est dit *propre* si le bord à l'infini dans $\overline{\mathcal{G}}$ de chacun des demi-espaces qu'il définit n'est pas égal à tout $\partial\mathcal{G}$. Une chaîne $C=(A_i)_{i\in\mathbb{Z}}$ est *propre* si chaque mur $M_i=\{A_i,X\setminus A_i\}$ est propre. Dans la condition (H), nous pouvons de plus supposer que les murs définissant les demi-espaces A sont propres. Si G est un groupe d'automorphismes de $(X,\mathcal{M})$, nous noterons G^+ le sous-groupe de G engendré par les fixateurs stricts de murs propres.

Lemme 3.6 *Soit* $(X,\mathcal{M})$ *un espace à murs hyperbolique, de graphe associé* $\mathcal{G}$ *localement fini, et* G *un groupe d'automorphismes de* $(X,\mathcal{M})$, *fermé vu comme sous-groupe du groupe des automorphismes de* $\mathcal{G}$, *ayant la propriété* (P), *agissant de manière non élémentaire sur* $\mathcal{G}$ *et d'ensemble limite égal à* $\partial\mathcal{G}$. *Si* G^+ *est non trivial, alors* G^+ *est non dénombrable.*

Preuve Soit M un mur propre, de fixateur strict non trivial. Soit G_M^+ le sous-groupe de G fixant strictement le mur M et fixant l'un des demi-espaces, disons A, définis par M. Le sous-groupe G_M^+ est fermé dans G, donc dans

$\mathrm{Aut}(\mathcal{G})$. On en déduit que G_M^+ est localement compact. Pour montrer qu'il est non dénombrable, il suffit de montrer qu'il n'a pas de point isolé, et comme c'est un groupe topologique, que l'identité n'est pas isolée.

Soit g un élément non trivial de G_M^+, qui existe par la propriété (P) quitte à échanger A et $X \setminus A$, et K une partie compacte arbitraire de $\mathcal{G}$. Puisque M est propre, soit x un point de $\partial X \setminus \partial A$. Soit U un ouvert, contenu dans $\overline{X} \setminus (\overline{A} \cap K)$, contenant x. Puisque G est non élémentaire, il existe un élément hyperbolique h dans G dont le point fixe attractif est contenu dans U et le point fixe répulsif dans $\partial X \setminus \partial(X \setminus A)$. Si n est assez grand, alors $h^n(\overline{X \setminus A})$ est contenu dans U. Posons $g_n = h^n g h^{-n}$, qui appartient à G et même à G_M^+. Comme g vaut l'identité sur A, l'élément g_n vaut l'identité sur $h^n(A)$, donc sur K. Puisque g est non trivial, g_n l'est aussi. On en déduit que l'identité n'est pas isolée dans G_M^+. □

3.2 L'exemple classique des systèmes de Coxeter

Adoptons un premier point de vue algébrique (on trouvera dans [5, Chapitre IV, Section 1, Exemple 16], [27] toutes les justifications des affirmations ci-dessous). Soient (W, S) un système de Coxeter, $\mathcal{T}$ l'ensemble de ses *réflexions* (i.e. des conjugués dans W des éléments de S), et $\ell(w)$ la longueur minimale d'une écriture de $w \in W$ comme mot sur S. Pour $t \in \mathcal{T}$, posons:

$$A_t^+ = \{w \in W, \ell(w) < \ell(tw)\} \ \text{ et } \ A_t^- = \{w \in W, \ell(w) > \ell(tw)\}.$$

Alors A_t^+ contient 1_W et A_t^- contient t. De plus, $\ell(w)$ et $\ell(tw)$, n'ayant pas la même parité, sont toujours différents. Donc $\{A_t^+, A_t^-\}$ est un mur de W (les demi-espaces $A_t^\pm$ sont appelées moitiés dans [5]). Notons $\mathcal{M}(W,S) = \mathcal{M}$ l'ensemble des murs ainsi obtenus (en correspondance biunivoque avec $\mathcal{T}$). Montrons que $\mathcal{M}(W,S)$ vérifie l'axiome (M).

Pour $w', w'' \in W$, l'ensemble des murs séparant w' de w'' correspond à l'ensemble des réflexions t de la forme $s_1 \dots s_{i-1} s_i s_{i-1} \dots s_1$, pour une écriture géodésique fixée $w'^{-1}w'' = s_1 \dots s_n$, avec $n = \ell(w'^{-1}w'')$. Il y a n telles réflexions, autrement dit $\mathrm{card}\, \mathcal{M}(w', w'') = \ell(w'^{-1}w'')$. En particulier, l'axiome (M) est vérifié, et le graphe de l'espace à murs $(W, \mathcal{M})$ s'identifie au graphe de Cayley de (W, S). Cette identification est W–équivariante (l'image par w de A_t^+ est $A_{t'}^\varepsilon$, où $t' = w^{-1}tw$ et $\varepsilon = +$ si $w \in A_t^+$, $\varepsilon = -$ sinon).

On peut aussi définir le système de murs $\mathcal{M}$ sur W en considérant diverses actions de W sur des complexes polyédraux.

Si W agit sur un espace P et si t est une réflexion, appelons *mur* de t dans P, et notons $M(t,P)$, l'ensemble des points fixes de t dans P. Pour P, prenons successivement le graphe de Cayley de (W,S) (noté $\mathcal{G}(W,S)$), la réalisation géométrique standard de (W,S) (notée $|W|$, voir [27]), et enfin sa réalisation géométrique au sens de Davis–Moussong (notée $|W|_0$). Chacun de ces trois complexes est un "appartement" au sens de Davis, voir [15] pour les définitions et propriétés concernant ces espaces W–homogènes; le complexe $|W|_0$ est introduit dans [15], et muni d'une métrique CAT(0) dans [25].

Notons que $|W|$ est un complexe simplicial de dimension card $S-1$ sur lequel W agit, de manière simplement transitive sur les simplexes de dimension maximale. On identifie les éléments de W aux centres de ces simplexes maximaux.

(Rappelons brièvement la construction de $|W|_0$. Soit Δ_S le simplexe standard d'ensemble de sommets S, dont les faces s'identifient aux parties de S. Si T est une partie de S, on note W_T le *sous-groupe spécial* de W engendré par T. Soit $N = N(W,S)$ le sous-complexe simplicial de Δ_S, appelé *nerf fini* de (W,S), dont les simplexes sont les parties T de S telles que W_T soit fini. En particulier, N contient tous les sommets de Δ_S. Soit $C(W,S) = x_0 * N'$ le cône simplicial (de sommet x_0) sur la subdivision barycentrique N' de N. Pour tout sommet s de N, on note F_s l'étoile de s dans N', naturellement contenu dans $C(W,S)$. On considère alors le quotient

$$W \times C(W,S)/\sim$$

où $\sim$ est la relation d'équivalence engendrée par $(w,x) \sim (w',x')$ s'il existe $s \in S$ tel que $w' = ws$ et $x' = x \in F_s$. On montre (voir [25]) que ce quotient admet une structure de subdivision barycentrique d'un complexe polyédral euclidien CAT(0) $|W|_0$, d'ensemble de sommets l'image de $W \times \{*\}$, que l'on identifie avec W.)

Pour chacune des trois actions considérées,

- le mur M d'une réflexion de W sépare P en deux composantes connexes, appelées *demi-espaces* de P définis par M;
- dans P, il y a un plongement W–équivariant de $\mathcal{G}(W,S)$, étendant celui de W (c'est le 1–squelette de $|W|_0$ par construction, et le graphe dual de $|W|$);
- si t est une réflexion, son mur dans P évite W, et deux éléments de W sont dans une même composante connexe de $P - M(t,P)$ si et seulement s'ils le sont dans $\mathcal{G}(W,S) - M(t,\mathcal{G}(W,S))$.

C'est pourquoi, pour chaque réflexion t, les intersections de W avec les deux demi-espaces de P définis par $M(t,P)$ donnent un mur de W indépendant de P. D'autre part, on vérifie que, si $P=\mathcal{G}(W,S)$, l'ensemble de murs ainsi obtenu est $\mathcal{M}(W,S)$.

Puisque le graphe de l'espace à murs $(W,\mathcal{M})$ s'identifie au graphe de Cayley de (W,S), il est hyperbolique (au sens de Gromov) si et seulement si W est un groupe hyperbolique. Nous vérifierons dans la section suivante que la condition (H) est satisfaite. Pour information, par un théorème de G. Moussong [25], les conditions suivantes sont équivalentes:

(1) W est un groupe hyperbolique;

(2) W ne contient pas de sous-groupe isomorphe à $\mathbb{Z}\times\mathbb{Z}$;

(3) il n'existe pas de partie T de S telle que (W_T,T) soit un système de Coxeter affine de rang au moins 3, ni de paires de parties T_1,T_2 de S, disjointes, avec W_{T_1},W_{T_2} commutants et infinis.

Cas particuliers (Complexes de Benakli–Haglund, voir [4, 23]) Soit k un entier pair au moins 4, et L un graphe fini (sans boucle ni arête double), de maille (i.e. la plus petite longueur d'un cycle) au moins 5 si $k=4$, et 4 si $k=6$. Soit $(W(k,L),S(k,L))$ le système de Coxeter de matrice de Coxeter la matrice d'adjacence du graphe L, avec les 1 et 0 remplacés respectivement par $\frac{k}{2}$ et ∞. Il vérifie clairement la condition (3) ci-dessus.

Nous noterons $A(k,L)$ la réalisation géométrique au sens de Davis–Moussong de ce système de Coxeter. Alors (voir [23]) $A(k,L)$ est un complexe polygonal $\mathrm{CAT}(-1)$, dont les polygones sont des k–gones hyperboliques, le link de chaque sommet étant isomorphe à L.

Si p est un entier pair et $L_{q,q}$ est le graphe biparti complet sur $q+q$ sommets, alors l'immeuble de Bourdon $I_{p,q}$ est isomorphe, en tant que complexe polygonal, à $A(p,L_{q,q})$.

Un autre exemple d'espace à murs Par contre, si $p=2m+1$ est impair et $q\geq 5$, l'immeuble de Bourdon $I_{p,q}$ n'est isomorphe ni à un complexe polygonal $A(k,L)$ ni à un complexe cubique (sauf à passer à une subdivision). Supposons $p\geq 7$. Pour chaque côté fixé A du p–gone régulier à angles droits P, numérotons cycliquement $A=A_1,A_2,\cdots,A_p$ les côtés de P. Considérons les deux segments de perpendiculaire commune aux paires de côtés respectivement $A,A_{p-1/2}$ et $A,A_{p+1/2}$. Notons α_1,α_2 ces segments.

Nous appellerons *mur* de $I_{p,q}$ toute partie M de $I_{p,q}$ ainsi obtenue. Pour tout $i=1,2$ et pour toute identification isométrique d'un polygone de $I_{p,q}$ avec

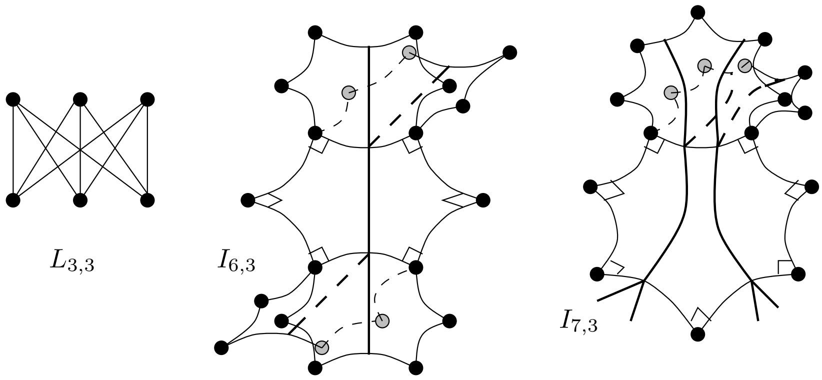


Figure 2: L'immeuble de Bourdon: son link, cas p pair, cas p impair

P, on considère la réunion M de toutes les géodésiques de $I_{p,q}$ passant par le segment α_i. Nous notons $X_{p,q}$ l'ensemble des sommets de $I_{p,q}$, et *mur* de $X_{p,q}$ la partition de $X_{p,q}$ obtenue en prenant l'intersection de $X_{p,q}$ avec les deux composantes connexes du complémentaire d'un mur de $I_{p,q}$. (Comme $I_{p,q}$ est simplement connexe, et qu'un mur sépare localement en deux composantes connexes, il sépare globalement en deux composantes connexes.)

Il est facile de montrer que l'espace à murs $(X_{p,q}, \mathcal{M}_{p,q})$ ainsi défini vérifie l'axiome (M). Le graphe associé $\mathcal{G}$ s'identifie avec le 1–squelette de $I_{p,q}$, mais les deux sommets de chaque arête de $I_{p,q}$ sont séparés par exactement deux murs. Ce système de demi-espace est différent de celui obtenu par subdivision en complexe cubique. Comme $I_{p,q}$ est CAT(-1), son 1–squelette est un espace métrique hyperbolique, de même bord que $I_{p,q}$. La condition (H) est facile à vérifier.

Le groupe Aut $I_{p,q}$ des automorphismes de complexe polygonal de l'immeuble de Bourdon $I_{p,q}$ s'identifie naturellement à Aut $(X_{p,q}, \mathcal{M}_{p,q})$. En effet, tout automorphisme de $I_{p,q}$ est une isométrie pour la distance de $I_{p,q}$, et donc envoie tout segment de perpendiculaire commune entre deux arêtes à distance cyclique $q-1/2$ ou $q+1/2$ sur le bord d'un polygone de $I_{p,q}$ sur un tel autre segment. Donc il préserve l'ensemble des sommets $X_{p,q}$ de $I_{p,q}$, ainsi que l'ensemble $\mathcal{M}_{p,q}$ des demi-espaces, et Aut $I_{p,q}$ est contenu dans Aut $(X_{p,q}, \mathcal{M}_{p,q})$.

Comme les seuls cycles de longueur p dans le 1–squelette de $I_{p,q}$ sont les bords des polygones, il en découle que Aut $I_{p,q}$ est égal à Aut $(X_{p,q}, \mathcal{M}_{p,q})$

4 Complexes polyèdraux pairs à courbure négative ou nulle

4.1 Polyèdres pairs

Un polyèdre (compact convexe) C d'une variété riemanienne (complète, simplement connexe) à courbure constante ≤ 0 est *pair* si

> pour toute arête a de C, l'unique réflexion $\sigma_{a,C}$ de l'espace ambiant échangeant les extrémités de a préserve C, mais ne fixe aucun sommet de C.

Par exemple, si C est un polygone régulier, il est pair si et seulement s'il a un nombre pair de côtés. Un cube euclidien régulier de dimension quelconque est pair. Plus généralement, le produit de deux polyèdres euclidiens pairs est un polyèdre euclidien pair. Voir figure 3 pour d'autres exemples. Nous donnons ci-dessous une caractérisation constructive de tous les polyèdres pairs.

Soit $\mathbb{X}_\kappa$ l'espace à courbure constante $\kappa \leq 0$ de dimension n. Si $\kappa = 0$, nous prendrons $\mathbb{X}_\kappa = \mathbb{R}^n$. Si $\kappa < 0$, nous utiliserons le modèle de la boule de Poincaré pour l'espace hyperbolique $\mathbb{X}_\kappa$ à courbure constante κ. Le groupe des isométries de $\mathbb{X}_\kappa$ fixant l'origine s'identifie alors avec $O(n)$. Notons $\phi\colon \mathbb{R}^n \to \mathbb{X}_\kappa$ l'exponentielle riemannienne en l'origine (l'identité si $\kappa = 0$). Soit W un groupe fini engendré par des réflexions sur des hyperplans vectoriels de $\mathbb{R}^n$. L'application ϕ permet alors de définir les notions de *chambres, murs ... dans* $\mathbb{X}_\kappa$ pour l'action isométrique de W sur $\mathbb{X}_\kappa$.

Proposition 4.1 *Un polyèdre (compact convexe) C d'un espace $\mathbb{X}_\kappa$ à courbure constante $\kappa \leq 0$ est pair si et seulement s'il existe un point x dans $\mathbb{X}_\kappa$, un système de Coxeter fini (W, S) et une représentation (injective, envoyant chaque élément de S sur une réflexion) ρ de W dans le groupe des isométries de $\mathbb{X}_\kappa$ fixant x telle que C est l'enveloppe convexe de l'orbite par W d'un point y de l'intérieur d'une chambre. De plus, le 1–squelette de C est isomorphe au graphe de Cayley de (W, S).*

Preuve Supposons tout d'abord que C est pair. Notons W le groupe engendré par les réflexions dans $\mathbb{X}_\kappa$ par rapport aux hyperplans médiateurs des arêtes de C. Puisque C est invariant par W, le groupe W est fini et admet au moins un point fixe, le *centre métrique* x de la cellule C (c'est le centre de l'unique plus petite boule de $\mathbb{X}_\kappa$ contenant C). Nous supposerons que x est l'origine de $\mathbb{X}_\kappa$.

Fixons y un sommet de C, et notons S l'ensemble des réflexions dans $\mathbb{X}_\kappa$ par rapport aux hyperplans médiateurs des arêtes de C ayant y pour sommet. Par connexité du 1–squelette de C, le groupe W est engendré par S. Puisque c'est vrai au niveau de l'espace tangent en x (voir [5] par exemple), le groupe W agit simplement transitivement sur les chambres dans $\mathbb{X}_\kappa$ (qui sont les composantes connexes du complémentaire des hyperplans médiateurs des arêtes). Tout sommet de C est contenu dans une chambre, et la chambre contenant y ne contient pas d'autre sommet de C. Donc le groupe W agit simplement transitivement sur les sommets de C. Le sommet y de C est joint par une arête précisément aux sommets sy avec s dans S. Par définition du graphe de Cayley, le 1–squelette de C s'identifie donc au graphe de Cayley de (W, S). Comme C est l'enveloppe convexe de ses sommets, C est bien l'enveloppe convexe de l'orbite de y par W.

Réciproquement, soit C l'enveloppe convexe de l'orbite par W d'un point y de l'intérieur d'une chambre pour une représentation comme dans l'énoncé d'un système de Coxeter fini (W, S). Montrons que C est pair. Puisque toutes les images de y par W sont à la même distance de x, par convexité stricte des sphères, les sommets de C sont exactement les images de y par W. Le même argument de convexité stricte montre que le point y est strictement au-dessus de l'hyperplan affine passant par les sy pour s dans S. Donc les segments de droites entre y et les sy sont des arêtes de C. □

Proposition 4.2 *Soit C un polyèdre pair d'un espace à courbure constante négative ou nulle. Alors C est simple, i.e. les links de ses sommets sont des simplexes (sphériques). Si C est euclidien, alors les longueurs des arêtes des links de faces de C sont dans $[\frac{\pi}{2}, \pi]$ (et en particulier ses angles dièdres sont obtus).*

Preuve Comme le type combinatoire des polyèdres pairs ne dépend pas de la courbure, nous pouvons supposer C euclidien. Si la dimension n de C est égale à celle de l'espace ambiant (ce que nous pouvons toujours supposer), le groupe fini engendré par des réflexions W construit ci-dessus est *essentiel* (i.e. il ne fixe aucun vecteur tangent au centre métrique de C non nul). Si v est un sommet de C, alors les sommets du link de v sont en bijection avec les murs de la chambre contenant v. Or (voir [5, Ch. V, section 3, Prop. 7]) les chambres sont des cônes simpliciaux. Donc le link de v (qui est de dimension $n-1$) a exactement $n-1$ sommets, et est donc un simplexe.

Si a, b sont deux arêtes de C, le plan P qui les contient rencontre perpendiculairement les hyperplans médiateurs de a, b en deux droites α, β. Les arêtes

a, b et les droites α, β définissent un quadrilatère dont deux angles sont droits et l'un des deux autres est l'angle dièdre entre les hyperplans médiateurs de a, b. L'angle dièdre entre deux murs d'une même chambre est dans $[0, \frac{\pi}{2}]$, donc l'angle entre deux arêtes de C est dans $[\frac{\pi}{2}, \pi]$. La longueur de toute arête du link de tout sommet s de C est donc dans $[\frac{\pi}{2}, \pi]$. Par les formules de trigonométrie sphérique, il en découle que l'angle en un sommet d'une 2–face du link de s est au moins $\frac{\pi}{2}$, donc que la longueur des arêtes des links de face de dimension 1 est au moins $\frac{\pi}{2}$. Le résultat en découle par récurrence sur la dimension de la face. □

Nous donnons ci-dessous la liste complète des polyèdres hyperboliques pairs qui sont des polyèdres de Coxeter, en dimension 2 et 3. Dans le tableau suivant, m est un entier, avec $m = 5$ ou $m \geq 7$. À tout polyèdre pair P de dimension n, et à tout sommet x_0 de celui-ci, est associé par la proposition 4.1 un système de Coxeter fini (W, S) de rang n, dont nous donnons le type et le diagramme de Coxeter. Les arêtes de P issues de x_0 sont en bijection avec les éléments de S. Si P est de dimension 3, nous donnons les angles dièdres $(\alpha_a, \alpha_b, \alpha_c)$ des arêtes issues de x_0 correspondant aux éléments de $S = \{a, b, c\}$. Par la formule de Gauss–Bonnet, un polygone hyperbolique pair est déterminé à isométrie près par (m, α, ℓ) dans $\mathbb{N} \setminus \{0,1\} \times]0, \frac{(p-1)\pi}{p}[\times]0 + \infty[$, avec $2m$ son nombre de côtés, α l'angle en chacun de ses sommets, et ℓ la longueur d'un de ses côtés (et donc des côtés à distance paire de celui-ci).

Rang 2			Rang 3		
(W,S)		α	(W,S)		$(\alpha_a, \alpha_b, \alpha_c)$
$A_1 \times A_1$	• •	$\frac{\pi}{n}, n \geq 3$	$A_1 \times W$	a • b •—m—• c	$(\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{n}), n = 3, 4, 5$
A_2	•—•	$\frac{\pi}{n}, n \geq 2$	où $W =$	A_2, B_2, G_2	ou $I_2(m)$
B_2	•—4—•	$\frac{\pi}{n}, n \geq 2$	A_3	a •—b •—• c	$(\frac{\pi}{2}, \frac{\pi}{n}, \frac{\pi}{2}), n \geq 3$ $(\frac{\pi}{2}, \frac{\pi}{n}, \frac{\pi}{3}), n = 3, 4, 5$
G_2	•—6—•	$\frac{\pi}{n}, n \geq 2$	B_3	a •—b •—4—• c	$(\frac{\pi}{2}, \frac{\pi}{n}, \frac{\pi}{2}), n \geq 3$ $(\frac{\pi}{2}, \frac{\pi}{n}, \frac{\pi}{3}), n = 3, 4, 5$
$I_2(m)$	•—m—•	$\frac{\pi}{n}, n \geq 2$	H_3	a •—b •—5—• c	$(\frac{\pi}{2}, \frac{\pi}{n}, \frac{\pi}{2}), n \geq 3$ $(\frac{\pi}{2}, \frac{\pi}{n}, \frac{\pi}{3}), n = 3, 4, 5$

Proposition 4.3 *À isométrie près, un polyèdre hyperbolique pair de dimension 2 ou 3 qui est un polyèdre de Coxeter est donné à isométrie près par le*

tableau précédent (avec un paramètre libre $\ell \in]0, +\infty[$ *en rang* 2*).*

Preuve Soit (W, S) un système de Coxeter de rang 3. Soit Z la cellulation duale de la subdivision barycentrique τ de la cellulation de la sphère $\mathbb{S}^2$ décrite ci-dessous:

- la cellulation de la sphère $\mathbb{S}^2$ par $4, 6, 8, 12, 2m$ bigones si (W, S) est de type $A_1 \times W$ avec W le groupe de Coxeter de rang 2 de type $A_1 \times A_1, A_2, B_2, G_2, I_2(m)$ respectivement;
- la cellulation bord du tétraèdre, cube, dodécaèdre si (W, S) est de type A_3, B_3, H_3 respectivement.

Notons que si P est un polyèdre hyperbolique pair construit à partir de (W, S) comme dans la proposition 4.1, alors son bord est isomorphe à la cellulation Z.

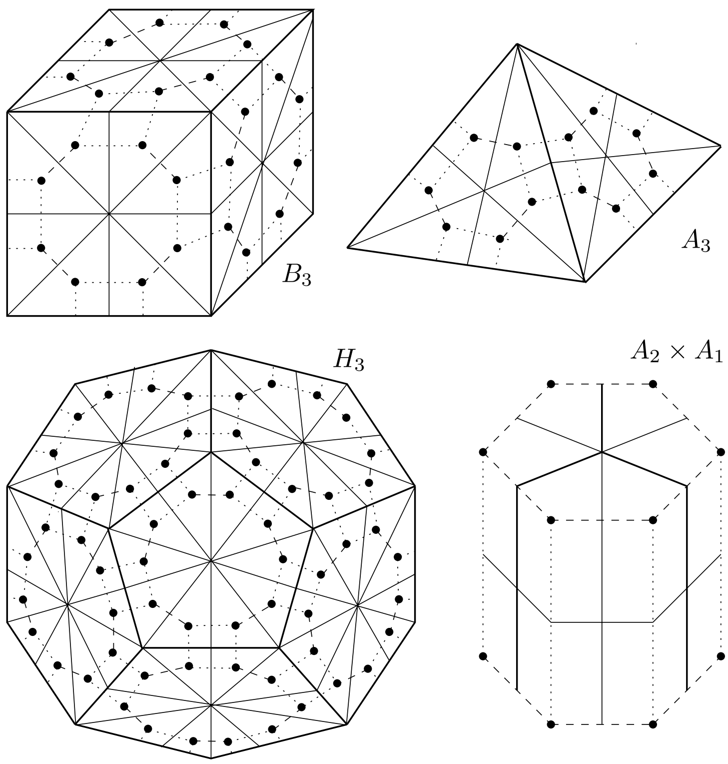


Figure 3: Polyèdres de Coxeter hyperboliques pairs de dimension 3

Par le théorème d'Andréev [1], si α est une application de l'ensemble des arêtes de Z dans $]0, \frac{\pi}{2}]$, alors il existe un polyèdre hyperbolique (compact), unique à isométrie près, dont la cellulation du bord est isomorphe à Z, avec angle dièdre $\alpha(z)$ le long d'une arête z si et seulement si

(1) la somme des angles le long d'un cycle de longueur 3 dans τ qui ne borde pas un triangle de τ est strictement inférieure à π,

(2) la somme des angles le long d'un cycle de longueur 3 dans τ qui borde un triangle de τ est strictement supérieure à π,

(3) la somme des angles le long d'un cycle de longueur 4 dans τ qui ne borde pas la réunion de deux triangles de τ est strictement inférieure à 2π.

Comme il n'existe pas de cycle de longueur 3 dans τ qui ne borde pas un triangle, et que les seuls triangles sphériques de Coxeter ont pour angles $\{\frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{n}\}$, $n \geq 2$ ou $\{\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{n}\}, n = 3, 4, 5$, le résultat en découle par examination des divers cas possibles. L'unicité découle de l'unicité dans le théorème d'Andreev, en remarquant que ces polyèdres ont une symétrie supplémentaire (i.e. qui n'est pas dans W), par rapport à un hyperplan passant par des sommets. □

Définissons maintenant la notion de parallélisme d'arêtes. Soit C un polyèdre pair de dimension quelconque. Si a est une arête de C, nous noterons $M(a, C)$ l'ensemble des points de C fixes par $\sigma_{a,C}$. C'est un convexe compact de codimension 1 dans C, séparant C en deux composantes connexes. Il ne peut rencontrer une arête b de C qu'en son milieu, et perpendiculairement: dans ce cas $M(a, C) = M(b, C)$. Deux arêtes a, b de C sont dites *parallèles dans* C si $M(a, C) = M(b, C)$. La relation de parallélisme dans C est une relation d'équivalence sur les arêtes de C.

4.2 L'espace à murs d'un complexe polyédral pair

Soit P un complexe polyédral, n'ayant qu'un nombre fini de types d'isométrie de cellules. Nous dirons que P est un *complexe polyédral pair* si toute cellule C de P est paire. Par exemple, un arbre, ou plus généralement un complexe cubique (voir [21, 28, 26]) est un complexe polyédral pair.

La réunion des relations de parallélisme sur les arêtes d'une même cellule de P engendre une relation d'équivalence sur l'ensemble de toutes les arêtes de P, que nous appelerons *parallélisme entre arêtes* dans P. (Voir [28, section 2.4] pour le cas des complexes cubiques.) Définissons alors le *mur de* P *transverse à une arête* a comme l'union des $M(b, C')$, avec b une arête parallèle à a contenue dans une cellule (maximale pour l'inclusion) C' de P.

Puisque P n'a qu'un nombre fini de types d'isométrie de cellules, et les compacts $M(b, C')$ ne contenant aucun sommet de P, car C' est pair, il vient:

- tout mur de P est fermé, (localement compact si le link de toute cellule de P de dimension > 0 est compact) et évite l'ensemble X des sommets de P;
- l'ensemble des murs de P est localement fini.

Comme dans le cas des complexes cubiques [28, Theo. 4.10], le premier résultat est le suivant.

Lemme 4.4 *Soit P un complexe polyédral pair* $\mathrm{CAT}(0)$ *et M le mur de P transverse à une arête a. Alors M est convexe dans P, et sépare P en deux composantes connexes.*

Preuve Soit $V(M)$ l'union des cellules de P contenant une arête parallèle à a. Donnons d'abord une description du revêtement universel de $V(M)$.

Soit $\mathcal{C}(a)$ l'ensemble des suites de la forme $(a_0, a_1, \dots, a_n, C)$, où les $(a_i)_{0 \le i \le n}$ sont des arêtes de P, avec $a_0 = a$, a_i parallèle à a_{i+1} dans une cellule de P, et C est une cellule de P contenant a_n. Si a_i, a_{i+1} et a_{i+2} sont trois arêtes parallèles à a dans une même cellule C', nous dirons qu'il y a entre $(a_0, \dots, a_i, a_{i+1}, a_{i+2}, \dots, a_n, C)$ et $(a_0, \dots, a_i, a_{i+2}, \dots, a_n, C)$ une *homotopie élémentaire* (à extrémités fixées). Les homotopies élémentaires engendrent une relation d'équivalence sur $\mathcal{C}(a)$: nous noterons $[a_0, \dots, a_n, C]$ la classe d'équivalence de $(a_0, \dots, a_n, C)$ pour cette relation.

Soit $\overline{V}(M)$ le complexe polyédral obtenu à partir de l'union disjointe des cellules de la forme $[a_0, \dots, a_n, C] \times C$ en identifiant deux points de la forme $([a_0, \dots, a_n, C'], x')$ et $([a_0, \dots, a_n, C''], x'')$ lorsque $x' = x''(\in C' \cap C'')$. Notons p: $\overline{V}(M) \to V(M)$ l'application polyédrale naturelle. Alors p est surjective et un isomorphisme sur chaque cellule. Via p, le complexe $\overline{V}(M)$ hérite d'une structure de complexe polyédral, n'ayant qu'un nombre fini de types d'isométrie de cellules (pour laquelle p est une isométrie sur chaque cellule). Montrons que sur $\overline{M} = p^{-1}(M)$, l'application p est une isométrie.

D'abord, $\overline{M}$ est localement convexe dans $\overline{V}(M)$. En effet, $\overline{V}(M)$ possède une réflexion $\overline{\sigma}_a$ (obtenue sur chaque cellule $\overline{C}$ de $\overline{V}(M)$ image de $[a_0, \dots, a_n, C] \times C$ en conjugant $\sigma_{a_n, C}$ par $p_{|\overline{C}}$). L'ensemble des points fixes de l'isométrie $\overline{\sigma}_a$ est précisément $\overline{M}$. Or la métrique de $\overline{V}(M)$ est localement convexe. Il en résulte que $\overline{M}$ est localement convexe dans $\overline{V}(M)$.

Ensuite, l'image d'une géodésique $\overline{\gamma}$ de $\overline{M}$ par p est une géodésique de P contenue dans M. En effet, on remarque d'abord que $\overline{\gamma}$ est une géodésique locale de $\overline{V}(M)$, puis que p est une isométrie locale au voisinage de $\overline{M}$. Donc

$p(\overline{\gamma})$ est une géodésique locale de P. Mais comme P est CAT(0), ceci implique que $p(\overline{\gamma})$ est une géodésique globale de P.

Puisque $\overline{M}$ est évidemment connexe, M est convexe dans P, et p induit une isométrie de $\overline{M}$ sur M.

En fait, p: $\overline{V}(M) \to V(M)$ est un homéomorphisme. En effet, notons d'abord qu'un point $\overline{x}$ de $\overline{V}(M)$ est dans une cellule minimale $\overline{C}_{\overline{x}}$ de $\overline{V}(M)$ rencontrant $\overline{M}$. Si $\overline{x}'$ désigne la projection orthogonale de $\overline{x}$ sur $\overline{M} \cap \overline{C}_{\overline{x}}$, alors toute géodésique de $\overline{M}$ issue de $\overline{x}'$ fait avec $[\overline{x}', \overline{x}]$ un angle au moins égal à $\frac{\pi}{2}$. Maintenant, si deux points $\overline{x}$ et $\overline{y}$ de $\overline{V}(M) - \overline{M}$ sont identifiés par p, il apparaît dans P un triangle de sommets $p(\overline{x}) = p(\overline{y})$, $p(\overline{x}')$ et $p(\overline{y}')$, avec des angles à la base supérieurs ou égaux à $\frac{\pi}{2}$. Comme P est CAT(0), cela n'est possible que si $p(\overline{x}') = p(\overline{y}')$. Donc $x' = \overline{y}'$, et $\overline{C}_{\overline{x}} = \overline{C}_{\overline{y}}$. Or p est un plongement sur chaque cellule: donc $\overline{x} = \overline{y}$.

Après avoir vérifié que $\overline{M}$ sépare $\overline{V}(M)$ en deux composantes connexes, on en déduit que M sépare P en deux composantes connexes (parce qu'il sépare son voisinage $V(M)$, et que P est simplement connexe). □

Le résultat suivant découle aussi de la preuve du lemme précédent.

Lemme 4.5 *Pour toute cellule C de P maximale pour l'inclusion, le mur de P transverse à une arête a de C est la réunion de tous les segments géodésiques rencontrant $M(a, C)$ en un intervalle d'intérieur non vide.* □

Soient $X = X_P$ l'ensemble des sommets de P et M un mur de P transverse à une arête; notons $P^+(M)$ et $P^-(M)$ les deux composantes connexes de $P - M$. Comme $X \cap M = \emptyset$, la paire $\{X \cap P^+(M), X \cap P^-(M)\}$ est une partition de X. Nous noterons encore M ce mur de X, et $\mathcal{M} = \mathcal{M}_P$ l'ensemble des murs de X ainsi défini.

Proposition 4.6 *Soit P un complexe polyédral pair* CAT(0).
Alors $(X_P, \mathcal{M}_P)$ est un espace à murs.

Preuve Vérifions que $\mathcal{M}$ satisfait l'axiome (M).

Soient x et y deux sommets de P, et γ la géodésique de P qui les joint. Tout mur de $\mathcal{M}(x, y)$ correspond à un mur de P séparant topologiquement x et y, donc coupant γ. L'ensemble des murs de P étant localement fini, on en déduit que $\mathcal{M}(x, y)$ est fini.

D'autre part, γ part de x par l'intérieur d'une (unique) cellule C, elle doit traverser un des murs $M(a, C)$ (avec a une arête issue de x) avant de retoucher ∂C: donc $\mathcal{M}(x, y)$ est non vide. □

Avant de poursuivre l'étude de cet exemple fondamental, il convient de faire quelques remarques.

Remarque 1 Le système de murs d'un système de Coxeter (W,S) défini dans la section 3 peut s'obtenir par la présente construction, en prenant pour P la réalisation géométrique au sens de Davis–Moussong $|W|_0$ de (W,S). Par construction même (voir [25]), une cellule de $|W|_0$ est paire, le groupe engendré par les réflexions orthogonales le long des arêtes de la cellule étant isomorphe à un sous-groupe spécial fini de (W,S); d'autre part, $|W|_0$ est bien CAT(0) (voir [25]).

Remarque 2 De nombreux complexes polyèdraux CAT(0) admettent des subdivisions régulières cubiques qui restent CAT(0) lorsqu'on munit les cubes de leurs métriques euclidiennes standard. Par exemple, si P est un complexe polygonal CAT(0) sans triangle tel que le link d'un sommet de P ne contient aucun circuit de longueur 3, alors la subdivision de chaque k–gone de P en k carrés, identifiés au carré euclidien unité, fournit un complexe carré encore CAT(0). Voir par exemple l'exemple à la fin de la section 3.2, où le système de murs est toutefois différent de celui obtenu par subdivision cubique. Ce genre de subdivision permet d'appliquer nos résultats de simplicité à des complexes polyèdraux CAT(0) non nécessairement pairs (comme l'immeuble de Bourdon avec p impair).

Remarque 3 On pourrait penser que tout complexe polyédral pair CAT(0) peut être subdivisé en cubes, tout en restant CAT(0), et donc qu'il suffit d'étudier les complexes cubiques. Mais il n'en est rien, comme le montre l'exemple suivant en dimension 2.

Soient ℓ et m deux entiers supérieurs ou égaux à 3. Considérons un ensemble $S_{\ell,m}$ de ℓm points, répartis en ℓ colonnes de m points chacune. Relions deux points de $S_{\ell,m}$ si et seulement s'ils n'appartiennent pas à la même colonne. Nous noterons $K_{\ell,m}$ le graphe ainsi obtenu (dont le graphe complémentaire est donc une union disjointe de ℓ graphes complets sur m sommets). Comme $\ell \geq 3$, ce graphe contient des circuits de longueur 3. Fixons d'autre part un entier $k \geq 4$. Nous pouvons considérer le système de Coxeter $(W_{k,\ell,m}, S_{\ell,m})$ dont le graphe de Coxeter a des arêtes de poids infini entre points d'une même colonne, et des arêtes de poids k entre points n'appartenant pas à la même colonne. Alors la réalisation géométrique de Davis–Moussong de $(W_{k,\ell,m}, S_{\ell,m})$ est (la subdivision barycentrique d')un complexe polygonal $W_{k,\ell,m}$–homogène X, dont les polygones sont hyperboliques réguliers à $2k$ côtés, d'angle aux sommets $\frac{2\pi}{3}$, et tel que le link de chaque sommet est isomorphe à $K_{\ell,m}$ (voir

[23]). Donc X est un complexe polyédral pair CAT(0). Une subdivision en carrés de X donne alors des angles aux sommets égaux à $\frac{\pi}{2}$, donc des circuits de longueur égale à $\frac{3\pi}{2} < 2\pi$ dans le link métrique des sommets, ce qui empêche X d'être CAT(0).

4.3 Le graphe associé à l'espace à murs d'un complexe polyédral pair

Soit à nouveau P un complexe polyédral pair CAT(0) et $\mathcal{M} = \mathcal{M}_P$ son système de murs sur l'ensemble $X = X_P$ de ses sommets. Nous étudions maintenant les géodésiques du 1–squelette $\mathcal{G}$ de P, pour la métrique géodésique sur $\mathcal{G}$ rendant chaque arête isométrique au segment unité (qui n'est pas forcément celle induite par P). Nous allons voir que cette métrique sur $\mathcal{G}$ vérifie des propriétés analogues à la métrique des mots d'un système de Coxeter.

Si $c = (a_0, a_1, \ldots, a_n)$ est un chemin combinatoire de $\mathcal{G}$ empruntant les $n+1$ arêtes $a_0, a_1, \ldots, a_n$, nous noterons $M(c)$ la suite $M(a_0), M(a_1), \ldots, M(a_n)$ des murs traversés par c.

Lemme 4.7 *Soit c un chemin combinatoire de $\mathcal{G}$ d'extrémités x et y.*

a) *Un mur M sépare x de y si et seulement s'il apparaît un nombre impair de fois dans la suite $M(c)$.*

b) *Si la suite $M(c)$ est sans répétition, alors c est une géodésique de $\mathcal{G}$.*

Preuve a) D'une part, tout mur séparant x de y est traversé par c. D'autre part, si un mur M est traversé un nombre pair de fois par c, c'est donc que x et y sont dans la même composante connexe de $P - M$.

b) Il résulte du a) que, pour un tel chemin, l'ensemble des murs traversés par c est $\mathcal{M}(x, y)$, et la longueur de c est le cardinal de $\mathcal{M}(x, y)$. Si c' est un autre chemin d'extrémités x et y, sa longueur est égale au nombre de murs qu'il traverse, donc au moins égale au nombre de murs qu'il traverse un nombre impair de fois. Donc c' est au moins aussi long que c. □

Lemme 4.8 *Soit $\mathcal{O}$ l'ouvert des points de P qui ne sont sur aucun mur de P. Alors toute composante connexe de $\mathcal{O}$ contient un et un seul sommet de P.*

Preuve Puisque deux sommets distincts de P sont toujours séparés par un mur, il y a au plus un sommet de P par composante connexe de $\mathcal{O}$.

Pour la réciproque, il suffit de considérer le cas où P est réduit à une cellule C. Par la proposition 4.1, ceci découle du fait qu'un groupe de Coxeter fini agit simplement transitivement sur ses chambres. □

Le résultat suivant montre qu'on peut accompagner une géodésique de P par une géodésique de son 1–squelette.

Lemme 4.9 *Soient x' et y' deux points du polyèdre P n'appartenant à aucun mur de P, et γ la géodésique qui les joint dans P. Alors il existe un chemin combinatoire c de $\mathcal{G}$ tel que $M(c)$ est sans répétition, contenu dans la réunion $V(\gamma)$ des cellules de P touchant γ, et d'extrémités x et y définis par: x' et x (resp. y' et y) sont dans la même composante connexe de $\mathcal{O}$.*

Preuve D'abord, par le lemme 4.5, la géodésique γ rencontre un nombre fini de murs, en des points distincts $z_1, z_2, \ldots, z_n$. Pour prouver le lemme, il suffit de l'établir lorsque $n = 1$. En effet, pour n quelconque, on découpe γ en n segments géodésiques successifs γ_i, contenant z_i, d'extrémités x'_i et x'_{i+1} contenues dans aucun mur de P. On applique le lemme pour $n = 1$ à chacun de ces segments, ce qui fournit n chemins combinatoires $c_1, \ldots, c_n$, les extrémités de c_i étant x_i et x_{i+1}, seuls sommets de P appartenant à la même composante connexe de $\mathcal{O}$ que x'_i et x'_{i+1} respectivement. Ainsi, les c_i se raccordent pour former un chemin c de x à y. De plus, la suite des murs traversés par c_i est sans répétition. En effet, d'après le lemme 4.7 b), l'ensemble des murs traversés est l'ensemble des murs séparant x_i de x_{i+1}, ou encore l'ensemble des murs séparant x'_i de x'_{i+1}, c'est-à-dire précisément l'ensemble des murs passant par z_i. Un mur étant convexe, il ne peut contenir z_i et z_j pour $i \neq j$ (sinon, il contiendrait tous les points entre z_i et z_j). Ceci achève de prouver que la suite des murs traversés par c est sans répétition. Enfin, $c \subset V(\gamma_1) \cup \ldots \cup V(\gamma_n) \subset V(\gamma)$.

Considérons donc une géodésique γ entre deux points x' et y' n'appartenant à aucun mur, de sorte que γ quitte $\mathcal{O}$ en un seul point z.

Juste avant z (resp. juste après z), la géodésique γ est dans l'intérieur d'une unique cellule C_- (resp. C_+). Les points de γ avant (resp. après) z sont dans une même composante connexe de $\mathcal{O}$, celle de x (resp. de y). Donc $x \in C_-$ et $y \in C_+$. En revanche, x' n'est pas nécessairement dans C_- (ni y' dans C_+).

Si C désigne la plus petite cellule contenant z, on a $C \subset C_-$ (resp: $C \subset C_+$), mais pas nécessairement égalité. Cependant, nous allons montrer que $x \in C$ et $y \in C$ (même lorsque C est une face stricte de C_- ou C_+).

Raisonnons par récurrence sur $dim(C_-) - dim(C)$. Si ce nombre est nul, il n'y a rien à prouver. Sinon C est contenu dans le bord de C_-, et nous pouvons projeter radialement à partir du centre métrique de C_- sur ∂C_- la partie de γ contenue dans C_-. Nous obtenons une géodésique par morceaux γ_z de ∂C_- aboutissant à z. Mis à part z, aucun point de γ_z n'est sur un mur de C_-, sinon, par convexité des murs, le point de γ correspondant serait sur le même mur. La partie de γ_z juste avant z (notée γ_z^-) est une géodésique aboutissant à z dans une face stricte de C_-: on peut lui appliquer l'hypothèse de récurrence, assurant que l'unique sommet x_z de P contenu dans la composante connexe de $\mathcal{O}$ contenant γ_z^- est un sommet de C. D'autre part, un point de $\gamma \cap C_-$ (différent de z) et sa projection sur γ_z ne sont séparés par aucun mur (par convexité, un tel mur, qui passe par le centre métrique de C_-, devrait contenir le point de γ). Ce qui prouve que $x = x_z$ et achève la récurrence.

Les deux sommets x et y appartenant à une même cellule C (rencontrant γ en z, et engendrée par ce point), nous pouvons considérer une géodésique c du 1–squelette de C entre x et y. Nous avons déjà $c \subset V(\gamma)$. Il reste à prouver que la suite des murs de P traversés par c est sans répétition. Raisonnons par l'absurde: si c'est le cas, il existe deux arêtes a et b de c définissant un même mur M de C, et dont les milieux sont les extrémités d'une composante connexe c_0 de $c - M$. Alors, en remplaçant c_0 par $\sigma_{a,C}(c_0)$, on obtient un chemin du 1–squelette de C de même longueur et mêmes extrémités que c, mais avec deux allers-retours dans les arêtes a et b. Ceci contredit le fait que c est géodésique. □

Le corollaire suivant nous permet d'identifier par la suite le 1–squelette de P au graphe de l'espace à murs $(X_P, \mathcal{M}_P)$.

Corollaire 4.10 *Deux sommets de P sont liés par une arête de P si et seulement s'ils sont liés dans $\mathcal{G}(X_P, M_P)$.*

Preuve La condition est bien sûr nécessaire. Réciproquement, soient x et y deux sommets de P à distance combinatoire $n > 1$. Il s'agit de montrer que x et y ne sont pas liés dans $\mathcal{G}(X, M)$, autrement dit qu'il existe un sommet z de P entre x et y (au sens des murs). Considérons la géodésique de P entre x et y. Appliquons-lui le lemme 4.9. Nous trouvons un chemin c de $\mathcal{G}$ entre x et y, tel que la suite des murs traversés par c est sans répétition. En particulier, d'après le lemme 4.7 b), le chemin c est géodésique. Comme $n > 1$, le chemin c contient un point z différent de ses extrémités, qui découpe c en deux sous-chemins c_- et c_+. Si c est constitué des arêtes $a_1, \dots, a_n$, avec

$z = a_i \cap a_{i+1}, i < n$, on obtient, grâce au lemme 4.7 b):

$$\mathcal{M}(x,y) = \{M(a_1), \ldots, M(a_i), M(a_{i+1}), \cdots, M(a_n)\},$$
$$\mathcal{M}(x,z) = \{M(a_1), \ldots, M(a_i)\}$$
$$\mathcal{M}(z,y) = \{M(a_{i+1}), \ldots, M(a_n)\}.$$

Donc $\mathcal{M}(x,y)$ est bien l'union (disjointe) de $\mathcal{M}(x,z)$ et $\mathcal{M}(z,y)$: le point z est entre x et y dans $(X,\mathcal{M})$. □

Le résultat suivant est analogue à celui des complexes de Coxeter (voir [27]) et des complexes cubiques (voir [28]).

Proposition 4.11 *Un chemin combinatoire du* 1*–squelette est une géodésique si et seulement si la suite des murs qu'il traverse est sans répétition.*

Preuve Compte tenu du lemme 4.7 b), il ne reste que le sens "seulement si" à démontrer. Commençons par un analogue combinatoire de la convexité des murs de P.

Lemme 4.12 *Soient* M *un mur de* P, $V(M)$ *la réunion des cellules touchant* M, x *et* y *deux sommets de* $V(M)$. *Alors il existe une géodésique de* $\mathcal{G}$ *d'extrémités* x *et* y *contenue dans* $V(M)$.

Preuve D'abord, d'après les hypothèses de finitude sur les types d'isométrie des cellules de P, il existe un $\varepsilon > 0$ tel que toute cellule de P passant à distance inférieure ou égale à ε de M coupe M.

Soit alors C_x une cellule de P contenant x et touchant M. Le centre métrique $\hat{C}_x$ de C_x est dans M, mais le segment de x à $\hat{C}_x$ ne touche aucun mur de P entre ses extrémités (sinon x serait dans ce mur). Nous pouvons donc trouver sur ce segment un point x' distinct de $\hat{C}_x$, mais ε proche de celui-ci, donc ε proche de M. Il faut noter que x et x' sont dans la même composante connexe de $\mathcal{O}$. De même, il existe un point y' n'appartenant à aucun mur, dans la même composante connexe de $\mathcal{O}$ que y, et ε proche de M. Par convexité (de l'espace P et de M dans P), la géodésique γ de x' à y' reste à distance inférieure ou égale à ε de M. Par définition de ε, cela entraîne que $V(\gamma) \subset V(M)$. Donc la géodésique de $\mathcal{G}$ fournie par le lemme 2 entre x et y reste dans $V(M)$. □

Pour montrer la proposition, considérons un chemin c qui traverse (au moins) deux fois un mur M de P, et prouvons que c n'est pas géodésique. Nous pouvons trouver un sous-chemin c_0 de c qui ne traverse pas M, mais dont

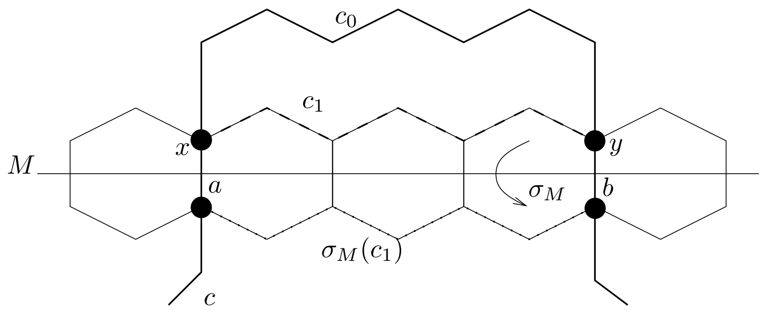


Figure 4: Comment raccourcir les chemins par réflexion

les extrémités sont des sommets x et y d'arêtes a et b transverses à M et contenues dans c.

En appliquant le lemme 4.12, nous remplaçons c_0 par une géodésique c_1 de $\mathcal{G}$ contenue dans $V(M)$ et d'extrémités x et y. Le chemin c' ainsi obtenu a les mêmes extrémités que c, il n'est pas plus long, et il contient comme sous-chemin (a, c_1, b). Or $V(M)$ possède une réflexion σ_M par rapport à M: le chemin $\sigma_M(c_1)$ a les mêmes extrémités que (a, c_1, b), mais il est plus court de deux unités. Ceci prouve que ni c', ni a fortiori c, ne sont géodésiques. □

Compte tenu de la proposition 4.11, la preuve du théorème 1.5 est exactement la même que celle du théorème B de [26].

4.4 Hyperbolicité de l'espace à murs d'un complexe polyédral pair

Soit P un complexe polyédral pair CAT(0), dont la métrique est hyperbolique au sens de Gromov (par exemple, P est CAT(-1)). Comme P n'a qu'un nombre fini de types d'isométrie de cellules, le diamètre des cellules est uniformément majoré. Donc l'inclusion du 1–squelette $\mathcal{G}$ dans P est une quasi-isométrie (quasi-surjective), et $\mathcal{G}$ est hyperbolique.

Nous allons montrer que la condition (H) est remplie dans $(X_P, \mathcal{M}_P)$, en établissant son analogue dans P. Comme d'habitude, nous notons $\overline{P}$ le compactifié de Gromov de P (donc $\overline{P} = P \cup \partial P$), et si E est une partie de P, nous notons $\overline{E}$ son adhérence dans $\overline{P}$. Compte tenu des lemmes 4.5 et 4.4, le premier lemme suivant est clair.

Lemme 4.13 *Soient M un mur de P, x un point de M et $p_x\colon \overline{P} \to lk(x,P)$ la projection qui à un rayon de P d'origine x associe la direction qu'il définit en partant de x. Alors $\overline{M}$ sépare $\overline{P}$ en deux composantes connexes, images réciproques par p_x des deux composantes connexes de $lk(x,P) - lk(x,M)$.* □

Lemme 4.14 *Il existe une constante $D > 0$ telle que deux points de P à distance supérieure ou égale à D sont séparés par au moins un mur de P.*

Preuve Puisque P n'a qu'un nombre fini de types d'isométrie de cellules, il existe un entier N bornant le nombre de murs susceptibles de traverser une cellule donnée de P. Soient x et y deux points quelconques de P, et considérons deux sommets x_0 et y_0 contenus dans une même cellule que x et y respectivement. Le nombre des murs séparant x_0 de x ou y_0 de y est inférieur à $2N$. D'autre part, d'après l'étude de la distance combinatoire sur $\mathcal{G}$, nous savons que le nombre de murs séparant x_0 de y_0 vaut la distance entre x_0 et y_0 dans $\mathcal{G}$. Cette distance tend vers l'infini avec la distance dans P entre x et y, par quasi-isométrie entre P et X, et puisque le diamètre des cellules est uniformément borné. En particulier, il existe un nombre $D > 0$ tel que, si $d_P(x,y) > D$, alors x_0 et y_0 sont séparés par au moins $2N+1$ murs de P. L'un de ces murs ne sépare ni x_0 de x, ni y_0 de y. Donc il sépare x de y. □

Si x_0 est un point base de P, ξ un point de ∂P et r_0 l'unique rayon géodésique de P joignant x_0 à ξ, nous notons $\mathcal{M}(r_0)$ l'ensemble des murs M de P tels que $\overline{M}$ sépare x_0 de ξ.

Lemme 4.15 *Pour tout rayon géodésique r de P, l'ensemble $\mathcal{M}(r)$ est infini.*

Preuve Considérons la suite de points $(x_k)_{k\geq 0}$ du rayon r définie par: x_0 est l'origine de r, et x_k est le point de r à distance kD de x_0 — où D est la constante du lemme 4.14 précédent. Il existe donc pour $k > 0$ un mur M_k séparant x_{k-1} de x_k. Pour $k < \ell$, on a nécessairement $M_k \neq M_\ell$ (sinon, par convexité, ce mur contiendrait les points x_{k+1} et $x_{\ell-1}$).

Le mur M_k et le point ξ ne sont pas adhérents. En effet, si m_k désigne le point d'intersection de M_k avec le sous-segment de r entre x_{k-1} de x_k, la projection de ξ dans $lk(m_k,P)$ correspond à la géodésique $[m_k,x_k]$, non tangente à M. Le lemme 4.13 entraîne bien que $\xi \notin \overline{M}$. La portion de r de x_0 à x_{k-1} ne coupe pas M_k (sinon, par convexité, M_k contiendrait x_{k-1}). De même, la portion de r de x_k à l'infini ne coupe pas M_k. Mais M_k sépare x_{k-1} de x_k. Donc $M_k \in \mathcal{M}(r)$. □

Soient M un mur de P et ξ un point de ∂P non adhérent à M dans $\overline{P}$. Nous noterons $V(\xi, M)$ la composante connexe de $\overline{P} - \overline{M}$ contenant ξ.

Proposition 4.16 *La famille $(V(M,\xi))_{M\in\mathcal{M}(r_0)}$ est une base de voisinages de ξ dans $\overline{P}$.*

Preuve Remarquons tout d'abord que $V(M,\xi)$ est bien un voisinage de ξ. Pour montrer que la famille est une base de voisinages, raisonnons par l'absurde. Par définition de la topologie de $\mathcal{G}\cup\partial\mathcal{G}$, supposons que les distances $d_P(x_0, V(M,\xi))$ restent bornées pour $M\in\mathcal{M}(r_0)$.

En fait, x_0 n'appartient à aucun des voisinages $V(M,\xi)$ pour $M\in\mathcal{M}(r_0)$. Donc $d_P(x_0, V(M,\xi))$ est atteinte sur le bord de $V(M,\xi)$, c'est-à-dire sur M. Nous sommes donc en train de supposer que tous les murs de $\mathcal{M}(r)$ rencontrent une certaine boule fermée B de centre x_0 et de rayon R.

Si P est supposé localement compact, nous obtenons immédiatement une contradiction entre la locale finitude de l'ensemble des murs et le fait que $\mathcal{M}(r)$ est infini.

Donnons un raisonnement général, où l'on ne suppose plus les links de sommet de P compacts. Dans ce cas les boules de P de rayons trop grands peuvent rencontrer une infinité de murs. Cependant, par finitude du nombre de types d'isométrie de cellules de P, il existe un $\varepsilon_0 > 0$ (qu'on peut choisir strictement inférieur à R) et un entier $N_0 > 0$ tels que toute ε_0–boule fermée de P rencontre un nombre de murs strictement inférieur à N_0.

Pour un entier $N\geq N_0$, posons $t_N = ND(\frac{R}{\varepsilon_0}+1)$ et $s_N = ND\frac{R}{\varepsilon_0}$ (le nombre D est celui qui apparaît dans le lemme 4.14). Appelons x_N (resp. y_N) le point du rayon r à distance s_N (resp. t_N) de l'origine x_0. Montrons tout d'abord que toute géodésique γ de P joignant un point u de la boule B à un point v de r entre x_N et y_N passe par la ε_0–boule fermée de P de centre x_N.

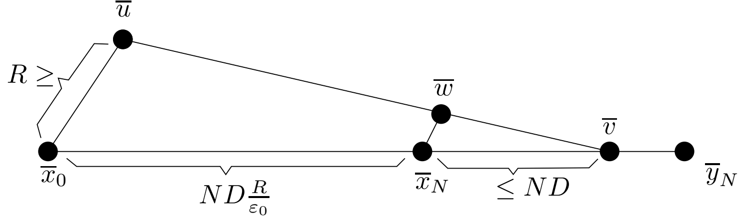


Figure 5: Triangle de comparaison

En effet, considérons le triangle géodésique de P dont les sommets sont x_0 et les extrémités u, v de γ. Soient $\overline{x_0}, \overline{u}, \overline{v}$ les sommets correspondants d'un triangle euclidien de comparaison. Si $\overline{x_N} \in [\overline{x_0}, \overline{v}]$ est le point correspondant à x_N, alors:

$$d(\overline{x_0}, \overline{u}) \leq R, d(\overline{x_0}, \overline{v}) \geq ND\frac{R}{\varepsilon_0}, d(\overline{x_N}, \overline{v}) \leq ND.$$

Soit $\overline{w}$ le point de $[\overline{u}, \overline{v}]$ situé sur la parallèle au côté $[\overline{x_0}, \overline{u}]$ passant par $\overline{x_N}$. Alors par le théorème de Thalès, il vient

$$\frac{d(\overline{w}, \overline{x_N})}{ND} \leq \frac{R}{ND\frac{R}{\varepsilon_0}}.$$

Par l'inégalité CAT(0), la distance de x_N au point w de γ correspondant à $\overline{w}$ est donc inférieure à ε_0.

Pour achever la démonstration de la proposition, découpons le sous-segment de r entre x_N et y_N en N intervalles de longueur D. Par le lemme 4.14, on trouve N murs deux à deux distincts séparant les extrémités de ces intervalles. Ces N murs sont dans $\mathcal{M}(r_0)$ (voir preuve du lemme 4.15). Ils passent par un point du sous-segment de r entre x_N et y_N, et d'autre part ils coupent la boule B par hypothèse. Par convexité des murs et ce qui précède, chacun de ces murs coupe la ε_0–boule fermée de P de centre x_N. Ainsi cette boule est coupée par N murs, avec $N \geq N_0$, en contradiction avec les définitions de ε_0 et N_0. □

L'image réciproque par l'inclusion canonique de $\mathcal{G}$ dans P est une quasi-isométrie, se prolongeant en un homéomorphisme entre les bords. On obtient ainsi un plongement $\overline{i}$: $\overline{\mathcal{G}} \to \overline{P}$. De plus l'image réciproque par $\overline{i}$ d'un voisinage d'un point de ∂P est un voisinage du point correspondant sur $\partial\mathcal{G}$. La proposition précédente entraîne donc que $(X, \mathcal{M})$ vérifie l'axiome (H).

Nous résumons les résultats 4.6, 4.10, 4.16 dans l'énoncé suivant.

Théorème 4.17 *Soit P un complexe polyédral pair* CAT(0), *hyperbolique au sens de Gromov. Alors $(X_P, \mathcal{M}_P)$ est un espace à murs hyperbolique, dont le graphe associé est le 1–squelette de P.*

Un mur M de P est dit *propre* si $\partial P \setminus \partial A$ est non vide pour chacune des composantes connexes A de $P \setminus M$. Ceci équivaut au fait que le mur correspondant de l'espace à murs $(X_P, \mathcal{M}_P)$ est propre.

Lemme 4.18 *Supposons que chaque arête de P soit contenue dans une droite géodésique. Alors tout mur de P est propre.*

Preuve Soit M un mur transverse à une arête d, et A, B les deux composantes connexes de $P \setminus M$. Soit D une droite géodésique contenant d, et a, b l'extrémité du rayon géodésique $D \cap A, D \cap B$ respectivement. Alors puisque P est CAT(0) et que l'angle entre M et d est droit au point d'intersection, le point a n'appartient pas à ∂B, ni b à ∂A. Donc M est propre. □

5 Groupes d'automorphismes d'un complexe polyédral pair

Nous fixons P un complexe polyédral pair CAT(0). Nous notons $(X, \mathcal{M}) = (X_P, M_P)$ son espace à murs associé et $\mathcal{G}$ le 1–squelette de P.

5.1 Automorphismes de l'espace à murs d'un complexe polyédral pair

Le but de cette section est de montrer que le groupe des automorphismes de P et celui de $(X, \mathcal{M})$ coïncident.

Si f est un automorphisme isométrique de P, C une cellule de P et a une arête de C, alors $f(M(a, C)) = M(f(a), f(C))$. Aussi, tout automorphisme isométrique de P agit sur l'ensemble des murs de P. Plus généralement, un isomorphisme (non nécessairement isométrique) entre deux cellules paires préserve le parallélisme entre arêtes. En effet, deux arêtes a et b d'une cellule paire C sont parallèles si et seulement s'il existe une géodésique combinatoire γ du 1–squelette de C joignant une extrémité de a à une extrémité de b, de sorte que a suivie de γ, ainsi que γ suivie de b, soit encore géodésique, mais (a, γ, b) n'est plus géodésique. D'autre part, deux sommets x et y sont du même côté d'un mur M si et seulement si une géodésique de x à y ne contient pas d'arête transverse à M.

Ainsi, le parallélisme des arêtes est une notion ne faisant appel qu'à la combinatoire de C, et même seulement de son 1–squelette. Si f est un isomorphisme (polyédral) d'une cellule paire C sur une autre cellule paire C', et si M est un mur de C, alors les arêtes de C' images par f des arêtes de C transverses à M sont toutes transverses à un même mur de C', qu'on notera $f(M)$. Et deux sommets x et y de C sont du même côté de M si et seulement si $f(x)$ et $f(y)$ sont du même côté de $f(M)$.

Les résultats précédents restent valables pour P tout entier. Il y a donc un morphisme canonique (d'ailleurs clairement injectif) du groupe $\mathrm{Aut}(P)$ des automorphismes (polyèdraux) de P dans $\mathrm{Aut}(X, \mathcal{M})$.

Théorème 5.1 *Soit P un complexe polyédral pair* $\mathrm{CAT}(0)$. *Alors le morphisme de* $\mathrm{Aut}(P)$ *dans* $\mathrm{Aut}(X_P, \mathcal{M}_P)$ *ci-dessus est un isomorphisme.*

Preuve Si $\mathcal{G}(X,\mathcal{M})$ est le graphe associé à $(X,\mathcal{M})$, alors nous avons défini un morphisme injectif $\mathrm{Aut}(X,\mathcal{M}) \to \mathrm{Aut}\ \mathcal{G}(X,\mathcal{M})$. Comme $\mathcal{G}(X,\mathcal{M})$ s'identifie avec le 1–squelette combinatoire $\mathcal{G}$ de P, si ρ: $\mathrm{Aut}\ P \to \mathrm{Aut}\ \mathcal{G}$ est l'application de restriction d'un automorphisme de P à son 1–squelette, alors le diagramme suivant est commutatif:

$$\begin{array}{ccccc} & & \mathrm{Aut}\ (X,\mathcal{M}) & & \\ & \nearrow & & \searrow & \\ \mathrm{Aut}\ P & & \xrightarrow{\rho} & & \mathrm{Aut}\ \mathcal{G} \end{array}$$

Pour établir que tous ces morphismes injectifs sont des isomorphismes, il suffit de montrer que ρ est surjective, i.e. que l'on peut construire un automorphisme (polyédral) de P à partir d'un automorphisme de son 1–squelette $\mathcal{G}$.

Lemme 5.2 *Soient C une cellule de P et a une arête de P telle que l'intersection $a \cap C$ est réduite à un sommet x_0. Alors le mur transverse à a ne coupe pas C.*

Preuve Supposons, par l'absurde, qu'il existe une cellule C, une arête a et un sommet x_0 tels que $a \cap C = \{x_0\}$ et $M = M(a)$ coupe C. Soient y_0 le sommet de C symétrique de x_0 par rapport à M, et p le point où la géodésique qui joint x_0 à y_0 (dans C) coupe M. Alors p est le point de $M \cap C$ le plus proche de x_0.

En fait, pour toute cellule D dont C est une face, p est encore le point de $M \cap D$ le plus proche de x_0: donc p est un minimum local (strict) pour la fonction qui à un point q de M associe sa distance à x_0 dans P. Mais il en va de même pour le point p', milieu de l'arête a. Or $p' \neq p$, puisque $a \not\subset C$, ce qui donne deux minimaux locaux sur M à la fonction "distance à x_0", en contradiction avec la convexité de cette fonction et celle de M dans P. □

Corollaire 5.3 *Le 1–squelette d'une cellule C est convexe dans $\mathcal{G}$, le 1–squelette de P.*

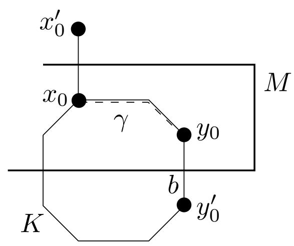
Preuve Soient x_0, y_0 deux sommets de C, et γ un chemin de $\mathcal{G}$ entre x_0 et y_0, qui sort de C. D'après le lemme précédent, la suite des murs traversés par γ contient un mur ne coupant pas C. Or l'ensemble des murs qui sépare x_0, y_0 est contenu dans l'ensemble des murs coupant C. Donc, d'après la proposition 4.11, γ ne peut être géodésique. □

Notons $\mathcal{E}$ l'ensemble des cellules de P et $\mathcal{F}$ l'ensemble des sous-graphes convexes de $\mathcal{G}$ isomorphes au graphe de Cayley d'un système de Coxeter fini.

Comme le 1–squelette d'une cellule paire est le graphe de Cayley d'un système de Coxeter fini (Proposition 4.1), le corollaire ci-dessus montre que l'application $i\colon C \mapsto C \cap \mathcal{G}$ est une application (injective) de $\mathcal{E}$ dans $\mathcal{F}$. Pour retrouver les cellules de P à partir de son 1–squelette, nous allons montrer que $i(\mathcal{E}) = \mathcal{F}$.

Lemme 5.4 *Soient K un élément de $\mathcal{F}$ et a une arête de P telle que l'intersection $a \cap K$ est réduite à un sommet x_0. Alors le mur M transverse à l'arête a ne recoupe pas K.*

Preuve Raisonnons par l'absurde. Soit b une arête de K transverse à M. Notons y_0 l'extrémité de b du même côté de M que x_0, puis x_0' et y_0' les images de x_0 et y_0 par la réflexion σ_M du voisinage $V(M)$ de M. Comme le 1–squelette de $V(M)$ est géodésique dans le 1–squelette de P (voir lemme 4.12), il existe une géodésique γ de $\mathcal{G}$ entre x_0 et y_0 contenue dans $V(M)$. Mais comme K est convexe dans $\mathcal{G}$, on a $\gamma \subset K$.



Puisque x_0 et y_0 ne sont pas séparés par M, le chemin γ ne coupe pas le mur M. Donc (γ, b) est une géodésique de $\mathcal{G}$ entre x_0 et y_0' (voir lemme 4.11). Le chemin $(a, \sigma_M(\gamma))$ a les mêmes extrémités et la même longueur, mais il passe par $x_0' \notin K$: ceci contredit la convexité de K dans $\mathcal{G}$. □

Lemme 5.5 *Soient K un élément de $\mathcal{F}$ et M un mur coupant une arête a de K. Alors chaque arête de K touchant a est contenue dans $V(M)$, et l'ensemble de ces arêtes est invariant par σ_M.*

Preuve Soient x_0 et y_0 les extrémités de a, et b une arête de K distincte de a, contenant y_0. Il s'agit de montrer que b est dans $V(M)$, et que $\sigma_M(b)$ est dans K.

Soit (W, S) le système de Coxeter de graphe de Cayley $\mathcal{G}(W, S)$ isomorphe à K. Puisque W est transitif sur les sommets de $\mathcal{G}(W, S)$, on peut trouver un isomorphisme φ de $\mathcal{G}(W,S)$ sur K envoyant 1 sur x_0. Soient s et w les éléments de W dont l'image par φ sont y_0 et z_0, la deuxième extrémité de b. D'abord, $s \in S$, puisqu'il est lié à 1 dans $\mathcal{G}(W,S)$ par l'arête $\varphi^{-1}(a)$. Ensuite, il existe $t \neq s$, $t \in S$ tel que $w = st$. Considérons $\mathcal{G}_{s,t}$, le sous-graphe plein de $\mathcal{G}(W,S)$ dont les sommets sont $1, s, st, sts, \ldots, 1$. C'est un graphe homéomorphe à un cercle, contenant $2m_{s,t}$ arêtes, où $m_{s,t}$ désigne l'ordre du produit st dans W. Ce sous-graphe est une maille de $\mathcal{G}(W,S)$, au sens suivant: une *maille* est un circuit de longeur $2m$ totalement géodésique dans $\mathcal{G}(W,S)$, tel que si deux de ses sommets sont à distance strictement inférieure à m, il y a une unique géodésique de $\mathcal{G}(W,S)$ les joignant (alors nécessairement contenue dans le circuit).

L'image de $\mathcal{G}_{s,t}$ dans K est une maille $K_{s,t}$ de K; par convexité de K dans $\mathcal{G}$, c'est aussi une maille de $\mathcal{G}$. L'arête a' de $K_{s,t}$ la plus éloignée de a est caractérisée par l'existence d'un sous-segment c de $K_{s,t}$, tel que c joint y_0 à une extrémité y_0' de a', (a,c) et (c,a') sont géodésiques, mais (a,c,a') ne l'est pas.

Des trois dernières propriétés et de la proposition 4.11, il résulte que $M(a) = M(a')$.

D'après le lemme 4.12, il existe une géodésique de y_0 à y_0' contenue dans $V(M)$. Mais comme $K_{s,t}$ est une maille, cette géodésique est c. Alors $\sigma_M(c)$ est une géodésique entre deux points de $K_{s,t}$ à distance strictement inférieure à $m_{s,t}$, donc $\sigma_M(c) \subset K_{s,t}$. □

Lemme 5.6 *Soient K un élément de $\mathcal{F}$ et C une cellule de P dont le 1–squelette contient un sommet x_0 de K tel que $St(x_0,K) = St(x_0,i(C))$. Alors $K = i(C)$.*

Preuve On peut supposer la dimension de C au moins égale à deux, sinon il n'y a rien à montrer. Par connexité de C, il suffit de montrer que K est un ouvert de $i(C)$. Par connexité de K, il suffit de montrer que si x_0 est un sommet de K tel que $St(x_0,K) = St(x_0,i(C))$, alors pour tout voisin y_0 de x_0 dans K, on a encore $St(y_0,K) = St(y_0,i(C))$.

Soit a l'arête de K d'origine x_0 et d'extrémité y_0, et M le mur transverse à a. Comme σ_M préserve $i(C)$ et l'ensemble des arêtes de K touchant a (d'après le lemme 5.5), on a:

$$St(y_0,K) = St(\sigma_M(x_0),K) = \sigma_M(St(x_0,K)) = \sigma_M(St(x_0,i(C))) = St(y_0,i(C)).$$ □

Proposition 5.7 *L'application* $i\colon \mathcal{E} \to \mathcal{F}$ *est surjective.*

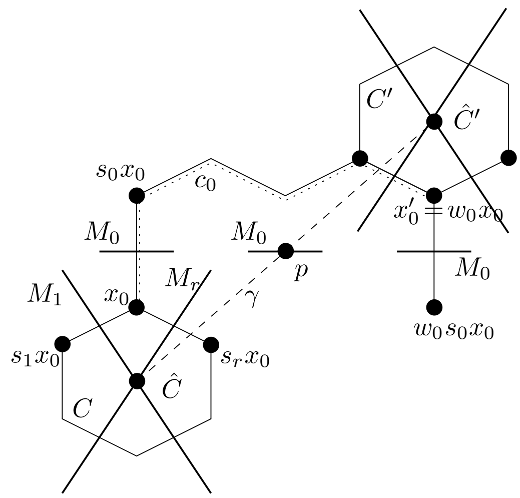
Preuve On raisonne par récurrence sur le rang du système de Coxeter dont K est le graphe de Cayley (cela correspond au degré du graphe régulier K). Il n'y a rien à dire en rang 1.

Soit $K \in \mathcal{F}$ de rang $r+1$ supérieur ou égal à 2. Considérons un sommet x_0 de K, et soient $a_0, a_1, \ldots, a_r$ les arêtes issues de x_0; nous noterons M_i le mur transverse à l'arête a_i. Alors il existe un système de Coxeter fini $(W, S = \{s_0, s_1, \ldots, s_r\})$ et un isomorphisme de son graphe de Cayley $\mathcal{G}(W,S)$ sur K envoyant 1 sur x_0 et l'arête issue de 1 préservée par s_i sur a_i. Considérons maintenant V, le sous-groupe spécial de (W,S) engendré par $T = \{s_1, \ldots, s_r\}$. Il y a une unique copie de son graphe de Cayley contenue dans le graphe de Cayley de (W,S) et passant par 1; à ce sous-graphe correspond un sous-graphe L de K.

Un résultat classique sur les sous-groupes spéciaux (cf. [5]) entraîne que L est convexe dans $\mathcal{G}$. On peut donc appliquer l'hypothèse de récurrence à L, et trouver une cellule C de P dont L est le 1–squelette.

Comme (W,S) est fini, il possède un unique élément w_0 de longueur maximale: soit x_0' le sommet correspondant de K.

Fait 1 Notons d'abord que les arêtes de K issues de x_0' sont traversées par les murs M_i, qui de plus séparent x_0' de x_0.



Preuve Pour tout s_i de S, l'élément $s_i w_0$ doit être lié à w_0 dans $\mathcal{G}(W,S)$, ce qui signifie qu'il existe un $s_j \in S$ tel que $s_i w_0 = w_0 s_j$. Soit c_i une géodésique de $\mathcal{G}(W,S)$ de 1 à w_0 commençant par l'arête de 1 à s_i: la suite des murs de $\mathcal{G}(W,S)$ traversés par c_i est sans répétition. Alors le chemin γ_i formé de c_i suivi de l'arête de w_0 à $s_i.w_0 = w_0.s_j$ n'est pas géodésique, car le premier mur qu'elle traverse est $M(s_i) = M(w_0.s_j.w_0^{-1})$, donc égal au dernier. On en déduit que le chemin c_i' tel que l'arête de 1 à s_i suivie de c_i' égale γ_i est géodésique.

En prenant les images de ces trois chemins dans K, en utilisant la convexité de K dans $\mathcal{G}$ et la caractérisation des géodésiques combinatoires par la suite des murs traversés, on voit que le mur M_i coupe une arête issue de x_0', et sépare x_0 de x_0'. □

Comme ci-dessus, il y a une unique copie convexe de L dans K passant par x_0', coupée par les murs $M_1, \dots M_r$: nous la noterons L' et C' sera la cellule de P dont le 1–squelette est L'.

Les centres métriques des cellules C et C' sont des points $\hat{C}$ et $\hat{C}'$ de $M_1 \cap \dots \cap M_r$; par convexité, la géodésique γ qui les joint est aussi dans cette intersection. D'autre part, la géodésique joignant $\hat{C}$ à x_0 ne coupe que les murs de C, donc pas M_0 (d'après le lemme 5.2). Un résultat analogue étant vrai pour x_0', et M_0 séparant x_0 de x_0', la géodésique γ doit couper M_0. Comme $\gamma \not\subset M_0$, l'intersection de γ avec M_0 ne contient qu'un point p.

Soit D la cellule de P engendrée par γ juste après $\hat{C}$; comme $\gamma \cap C = \hat{C}$, la cellule D contient C comme face stricte.

Fait 2 Le point p appartient à D.

Preuve Par l'absurde, supposons que p n'est pas dans D. Alors γ ressort de D par un point q de son bord; ce point est dans $M_1 \cap \dots \cap M_r$. Comme γ reste dans l'intérieur de D entre $\hat{C}$ et q, ces deux points ne peuvent être sur une même face du bord de D. Soit F la face stricte de D engendrée par q; cette cellule paire est coupée par les murs $M_i, 1 \leq i \leq r$, donc invariante par les réflexions σ_{M_i}, tout comme C.

Montrons que F est disjointe de C. Si F contenait un sommet de C, elle contiendrait toutes ses images par le groupe d'isométrie de D engendrée par les réflexions $\sigma_{M_i}, 1 \leq i \leq r$. Mais ce groupe est (simplement) transitif sur l'ensemble des sommets de C. Donc F contiendrait tous les sommets de C, autrement dit C elle-même. Mais alors $\hat{C}$ et q seraient dans une même face F du bord de D, ce qui n'est pas.

Considérons une géodésique combinatoire c de $x_0 \in C$ à un sommet de F, de longueur minimale. Par convexité de $i(D)$, on a $c \subset D$. Comme $C \cap F = \emptyset$, la longueur de c est non nulle: donc $c = (b, \ldots)$, où b est une arête de D issue de x_0. Comme σ_{M_i} préserve F, il est évident, par minimalité, que $b \not\subset C$. Donc, d'après le lemme 5.2, $M(b)$ ne peut pas couper C.

Maintenant, le mur $M(b)$ ne peut pas non plus couper F: sinon en appliquant $\sigma_{M(b)}$ au sous-segment de c après b, on trouverait une géodésique de x_0 à un sommet de F, de longueur inférieure à celle de c, en contradiction avec la minimalité de celle-ci.

Il en résulte que $M(b)$ sépare les deux cellules C et F, donc en particulier les deux points $\hat{C}$ et q. Alors $M(b)$ sépare $\hat{C}$ et $\hat{C}'$, C et C', donc x_0 et x'_0.

Le mur $M(b)$ n'est pas le mur M_0: car celui-là coupe la géodésique γ dans D, alors que celui-ci la coupe en p, supposé extérieur à D. Nous nous retrouvons avec un élément K de $\mathcal{F}$ et une arête b de P contenant un sommet de K, mais non contenue dans K, telle que $M(b)$ sépare deux points de K: une contradiction avec le lemme 5.4. Cette absurdité prouve que $p \in D$. □

Puisque l'arête a_0 est issue d'un sommet x_0 de D et que le mur $M_0 = M(a_0)$ recoupe D (en p), le lemme 5.2 entraîne que $a_0 \subset D$. Alors la sous-cellule E de D engendrée par les arêtes $a_0, a_1, \ldots a_r$ vérifie $St(x_0, K) = St(x_0, i(E))$, donc $K = i(E)$ d'après le lemme 5.6. □

Corollaire 5.8 *Le morphisme de restriction de* $\mathrm{Aut}(P)$ *dans* $\mathrm{Aut}(\mathcal{G})$ *est un isomorphisme.*

Preuve Il suffit de montrer la surjectivité. Si φ est un automorphisme de $\mathcal{G}$, définissons un automorphisme $\overline{\varphi}$ de P de la façon suivante. Pour une cellule C de P, considérons l'élément K' de $\mathcal{F}$ défini par $K' = \varphi(i(C))$. D'après la proposition précédente, il existe une (unique) cellule C' dont le 1–squelette est K'. Alors il existe un unique isomorphisme polyédral de C sur C' prolongeant $\varphi|_{i(C)}$.

La collection d'isomorphismes polyédraux locaux $\overline{\varphi}_C$ ainsi obtenue se recolle pour donner l'automorphisme $\overline{\varphi}$. □

Ce corollaire termine la preuve du théorème 5.1. □

5.2 Existence d'automorphisme non trivial fixant strictement un mur propre

Un automorphisme de P *fixe strictement* un mur M de P si et seulement s'il fixe M (point par point) et préserve chacune des deux composantes connexes de $P \setminus M$.

Le but de cette section est de donner des exemples de P dont le groupe $\mathrm{Aut}^+(P)$, sous-groupe de $\mathrm{Aut}(P)$ engendré par les stabilisateurs stricts de murs propres est très gros.

Remarque (1) L'automorphisme f fixe strictement le mur M si et seulement s'il fixe point par point $M \cup a$, où a est une arête transverse à M. Une condition équivalente est que f fixe $V(M)$ point par point. Et un automorphisme de P fixe strictement un mur M si et seulement si l'automorphisme correspondant de $(X_P, \mathcal{M}_P)$ fixe strictement le mur correspondant à M.

(2) Soient P^+ et P^- les adhérences des deux composantes connexes de $P \setminus M$. Alors le sous-groupe de $\mathrm{Aut}(P)$ formé des automorphismes fixant strictement M est le produit direct de $\mathrm{Fix}(P^+)$ et de $\mathrm{Fix}(P^-)$.

Lemme 5.9 *Soit P un complexe polyédral pair* $\mathrm{CAT}(0)$. *Alors son espace à murs (X_P, M_P) vérifie la propriété* (M').

Preuve Soit f un automorphisme de P fixant strictement un mur M et A une des deux moitiés de X définies par M. Soit B une moitié de X telle que $A \cap B$ et $(X \setminus A) \cap B$ sont non vides. Notons N le mur de P dont le mur associé sur X est $(B, X \setminus B)$. Alors on voit que N contient des points séparés par M. Donc, par convexité, $M \cap N$ est non vide. En particulier, il existe une cellule C de P coupée par M et N. Puisque f fixe strictement M, elle vaut l'identité sur C. Donc f fixe une arête transverse à N: f préserve globalement N, ainsi que les deux composantes connexes de $X \setminus N$. □

Nous allons étudier le cas où P est la réalisation géométrique de Davis–Moussong d'un système de Coxeter.

Soit (W, S) un système de Coxeter. Nous noterons $N = N(W, S)$ le nerf fini de (W, S). Nous munissons la première subdivision barycentrique N' de N d'une fonction m, définie sur l'ensemble des milieux $\hat{a}$ des arêtes a de N par la formule: $m(\hat{a})$ est l'ordre du produit st, avec s et t les réflexions de S correspondant aux extrémités de a. Il est alors immédiat que les automorphismes du graphe de Coxeter de (W,S) correspondent aux automorphismes de N' qui proviennent d'un automorphisme de N et préservent la fonction m.

Notons $P = |W|_0$ la réalisation de Davis–Moussong de (W,S). On a $P' = (W \times (x_0 * N'))/\sim$ (voir section 3.2), et nous noterons $[w,x]$ la classe de (w,x). Les sommets de P sont les points $[w,x_0]$ pour $w \in W$. Nous identifierons un point x de $x_0 * N'$ avec son image $[id,x]$ dans P'. En particulier, le link de x_0 dans P' s'identifie avec N'. L'action à gauche de W sur le produit passe au quotient, en une action simplement transitive sur les sommets wx_0 de P. Mais on peut aussi construire, à partir de (W,S), des éléments de $\mathrm{Aut}(P)$ fixant le sommet x_0.

Soit $G(W,S)$ le groupe des automorphismes du diagramme de Coxeter de (W,S). Tout élément f de $G(W,S)$ agit sur N' (en préservant m), donc nous pouvons considérer son prolongement conique à $x_0 * N'$, encore noté f. D'autre part, f induit naturellement un automorphisme du groupe W (permutant S), que nous noterons $\overline{f}$. Alors l'application $(w,x) \mapsto (\overline{f}(w), f(x))$ est compatible avec $\sim$, donc induit un automorphisme $\hat{f}$ de P'. On a $\hat{f}([w,x]) = [\overline{f}(w), f(x)]$, donc $\hat{f}(x_0) = x_0$, et $\hat{f}$ agit sur le link de x_0 comme f sur N'. Enfin, $\hat{f}$ provient d'un automorphisme de P (car f provient d'un automorphisme de N).

Nous obtenons ainsi une représentation fidèle de $G(W,S)$ dans $\mathrm{Aut}(P)$, d'image contenue dans le stabilisateur de x_0. D'après la formule $\hat{f}(w \cdot [w',x]) = \overline{f}(w) \cdot \hat{f}([w',x])$, si $\overline{f}$ fixe point par point un sous-ensemble T de S, alors $\hat{f}$ commute avec l'action sur P du sous-groupe spécial engendré par T.

Définition 5.10 Soient Q un complexe polyédral pair CAT(0), et a une arête de Q. La *facette* de Q transverse à a est la réunion des simplexes de Q' (la première subdivision barycentrique de Q) qui contiennent le milieu de l'arête a, mais aucune de ses extrémités. Nous la noterons $\phi(a)$. Si x_0 est un sommet de Q, le *bloc de centre* x_0 est l'étoile de x_0 dans Q'.

Lemme 5.11 *Le mur transverse à l'arête a est la réunion des facettes $\phi(b)$, avec b parallèle à a. Les deux blocs centrés sur les extrémités d'une arête a ont pour intersection la facette $\phi(a)$.*

Preuve Vérifions d'abord que $\phi(a) \subset M(a)$. Soit Δ un simplexe de $\phi(a)$. Considérons la plus petite cellule C contenant Δ: les sommets de Δ sont les centres métriques de certaines faces de C contenant l'arête a. Donc chacun de ces sommets est invariant par $\sigma(a,C)$: autrement dit $\Delta \subset M(a,C)$.

Pour achever de montrer la première assertion, il suffit de prouver que si C est un polyèdre pair et a une arête de C, alors $M(a,C)$ est contenu dans l'union des facettes $\phi(b)$, avec b parallèle à a dans C. On raisonne par récurrence sur $dim(C)$, la propriété étant évidente en dimension 1.

Soit x un point de $M(a,C)$. Si $x = \hat{C}$ (le centre métrique de C), alors x est dans toutes les facettes de C, en particulier dans $\phi(a)$. Si $x \neq \hat{C}$, nous pouvons considérer la géodésique de $\hat{C}$ à x, et la prolonger jusqu'au bord de C, qu'elle touche en un point y. Comme x et $\hat{C}$ sont dans $M(a,C)$, le point y est aussi dans ce mur. Cela signifie que D, la face stricte de C engendrée par y, est coupée par $M(a,C)$. On applique alors l'hypothèse de récurrence à $y \in D$: il existe un simplexe Δ_y de la facette d'une arête b de D telle que $M(b,D) = M(a,C) \cap D$ qui contient y. Alors b est parallèle à a, et x est dans Δ_x, le joint de Δ_y avec $\hat{C}$. Ceci conclut, car Δ_x est dans la facette de b dans C.

Pour la seconde assertion, soit a une arête de Q d'extrémités x_0 et y_0. Un sommet de Q' est joignable aux extrémités de a si et seulement s'il est le centre métrique d'une face C contenant x_0 et y_0. Ceci, par convexité, équivaut à dire que C contient a, autrement dit $\hat{C} \in \phi(a)$. □

Si (W,S) est un système de Coxeter, nous appellerons *facette de* (W,S) *(au sens de Davis–Moussong)* l'étoile dans N' d'un sommet de N; si ce sommet correspond à la réflexion s, nous noterons ϕ_s cette facette. Le système de Coxeter est dit *rigide* si le fixateur dans $G(W,S)$ de toute facette de (W,S) est trivial. Tous les blocs de la réalisation géométrique de Davis–Moussong P sont isomorphes au cône sur N'; l'intersection de deux blocs centrés sur des sommets voisins de P est donc une facette de (W,S).

Théorème 5.12 *Si* (W,S) *est rigide, alors* $\mathrm{Aut}(P)$ *est discret: c'est le produit semi-direct de* W *et de* $G(W,S)$.

Supposons que (W,S) *n'est pas rigide, et que* (W,S) *est hyperbolique au sens de Gromov. Alors, pour tout automorphisme non trivial* f *de* $G(W,S)$ *fixant une facette* ϕ_s*, le mur* M_s *passant par* ϕ_s *est propre. De plus, il existe un automorphisme* φ *de* P*, dont la restriction à l'étoile de* x_0 *dans* P *est* $\hat{f}$*, et qui fixe strictement* M_s*. En particulier,* $\mathrm{Aut}^+(P) \neq \{1\}$*, et* $\mathrm{Aut}(P)$ *est non discret.*

Preuve Supposons d'abord que (W,S) est rigide. Il s'agit de montrer que le stabilisateur de x_0 est $G(W,S)$.

D'abord le fixateur de $St(x_0,P')$ dans $\mathrm{Aut}(P)$ est trivial: car si $F \in \mathrm{Aut}(P)$ fixe l'étoile d'un sommet dans P', alors par rigidité F fixe l'étoile de tout sommet voisin.

Ensuite, si $F \in \mathrm{Aut}(P)$ fixe x_0, il induit un automorphisme du link de x_0 dans P' (isomorphe à N'), provenant d'un automorphisme de N, et préservant la fonction m. En effet, cette fonction a une interprétation géométrique: $m(x)$ est simplement le diamètre combinatoire du bord de la 2–face dont x est le centre. Il existe donc un $f \in G(W,S)$ tel que $\hat{f}$ coïncide avec F sur le bloc de P de centre x_0. D'après la première partie, $\hat{f} = F$.

Lemme 5.13 *Soit (W,S) un système de Coxeter hyperbolique. Alors l'ensemble des $w \in W$ qui agissent trivialement au bord de W est un sous-groupe spécial fini W_F tel que le système $(W_{S\backslash F}, S\backslash F)$ est irréductible, et tout élément de F commute avec tout élément de $S\backslash F$. En particulier, si W est irréductible, alors W agit fidèlement sur son bord.*

Preuve Soit G le sous-groupe de W agissant trivialement sur ∂W : c'est un sous-groupe distingué fini de W (voir [13]). En tant que sous-groupe fini, G est contenu dans un conjugué d'un sous-groupe spécial fini W_T. Mais comme G est distingué, on a $G \subset W_T$, avec toujours G distingué dans W. En prenant l'intersection des sous-groupes spéciaux finis contenant G, on trouve un sous-groupe spécial fini W_F contenant G, et tel que, pour tout $t \in F$, il existe $g \in G$ tel que t apparaisse dans une écriture de longueur minimale de g. Pour s n'appartenant pas à F et $g \in G$, on a $s.g.s \in G \subset W_F$. Donc s commute avec tous les éléments t de F apparaissant dans une écriture géodésique de g. On en déduit que tout élément de F commute avec tout élément de $S \setminus F$.

Il reste à montrer que $(W_{S\setminus F}, S \setminus F)$ est irréductible. Supposons que $S \setminus F = T_1 \cup T_2$, avec $T_1 \cap T_2 = \emptyset$ et tout élément de T_1 commute avec tout élément de T_2. On ne peut avoir W_{T_1} et W_{T_2} infinis, puisque W est hyperbolique et contient $W_{T_1} \times W_{T_2}$. Si par exemple W_{T_1} est fini, il commute à $W_{T_2 \cup F}$, donc agit trivialement au bord: d'où $T_1 \subset W_{T_1} \subset G \subset W_F$, et donc $T_1 = \emptyset$. □

Supposons maintenant que (W,S) est non rigide. Soient $f \in G(W,S)$ et $s \in S$ tels que f est non trivial et f fixe ϕ_s (point par point). Le fait que M_s soit un mur propre de P résulte du lemme précédent. En effet, si M_s n'est pas propre, comme s permute les deux demi-espaces définis par M_s, le bord de M_s est égal à tout le bord de P, donc s agit trivialement sur le bord de P. Par le lemme, s appartient à F, et son étoile est égale à tout le nerf fini de (W,S), ce qui contredit la non-trivialité de f.

Comme f fixe la facette ϕ_s, $\overline{f}$ fixe tous les $t \in S$ tels que $m_{s,t} < \infty$. Donc $\hat{f}$ commute à tout produit de telles réflexions. Comme d'autre part $\hat{f}$ fixe l'arête transverse à M_s passant par x_0, c'est donc que $\hat{f}$ fixe toutes les arêtes de la forme $w.a_s$, avec $w \in W_{T_s}$, où $T_s = \{t \in S \,/\, m_{s,t} < \infty\}$, et a_s est l'arête de P entre x_0 et sx_0.

Soient $\partial\mathcal{H}$ la réunion de ces arêtes, $\mathcal{H}$ la réunion des chemins d'origine x_0 dans le 1–squelette $\mathcal{G}$ de P, qui ne traversent pas $\partial\mathcal{H}$, et $\mathcal{H}^c$ le sous-graphe de $\mathcal{G}$ réunion des arêtes non dans $\mathcal{H}$. La proposition 5.17 de la section suivante dit que $\mathcal{H}^c$ contient le demi-espace A de W défini par M_s et contenant s. L'automorphisme $\hat{f}$ vaut l'identité sur $\partial\mathcal{H}$, et $\mathcal{H} \cap \mathcal{H}^c$ est contenu dans $\partial\mathcal{H}$. Donc on peut définir un automorphisme φ de $\mathcal{G}$ qui coïncide avec l'identité sur $\mathcal{H}^c$, et avec $\hat{f}$ sur $\mathcal{H}$. Comme A est contenu dans $\mathcal{H}^c$, φ vaut l'identité sur le demi-espace A, donc fixe strictement le mur M_s. Enfin, $\hat{f}$ agit sur l'ensemble $\bigcup_{t \in S \setminus \{s\}} a_t$ des arêtes à la fois dans $\mathcal{H}$ et dans l'étoile de x_0, comme f sur $S \setminus \{s\}$. Donc φ, qui coïncide avec $\hat{f}$ sur l'étoile de x_0, est non trivial. □

5.3 Un résultat technique sur les groupes de Coxeter

Soit (W, S) un système de Coxeter, notons 1 son élément neutre et $\mathcal{G} = \mathcal{G}(W, S)$ son graphe de Cayley.

Si $t \in S$ et $T \subset S$, nous noterons a_t l'arête de $\mathcal{G}(W, S)$ entre 1 et t et $\mathcal{G}_T$ le sous-graphe de $\mathcal{G}$ réunion des arêtes reliant deux sommets de $\mathcal{G}$ appartenant au sous-groupe spécial W_T engendré par T. Alors $\mathcal{G}_T$ est isomorphe à $\mathcal{G}(W_T, T)$, et c'est un sous-graphe convexe de $\mathcal{G}$ (voir [5]). On peut aussi voir $\mathcal{G}_T$ comme la réunion des chemins de $\mathcal{G}$ d'origine 1, et dont toutes les arêtes ont un label dans T (i.e. sont de la forme wa_t, avec $t \in T$).

Pour $s \in S$ quelconque, soit T_s la partie de S formée des réflexions t telles que $m_{s,t} < \infty$. Notons alors $\mathcal{H} = \mathcal{H}_s$ la réunion des chemins c de $\mathcal{G}(W, S)$ d'origine 1, et n'empruntant que des arêtes de la forme wa_t, avec $t \neq s$, ou de la forme wa_s, avec $w \notin W_{T_s}$. Introduisons enfin $A = A_s$, l'ensemble des éléments de W séparés de 1 par le mur M_s de s.

Notre but est de montrer que $\mathcal{H}$ et A sont disjoints, ce qui est le résultat voulu dans la preuve du théorème 5.12.

Commençons par donner une description plus constructive de $\mathcal{H}$. Posons

$$\mathcal{K}_1 = \{1\}, \mathcal{H}_1 = \mathcal{G}_{S \setminus \{s\}}, \ldots,$$

$$\mathcal{K}_{n+1} = \bigcup_{w \in \mathcal{H}_n, w \notin W_{T_s}} w.\mathcal{G}_{T_s}, \mathcal{H}_{n+1} = \bigcup_{w \in \mathcal{K}_{n+1}} w.\mathcal{G}_{S \setminus \{s\}}.$$

Alors $\mathcal{H}$ est la réunion croissante des $\mathcal{H}_n$.

Lemme 5.14 *Soient $w \in \mathcal{H}$ et $n = n(w)$ le plus petit indice i tel que $w \in \mathcal{H}_i$. Alors il existe deux suites $v_1^-, \dots, v_{n-1}^-$ et $v_1^+, \dots, v_{n-1}^+$ d'éléments de W appartenant à $\mathcal{H}$, et une suite $M_1, \dots, M_{n-1}$ de murs de (W, S) tels que, pour $1 \leq i < n$,*

- *$n(v_i^-) = i$ et $n(v_i^+) = i + 1$;*
- *v_i^- et v_i^+ sont congrus modulo W_{T_s}, v_{i-1}^+ et v_i^- sont congrus modulo $W_{S\setminus\{s\}}$ (en posant $v_0^+ = 1$), et v_{n-1}^+ est congru à w modulo $W_{S\setminus\{s\}}$;*
- *M_i est transverse à l'arête de type s d'origine v_i^-, et sépare $v_i^-.\mathcal{G}_{S\setminus\{s\}}$ de $v_i^+.\mathcal{G}_{S\setminus\{s\}}$.*

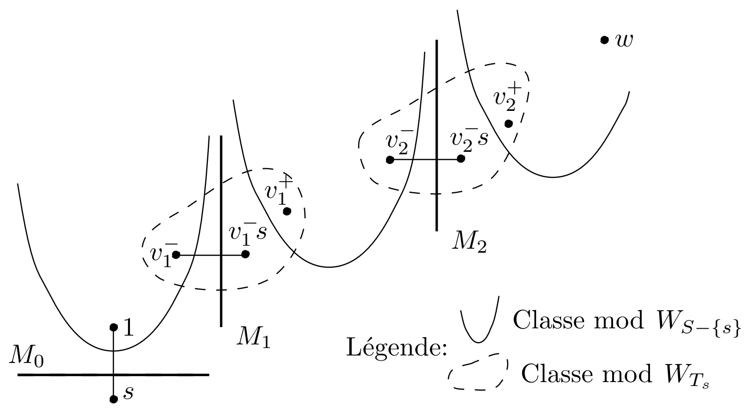


Figure 6: Description constructive de $\mathcal{H}$

Preuve Par récurrence sur n. Si $n = 1$, il n'y a rien à démontrer.

Supposons donc $n > 1$. Comme $w \in \mathcal{H}_n$, il existe un v_{n-1}^+ de $\mathcal{K}_n$ congru à w modulo $W_{S\setminus\{s\}}$; puis il existe un v_{n-1}^- de $\mathcal{H}_{n-1}$ auquel v_{n-1}^+ est congru modulo W_{T_s}. Quitte à multiplier v_{n-1}^- et v_{n-1}^+ par des éléments convenables de $W_{T_s\setminus\{s\}}$ (ce qui ne change ni les classes modulo W_{T_s}, ni les classes modulo $W_{S\setminus\{s\}}$), on peut supposer que $d_W(v_{n-1}^-.W_{T_s\setminus\{s\}}, v_{n-1}^+.W_{T_s\setminus\{s\}}) = d_W(v_{n-1}^-, v_{n-1}^+)$.

Si v_{n-1}^- était dans un $\mathcal{H}_i$ avec $i < n-1$, w serait dans $\mathcal{H}_{n-1}$, en contradiction avec la définition de n. De même, $v_{n-1}^+ \in \mathcal{H}_n \setminus \mathcal{H}_{n-1}$ (en particulier, $v_{n-1}^-.W_{S\setminus\{s\}} \neq v_{n-1}^+.W_{S\setminus\{s\}}$). Donc, si on complète les suites fournies par la récurrence appliquée à v_{n-1}^- à l'aide de v_{n-1}^- et v_{n-1}^+ d'une part, et d'autre part à l'aide du mur M_{n-1} fourni par le lemme suivant d'autre part, on obtient le résultat au rang n. □

Lemme 5.15 *Supposons que $v^-.W_{T_s} = v^+.W_{T_s}$ et $v^-.W_{S\setminus\{s\}} \neq v^+.W_{S\setminus\{s\}}$. Si de plus $d_W(v^-.W_{T_s\setminus\{s\}}, v^+.W_{T_s\setminus\{s\}}) = d_W(v^-, v^+)$, alors le mur transverse à l'arête $v^-.a_s$ issue de v^- sépare $v^-.\mathcal{G}_{S\setminus\{s\}}$ de $v^+.\mathcal{G}_{S\setminus\{s\}}$.*

Preuve Quitte à multiplier par l'inverse de v^-, on peut supposer $v^- = 1$. On a alors $v = v^+ \in W_{T_s} \setminus W_{S\setminus\{s\}}$, et v est l'élément de plus petite longueur dans sa classe modulo $W_{T_s\setminus\{s\}}$ (cette longueur est non nulle, sinon v serait dans $W_{T_s\setminus\{s\}}$, donc dans $W_{S\setminus\{s\}}$). En particulier, toute géodésique de 1 à v passe par a_s, et le mur M_s transverse à a_s sépare 1 de v.

La convexité de $\mathcal{G}_{S\setminus\{s\}}$ l'empêche d'être coupée par le mur M_s. Supposons que M_s soit transverse à une arête de $v.\mathcal{G}_{S\setminus\{s\}}$. Cela signifie qu'il existe $t \in S \setminus \{s\}$ et $w \in W_{S\setminus\{s\}}$ tels que $s(vw) = (vw)t$. Donc $v^{-1}sv = wtw^{-1}$: par convexité des sous-groupes spéciaux, la réflexion $v^{-1}sv$ est donc dans $W_{T_s} \cap W_{S\setminus\{s\}} = W_{T_s\setminus\{s\}}$. Alors l'élément $v' = v(v^{-1}sv) = s.v$ est congru à v modulo $W_{T_s\setminus\{s\}}$, mais il est de longueur 1 de moins que v, puisque toute géodésique de 1 à v commence par s. Ceci contredit la minimalité supposée de $|v|$. □

Nous allons montrer que, vus dans la réalisation de Davis-Moussong P de (W,S), les murs apparaissant dans le lemme 5.14 sont disjoints, et ne séparent pas deux points de A. Pour cela, nous faisons agir W sur un certain arbre.

Soient s,t dans S tels que $m_{s,t} = \infty$. Alors W est le produit amalgamé $W_{S\setminus\{s\}} *_{W_{S\setminus\{s,t\}}} W_{S\setminus\{t\}}$. Considérons le graphe biparti $\mathcal{T}_{s,t}$ ayant un sommet de type s pour chaque classe de W modulo $W_{S\setminus\{s\}}$, un sommet de type t pour chaque classe de W modulo $W_{S\setminus\{t\}}$, avec une arête entre une classe modulo $W_{S\setminus\{s\}}$ et une classe modulo $W_{S\setminus\{t\}}$ lorsque ces deux classes ne sont pas disjointes. Notons que si w appartient à $uW_{S\setminus\{s\}} \cap vW_{S\setminus\{t\}}$, alors $wW_{S\setminus\{s,t\}}$ est contenu dans $uW_{S\setminus\{s\}} \cap vW_{S\setminus\{t\}}$. La convexité des sous-groupes spéciaux entraîne alors que $wW_{S\setminus\{s,t\}}$ est égal à $uW_{S\setminus\{s\}} \cap vW_{S\setminus\{t\}}$. Ainsi, les arêtes de $\mathcal{T}_{s,t}$ correspondent bijectivement aux classes de W modulo $W_{S\setminus\{s,t\}}$.

Le groupe W agit sur $\mathcal{T}_{s,t}$ par multiplication à gauche. Cette action est transitive sur les arêtes de $\mathcal{T}_{s,t}$, le stabilisateur de $x_s = W_{S\setminus\{s\}}$ est $W_{S\setminus\{s\}}$, le stabilisateur de $x_t = W_{S\setminus\{t\}}$ est $W_{S\setminus\{t\}}$, et le stabilisateur de l'arête joignant ces deux sommets est $W_{S\setminus\{s,t\}}$. Il résulte alors de la théorie de Bass–Serre [29] que $\mathcal{T}_{s,t}$ est un arbre.

Lemme 5.16 *Vus dans P, les murs M_{i-1} et M_i apparaissant dans le lemme 5.14 sont disjoints. Le mur M_i ne sépare pas $v^-_{i-1}.a_s$ de v^-_i. Enfin $M_1 \cap M(a_s) = \emptyset$.*

Preuve Notons que v_{i-1}^+ et v_i^- ne peuvent être dans la même classe modulo W_{T_s} (sinon $w \in \mathcal{H}_{n-1}$). Par W–homogénéité, il suffit donc de montrer le résultat suivant (lequel donne du même coup la dernière partie du lemme).

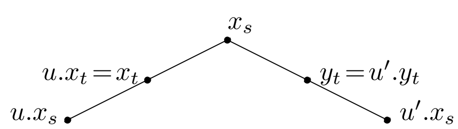
Soient $v^- \in W_{S\setminus\{s\}} \setminus W_{T_s}$ et M le mur d'une réflexion u de W_{T_s} ne coupant pas $\mathcal{G}_{S\setminus\{s\}}$. Alors M est disjoint du mur M' transverse à $v^-.a_s$.

Puisque $v^- \notin W_{T_s}$, il existe un $t \in S$ tel que $m_{s,t} = \infty$ et $v^- \notin W_{S\setminus\{t\}}$. Nous raisonnons en considérant l'action de W sur l'arbre $\mathcal{T}_{s,t}$.

D'abord, dire que $M \cap \mathcal{G}_{S\setminus\{s\}} = \emptyset$, c'est dire que $u \notin W_{S\setminus\{s\}}$. Autrement dit, $u.x_s \neq x_s$. De même, si u' est la réflexion par rapport à M', on a $u'.x_s \neq x_s$.

Soit y_t la classe à gauche de v^- modulo $W_{S\setminus\{t\}}$. Alors $u'.y_t = y_t$. D'autre part, comme $T_s \subset S \setminus \{t\}$, on a aussi $u.x_t = x_t$.

Enfin, $y_t \neq x_t$, et x_s est lié dans $\mathcal{T}_{s,t}$ à x_t et y_t.



Donc le produit $u'u$ agit comme une translation non triviale de l'arbre $\mathcal{T}_{s,t}$, et est nécessairement d'ordre infini. Or, si les murs des deux réflexions u et u' se coupaient dans P, le produit $u'u$ aurait un point fixe, donc devrait être d'ordre fini (l'action de W sur P est propre).

En fait, non seulement $M \cap M' = \emptyset$, mais de plus M' ne sépare pas 1 d'une arête a transverse à M et contenue dans $\mathcal{G}_{T_s}$ (ce qui achève de prouver le lemme). Car si c'était le cas, par convexité, M' serait transverse à une arête a' de $\mathcal{G}_{T_s}$, et u' serait une réflexion de W_{T_s}. Donc u' fixerait x_t. Comme u' fixe déjà y_t, elle fixerait l'unique sommet de type s lié à la fois à x_t et à y_t, c'est à dire x_s. Or nous avons vu que ce n'était pas le cas. □

Proposition 5.17 *$\mathcal{H}$ est disjoint de A.*

Preuve Si $w \in \mathcal{H}$, appliquons le lemme 5.14 pour trouver une suite $s = v_0^-, v_1^-, \ldots, v_{n-1}^-$, $v_n^- = w$ et une suite de murs $M_0 = M_s, M_1, \ldots, M_{n-1}$ tels que M_i sépare v_i^- de v_{i+1}^- et M_i est transverse à $v_i^-.a_s$. D'après le lemme 5.16, les murs M_i et M_{i+1} sont disjoints, et M_{i+1} ne sépare pas v_{i+1}^- de M_i (voir figure 6). Soit A_i la moitié de W définie par M_i contenant v_i^-. Il est maintenant immédiat que la suite des moitiés A_i est (strictement) croissante, avec $A_0 = A$, et $w \notin A_{n-1}$. Donc $w \notin A$. □

5.4 Exemples de complexes polyédraux pairs CAT(-1)

(1) Soient k un entier pair avec $k \geq 4$, et L le graphe d'incidence d'un plan projectif sur un corps fini, ou plus généralement n'importe quel immeuble épais fini de rang 2 vérifiant la condition de Moufang (voir [27]). Cette condition (plus le fait que L soit épais) implique en particulier que le fixateur de l'étoile d'un sommet de L est non trivial. Donc $\mathrm{Aut}^+(A(k,L))$ est non trivial, dès que $W(k,L)$ est hyperbolique (au sens de Gromov), c'est-à-dire si $k \geq 6$ ou $k = 4$ et L n'est pas de type $A_1 \times A_1$. Ceci concerne donc l'immeuble de Bourdon $I_{p,q}$, avec p pair, $p \geq 6$ et $q \geq 3$.

(2) Étant donné un polyèdre pair C, nous allons montrer comment construire un complexe polyédral pair CAT(-1) ayant un gros groupe d'automorphismes, et dont toute cellule maximale est isomorphe (combinatoirement) à C.

Proposition 5.18 *Pour tout polyèdre pair C, il existe un complexe polyédral pair localement compact* CAT(-1), *dont les cellules maximales sont combinatoirement isomorphes à C, admettant un groupe discret cocompact d'automorphismes, et dont le groupe des automorphismes engendré par les fixateurs stricts de murs propres est non dénombrable. Si C n'est pas combinatoirement un produit, alors on peut de plus supposer que tous les murs sont propres.*

Preuve Soit (W,S) le système de Coxeter fini associé à C par la proposition 4.1. Considérons une fonction $\overline{n}$ de S dans l'ensemble des entiers strictement positifs, telle que, si $\overline{n}(s) > 1$ et $\overline{n}(t) > 1$, on a $m_{s,t} > 2$ (c'est-à-dire s et t sont liés par une arête dans le graphe de Coxeter de (W,S)). Nous noterons $K_{\overline{n}}$ le sous-graphe complet du graphe de Coxeter de (W,S) dont les sommets s vérifient $\overline{n}(s) > 1$. Remarquons que par le théorème de classification des systèmes de Coxeter fini (voir par exemple [5, p.193]), $K_{\overline{n}}$ est réduit à un seul sommet ou à une seule arête.

Définissons $(\overline{W},\overline{S})$, l'unique système de Coxeter tel qu'il existe une application $\tau\colon \overline{S} \to S$ avec

i) $\tau^{-1}(\{s\})$ possède $\overline{n}(s)$ éléments;

ii) si $\overline{s} \neq \overline{t}$, ou bien $\tau(\overline{s}) = \tau(\overline{t})$, et dans ce cas $m_{\overline{s},\overline{t}} = \infty$, ou bien $\tau(\overline{s}) \neq \tau(\overline{t})$, et dans ce cas $m_{\overline{s},\overline{t}} = m_{\tau(\overline{s}),\tau(\overline{t})}$.

Il est immédiat que τ s'étend en un homomorphisme de groupes de $\overline{W}$ dans W, et est injective sur les parties $\overline{T}$ de $\overline{S}$ telles que $\overline{W}_{\overline{T}}$ est fini. Donc les simplexes du nerf fini de $(\overline{W},\overline{S})$ sont les parties de $\overline{S}$ sur lesquelles τ est injective.

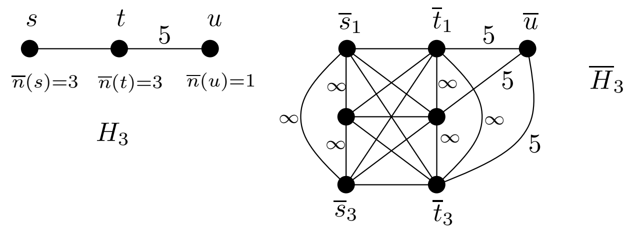


Figure 7: Exemple de système de Coxeter hyperbolique non rigide

Les permutations de $\overline{S}$ laissant τ invariante donnent des automorphismes de $(\overline{W},\overline{S})$. Si on suppose que $\overline{n}$ atteint une valeur supérieure ou égale à 3, on en déduit que $(\overline{W},\overline{S})$ n'est pas rigide.

Montrons que $(\overline{W},\overline{S})$ est hyperbolique. Si ce n'est pas le cas, d'après Moussong, $(\overline{W},\overline{S})$ contient un sous-groupe spécial affine de rang au moins 3, ou bien deux sous-groupes spéciaux infinis qui commutent. Dans le premier cas, τ est nécessairement injective sur le sous-groupe spécial (car le graphe d'un tel système de Coxeter ne contient pas d'∞), donc (W,S) est infini, contradiction. Dans le deuxième cas, un argument analogue au précédent montre qu'il existe $\overline{s_1}$, $\overline{s_2}$, $\overline{t_1}$ et $\overline{t_2}$ tels que $m_{\overline{s_1},\overline{s_2}} = m_{\overline{t_1},\overline{t_2}} = \infty$ et $m_{\overline{s_i},\overline{t_j}} = 2$ pour tous $i,j = 1,2$. Mais alors $\tau(\overline{s_1}) = \tau(\overline{s_2})$ commute avec $\tau(\overline{t_1}) = \tau(\overline{t_2})$, en contradiction avec l'hypothèse de départ sur $\overline{n}$.

Enfin, notons $P(C,\overline{n})$ la réalisation géométrique de Davis–Moussong de $(\overline{W},\overline{S})$. Alors les cellules maximales de $P(C,\overline{n})$ correspondent aux sous-groupes spéciaux finis maximaux de $(\overline{W},\overline{S})$, lesquels sont tous isomorphes à (W,S). Donc toutes les cellules maximales de $P(C,\overline{n})$ sont isomorphes à C.

Supposons (W,S) irréductible. Par classification, son graphe de Coxeter contient au plus une arête ayant un label pair. Alors il existe une application $\overline{n}$ telle que $\overline{n}(s) \geq 3$ si s appartient à $K_{\overline{n}}$, et telle que s'il existe une arête de label pair (différent de 2), alors $K_{\overline{n}}$ consiste en cette arête. Rappelons que si deux sommets d'un graphe de Coxeter peuvent être joints par un chemin d'arêtes dont tous les labels sont impairs, alors les deux réflexions correspondantes sont conjuguées dans le groupe de Coxeter (voir [5]). Donc toute réflexion de $(\overline{W},\overline{S})$ est conjuguée à un élément de $\tau^{-1}(K_{\overline{n}})$. Or le mur de toute réflexion dans $\tau^{-1}(K_{\overline{n}})$ est propre. Par conséquent, tout mur est propre. □

Par exemple, lorsque C est le polygone à $p = 2k$ côtés et $\overline{n}$ est constante égale à $q \geq 3$, le polyèdre $P(C,\overline{n})$ est l'immeuble de Bourdon $I_{p,q}$.

Lorsque C est un cube de dimension 3, $P(C,\overline{n})$ est le produit d'un arbre régulier par un carré.

Lorsque C est le polyèdre pair du groupe H_3, définissons $\overline{H_3}$ comme dans la figure précédente. Alors toutes les 2–faces de $P(C,\overline{n})$ sont contenues dans 3 copies de C, sauf les décagones, qui ne sont contenus que dans une copie de C.

5.5 Automorphismes préservant le type de complexes polyédraux pairs

Dans toute cette section, P est un complexe polyédral pair CAT(0) dont toutes les cellules maximales (appelées *chambres* par la suite) sont isométriques à une cellule C fixée (par exemple, P est un (k,L)–complexe, au sens de [23, 4], voir aussi [3]). La codimension des faces de P est maintenant bien définies.

Définition 5.19 Une *fonction type* de P dans C est une application polyédrale $\tau\colon P \to C$ dont la restriction à chaque chambre de P est une isométrie.

Exemples (1) Supposons que C soit une cellule paire de l'espace E_χ à courbure constante $\chi \leq 0$, dont les faces de codimension 1 font des angles dièdres de la forme $\frac{\pi}{n}$, avec $n \geq 2$. Alors, par le théorème de Poincaré (voir par exemple [24]), le sous-groupe $W(C)$ des isométries de E_χ engendré par les réflexions par rapport aux faces de codimension 1 de C est discret, et le quotient de E_χ par $W(C)$ s'identifie naturellement à C. Cela signifie que le pavage $P(C)$ de E_χ donné par les wC, avec $w \in W(C)$, admet une fonction type dans C.

(2) Plus généralement, tout immeuble P dont les appartements sont isométriques à $P(C)$ admet une fonction type (on fixe une certaine copie A_0 de $P(C)$ dans P, ainsi qu'une certaine chambre C_0 de A_0, puis on considère la rétraction de P sur A_0 basée en C_0, et on la compose par une quelconque fonction type sur A_0).

(3) Enfin, un arbre quelconque admet toujours une fonction type à valeur dans l'une de ses arêtes.

Appelons *galerie de P* toute suite de chambres $(C_0, C_1, \ldots, C_n)$ telles que $C_i \cap C_{i+1}$ contient une cellule de codimension 1. Nous laissons au lecteur le soin de démontrer la proposition suivante, qui ne servira pas dans ce texte.

Proposition 5.20 *Supposons que deux chambres de P sont jointes par au moins une galerie. Deux fonctions de type égales sur une chambre C_0 de P*

sont égales. S'il est non vide, l'ensemble des fonctions types sur P s'identifie avec l'ensemble (fini) des isométries de C_0 sur C. Dans ce cas, le link d'une face de codimension 2 de P est biparti.

Réciproquement, si P est de dimension 2 avec 2-cellules régulières, et si le link de chaque sommet de P est un graphe biparti connexe, alors l'ensemble des fonctions types sur P est non vide. □

A partir de maintenant, nous supposerons que P admet une fonction type dans un polyèdre pair C, et que deux chambres quelconques de P sont jointes par une galerie.

Définition 5.21 Nous noterons $\mathrm{Aut}_0(P)$ le noyau de l'action par précomposition du groupe $\mathrm{Aut}(P)$ sur l'ensemble des fonctions types de P dans C. Nous dirons que ses éléments *préservent le type*.

Remarque Si C' est isomorphe à C, un élément de $\mathrm{Aut}(P)$ préserve le type dans C si et seulement s'il préserve le type dans C'. C'est ce qui justifie l'omission de C dans la notation $\mathrm{Aut}_0(P)$. Remarquons que $\mathrm{Aut}_0(P)$ est d'indice fini dans $\mathrm{Aut}(P)$.

Notons $\mathrm{Aut}_F(P)$ le sous-groupe caractéristique de $\mathrm{Aut}(P)$ engendré par les fixateurs de facettes (au sens de la définition 5.10). Ses éléments seront appelés *F-automorphismes*. Notons G_0 et G_1 les sous-groupes de $G = \mathrm{Aut}(P)$ engendrés par les intersections avec $\mathrm{Aut}_F(P)$ des fixateurs de chambres d'une part, et des fixateurs de cellules de codimension 1 d'autre part.

Il est clair que $G_0 \subset G_1 \subset \mathrm{Aut}_F(P)$. Si M est un mur propre de P, son fixateur strict est dans G_0. Donc $\mathrm{Aut}^+(P)$ est contenu dans G_0. Si $f \in \mathrm{Aut}(P)$ fixe une face F de codimension 1 et envoie une chambre C_2 contenant F sur C_1, alors $f|_{C_2}$ commute avec la fonction type de P restreinte à C_1 et à C_2. Par connexité par galeries de P, f préserve alors le type. Donc $G_1 \subset \mathrm{Aut}_0(P)$. En résumé, $\mathrm{Aut}^+(P) \subset G_0 \subset G_1 \subset \mathrm{Aut}_0(P) \cap \mathrm{Aut}_F(P)$.

Introduisons des propriétés de transitivité, globales ou locales:

- (T_0) L'action de G_0 sur l'ensemble des chambres de P est transitive.
- (T_1) L'action de G_1 sur l'ensemble des chambres de P est transitive.
- (TL_1) Pour toute face σ de codimension 1, le fixateur de σ dans $\mathrm{Aut}_F(P)$ agit transitivement sur les chambres contenant σ.

(TL_0) Pour toute face σ de codimension 1, le sous-groupe de $\mathrm{Fix}(\sigma)\cap\mathrm{Aut}_F(P)$ engendré par les $\mathrm{Fix}(\sigma)\cap\mathrm{Aut}_F(P)\cap\mathrm{Fix}(C)$, où C est une chambre de P, agit transitivement sur les chambres contenant σ.

Il est immédiat que (T_0) implique (T_1) et (TL_0) implique (TL_1). D'autre part:

Lemme 5.22 *Pour $i=0,1$, la condition (TL_i) implique (T_i), qui implique que $G_i=\mathrm{Aut}_0(P)\cap\mathrm{Aut}_F(P)$.*

Preuve La première implication découle de la connexité par galerie de l'ensemble des chambres de P. La deuxième de ce qu'un élément de $\mathrm{Aut}_0(P)$ préservant une chambre la fixe nécessairement. □

Si a et b sont deux arêtes adjacentes à un sommet x_0 et contenues dans un même polygone de P, nous noterons $m_{a,b}$ la moitié du nombre de côtés de ce polygone. Nous obtenons ainsi une fonction de l'ensemble des arêtes de $lk(x_0,P)$ dans l'ensemble des entiers supérieurs ou égaux à 2. Il est clair qu'un automorphisme f de P envoie la fonction m du sommet x_0 sur la fonction m du sommet $f(x_0)$. Nous noterons G_{x_0} le groupe des automorphismes de $(lk(x_0,P),m)$ engendré par les fixateurs de facettes de $lk(x_0,P)$. Remarquons que si P est de dimension 2, alors m est constant (car tous les polygones ont le même nombre de côtés).

Voici maintenant deux propriétés de prolongement:

(P_0) Pour tout sommet x_0 de P, tout élément de G_{x_0} s'étend à P.

(P^+) Pour tout sommet x_0 de P, tout élément de G_{x_0} fixant une facette (voir définition 5.10) ϕ (transverse à une arête issue de x_0) s'étend à P en un automorphisme fixant le mur M passant par ϕ et fixant toute la moitié de P définie par M et ne contenant pas x_0.

Nous avons maintenant des conditions permettant d'identifier $\mathrm{Aut}^+(P)$ et $\mathrm{Aut}_0(P)\cap\mathrm{Aut}_F(P)$, dans le cas où P est de dimension 2.

Proposition 5.23 *Soit P un complexe polyédral pair* $\mathrm{CAT}(0)$ *de dimension 2 admettant un type et dont deux chambres sont jointes par au moins une galerie.*

(1) *Supposons que, pour tout sommet x_0 de P, le stabilisateur dans G_{x_0} d'un sommet de $lk(x_0,P)$ agit transitivement sur les arêtes issues de ce sommet. Si P vérifie (P_0), alors $\mathrm{Aut}_0(P)\cap\mathrm{Aut}_F(P)=G_1$.*

(2) *Supposons que, pour tout sommet* x_0 *de* P, *pour toute arête* a *issue de* x_0, *l'ensemble* $E(a)$ *des polygones de* P *contenant* a *est de cardinal au moins trois, et que pour tout polygone* c *contenant* a, *le stabilisateur dans* G_{x_0} *de* c *agit transitivement sur* $E(a)\setminus\{c\}$. *Si* P *vérifie* (P_0), *alors* $G_0 = G_1 = \mathrm{Aut}_0(P) \cap \mathrm{Aut}_F(P)$.

(3) *Supposons que, pour tout sommet* x_0 *de* P, *toute arête* c *de* $lk(x_0, P)$, *et tout* $f \in G_{x_0}$ *fixant* c, *on a une décomposition* $f = f_1 \circ f_2$, *où* $f_1 \in G_{x_0}$ *fixe toute une facette de* $lk(x_0, P)$ *contenant une extrémité de* c, *et* $f_2 \in G_{x_0}$ *fixe toute la facette de* $lk(x_0, P)$ *contenant l'autre extrémité de* c. *Supposons que la restriction d'un* F*-automorphisme fixant un sommet* x_0 *à* $lk(x_0, P)$ *est dans* G_{x_0}. *Si* P *vérifie* (P^+) *et si tous ses murs sont propres, alors* $\mathrm{Aut}^+(P) = G_0$.

Preuve Pour la première assertion, il suffit de remarquer que l'hypothèse, plus la propriété (P_0), entraînent la propriété (TL_1). On applique alors le lemme 5.22 précédent.

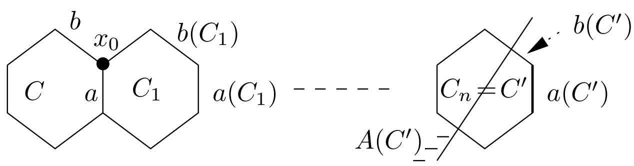
Pour la deuxième, par le lemme 5.22, il suffit de vérifier que P satisfait (TL_0). Soit a une arête de P contenues dans deux polygones c_1, c_2. Fixons un sommet x_0 de a et un troisième polygone c contenant a distinct de c_1, c_2. Par hypothèse, soit f dans G_{x_0} fixant c et envoyant c_1 sur c_2. La propriété (P_0) permet d'étendre f en un F-automorphisme de P, qui fixe a et c, et envoie c_1 sur c_2, ce qui montre (TL_0).

Montrons la troisième assertion. Comme $\mathrm{Aut}^+(P)$ est contenu dans le groupe engendré par les fixateurs de chambres, il suffit de montrer que pour toute chambre C de P, le groupe $\mathrm{Fix}(C)$ est contenu dans $\mathrm{Aut}^+(P) \cap \mathrm{Aut}_F(P)$. En fait, nous allons montrer que si $\overline{f}$ est dans $\mathrm{Fix}(C)$ et si a et b sont deux arêtes du polygone C adjacentes en un sommet x_0, alors il existe $\overline{f}_a$ et $\overline{f}_b$ fixant strictement les murs $M(a)$ et $M(b)$ tels que $\overline{f} = \overline{f}_b \circ \overline{f}_a$.

Affirmation 1 Il existe un automorphisme $\overline{f}_b$ de P fixant $M(b)$ et toute la moitié de P définie par $M(b)$ ne contenant pas x_0, tel que $\overline{f}$ coïncide avec $\overline{f}_b$ sur l'ensemble des chambres de P contenant l'arête a.

Preuve L'automorphisme $\overline{f}$ fixe x_0 et C, donc induit un automorphisme f de G_{x_0} fixant l'arête c entre les sommets du link correspondant aux arêtes a et b de P. Vu l'hypothèse sur P, il existe f_a et f_b dans G_{x_0} tels que $f = f_b \circ f_a$. D'après (P^+), on peut prolonger ces deux automorphismes locaux en éléments $\overline{f}_a$ et $\overline{f}_b$ fixant strictement les murs $M(a)$ et $M(b)$ (ainsi que les moitiés convenables). Maintenant l'égalité $\overline{f} = \overline{f}_b \circ \overline{f}_a$ sur l'étoile de x_0 entraîne $\overline{f} = \overline{f}_b$ sur l'ensemble des chambres de P contenant l'arête a, puisque $\overline{f}_a$ agit trivialement sur cet ensemble. □

Appelons *galerie géodésique de* $M(a)$ *d'origine* (C,a) toute galerie sans répétition $(C_0, C_1, \ldots, C_n)$ telle que $C_0 = C$, $C_0 \cap C_1 = a$, l'arête $C_i \cap C_{i+1}$ est parallèle à a et distincte de a_{i-1}. Comme $M(a)$ est un arbre, deux galeries géodésiques de $M(a)$ d'origine (C,a) et de mêmes extrémités sont égales. Nous noterons $\delta(C_n)$ la longueur n de cette galerie. Soit alors $\mathcal{B}_n^+(C)$ l'ensemble des polygones C' de P qui sont extrémités d'une galerie géodésique de $M(a)$



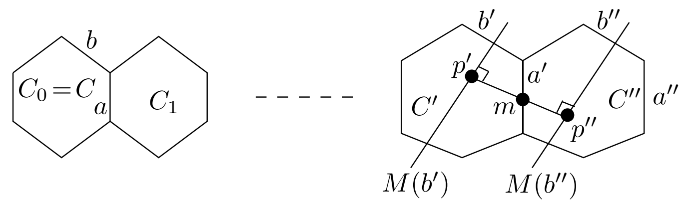
d'origine (C,a) de longueur au plus n. Nous noterons $a(C)$ l'arête a; pour $C' \in \mathcal{B}_n^+(C)$, avec $\delta(C') = n > 0$, soit $a(C')$ l'arête de C' parallèle à a, non contenue dans un polygone de $\mathcal{B}_{n-1}^+(C)$. Nous pouvons ensuite définir $b(C')$ comme l'arête de C' adjacente à $a(C')$, non séparée de b par $M(a) = M(a(C'))$.

Soient enfin $A(C')$ la moitié fermée de P définie par $M(b(C'))$ ne contenant pas $a(C')$, et $\mathcal{B}^+(C)$ l'union des $\mathcal{B}_n^+(C)$.

Affirmation 2 Les murs $M(b(C'))$ sont deux à deux disjoints; la moitié $A(C')$ contient strictement le mur $M(b(C''))$ (donc la chambre C''), dès que la galerie géodésique de $M(a)$ d'origine (C,a) et d'extrémité C'' ne passe pas par C'.

Preuve Prouvons d'abord que $M(b(C')) \cap M(b(C'')) = \emptyset$, lorsque C' et C'' sont deux chambres de $\mathcal{B}^+(C)$ telles que $C' \cap C''$ est une arête, et $\delta(C') \neq \delta(C'')$. Nous pouvons supposer les notations telles que l'arête a' commune à C' et C'' est l'arête $a(C')$ (autrement dit, $\delta(C') < \delta(C'')$). Pour alléger, nous notons alors a'', b' et b'' les arêtes $a(C'')$, $b(C')$ et $b(C'')$. Pour voir que $M(b') \cap M(b'') = \emptyset$, il suffit de voir que les deux murs ont une perpendiculaire commune (dans $C' \cup C''$): l'inégalité CAT(0) permet alors de conclure.

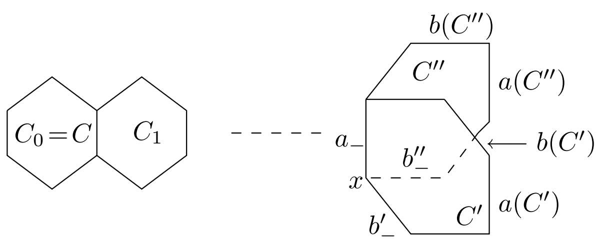
Soient m le milieu de a', et p' (resp: p'') la projection orthogonale de m sur $M(b')$ (resp: $M(b'')$). Alors p' s'obtient comme l'intersection avec $M(b', C')$ de la géodésique de C' joignant m à son image par $\sigma(b', C')$. En particulier, P' est à l'intérieur de C', et $p'' \neq p'$. Il reste à montrer que la géodésique de P joignant p' et p'' passe par m.



La réunion de C' et C'' admet deux réflexions orthogonales: $\sigma_{a'}$ qui échange les extrémités de a', et $\rho_{a'}$ qui fixe a' en échangeant les deux chambres C' et C'' (rappelons que P admet un type). Il est alors immédiat que la symétrie centrale $\rho_{a'} \circ \sigma_{a'}$ envoie $M(b', C')$ sur $M(b'', C'')$ en fixant m, donc envoie le segment le m à p' sur le segment de m à p'', de sorte que l'union de ces deux segments est encore une géodésique.

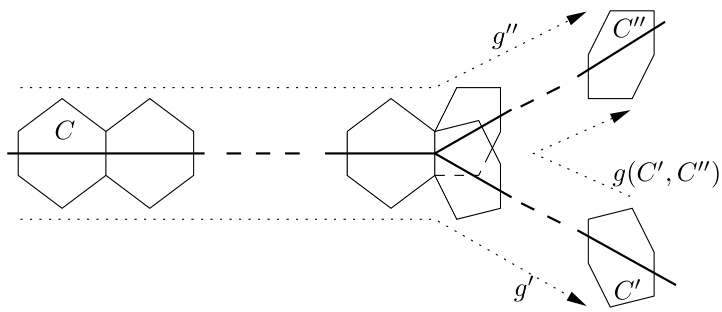
Il est maintenant clair que, si $(C_0, C_1, \ldots, C_n)$ est une galerie géodésique de $M(a)$ d'origine (C, a), la suite des demi-espaces fermés $A(C_i)$ est strictement croissante. En particulier, $A(C_n)$ contient strictement les murs $M(b(C_i))$, pour $0 \leq i < n$.

Montrons maintenant que $M(b(C')) \cap M(b(C'')) = \emptyset$, lorsque C' et C'' sont deux chambres de $\mathcal{B}^+(C)$ telles que $C' \cap C''$ est une arête, et $\delta(C') = \delta(C'')$. Dans ce cas, l'arête a_- formant $C' \cap C''$ est opposée à $a(C')$ et $a(C'')$ dans C' et C'' respectivement. Il existe alors deux arêtes b'_- et b''_- de C' et C'', opposées à $b(C')$ et $b(C'')$ respectivement, donc adjacentes à a_- en un sommet x, avec $M(b'_-) = M(b(C'))$ et $M(b''_-) = M(b(C''))$.



Pour montrer que ces deux murs sont disjoints, on exhibe là aussi une perpendiculaire commune. Auparavant, on modifie la métrique CAT(0) sur P, en rendant tous les polygones de P réguliers à angle droit (donc hyperboliques, sauf si au départ on avait des carrés). La nouvelle métrique est bien encore CAT(0) (et même souvent CAT(−1)), puisque tous les links de P sont des graphes bipartis (P admet un type), donc ont des circuits de longueur au moins 4. Alors $b'_- \cup b''_-$ est géodésique même en x, et perpendiculaire aux deux murs.

Il est alors évident que $A(C')$ contient strictement la moitié fermée de P définie par $M(b(C''))$ et contenant $a(C'')$ (i.e. dont la réunion avec $A(C'')$ est P entier).



Pour achever la preuve de l'affirmation, soient C' et C'' deux chambres distinctes de $B^+(C)$, telles que la galerie géodésique de $M(a)$ d'origine (C,a) et d'extrémité C'' ne passe pas par C'. Si g' et g'' sont les galeries géodésiques de $M(a)$ d'origine (C,a) et d'extrémité C' et C'' respectivement, la galerie $g(C',C'')$ obtenue à partir de $g'^{-1}.g''$ en ôtant les répétitions permet, compte tenu des résultats préliminaires ci-dessus, de construire une suite strictement décroissante de moitiés fermées dont la première est $A(C')$ et la dernière est la moitié complémentaire de $A(C'')$. Ceci conclut. $\square$

Revenons à la preuve de la proposition. Pour $\overline{f} \in \mathrm{Fix}(C) \cap \mathrm{Aut}_F(P)$, supposons avoir construit $\overline{f}_b^n$ fixant strictement le mur $M(b)$, et coïncidant avec $\overline{f}$ sur chaque chambre de $\mathcal{B}_n^+(C)$ (c'est vrai pour $n=1$, d'après l'affirmation 1). Alors $(\overline{f}_b^n)^{-1} \circ \overline{f}$ agit trivialement sur chaque chambre de $\mathcal{B}_n^+(C)$. Soient $C^1, \dots, C^k$ les chambres de $\mathcal{B}^+(C)$ avec $\delta(C^i) = n$. En appliquant l'affirmation 1 à la chambre C^1, aux arêtes $a(C^1)$ et $b(C^1)$, on trouve $\overline{f}^1$ fixant toute la moitié $A(C^1)$ et coïncidant avec $(\overline{f}_b^n)^{-1} \circ \overline{f}$ sur l'ensemble des chambres contenant $a(C^1)$. D'après l'affirmation 2, $\overline{f}^1$ fixe strictement le mur $M(b)$, agit trivialement sur toutes les chambres de $\mathcal{B}_n^+(C)$, et même sur les chambres de $\mathcal{B}^+(C)$ adjacentes à $C^2, C^3, \dots$ ou C^k. Alors $(\overline{f}^1)^{-1} \circ (\overline{f}_b^n)^{-1} \circ \overline{f}$ agit trivialement sur chaque chambre de $\mathcal{B}_n^+(C)$, et sur chaque chambre contenant $a(C^1)$. En réutilisant les affirmations 1 et 2, on trouve des automorphismes $\overline{f}^2, \overline{f}^3, \dots, \overline{f}^k$ fixant tous strictement $M(b)$, tels que $\overline{f}_b^{n+1} = \overline{f}_b^n \circ \overline{f}^1 \circ \overline{f}^2 \circ \dots \circ \overline{f}^k$ agit comme $\overline{f}$ sur $\mathcal{B}_{n+1}^+(C)$. L'automorphisme $\overline{f}_b^{n+1}$ fixe strictement $M(b)$ et coïncide avec $\overline{f}$ sur chaque chambre de $B_{n+1}^+(C)$. En itérant ce processus, et quitte à extraire une sous-suite convergente, on trouve à la limite un $\overline{f}_b^+$ fixant strictement le mur $M(b)$ et coïncidant avec $\overline{f}$ sur chaque chambre de $\mathcal{B}^+(C)$. On peut imposer que $\overline{f}_b^+$ fixe point par point la moitié de P définie par $M(b)$ et ne contenant pas a.

En appliquant la construction précédente sur l'autre moitié de $M(b)$, on trouve un $\overline{f}_b^-$ coïncidant avec $\overline{f}$ sur chaque chambre de $\mathcal{B}^-(C)$, et fixant point par point la moitié de P définie par $M(b)$ et contenant a. Si on pose $\overline{f}_b = \overline{f}_b^+ \circ \overline{f}_b^-$ et $\overline{f}_a = (\overline{f}_b)^{-1} \circ \overline{f}$, on a $\overline{f} = \overline{f}_b \circ \overline{f}_a$, avec $\overline{f}_b$ fixant strictement le mur $M(b)$, et $\overline{f}_a$ fixant strictement le mur $M(a)$. □

Soit k un entier pair au moins 4 et L un graphe fini de maille au moins 5 si $k = 4$ et 4 si $k \geq 6$. Pour tout bloc B de $A(k, L)$ (au sens de la définition 5.10), notons F_B le sous-groupe caractéristique des automorphismes de B engendré par les fixateurs de facettes dans B. En fait, si x est le centre du bloc B, alors $F_B = G_x$ avec les notations précédant la proposition 5.23. Remarquons que $W(k, L)$ est un sous-groupe de $\mathrm{Aut}_F A(k, L)$.

Lemme 5.24 *Soit B_0 un bloc de $A(k, L)$. Si $\rho = \rho_{B_0}$ désigne le morphisme de restriction de* $\mathrm{Stab}(B_0, \mathrm{Aut}A(k, L))$ *dans* $\mathrm{Aut}(B_0)$, *alors*

$$\rho(\mathrm{Stab}(B_0, \mathrm{Aut}_F A(k, L))) = F_{B_0}.$$

Preuve Pour tout bloc B de $A(k, L)$, notons $\underline{F}_B$ l'image réciproque de F_B par ρ_B. Par le prolongement $W(k, L)$-équivariant (voir les remarques avant la définition 5.10), on a $\rho_B(\underline{F}_B) = F_B$.

D'autre part, si $\overline{F}_B$ est le stabilisateur de B dans $\mathrm{Aut}_F A(k, L)$, alors $\underline{F}_B \subset \overline{F}_B$. En effet, si $\hat{\varphi} \in \underline{F}_B$, par définition $\rho_B(\hat{\varphi})$ s'écrit $\rho_B(\hat{\varphi}) = \varphi_1 \cdots \varphi_n$, où les φ_i sont des automorphismes de B fixant une facette de B. Comme $\rho_B(\underline{F}_B) = F_B$, il existe $\hat{\varphi}_1, \cdots, \hat{\varphi}_n$ éléments de $\underline{F}_B$ prolongeant les φ_i. On a donc $\hat{\varphi} = \hat{\varphi}_1 \cdots \hat{\varphi}_n \varepsilon$, où ε vaut l'identité sur B. Chaque terme de la décomposition fixant une facette de B, on a $\hat{\varphi} \in \mathrm{Aut}_F A(k, L)$.

Pour montrer l'inclusion réciproque $\underline{F}_{B_0} \supset \overline{F}_{B_0}$, introduisons le sous-groupe H de $\mathrm{Aut}A(k, L)$ engendré par $W(k, L)$ et $\underline{F}_{B_0}$. Nous allons d'abord montrer que $H = \mathrm{Aut}_F A(k, L)$, puis que $\mathrm{Stab}(B_0, H) = \underline{F}_{B_0}$, ce qui achèvera la preuve du lemme.

D'abord, comme $\underline{F}_{B_0} \subset \overline{F}_{B_0}$ et $W(k, L) \subset \mathrm{Aut}_F(A(k, L))$, on a bien $H \subset \mathrm{Aut}_F A(k, L)$. Réciproquement si f est un automorphisme de $A(k, L)$ fixant une facette ϕ, il existe w dans $W(k, L)$ tel que $w(\phi) \subset B_0$. Alors l'automorphisme wfw^{-1} fixe une facette ϕ' de B_0. Si s désigne la réflexion de

$W(k,L)$ par rapport au mur passant par cette facette, il existe un $k\in\{0,1\}$ tel que $s^k wfw^{-1}$ préserve B_0 et fixe une facette ϕ' de B_0, donc est dans $\underline{F}_{B_0}$. Ainsi $f\in H$, et H contient $\mathrm{Aut}_F A(k,L)$.

Montrons maintenant que $\mathrm{Stab}(B_0,H)=\underline{F}_{B_0}$, c'est-à-dire $\mathrm{Stab}(B_0,H)\subset\underline{F}_{B_0}$. Si nous vérifions que tout $h\in H$ peut s'écrire $h=wf$, avec $w\in W(k,L)$ et $f\in\underline{F}_{B_0}$, alors on aura $h(B_0)=B_0$ implique $w=1$, donc $h\in\underline{F}_{B_0}$. Pour établir que H coïncide avec l'ensemble H' des automorphismes f de $A(k,L)$ tels que $w_0^{-1}f\in\underline{F}_{B_0}$, pour w_0 l'unique élément de $W(k,L)$ tel que $w_0(B_0)=f(B_0)$, introduisons l'ensemble H'' des automorphismes f de $A(k,L)$ tels que, pour tout bloc B, $w^{-1}f\in\underline{F}_B$, avec w l'unique élément de $W(k,L)$ tel que $w(B)=f(B)$. Il est clair que H'' est un sous-groupe de H contenu dans H'. Pour conclure, montrons que $H''=H$. Comme $W(k,L)\subset H''$, il suffit de montrer que $\underline{F}_{B_0}\subset H''$, ce qui découle de l'affirmation suivante: si $w_1^{-1}f$ est dans $\underline{F}_{B_1}$ et si $B_1\cap B_2$ est une facette ϕ, alors $w_2^{-1}\circ f$ est dans $\underline{F}_{B_2}$ (avec $w_i(B_i)=f(B_i)=B_i'$). Pour voir ceci, soit s (resp. s') la réflexion de $W(k,L)$ échangeant B_1 et B_2 (resp. B_1' et B_2'), alors $w_2=s'w_1s$. Posons $\varepsilon_i=w_i^{-1}f$. Par hypothèse sur l'automorphisme f, on a ε_1 est dans $\underline{F}_{B_1}$. Donc $s\varepsilon_1 s$ est dans $\underline{F}_{B_2}$. Or $(s\varepsilon_1 s)^{-1}\varepsilon_2=sf^{-1}w_1ssw_1^{-1}s'f=sf^{-1}s'f$. Ce dernier automorphisme fixe la facette ϕ et préserve le bloc B_2, donc est lui aussi dans $\underline{F}_{B_2}$, ce qui conclut. □

Appelons *facette de* L l'étoile d'un sommet de L dans la subdivision barycentrique L'. Soit F le sous-groupe caractéristique de $\mathrm{Aut}(L)$ engendré par les fixateurs de facettes de L. Si L est le graphe biparti complet sur $p+q$ sommets avec $p,q\geq 3$, alors $\mathrm{Aut}_0(L)(\simeq S_p\times S_q)=F$.

Corollaire 5.25 *Le quotient de* $\mathrm{Aut}A(k,L)$ *par son sous-groupe distingué* $\mathrm{Aut}_F A(k,L)$ *est isomorphe au quotient de* $\mathrm{Aut}(L)$ *par son sous-groupe distingué* F.

Preuve Le groupe $\mathrm{Aut}A(k,L)$ est transitif sur les sommets de $A(k,L)$ car $W(k,L)$ l'est. Donc pour tout sommet x_0, centre du bloc B_0, le quotient $\mathrm{Aut}A(k,L)/\mathrm{Aut}_F A(k,L)$ est isomorphe à $\mathrm{Fix}\, x_0/\mathrm{Fix}\, x_0\cap\mathrm{Aut}_F A(k,L)$. Par restriction, on a un morphisme $\mathrm{Fix}\, x_0\to\mathrm{Aut}(B_0)$, qui est surjectif par le paragraphe précédant la définition 5.10. Son noyau est contenu dans $\mathrm{Fix}\, x_0\cap\mathrm{Aut}_F A(k,L)$. De plus, par le lemme précédent, l'image de $\mathrm{Fix}\, x_0\cap\mathrm{Aut}_F A(k,L)$ est exactement F_{B_0}. Donc $\mathrm{Fix}\, x_0/\mathrm{Fix}\, x_0\cap\mathrm{Aut}_F A(k,L)$ est isomorphe au quotient $\mathrm{Aut}(B_0)/F_{B_0}$. □

Soit G un groupe de Chevalley fini de rang 2, sur le corps fini K, de système de racines Φ, de racines fondamentales α_1, α_2, de racines positives Φ^+ et de groupes de racines

$$X_\alpha = \{x_\alpha(t) \,/\, t \in K\}$$

pour $\alpha \in \Phi$. Nous utiliserons les notations de [12]. En particulier, U est le sous-groupe de G engendré par les racines positives. On a un morphisme $h : \mathrm{Hom}(\mathbb{Z}\alpha_1 \oplus \mathbb{Z}\alpha_2, K^\times) \rightarrow \mathrm{Aut}(G)$ qui, à un caractère χ du réseaux des racines à valeurs dans le groupe multiplicatif de K, associe l'automorphisme de G induit par l'automorphisme

$$h(\chi) : x_\alpha(t) \mapsto x_\alpha(\chi(\alpha)t)$$

sur chaque groupe de racine de G. On rappelle (voir [12]) que G est sans centre, est engendré par les groupes de racines X_α, et que chaque racine est combinaison linéaire à cœfficients entiers (tous du même signe) de α_1, α_2.

On identifie G à son image dans $\mathrm{Aut}(G)$ par les automorphismes intérieurs. On note $\widehat{H}$ l'image de h, $H = G \cap \widehat{H}$ et $B = UH$. Il existe alors (voir [12, page 101]) un sous-groupe N de G tel que (B, N) est une BN-paire de G. Soit L le m-gone généralisé associé à cette BN-paire, muni de son action de G, de sa chambre fondamentale c de fixateur B, et de son appartement fondamental Σ de fixateur H [12, page 102]. On identifie Φ avec l'ensemble des demi-appartements de Σ, de sorte que Φ^+ corresponde à ceux contenant c, et que X_α soit le fixateur de la réunion de α et des arêtes de L rencontrant α en un sommet intérieur de α.

Notons que $\widehat{H}$ préserve chaque groupe de racine X_α. Par conséquent, il agit sur L en fixant Σ (et en particulier en préservant le type). Pour $i = 1, 2$, notons x_i le sommet de c appartenant au bord de α_i, c_i la chambre de Σ adjacente à c en x_i, et ϕ_i la facette de L de centre x_i. Les arêtes de ϕ_i sont les moitiés contenant x_i des chambres disjointes $\{c\} \cup \{x_{\alpha_i}(t)c_i \,/\, t \in K\}$, car X_{α_i} agit simplement transitivement sur l'ensemble des chambres contenant x_i différentes de c. Tout caractère $\chi : \mathbb{Z}\alpha_1 \oplus \mathbb{Z}\alpha_2 \rightarrow K^\times$ s'écrit comme un produit de caractères $\chi_1\chi_2$ avec χ_i valant 1 sur α_i. Comme h est un morphisme, on a donc $h(\chi) = h(\chi_1)h(\chi_2)$. De plus $h(\chi_i)$ fixe ϕ_i par la description précédente.

Proposition 5.26 *Si F est le sous-groupe caractéristique de* $\mathrm{Aut}(L)$ *engendré par les fixateurs de facettes de L, alors*

(1) $F = G\widehat{H}$,

(2) *Le fixateur* $\mathrm{Fix}_F(c)$ *de c dans F est* $U\widehat{H}$,

(3) $\mathrm{Aut}_0(L)/F$ *est isomorphe au groupe* $\mathrm{Aut}(K)$ *des automorphismes du corps* K*.*

Preuve (1) D'après les rappels précédents, l'inclusion de $G\widehat{H}$ dans F est claire. Soit ϕ une facette de L et f un automorphisme de L fixant ϕ. On veut montrer que f appartient à $G\widehat{H}$. Quitte à composer à gauche par un élément de G, on peut supposer que f fixe Σ et l'une des facettes ϕ_1 ou ϕ_2, disons ϕ_1. Comme f préserve ϕ_2 en fixant c et c_2, on peut écrire

$$f(x_{\alpha_2}(1)c_2) = x_{\alpha_2}(\xi)c_2$$

pour un certain ξ dans $K - \{0\}$. Soit χ le caractère qui à α_1 associe 1 et à α_2 associe ξ. Montrons alors que $f = h(\chi)$. Posons $\theta = h(\chi)^{-1}f$. C'est un automorphisme de L fixant Σ, ϕ_1 et $x_{\alpha_2}(1)c_2$. Notons que θ normalise G, et notons encore θ l'automorphisme de G induit. Alors θ préserve chaque X_α, $\theta(x_{\alpha_1}(t)) = x_{\alpha_1}(t)$ pour tout $t \in K$ et $\theta(x_{\alpha_2}(1)) = x_{\alpha_2}(1)$. Il découle de la preuve du théorème 12.5.1 de [12, page 211] que l'ensemble des automorphismes θ de G qui préservent chaque X_α, et fixent $x_{\alpha_i}(1)$ pour $i = 1, 2$, est un sous-groupe de G isomorphe à $\mathrm{Aut}(K)$, et que si de plus $\theta(x_{\alpha_1}(t)) = x_{\alpha_1}(t)$ pour tout $t \in K$, alors θ vaut l'identité.

(2) L'inclusion de $U\widehat{H}$ dans $\mathrm{Fix}_F(c)$ est claire. Réciproquement, soit f dans F fixant c. Alors $f = g\widehat{f}$ avec $g \in G$ et $\widehat{f} \in \widehat{H}$ par (1). Comme $\widehat{f}$ fixe c, on en déduit que g fixe c. Or le fixateur de c dans G est $B = UH$, et comme $H \subset \widehat{H}$, le résultat en découle.

(3) Il est clair que F est contenu et distingué dans $\mathrm{Aut}_0(L)$. Soit $\theta \in \mathrm{Aut}_0(L)$. Quitte à le multiplier par un élément de G, on peut supposer que θ fixe Σ. Quitte à le multiplier par un élément de $\widehat{H}$, on peut supposer que θ fixe $x_{\alpha_i}(1)$ pour $i = 1, 2$. Soit Z le fixateur dans $\mathrm{Aut}_0(L)$ de $\Sigma \cup \{x_{\alpha_1}(1), x_{\alpha_2}(1)\}$. On a donc un isomorphisme entre $\mathrm{Aut}_0(L)/F$ et $Z/Z \cap F$. Or si $f = g\widehat{f}$ fixe Σ, avec $g \in G$ et $\widehat{f} \in \widehat{H}$, alors g fixe Σ. Donc g appartient au fixateur de Σ dans G, qui est H. Par conséquent $f \in \widehat{H}$. Or un élément $h(\chi)$ de $\widehat{H}$ fixant $x_{\alpha_1}(1)$ et $x_{\alpha_2}(1)$ vaut l'identité, car on aurait $\chi(\alpha_1) = 1$ et $\chi(\alpha_2) = 1$. D'où $Z \cap F = \{1\}$, ce qui montre le résultat, Z étant isomorphe à $\mathrm{Aut}(K)$, d'après le dernier argument de (1). □

Corollaire 5.27 *Pour* $i = 1, 2$*, soit* F_i *le fixateur de la facette* ϕ_i *dans* $\mathrm{Aut}(L)$*. Alors*

$$\mathrm{Fix}_F(c) = F_1F_2 = F_2F_1.$$

Preuve L'égalité $F_1F_2 = F_2F_1$ vient du fait que F_1 et F_2 fixent $c = \phi_1 \cap \phi_2$ donc F_i préserve ϕ_{3-i}. L'inclusion de F_1F_2 dans $\mathrm{Fix}_F(c)$ est claire. Pour montrer l'inclusion inverse, comme $\mathrm{Fix}_F(c) = U\widehat{H}$, il suffit de le faire pour U et pour $\widehat{H}$. Or U est engendré par les X_α pour α racine positive, et un tel X_α est contenu soit dans F_1, soit dans F_2. De plus, on a vu avant la proposition 5.26 que pour tout caractère χ, $h(\chi) = h(\chi_1)h(\chi_2)$ avec $h(\chi_i)$ fixant ϕ_i. □

Corollaire 5.28 *Si L est un m–gone généralisé épais fini classique, alors le groupe $\mathrm{Aut}_0 A(k,L) \cap \mathrm{Aut}_F(P)$ des F-automorphismes préservant le type de $A(k,L)$, coïncide avec le groupe $\mathrm{Aut}^+ A(k,L)$ des automorphismes de $A(k,L)$ engendré par les fixateurs stricts de murs propres, et est distingué dans $\mathrm{Aut}_0 A(k,L)$, de quotient trivial si $m = 2$, et sinon isomorphe au groupe fini des automorphismes de corps du corps fini de définition de L.*

Preuve Nous allons vérifier les hypothèses de la proposition 5.23 (2) et (3) pour montrer que $\mathrm{Aut}^+ A(k,L) = G_0 = \mathrm{Aut}_0(P) \cap \mathrm{Aut}_F(P)$. Puisque $A(k,L)$ est un immeuble, il admet un type. Puisque $A(k,L)$ est la réalisation géométrique de Davis–Moussong d'un système de Coxeter, il vérifie la propriété (P^+). Tous ses murs sont propres par le lemme 4.18. Par hypothèse, L est épais et de Moufang, et pour tout sommet x_0 de $A(k,L)$, le bord du bloc de centre x_0 s'identifie avec L, donc l'hypothèse de 5.23 (2) est vérifiée.

Par le lemme 5.24, si ρ est le morphisme de restriction à un bloc B_0 de centre x_0 des automorphismes de $A(k,L)$ fixant x_0, alors l'image I par ρ du fixateur de x_0 dans $\mathrm{Aut}_F A(k,L)$ est exactement G_{x_0}. L'inclusion de G_{x_0} dans I montre la propriété (P_0) et l'inclusion réciproque montre la deuxième hypothèse de 5.23 (3).

La première hypothèse de 5.23 (3) découle du corollaire 5.27 si $m \geq 3$, et est claire si $m = 2$.

Enfin, par le corollaire 5.25 et la proposition 5.26 (3) si $m \geq 3$, le quotient $\mathrm{Aut}_0 A(k,L)/\mathrm{Aut}^+ A(k,L)$ est isomorphe à $\mathrm{Aut}(K)$. □

6 Simplicité de groupes d'automorphismes d'espaces à murs

Théorème 6.1 *Soient $(X, \mathcal{M})$ un espace à murs hyperbolique, de graphe associé $\mathcal{G}$, et G un groupe d'automorphismes de $(X, \mathcal{M})$, dont l'action sur $\mathcal{G}$ est non élémentaire, d'ensemble limite égal à $\partial\mathcal{G}$. Supposons que G vérifie la* *condition* (P)*. Soit* G^+ *le sous-groupe de* G *engendré par les fixateurs stricts de murs propres et* H *un sous-groupe distingué de* G^+*. Alors ou bien* H *est contenu dans le noyau de l'action de* G^+ *sur* $\partial\mathcal{G}$*, ou bien* H *est égal à* G^+*.*

Corollaire 6.2 *Si l'action de* G^+ *sur* $\partial\mathcal{G}$ *est fidèle, alors* G^+ *est simple.* □

Remarquons que le sous-groupe G^+ est distingué dans G, et qu'il peut être trivial.

Preuve Soit H un sous-groupe distingué non trivial de G^+. Supposons que H n'est pas contenu dans le noyau de l'action de G^+ sur $\partial\mathcal{G}$. Rappelons que X est le sous-ensemble des sommets de $\mathcal{G}$.

Lemme 6.3 *Pour tout demi-espace* A *avec* $\partial X \setminus \partial A$ *non vide, il existe une chaîne propre* $(A_i)_{i\in\mathbb{Z}}$ *et un élément* h *dans* H *tels que* $A \subset A_0 \setminus A_1$*,* $h(A_i) = A_{i+1}$ *pour tout* i*.*

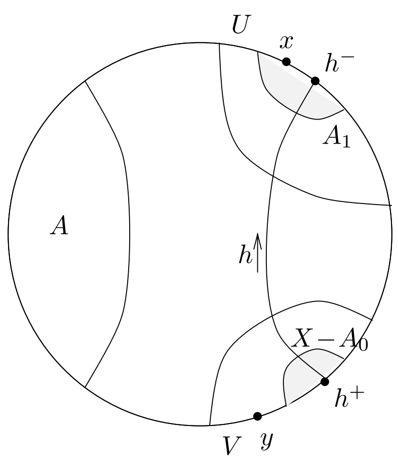


Figure 8: Construction de chaîne invariante par un élément hyperbolique

Preuve Puisque $\Lambda G = \partial\mathcal{G}$ n'a pas de point isolé (G est non élémentaire), et par la condition (H), il existe (voir figure 8):

- x, y deux points distincts dans l'ouvert $\partial X \setminus \partial A = (\overline{X} \setminus \overline{A}) \cap \partial X$ de ∂X,
- U, V deux ouverts disjoints de $\overline{X}$, contenus dans $\overline{X} \setminus \overline{A}$ et contenant respectivement x, y,

- A_1 un demi-espace contenu dans U, avec $\overline{A_1}$ un voisinage de x, et dont le mur est propre.

Par une application double du lemme 2.1 (à $G^+ \subset G$ et à $H \subset G^+$), les couples des points fixes d'éléments hyperboliques de H sont denses dans $\partial^2\mathcal{G}$. Soit donc h un élément hyperbolique de H dont un point fixe au bord est contenu dans l'intérieur de ∂A_1, et l'autre dans V. Quitte à remplacer h par une puissance suffisamment grande (en valeur absolue), pour que $h(A_1)$ soit strictement contenu dans A_1 et que $h^{-1}(X \setminus A_1)$ soit contenu dans V, la suite de demi-espaces $(h^{i-1}(A_1))_{i\in\mathbb{Z}}$ est une chaîne. Le lemme est alors facile à vérifier. □

Lemme 6.4 *Soient $h \in H$ et $C = (A_i)_{i\in\mathbb{Z}}$ une chaîne propre tels que $h(A_i) = A_{i+1}$ pour tout i. Pour tout $g \in G$ fixant strictement C, il existe $f \in G^+$ tel que $g = [h, f]$.*

Preuve On note $[u, v] = uvu^{-1}v^{-1}$. Soient h, g comme dans l'énoncé. Si $u \in G$ fixe strictement les $M_i = \{A_i, X \setminus A_i\}$, notons u_i la restriction de u à $A_i \setminus A_{i+1}$. Alors $g = [h, f]$ si et seulement si, pour tout $i \in \mathbb{Z}$,

$$g_i = hf_{i-1}h^{-1}f_i^{-1}$$

ou encore

$$f_i = g_i^{-1}hf_{i-1}h^{-1}.$$

Posons f_0 la restriction à $A_0 \setminus A_1$ de l'identité de G. Alors la relation de récurrence ci-dessus (ou $f_{i-1} = h^{-1}g_if_ih$ pour les i strictement négatifs) permet de définir une application f_i sur $A_i \setminus A_{i+1}$, qui est par récurrence restriction à $A_i \setminus A_{i+1}$ d'un élément $\tilde{f}_i$ de G fixant strictement C. En effet, le fixateur strict de C, qui contient $\tilde{f}_{i-1}$, est distingué dans le stabilisateur de C (qui contient h). Par la propriété (P), il existe un élément f dans G fixant strictement C, dont les restrictions sont les f_i, et la remarque préliminaire montre que $g = [h, f]$.

Par définition, le fixateur strict dans G d'une chaîne propre est contenu dans G^+. Ceci conclut la preuve. □

Corollaire 6.5 *Le groupe H contient le fixateur strict dans G de tout mur propre.*

Preuve Soit $M = \{A, X \setminus A\}$ un mur propre, donc tel que $\partial X \setminus \partial A$ est non vide. Soit $g \in G$ fixant strictement M. Par le lemme 3.4, pour montrer que g

appartient à H, il suffit de le montrer en supposant de plus que g fixe (point par point) $X\backslash A$. Par le lemme 6.3, il existe $h \in H$ et une chaîne propre $C = (A_i)_{i\in\mathbb{Z}}$ tels que $h(A_i) = A_{i+1}$ et $A \subset A_0 \backslash A_1$. En particulier g fixe strictement C. Par le lemme 6.4, il existe f dans G^+ tel que $g = [h, f] = h(fh^{-1}f^{-1})$. Comme H est distingué dans G^+, il contient g, d'où le résultat. □

Le corollaire 6.5 démontre le théorème. □

7 Applications

Théorème 7.1 *Soit P un complexe polyédral pair* CAT(0), *dont la métrique est hyperbolique au sens de Gromov, dont le groupe des automorphismes est non élémentaire et d'ensemble limite égal à ∂P. Soit* $\mathrm{Aut}^+(P)$ *le sous-groupe de* $\mathrm{Aut}(P)$ *engendré par les fixateurs stricts de murs propres et H un sous-groupe distingué de* $\mathrm{Aut}^+(P)$. *Alors ou bien H est contenu dans le noyau de l'action de G sur $\partial\mathcal{G}$, ou bien H est égal à* $\mathrm{Aut}^+(P)$.

Preuve D'après le théorème 4.17, l'espace à murs $(X_P, \mathcal{M}_P)$ associé à P est un espace à murs hyperbolique, et le bord de P s'identifie au bord du graphe associé à $(X_P, \mathcal{M}_P)$. D'après le théorème 5.1, le groupe des automorphismes de P (resp. le groupe engendré par les fixateurs stricts de murs propres de P) coïncide avec le groupe G des automorphismes de l'espace à murs $(X_P, \mathcal{M}_P)$ (resp. le groupe engendré par les fixateurs stricts de murs propres de $(X_P, \mathcal{M}_P)$). Par le lemme 5.9, l'espace à murs $(X_P, \mathcal{M}_P)$ vérifie la condition (M′). Donc G vérifie la condition (P) par le lemme 3.5. Le résultat découle alors du théorème 6.1. □

Corollaire 7.2 *Sous les hypothèses du théorème précédent:*

(1) *Si P est localement compact alors H est relativement compact, ou égal à* $\mathrm{Aut}^+ P$.

(2) *Si le seul élément de* $\mathrm{Aut}^+ P$ *agissant trivialement sur le bord de P est l'identité, alors* $\mathrm{Aut}^+ P$ *est simple.*

(3) *Si P est* CAT(-1) *et tout point de P est contenu dans une droite géodésique, alors* $\mathrm{Aut}^+ P$ *est simple.*

Preuve Si H est contenu dans le noyau de l'action sur le bord, et si P est localement compact, alors par le lemme 2.2, H est relativement compact.

Sinon, par le théorème précédent on a $H = \mathrm{Aut}^+ P$, ce qui montre (1) et (2). L'assertion (3) découle de (2) par le lemme 2.2, car puisque $\mathrm{Aut}(P)$ est non élémentaire, d'ensemble limite égal à tout $\partial\mathcal{G}$, il n'y a pas de point isolé dans $\partial\mathcal{G}$. □

Le théorème 1.4 de l'introduction découle de ce corollaire et de la remarque précédant le lemme 2.2.

Corollaire 7.3 *Soit (W, S) un système de Coxeter, avec W hyperbolique au sens de Gromov. Alors le quotient, par son sous-groupe distingué localement compact formé des éléments fixant l'infini, du sous-groupe G^+ des automorphismes du graphe de Cayley de (W, S) engendré par les fixateurs stricts de murs propres, est simple. Il est non trivial (et donc non dénombrable) si et seulement si (W, S) est non rigide.*

Preuve D'après la remarque (1) de la section 4.2, le complexe polyédral $|W|_0$ est pair. Il est localement compact, et W agit discrètement avec quotient compact sur lui. Le résultat de simplicité découle du théorème précédent. La dernière assertion découle du théorème 5.12, la non trivialité de $\mathrm{Aut}^+|W|_0$ impliquant sa non dénombrabilité par le lemme 3.6. □

Le théorème 1.3 de l'introduction découle de ce corollaire, car le groupe des automorphismes du graphe de Cayley de (W, S) s'identifie avec le groupe des automorphismes polyédraux de la réalisation géométrique au sens de Davis–Moussong de (W, S) (voir section 5.1).

Pour terminer, démontrons les théorèmes 1.1 et 1.2 de l'introduction. Par le lemme 5.28, le groupe des F-automorphismes préservant le type des immeubles hyperboliques $A(k, L)$ coïncide avec le groupe engendré par les fixateurs strict de murs propres, est d'indice fini dans $\mathrm{Aut}_0 A(k, L)$ et est simple par le corollaire 7.2 (3). Comme L est non rigide (par exemple si $m \geq 3$, un groupe de racine est non trivial et fixe l'étoile d'un sommet), il est non dénombrable, par le lemme 3.6. Il est évidemment fermé dans le groupe de tous les automorphismes, donc est localement compact.

Enfin, pour montrer que $\mathrm{Aut}^+ A(k, L)$ est non linéaire, il suffit, par le théorème de Schur–Kaplansky (voir par exemple [18, page 154]), de montrer qu'il contient un sous-groupe de type fini, de torsion et infini. Supposons que k est multiple de 4 et que L est ou bien un graphe biparti complet $K_{p,p'}$, ou l'immeuble sphérique d'un groupe de Chevalley fini sur un corps F_q de caractéristique p différente de 2. En utilisant les méthodes de l'affirmation 2 de la proposition 5.23, il est alors possible de montrer que G contient une copie du p–groupe infini à deux générateurs τ, α de Grigorchuk–Gupta–Sidki (voir [2, page 19]).

Bibliographie

[1] **E M Andreev**, *On convex polyhedra in Lobachevskii space*, Math. USSR Sb. 12 (1970) 413–440

[2] **G Baumslag**, *Topics in Combinatorial Group Theory*, Lectures in Math. Birkhauser (1993)

[3] **W Ballmann**, **M Brin**, *Polygonal complexes and combinatorial group theory*, Geom. Dedicata 50 Série I (1994) 165-191

[4] **N Benakli**, *Polygonal complexes I: Combinatorial and geometric properties*, J. Pure Appl. Alg. 97 (1994) 247–263

[5] **N Bourbaki**, *Groupes et algèbres de Lie*, chap. 4,5,6, Hermann, Paris (1968)

[6] **M Bourdon**, *Structure conforme au bord et flot géodésique d'un* $CAT(-1)$ *espace*, L'Ens. Math. 41 (1995) 63–102

[7] **M Bourdon**, *Immeubles hyperboliques, dimension conforme et rigidité de Mostow*, GAFA 7 (1997) 245–268

[8] **M R Bridson**, *Geodesics and curvature in metric simplicial complexes*, dans: "Group theory from a geometrical viewpoint" (E Ghys, A Haefliger, A Verjovsky, editeurs) World Scientific (1991) 373–463

[9] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grund. math. Wiss. 319, Springer–Verlag (1998)

[10] **K Brown**, *Buildings*, Springer–Verlag (1989)

[11] **M Burger**, **S Mozes**, *Finitely presented simple groups and products of trees*, C. R. Acad. Sci. Paris, Ser. I, 324 (1997) 747–752

[12] **R W Carter**, *Simple groups of Lie type*, Pure Appl, Math. 28, Wiley (1972)

[13] **C Champetier**, *L'espace des groupes de type fini*, à paraître dans Topology

[14] **R Charney**, **M Davis**, *Singular metrics of nonpositive curvature on branched covers of Riemannian manifolds*, Amer. J. Math. 115 (1993) 929–1009

[15] **M Davis**, *Groups generated by reflections and aspherical manifolds not covered by Euclidean space*, Annals of Math. 117 (1983) 293–324

[16] **M Davis**, *Negative curvature and reflection groups*, to appear in Handbook of Geometric Topology

[17] **M Davis**, **T Januszkiewicz**, **R Scott**, *Nonpositive curvature of blow-ups*, to appear in Selecta Math.

[18] **J Dixon**, *The structure of linear groups*, Van Nostrand (1971)

[19] **D Gaboriau**, **F Paulin**, *Sur les immeubles hyperboliques*, prépublication 18, Univ. Orsay (Janv 1998)

[20] **E Ghys**, **P de La Harpe**, editeurs, *Sur les groupes hyperboliques d'après Mikhael Gromov*, Prog. in Math. 83, Birkhäuser (1990)

[21] **M Gromov**, *Hyperbolic groups*, dans: "Essays in group theory", (S Gersten, editeur) MSRI Pub. 8, Springer–Verlag (1987) 75–263

[22] **A Haefliger**, *Complexes of groups and orbihedra*, dans: "Group theory from a geometrical viewpoint" (E Ghys, A Haefliger, A Verjovsky, editeurs) World Scientific (1991) 504–540

[23] **F Haglund**, *Réseaux de Coxeter–Davis et commensurateurs*, Ann. Inst. Fourier 48 (1998) 649-666

[24] **P de la Harpe**, *An invitation to Coxeter groups*, dans: "Group theory from a geometrical viewpoint", (E Ghys, A Haefliger, A Verjovsky, editeurs) World Scientific (1991) 193–253

[25] **G Moussong**, *Hyperbolic Coxeter group*, Doctoral Dissertation, Ohio State University (1988)

[26] **G Niblo**, **L Reeves**, *Groups acting on $CAT(0)$ cube complexes*, Geometry and Topology 1 (1997) 1–7

[27] **M A Ronan**, *Lectures on buildings*, Persp. Math. 7, Academic Press (1989)

[28] **M Sageev**, *Ends of group pairs and non-positively curved cube complex*, Proc. London Math. Soc. 71 (1995) 585–617

[29] **J-P Serre**, *Arbres, amalgames, SL_2*, Astérisque 46, Soc. Math. France (1983)

[30] **E H Spanier**, *Algebraic topology*, Tata–McGraw–Hill (1966)

[31] **J Tits**, *Sur le groupe des automorphismes d'un arbre*, dans: "Essays on Topology (Mémoires dédiées à G de Rham)", Springer–Verlag (1970) 188–211

Laboratoire de Topologie et Dynamique URA 1169 CNRS
Université Paris-Sud
Bât. 425 (Mathématiques)
91405 ORSAY Cedex
FRANCE

Email: haglund@math.u-psud.fr, Frederic.Paulin@math.u-psud.fr

Received: 17 November 1997 Revised: 29 November 1998

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 249–260

Automatic groups, subgroups and cosets

Derek F Holt

Abstract The history, definition and principal properties of automatic groups and their generalisations to subgroups and cosets are reviewed briefly, mainly from a computational perspective. A result about the asynchronous automaticity of an HNN extension is then proved and applied to an example that was proposed by Mark Sapir.

AMS Classification 20F32; 20F05

Keywords Automatic Groups, HNN extensions

The concept of an automatic group was introduced in 1986 by Thurston, motivated by some results of Jim Cannon on hyperbolic groups. Much of the basic theory of this important class of groups was developed by David Epstein during the following few years.

In the first section of this paper, we review briefly the history, definition and properties of automatic groups and their generalisation to subgroups and cosets, mainly from a perspective of carrying out efficient computations within such groups and their subgroups. In the second section, we prove a result about the (asynchronous) automaticity of an HNN extension, and use it, together with the results of some machine computations, to prove that a particular group, defined by Mark Sapir, is asynchronously automatic.

1 Definitions and discussion

1.1 Automatic groups

In [2], J.W. Cannon proved certain geometrical properties of the Cayley graph of cocompact discrete hyperbolic groups. Two years later, in 1986, W. Thurston noticed that some of these properties could be reformulated in terms of finite state automata (fsa; this abbreviation will be used for both the singular and plural).

In particular, the goedesic paths in the Cayley graph that start at the origin form a regular set or, equivalently, they form the language of an fsa. Furthermore, any pair of such geodesic paths that end at the same or neighbouring vertices lie within a bounded distance of each other. It can be deduced that such geodesic pairs also form the language of an fsa. This led Thurston to formulate the following general definition.

Definition 1.1 Let G be a group with finite generating set X, let $A = X \cup X^{-1}$, and let $A' = A \cup \{\$\}$, where $\$ \notin A$. Then G is said to be *automatic* (with respect to X), if there exist fsa W and M_a for each $a \in A'$, such that

(i) W has input alphabet A, and accepts at least one word in A^* mapping onto each element of G.

(ii) Each M_a has input alphabet $A' \times A'$, it accepts only padded pairs, and it accepts the padded pair (w^+, x^+) for $w, x \in A^*$ if and only if $w, x \in L(W)$ and $wa =_G x$.

Here A^* as usual denotes the set of words in A. For $w \in A^*$, $\overline{w}$ denotes the element of G onto which w maps; for $w, x \in A^*$, we also use $w =_G x$ to mean that w, x map onto the same element of G. The extra symbol \$ maps onto the identity element of G. For $w, x \in A^*$, the associated *padded pair* $(w^+, x^+) \in (A' \times A')^*$ is obtained by adjoining symbols \$ to the end of the shorter of w and x to make them have equal length. The language of the fsa W is denoted by $L(W)$. For general properties of finite state automata, the user is referred to any textbook on automata or formal language theory, such as [10].

In the definition, W is called the *word-acceptor* and the M_a the *multiplier* automata. The complete collection $\{W, M_a\}$ is known as an *automatic structure* for G. Note that the multiplier $M_\$$ recognises equality in G between words in $L(W)$. From a given automatic structure, we can always use $M_\$$ to construct another one such that W accepts a unique word mapping onto each element of G; we simply choose the lexicographically least amongst the shortest words that map onto each element as the 'normal form' representative of that element. We shall call such a W a word-acceptor with uniqueness.

The best general reference for the theory of automatic groups is the multi-author book [3]. In particular, it turns out that the automaticity of G is independent of the choice of generating set X. This immediately suggests that the definition is a sensible one, because it means that automaticity is an algebraic property of the group, rather than just a geometrical property of its Cayley graph.

All finite groups are easily seen to be automatic; in fact the class of automatic groups is invariant under finite variations, such as sub- and super-groups of finite index. It is also closed under direct and free products, and includes, for example, all word-hyperbolic groups, braid groups, Coxeter groups and Artin groups of finite and of 'large' type. All automatic groups have finite presentations.

Some of the most important and useful applications of this theory only involve an explicit knowledge of a word acceptor with uniqueness, particularly in the frequently occurring case when the accepted words are all geodesics in the Cayley graph. From such a word-acceptor, one can quickly enumerate unique representatives of all words up to a given length. This can serve as an invaluable time-saving device in certain computer graphics applications, such as drawing tessellations of hyperbolic space on which these groups act freely. One can also use W to compute the growth function for the group (see [5]).

Another important application of automatic structures for groups G is their use for the efficient (quadratic time) solution of the word problem in G. More precisely, the multiplier automata can be used to reduce an arbitrary word in A^* in quadratic time to the G–equivalent word in $L(W)$.

With these applications in mind, a collection of programs was written at Warwick in the late 1980's for computing automatic structures. These programs take a finite presentation of the group G as input. Currently, they only work for so-called *shortlex* structures, which are those in which $L(W)$ consists of the lexicographically least amongst the shortest words that map onto each group element. (So W depends upon the order of A as well as on A itself.) Many, but not all, of the known classes of automatic groups are known to possess shortlex structures. The programs are described in some detail in [4] and [8], and in a much more general setting in [3]. The latest version is part of a package called `kbmag` and is available by anonymous `ftp` from `ftp.maths.warwick.ac.uk` in the directory `people/dfh/kbmag2`.

From an algorithmic point of view, there is a close connection between automatic groups and rewriting systems for groups, and the programs used make use of the Knuth–Bendix completion process in groups. However, typically, this process alone would not terminate and in fact automatic groups normally have infinite regular rather than finite complete rewriting systems. When the automatic structure is successfully computed it is, in some sense, enabling this infinite regular system to be used to solve the word problem in a manner that is typically at least as efficient as could be done with a finite rewriting system. The idea of trying to use infinite regular rewriting systems for this purpose was first proposed by Gilman in [7].

Given a word-acceptor automaton for a group, it turns out that the existence and properties of the multiplier automata are equivalent to the so-called (synchronous) fellow-traveller property, which was one of the geometrical properties of hyperbolic groups observed originally by J. W. Cannon, and is defined as follows.

For a word $w \in A^*$ we denote the length of w by $l(w)$ and, for $g \in G$, $l(g)$ (or more precisely $l_A(g)$) denotes the length of the shortest word $w \in A^*$ with $\overline{w} = g$. For $t \geq 0$, $w(t)$ denotes the prefix of w of length t when $t \leq l(w)$, and $w(t) = w$ for $t \geq l(w)$. The fellow-traveller property asserts that there exists a constant k such that, for all $w, x \in L(W)$ and $a \in A$ such that $wa =_G x$, and all $t \geq 0$, we have $l_A(\overline{w(t)}^{-1}\overline{x(t)}) \leq k$. In other words, two travellers proceeding at the same speed along the words w and x from the base point in the Cayley graph of G would always remain a bounded distance away from each other.

The fellow-traveller property enables the multiplier automata M_a to be defined in a uniform manner (see Definition 2.3.3 of [3]). Their state set is the set of triples (s_1, s_2, g), where s_1, s_2 are states of W, and $g \in G$ with $l(g) \leq k$. The start state is $(s_0, s_0, 1)$, where s_0 is the start state of W. For $(a_1, a_2) \in A \times A$, there is a transition from (s_1, s_2, g) to (t_1, t_2, h) with label (a_1, a_2) if and only if there are transitions $s_1 \to t_1$ and $s_2 \to t_2$ in W with labels a_1 and a_2, respectively, and if $a_1^{-1} g a_2 =_G h$. The state (s_1, s_2, g) is a success state of M_a if and only if s_1 and s_2 are success states of W, and $g =_G a$. Thus the M_a differ only in their accept states. (We have omitted a technicality from this definition. To deal with the padding symbol, we have to add an extra state to W which is reached when W is in an accept state and the padding symbol is read.) It is clear that the M_a behave precisely according to Condition (ii) of Definition 1.1. This method is used to construct the M_a in the programs mentioned above.

Note also that it follows from the fellow-traveller property that if g is any fixed element of G and $w, x \in L(W)$ with $wg =_G x$, then w and x fellow-travel with constant at most $kl_A(g)$.

Finally, we must mention the weaker concept of an asynchronously automatic group, because it will arise in the next section. The definition is the same as before, except that the multiplier automata are allowed to read their two input strings at different rates. More precisely, rather than reading one symbol from each of the two input words at each transition, they read a symbol from one of the two words only, where the choice of which word to read is a function of the state of M_a. Of course, when the end of one of the words is reached, the other word must be selected. See Chapter 7 of [3] for the formal definition. Again there is a corresponding fellow-traveller property, in which the imaginary travellers are allowed to move at different speeds. See [3] or Section 7, Part II of [1] for details.

The word problem is still solvable for asynchronously automatic groups, but it is unknown whether this can be done in polynomial time. There are examples known, such as the Baumslag–Solitar groups $\langle x, y \mid y^{-1}x^py = x^q \rangle$ with $p \neq q$, which are asynchronously automatic but not automatic.

There is a more detailed treatment, with references to the literature, of the synchronous and asynchronous fellow-traveller properties in groups in the article [16] in these proceedings.

1.2 Subgroups

Let $L = L(W)$ be the language of the word-acceptor in an automatic structure of a group G. A subgroup H of G is called L–rational if $L \cap H$ is a regular language (ie the language of an fsa). Such subgroups were studied in [6], where it is proved that L–rational is equivalent to L–quasiconvex. This means that any prefix of a word in $L \cap H$ lies within a bounded distance of H in the Cayley graph of G. Such subgroups are always finitely generated.

An algorithm for constructing an fsa W_H with language $L \cap H$, which takes as input an automatic structure for G and a set of generators for an L–rational subgroup H of G, is described in [12]. A practical and efficient version is described in [11], and an implementation is available in kbmag.

The fsa W_H can be used together with the automatic structure to determine whether a given word in A^* lies in H; that is, to solve the generalised word problem for H in G. First use the the multiplier automata to reduce the word to one in L, and then use W_H to test whether it lies in H. Given W_H and W_K for two subgroups H and K of G, it is easy to intersect their languages to obtain a fsa $W_{H \cap K}$ for their intersection, which can then be used to construct a finite generating set for $H \cap K$.

1.3 Cosets

It is possible to generalise the concept of an automatic group from a notion about the elements of the group to one about the cosets of a given subgroup H of G. This has been carried out by two doctoral students of the author (see [15] and [11]). The definition is as follows.

Definition 1.2 Let G be a group with finite generating set X, let $A = X \cup X^{-1}$, $A' = A \cup \{\$\}$, and let H be a subgroup of G. Then G is said to be *coset automatic* with respect to H, if there exist fsa W, and M_a for each $a \in A'$, such that:

(i) W has input alphabet A, and accepts at least one word in each right coset of H in G;

(ii) Each M_a has input alphabet $A' \times A'$, it accepts only padded pairs, and it accepts the padded pair (w^+, x^+) for $w, x \in A^*$ if and only if $w, x \in L(W)$ and $H\overline{wa} = H\bar{x}$.

Here W is called the *coset word-acceptor* and the M_a the *coset multiplier* automata. The complete collection $\{W, M_a\}$ is known as an *automatic coset system* for the pair (G, H). Again the existence of such a system turns out to be independent of the generating set X of G, and we can, if we wish, always find a new system in which W accepts a unique word in each right coset.

It is proved in [15] that if L is the language of the shortlex automatic structure of a word-hyperbolic group G (or even the set of all geodesics in the Cayley graph of G), and if the subgroup H is L–quasiconvex, then G is coset automatic with respect to H. In [11] the converse is proved for word-hyperbolic groups, although we shall see from the example in the next section that the converse does not hold in general.

An interesting application to the drawing of limit sets of Kleinian groups is described in [14]. As in the graphical applications of ordinary automatic structures, this involves only the use of W to enumerate unique shortest words in each coset.

An algorithm for computing automatic coset systems in the shortlex case was first described in [15], and was implemented by him as a standalone program. It has the disadvantage that it is not usually possible to prove conclusively that the system computed is correct. A different approach is described in [11]. This does enable the output to be proved correct, but it requires an additional hypothesis, to be described below, for it to work at all. It has the further advantage that it has an optional extension to compute a finite presentation for the subgroup H of G after the automatic coset system has been found. This second algorithm, together with the subgroup presentation facility, has been implemented and is available in kbmag. The theory, implementation details and performance statistics can also be found in [9].

These algorithms provide an alternative method to that described in the previous subsection for solving the generalised word problem for H in G. The given word in $w \in A^*$ is reduced (in quadratic time, using the coset multiplier automata) to the unique word w' in the language of the coset word-acceptor for which $H\overline{w} = H\overline{w'}$. Then $w \in H$ if and only if w' is the empty word. The two methods of solving the generalised word problem are to some extent complementary to each other, since there can exist L–quasiconvex subgroups that are not coset automatic and vice versa, although the two concepts are equivalent in word-hyperbolic groups.

The additional hypothesis required for the algorithm developed by Hurt is the following generalisation of the fellow-traveller condition. Let $\{W, M_a\}$ be the shortlex automatic coset system for (G, H) that we are trying to compute. Then, if $(w^+, x^+) \in L(M_a)$ for some $a \in A$, there exists $h \in H$ such that $wa =_G hx$. The hypothesis is that there exists a constant $k \geq 0$ such that for all such w, x, a and h, and all $t \geq 0$, we have $l_A(\overline{w(t)}^{-1} h\overline{x(t)}) \leq k$. In particular, taking $t = 0$, we get $l_A(h) \leq k$, and so in all such equations, only a finite number of elements h occur.

One step in the algorithm is to define the states of the M_a as triples (s_1, s_2, g), as in the automatic group case, but now the initial states are (s_0, s_0, h), where s_0 is the initial state of W, and h is one of the elements of H occurring in the above equations. So the M_a are in fact constructed initially as non-deterministic automata with multiple initial states,

If the hypothesis holds, then we shall say that G is *strongly coset automatic* with respect to H, and call $\{W, M_a\}$ a *strong automatic coset system* for (G, H). It is proved in [11] that word-hyperbolic groups are always strongly coset automatic with respect to their quasiconvex subgroups. It is easy to construct examples in which the hypothesis does not hold, by choosing H to be normal in G, in which case G coset automatic with respect to H is equivalent to G/H automatic, but we do not know of any example in which $\mathrm{Core}_G(H) = 1$.

2 HNN extensions and an example

For the application to be described in this section, we need to strengthen the hypothesis defined at the end of the preceding section for strong automatic coset systems.

Definition 2.1 Let $\{W, M_a\}$ be a strong automatic coset system for (G, H) with respect to the generating set X of G. Let Y be a finite set of generators of H, and let $B = Y \cup Y^{-1}$. Then Y is said to be efficient with respect to

$\{W, M_a\}$ if, for any $w, x \in L(W)$ and any $b \in B, h \in H$ such that $wb =_G hx$, we have either $h = 1$ or $h \in B$.

We are not currently aware of any particular situations under which an efficient generating set could be shown to exist; it would be interesting to investigate this question. In specific examples of automatic coset systems that we have calculated with the programs, it is often possible to observe directly from the calculation that a particular Y is efficient. The concept is useful to us here, because it enables us to prove the following result about HNN extensions, which can then be applied to a specific example. Note that a rather different condition under which an HNN extension of an automatic group is asynchronously automatic has been proved by Shapiro in [17], and results of a similar nature for amalgamated free products are proved in [1].

Theorem 2.2 *Let $\{W, M_a\}$ be a strong automatic coset system for (G, H), let $G = \langle X \mid R\rangle$ be a finite presentation of G, and suppose that H has the efficient generating set Y. Suppose also that H is automatic, and let α be an automorphism of H such that $\alpha(Y) = Y$.*

Then the HNN extension

$$K = \langle X, z \mid R,\ z^{-1}yz = \alpha(y)\ (y \in Y)\rangle$$

is asynchronously automatic.

Proof Let T be a right transversal for H in G. Then by the normal form theorem for HNN extensions (see, for example, Theorem 2.1 (II), page 182 of [13]), each element of $g \in K$ has a unique expression of the form

$$k = ht_1z^{n_1}t_2z^{n_2}\ldots t_rz^{n_r},$$

where $h \in H$, $t_i \in T$, $n_i \in \mathbb{Z}$, $t_i \notin H$ for $i > 1$ and $n_i \neq 0$ for $i < r$.

We use this normal form in the natural manner to construct a regular language L_K for K on the alphabet $A \cup B \cup \{z^{\pm 1}\}$ where, as before, $A = X \cup X^{-1}$ and $B = Y \cup Y^{-1}$. We are assuming that H is automatic, so we can use the language L_H of the word-acceptor from an associated automatic structure with alphabet B to obtain a word $w_h \in L_H$ for the element $h \in H$ in the normal form. For T we choose the image in G of $L(W)$, and to represent t_i, we choose the unique word $w_i \in L(W)$ with $\overline{w_i} = t_i$. This clearly yields a regular language L_K mapping bijectively onto K.

We now have to show how to construct the asynchronous multiplier automata M_c for $c \in A \cup B \cup \{z^{\pm 1}\}$. Since this is fairly routine, we describe the construction in outline only. Suppose that $u, v \in L_K$ and $uc =_K v$, and let the

HNN normal form of $k = \overline{u}$ be $ht_1z^{n_1}t_2z^{n_2}\ldots t_rz^{n_r}$, as above. If $c = z$ or z^{-1}, then the HNN normal form for kc in K is just $ht_1z^{n_1}\ldots t_rz^{n_r\pm 1}$, and it is easy to construct M_c. So suppose $c \in A \cup B$. We shall suppose that $n_r \neq 0$ and omit the details of the case $n_r = 0$, which are similar. There exist words $c_1 \in B^*$ and $c_2 \in L(W)$ such that $c =_G c_1c_2$. Let $l_B(c_1) = k$. Then, from the assumptions that the generating set Y of H is efficient and that $\alpha(Y) = Y$, it follows that the HNN normal form in K for kc is

$$kc = h't'_1z^{n_1}t'_2z^{n_2}\ldots t'_rz^{n_r}\overline{c_2},$$

where there are elements $x_i, y_i \in H$ $(1 \leq i \leq r)$, all having B–length at most k, such that $z^{n_r}\overline{c_1} = y_rz^{n_r}$, $t_iy_i = x_it'_i$ for $1 \leq i \leq r$, $z^{n_i}x_{i+1} = y_iz^{n_i}$ for $1 \leq i < r$, and $hx_1 = h'$. Thus we have $u = w_hw_1z^{n_1}\ldots w_rz^{n_r}$ and $v = w_{h'}w'_1z^{n_1}\ldots w'_rz^{n_r}c_2$, where $w_h, w_{h'} \in L_H$ map onto $h, h' \in H$, and $w_i, w'_i \in L(W)$ map onto $t_i, t'_i \in T$ for $1 \leq i \leq r$.

The multiplier M_c proceeds by reading the words w_h and $w_{h'}$ in parallel at the same rate, then the z^{n_1} together, then t_1 and t'_1 together, and so on. If either of w_h or $w_{h'}$ is longer than the other, then it will wait at the end of the shorter one until the longer word has been read, and similarly for t_i and t'_i. (This explains why M_c needs to be asynchronous. Although $|l(w_h) - l(w_{h'})|$ and $|l(t_i) - l(t'_i)|$ are all bounded, there is no bound on r, and so one of the two tapes of the input of M_c may conceivably get indefinitely ahead of the other; indeed, we have verified that this really can happen in the example below.)

Of course, if either of the two words input to M_c is not in L_K, or if they do not both have the same pattern with respect to the occurrences of z, then they are rejected. Otherwise, if after t transitions, M_c has read $\phi(t)$ symbols from u and $\psi(t)$ from v, then the element $g(t) = \overline{u(\phi(t))}^{-1}\overline{v(\psi(t))}$ of K is remembered as a function of the state of M_c. As in the synchronous case, it is sufficient to show that $l(g(t))$ is bounded.

There are four essentially different situations that occur as the words u, v are read.

(i) $u(\phi(t))$ and $v(\psi(t))$ are prefixes of w_h and ${}_{h'}$, where $|\phi(t) - \psi(t)|$ is bounded. Then the the boundedness of $l(g(t))$ from the automaticity of H, and the fact that $hx_1 = h'$ with $l(x_1) \leq k$.

(ii) $u(\phi(t)) = w_hw_1z^{n_1}\ldots w_i(s_1)$ for some i and some prefix $w_i(s_1)$ of w_i, and $v(\psi(t)) = w_{h'}w'_1z^{n_1}\ldots w'_i(s_2)$, where $|s_1 - s_2|$ is bounded. Then $g(t) = \overline{w_i(s_1)}^{-1}x_i\overline{w'_i(s_2)}$, and its boundedness follows from the assumptions that $l_B(x_i) \leq k$ and that $\{W, M_a\}$ is a strong automatic coset system for (G, H).

(iii) $u(\phi(t)) = w_h w_1 z^{n_1} \ldots w_i z^{m_1}$ for some i and some $m_1 \le n_i$, and $v(\psi(t)) = w_{h'} w'_1 z^{n_1} \ldots w'_i z^{m_2}$, where $|m_1 - m_2| \le 1$. Then $g(t) z^{-m_1} y_i z^{m_2}$, and its boundedness follows from $l_B(y_i) \le k$ and the assumption that $\alpha(Y) = Y$.

(iv) $\phi(t) > l(u)$ and $\psi(t) \ge l(v) - l(c_2)$. Then $l(g(t)) \le l(c_2)$ which is clearly bounded.

This completes the proof of the theorem. □

As an application, we shall use this theorem together with the results of some machine computations that were done with kbmag, to prove that the group defined by the presentation

$$\langle\, a, b, r, t, x, z \mid xaxa = t, bxbx = t, bbtaa = t, a^{-1}br = ra^{-1}b, zt = tz, btaz = zbta \,\rangle$$

is asynchronously automatic.

This group, which we shall denote by K, was originally proposed by Mark Sapir as a possible building block in his attempts to construct groups with given Dehn functions. However, he later found a different approach to his problem, and so the example is no longer relevant from that viewpoint. He had hoped that it could be proven automatic, but the methods we have been discussing in this paper only appear to be sufficient to prove it asynchronously automatic.

The computer programs could make no progress with the presentation as given above, but matters improved after manipulating it a little. Eliminating $t = bxbx$, we get

$$\langle\, a, b, r, x, z \mid xaxa = bxbx, bbxbxaa = xbx, a^{-1}br = ra^{-1}b, zbxbx = bxbxz, bbxbxaz = zbbxbxa \,\rangle.$$

Now, putting $u = xa$ and $v = bx$, and eliminating $a = x^{-1}u = v^{-1}bu$ and $x = b^{-1}v$, we get

$$\langle\, u, v, b, r, z \mid u^2 = v^2, bvbuv^{-1}bu = b^{-1}v^2, u^{-1}b^{-1}vbr = ru^{-1}b^{-1}vb, zv^2 = v^2z, bvbuz = zbvbu \,\rangle.$$

Finally, using $u^2 = v^2$ to simplify the second relation , we get

$$\langle\, u, v, b, r, z \mid u^2 = v^2, bvbu = b^{-1}ub^{-1}v, u^{-1}b^{-1}vbr = ru^{-1}b^{-1}vb, zu^2 = u^2z, zbvbu = bvbuz \,\rangle,$$

This is now visibly an HNN extension of the group

$$G = \langle\, u, v, b, r \mid u^2 = v^2, bvbu = b^{-1}ub^{-1}v, u^{-1}b^{-1}vbr = ru^{-1}b^{-1}vb \,\rangle.$$

with respect to the subgroup $H = \langle u^2, bvbu \rangle$, where H is centralised by the new generator z. (In fact G is itself an HNN extension with extra generator r, but we shall not make use of that fact.)

Running the automatic coset system program from kbmag on the subgroup H of G verifies that G is strongly coset automatic with respect to H. (The coset word acceptor has 302 states, and the coset multiplers about 1400 states.) The presentation of H computed by the program proves that H is free of rank 2, and so it is certainly automatic. The programs can also be used to verify that the set $Y = \{u^2, bvbu^{-1}\}$ is an efficient generating set for H. (Briefly, this is done by constructing the multiple initial state multiplier automata for the elements u^2 and $bvbu^{-1}$. The elements of H corresponding to the initial states of these automata can then be inspected from the output, and it turns out that these are just the identity and elements of $B = Y \cup Y^{-1}$.) We can now deduce from the theorem that Sapir's group K is asynchronously automatic.

As a final remark about this example, it turns out (again using calculations carried out by kbmag) that the subgroup H is not L–quasiconvex, where L is the language of the word-acceptor of the shortlex automatic structure of G. The element $(bub^{-1}v^{-1})^n(b^{-1}vbu^{-1})^n$ of L lies in H for all $n \geq 0$, but the coset representative of $(bub^{-1}v^{-1})^n$ in the language of the coset word acceptor is b^{2n}.

References

[1] **G Baumslag**, **S M Gersten**, **M Shapiro**, **H Short**, *Automatic groups and amalgams*, J. of Pure and Applied Algebra 76 (1991) 229–316

[2] **J W Cannon**, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata 16 (1984) 123–148

[3] **David B A Epstein**, **J W Cannon**, **D F Holt**, **S Levy**, **M S Patterson**, **W Thurston**, *Word Processing in Groups*, Jones and Bartlett (1992)

[4] **D B A Epstein**, **D F Holt**, **S E Rees**, *The use of Knuth–Bendix methods to solve the word problem in automatic groups*, J. Symbolic Computation 12 (1991) 397–414

[5] **D B A Epstein**, **A R Fletcher**, **U Zwick**, *Growth functions and automatic groups*, Experimental Mathematics 5 (1996) 297–315

[6] **S M Gersten**, **H B Short**, *Rational Subgroups Of Biautomatic Groups*, Annals of Math. 134 (1991) 125–158

[7] **R H Gilman**, '*Presentations of groups and monoids*, J. Algebra 57 (1979) 544–554

[8] **Derek F Holt**, *The Warwick automatic groups software*, from: "Geometrical and Computational Perspectives on Infinite Groups", Gilbert Baumslag et al (editors), DIMACS Series in Discrete Mathematics and Theoretical Computer Science 25 (1995) 69–82

[9] **Derek F Holt**, **Darren F Hurt**, *Computing automatic coset systems and subgroup presentations*, submitted to J. Symbolic Computation

[10] **John E Hopcroft**, **Jeffrey D Ullman**, *Introduction to automata theory, languages and computation*, Addison–Wesley (1979)

[11] **Darren F Hurt**, *The Use of Knuth–Bendix Methods and Automatic Coset Systems for Solving the Generalized Word Problem and Finding Subgroup Presentations*, PhD Thesis, University of Warwick (1996)

[12] **Ilya Kapovich**, *Detecting quasiconvexity: algorithmic aspects*, from: "Geometric and Computational Perspectives on Infinite Groups", Gilbert Baumslag et al (editors), DIMACS Series in Discrete Mathematics and Theoretical Computer Science 25 (1995) 91–99

[13] **R C Lyndon**, **P E Shupp**, *Combinatorial Group Theory*, Springer–Verlag, Berlin–Heidelberg–New York (1977)

[14] **G McShane**, **J Parker**, **I Redfern**, *Drawing Limit Sets of Kleinian Groups Using Finite State Automata*, Experimental Mathematics 3 (1994) 153–172

[15] **I D Redfern**, *Automatic Coset Systems*, PhD Thesis, University of Warwick (1993)

[16] **Sarah Rees**, *Hairdressing in groups: a survey of combings and formal languages*, these proceedings

[17] **M Shapiro**, *Graph products of automatic groups*, unpublished

Mathematics Institute, University of Warwick
Coventry, CV4 7AL, UK.

Email: `dfh@maths.warwick.ac.uk`

Received: 26 August 1997

ISSN 1464-8997 **261**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 261–293

Minimal Seifert manifolds for higher ribbon knots

James Howie

Abstract We show that a group presented by a labelled oriented tree presentation in which the tree has diameter at most three is an HNN extension of a finitely presented group. From results of Silver, it then follows that the corresponding higher dimensional ribbon knots admit minimal Seifert manifolds.

AMS Classification 57Q45; 20E06, 20F05, 57M05

Keywords Ribbon knots, Seifert manifolds, LOT groups

1 Introduction

It is well known that every classical knot k (knotted circle in S^3) bounds a compact orientable surface, known as a *Seifert surface* for the knot. A Seifert surface Σ of minimal genus (among all Seifert surfaces for the given knot k) is called *minimal*, and satisfies the following property: the inclusion-induced map $\pi_1(\Sigma\backslash k) \to \pi_1(S^3\backslash k)$ is injective.

For a higher dimensional knot, or more generally a knotted (closed, orientable) n–manifold M in S^{n+2}, a *Seifert manifold* is defined to be a compact, orientable $(n+1)$–manifold W in S^{n+2}, such that $\partial W = M$. A Seifert manifold W for M is defined to be *minimal* if the inclusion-induced map $\pi_1(W\backslash M) \to \pi_1(S^{n+2}\backslash M)$ is injective. In general, any M will always admit Seifert manifolds, but not necessarily minimal Seifert manifolds. For example, Silver [13] has shown that, for any $n \geq 3$, there exist n–knots in S^{n+2} with no minimal Seifert manifolds, and Maeda [9] has constructed, for all $g \geq 1$, a knotted surface of genus g in S^4 that has no minimal Seifert manifold. Further examples of knotted tori in S^4 without minimal Seifert manifolds are constructed by Silver [16].

A theorem of Silver [14] says that, for $n \geq 3$, a knotted n–sphere K in S^{n+2} has a minimal Seifert manifold if and only if its group $G_K = \pi_1(S^{n+2}\backslash K)$ can be expressed as an HNN extension with a *finitely presented* base group. (It is standard that any higher knot group can be expressed as an HNN extension with a *finitely generated* base group.)

As Silver remarks, the proof of his theorem does not extend to the case $n = 2$. However, it remains a *necessary* condition for the existence of a minimal Seifert manifold that the group be an HNN extension with finitely presented base group. This applies also to knotted n–manifolds in S^{n+2}, a fact which is used implicitly by Maeda in the result mentioned above. It remains an open question whether every 2–knot in S^4 has a minimal Seifert manifold. This seems unlikely, however. For example Hillman [5], p. 139 shows that, provided the 3–dimensional Poincaré Conjecture holds, there is an infinite family of distinct 2–knots, all with the same group G, such that the commutator subgroup of G is finite of order 3; and at most one of these knots can admit a minimal Seifert manifold.

In the present article we consider the case of higher dimensional *ribbon knots*, for which the existence of minimal Seifert manifolds is also an open question. Indeed, as we shall point out in the next section, higher ribbon knot groups are special cases of *knot-like groups*, in the sense of Rapaport [12], and Silver [15] has conjectured that every finitely generated HNN base for a knot-like group is finitely presented. It would therefore follow from Silver's conjecture (and his Theorem) that every higher ribbon knot has a minimal Seifert manifold.

Now any higher ribbon knot group has a Wirtinger-like presentation that can be encoded in the form of a *labelled oriented tree* (LOT) [7]. Indeed the LOT encodes not only a presentation for the knot group, but the complete homotopy type of the knot complement. In [7] it was shown that, if the diameter of the tree is at most 3, then the group is locally indicable, and using this that the 2–complex model of the associated Wirtinger presentation is aspherical. A shorter proof of this fact is given in [8], where it is shown that the presentation is in fact diagrammatically aspherical.

In the present paper, we show that, under the same hypothesis on the diameter of the tree, the group is an HNN extension with finitely presented base group, and hence that the higher ribbon knot has a minimal Seifert manifold.

Theorem 1.1 *Let Γ be a labelled oriented tree of diameter at most 3, and $G = G(\Gamma)$ the corresponding group. Then G is an HNN extension with finitely presented base group.*

Corollary 1.2 *Let K be a ribbon n–knot in S^{n+2}, where $n \geq 3$, such that the associated labelled oriented tree has diameter at most 3. Then K admits a minimal Seifert manifold.*

The paper is arranged as follows. In section 2 we recall some basic definitions relating to LOTs and higher ribbon knots. In section 3 we prove some preliminary results about HNN bases for one-relator products of groups, which will allow us to simplify the original problem. In section 4 we reduce the problem to the study of *minimal* LOTs, In section 5 we construct a finitely generated HNN base B for G, and describe a finite set of relators in these generators. In section 6 we prove some technical results about the structure of these relations, which we apply in section 7 to complete the proof of Theorem 1.1 by proving that this finite set is a set of defining relators for B. We close, in section 8, with a geometric description of our generators and relators for the HNN base, and a discussion of how this might be used to generalise Theorem 1.1.

Acknowledgements

In the course of this work I have received useful comments and advice from Nick Gilbert and from Dan Silver. I am grateful to them for their help.

2 LOTs and higher ribbon knots

A *labelled oriented tree* (LOT) is a tree Γ, with vertex set $V=V(\Gamma)$, edge set $E=E(\Gamma)$, and initial and terminal vertex maps $\iota,\tau\colon E\to V$, together with an additional map $\lambda\colon E\to V$. For any edge e of Γ, $\lambda(e)$ is called the *label* of e. In general, one can consider LOTs of any cardinality, but for the purposes of the present paper, every LOT will be assumed to be finite.

To any LOT Γ we associate a presentation

$$\mathcal{P}=\mathcal{P}(\Gamma):\quad \langle\ V(\Gamma)\mid \iota(e)\lambda(e)=\lambda(e)\tau(e)\ \rangle$$

of a group $G=G(\Gamma)$, and hence also a 2–complex $K=K(\Gamma)$ modelled on $\mathcal{P}$. The 2–complex K is a spine of a *ribbon disk complement* $D^4\backslash k(D^2)$ [7], that is the complement of an embedded 2–disk in D^4, such that the radial function on D^4 composed with the embedding k is a Morse function on D^2 with no local maximum. Conversely, any ribbon disk complement has a 2–dimensional spine of the form $K(\Gamma)$ for some LOT Γ.

By doubling a ribbon disk, we obtain a ribbon 2–knot in S^4, and by successively spinning we can obtain ribbon n–knots in S^{n+2} for all $n\geq 2$. In each case the group of the knot is isomorphic to the fundamental group of the ribbon disk complement that we started with. Conversely, every ribbon n–knot (for $n \geq 2$) can be constructed this way, so that higher ribbon knot groups and LOT groups are precisely the same thing.

Recall [12] that a group G is *knot-like* if it has a finite presentation with deficiency 1 (in other words, one more generator than defining relator), and infinite cyclic abelianisation. It is clear that every LOT group has these properties, so LOT groups are special cases of knot-like groups.

The *diameter* of a finite connected graph Γ is the maximum distance between two vertices of Γ, in the edge-path-length metric. A key factor in our situation is the special nature of trees of diameter 3 or less. For any LOT Γ of diameter 0 or 1, it is easy to see that $G(\Gamma)$ is infinite cyclic, so such LOTs are of little interest.

Remark Every tree of diameter 2 has a single non-extremal vertex. Every tree of diameter 3 has precisely 2 non-extremal vertices.

We recall from [7] that a LOT Γ is *reduced* if:

(i) for all $e \in E$, $\iota(e) \neq \lambda(e) \neq \tau(e)$;

(ii) for all $e_1 \neq e_2 \in E$, if $\lambda(e_1) = \lambda(e_2)$ then $\iota(e_1) \neq \iota(e_2)$ and $\tau(e_1) \neq \tau(e_2)$;

(iii) every vertex of degree 1 in Γ occurs as a label of some edge of Γ.

For every LOT Γ there is a reduced LOT Γ' with the same group as Γ, and the same or smaller diameter, so we may also restrict our attention to reduced LOTs.

A subgraph Γ' of a LOT Γ is *admissible* if $\lambda(e) \in V(\Gamma')$ for all $e \in E(\Gamma')$. If Γ' is connected and admissible, then it is also a LOT. A LOT is *minimal* if every connected admissible subgraph consists only of a single vertex.

If Γ is a LOT and $A \subseteq V(\Gamma)$, we define the *span* of A (in Γ) to be the smallest subgraph Γ' of G such that:

(i) $A \subseteq V(\Gamma')$; and

(ii) if $e \in E(\Gamma)$ with $\lambda(e) \in V(\Gamma')$ and at least one of $\iota(e)$, $\tau(e)$ belongs to $V(\Gamma')$, then $e \in E(\Gamma')$.

We write span(A) for the span of A, and say that A *spans*, or *generates* Γ' if $\Gamma' = \text{span}(A)$. The following is essentially Proposition 4.2 of [7].

Lemma 2.1 *If Γ is a LOT spanned by A, then $\mathcal{P}(\Gamma)$ is Andrews–Curtis equivalent to a presentation with generating set A. If Γ' is an admissible subgraph of Γ with $V(\Gamma') \subseteq A$, then the presentation may be chosen to contain $\mathcal{P}(\Gamma')$, and the Andrews–Curtis moves can be taken relative to $\mathcal{P}(\Gamma')$.*

Corollary 2.2 *If Γ is a LOT spanned by two vertices, then $G(\Gamma)$ is a torsion-free one-relator group.*

Proof Let A be a set of two vertices spanning Γ. Then $\mathcal{P}(\Gamma)$ is Andrews–Curtis equivalent to a presentation $\langle A|R\rangle$. Since $\mathcal{P}(\Gamma)$ has deficiency 1, the same is true of the equivalent presentation $\langle A|R\rangle$. In other words $|R| = 1$, and $G(\Gamma)$ is a one-relator group. But the abelianisation G^{ab} of G is infinite cyclic, so the relator $r \in R$ cannot be a proper power, and so G is torsion-free. □

We will require the following generalisation of Corollary 2.2. Recall that a *one-relator product* of two groups A, B is the quotient of the free product $A * B$ by the normal closure of a single word w, called the *relator*.

Corollary 2.3 *If Γ is a LOT spanned by $V(\Gamma') \cup \{x\}$, where Γ' is an admissible subgraph of Γ and x is a vertex in $V(\Gamma) \backslash V(\Gamma')$, then $G(\Gamma)$ is a one-relator product of $G(\Gamma')$ and $\mathbb{Z}$, where the relator is not a proper power.*

Proof Let $A = V(\Gamma') \cup \{x\}$ and apply the Theorem. Then $\mathcal{P}(\Gamma)$ is equivalent, relative to $\mathcal{P}(\Gamma')$, to a presentation $\mathcal{Q}$ with generating set A and containing $\mathcal{P}(\Gamma')$. Now each of $\mathcal{P}(\Gamma)$, $\mathcal{P}(\Gamma')$ and $\mathcal{Q}$ has deficiency 1. Moreover, $\mathcal{Q}$ has one more generator than $\mathcal{P}(\Gamma')$, so $\mathcal{Q}$ also has one more defining relator than $\mathcal{P}(\Gamma')$. It follows that $G(\Gamma)$ is a one relator product of $G(\Gamma')$ with the infinite cyclic group $\langle x\rangle$. Finally, since the abelianisations of $G(\Gamma)$, $G(\Gamma')$ and $\langle x\rangle$ are all infinite cyclic, it follows that the relator cannot be a proper power. □

3 One-relator groups and one-relator products

The following result is merely a summary of some well-known properties of one-relator groups, which have useful applications to our situation. Recall that a group G is *locally indicable* if, for every nontrivial, finitely generated subgroup H of G, there exists an epimorphism $H \to \mathbb{Z}$.

Theorem 3.1 *Let G be a finitely generated one-relator group. Then*

(i) *G is either a finite cyclic group, or an HNN extension of a finitely presented, one-relator group (with shorter defining relator);*

(ii) *if the defining relator of G is not a proper power, then G is locally indicable.*

Proof See [11] and [3] respectively. □

In order to complete the process of reducing ourselves to a simple special case, we require a generalisation of the above theorem to one-relator products. Suppose that A and B are locally indicable groups, and $N = N(w)$ is the normal closure in $A * B$ of a cyclically reduced word w of length at least 2 that is not a proper power. Then the one-relator product $G = (A * B)/N$ is known [6] to be locally indicable. We show also that G has a finitely presented HNN base, provided that A and B also have this property.

Theorem 3.2 *Let $G = (A * B)/N(w)$ be a one-relator product of two finitely presented, locally indicable groups A and B, each of which has a finitely presented HNN base. Suppose also that G^{ab} is infinite cyclic, with each of the natural maps $A^{ab} \to G^{ab}$ and $B^{ab} \to G^{ab}$ an isomorphism. Then G is a finitely presented, locally indicable group with a finitely presented HNN base.*

Remark The condition on G^{ab} in this theorem is unnecessary for the proof that G has a finitely presented HNN base. It can be removed at the expense of a less straightforward proof. However the condition does hold for all the groups that we are considering in this paper, so there is no loss of generality for us in imposing that condition. The condition also ensures that w cannot be a proper power, so that G is locally indicable by the results of [6].

Proof A presentation for G can be obtained by taking the disjoint union of finite presentations for A and for B, and imposing the single additional relation $w = 1$. Hence G is finitely presented. As pointed out in the remark above, w cannot be a proper power, so G is locally indicable by [6]. It remains only to prove that G has a finitely presented HNN base.

Let

$$A = \langle A_0, a | a^{-1} g a = \alpha(g) \ (g \in A_1) \rangle$$

and

$$B = \langle B_0, b | b^{-1} h b = \beta(h) \ (h \in B_1) \rangle$$

be HNN presentations for A and B with finitely presented bases A_0 and B_0 respectively. Since A and B are finitely presented, it follows also that the associated subgroups A_1 and B_1 are finitely generated.

The commutator subgroup G' of G can be expressed in the form

$$(A' * B' * \langle\ c_n\ (n \in \mathbb{N})\rangle)/N(\{w_n\ (n \in \mathbb{N})\}),$$

where $c_n = a^{n+1}b^{-1}a^{-n}$ and $w_n = a^{-n}wa^n$.

Now A' is an infinite stem product

$$\cdots \quad (a^{-1}A_0a) \underset{(a^{-1}A_1a)}{*} A_0 \underset{A_1}{*} (aA_0a^{-1}) \quad \cdots$$

Since A_0 is finitely presented and A_1 is finitely generated, the subgroup

$$(a^{-k}A_0a^k) \underset{(a^{-k}A_1a^k)}{*} \cdots\cdots \underset{(a^{k-1}A_1a^{1-k})}{*} (a^kA_0a^{-k})$$

is finitely presented for each k. Moreover it is also an HNN base for A. Replacing A_0 by this subgroup, for any sufficiently large k, we may assume that $w_0 \in A_0 * B' * \langle\ c_n\ (n \in \mathbb{N})\rangle$.

Similarly, possibly after replacing B_0 by a sufficiently large finitely presented HNN base for B, we may assume that $w_0 \in A_0 * B_0 * \langle\ c_n\ (n \in \mathbb{N})\rangle$. Now let μ and ν be the least and greatest indices i such that c_i occurs in w_0. (Note that at least one c_i occurs in w_0, for otherwise $w_0 \in A_0 * B_0$, so $w \in A' * B'$, whence $G^{ab} \cong A^{ab} \times B^{ab} \ncong \mathbb{Z}$, a contradiction.) Define $G_0 = (A_0 * B_0 * \langle c_\mu, \ldots, c_\nu\rangle)/N(w_0)$ and $G_1 = A_0 * B_0 * \langle c_\mu, \ldots, c_{\nu-1}\rangle$, and observe that G_0 is a finitely presented HNN base for G, with associated subgroup G_1. □

4 Reduction of the problem

Recall from section 2 that a LOT Γ is *minimal* if it contains no admissible subtree with more than one vertex. In this section we reduce the proof of the main theorem to the case of a minimal LOT of diameter 3, using the results of section 3. The key point is that a non-minimal LOT can be obtained from a minimal admissible subtree by successively expanding to the span of the existing tree with one extra vertex. By Corollary 2.3, this construction corresponds at the group level to taking a one-relator product of a given group with an infinite cyclic group.

Lemma 4.1 *Let Γ be a LOT of diameter at most 3, containing a proper admissible subtree with more than one vertex. Then there is such an admissible subtree Γ' and a vertex $x \in V(\Gamma)\backslash V(\Gamma')$ such that Γ is spanned by $V(\Gamma')\cup\{x\}$.*

Proof Suppose first that some extremal vertex x of Γ does not occur as a label of any edge of Γ. In this case we take Γ' to consist of Γ with the vertex x and the edge incident to x removed. Clearly Γ' is connected, so a subtree of Γ. Since x is not the label of any edge in $E(\Gamma')$, it follows that Γ' is admissible. Moreover Γ is spanned by $V(\Gamma) = V(\Gamma') \cup \{x\}$, as required.

We may therefore assume that every extremal vertex of Γ occurs at least once as the label of an edge of Γ.

Next suppose that Γ has a proper admissible subtree that contains all the non-extremal vertices of Γ. Let Γ' be a maximal such admissible subtree. The vertices in $V(\Gamma)\backslash V(\Gamma')$ are all extremal in Γ, so occur as labels of edges of Γ. But since Γ' is admissible, no such vertex can be a label of an edge of Γ'. Since the finite sets $V(\Gamma)\backslash V(\Gamma')$ and $E(\Gamma)\backslash E(\Gamma')$ have the same cardinality, it follows that each vertex in $V(\Gamma)\backslash V(\Gamma')$ is the label of precisely one edge in $E(\Gamma)\backslash E(\Gamma')$. In turn, this edge has precisely one endpoint in $V(\Gamma)\backslash V(\Gamma')$, so we can define a permutation σ on $V(\Gamma)\backslash V(\Gamma')$ by defining $\sigma(x)$ to be the extremal endpoint of the unique edge labelled x, for all $x \in V(\Gamma)\backslash V(\Gamma')$. Now fix some vertex $x \in V(\Gamma)\backslash V(\Gamma')$, let t be the size of the orbit of σ that contains x, and define $x_i = \sigma^i(x)$, $i = 1, \ldots, t$. Now $\Delta = \mathrm{span}(V(\Gamma') \cup \{x\})$ contains the vertex $x = x_t$, together with any non-extremal vertex of Γ. Hence Δ contains the edge labelled x_t, and hence its endpoint x_1. Similarly Δ contains $x_2, \ldots, x_{t-1}$, as well as the edges labelled $x_1, \ldots, x_{t-1}$. On the other hand, The vertices $x_1, \ldots, x_t$, the edges labelled by them, and the vertices and edges of Γ' together form an admissible subtree of Γ, which by maximality of Γ' must be the whole of Γ. Hence $\Delta = \Gamma$, in other words Γ is spanned by $V(\Gamma') \cup \{x\}$.

Finally, suppose that no proper admissible subtree of Γ contains all the non-extremal vertices of Γ. In particular, Γ must have more than one non-extremal vertex, so has diameter 3. By hypothesis, there is a proper admissible subtree Γ' of Γ that contains more than one vertex. Hence Γ' contains precisely one of the two nonextremal vertices of Γ, say u. As an abstract graph, Γ is the union of Γ' with another tree Γ'', such that $\Gamma' \cap \Gamma'' = \{u\}$. Note that Γ'' contains both of the non-extremal vertices of Γ, so cannot be an admissible subtree, by hypothesis. Hence at least one edge f of Γ'' is labelled by a vertex a of Γ' (other than u). Let e be the edge of Γ that joins the two non-extremal vertices u, v, and let $\Delta = \mathrm{span}(V(\Gamma') \cup \{\lambda(e)\})$. Then Δ contains Γ' and the edge e, and hence v, and hence the edge f. Each extremal vertex of Δ is the label of an edge of Γ, and hence of Δ, since Δ contains at least one endpoint (namely u or v) of every edge of Γ. Moreover there are $|E(\Gamma')|+1$ edges of Δ labelled by the $|V(\Gamma')| = |E(\Gamma')|+1$ vertices of Γ', so an easy counting argument shows that there must be at least $|V(\Delta)|-1$ edges in Δ. In other words Δ is a tree, so the whole of Γ. In other words Γ is spanned by $V(\Gamma') \cup \{\lambda(e)\}$. □

Remark If Γ is a minimal LOT of diameter 2, then the above argument still applies (to the subtree consisting of only the unique non-extremal vertex). In this case we see that the permutation σ is transitive, and that Γ is spanned by two vertices.

Lemma 4.2 *Let Γ be a minimal LOT of diameter 3, and let u, v be the two non-extremal vertices of Γ. Then one of the following holds:*

(i) *One of u, v is a label in Γ, and Γ is spanned by $\{u, v\}$;*

(ii) *Some vertex a occurs twice as a label in Γ, and Γ is spanned by $\{a, u, v\}$.*

Proof By minimality of Γ, every extremal vertex of Γ occurs as a label. There are $|V|-2$ extremal vertices, and $|V|-1$ edges, so either one of u, v occurs as a label or some unique extremal vertex a occurs twice as a label. Note that every edge of Γ is incident to at least one of u, v, so if $u, v \in A \subset V$ then every edge labelled by a vertex of $\mathrm{span}(A)$ is an edge of $\mathrm{span}(A)$.

(i) Suppose that u occurs as a label, and let $\Gamma' = \mathrm{span}(\{u, v\})$. If Γ' has $k+2$ vertices $u, v, x_1, \ldots, x_k$, then $x_1, \ldots, x_k$ are all extremal in Γ, so each of $u, x_1, \ldots, x_k$ is a label of an edge of Γ, which must therefore be an edge of Γ'. Hence Γ' has at least $k-1$ edges, so is connected. By minimality of Γ we have $\Gamma = \Gamma' = \mathrm{span}(\{u, v\})$.

(ii) Suppose that an extremal vertex a appears twice as a label, and let $\Gamma' = \mathrm{span}(\{a, u, v\})$. If Γ' has $k+3$ vertices $a, u, v, x_1, \ldots, x_k$, then each of $x_1, \ldots, x_k$ is extremal, so the label of an edge of Γ, while a is the label of 2 edges of Γ. Each of these $k+2$ edges is an edge of Γ', so Γ' is connected, and by minimality again we have $\Gamma = \Gamma' = \mathrm{span}(\{a, u, v\})$. □

Corollary 4.3 *If Γ is either a minimal LOT of diameter 2, or a minimal LOT of diameter 3 in which no vertex occurs twice as a label, then $G(\Gamma)$ is a locally indicable group with a finitely presented HNN base.*

Proof By Lemma 4.2 or the remark following Lemma 4.1, Γ is spanned by two vertices. Hence $G = G(\Gamma)$ is a 2–generator, one-relator group. Since G^{ab} is infinite cyclic, G is not finite, and the relator of G cannot be a proper power. The result follows immediately from Theorem 3.1. □

Using the above results, we can reduce our problem to the case of a minimal LOT of diameter 3 that is not spanned by two vertices. In particular, some extremal vertex must occur twice as a label.

Corollary 4.4 *If the group of every reduced, minimal LOT of diameter 3 which is not spanned by two vertices is locally indicable with finitely presented HNN base, then the same is true for every LOT of diameter 3 or less.*

Recall [7] that the *initial graph* $I(\Gamma)$ of Γ is the graph with the same vertex and edge sets as Γ, but with incidence maps ι, λ. Similarly the *terminal graph* $T(\Gamma)$ of Γ has the same vertex and edges sets as Γ, but incidence maps λ, τ. It was shown in [7] that the commutator subgroup of $G(\Gamma)$ is locally free if either $I(\Gamma)$ or $T(\Gamma)$ is connected. (If $I(\Gamma)$ and $T(\Gamma)$ are both connected, then $G(\Gamma)'$ is free of finite rank.) In particular, any finitely generated HNN base for $G(\Gamma)$ is free, so automatically finitely presented.

Hence we can concentrate attention on the case of a minimal LOT Γ of diameter 3, not spanned by any two of its vertices, such that neither $I(\Gamma)$ nor $T(\Gamma)$ is connected. Our next result gives a detailed description of the structure of $I(\Gamma)$. In particular it will show us that $I(\Gamma)$ has precisely two connected components, one containing each of the nonextremal vertices of Γ. A similar statement holds for $T(\Gamma)$.

Lemma 4.5 *Let Γ be a minimal LOT of diameter 3, with nonextremal vertices u and v, and an extremal vertex a that occurs twice as a label of edges of Γ. Then:*

(i) *u and v are sources in $I(\Gamma)$;*

(ii) *no vertex other than u or v is the initial vertex of more than one edge of $I(\Gamma)$;*

(iii) *a is the terminal vertex of precisely two edges of $I(\Gamma)$;*

(iv) *each vertex other than a, u, v is the terminal vertex of precisely one edge of $I(\Gamma)$;*

(v) *any directed cycle in $I(\Gamma)$ contains a;*

(vi) *each component of $I(\Gamma)$ contains at least one of u, v;*

(vii) $I(\Gamma)$ *has at most two connected components.*

Proof (i) Since $\lambda(e) \neq u$ for all $e \in E(\Gamma)$, u is not the terminal vertex of any edge in $I(\Gamma)$, in other words u is a source. Similarly v is a source in $I(\Gamma)$.

(ii) Any vertex x of Γ, with the exception of u and v, is extremal in Γ, so the initial vertex of at most one edge of Γ. Hence x is also the initial vertex of at most one edge in $I(\Gamma)$.

(iii) $a = \lambda(e)$ for precisely two edges $e \in E(\Gamma)$.

(iv) If $x \in V(\Gamma)\backslash\{a, u, v\}$ then $x = \lambda(e)$ for precisely one edge $e \in E(\Gamma)$.

(v) Suppose $(e_1, e_2, \ldots, e_n)$ is a directed cycle in $I(\Gamma)$. Then there are vertices $x_1, \ldots, x_n \in V(\Gamma)$ with $x_i = \iota(e_i)$ for all i, $\lambda(e_i) = x_{i+1}$ for $i < n$, and $\lambda(e_n) = x_1$. Now each x_i is extremal since it occurs as a label. If no x_i is equal to a then we can remove the vertices $x_1, \ldots, x_n$ and the edges $e_1, e_2, \ldots, e_n$ from Γ to form a connected, admissible subgraph Γ' that contains at least three vertices (a, u, v). This contradicts the minimality of Γ, and so $x_i = a$ for some i, as claimed.

(vi) By (iv) if $x \notin \{a, u, v\}$ then x is the terminal vertex in $I(\Gamma)$ of a unique edge. If the initial vertex of this edge is not one of a, u, v then it also is the terminal vertex of a unique edge. Continuing in this way, we can construct a directed path that ends at x, and either begins at one of a, u, v or contains a cycle. By (v) any directed cycle contains a, so in any case we have a directed path from one of a, u, v to x. It suffices therefore to find a path in $I(\Gamma)$ from u or v to a. But a is the terminal vertex in $I(\Gamma)$ of precisely two edges, with initial vertices x_1 and x_2 say. Now apply the above argument to each of x_1, x_2. If there is a path from u or v to x_1 or x_2 then we are done. Otherwise there are directed paths from a to each of x_1, x_2. Neither u nor v can belong to these paths, since they are sources in $I(\Gamma)$. But then from (ii) it follows that there is at most one directed path of any given length beginning at a, whence $x_1 = x_2$, a contradiction. Hence there is a directed path in $I(\Gamma)$ from u or v to a, as claimed.

(vii) This follows immediately from (vi). □

A similar result holds for $T(\Gamma)$.

Lemma 4.6 *Let Γ be a minimal LOT of diameter 3, with nonextremal vertices u and v, and an extremal vertex a that occurs twice as a label of edges of Γ. Then:*

(i) *u and v are sinks in $T(\Gamma)$;*

(ii) *no vertex other than u or v is the terminal vertex of more than one edge of $T(\Gamma)$;*

(iii) *a is the initial vertex of precisely two edges of $T(\Gamma)$;*

(iv) *each vertex other than a,u,v is the initial vertex of precisely one edge of $T(\Gamma)$;*

(v) *any directed cycle in $T(\Gamma)$ contains a;*

(vi) *each component of $T(\Gamma)$ contains at least one of u,v;*

(vii) *$T(\Gamma)$ has at most two connected components.*

Corollary 4.7 *Suppose that Γ is a reduced, minimal LOT of diameter 3, which is not spanned by two vertices, and such that neither $I(\Gamma)$ nor $T(\Gamma)$ is connected. Then*

(i) *There is a unique extremal vertex a of Γ that is the label of two distinct edges of Γ. One of these edges has an extremal initial vertex, and the other has an extremal terminal vertex.*

(ii) *$I(\Gamma)$ has precisely two connected components, each containing one of the two nonextremal vertices u,v of Γ.*

(iii) *There is a unique cycle in $I(\Gamma)$, which is either a directed cycle containing a, or consists of two directed paths (one of length 1, the other of length at least 2), from u or v to a.*

(iv) *$T(\Gamma)$ has precisely two connected components, each containing one of the two nonextremal vertices u,v of Γ.*

(v) *There is a unique cycle in $T(\Gamma)$, which is either a directed cycle containing a, or consists of two directed paths (one of length 1, the other of length at least 2), from a to u or v.*

(vi) *The cycles in $I(\Gamma)$ and $T(\Gamma)$ are not both directed.*

Proof (i) We already know that there is an extremal vertex a occurring twice as a label, by Lemma 4.2, since otherwise Γ can be spanned by two vertices. We also know that a is unique, since every extremal vertex occurs at least once as a label. Now suppose that neither of the edges labelled a has extremal initial vertex. The initial vertices of these two edges must be distinct, since Γ is reduced, and so must be the two nonextremal vertices u,v of Γ. But then there are edges of $I(\Gamma)$ from both u and v to a. Hence u and v belong to the same connected component of $I(\Gamma)$. By Lemma 4.5, (vi) it follows that $I(\Gamma)$ is connected, a contradiction.

A similar contradiction arises if neither edge has an extremal terminal vertex.

(ii) This is just a restatement of Lemma 4.5, (vi), together with the hypothesis that $I(\Gamma)$ is not connected.

(iii) Since $I(\Gamma)$ has the same vertex and edge sets as Γ, it has the same euler characteristic, namely 1. Since $I(\Gamma)$ has two components, it follows that $H_1(\Gamma) \cong \mathbb{Z}$, so there is a unique cycle in $I(\Gamma)$. If this cycle is directed, then it must contain a, by Lemma 4.5, (v). Otherwise it must contain at least two vertices at which the orientation of the edges of the cycle changes. This is possible only at a vertex which is either the initial vertex of at least two edges or the terminal vertex of at least two edges, and by Lemma 4.5 the only such vertices are a, u, v. Let us assume that a is in the same component of $I(\Gamma)$ as u. Then the cycle must contain both a and u, and indeed must consist of two directed paths from u to a. By uniqueness of the cycle (or directly from Lemma 4.5), we see that there only two directed paths in $I(\Gamma)$ from u to a. Moreover, precisely one of these paths is of length 1, since precisely one of the edges of Γ labelled a has a nonextremal initial vertex.

(iv) Similar to (ii).

(v) Similar to (iii).

(vi) If the cycle in $I(\Gamma)$ is directed, then there is an edge of $I(\Gamma)$ with initial vertex a, and so also there is an edge of Γ with initial vertex a. Similarly, if the cycle in $T(\Gamma)$ is directed, then there is an edge of Γ with terminal vertex a. Since a is extremal in Γ, these cannot both occur. □

5 Construction of the HNN base

In this section, we construct a presentation of a group that will turn out to be an HNN base for G. As a first step, we fix names for the various vertices of Γ. Throughout we make the following assumptions:

- Γ is a minimal LOT of diameter 3, which cannot be spanned by fewer than three vertices.
- The non-extremal vertices of Γ are u and v.
- The unique vertex of Γ that appears twice as a label is a.
- Of the edges labelled a, one has its initial vertex in $\{u, v\}$ and its terminal vertex extremal, while the other has its initial vertex extremal and its terminal vertex in $\{u, v\}$.

- Neither $I(\Gamma)$ nor $T(\Gamma)$ is connected.

We know from Lemma 4.2 that Γ is then spanned by $\{a,u,v\}$. Let Δ denote the subtree of Γ whose vertex set is $\{a,u,v\}$. We give inductive definitions of two sequences $\{b_1,b_2,\dots,b_P\}$ and $\{c_1,c_2,\dots,c_Q\}$ of vertices of Γ, and two sequences $\{e_0,\dots,e_P\}$, $\{f_0,\dots,f_Q\}$ of edges of Γ as follows.

Define e_0 to be the edge of Γ whose label is a and whose terminal vertex is in $\{u,v\}$. For $i\geq 0$, assume inductively that e_i has been defined. If e_i is an edge of Δ, then we define $P=i$ and stop the construction of the sequences $\{b_1,b_2,\dots,b_P\}$ and $\{e_0,\dots,e_P\}$. Otherwise e_i joins one of $\{u,v\}$ to an extremal vertex other than a, and we define b_{i+1} to be that extremal vertex, and e_{i+1} to be the unique edge of Γ labelled b_{i+1}.

Similarly, define f_0 to be the edge of Γ whose label is a and whose initial vertex is in $\{u,v\}$. For $i\geq 0$, assume inductively that f_i has been defined. If f_i is an edge of Δ, then we define $Q=i$ and stop the construction of the sequences $\{c_1,c_2,\dots,c_Q\}$ and $\{f_0,\dots,f_Q\}$. Otherwise f_i joins one of $\{u,v\}$ to an extremal vertex other than a, and we define c_{i+1} to be that extremal vertex, and f_{i+1} to be the unique edge labelled by c_{i+1}.

Note that the $P+Q+3$ vertices $\{u,v,a,b_1,\dots,b_P,c_1,\dots,c_Q\}$ and the $P+Q+2$ edges $\{e_0,\dots,e_P,f_0,\dots,f_Q\}$ together form an admissible subgraph of Γ, which has euler characteristic 1 and hence is connected, and hence by minimality of Γ must be the whole of Γ. In other words

$$V=V(\Gamma)=\{u,v,a,b_1,\dots,b_P,c_1,\dots,c_Q\},$$

and

$$E=E(\Gamma)=\{e_0,\dots,e_P,f_0,\dots,f_Q\}.$$

We also introduce the following notation. For $i=1,\dots,P$, x_i denotes the unique non-extremal vertex of Γ (ie $x_i\in\{u,v\}$) incident with the edge e_{i-1}. For $i=1,\dots,Q$, y_i denotes the unique non-extremal vertex of Γ incident with the edge f_{i-1}. In other words, x_i is the vertex adjacent to b_i in Γ, and y_i is the vertex adjacent to c_i.

Lemma 5.1 (i) *If $x_2=\ldots=x_P=u$, then $x_1=v$ and e_P is incident at v.*

(ii) *If $x_2=\ldots=x_P=v$, then $x_1=u$ and e_P is incident at u.*

(iii) *If $y_2=\ldots=x_Q=u$, then $y_1=v$ and f_Q is incident at v.*

(iv) *If $y_2=\ldots=y_Q=v$, then $y_1=u$ and f_Q is incident at u.*

Proof We prove (i). The other proofs are similar.

Suppose first that $x_1 = x_2 = \ldots = x_P = u$, and consider the subgraph $\Gamma_0 = \mathrm{span}\{a, u\}$ of Γ. Since $\lambda(e_0) = a$ and e_0 is incident to u, we have $e_0 \in E(\Gamma_0)$, and since b_1 is an endpoint of e_0 we have $b_1 \in V(\Gamma_0)$. Similarly $e_1 \in E(\Gamma_0)$ and $b_2 \in V(\Gamma_0)$, and so on, until $e_P \in E(\Gamma_0)$. If e_P is incident with v, then $v \in V(\Gamma_0)$, and since Γ is spanned by $\{a, u, v\}$ it follows that $\Gamma = \Gamma_0$ is spanned by $\{a, u\}$, a contradiction. Otherwise, e_P joins a to u, in which case the vertices $a, u, p_1, \ldots, b_P$ and the edges $e_0, \ldots, e_P$ form an admissible subtree of Γ of diameter two, which again is a contradiction.

Now suppose that $x_1 = v$ and $x_2 = \ldots = x_P = u$, and let $\Gamma_0 = \mathrm{span}\{b_1, u\}$. Arguing as above, we see that Γ_0 contains the edges $e_1, \ldots, e_{P-1}$ and the vertices $u, b_1, \ldots, b_P$. If e_P is not incident at v, then it joins u to a, so e_P and a also belong to Γ_0. But then e_0 joins b_1 to v and has label a, so we also have $v \in V(\Gamma_0)$. Hence $\Gamma = \Gamma_0$ since Γ is spanned by $\{a, u, v\}$, and so Γ is spanned by $\{b_1, u\}$, a contradiction. □

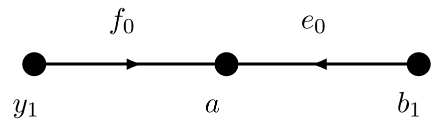
We next subdivide each of the sequences $\{b_i\}$, $\{c_i\}$ into two subsequences, depending on the orientation of the edges labelled by these vertices. Specifically, let:

- $p(1), \ldots, p(s)$ be the sequence, in ascending order, of integers i such that $0 < i \leq P$ and $b_i = \tau(e_{i-1})$;
- $p'(1), \ldots, p'(s')$ be the sequence, in ascending order, of integers i such that $0 < i \leq P$ and $b_i = \iota(e_{i-1})$;
- $q(1), \ldots, q(t)$ be the sequence, in ascending order, of integers i such that $0 < i \leq Q$ and $c_i = \iota(f_{i-1})$; and
- $q'(1), \ldots, q'(t')$ be the sequence, in ascending order, of integers i such that $0 < i \leq Q$ and $c_i = \tau(f_{i-1})$.

For consistency of notation in what follows, we set $p(0) = p'(0) = q(0) = q'(0) = 0$.

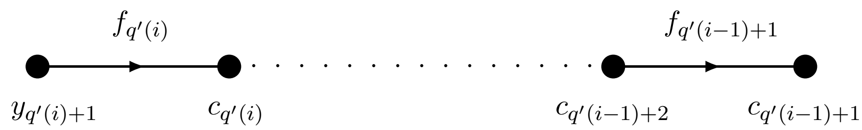
Thus each b_i, for $i = 1, \ldots, P$, can be written uniquely as $b_{p(j)}$ or as $b_{p'(j)}$, and each c_i, for $i = 1, \ldots, Q$, can be written uniquely as $c_{q(j)}$ or as $c_{q'(j)}$.

This notation allows us to give a more precise description of the structure of the initial and terminal graphs of Γ. Specifically, $I(\Gamma)$ is constructed from the vertices $\{a, u, v\}$ by adding two edges

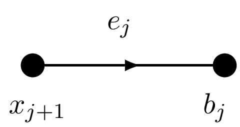


together with directed chains

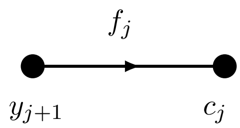
for each $i = 1, \dots, s$, and



for each $i = 1, \dots, t'$; and finally single edges



for $p(s) < j \leq P$ and



for $q'(t') < j \leq Q$.

In the above diagrams x_{P+1} and y_{Q+1} (which have not been defined) should be interpreted as $\iota(e_P)$ and $\iota(f_Q)$ respectively. Note that at most one of these is equal to a. (This happens if and only if a is the initial vertex of its incident edge in Γ.) All other x_j and y_j belong to $\{u, v\}$.

If $I(\Gamma)$ contains a directed cycle, for example, then this cycle must contain a. From the above, we see that this can happen only if $s = 1$, $p(1) = P$, and $x_{P+1} = a$.

The structure of $T(\Gamma)$ is entirely analogous, and similar remarks apply. We omit the details.

Now we are ready to construct a specific presentation for an HNN base for $G = G(\Gamma)$. Recall that G is given by a finite presentation

$$\mathcal{P}(\Gamma) = \langle V(\Gamma) \mid \iota(e)\lambda(e) = \lambda(e)\tau(e),\ e \in E(\Gamma)\rangle.$$

Since Γ is connected, we have $G^{ab} \cong \mathbb{Z}$, and the commutator subgroup G' is the normal closure in G of the subgroup $B = B(\Gamma)$ generated by the finite set $\{xy^{-1}\ ;\ x,y \in V(\Gamma)\}$. A theorem of Bieri and Strebel [2] says that G is an HNN extension of B with stable letter t (which can be taken to be any element of $V(\Gamma)$) and associated subgroups $A_0 = B \cap tBt^{-1}$ and $A_1 = B \cap t^{-1}Bt$:

$$G = \langle B, t \mid t^{-1}\alpha t = \phi(\alpha),\ \alpha \in A_0\rangle,$$

where $\phi\colon A_0 \to A_1$ is the isomorphism induced by conjugation by t.

Clearly B is finitely generated. It remains to prove that B is finitely presentable, and we do this by constructing an explicit set of defining relators.

Recall that our assumptions on Γ imply that each of $I(\Gamma)$ and $T(\Gamma)$ has precisely two connected components, with the vertices u, v belonging to separate components in each case.

Let F denote the subgroup of the free group on $V(\Gamma)$ generated by

$$\{xy^{-1}\ ;\ x,y \in V(\Gamma)\}.$$

Then F is free of rank $|V(\Gamma)| - 1 = |E(\Gamma)|$, and any basis for F can be chosen as a finite generating set for B. Rather than fix a specific basis for F, we proceed as follows. Let $\bar{K} = \bar{K}(\Gamma)$ be the maximal abelian cover of the 2–complex $K = K(\Gamma)$ associated to Γ (which is the standard 2–complex model of the presentation $\mathcal{P}(\Gamma)$). Then since K has a single 0–cell, we identify the 0–cells of $\bar{K}$ with integers, via the isomorphism $H_1(K) \cong G^{ab} \cong \mathbb{Z}$. The 1–cells of $\bar{K}$ with initial vertex $i \in \mathbb{Z}$ can be denoted w_i, where $w \in V(\Gamma)$, and each w_i has terminal vertex $i+1 \in \mathbb{Z}$. Let L be the 1–subcomplex of $\bar{K}$ with 0–cells $0, 1$ and 1–cells $\{w_0,\ w \in V(\Gamma)\}$. Then F is naturally identified with $\pi_1(L, 0)$.

We also construct a graph $\hat{L}$ and an immersion $\pi\colon \hat{L} \to L$ as follows. $V(\hat{L}) = \{0,1\} \times \{u,v\}$, $E(\hat{L}) = E(L)$, $\iota(w_0) = (0,x)$ where $x \in \{u,v\}$ belongs to the same component of $I(\Gamma)$ as w, and $\tau(w_0) = (1,y)$ where $y \in \{u,v\}$ belongs to the same component of $T(\Gamma)$ as w. The graph homomorphism π is defined to be the identity map on edges, and is defined on vertices by $\pi(i,u) = \pi(i,v) = i$, $i = 0,1$. It is not difficult to see that $\hat{L}$ is connected. Indeed, if the edge of

Γ between u and v has label w, then the edges u, v, w of $\hat{L}$ form a spanning tree. Since π is bijective on edges, it is an immersion, and hence injective on fundamental groups. Indeed, the fundamental group $\hat{F}$ of $\hat{L}$ embeds as a free factor of $F = \pi_1(L)$ via π_*, as we can see by the following construction: add an edge X to $\hat{L}$ with $\iota(X) = (0, u)$ and $\tau(X) = (0, v)$, and an edge Y with $\iota(Y) = (1, u)$, $\tau(Y) = (1, v)$, to form a larger graph $\tilde{L}$. The immersion $\pi\colon \hat{L} \to L$ extends to a homotopy equivalence $\pi\colon \tilde{L} \to L$ that shrinks the edge X to the vertex 0, and the edge Y to the vertex 1. Hence we have

$$F = \pi_1(L) \cong \pi_1(\tilde{L}) = \pi_1(\hat{L}) * \langle X, Y \rangle.$$

Since the map $\pi\colon \hat{L} \to L$ is bijective on edges, any path in L which lifts to a path in $\hat{L}$ does so uniquely. Given a closed path C in L that lifts to a closed path $\hat{C}$ in $\hat{L}$, we define two related paths in L, namely the *forward derivative* $\partial_+ C$ of C and the *backward derivative* $\partial_- C$ of C, as follows. For $\partial_+ C$ we first fix a maximal subforest Φ_I of $I(\Gamma)$. Next, we cyclically permute $\hat{C}$ so that it begins and ends at one of the vertices $(1, u)$ or $(1, v)$. Hence $\hat{C}$ is a concatenation of length two subpaths of the form $x^{-1}y$, where $x, y \in E(\hat{L}) = V(\Gamma)$ belong to the same component of $I(\Gamma)$. The next step is to replace each such subword $x^{-1}y$ by the product

$$(x^{-1}z_0)(z_0^{-1}z_1)\dots(z_m^{-1}y),$$

where $(x, z_0, z_1, \dots, z_m, y)$ is the geodesic from x to y in Φ_I. We now have a concatenation of length 2 subwords of the form $x^{-1}y$ where x and y are joined by an edge in Φ_I. This edge corresponds to an edge of Γ, and hence to a defining relation in $\mathcal{P}(\Gamma)$ that can be written

$$x^{-1}y = gh^{-1}$$

for some $g, h \in V(\Gamma)$. The final step is to replace each such word $x^{-1}y$ by the corresponding word gh^{-1}. The result is a closed path $\partial_+ C$ in L.

Remarks (i) $\partial_+ C$ depends on the choice of maximal forest Φ_I, and then is well-defined only up to cyclic permutation.

(ii) If C' is a cyclic permutation of C, then C' also lifts to a closed path in $\hat{L}$, so $\partial_+ C'$ is defined. It is equal to (a cyclic permutation of) $\partial_+ C$.

(iii) The definition of $\partial_+ C$ does not depend on C being (cyclically) reduced. Indeed the insertion into C of a cancelling pair xx^{-1} may alter $\partial_+ C$. However, the insertion of a cancelling pair $x^{-1}x$ will *not* alter $\partial_+ C$.

(iv) C and $\partial_+ C$ are (freely) homotopic in $\bar{K}$ (since the last part of the construction involves replacing a path $x^{-1}y$ by a homotopic path gh^{-1}). In particular, if C is nullhomotopic in $\bar{K}$, then so is $\partial_+ C$.

(v) The unique lift of $\partial_+ C$ in $\tilde{L}$ does not contain the edge Y.

The backward derivative $\partial_- C$ is defined similarly. This time we fix a maximal forest Φ_T of $T(\Gamma)$, and choose a cyclic permutation of $\hat{C}$ beginning at $(0,u)$ or $(0,v)$, split $\hat{C}$ into subpaths of the form xy^{-1} with x,y in the same component of $T(\Gamma)$, and then use relations of $\mathcal{P}$ corresponding to edges of Φ_T to transform $\hat{C}$. Remarks analogous to the above hold also for $\partial_- C$.

Now consider the unique cycle in $T(\Gamma)$. If $z_0,\dots,z_m$ are the vertices of this cycle in cyclic order, define $\hat{R}_0$ to be the nullhomotopic path

$$(z_m z_0^{-1})(z_0 z_1^{-1})\dots(z_{m-1} z_m^{-1})$$

in $\hat{L}$ and $R_0 = \pi(\hat{R}_0)$ the corresponding nullhomotopic path in L. Now define $R_1 = \partial_- R_0$. If R_1 lifts to $\hat{L}$ then define $R_2 = \partial_- R_1$, and so on. In this way we obtain either an infinite sequence $R_1, R_2, \dots$ of paths in L, or a finite sequence $R_1,\dots,R_M$ such that R_M does not lift to $\hat{L}$.

In a similar way, the unique cycle in $I(\Gamma)$ determines a nullhomotopic closed path S_0 in L that lifts to $\hat{L}$, so a sequence $S_1,\dots$ of closed paths in L (finite or infinite), such that $S_i = \partial_+ S_{i-1}$ for each $i \geq 1$, and if the sequence is finite with final term S_N then S_N does not lift to $\hat{L}$.

Lemma 5.2 *The paths R_i and S_j are all nullhomotopic in $\bar{K}$.*

Proof This follows by induction and Remark (iv) above, since R_0 and S_0 are nullhomotopic. □

Now suppose that the sequence $\{R_i\}$ contains at least m terms. We construct a 2–complex L_m as follows. The 1–skeleton of L_m is the subcomplex of $\bar{K}$ consisting of L, together with the 0–cells $2,\dots,m+1$ and the 1–cells $u_1, v_1, \dots, u_m, v_m$. Then L_m has precisely m 2–cells attached to L using the paths $R_1,\dots,R_m$. We also consider the full subcomplex $\bar{K}_m$ of $\bar{K}$ on the set $\{0,1,\dots,m+1\}$ of 0–cells.

Lemma 5.3 *The 2–complexes L_m and $\bar{K}_m$ are homotopy equivalent.*

Proof We argue by induction on m, there being nothing to prove in the case $m=0$. Let γ denote the covering transformation of $\bar{K}$ that sends a 0–cell $n\in\mathbb{Z}$ to $n+1$. Note that the link of the 0–cell $m+1$ in $\bar{K}_m$ is naturally identifiable with the graph $T(\Gamma)$. Let d be the unique edge in $E(\Gamma)=E(T(\Gamma))$ that does not belong to the maximal forest $\Phi_T \subset T(\Gamma)$. Then d is contained in the unique cycle in $T(\Gamma)$, so R_0 has a subword xy^{-1}, where x, y are the endpoints of d in $T(\Gamma)$. Corresponding to d is a relator $xy^{-1}h^{-1}g$ in $\mathcal{P}$, which lifts to a 2–cell α with boundary path $x_m y_m^{-1} h_{m-1}^{-1} g_{m-1}$ in $\bar{K}_m$. Modulo the other 2–cells of $\bar{K}_m$, the boundary path of α is homotopic to $\gamma^m(R_0)^{-1} \cdot \gamma^{m-1}(R_1)$. Since R_0 is nullhomotopic in the 1–skeleton of $\bar{K}$, this is in fact homotopic to $\gamma^{m-1}(R_1)$. This in turn is homotopic (in $\bar{K}_{m-1}$) to $\gamma^{m-2}(R_2)$, etc. Repeating this argument, we see that the boundary path of α is homotopic in $\bar{K}_m \backslash \alpha$ to R_m. A simple homotopy move allows us to replace α by a 2–cell whose boundary path is R_m.

The link of $m+1$ in the resulting 2–complex K' is then isomorphic to $T(\Gamma)\backslash d = \Phi_T$. Since Φ_T is a forest with two components (one containing u and the other containing v), it collapses to the graph with no edges and vertex set $\{u, v\}$. Each move in this collapsing process (removing a vertex and an edge from the graph) can be mirrored by a collapse in the 2–complex K' (removing a 1–cell and a 2–cell that are incident at the 0–cell $m+1$). After performing all these collapsing moves, we are left with a 2–complex K'', simple homotopy equivalent to $\bar{K}_m$. By inspection, K'' is formed from $\bar{K}_{m-1}$ by adding a 2–cell with boundary path R_m, a 0–cell $m+1$, and two 1–cells u_m, v_m, each joining m to $m+1$.

By inductive hypothesis, $\bar{K}_{m-1}$ is homotopy equivalent to L_{m-1}, so $\bar{K}_m$ is homotopy equivalent to the 2–complex obtained from L_{m-1} by adding a 2–cell with boundary path R_m, a 0–cell $m+1$, and two 1–cells u_m, v_m, each joining m to $m+1$. But this 2–complex is precisely L_m, and the proof is complete. □

Remark An analogous result holds for the S_j. We omit the details, but will use this result implicitly in what follows.

Corollary 5.4 *If $R_1, \ldots, R_m$ and $S_1, \ldots, S_n$ are all defined, then $m + n < |V(\Gamma)|$.*

Proof By the Lemma and its analogue for the S_j, $\bar{K}_m$ is homotopy equivalent to a 2–complex formed from L by attaching m 2–cells and then wedging on m circles; and $\gamma^{-n}(\bar{K}_n)$ is homotopy equivalent to a complex obtained from L by adding n 2–cells and then wedging on n circles. Since $\gamma^{-n}(\bar{K}_{m+n}) = \gamma^{-n}(\bar{K}_n) \cup \bar{K}_m$, with $\gamma^{-n}(\bar{K}_n) \cap \bar{K}_m = \bar{K}_1 = L$, it follows that $\gamma^{-n}(\bar{K}_{m+n})$ is homotopy equivalent to a complex formed from L by adding $m+n$ 2–cells and then wedging on $m+n$ circles. Hence $\beta_1(\bar{K}_{m+n}) \geq m+n$. Now $H_2(K) = 0$, and $\bar{K}$ is a $\mathbb{Z}$–cover of K, so $H_2(\bar{K}) = 0$ by [1], Proposition 1. Hence also $H_2(K') = 0$

for any subcomplex $K' \subseteq K$. In particular $H_2(\bar{K}_{m+n}) = 0 = H_2(L)$. Since also $H_0(\bar{K}_{m+n}) = \mathbb{Z} = H_0(L)$ and $\chi(\bar{K}_{m+n}) = \chi(L) = 2 - |V(\Gamma)|$, it follows that

$$m + n \leq \beta_1(\bar{K}_{m+n}) = \beta_1(L) = |V(\Gamma)| - 1. \qquad \square$$

Corollary 5.5 *Each of the sequences $\{R_i\}$ and $\{S_j\}$ are finite, and if the final terms are R_M and S_N respectively then $M + N < |V(\Gamma)|$.*

We claim that the finite sequences $\{R_i\}$ and $\{S_j\}$ form a full set of defining relators for the HNN base B of G, which completes the proof of our Theorem 1.1. In order to prove this claim, we need to derive some further information about the structure of the words R_i and S_j.

Remark The definitions of R_i and S_i depend, *a priori*, on specific choices for the maximal forests Φ_T and Φ_I respectively. Suppose we were to choose a different maximal tree Φ'_I in $I(\Gamma)$, for example. Then geodesics in Φ_I and Φ'_I would differ at most by the unique cycle in $I(\Gamma)$. It follows from this that the resulting definitions of $\partial_+ C$, for any closed path C in L that lifts to $\hat{L}$, are equal modulo the normal closure of S_1. An easy induction shows that, for any i, the definitions of S_i resulting from different choices of Φ_I are equal modulo the normal closure of $\{S_1, \ldots, S_{i-1}\}$. Hence our set of defining relators does not depend in an essential way upon the choices of maximal forests Φ_I and Φ_T.

6 Structure of the relations

In this section we examine the structure of the proposed defining relators R_i and S_i of the HNN base B for G. Recall that each of R_i and S_i is a closed path in the 2–complex L, and that we have a homotopy equivalence $\pi\colon \tilde{L} \to L$, which restricts to an edge-bijective graph immersion on $\hat{L} = \tilde{L}\backslash\{X, Y\}$ and shrinks each of the 1–cells X, Y to a point. Let $\tilde{C}$ denote the unique (up to cyclic permutation) cyclically reduced closed path in $\tilde{L}$ that maps to a given cyclically reduced closed path C in L. Then C lifts to $\hat{L}$ if and only if $\tilde{C}$ is a path in $\hat{L}$, in which case $\tilde{C}$ is the unique lift. By definition, each R_i (resp S_i) is defined if and only if R_{i-1} (resp S_{i-1}) lifts to $\hat{L}$. Hence $\tilde{R}_i$ is a path in $\hat{L}$ for $1 \leq i \leq M-1$, and S_i is a path in $\hat{L}$ for $1 \leq i \leq N-1$. Moreover, the path $\tilde{R}_M$ involves Y but not X, while the path $\tilde{S}_N$ involves X but not Y.

For any group A and letter Z, we say that a word $w \in A * \langle Z \rangle$ is *positive* (resp *negative*) in Z if only positive (resp negative) powers of Z occur in w. We say that w is *strictly positive* (resp *strictly negative*) if in addition at least one positive (resp negative) power of Z does occur in w, in other words $w \notin A$.

We will concentrate our attention on the relators S_i. The analysis of the R_i is entirely analogous.

We first treat the case where $I(\Gamma)$ contains a directed cycle C.

Theorem 6.1 *Suppose that the unique cycle C in $I(\Gamma)$ is directed. Then:*

- $N = 1$;
- $\tilde{S}_1$ *is either strictly positive or strictly negative in* X;
- S_1 *involves each of* $a, b_1, \dots, b_P$ *exactly once, and no* c_j;
- *each of* $a, b_1, \dots, b_P$ *is an extremal source in* Γ.

Proof The vertex a is contained in C, by Lemma 4.5, (v). Since $\iota(f_0) \in \{u, v\}$, f_0 is not an edge of C, so the edge of C coming into a is e_0. Hence $b_1 = \iota(e_0)$ is a vertex of C, and since e_1 is the only edge with $\lambda(e_1) = b_1$, it is also an edge of C, and so on. Hence each of $b_1, \dots, b_P$ are vertices of C, $\iota(e_P) = a$, and the edges of C are precisely $e_P, \dots, e_0$ (in the order of the orientation of C). Each of the vertices of C is extremal in Γ, and since it is the initial vertex of an edge of $I(\Gamma)$ it is also the initial vertex of an edge of Γ, ie a source in Γ. Moreover

$$S_0 = (a^{-1} b_P)(b_P^{-1} b_{P-1}) \dots (b_1^{-1} a),$$

so

$$S_1 = \partial_+ S_0 = (b_P \tau(e_P)^{-1})(b_{P-1} x_P^{-1}) \dots (b_1 x_2^{-1})(a x_1^{-1}),$$

where each $x_i \in \{u, v\}$.

Suppose that S_1 lifts to $\hat{L}$. Then $\tau(e_P)$ belongs to the same component of $I(\Gamma)$ as b_{P-1}, x_P to the same component as b_{P-2}, and so on. Since $a, b_1, \dots, b_P$ all belong to the same component of $I(\Gamma)$, it follows that the x_i also all belong to the same component. But u and v belong to different components of $I(\Gamma)$, and so the x_i are all equal, which contradicts Lemma 5.1.

Hence S_1 does not lift to $\hat{L}$, and so $N = 1$. Moreover, by the above argument, some of the x_i belong to the opposite component of $I(\Gamma)$ from a. If a, u belong to the same component of $I(\Gamma)$, this means that some of the x_i are equal to v. Then $\tilde{S}_1$ is formed from S_1 by replacing each occurrence of v^{-1} by $v^{-1} X^{-1}$, and so $\tilde{S}_1$ is strictly negative in X. Similarly, if a, v belong to the same component of $I(\Gamma)$, then $\tilde{S}_1$ is strictly positive in X. □

For the rest of the section, we can assume that the cycle C is not directed. Then $y_1 = \iota(f_0) = \iota(e_{p(1)}) \in \{u, v\}$. We may assume that $y_1 = u$. Then C has the form

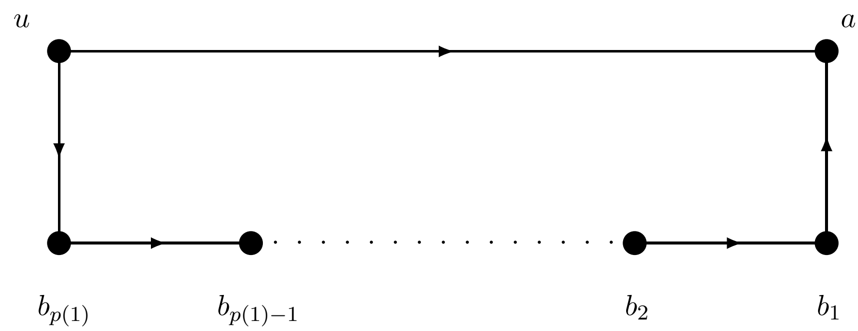


Figure 1

For the purpose of defining forward derivatives, and hence the S_i, we fix Φ_I to be the maximal subforest of $I(\Gamma)$ obtained by removing the edge f_0 (the edge joining u to a in C).

For $k \le \min(s, t' + 1)$, let $I_k(\Gamma)$ denote the subgraph of Φ_I consisting of the edges $\{e_i, 0 \le i \le p(k)\}$ and $\{f_i, 1 \le i \le q'(k-1)\}$, together with all their incident vertices. Note that I_k contains no more than two components, one contained in each component of Φ_I. Hence whenever two vertices of I_k belong to the same component of Φ_I, then the geodesic between them is also contained in I_k.

Theorem 6.2 *Suppose that the cycle in $I(\Gamma)$ has the form shown in Figure 1. Then:*

(i) *Each S_i can be written, up to cyclic permutation, in the form $aU_ia^{-1}V_i$, where U_i is a word in*

$$\{a, u, v, c_1, \ldots, c_{q'(i-1)+1}\};$$

and V_i is a word in

$$\{a, u, v, b_1, \ldots, b_{p(i)+1}\}.$$

(ii) *If $p(i) < P$, then V_i contains a single occurrence of $b_{p(i)+1}$ and does not contain a.*

(iii) *If $q'(i-1) < Q$, then U_i contains a single occurrence of $c_{q'(i-1)+1}$ and does not contain a.*

(iv) *Every letter occurring in S_i, other than $b_{p(i)+1}$ and $c_{q'(i-1)+1}$, is a vertex of the subgraph $I_i \subseteq I(\Gamma)$.*

(v) *If* $p(i)=P$ *or* $q'(i-1)=Q$ *then* $i=N$.

Proof We prove this by induction on i, the initial case being when $i=1$. We have

$$S_0=(u^{-1}a)(a^{-1}b_1)(b_1^{-1}b_2)\dots(b_{p(1)}^{-1}u),$$

so

$$S_1=\partial_+S_0=(ac_1^{-1})(x_1a^{-1})(x_2b_1^{-1})\dots(x_{p(1)}b_{p(1)-1}^{-1})(b_{p(1)+1}b_{p(1)}^{-1})$$

(if $p(1)<P$). The vertices $a,u,b_1,\dots,b_{p(1)}$ are contained in I_1, but not c_1, $b_{p(1)+1}$. The first four statements of the result (for $i=1$) follow, setting $U_1=c_1^{-1}x_1$ and

$$V_1=(x_2b_1^{-1})\dots(x_{p(1)}b_{p(1)-1}^{-1})(b_{p(1)+1}b_{p(1)}^{-1}).$$

For the last statement, certainly $Q>0=q'(0)$. Suppose that $p(1)=P$ and $i<N$. Then

$$S_1=(ac_1^{-1})(x_1a^{-1})(x_2b_1^{-1})\dots(x_Pb_{P-1}^{-1})(\tau(e_P)b_P^{-1})$$

lifts to $\hat{L}$, so each of $x_2,\dots,x_P$ belongs to the same component of $I(\Gamma)$ as $a,b_1,\dots,b_{P-1}$, in other words $x_2=\dots=x_P=u$. By Lemma 5.1 we have $x_1=v$ and e_P incident with v. But $\iota(e_P)=u$ so $\tau(e_P)=v$, which does not belong to the same component of $I(\Gamma)$ as b_{P-1}. It follows that S_1 does not, after all, lift to $\hat{L}$, a contradiction.

This completes the proof of the initial case of the induction.

Now assume inductively that $i>1$ and the result is true for $i-1$. In particular, $i-1<N$, so $p(i-1)<P$ and $q'(i-2)<Q$. Hence U_{i-1} contains a single occurrence of $c_{q'(i-2)+1}$, V_{i-1} contains a single occurrence of $b_{p(i-1)+1}$, and every other letter occurring in S_{i-1} is a vertex of the subgraph I_{i-1} of $I(\Gamma)$. Consider the construction of $S_i=\partial_+S_{i-1}$ from S_{i-1}. We first write a suitable cyclic permutation of S_{i-1} as a product of length two subwords of the form $g^{-1}h$. For all but two of these subwords, both g and h are vertices of I_{i-1}. (There are precisely two exceptions, since the occurrences of $b_{p(i-1)+1}$ and $c_{q'(i-2)+1}$ in S_{i-1} are separated at least by an occurrence of $a^{\pm 1}$.)

Suppose first that g,h are vertices of I_{i-1}. The next step is to replace $g^{-1}h$ by the product

$$(g^{-1}z_1)(z_1^{-1}z_2)\dots(z_t^{-1}h)$$

where $g,z_1,z_2,\dots,z_t,h$ are the vertices on the geodesic from g to h in Φ_I. This geodesic is contained in I_{i-1}, so each bracketed term here is $(\iota(e)^{-1}\lambda(e))^{\pm 1}$ for some edge e of I_{i-1}. The final step is to replace this by $(\lambda(e)\tau(e)^{-1})^{\pm 1}$. Note that $\tau(e)$ is a vertex of I_i, and $\tau(e) \neq a$. Also, none of the intermediate vertices z_i in the geodesic is equal to a, since a is an extremal vertex of Φ_I. Note that, if $g^{-1}h$ is a subword of U_{i-1}, then all letters in the resulting subword of S_i come from $\{u, v, c_1, \ldots, c_{q'(i-1)}\}$, while if it is a subword of $a^{-1}V_{i-1}a$ then all letters come from $\{a, u, v, b_1, \ldots, b_{p(i)}\}$.

A similar argument holds if, say $g = b_{p(i-1)+1}$. Here, however, the geodesic from g to h is not contained in I_{i-1}. It is the union of the geodesic from $b_{p(i-1)+1}$ to z in I_i, where $z \in \{u, v\}$, with the geodesic (in I_{i-1}) from z to h. Edges in I_{i-1} give rise to length 2 subwords of S_i consisting of letters which are vertices in I_i. The same is true for an edge e_j from b_j to b_{j+1}, for $p(i-1) < j < p(i)$. (The corresponding word is $x_j b_j^{-1}$.) Finally, the edge $e_{p(i)}$ (from $b_{p(i)}$ to z) contributes a subword $\tau(e_{p(i)}) b_{p(i)}^{-1}$. If $p(i) < P$ then $\tau(e_{p(i)}) = b_{p(i)+1}$; otherwise $\tau(e_{p(i)}) \in \{a, u, v\}$.

The analysis if $h = b_{p(i-1)+1}$, or if one of g, h is $c_{q'(p-2)+1}$ is similar to the above.

Each of the two subwords $g^{-1}h$ of S_{i-1} that contain the letter a gives rise to a subword of S_i containing an occurrence of a with the same exponent. If $g = a$ then the subword begins $(x_1 a^{-1})\ldots$, while if $h = a$ then the subword ends $\ldots(a x_1^{-1})$. If $p(i) < P$ and $q'(i-1) < Q$ then this will be the only occurrence of a in this subword of S_i.

Statements (i)–(iv) follow.

To prove (v), suppose for example that $i < N$ and $p(i) = P$. Another induction on i shows that $x_2 = \ldots = x_P = u$. An argument similar to that given above in the initial case of the induction again gives rise to a contradiction: by Lemma 5.1, $\tau(e_P) = v$, which does not belong to the same component of $I(\Gamma)$ as b_{P-1}, so S_i does not lift to $\hat{L}$ and $i = N$.

If $i < N$ and $q'(i-1) = Q$ then a similar argument applies. Here we can show that $y_1 = \ldots = y_Q = x_1 \in \{u, v\}$, which contradicts Lemma 5.1. □

This result contains all the necessary information about S_i if $i < N$. We now need to investigate further the structure of $\tilde{S}_N$, particularly as regards occurrences of X. Note that, up to cyclic permutation, we have $\tilde{S}_N = a\tilde{U}_N a^{-1}\tilde{V}_N$, by Theorem 6.2 (i).

Lemma 6.3 *Each of $\tilde{U}_N$, $\tilde{V}_N$ is either positive or negative in X.*

Proof As indicated in the proof of Theorem 6.2, all of V_N, except for the part arising from the geodesic γ from $b_{p(N-1)+1}$ to u, consists of letters which are vertices in I_{N-1}. All of these vertices are in the same component of $I(\Gamma)$ as u. The part of V_N arising from γ is

$$[(x_{p(N-1)+2}b^{-1}_{p(N-1)+1})\dots(x_{p(N)}b^{-1}_{p(N)-1})(\tau(e_{p(N)})b^{-1}_{p(N)})]^{\pm 1},$$

or, if γ passes through a (ie if $\iota(e_{p(N)})=a$):

$$[(x_{p(N-1)+2}b^{-1}_{p(N-1)+1})\dots(\tau(e_{p(N)})b^{-1}_{p(N)})(x_1a^{-1})\dots(b_{p(1)+1}b^{-1}_{p(1)})]^{\pm 1}.$$

The expression in square brackets is a product of terms gh^{-1} with h in the same component of $I(\Gamma)$ as u. To lift to $\tilde{L}$, we replace $h^{-1}g$ by $h^{-1}Xg$ whenever g belongs to the same component of $I(\Gamma)$ as v and h to the same component as u, and by $h^{-1}X^{-1}g$ if g belongs to the same component as u and h to the same component as v. Hence $\tilde{V}_N$ is either positive or negative in X

A similar argument applies to $\tilde{U}_N$, replacing u by x_1 in the above. □

We will also need to investigate possible occurrences of a in S_N other than those indicated in Theorem 6.2.

Lemma 6.4 *The words $\tilde{U}_N$ and $\tilde{V}_N$ contain in total at most one occurrence of a.*

Proof From the discussion in the proof of Lemma 6.3, the word V_N (and hence also $\tilde{V}_N$) contains a single occurrence of a if $e_{p(N)}$ is incident with a in Γ, and no occurrence of a otherwise. Similarly U_N (and hence also $\tilde{U}_N$) contains a single occurrence of a if $f_{q'(N-1)}$ is incident with a in Γ, and no occurrence of a otherwise. The result now follows from the fact that a is extremal in Γ. □

7 Completion of the proof

Define

$$G_0=\pi_1(\hat{L})/\{R_1,\dots,R_{M-1},S_1,\dots,S_{N-1}\},$$
$$G_+=(G_0*\langle X\rangle)/\{\tilde{S}_N\},$$
$$G_-=(G_0*\langle Y\rangle)/\{\tilde{R}_M\},$$

and

$$G_1=(G_0*\langle X,Y\rangle)/\{\tilde{R}_M,\tilde{S}_N\}\cong(\pi_1(L))/\{R_1,\dots,R_M,S_1,\dots,S_N\}.$$

Lemma 7.1 *The group G_0 is free.*

Proof By Theorems 6.1 and 6.2, and the analogous results for the R_i, the set of $M+N-2$ distinct numbers $\mathcal{B} = \{p(1)+1, \ldots, p(N-1)+1, p'(0)+1, \ldots, p'(M-2)+1\}$ has the property that each $j \in \mathcal{B}$ is the greatest index of a b–letter occurring in a unique relator R_i or S_i, and moreover that relator contains precisely one occurrence of b_j.

It follows that the 1–complex L' obtained from $\hat{L}$ by removing the 1–cells b_j, $j \in \mathcal{B}$ is connected, with fundamental group isomorphic to G_0. □

Lemma 7.2 *The natural maps $G_0 \to G_+$ and $G_0 \to G_-$ are injective.*

Proof We show that the map $G_0 \to G_+$ is injective. The proof of injectivity of $G_0 \to G_-$ is entirely analogous. Since G_0 is a free group and G_+ is a one-relator group $G_+ = (G_0 * \langle X \rangle)/\{\tilde{S}_N\}$, we need only show that $\tilde{S}_N$, regarded as a word in $(G_0 * \langle X \rangle)$, genuinely involves X. The result then follows from the Freiheitssatz for one-relator groups [10].

Consider the various possibilities for the structure of $\tilde{S}_N$. If the initial graph $I(\Gamma)$ contains a directed cycle, then $N=1$ and $\tilde{S}_1$ is a strictly positive (or strictly negative) word in X, by Theorem 6.1. Thus $\tilde{S}_1$, regarded as a word in the free product $G_0 * \langle X \rangle$, is also strictly positive (or strictly negative) in X, and so genuinely involves X.

Suppose then that $I(\Gamma)$ does not contain a directed cycle. By Theorem 6.2 (i) and Corollary 6.3 we have (up to cyclic permutation) $\tilde{S}_N = a\tilde{U}_N a^{-1}\tilde{V}_N$, with each of $\tilde{U}_N$ and $\tilde{V}_N$ being either positive or negative in X. We also have $\tilde{S}_N$ definitely involving X, since otherwise S_N would lift to $\hat{L}$.

If X occurs in $\tilde{S}_N$ with nonzero exponent-sum, then occurrences of X survive modulo the relators $R_1, \ldots, R_{M-1}, S_1, \ldots, S_{N-1}$, so we may assume that X appears with exponent-sum zero. Thus one of $\tilde{U}_N$, $\tilde{V}_N$ is strictly positive, and the other is strictly negative, with precisely the same number of occurrences of $X^{\pm 1}$. We may rewrite $\tilde{S}_N$ (again, up to cyclic permutation) as

$$\tilde{S}_N = XA_1X \ldots A_tXW_1X^{-1}B_tX^{-1} \ldots B_1X^{-1}W_2$$

for some $t \geq 0$ and words A_i, B_i and W_1, W_2 that do not involve X. If we can show that neither W_1 nor W_2 is equal to the identity element in G_0, then it will follow that the above expression for $\tilde{S}_N$ does not allow for cancellation of X–symbols, when reducing modulo the relators of G_0. The result will follow.

Now a occurs with exponent-sum zero in each of the relators $R_1,\dots,R_{M-1}$ and $S_1,\dots,S_{N-1}$ of the group G_0, by Theorem 6.2. If neither U_N nor V_N contains the letter a, then each of W_1, W_2 contains precisely one occurrence of a, and so has infinite order in G_0. In particular, they are nontrivial in G_0, as required.

This reduces us to the case where one of U_N, V_N involves the letter a. By Corollary 6.4 we know that this can happen for only one of U_N, V_N.

First suppose that a occurs in U_N. Then $q'(N-1)=Q$ (and so also $N>1$). As in the proof of Corollary 6.3, the part of U_N that gives rise to occurrences of X comes from the geodesic δ in Φ_I from $c_{q'(N-2)+1}$ to x_1. The relevant subword of U_N has the form:

$$[(y_{q'(N-2)+2}c^{-1}_{q'(N-2)+1})\dots(y_Qc^{-1}_{Q-1})(\tau(f_Q)c_Q^{-1})]^{\pm1},$$

or, if δ passes through a:

$$[(y_{q'(N-2)+2}c^{-1}_{q'(N-2)+1})\dots(\tau(f_Q)c_Q^{-1})(x_1a^{-1})\dots(b_{p(1)+1}b^{-1}_{p(1)})]^{\pm1}.$$

The occurrences of X in $\tilde{U}_N$ correspond to those y_j, $j\geq q'(N-2)+2$ that are not equal to x_1, and also from $\tau(f_Q)$ if this is not in the same component of $I(\Gamma)$ as x_1. In the case where δ passes through a, we see that, in $\tilde{S}_N=a\tilde{U}_Na^{-1}\tilde{V}_N$ the a–letters that occur in the same W_i have the same exponent, and hence the W_i are both nontrivial in G_0, as required. In the other case, $\tau(f_Q)=a$ and the unique occurrence of c_Q in $\tilde{V}_N$ lies on the same side of all the X–letters as the unique occurrence of a. Hence c_Q occurs (precisely once) in the same W_i that contains two a–letters. To prove that this W_i is nontrivial in G_0, it suffices to show that c_Q does not occur in any of the relators $R_1,\dots,R_{M-1}$ or $S_1,\dots,S_{N-1}$. But c_Q can occur in S_j $(j<N)$ only if $j=N-1$ and $q'(N-2)=Q-1$, while c_Q can occur in R_j $(j<M)$ only if $j=M-1$ and $q(M-1)=Q-1$. In either case $y_2=\dots=y_Q=x_1$ (since R_{M-1} and S_{N-1} lift to $\hat{L}$) and f_Q joins a to x_1, which contradicts Lemma 5.1.

Suppose next that a occurs in V_N. Then $p(N)=P$. The occurrences of X in $\tilde{V}_N$ arise as indicated in the proof of Corollary 6.3. The relevant subword of V_N has the form:

$$[(x_{p(N-1)+2}b^{-1}_{p(N-1)+1})\dots(x_Pb^{-1}_{P-1})(\tau(e_P)b_P^{-1})]^{\pm1},$$

or, if γ passes through a:

$$[(x_{p(N-1)+2}b^{-1}_{p(N-1)+1})\dots(\tau(e_P)b_P^{-1})(x_1a^{-1})\dots(b_{p(1)+1}b^{-1}_{p(1)})]^{\pm1}.$$

The occurrences of X in $\tilde{V}_N$ correspond to those x_j, $j \geq p(N-1)+2$ in this subword that are equal to v, and also to $\tau(e_P)$ if $\tau(e_P) = v$. If $a = \tau(e_P)$ then since

$$\tilde{S}_N \sim a\tilde{U}_N a^{-1}\tilde{V}_N \sim XA_1X \dots A_tXW_1X^{-1}B_tX^{-1} \dots B_1X^{-1}W_2$$

we see that the two a–letters that occur in the same W_i have the same exponent, and hence both W_i are nontrivial in G_0, as required.

If $a = \iota(e_P)$ then γ passes through a. Assume for the moment that $x_1 = u$. Then the unique occurrence of b_P in $\tilde{U}_N$ lies on the same side of all the X–letters as the unique occurrence of a. Hence the W_i that contains two a–letters also contains a single occurrence of b_P. To prove that this W_i is nontrivial in G_0, it suffices to show that b_P does not occur in any of the relators $R_1, \dots, R_{M-1}$ or $S_1, \dots, S_{N-1}$ of G_0. But b_P can occur in S_j $(j < N)$ only if $j = N-1$ and $p(N-1) = P-1$, while if b_P occurs in R_j $(j < M)$, then $j = M-1$ and $p'(M-2) = P-1$. In either case $x_1 = \dots = x_P = u$, contradicting Lemma 5.1.

This last argument does not apply if $x_1 = v$. In this case we still have $x_2 = \dots = x_P = u$, and since $a = \iota(e_P)$ it follows from Lemma 5.1 that $\tau(e_P) = v$. If, say, $W_1 = 1$ in G_0, then $A_t = vb_P^{-1}$ and $A_tW_1B_t = A_tB_t \neq 1$ in G_0, since this word contains a single occurrence of b_P, which by similar arguments to the above cannot occur in any of the relators of G_0. Hence no more than one pair of letters $X^{\pm 1}$ in S_N can cancel modulo the relators of G_0, and so S_N, as a word in $G_0 * \langle X \rangle$, definitely involves X, as required.

This completes the proof of the Lemma. □

Corollary 7.3 *The maps $G_\pm \to G_1$ are injective.*

Proof The commutative square

$$\begin{array}{ccc} G_0 & \longrightarrow & G_+ \\ \downarrow & & \downarrow \\ G_- & \longrightarrow & G_1 \end{array}$$

is a pushout, and the maps $G_0 \to G_\pm$ are injective by the lemma. Hence G_1 is the free product of G_+ and G_-, amalgamated over G_0. □

Let L_+ be the 1–complex obtained from $\hat{L}$ by identifying the 0–cells $(0,u)$ and $(0,v)$ to a single 0–cell 0. Then L_+ is homotopy equivalent to the subcomplex $\hat{L}\cup X$ of $\tilde{L}$, and G_+ is a homomorphic image of the free group $\pi_1(\hat{L})*\langle X\rangle$, which is naturally identifiable with $\pi_1(L_+)$. Let us fix the 0–cell 0 as a basepoint for L_+, and consider the generating set

$$B_+=\{\theta_e=\tau(e)\lambda(e)^{-1}\ ;\ e\in E(\Gamma)\}$$

for $\pi_1(L_+,0)$. Note that B_+ is not a basis, since the unique cycle in $T(\Gamma)$ gives rise to a relation R_0 among the θ_e. However, this is the only relation, in the sense that $\pi_1(L_+,0)$ has a one-relator presentation $\langle B_+ \mid R_0\rangle$.

Similarly, if L_- is obtained from $\hat{L}$ by identifying the 0–cells $(1,u)$ and $(1,v)$ to a single 0–cell 1, then G_- is a homomorphic image of the free group $\pi_1(L_-,1)$, which is generated by

$$B_-=\{\phi_e=\lambda(e)^{-1}\iota(e)\ ;\ e\in E(\Gamma)\}$$

modulo a single relator S_0 arising from the unique cycle in $I(\Gamma)$.

Theorem 7.4 *The correspondence $\theta_e\leftrightarrow\phi_e$ $(e\in E(\Gamma))$ induces a group isomorphism $G_+\leftrightarrow G_-$.*

Proof The relation R_0 among the generators B_+ is precisely the nullhomotopic path R_0 in L, which lifts to L_+ (indeed to $\hat{L}$). Under the isomorphism $\Psi\colon F(B_+)\to F(B_-)$ induced by the map $\theta_e\mapsto\phi_E$, this relation R_0 is mapped to $\partial_-R_0=R_1$, which is a relation in G_-. Hence we have an induced homomorphism $\pi_1L_+\to G_-$. In order to show that this in turn induces a homomorphism $G_+\to G_-$, we must show that each relation of G_+ is mapped to a relation of G_-.

Each word R_i, $1\le i\le M-1$ is mapped under Ψ to $\partial_+R_i=R_{i+1}$, which is a relation in G_-. Similarly, for $1\le j\le N$ we have $\Psi^{-1}(S_{j-1})=\partial_-S_{j-1}=S_j$, so $\Psi(S_j)=S_{j-1}$, which is also a relation in G_-. Hence Ψ induces a group homomorphism $G_+\to G_-$, as claimed. Similarly Ψ^{-1} induces a group homomorphism $G_-\to G_+$, and these homomorphisms are mutually inverse isomorphisms, by standard arguments. □

Corollary 7.5 *$G(\Gamma)$ is isomorphic to an HNN extension of the finitely presented group G_1, with associated subgroups $G_\pm$.*

Proof This is an easy exercise, given the isomorphism described in the previous lemma. □

This completes the proof of our main result, Theorem 1.1.

8 Further remarks

In the proof of Theorem 1.1, we have relied heavily on one-relator theory to show that our HNN base G_1 is indeed defined by the relators R_i and S_i. If we look at LOTs of larger diameter, we no longer have these tools at our disposal.

As long as $I(\Gamma)$ and $T(\Gamma)$ each have only two components (and hence only one cycle), a great deal of the proof goes through. Certainly the forward and backward derivatives give rise to two finite sequences R_i and S_i of relators for G_1, but in order to prove that these relations are sufficient to define G_1 we would need to prove a Freiheitssatz for the one-relator products $(G_0 * \langle X \rangle)/S_N$ and $(G_0 * \langle Y \rangle)/R_M$. In our case, we have used the combinatorics of the diameter 3 situation in a nontrivial way to show that G_0 is free and that S_N properly involves X (resp R_M properly involves Y) modulo the relations of G_0, from which the Freiheitssatz follows.

It seems reasonable to conjecture in more generality that the HNN base B for G, generated by $\{xy^{-1},\ x, y \in V\}$ will be finitely presented. One may construct sets of relations on this generating set analogous to the R_i and S_i above, by repeatedly applying the forward derivative construction to nullhomotopic paths arising from closed paths in $I(\Gamma)$ (analogous to our S_0), and the backward derivative construction to nullhomotopic paths arising from closed paths in $T(\Gamma)$ (analogous to our R_0). Provided we restrict attention to simple closed paths, only finitely many relations arise in this way, and one can conjecture that these form a set of defining relators for B.

Before making this conjecture precise, let us first give a geometric interpretation of these relations. On the 2–complex $K = K(\Gamma)$ we define a *track* $\mathbf{T}$ in the sense of Dunwoody [4] as follows: $\mathbf{T}$ intersects each 1–cell in a single point, and each 2–cell in two arcs as in the diagram below.

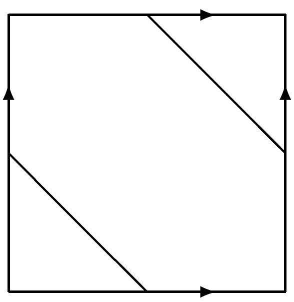


Figure 2

The initial graph $I(\Gamma)$ is naturally embedded as a subgraph of the link of the 0–cell in K. Corresponding to a cycle

$$C = (x_1, \ldots, x_n)$$

in $I(\Gamma)$ is a Dehn diagram D_1 over $\mathcal{P}(\Gamma)$ with a single interior vertex (whose link maps isomorphically to C). We also have a nullhomotopic closed path

$$S_0 = (x_1^{-1}x_2)\ldots(x_n^{-1}x_1)$$

in $K^{(1)}$. The boundary label of D_1 is $S_1 = \partial_+ S_0$. Moreover, if we regard D_1 as a map from the disc D^2 to K, then the track $\mathbf{T}$ on K induces a track on D^2. This track consists of a single circle in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2.

Now suppose that S_1 lifts to $\hat{L}$. Then the Dehn diagram D_1 can be extended to a diagram D_2 with boundary label $S_2 = \partial_+ S_1$, and so on. On any Dehn diagram arising in this way, the track induced by $\mathbf{T}$ consists of a collection of concentric circles in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2.

Dual to the track $\mathbf{T}$ is a flow on K, indicated on the boundary of the 2–cells by the arrows in Figure 2. The flow induced on D^2 by any of the Dehn diagrams obtained as above has only one singular point in the interior of D^2, which is a sink.

We can perform a similar construction for any cycle in $T(\Gamma)$. The boundary label of the resulting Dehn diagram is obtained by repeatedly applying the backward derivative operator to a nullhomotopic closed path in $K^{(1)}$. Again, the induced track on D^2 consists of a collection of concentric circles in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2. The induced flow has only one singular point in the interior of D^2, which is a source.

Let us define a Dehn diagram to be *tame* if the induced track on D^2 consists of a collection of concentric circles in the interior of D^2, together with a collection of arcs, each connecting two adjacent track points on ∂D^2. This is equivalent to the induced flow having only one singular point in the interior of D^2, which is either a sink or a source. It is not difficult to show that every tame Dehn diagram arises by the above construction from a cycle in $I(\Gamma)$ or $T(\Gamma)$, and that its boundary label is an alternating word in the generators $V(\Gamma)$ of $G(\Gamma)$.

Conjecture 8.1 *Let B be the subgroup of $G(\Gamma)$ generated by the alternating words in $V(\Gamma)$. Then B has a finite presentation in which the defining relators are the boundary labels of tame Dehn diagrams.*

References

[1] **J F Adams**, *A new proof of a theorem of W H Cockcroft*, J. LMS 49 (1955) 482–488

[2] **R Bieri**, **R Strebel**, *Almost finitely presented soluble groups*, Comment. Math. Helv. 65 (1990) 243–254

[3] **S D Brodskiĭ**, *Equations over groups and groups with a single defining relator*, Siberian Mathematical Journal 25 (1984) 231–251

[4] **M J Dunwoody**, *The accessibility of finitely presented groups*, Invent. Math. 81 (1985) 449–457

[5] **J Hillman**, *2-Knots and their Groups*, Austral. Math. Soc. Lecture Series 5, Cambridge University Press (1989)

[6] **J Howie**, *On locally indicable groups*, Math. Z. 180 (1982) 445–461

[7] **J Howie**, *The asphericity of ribbon disc complements*, Trans. AMS 289 (1985) 281–302

[8] **S V Ivanov**, *On asphericity of group presentations given by labelled oriented trees*, preprint (1996)

[9] **T Maeda**, *Knotted surfaces in the 4-sphere with no minimal Seifert manifolds*, from: "Combinatorial and Geometric Group Theory" (A J Duncan, N D Gilbert and J Howie, editors), LMS Lecture Note Series 204, Cambridge University Press (1994) 239–246

[10] **W Magnus**, *Über diskontinuierliche Gruppen mit einer definierenden Relation (der Freiheitssatz)*, J. reine angew. Math. 163 (1930) 141–165

[11] **D I Moldavanskiĭ**, *Certain subgroups of groups with one defining relation* (Russian), Sibirsk. Mat. Zh. 8 (1967) 1370–1384

[12] **E S Rapaport**, *Knot-like groups*, from: "Knots, Groups and 3-Manifolds" (L P Neuwirth, editor), Annals of Mathematics Studies, 84 Princeton University Press (1975) 119–133

[13] **D Silver**, *Examples of 3-knots with no minimal Seifert manifolds*, Math. Proc. Cambridge Phil. Soc. 110 (1991) 417–420

[14] **D Silver**, *On the existence of minimal Seifert manifolds*, Math. Proc. Cambridge Phil. Soc. 114 (1993) 103–109

[15] **D Silver**, *HNN bases and high-dimensional knots*, Proc. Amer. Math. Soc. 124 (1996) 1247–1252

[16] **D Silver**, *Free group automorphisms and knotted tori in* S^4, J. Knot Theory Ramifications (1997) 95–103

Department of Mathematics, Heriot-Watt University
Edinburgh EH14 4AS, Scotland

Email: J.Howie@hw.ac.uk

Received: 2 June 1997

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997 **295**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 295–301

On the fixed-point set of automorphisms of non-orientable surfaces without boundary

M Izquierdo
D Singerman

Abstract Macbeath gave a formula for the number of fixed points for each non-identity element of a cyclic group of automorphisms of a compact Riemann surface in terms of the universal covering transformation group of the cyclic group. We observe that this formula generalizes to determine the fixed-point set of each non-identity element of a cyclic group of automorphisms acting on a closed non-orientable surface with one exception; namely, when this element has order 2. In this case the fixed-point set may have simple closed curves (called *ovals*) as well as fixed points. In this note we extend Macbeath's results to include the number of ovals and also determine whether they are twisted or not.

AMS Classification 20F10, 30F10; 30F35, 51M10, 14H99

Keywords Automorphism of a surface, NEC group, universal covering transformation group, oval, fixed-point set

For David Epstein on the occasion of his sixtieth birthday

1 Introduction

Let Y be a compact non-orientable Klein surface of genus $p \geq 3$. By genus here we mean the number of cross-caps of the surface. Let $t\colon Y \to Y$ be an automorphism of order M. If $1 \leq i < M$ and if $i \neq M/2$ then the fixed-point set of t^i consists of isolated fixed points and their number can be calculated, as described below, by a formula which is completely analogous to Macbeath's formula [5] concerning automorphisms of Riemann surfaces. However, if $M = 2N$ then the fixed-point set of the involution t^N consists of a finite number n of disjoint simple closed curves called *ovals* together with a finite number of isolated fixed points [2], [6]. The ovals may be *twisted* or *untwisted* which means that they have Möbius band or annular neigbourhoods respectively.

In this note we calculate the number of ovals and isolated fixed-points of t^N and whether the ovals are twisted or not.

The information is given, as in Macbeath [5] in terms of the universal covering transformation group.

The authors acknowledge Mälardalen University and the Swedish Natural Science Research Council for financial support.

2 The universal covering transformation group

If Y is a compact non-orientable Klein surface of genus $p \geq 3$ then the orientable two-sheeted covering surface of Y has genus ≥ 2, so that the universal covering space of Y is the upper half-plane H (with the hyperbolic metric) and the group of covering transformations is a non-orientable surface subgroup K generated by glide-reflections. If G is a group of automorphisms of Y then the elements of G lift to a *non-euclidean crystallographic (NEC) group* Γ acting on H. There is a smooth epimorphism

$$\theta \colon \Gamma \to G \tag{1}$$

whose kernel is K, where smooth means that θ preserves the orders of elements of finite order in Γ. The transformation group $(\Gamma, \mathcal{H})$ is called the *universal covering transformation group* of (G, Y).

Now let $G = \langle t | t^{2N} = 1 \rangle$ be a cyclic group of order $2N$. As θ is smooth we must have $\theta(c) = t^N$ for every reflection c in Γ. Also we cannot have two distinct reflections in Γ whose product has finite order. So it follows, in the canonical presentation of NEC groups as given in [4] or [3], that Γ has empty period cycles.

Thus Γ has signature of the form

$$s(\Gamma) = (g; \pm; [m_1, ..., m_n]; \{(\quad)^k\}) \tag{2}$$

with k empty period cycles; then Γ has one of the two presentations depending on whether there is a $+$ or a $-$ in the signature;

for the $(+)$ case

$$\begin{aligned} x_1, \ldots, x_n, e_1, \ldots, e_k, c_1, \ldots, c_k, a_1, b_1, \ldots, a_g, b_g \mid \\ x_i^{m_i} = 1, i = 1, ..., n, c_j^2 = c_j e_j^{-1} c_j e_j = 1, j = 1, ..., k, \\ x_1 ... x_n e_1 ... e_k a_1 b_1 a_1^{-1} b_1^{-1} ... a_g b_g a g h^{-1} b_g^{-1} \end{aligned} \tag{3}$$

for the $(-)$ case

$$x_1, \ldots, x_n, e_1, \ldots, e_k, c_1, \ldots, c_k, d_1, ..., d_g \mid$$
$$x_i^{m_i} = 1, i = 1, ..., n, c_j^2 = c_j e_j^{-1} c_j e_j = 1, j = 1, ..., k, x_1 ... x_n e_1 ... e_k d_1^2 ... d_g^2 \quad (4)$$

In these presentations the generators x_i are elliptic elements, the generators c_j are reflections, *the generating reflections* of Γ, and the generators e_j are orientation-preserving transformations called the *connecting generators*. Each empty period cycle corresponds to a conjugacy class of reflections in Γ.

One important fact to note about these presentations is that the connecting generator e_j commutes with the generating reflection c_j, and in fact the centralizer of c_j in Γ is just the group $gp\langle c_j,\ e_j\rangle \cong C_2 \times C_\infty$. (See [8])

3 The fixed-point set of a power of t

Let Y be a non-orientable surface of topological genus $p \geq 3$ and let t be an automorphism of order $2N$. If $1 \leq i < 2N$ and $i \neq N$ then the number of fixed points of the automorphism t^i is given by Macbeath's formula (see [5]). If t^i has order d than t^i has

$$2N \sum_{d|m_j} \frac{1}{m_j} \quad (5)$$

fixed points, where m_j runs over the periods in $s(\Gamma)$.

This is because Macbeath's proof (applying to Fuchsian groups) only uses the facts that each period corresponds to a unique conjugacy class of elliptic elements of Γ, and each elliptic element has a unique fixed point in H. Now, the number of isolated fixed points of t^i is independent of the smooth epimorphism θ above. However the epimorphism θ does play a part in the number of ovals of t^N.

Theorem 3.1 *Let Y be a non-orientable surface of topological genus $p \geq 3$. Let $G \cong C_{2N} = \langle t \mid t^{2N} = 1\rangle$ be a group of automorphisms of Y, and let θ and Γ be as described in equations 1 and 2. If $\theta(e_j) = t^{v_j}$ than the number of ovals of the involution t^N is*

$$\sum_{j=1}^{k} (N, v_j) \quad (6)$$

and the number of isolated fixed points of t^N is

$$2N \sum_{m_j\ even} \frac{1}{m_j}.$$

Proof Let $\Lambda = \theta^{-1}(\langle t^N \rangle)$ so that Λ contains the group $K = Ker\theta$ with index 2. Now, Λ must have signature of the form

$$s(\Gamma) = (g; \pm; [2^{(r)}]; \{(\quad)^s\}) \tag{7}$$

with r periods equal to 2 and s empty period cycles.

The reason that all periods in Λ are equal to 2 is because if m_j in $s(\Gamma)$ is even then $x_j^{m_j/2} \in \Lambda$ and any elliptic element of Λ are conjugate to some $x_j^{m_j/2}$ (see [7]).

By results in [2] (see also [3]), r is the number of isolated fixed points of t^N and is given by Macbeath's formula

$$2N \sum_{m_j \ even} \frac{1}{m_j}$$

It also follows from [2] that the number of ovals of t^N is just the number s of period cycles in Λ, which corresponds to the number of conjugacy classes of reflections in Λ. As a reflection c_j in Λ belongs also to Γ and the group Γ has k conjugacy classes of reflections, we just have to determine into how many Λ–conjugacy classes the Γ–conjugacy class of c_j splits. We shall use the epimorphism θ to calculate this number.

There is a transitive action of Γ on the Λ–conjugacy classes of c_j in Λ by letting $\gamma \in \Gamma$ map the reflection gc_jg^{-1} to $g\gamma c_j\gamma^{-1}g^{-1}$, with $g \in \Lambda$. (Because $\Lambda \lhd \Gamma$). Clearly, if $\lambda \in \Lambda$ then λ has a trivial action on these Λ–conjugacy classes. So we have an action of $\Gamma/\Lambda \cong C_{2N}/C_2 \cong C_N$ on these classes. As the centralizer of c_j in Γ is just $\langle c_j, e_j \rangle$, the stabilizer of the Λ–conjugacy classes of c_j in Λ are the cosets $\Lambda, \Lambda e_j, \ldots, \Lambda e_j^{\delta_j - 1}$, where $\delta_j = exp_\Lambda e_j$, the least positive power of e_j that belongs to Λ. Now, let $\varepsilon_j = exp_K e_j$. Then either $\varepsilon_j = \delta_j$ or $\varepsilon_j = 2\delta_j$.

The additive group Z_{2N} contains a subgroup isomorphic to Z_N and $a \in Z_N$ has order $\frac{N}{(N,a)}$ in Z_N so that a has the same order in Z_{2N} if and only if $(2N, a) = 2(N, a)$. If $(2N, a) = (N, a)$ then the order of a in Z_{2N} is twice the order of a in Z_N and we then find that

$$\varepsilon_j = \delta_j \qquad \text{if} \qquad (2N, v_j) = 2(N, v_j)$$

and

$$\varepsilon_j = 2\delta_j \qquad \text{if} \qquad (2N, v_j) = (N, v_j),$$

where $\theta(e_j) = t^{v_j}$.

By the above argument on the action of Γ/Λ on the Λ–conjugacy classes of c_j we see that the number of such classes is N/δ_j, which is

if $\varepsilon_j = \delta_j$

$$\frac{N}{\delta_j} = \frac{N}{\varepsilon_j} = \frac{N(2N, v_j)}{2N} = \frac{(2N, v_j)}{2} = (N, v_j),$$

or if $\varepsilon_j = 2\delta_j$

$$\frac{N}{\delta_j} = \frac{2N}{\varepsilon_j} = \frac{2N(2N, v_j)}{2N} = (2N, v_j) = (N, v_j)$$

Thus in both cases the generating reflection c_j of Γ induces (N, v_j) conjugacy classes of reflections in Λ. Thus the number of ovals of t^N in Y is

$$\sum_{j=1}^{k} (N, v_j) \tag{8}$$

□

Theorem 3.2 *The ovals of t^N in Y induced by the jth period cycle in Γ are twisted if $(2N, v_j) = (N, v_j)$ and untwisted if $(2N, v_j) = 2(N, v_j)$.*

Proof As we have found in Theorem 3.1, the jth empty period cycle in Γ induces (N, v_j) empty period cycles in Λ. The generating reflections of these period cycles are just conjugates of c_j in Γ and, as the corresponding connecting generator e_j is just the orientation-preserving element generating the centralizer of c_j in Γ, we see that the connecting generator of each of the period cycles in Λ induced by the jth period cycle in Γ is just conjugate to $e_j^{\delta_j}$, $\delta_j = exp_\Lambda e_j$ as in the proof of Theorem 3.1. Now, let $\theta'\colon \Lambda \to C_2 = gp\langle\xi\rangle$, where $\xi = t^N$, be the restriction of the epimorphism $\theta\colon \Gamma \to C_{2N}$. Then

if $\varepsilon_j = \delta_j$

$$\theta'(e_j^{\delta_j}) = \theta'(e_j^{\varepsilon_j}) = \theta(e_j^{\varepsilon_j}) = 1$$

if $\varepsilon_j = 2\delta_j$

$$\theta'(e_j^{\delta_j}) = \theta'(e_j^{\frac{\varepsilon_j}{2}}) = \theta(e_j^{\frac{\varepsilon_j}{2}}) = \xi,$$

ξ the generator of C_2. Generally, if c is the generating reflection of an empty period cycle of Λ and e is the corresponding connecting generator then figures 1 and 2 show that $\theta'(e) = 1$ corresponds to an untwisted oval while $\theta'(e) = \xi$ corresponds to a twisted oval.

However, as in the proof of Theorem 3.1 $\varepsilon_j = \delta_j$ if and only if $(2N, v_j) = 2(N, v_j)$ and hence we have untwisted ovals while $\varepsilon_j = 2\delta_j$ if and only if $(2N, v_j) = (N, v_j)$ and we have twisted ovals. □

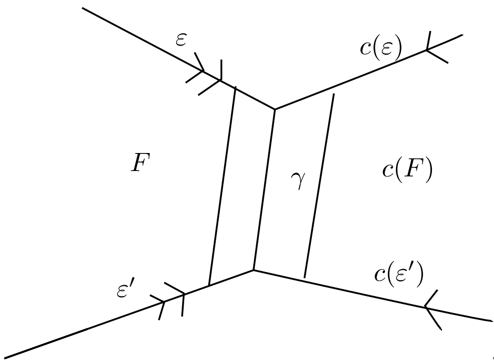


Figure 1: $\theta'(e) = 1$ so $e \in K$

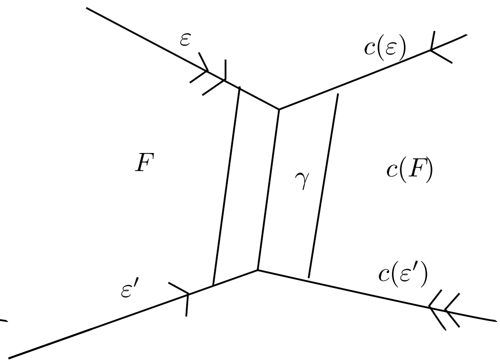


Figure 2: $\theta'(e) = \xi$ so $ce \in K$

4 Bounds and examples

In [6] (also see [2]) Scherrer showed that that if an involution of a non-orientable surface of genus p has $\mid F \mid$ fixed points and $\mid V \mid$ ovals then

$$\mid F \mid +2 \mid V \mid \leq p+2.$$

In our examples we will show that for any integer N we can find a non-orientable surface of genus p admitting a C_{2N} action with generator t such that t^N attains the Scherrer bound.

Example 1 Bujalance [1] found the maximum order for an automorphism t of a non-orientable surface Y of genus $p \geq 3$; it is $2p$ for odd p and $2(p-1)$ for even p. The universal covering transformation group Γ has signature $s(\Gamma) = (0;[2,p];\{(\ \ )\})$ for odd p, and signature $s(\Gamma) = (0;[2,2(p-1)];\{(\ \ )\})$ for even p. There is, essentially, only one way of defining the epimorphism θ in each case:

if p is odd, we define $\theta\colon \Gamma \to C_{2p}$ by $\theta(x_1) = t^p$, $\theta(x_2) = t^2$, $\theta(c) = t^p$, and $\theta(e) = t^{p-2}$,

if p is even, we define $\theta\colon \Gamma \to C_{2(p-1)}$ by $\theta(x_1) = t^{p-1}$, $\theta(x_2) = t^1$, $\theta(c) = t^{p-1}$, and $\theta(e) = t^{p-2}$.

Using Macbeath's formula (5) we see that the involution t^p has p fixed points for surfaces of both odd and even genera. Now, if p is odd then the involution t^p also has, by Theorems 3.1 and 3.2, one twisted oval if p is odd as $(p,p-2) = (2p,p-2) = 1$. If p is even then the involution t^{p-1} has, by Theorems 3.1 and 3.2, one untwisted oval as $(p-1,p-2) = 1$ and $(2(p-1),p-2) = 2(p,p-2) = 2$. We note that the involution t^p obeys the Scherrer bound. Note that the orders of the cyclic groups in Bujulance's examples are $\equiv 2 \bmod 4$. Our second example shows that the Scherrer bound can be obtained for the involution in a C_4 action.

Example 2 Let Y be a non-orientable surface of genus $p \geq 3$, and let t be an automorphism of Y of order 4. Let Γ have signature

$$(0; +; [2^{(r)}, 4, 4]; (\quad)^k)$$

and define a smooth epimorphism $\theta\colon \Gamma \to C_4$ by mapping the generators of order two to t^2, the two generators of order 4 to t and t^{-1} and the connecting generators to the identity. We then find that for the involution t^2, $\mid F \mid = 2r+2$, and $\mid V \mid = 2k$,and $p = 4k + 2r$, so that we find infinitely many surfaces where the Scherrer bound is attained for the involution in C_4. This is easily extended to groups of order $4m$ by replacing the two periods 4 in the signature of Γ by $4m$.

References

[1] **E Bujalance**, *Cyclic groups of automorphisms of compact non-orientable Klein surfaces without boundary*, Pac. J. of Math. 109 (1983) 279–289

[2] **E Bujalance**, **A F Costa**, **S Natanzon**, **D Singerman**, *Involutions of compact Klein surfaces*, Math. Z. 211 (1992) 461–478

[3] **E Bujalance**, **J J Etayo**, **J M Gamboa**, **G Gromadzki**, *A combinatorial approach to groups of automorphisms of bordered Klein surfaces*, Lect. Notes in Math. vol. 1439, Springer–Verlag (1990)

[4] **A M Macbeath**, *The classification of non-euclidean plane crystallographic groups*, Canad. J. Math. 19 (1967) 1192–1205

[5] **A M Macbeath**, *Action of automorphisms of a compact Riemann surface on the first homology group*, Bull. London. Math. Soc. 5 (1973) 103–108

[6] **W Scherrer**, *Zur Theorie der endlichen Gruppen topologischer Abbildungen von geschlossenen Flächen in sich*, Comment. Math. Helv. 1 (1929) 60–119

[7] **D Singerman**, *Subgroups of Fuchsian groups and finite permutations groups*, Bull. London. Math. Soc. 2 (1970) 319–323

[8] **D Singerman**, *On the structure of non-euclidean crystallographic groups*, Proc. Camb. Phil. Soc. 76 (1974) 233–240

Department of Maths, Mälardalen University, 721 23 Västerås, Sweden
and
Department of Maths, University of Southampton, Southampton, SO17 1BJ, UK

Email: `mio@mdh.se, ds@maths.soton.ac.uk`

Received: 15 November 1997

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997 303

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 303–316

The Riley slice revisited

Yohei Komori
Caroline Series

Abstract In [4], Keen and Series analysed the theory of pleating coordinates in the context of the Riley slice of Schottky space $\mathcal{R}$, the deformation space of a genus two handlebody generated by two parabolics. This theory aims to give a complete description of the deformation space of a holomorphic family of Kleinian groups in terms of the bending lamination of the convex hull boundary of the associated three manifold. In this note, we review the present status of the theory and discuss more carefully than in [4] the enumeration of the possible bending laminations for $\mathcal{R}$, complicated in this case by the fact that the associated three manifold has compressible boundary. We correct two complementary errors in [4], which arose from subtleties of the enumeration, in particular showing that, contrary to the assertion made in [4], the *pleating rays*, namely the loci in $\mathcal{R}$ in which the projective measure class of the bending lamination is fixed, have two connected components.

AMS Classification 30F40; 32G05

Keywords Kleinian group, Schottky Group, Riley slice, pleating coordinates

In [4], L Keen and C Series used their theory of pleating invariants to study the so called *Riley slice of Schottky space*. The Riley slice is shown to be foliated by *pleating rays* on which the geometry of the limit set has fixed combinatorial properties. There are two rather subtle errors in [4], concerning the labelling and connectivity of these rays. While the errors do not substantially affect the main results, the correct picture illustrates the interesting new phenomenon that pleating varieties may not be connected, as well as some delicate points related to the marking of the group. The ideas involved will apply to other examples, and we consider them to be of sufficient interest to be worth discussing at some length. Besides explaining and correcting the errors, we take the opportunity to review the background and discuss some of the techniques used in [4] in more detail.

A Kleinian group is a discrete subgroup G of $PSL(2,\mathbf{C})$. It acts on the Riemann sphere $\hat{\mathbf{C}}$ by Möbius transformations and on hyperbolic 3–space $\mathbf{H}^3$ by isometries. The regular set $\Omega(G)$ is the subset of $\hat{\mathbf{C}}$ on which the elements of G form a normal family and the limit set $\Lambda(G)$ is its complement. The quotient $\Omega(G)/G$ is a (possibly disconnected) Riemann surface and $\mathbf{H}^3/G$ is a hyperbolic 3–manifold whose ends are exactly the components of $\Omega(G)/G$. Let $\mathcal{C}$ be the hyperbolic convex hull of $\Lambda(G)$ in $\mathbf{H}^3$; $\mathcal{C}/G$ is the *convex core* of the hyperbolic manifold $\mathbf{H}^3/G$. The boundary $\partial\mathcal{C}/G$ of $\mathcal{C}$ the convex core of $\mathbf{H}^3/G$ is a (possibly disconnected) pleated surface homeomorphic to $\Omega(G)/G$. We denote the geodesic lamination along which this surface is pleated by $pl(G)$.

Let G_μ be a family of Kleinian groups depending holomorphically on a parameter μ which varies over a complex manifold D, and such that the groups G_μ are all quasiconformally conjugate. The theory of pleating invariants analyses D in terms of the *pleating varieties* $\mathcal{P}_\lambda = \{\mu \in D : pl(G_\mu) = \lambda\}$, where λ is a fixed geodesic lamination on $\partial\mathcal{C}/G$.

Let $f\colon U \mapsto \mathbf{C}$ be a holomorphic function defined on a subset $U \subset D$. The *real locus* of f in U is the set $f^{-1}(\mathbf{R}) \cap U$. A geodesic lamination is called *rational* if all its leaves are closed.

In all cases studied so far, [3, 4, 5], D has one or two complex dimensions and it has been shown that:

(1) All geometrically possible pleating varieties are non-empty.

(2) The pleating variety $\mathcal{P}_\lambda$ is a union of connected components of the real loci of a (finite) collection of non-constant holomophic functions $f_{i,\lambda}$ in the part of D on which G_μ is non-Fuchsian.

(3) The pleating variety $\mathcal{P}_\lambda$ is a submanifold of appropriate dimension.

(4) The pleating varieties $\mathcal{P}_\lambda$ for which λ is rational are dense in D.

The pleating varieties foliate D, possibly omitting an exceptional set on which G_μ is Fuchsian.

We say that $g_1, g_2 \in G$ are *I–equivalent* if g_1 is conjugate in G to either g_2 or g_2^{-1}, and write $g_1 \sim g_2$. We denote the equivalence class of g by $C(g)$ and note that the trace function $\operatorname{Tr} g$ is constant on $C(g)$. An oriented closed geodesic in $\mathbf{H}^3/G$ corresponds to a conjugacy class in G, however a closed leaf of a geodesic lamination is unoriented and hence defines only an an I–equivalence class in G.

Suppose that the lamination λ is rational. The functions $f_{i,\lambda}$ of (2) above may be taken to be the set of trace functions $\operatorname{Tr} g_{i,\lambda}$ as $g_i = g_{i,\lambda}, i = 1, \ldots, k$ ranges over a full set of representatives of the I–equivalence classes corresponding to leaves of λ. All of these trace functions are, in principle, computable holomophic functions on the parameter space D. (In any specific example we have to discuss how to make a consistent choice of sign for the trace corresponding to a lifting from $PSL(2,\mathbf{C})$ to $SL(2,\mathbf{C})$; in the case of this paper the problem does not arise since the deformation space is defined as a set of subgroups of $SL(2,\mathbf{C})$.) It follows that, in order to find the foliation by pleating varieties, and hence to compute D, it suffices to enumerate the possible rational pleating laminations $\mathcal{P}_\lambda$ for $\partial\mathcal{C}/G$, and then to identify $\mathcal{P}_\lambda$ among the components of the real loci of the associated trace functions $\operatorname{Tr} g_{i,\lambda}$.

In the case of the Riley slice, this programme was carried out in [4].

Consider the set of discrete subgroups of $SL(2,\mathbf{C})$ which are freely generated by two non-commuting parabolics. Up to conjugation in $SL(2,\mathbf{C})$, any such subgroup can be put in the form $G = G_\rho = \langle X, Y_\rho\rangle$ where X, Y_ρ are the matrices

$$X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad Y_\rho = \begin{bmatrix} 1 & 0 \\ \rho & 1 \end{bmatrix}.$$

The Riley slice $\mathcal{R}$ is defined by:

$$\mathcal{R} = \{\rho \in \mathbf{C} : \Omega(G_\rho)/G_\rho \text{ is a four times punctured sphere }\}.$$

It is known that the deformation space $D = \mathcal{R}$ is topologically an annulus in $\mathbf{C}$ (see [2]). Thus the real locus of a holomorphic function on $\mathcal{R}$ has one real dimension. In this case, (2) is theorem 3.7 in [4] , (1) and (3) follow from theorem 4.1 and (4) is theorem 5.2.

The first error in [4] concerns the enumeration of rational laminations on $\partial\mathcal{C}/G$. Let S denote a four times punctured sphere. In the Riley slice setup, the convex hull boundary $\partial\mathcal{C}/G$ is compressible, in other words, the induced map $\pi_1(S) \to G$ is not injective. This means that to enumerate correctly the possible pleating laminations, one has to determine when two distinct laminations on S define the same family of geodesics in $\mathbf{H}^3/G$. In particular, one has to determine when the images of distinct simple closed curves on S are equal in G. This was not handled quite correctly in [4]. Curves counted there as distinct are actually equivalent in pairs. (For a more general discussion of the pleating locus with compressible boundary, we refer to Otal's thesis [9].)

The second error concerns connectivity of the pleating varieties; contrary to the assertion in [4], except for degenerate cases, each pleating variety has *two* connected components. In fact, the inductive proof of theorem 4.1 of [4] is not quite correct.

At the time of writing [4], these errors went unnoticed as they are in some sense complementary. That errors were present, was first deduced as follows. Let γ be a simple closed curve on S and let $\rho^*(\gamma)$ be the endpoint on $\partial\mathcal{R}$ of the pleating ray $\mathcal{P}_\gamma$, which we assume for the moment has one connected component as in theorem 4.1 of [4]. In the group $G_{\rho^*(\gamma)}$, the element γ is represented by an accidental parabolic with trace equal -2 (see [4] proposition 4.2). Since the trace is a polynomial in ρ with integer coefficients, the element γ in the group $G_{\overline{\rho^*(\gamma)}}$ is also an accidental parabolic so that one would expect $\overline{\rho^*(\gamma)}$ also to be an endpoint of $\mathcal{P}_\gamma$. This appears to contradicts the assertion of theorem 4.1 that $\mathcal{P}_\gamma$ has a unique branch. Again, according to [4], the possible rational pleating laminations are enumerated by the rationals modulo 2 and the points $\rho^*(\gamma), \overline{\rho^*(\gamma)}$ should be the endpoints of a pair of *distinct* pleating laminations whose labels are p/q and $2-p/q$. However, this would imply that the distinct elements $\gamma(p/q)$ and $\gamma(2-p/q)$ are both pinched at $\rho^*(\gamma)$ (and also $\overline{\rho^*(\gamma)}$), which is impossible.

These contradictions are resolved simultaneously by showing that in fact (a) the laminations with labels p/q and $2-p/q$ are the same, and (b) the pleating locus has two connected components which are complex conjugate in the ρ–plane.

The details of how this works are explained below.

1 Enumeration

As explained above, to enumerate correctly the possible rational pleating laminations, one has to determine when two distinct homotopy classes of simple closed curves on S define the same geodesic in $\mathbf{H}^3/G_\rho$. This involves an implicit choice of marking on S (i.e. a choice of generators for $\pi_1(S)$), together with a choice of homomorphism $h\colon \pi_1(S) \to G$. To explain the error, we have to review the enumeration with some care.

The group $G_\rho = \langle G_\rho; X, Y_\rho\rangle$ should be thought of as *marked* by the ordered pair of generators (X, Y_ρ). Thus, although $Y_\rho^{-1} = Y_{-\rho}$ so that $G_\rho = G_{-\rho}$ as subgroups of $PSL(2,\mathbf{C})$, the marked groups $\langle G; X, Y_\rho\rangle$ and $\langle G; X, Y_\rho^{-1}\rangle$ are distinct. (In fact, it follows easily from lemma 1 of [2], that the only possible pair of parabolic generators of G_ρ are the (unordered) pair $X^{\pm 1}, Y_\rho^{\pm 1}$.) Thus we should always identify G_ρ and $G_{\rho'}$ by the isomorphism $X \mapsto X, Y_\rho \mapsto Y_{\rho'}$. We denote by $\langle G; X, Y\rangle$ an abstract two generator marked free group and always use the isomorphisms $X \mapsto X, Y \mapsto Y_\rho$ to identify G with G_ρ. With these identifications understood, abstract words in the symbols $X^{\pm}, Y^{\pm}$ represent elements in both the groups G and G_ρ.

The content of proposition 2.1 of [4] is that a line of rational slope $p/q \in \mathbf{Q} \cup \infty$ in $\mathbf{C}$ projects to a homotopy class of simple closed non-boundary parallel curves on S, and that every such homotopy class on S is obtained in this way. Further, the homotopy classes corresponding to distinct rationals are distinct.

Given a hyperbolic structure on S, there is a unique closed geodesic in each free homotopy class of simple closed non-boundary parallel curves. In [4], with the hyperbolic structure of $\partial\mathcal{C}/G$ understood, we denoted the geodesic corresponding to a line of slope p/q by $\gamma(p/q)$. There is some confusion at this point, which contributes to the error under discussion. Since the line described in proposition 2.1 is in fact unoriented, the correct statement is that a line of rational slope defines an I–equivalence class in $\pi_1(S)$. From now on, therefore, $\gamma(p/q)$ should be understood to denote a specific I–equivalence class in $\pi_1(S)$.

The essence of the proof of proposition 2.1 appears on page 78 of [4] as part of the proof of proposition 2.2, which describes an explicit word $V_{p/q}$ in the generating set $\{X^{\pm 1}, Y^{\pm 1}\}$ which represents $h(\gamma(p/q))$ in G. We need to go through the construction of $V_{p/q}$ with some care. The idea is a simple case of the method of π_1–train tracks introduced in [1], see also [7].

Following [4], let $\mathcal{L}$ denote the integer lattice in the complex plane $\mathbf{C}$; let β: $z \mapsto z + 2i$ and let ξ, η be the rotations by π about the points i and $i + 1$ respectively. Let $\Gamma_0 = \{\xi^{\pm}, \eta^{\pm}, \beta^{\pm}\}$. The surface S can be realised as the quotient of $\mathbf{C} - \mathcal{L}$ by the group Γ generated by the elements of Γ_0.

We shall compare the three diagrams in figure 2 in [4]. Figure 2a is a fundamental domain R for the action of $G = G_\rho$ on $\Omega = \Omega_\rho$. Figure 2b is a fundamental domain R' for the action of $\tilde{G} = \tilde{G}_\rho$, the Fuchsian uniformisation of $\pi_1(S)$, acting in the hyperbolic disc Δ, thought of as the universal cover of S. Figure 2c is a fundamental domain R'' for the action of Γ on $\mathbf{C} - \mathcal{L}$. The sides of each of these domains are supposed to be labelled by generators α of G, $\tilde{G}$ and Γ respectively in such a way that the label α on a side indicates that it is paired to the side labelled α^{-1} under the action of α. (We note that, although R'' is a rectangle in $\mathbf{C}$, for the purposes of this discussion it should be thought of as having *six* sides.) Denote by $\tilde{G}_0$ the generating set $\{X', X'^{-1}, Y', Y'^{-1}, B', B'^{-1}\}$ of $\tilde{G}$.

Unfortunately, there is a labelling error in figure 2 (but not in the text) which may obscure the explanation on pages 77–78 of [4]. The configuration in figure 2(a) refers to the case $\rho < -4$. The two circles shown are the isometric circles of

$Y_\rho^{\pm}$; since $\rho < -4$ the circle on the right has centre $-1/\rho$ and is the isometric circle of Y_ρ. This circle is identified with the circle on the left by Y_ρ and thus, with the convention explained above, the labels Y and Y^{-1} should be interchanged. This error carries through to figures 2 (b) and (c) in which we should interchange the labels Y' and Y'^{-1}, and η and η^{-1}, respectively.

We proceed with this change of labelling throughout.

Let γ be *any* simple closed non-boundary parallel loop on S. Its lift to any of the three covering spaces Ω, Δ or $\mathbf{C}-\mathcal{L}$ of S is simple and therefore appears on each region R, R', R'' as a collection of pairwise disjoint arcs with endpoints on the labelled sides. When the sides of one of the regions R, R', R'' are identified by the side pairings, there is a unique way to link the endpoints of the arcs to form a simple closed loop on S. This loop is well defined up to orientation and homotopy, and thus we obtain an I–equivalence class in $\pi_1(S)$.

Making a suitable homotopy, we may assume that none of these arcs join a side to itself. It is also clear that the total number of arcs meeting a side labelled α must equal the number meeting its paired side labelled α^{-1}. Let $n(\alpha,\beta)$ denote the number of arcs joining the sides with labels α,β. When the sides of R' are identified, any arc joining sides X' to X'^{-1} links up to form a loop round a puncture. Since γ is connected and non-boundary parallel, we conclude that $n(X', X'^{-1}) = 0$, and likewise that $n(Y', Y'^{-1}) = 0$. A similar argument (which makes crucial use of the fact that γ is simple) shows that at least one of $n(X', B')$ and $n(X'^{-1}, B'^{-1})$, and at least one of $n(Y', B')$ and $n(Y'^{-1}, B'^{-1})$, must vanish, see [1, 7].

Exactly the same constraints apply to the weights $n(\alpha,\beta)$, $\alpha,\beta \in \Gamma_0$, in figure 2c. Inserting these constraints, we obtain precisely either one of the three patterns shown in figure 3 of [4], or its reflection in the line $\Re z = 1/2$. In these diagrams, there is at most one line $l(\alpha,\beta)$ joining a pair of sides α,β and the integer label k on $l(\alpha,\beta)$ indicates that $n(\alpha,\beta) = k$. Conversely, given such a weighted diagram, we can recover a simple closed curve by replacing the line $l(\alpha,\beta)$ by $n(\alpha,\beta)$ parallel arcs joining the sides α,β. When the sides of R'' are identified, there is a unique way to link these arcs to form a union of simple closed loops on S. There is one connected loop if and only if the integers n, m appearing in figure 3 are relatively prime. Taking $(n,m) = 1$, we see that each of the patterns in figure 3 is exactly that obtained from a line of rational slope in $\mathbf{C}$, and that, provided we include reflections as above, every line of rational slope appears. This proves that every homotopy class of simple closed loops on S is the projection of a line of rational slope in the plane as claimed.

We want to show that lines of different slope correspond to non-homotopic loops on S. To do this, observe that each side σ of R' is a line joining two punctures on S, and that the number of arcs meeting σ is exactly the minimum geometric intersection number of loops homotopic to γ with σ. It is clear that these intersection numbers determine $n(\alpha, \beta),\ \alpha, \beta \in \tilde{G}$, which gives the result. It is also clear from the weighted diagrams in figure 3, that *all* lines of the same slope define the same class.

As noted above, this construction determines a curve only up to homotopy and orientation. The class corresponding to a line of slope p/q in figure 2c is exactly the I–equivalence class $\gamma(p/q)$ in $\pi_1(S)$ described above.

We now want to find a word $V_{p/q}$ representing $h(\gamma(p/q))$ in G_ρ. As indicated in the proof of 2.2 in [4], this is done by the method of cutting sequences, see for example [1, 10]. We explain the method in somewhat more detail here.

Consider first the tesselation $\mathcal{T}$ of the hyperbolic disc Δ by images of the region R' under the action of the group $\tilde{G}$. With the correction noted above, the sides of R' should be labelled, in anticlockwise order starting from 0 by $B', Y'^{-1}, Y', B'^{-1}, X'^{-1}, X'$. These labels are transported to the tesselation $\mathcal{T}$ by the action of $\tilde{G}$. Two copies R'_1, R'_2 of R' meet along each edge, and each edge carries two labels $\alpha, \alpha^{-1} \in \tilde{G}_0$, one label interior to R'_1 and the other interior to R'_2.

Let λ be an oriented geodesic segment in Δ and let $\alpha_1, \ldots, \alpha_k$ be the ordered sequence of labels of edges of $\mathcal{T}$ cut by λ, where if λ cuts successively adjacent regions $R'_i, i = 1, \ldots, k+1$ then α_i is the label of the common side of R'_i and R'_{i+1} which is *inside* R'_{i+1}. The sequence thus obtained is called the $\tilde{G}$*–cutting sequence* of λ. With the above labelling conventions, if $h \in \Gamma$ and $z \in \Delta$, then one can verify that the $\tilde{G}$–cutting sequence of the oriented geodesic from z to $h(z)$ is a word in the generators $\tilde{G}_0$ representing h, see [1] for details.

We define $G-$ and $\Gamma-$cutting sequences similarly. It is clear that the G*–cutting sequence* of the projection of the segment λ to Ω is obtained from the $\tilde{G}$–sequence by omitting the labels $B^{\pm}$ and replacing X' by X and Y' by Y. This specifies implicitly that the map $h\colon \pi_1(S) \to G = \pi_1(\mathbf{H}^3/G_\rho)$ is $h(X') = X, h(Y') = Y_\rho, h(B') = \mathrm{id}$. Likewise the $\Gamma-$cutting sequence of the projection of λ to $\mathbf{C} - \mathcal{L}$ is obtained from the $\tilde{G}$–sequence by replacing the labels $B'^{\pm}$ with $\beta'^{\pm}$, X' by ξ and Y' by η. Clearly, since the combinatorics of all three diagrams in figure 2 are the same, we can read off the G–sequence from the Γ–sequence by omitting the labels $\beta^{\pm}$ and replacing ξ by X and η by Y. This is a key point in our idea.

In practice, the cutting sequence is read off from weighted diagram by a simple combinatorial procedure. For definiteness, suppose we have a weighted diagram on the region R'' as in figure 3 of [4]. First, redraw the diagram replacing the line $l(\alpha, \beta)$ with weight $n(\alpha, \beta)$ by $n(\alpha, \beta)$ parallel and pairwise disjoint arcs joining the sides α, β. As explained above, these arcs link in a unique order to form a simple closed loop λ on S. Pick an orientation and initial point on λ. To follow the convention described above, every time λ crosses an edge s of R'', write down the label on s and *outside* R''. Thus, if an oriented arc of λ has initial point on an edge labelled α inside R'' and final point an edge labelled β inside R'', then its contribution to the cutting sequence is α, β^{-1}. The cutting sequence thus obtained is a word in the generators of the group Γ. Changing the initial point of λ cyclically permutes the cutting sequence, so that the corresponding words are conjugate elements in Γ, while reversing the orientation of λ produces the inverse word. Thus the loop λ defines an I–equivalence class in Γ.

Now let $p/q \in \mathbf{Q} \cup \infty$ and let $L_{p/q}$ denote some line of slope p/q in $\mathbf{C}$. Its Γ–cutting sequence is periodic, and the word $V_{p/q} \in G$ of proposition 2.2 representing $h(\gamma(p/q)) \in G$ is obtained by the procedure described above. Notice that $V_{p/q}$ is automatically cyclically reduced. Clearly, $h(\gamma(p/q))$ is equally represented by the word $V_{p/q}^{-1}$ corresponding to the cutting sequence of the line $L_{p/q}$ with its orientation reversed.

The remark on page 77 of [4] gives some examples. We note that the words given in the text are correct, but should be read off relative to the corrected labelling of figure 2 in which Y and Y^{-1} are interchanged.

That the words $V_{p/q}$ are defined only up to cyclic conjugation and inversion is another source of confusion in [4]. Only the I–equivalence class is well defined. As noted above, this equivalence class should also not change when $L_{p/q}$ is replaced by a parallel line of the same slope. In fact, it is clear that there are only a finite set of possible cutting sequences obtained from parallel translates of a line segment of finite length and that these sequences differ only by cyclic permutation. We denote the I–equivalence class in G thus obtained by $C_{p/q}$.

1.1 The enumeration error

In accordance with the comments in the introduction, our task is to identify when two equivalence classes $C_{p/q}$ and $C_{r/s}$ coincide. This problem is discussed in remark 2.5 on page 79 of [4], where it is stated correctly that $V_{p/q} \sim V_{r/s}$ if $r/s = p/q + 2n, n \in \mathbf{Z}$. However, the claim in that remark that if $0 \leq p/q <$

$r/s < 2$ then $\gamma(p/q)$ and $\gamma(r/s)$ are distinct is wrong; in fact, as explained in the proof of theorem 1.2 below, only $q > 0$ and $|p|$ are invariants of the class $C_{p/q}$. Thus, contrary to the claims implicit in [4], we have:

Lemma 1.1 *For $p/q \in \mathbf{Q}$, the classes $C_{p/q}$ and $C_{-p/q}$ coincide.*

Proof Let $L_{p/q}$ be a line of rational slope $p/q \in \mathbf{Q}$ with initial point on the edge of R'' joining vertices $0, i$. Its reflection $L_{-p/q}$ in the imaginary axis has slope $-p/q$; let $V_{\pm p/q}$ be the words obtained from the G-cutting sequences of $L_{\pm p/q}$ as above. It is easy to see that the Γ-sequences of $L_{\pm p/q}$ differ by interchanging ξ with ξ^{-1}, η with η^{-1}, and β with β^{-1}. (The interchange of β with β^{-1} happens because in the tesselation of $\mathbf{C} - \mathcal{L}$ by images of R'' under Γ, the labels β and β^{-1} alternate along horizontal lines.) Therefore $C_{p/q}(X^{-1}, Y^{-1}) = C_{-p/q}(X, Y)$.

Now compare two lines of the same slope p/q which differ by vertical translation by i. Their cutting sequences differ by interchanging ξ with ξ^{-1}, η with η^{-1}, and β with β^{-1}; in addition the position of the β terms in the sequence shifts relative to that of the $\xi's$ and $\eta's$. (For example the sequence for $1/1$ with initial point between 0 and i is $\xi\eta^{-1}\beta^{-1}$, while with initial point between i and $2i$ we get $\xi^{-1}\beta\eta$.) Since the position of $\beta^{\pm}$ relative to $\xi^{\pm}, \eta^{\pm}$ does not affect the X, Y sequence, we get $C_{p/q}(X^{-1}, Y^{-1}) = C_{p/q}(X, Y)$.

Combining these observations gives the proof. □

We also need to know there are no other identifications. We have:

Theorem 1.2 *The classes $C_{p/q}, C_{r/s}, p/q, r/s \in \mathbf{Q} \cup \infty$ are equivalent if and only if $r/s = p/q + 2n$ or $-r/s = p/q + 2n$, $n \in \mathbf{Z}$.*

Proof As discussed in remark 2.5 of [4] on page 79, a (left) Dehn twist about the curve $\gamma(\infty)$ represented by $\beta \in \Gamma$ induces an automorphism of Γ which maps $\gamma(p/q)$ to $\gamma(2+p/q)$. Since this automorphism induces the identity on G, we have $C_{p/q} = C_{2+p/q}$. This can also be seen by representing $C_{p/q}$ and $C_{2+p/q}$ by the cutting sequences of lines of slope p/q and $2 + p/q$ in $\mathbf{C}$. Reading off the two cutting sequences starting from the same inital point, it is easy to see that, while the Γ–sequences differ, the induced G–sequences are the same.

To complete the proof, it only remains to show that if $r/s \neq \pm(p/q + 2n)$ then $C_{p/q}, C_{r/s}$ are distinct.

As stated in remark 2.5 of [4], the cutting sequence of $L_{p/q}$ has length $2q$. Moreover, the words $V_{p/q}$ and $V_{r/s}$ are cyclically reduced and so, since G is a free group, are conjugate only if they have the same length. Thus, since we are assuming that $q, s \geq 0$, a necessary condition for $C_{p/q} = C_{r/s}$ is that $q = s$.

It is also stated in remark 2.5 that p can be deduced from number of sign changes in the exponents of X and Y in $V_{p/q}$. This is not quite correct, and herein lies the root of error number 1. Since the number of sign changes is necessarily non-negative, one verifies that in all cases we can only obtain $|p|$ and not p from $C_{p/q}$, in other words, $C_{p/q} = C_{r/s}$ implies $|p| = |r|$ but, contrary to the claim of remark 2.5, the number of sign changes cannot be used to distinguish the classes of $V_{p/q}$ and $V_{-p/q}$. This is correct, since by lemma 1.1 the two classes $C_{p/q}$ and $C_{-p/q}$ coincide. □

Remark 1.3 For future reference, we note that a similar argument to the above shows that $V_{1+p/q}$ can be obtained from $V_{p/q}$ by interchanging Y and Y^{-1}, more precisely, that $V_{1+p/q}(X, Y) = V_{p/q}(X, Y^{-1})$.

This completes the discussion of the first error.

2 Connectivity

Let $g \in G_\rho$ correspond to a simple closed geodesic γ on $\partial\mathcal{C}/G$. The trace $\mathrm{Tr}\, g$ is a polynomial in ρ with integer coefficients. It is claimed in [4] theorem 4.1 that the pleating ray $\mathcal{P}_\gamma$ has a unique connected component with a unique endpoint $\rho^* = \rho^*(\gamma)$ on $\partial\mathcal{R}$. At this endpoint, $\mathrm{Tr}\, g = \mathrm{Tr}\, g(\rho^*) = -2$ ([4] proposition 4.2) and g is an accidental parabolic. The group G_ρ is free, $\Omega(G_\rho) \neq \emptyset$, and therefore G_ρ is maximally parabolic as in [6], i.e., G_{ρ^*} contains the maximal number of rank 1 parabolic subgroups among subgroups of $PSL(2, \mathbf{C})$ isomorphic to G_{ρ^*}.

The map $\rho \mapsto \bar{\rho}$ induces the maps $X \mapsto X, Y_\rho \mapsto Y_{\bar{\rho}}$ and hence an isomorphism $J\colon G_\rho \to G_{\bar{\rho}}$; clearly, J is type preserving, i.e., $g \in G_\rho$ and $J(g) \in G_{\bar{\rho}}$ are either both parabolic or both loxodromic It is also clear that $\rho \mapsto \bar{\rho}$ maps $\mathcal{R}$ to itself. We note that this does not contradict the uniqueness of maximally pinched groups asserted in theorem III of [6], because the conjugation $G_{\rho^*} \to G_{\bar{\rho}^*}$ is *anti*holomorphic. However, it does mean that $\bar{\rho}^*$ should be also be an endpoint of the pleating ray $\mathcal{P}_\gamma$, which contradicts theorem 4.1 of [4].

This contradiction is resolved by the corollary to the following lemma.

Lemma 2.1 *Let $j\colon \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ be an conformal or anticonformal bijection and let $j_*(M) = jMj^{-1}, M \in SL(2, \mathbf{C})$. Let $\langle G_\rho; X, Y_\rho \rangle$ be the marked free group with generators X, Y_ρ. Suppose that the pleating locus $pl(G_\rho)$ consists of a simple closed geodesic represented by a word $W(X, Y_\rho)$. Let $j_*(G_\rho)$ denote the marked group with generators $j_*(X), j_*(Y_\rho)$. Then $pl(j_*(G_\rho))$ is a closed geodesic represented by the word $W(j_*(X), j_*(Y_\rho))$.*

Proof Recall from [4] that a subgroup of a Kleinian group G is called *F–peripheral* if it is Fuchsian and if one of the two open round discs bounded by its limit set contain no points of $\Lambda(G)$. If a geodesic γ is the pleating locus of $G_\rho, \rho \in \mathcal{R}$, then, as discussed on page 82 of [4], γ divides $\partial\mathcal{C}/G_\rho$ into two connected components each of which is a sphere with two punctures and one hole. The lifts of these components to $\mathbf{H}^3$ lie in hyperbolic planes which separate $\mathbf{H}^3$ into two open hyperbolic half spaces, one of which contains no points of $\mathcal{C}$. The half spaces meet $\hat{\mathbf{C}}$ in open discs which have empty intersection with the limit set $\Lambda(G_\rho)$, so that the subgroups of G_ρ which leave these discs invariant are F–peripheral and contain the element suitable conjugates of $W(X, Y_\rho)$. In particular, $W(X, Y_\rho)$ lies in in two non-conjugate F–peripheral subgroups of G_ρ.

We showed in [4] proposition 3.6 that conversely, if an element $g \in G_\rho$ which represents a simple closed geodesic γ on $\partial\mathcal{C}/G$ lies in two non-conjugate F–peripheral subgroups, then $pl(G_\rho) = \gamma$.

Now since the map j is conformal or anticonformal, it maps circles to circles. It is also clear that $\Lambda(j_*(G_\rho)) = j(\Lambda(G_\rho))$, and that if a subgroup of G_ρ is Fuchsian, so is its image in $j_*(G_\rho)$. Hence j_* preserves F–peripheral subgroups.

Thus $W(X, Y_\rho)$ lies in two non-conjugate F–peripheral subgroups of G_ρ if and only if $j_*(W(X, Y_\rho)) = W(j_*(X), j_*(Y_\rho))$ lies in two non-conjugate F–peripheral subgroups of $j_*(G_\rho)$. The result follows. □

Corollary 2.2 *Let $p/q \in \mathbf{Q} \cup \infty$ and let $j\colon \rho \mapsto \bar{\rho}$ be complex conjugation. Then $\mathcal{P}_{p/q} = j(\mathcal{P}_{p/q})$.*

Proof Let $\rho \in \mathcal{P}_{p/q}$ so that $pl(G_\rho) = \gamma_{p/q}$. As above, in the marked group $\langle G_\rho; X, Y_\rho \rangle$, the class $\gamma_{p/q}$ is represented by the word $V_{p/q}(X, Y_\rho)$ in X, Y_ρ. We apply lemma 2.1 to j and compute $j_*(X) = X$ and $j_*(Y_\rho) = Y_{\bar{\rho}}$. Thus $j_*(G_\rho)$ is the marked group $\langle G_{\bar{\rho}}; X, Y_{\bar{\rho}} \rangle$, and $pl(G_{\bar{\rho}})$ is represented by the word $V_{p/q}(X, Y_{\bar{\rho}})$ which corresponds in the marked group $G_{\bar{\rho}}$ to $\gamma_{p/q}$. □

Remark 2.3 We can also apply lemma 2.1 to the involution $k\colon \rho \mapsto -\rho$. We find $k_*(X) = X^{-1}$ and $k_*(Y_\rho) = Y_{-\rho} = Y_\rho^{-1}$. Thus $k_*(G_\rho)$ is the marked group $\langle G_{-\rho}; X^{-1}, Y_\rho^{-1}\rangle$. If as above, $pl(G_\rho)$ is represented by the word $V_{p/q}(X, Y_\rho)$ in X, Y_ρ, then $pl(G_{-\rho})$ is represented by the word $V_{p/q}(X^{-1}, Y_\rho^{-1})$. Now as in lemma 1.1 and remark 1.3 above, $V_{p/q}(X^{-1}, Y_\rho^{-1}) \sim V_{p/q}(X, Y_\rho)$ and $V_{p/q}(X, Y_\rho) \sim V_{1+p/q}(X, Y_\rho^{-1})$. Thus the pleating locus of the *marked* group $\langle G_{-\rho}; X, Y_{-\rho}\rangle = \langle G_{-\rho}; X, Y_\rho^{-1}\rangle$ is $\gamma_{1+p/q}$ which, using remark 2.3, is the same as $\gamma_{1-p/q}$.

Although as groups G_ρ and $k(G_\rho)$ are the same, k_* is *not* the the standard isomorphism and $k_*(V_{p/q}(X, Y_\rho)) \neq V_{p/q}(X, Y_{-\rho})$. This explains why the end-points of the rays $\mathcal{P}_{p/q}$ and $k(\mathcal{P}_{p/q}) = -\mathcal{P}_{p/q}$ correspond to *different* maximally pinched groups.

2.1 The connectivity error

We can now prove a correct form of theorem 4.1 of [4]. Recall that the hyperbolic locus of the trace polynomial $\operatorname{Tr} V_{p/q}$ is the set

$$\tilde{\mathcal{H}}_{p/q} = \{\rho \in \mathbf{C} : \Im \operatorname{Tr} V_{p/q} = 0, |\Re \operatorname{Tr} V_{p/q}| > 2\},$$

and that $\mathcal{P}_{p/q}$ is a union of connected components of $\tilde{\mathcal{H}}_{p/q}$.

Theorem 2.4 *For $0 < p/q < 1$, the rational pleating ray $\mathcal{P}_{p/q}$ consists of exactly two connected components of the hyperbolic locus $\tilde{\mathcal{H}}_{p/q}$. These rays are the branches which asympotically have arguments $-e^{\pi i p/q}$ and $-e^{-\pi i p/q}$. They are complex conjugate 1–manifolds, with unique and complex conjugate endpoints on $\partial\mathcal{R}$.*

Proof In [4], we argued by "induction on the Farey tree". Once again, there is an error in the argument which can be corrected using corollary 2.2.

For the rays $\mathcal{P}_{0/1}$ and $\mathcal{P}_{1/1}$ we argue exactly as in [4] proposition 3.8. (The assertion that these special rays have one connected component is correct; we note that since they are contained the real axis, they are invariant under complex conjugation so the contradiction explained above does not occur.)

Now suppose we have the result for $\mathcal{P}_{p/q}$ and $\mathcal{P}_{r/s}$ for which $ps - rq = -1$. Let $\mathbf{H}^+$ and $\mathbf{H}^-$ denote the upper and lower half planes respectively. The argument in [4] shows that an arc in $\mathbf{H}^+$ joining the components $\mathcal{P}^+_{p/q}$ to $\mathcal{P}^+_{r/s}$ of $\mathcal{P}_{p/q}$ to $\mathcal{P}_{r/s}$ in $\mathbf{H}^+$ must intersect $\mathcal{P}_{(p+r)/(q+s)}$. Also as in [4], the only branch of $\tilde{\mathcal{H}}_{(p+r)/(q+s)}$ whose asymptotic direction lies between directions $-e^{\pi ip/q}$ and $-e^{\pi ir/s}$ is the one with asymptotic direction $-e^{\pi i(p+r)/(q+s)}$, and this must therefore be coincident with a component of $\mathcal{P}_{(p+r)/(q+s)}$. Similarly an arc in $\mathbf{H}^-$ joining the components $\mathcal{P}^-_{p/q}$ to $\mathcal{P}^-_{r/s}$ of $\mathcal{P}_{p/q}$ to $\mathcal{P}_{r/s}$ in $\mathbf{H}^-$ must intersect a component of $\mathcal{P}_{(p+r)/(q+s)}$, with asymptotic direction $-e^{-\pi i(p+r)/(q+s)}$. This gives the result.

(The problem with the argument in [4] is that we forgot to consider arcs joining the components of $\mathcal{P}^+_{p/q}$ to $\mathcal{P}^+_{r/s}$ in $\mathbf{H}^+$ and running through $\mathbf{H}^-$.)

Notice that the picture obtained in this way is entirely consistent with remark 2.3 above. □

3 Conclusion

The errors above do not substantially effect any of the conclusions of [4]. Theorems 1.1 and 2.4 have obvious extensions to irrational laminations, which we shall not spell out here. The only other result which is changed in consequence of the errors is theorem 5.4.

In [4], to deal with irrational rays $\lambda \in \mathbf{R}$, we introduced the complex pleating length L_λ, and referred to the methods of [3], section 7.1 to show that these rays were 1–manifolds with the connectivity claimed. In fact, the argument in [3] has a gap: we omitted to show that the pleating variety $\mathcal{P}_\lambda$ is open in the real locus of L_λ. This crucial fact is proved in a more general context in [5]. For a corrected version of the arguments required in a one dimensional parameter space, we refer to [8]. We note also that by the improved techniques of [5], it follows that even on irrational rays $\lambda \in \mathbf{R}$, the range of the complex pleating length L_λ (see [4] page 88) is $\mathbf{R}^+$.

Let j denote complex conjugation and define an equivalence relation on $\mathbf{R}$ by $x \approx y$ if and only if $x = \pm y + 2n, n \in \mathbf{Z}$. We can think of the pleating locus $pl(\rho)$ as a $\approx$–equivalence class in $\mathbf{R}$. Then the map

$$\Pi\colon \mathcal{R} \to \mathbf{R}/\approx \times \mathbf{R}^+,\ \Pi(\rho) = (pl(\rho), L_{pl(\rho)}(\rho)),$$

factors through j. We denote the induced map, $\tilde{\Pi}$.

We obtain:

Theorem 3.1 *The map*

$$\tilde{\Pi}\colon \mathcal{R}/j \to \mathbf{R}/\approx \times \mathbf{R}^+$$

is a homeomorphism.

References

[1] **J Birman**, **C Series**, *Algebraic linearity for an automoprhism of a surface group*, Journal of Pure and Applied Algebra 52 (1988) 227–275

[2] **M Lyubich**, **V. Suvorov**, *Free subgroups of* $SL(2,\mathbf{C})$ *with two parabolic generators* , J. Soviet Math. 141 (1988) 976–979

[3] **L Keen**, **C Series**, *Pleating coordinates for the Maskit embedding of the Teichmüller space of punctured tori*, Topology 32 (1993) 719–749

[4] **L Keen**, **C Series**, *The Riley slice of Schottky space*, Proc. London Math. Soc. 69 (1994) 72–90

[5] **L Keen**, **C Series**, *Pleating invariants for punctured torus groups*, Warwick University preprint (1998)

[6] **L Keen**, **B Maskit**, **C Series**, *Geometric finiteness and uniqueness for Kleinian groups with circle packing limit sets*, J. Reine und Angew. Math. 436 (1993) 209–219

[7] **L Keen**, **J Parker**, **C Series**, *Combinatorics of simple closed curves on the twice punctured torus*, Israel J. Math. to appear

[8] **Y Komori**, **C Series**, *Pleating coordinates for the Earle embedding*, preprint

[9] **J-P Otal**, *Courants géodésiques et surfaces*, Thèse, Orsay, Univ. de Paris-Sud (1989)

[10] **C Series**, *The Geometry of Markoff Numbers*, Math. Intelligencer 7 (1985) 20–29

Department of Mathematics, Osaka City University, Osaka 558, Japan
and
Mathematics Institute, Warwick University, Coventry CV4 7AL, England

Email: `komori@sci.osaka-cu.ac.jp, cms@math.warwick.ac.uk`

Received: 27 November 1997

ISSN 1464-8997 **317**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 317–334

On the continuity of bending

Christos Kourouniotis

Abstract We examine the dependence of the deformation obtained by bending quasi-Fuchsian structures on the bending lamination. We show that when we consider bending quasi-Fuchsian structures on a closed surface, the conditions obtained by Epstein and Marden to relate weak convergence of arbitrary laminations to the convergence of bending cocycles are not necessary. Bending may not be continuous on the set of all measured laminations. However we show that if we restrict our attention to laminations with non negative real and imaginary parts then the deformation depends continuously on the lamination.

AMS Classification 30F40; 32G15

Keywords Kleinian groups, quasi-Fuchsian groups, geodesic laminations

The deformation of hyperbolic structures by bending along totally geodesic submanifolds of codimension one was introduced by Thurston in his lectures on *The Geometry and Topology of 3–manifolds*. The geometric and algebraic properties of the deformation were studied in [4] and [3]. Epstein and Marden [2] introduced the notion of a bending cocycle and used it to describe bending a hyperbolic surface along a measured geodesic lamination. The same notion was used in [5] to extend bending to a holomorphic family of local biholomorphic homeomorphisms of quasi-Fuchsian space $Q(S)$.

Epstein and Marden [2] give a careful analysis of the dependence of the bending cocycle on the measured lamination. They consider the set of measured laminations on $\mathcal{H}^2$ consisting of geodesics that intersect a compact subset $K \subset \mathcal{H}^2$. This is a subset of the space of measures on the space $G(K)$ of geodesics in $\mathcal{H}^2$ intersecting K, with the topology of weak convergence of measures. In this topology, the bending cocycle does not depend continuously on the lamination. One reason for this is the behaviour of the laminations near the endpoints of the segment over which we evaluate the cocycle. For example, consider the geodesic segment $[e^{i\theta}, i]$ in $\mathcal{H}^2$, for suitable θ in $[0, \pi/2]$, and the measured laminations μ_n, with weight 1 on the geodesic $(1/n, n)$ and weight -1 on the geodesic $(-1/n, -n)$. Then $\{\mu_n\}$ converges weakly to the zero lamination, but the cocycle of μ_n relative to $[e^{i\theta}, i]$ is approximately a hyperbolic isometry of translation length 1. Epstein and Marden find conditions under which a sequence of measured laminations gives a convergent sequence of cocycles relative to a given pair of points.

In this article we show that when the lamination is invariant by a discrete group and we only consider cocycles relative to points in the orbit of a suitable point $x \in \mathcal{H}^2$, any sequence of measured laminations $\{\mu_n\}$ which converges weakly gives rise to cocycles which converge up to conjugation. We show further that the same conjugating elements can be used for the cocycles for μ_n corresponding to the different generators of the group. Hence the laminations μ_n determine bending homomorphisms which, after conjugation by suitable isometries, converge to the bending homomorphism determined by μ_0. This implies that the deformations converge in $Q(S)$.

Theorem 1 *Let S be a closed hyperbolic surface and $Q(S)$ its space of quasi-Fuchsian structures. Let $\{\mu_n\}$ be a sequence of complex measured geodesic laminations, converging weakly to a lamination μ_0. Then the bending deformations*

$$B_{\mu_n}\colon\ \mathcal{D}_{\mu_n} \to Q(S)$$

converge to the deformation B_{μ_0}, uniformly on compact subsets of $\mathcal{D} = \mathcal{D}_{\mu_0} \cap (\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \mathcal{D}_{\mu_n})$.

We also state an infinitesimal version of the Theorem.

Theorem 2 *Let S be a closed hyperbolic surface and $Q(S)$ its space of quasi-Fuchsian structures. Let $\{\mu_n\}$ be a sequence of complex measured geodesic laminations, converging weakly to a lamination μ_0. Then the holomorphic bending vector fields T_{μ_n} on $Q(S)$ converge to T_{μ_0}, uniformly on compact subsets of $Q(S)$.*

These results do not necessarily imply the continuous dependence of the deformation on the bending lamination, because the space of measured laminations is not first countable. If however we restrict our attention to the subset of measured laminations with non negative real and imaginary parts, then we can apply results in [6] to obtain the following Theorem.

Theorem 3 *The mapping $\mathcal{ML}^{++}(S) \times Q(S) \to T(Q(S))\colon\ (\mu, [\rho]) \mapsto T_\mu([\rho])$ is continuous, and holomorphic in $[\rho]$.*

The proof of Theorem 1 is based on the observation that, when the lamination is invariant by a discrete group and we are considering cocycles with respect to points x and $g(x)$, for some g in the group, the effect of a lamination near the endpoints of the segment $[x, g(x)]$ is controlled by its effect near x, provided that the lamination does not contain geodesics very close to the geodesic carrying $[x, g(x)]$. This last condition can be achieved by choosing x to be a point not on the axis of a conjugate of g (see Corollary 2.12).

In Section 1 we describe the space of measured laminations and we recall the definition of bending. In the beginning of Section 2 we recall or modify certain results from [2] and [5] which provide bounds for the effect of bending along nearby geodesics. Lemma 2.11 and the results following it examine the consequences of the above condition on the choice of x.

The proof of Theorems 1, 2 and 3 is given in Section 3. The laminations μ_n are replaced by finite approximations. The main result is Lemma 3.1, which gives the basic estimate for the difference between the bending homomorphism of μ_0 and a conjugate of the bending homomorphism of μ_n. Then a diagonal argument is used to obtain the convergence of bending.

1 The setting

We consider a closed surface S of genus greater than 1. We fix a hyperbolic structure on S, and let $\rho_0\colon \pi_1(S) \to PSL(2,\mathbb{R})$ be an injective homomorphism with discrete image $\Gamma_0 = \rho_0(\pi_1(S)$, such that S is isometric to $\mathcal{H}^2/\Gamma_0$.

We consider the space R of injective homomorphisms $\rho\colon \Gamma_0 \to PSL(2,\mathbb{C})$ obtained by conjugation with a quasiconformal homeomorphism ϕ of $\widehat{\mathbb{C}}$: if $g \in \Gamma_0$, acting on $\widehat{\mathbb{C}}$ as Möbius transformations, then $\rho(g) = \phi \circ g \circ \phi^{-1}$.

$PSL(2,\mathbb{C})$ acts on the left on R by inner automorphisms. The quotient of R by this action is the *space* $Q(S)$ *of quasi-Fuchsian structures* on S, or *quasi-Fuchsian space* of S. We denote the equivalence class in $Q(S)$ of a homomorphism $\rho \in R$ by $[\rho]$. Then $[\rho]$ is a *Fuchsian point* if there is a circle in $\widehat{\mathbb{C}}$ left invariant by $\rho(\Gamma_0)$, so that $\rho(\Gamma_0)$ is conjugate to a Fuchsian group of the first kind. The subset of Fuchsian points in $Q(S)$ is the *Teichmüller* space of S, $T(S)$.

We fix a point $[\rho] \in Q(S)$, represented by the homomorphism $\rho\colon \Gamma_0 \to PSL(2,\mathbb{C})$ obtained by conjugation with the quasiconformal homeomorphism $\phi\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. We denote the image of ρ by Γ. The limit set of Γ_0 is $\widehat{\mathbb{R}}$. Then

$\phi(\widehat{\mathbb{R}})$ is the limit set of Γ. If γ is a geodesic in $\mathcal{H}^2$ with endpoints $u, v \in \widehat{\mathbb{R}}$, we denote by $\phi_*(\gamma)$ the geodesic in $\mathcal{H}^3$ with endpoints $\phi(u), \phi(v)$ in $\phi(\widehat{\mathbb{R}})$. In this way, geodesics on the surface $S \cong \mathcal{H}^2/\Gamma_0$ are associated to geodesics in the hyperbolic 3–manifold $\mathcal{H}^3/\Gamma$.

We want to study the deformation of quasi-Fuchsian structures by *bending*, [4], [2], [5]. Bending is determined by a geodesic lamination on S with a complex valued transverse measure.

A measured geodesic lamination on S lifts to a measured geodesic lamination on $\mathcal{H}^2$. The space $G(\mathcal{H}^2)$ of unoriented geodesics in $\mathcal{H}^2$ is homeomorphic to a Möbius strip without boundary. Let K be a compact subset of $\mathcal{H}^2$, projecting onto $\mathcal{H}^2/\Gamma_0$. The set $G(K)$ of geodesics in $\mathcal{H}^2$ intersecting K is a compact metrizable space.

A measured geodesic lamination on $\mathcal{H}^2$ determines a complex valued Borel measure μ on $G(K)$, with the property that if γ_1 and γ_2 are distinct geodesics in the support of μ, then they are disjoint. The set of measured geodesic laminations on S can be considered as a subset of $\mathcal{M}(G(K))$, the set of complex valued Borel measures on $G(K)$. The set $\mathcal{M}(G(K))$ has a norm, defined by

$$\|\mu\| = \sup\left\{\left|\int f\mu\right|, f \text{ continuous complex valued function on } G(K), |f| \le 1\right\}$$

We shall use the weak* topology on $\mathcal{M}(G(K))$, with basis the sets of the form

$$U(\mu, \varepsilon, f_1, \dots, f_m) = \left\{\nu \in \mathcal{M}(G(K)) : \left|\int f_i\mu - \int f_i\nu\right| < \varepsilon, i = 1, \dots, m\right\}$$

where $\mu \in \mathcal{M}(G(K))$, f_i, $i = 1, \dots, m$ are continuous functions on $G(K)$, and ε is a positive number.

A measured geodesic lamination μ on S is called *finite* if it is supported on a finite set of simple closed geodesics in S. Then, for any compact subset K of $\mathcal{H}^2$, the measure on $G(K)$ determined by the lift of μ to $\mathcal{H}^2$ has finite support.

Given a finite measured geodesic lamination μ on S, we define bending the quasi-Fuchsian structure $[\rho]$ on S as follows.

Let $g_1, \dots, g_k$ be a set of generators of Γ_0. Choose a point x on $\mathcal{H}^2$ and, for each g_j, consider the geodesic segment $[x, g_j(x)]$. Let $\gamma_1, \dots, \gamma_m$ be the geodesics in the support of μ intersecting $[x, g_j(x)]$, and let $z_1, \dots, z_m$ be the corresponding measures. If γ_1 (or γ_m) go through x (or $g_j(x)$ respectively), we replace z_1 (or z_m) by $\frac{1}{2}z_1$ (or $\frac{1}{2}z_m$).

If γ is an oriented geodesic in $\mathcal{H}^3$ and $z \in \mathbb{C}$, we denote by $A(\gamma, z)$ the element of $PSL(2, \mathbb{C})$ with axis γ and complex displacement z. We will use the same notation for one of the matrices in $SL(2,\mathbb{C})$ corresponding to $A(\gamma,z)$. In such cases either the choice of the lift will not matter, or there will be an obvious choice.

We orient the geodesics $\gamma_1,\dots,\gamma_m$ so that they cross the segment $[x,g_j(x)]$ from right to left, and define the isometry

$$C_{t\mu}(x,g_j(x)) = A(\phi_*(\gamma_1),tz_1)\cdots A(\phi_*(\gamma_m),tz_m).$$

For each generator $g_j,\ j=1,\dots,k$, define

$$\rho_{t\mu}(g_j) = C_{t\mu}(x,g_j(x))\ \rho(g_j).$$

For t in an open neighbourhood of 0 in $\mathbb{C}$, the representation $[\rho_{t\mu}]$ is quasi-Fuchsian, [4].

Any measured geodesic lamination μ on S can be approximated by finite laminations so that the corresponding bending deformations converge, [2], [5]. In this way, we obtain for any measured geodesic lamination on S a deformation B_μ defined on an open set $\mathcal{D}_\mu \subset Q(S)\times\mathbb{C}$,

$$B_\mu\colon\ \mathcal{D}_\mu \to Q(S)\colon\ ([\rho],t)\mapsto[\rho_{t\mu}].$$

B_μ is a holomorphic mapping.

2 The lemmata

In the vector space $\mathbb{C}^2$ we introduce the norm

$$\|(z_1,z_2)\| = \max\{|z_1|,|z_2|\}.$$

A complex matrix $A=\begin{pmatrix} a & b\\ c & d\end{pmatrix}$ acts on $\mathbb{C}^2$ and has norm

$$\|A\| = \max\{|a|+|b|,|c|+|d|\}.$$

We will use this norm on $SL(2,\mathbb{C})$.

Lemma 2.1 ([2], 3.3.1) *Let X be a set of matrices in $SL(2,\mathbb{C})$ and $c=(0,0,1)\in\mathcal{H}^3$. Then the following are equivalent.*

i) *The closure of X is compact.*

ii) *There is a positive number M such that if $A\in X$ then $||A||\le M$.*

iii) *There is a positive number M such that if $A\in X$ then $||A||\le M$ and $||A^{-1}||\le M$.*

iv) *There is a positive number* R *such that if* $A \in X$ *then* $d(c, A(c)) \leq R$. □

Let Λ be a maximal geodesic lamination on S, and $\psi\colon S \to \mathcal{H}^3/\Gamma$ the pleated surface representing the lamination Λ [1]. Let $\tilde{\psi}\colon \mathcal{H}^2 \to \mathcal{H}^3$ be the lift of ψ.

Lemma 2.2 ([5], 2.5) *Let* K *be a compact disc of radius* R *about* $c = (0,0,1) \in \mathcal{H}^3$, *and* M *a positive number. There is a positive number* N *with the following property. If* $[x,y]$ *is a geodesic segment in* $\mathcal{H}^2$ *such that* $\tilde{\psi}([x,y]) \subset K$ *and* $\{\gamma_i, z_i\}$, $i = 1, \ldots, m$ *is a finite measured lamination with support contained in* Λ, *whose leaves all intersect* $[x,y]$ *and are numbered in order from* x *to* y, *and such that* $\sum_{i=1}^{m} |\mathrm{Re}\, z_i| < M$, *then*

$$\|A(\gamma_1, z_1) \cdots A(\gamma_m, z_m)\| \leq N.$$ □

Lemma 2.3 ([2], 3.4.1, [5], 2.4) *Let* K *be a compact subset of* $SL(2,\mathbb{C})$, M *a positive number, and let* γ *be the geodesic* $(0,\infty)$. *Then there is a positive number* N *with the following property. For any* $B, C \in K$, *and* $z \in \mathbb{C}$ *with* $|z| \leq M$, *we have*

$$\left\|BA(\gamma,z)B^{-1} - CA(\gamma,z)C^{-1}\right\| \leq N \left\|B - C\right\| |z| \, .$$ □

In order to examine the effect of bending along nearby geodesics, in Lemma 2.5 and 2.6, we shall use the notion of a solid cylinder in hyperbolic space. A *solid cylinder* C *over a disk* D in $\mathcal{H}^n$ is the union of all geodesics orthogonal to a $(n-1)$–dimensional hyperbolic disc D in $\mathcal{H}^n$. The *radius* of the cylinder is the hyperbolic radius of the disc D. If x is the centre of D, we say that C is a solid cylinder *based* at x. The boundary of C at infinity consists of two discs D_1 and D_2 in $\partial\mathcal{H}^n$. We say that the solid cylinder C is *supported* by D_1 and D_2. The geodesic orthogonal to D through its centre is the *core* of the solid cylinder C. We shall denote the cylinder with core γ, basepoint $x \in \gamma$ and radius r by $C(\gamma, x, r)$.

Lemma 2.4 ([5], 2.6) *Let* L *be a compact set in* $\mathcal{H}^3$. *Then there exists a positive number* M *with the following property. If* D *is a disc of radius* r, *contained in* L, *and* α, β *are two geodesics contained in the solid cylinder over* D, *then there is an element* $A \in SL(2,\mathbb{C})$ *such that* $A(\alpha) = \beta$ *and* $||A - I|| \leq Mr$. □

If C is a solid cylinder supported on the discs D_1 and D_2, with $D_1 \cap D_2 = \emptyset$, and γ_1, γ_2 are two geodesics, each having one end point in D_1 and one in D_2, we say that γ_1 and γ_2 are *concurrently oriented* in C if their origins lie in the same component of $D_1 \cup D_2$.

Lemma 2.5 *Let m be a positive number and L a compact subset of $\mathcal{H}^3$. Then there are positive numbers M_1 and M_2 with the following property. If γ_1, γ_2 are concurrently oriented geodesics contained in a cylinder of radius r, based at a point in L, and z_1, z_2 are complex numbers such that $|z_i| \leq m$, then there are lifts of $A(\gamma_i, z_i)$ to $SL(2,\mathbb{C})$ such that*

$$\|A(\gamma_1, z_1) - A(\gamma_2, z_2)\| \leq M_1 r \min\{|z_1|, |z_2|\} + M_2|z_1 - z_2|.$$

Proof We assume that $|z_1| \leq |z_2|$. We have

$$\|A(\gamma_1, z_1) - A(\gamma_2, z_2)\| \leq \|A(\gamma_1, z_1) - A(\gamma_2, z_1)\| + \|A(\gamma_2, z_1) - A(\gamma_2, z_2)\|.$$

Let $B \in SL(2,\mathbb{C})$ be an element mapping the geodesic $(0,\infty)$ to γ_2, and mapping the point $c = (0,0,1)$ to a point in L. Then, by Lemma 2.1, there is a constant K_1 depending only on L, such that $||B|| \leq K_1$. By Lemma 2.4 there is an element $C \in SL(2,\mathbb{C})$ such that $C(\gamma_2) = \gamma_1$, and $||C - I|| \leq K_2 r$ for some constant K_2 depending only on L.

By Lemma 2.3 there is a constant K_3 such that

$$\|A(\gamma_1, z_1) - A(\gamma_2, z_1)\| \leq K_3 \|CB - B\| \, |z_1| \leq K_1 K_2 K_3 r |z_1|.$$

On the other hand,

$$\|A(\gamma_2, z_1) - A(\gamma_2, z_2)\| \leq \|B\| \, \|A((0,\infty), z_1 - z_2) - I\| \, \left\|B^{-1}\right\| \, \|A((0,\infty), z_2)\|.$$

By Lemma 2.1 and the fact that the entries of $A((0,\infty), z_1 - z_2)$ depend analytically on $z_1 - z_2$, there is a constant K_4, depending on L and m such that

$$\|A(\gamma_2, z_1) - A(\gamma_2, z_2)\| \leq K_4|z_1 - z_2|.$$

□

Lemma 2.6 ([5], 2.7) *Let m be a positive number and L a compact subset of $\mathcal{H}^3$. Then there is a positive number M with the following property. Let C be a solid cylinder of radius r based at a point in L. Let $\gamma_1, \ldots, \gamma_k$ be geodesics in C and $z_1, \ldots, z_k$ complex numbers with $\sum_{i=1}^k |\mathrm{Re}\,(z_i)| \leq m$. Then*

$$\left\| A(\gamma_1, z_1) \cdots A(\gamma_k, z_k) - A\left(\gamma_1, \sum_{i=1}^k z_i\right) \right\| \leq Mr \sum_{i=1}^k |z_i|.$$

□

We want to show that if two geodesics on S are sufficiently close, then the corresponding geodesics in $\mathcal{H}^3/\Gamma$ will also be close, (Lemma 2.10).

Lemma 2.7 *Let K be a compact subset of $\mathcal{H}^2$, and $\phi\colon \partial\mathcal{H}^2 \to \partial\mathcal{H}^3$ a homeomorphism onto its image. Then there is a compact subset L of $\mathcal{H}^3$ such that if γ is a geodesic of $\mathcal{H}^2$ intersecting K, then $\phi_*(\gamma)$ intersects L, i.e. $\phi_*(G(K)) \subset G(L)$.*

Proof We consider the Poincaré disk model of hyperbolic space. There, it is clear that if K is a compact subset of B^2, then there is a positive number m such that if γ is a geodesic in $G(K)$ with end-points u, v, then $|u - v| \geq m$. Since ϕ^{-1} is uniformly continuous, there is a positive number M such that $|\phi(u) - \phi(v)| \geq M$, and hence there is a compact subset of B^3 intersecting $\phi_*(\gamma)$. □

Lemma 2.8 ([5], 2.2) *Let ε and η be two positive numbers. Then there is a positive number δ with the following property. If D_1 and D_2 are discs in S^2, with spherical radius $\leq \delta$, and the spherical distance between D_1 and D_2 is $\geq \eta$, then the solid cylinder supported by D_1 and D_2 has hyperbolic radius $r \leq \varepsilon$.* □

Lemma 2.9 *Let K be a compact subset of B^n, and d a positive number. Then there is a positive number δ with the following property. If C is a solid cylinder in B^n, over a disc with radius $r \leq \delta$ and centre at a point in K, then the spherical radius of each of the discs supporting C is $\leq d$.*

Proof The radii of the supporting discs are given by continuous functions of the core geodesic, the base point and the radius of the cylinder. For a fixed base point, they tend to zero with the radius of the cylinder. The result follows by compactness. □

Lemma 2.10 *Let $[\rho]$ be a quasi-Fuchsian structure on S, K a compact subset of $\mathcal{H}^2$, and L a compact subset of $\mathcal{H}^3$ such that $\phi_*(G(K)) \subset G(L)$. Let r be a positive number. Then there is a positive number δ with the following property. If $\gamma \in G(K)$, $x \in \gamma \cap K$ and $0 \leq r_1 \leq \delta$, then there is some point $x' \in L$ such that for any geodesic α contained in the solid cylinder $C(\gamma, x, r_1)$, the geodesic $\phi_*(\alpha)$ is contained in the solid cylinder $C(\phi_*(\gamma), x', r) \subset \mathcal{H}^3$.*

Proof We work in the Poincaré disc model of the hyperbolic plane and space, B^2 and B^3. Since L is a compact subset of B^3, there is a number $\eta_2 > 0$ such that if u and v are the endpoints of any geodesic in B^3 intersecting L, then the spherical distance between u and v is $\geq \eta_2$. Then, by Lemma 2.8, there is a positive number δ_2, such that any solid cylinder with core a geodesic $\gamma \in G(L)$ and supported on discs of spherical radius $\leq \delta_2$, has hyperbolic radius $\leq r$.

Since $\phi\colon S^1 \to S^2$ is uniformly continuous, there is a positive number δ_1, such that any arc in S^1 of length $\leq \delta_1$ is mapped into a disc in S^2, of radius $\leq \delta_2$. Then, by Lemma 2.9, there is a positive number δ such that any solid cylinder of radius $\leq \delta$ and based at a point in K, is supported on two arcs of length $\leq \delta_1$. □

Recall that, if X is a subset of $\mathcal{H}^2$, we denote by $G(X)$ the set of geodesics in $\mathcal{H}^2$ which intersect X. To simplify notation, we will write $G(x)$ for the set of geodesics through the point $x \in \mathcal{H}^2$, and $G(x,y)$ for the set of geodesics intersecting the open geodesic segment (x,y).

If Γ is a group of isometries of $\mathcal{H}^2$, we denote by G'_Γ the set of geodesics in $\mathcal{H}^2$ which do not intersect any of their translates by Γ:

$$G'_\Gamma = \{\gamma \in G(\mathcal{H}^2) : \forall g \in \Gamma, g(\gamma) \cap \gamma = \emptyset \text{ or } g(\gamma) = \gamma\}.$$

In the following Lemma we consider the angle between unoriented geodesics to lie in the interval $[0, \frac{\pi}{2}]$.

Lemma 2.11 *Let ℓ and θ be positive numbers. Then there is a positive number ζ with the following property. Let $x, y \in \mathcal{H}^2$, γ the geodesic carrying the segment $[x,y]$, $g \in PSL(2,\mathbb{R})$ and $\gamma' \in G'_{\langle g \rangle}$, such that:*

i) *The hyperbolic distance $d(x,y) \leq \ell$.*
ii) *The geodesic segments $[x,y]$ and $[g(x), g(y)]$ intersect, and the angle between γ and $g(\gamma)$ is $\alpha \geq \theta$.*
iii) *γ' intersects the segment $[x,y]$ and the angle between γ and γ' is β.*

Then $\beta \geq \zeta$.

Proof Without loss of generality, we may asume that $x = i \in \mathcal{H}^2$ and $y = ti$. The angle of intersection between the geodesics δ and $g(\delta)$ is a continuous function of δ. Hence there is a neighbourhood U of $\gamma \in G(\mathcal{H}^2)$ disjoint from $G'_{\langle g \rangle}$, that is consisting of geodesics δ such that $g(\delta)$ intersects δ.

There is a positive number r such that the (two dimensional) solid cylinder $C(\gamma, i\sqrt{t}, r)$ has the property: if $\delta \subset C(\gamma, i\sqrt{t}, r)$ then $\delta \in U$. Then it is easy to show, using hyperbolic trigonometry, that there is a positive number ζ such that any geodesic δ intersecting $[x,y]$ at an angle $\leq \zeta$ is contained in $C(\gamma, i\sqrt{t}, r)$, and hence $\delta \notin G'_{\langle g \rangle}$. □

Corollary 2.12 *If g is a hyperbolic isometry of $\mathcal{H}^2$ and $x \in \mathcal{H}^2$ does not lie on the axis of g, then there is a positive number ζ with the following property. If μ is any geodesic lamination invariant by g, then no leaf of the lamination intersects the geodesic segment $[x, g(x)]$ at an angle smaller than ζ.* □

Lemma 2.13 *Let ℓ, θ and ε be positive numbers. Then there is a positive number r with the following property. Let $x, y \in \mathcal{H}^2$ with $d(x,y) \leq \ell$, and let γ be the geodesic carrying the segment $[x,y]$. Let $g \in PSL(2,\mathbb{R})$ be such that $[x,y]$ intersects $[g(x), g(y)]$ at the point x_0, and at an angle $\alpha \geq \theta$. If $\delta \in G'_{\langle g \rangle} \cap G(D(x_0, r))$, then δ intersects both γ and $g(\gamma)$, and the points of intersection lie in $D(x_0, \varepsilon)$.*

Proof Since $g^{-1}(x_0) \in [x,y]$, we have $d(g^{-1}(x_0), x_0) \leq \ell$. We consider the geodesic segment $[x', y']$ of length 3ℓ on the geodesic γ, centred at x_0.

Let U be a neighbourhood of $\gamma \in G(\mathcal{H}^2)$ disjoint from $G'_{\langle g \rangle}$. There is r_1 such that any geodesic which intersects $D(x_0, r_1)$ and does not intersect $[x', y']$, lies in U, and hence it is not in $G'_{\langle g \rangle}$. So, if $\delta \in G'_{\langle g \rangle} \cap G(D(x_0, r_1))$, δ intersects the segment $[x', y']$. Similarly, there is r_2 such that if $\delta \in G'_{\langle g \rangle} \cap G(D(x_0, r_2))$, δ intersects the segment $[g(x'), g(y')]$.

By Lemma 2.11, the angle at the points of intersection is greater than a constant ζ. If r satisfies $0 < r < \min(r_1, r_2)$ and $\sinh r < \sin\zeta \sinh\varepsilon$, then it has the required property. □

The following Lemma shows that, under certain conditions, taking integrals along geodesic segments describes weak convergence of measures.

Lemma 2.14 *Let $\{\mu_n\}$ be a sequence of measured geodesic laminations on $\mathcal{H}^2$, invariant by $g \in PSL(2,\mathbb{R})$, and assume that μ_n converge weakly to a measured lamination μ. Let γ be a geodesic in $\mathcal{H}^2$, such that γ and $g(\gamma)$ intersect at one point. Then, for every geodesic segment $[u,v]$ on γ and for every continuous function $f\colon [u,v] \to [0,1]$, with $f(u) = f(v) = 0$, the sequence $\int_{[u,v]} f\mu_n$ converges to $\int_{[u,v]} f\mu$.*

Proof Since γ intersects $g(\gamma)$ at one point, there is a neighbourhood U of γ in $G(\mathcal{H}^2)$ which is disjoint from $G'_{\langle g \rangle}$. We define a continuous function $\tilde{f}\colon G(\mathcal{H}^2) \to [0,1]$ by letting $\tilde{f}(\delta) = f(y)$ if $y \in [u,v]$ and $\delta \in G(y) - U$, and extending continuously to the rest of $G(\mathcal{H}^2)$. Then, for any measured geodesic lamination ν invariant by g,

$$\tilde{f}\nu(G(u,v)) = \int_{[u,v]} f\nu.$$ □

3 The theorems

We fix a reference point $[\rho_0] \in T(S)$, and we consider a point $[\rho] \in Q(S)$. Let $g_1, \ldots, g_k \in PSL(2,\mathbb{R})$ be a set of generators for $\Gamma_0 = \rho_0(\pi_1(S))$. Let $x \in \mathcal{H}^2$ be a point which does not lie on the axis of any conjugate of the generators g_j.

Let θ be the minimum of the angles between the geodesics carrying the segments $[g_j^{-1}(x), x]$ and $[x, g_j(x)]$, for $j = 1, \ldots, k$. Let d and d' be the maximum and the minimum, respectively, of the distances between x and $g_j(x)$, for $j = 1, \ldots, k$.

Let K be a compact disc in $\mathcal{H}^2$ containing in its interior the points x, $g_j(x)$, $g_j^{-1}(x)$, for $j = 1, \ldots, k$, and projecting onto $S_0 = \mathcal{H}^2/\Gamma_0$. Let L be a compact disc in $\mathcal{H}^3$ such that $\phi_*(G(K)) \subset G(L)$.

We consider a positive integer m, and a positive number $r(m)$ such that d/m is less than the number $\delta(K, L, r(m))$ given by Lemma 2.10.

Let μ be a complex measured geodesic lamination on $\mathcal{H}^2$, invariant by the group Γ_0, with $||\mu|| < M_0$. We consider one of the generators g_j, $j = 1, \ldots, k$, and to simplify notation we drop the suffix j for the time being. Let γ denote the geodesic carrying the segment $[x, g(x)]$. We divide the segment $[x, g(x)]$ into m equal subsegments, by the points

$$x = x_0, x_1, \ldots, x_{m-1}, x_m = g(x).$$

If $[x, y]$ is a geodesic segment in $\mathcal{H}^2$ and ν is a measure on a set of geodesics in $\mathcal{H}^2$, we introduce the notation

$$\int'_{[x,y]} \nu = \frac{1}{2}\nu\left(G\left(x\right)\right) + \nu\left(G\left(x,y\right)\right) + \frac{1}{2}\nu\left(G\left(y\right)\right)$$

We define two new measures on the set $G(\mathcal{H}^2)$ of geodesics in $\mathcal{H}^2$ in the following way. For every $i = 1, \ldots, m$, let $\tilde{\gamma}_i$ be a geodesic in $\operatorname{supp}\mu$, intersecting γ in $[x_{i-1}, x_i]$. We define, for $i = 1, \ldots, m$,

$$\tilde{\mu}(\tilde{\gamma}_i) = \int'_{[x_{i-1},x_i]} \mu.$$

For every $i = 1, \ldots, m-1$, let γ'_i be the geodesic in $\operatorname{supp}\mu$ intersecting the open segment (x_{i-1}, x_{i+1}) as near as possible to x_i. Let $\lambda_i\colon [x_0, x_m] \to [0,1]$, $i = 1, \ldots, m-1$, be continuous functions satisfying

(1) $\operatorname{supp}(\lambda_i) \subset [x_{i-1}, x_{i+1}]$ and

(2) $\sum_{i=1}^{m-1} \lambda_i(x) = 1$ for all $x \in [x_0, x_m]$.

Then, in particular, $[x_0, x_1] \subset \lambda_i^{-1}(1)$ and $[x_{m-1}, x_m] \subset \lambda_{m-1}^{-1}(1)$. We define, for $i = 1, \ldots, m-1$,

$$\mu'(\gamma_i') = \int_{[x_{i-1}, x_{i+1}]} \lambda_i \mu$$

Now we define

$$C_i = A(\phi_*(\tilde{\gamma}_i), \tilde{\mu}(\tilde{\gamma}_i)) \qquad \text{for } i = 1, \ldots, m$$

and

$$D_i = A(\phi_*(\gamma_i'), \mu'(\gamma_i')) \qquad \text{for } i = 1, \ldots, m-1.$$

We want to bound the norm $||C_1 C_2 \cdots C_m - D_1 D_2 \cdots D_{m-1}||$.

We put $a_i = \int'_{[x_{i-1}, x_i]} \lambda_i \mu$ and $b_i = \int'_{[x_i, x_{i+1}]} \lambda_i \mu$. Then $\mu'(\gamma_i') = a_i + b_i$, for $i = 1, \ldots, m-1$, and $\tilde{\mu}(\tilde{\gamma}_1) = a_1$, $\tilde{\mu}(\tilde{\gamma}_m) = b_{m-1}$, and for $i = 2, \ldots, m-1$, $\tilde{\mu}(\tilde{\gamma}_i) = b_{i-1} + a_i$.

We put $D_i^l = A(\phi_*(\gamma_i'), a_i)$ and $D_i^r = A(\phi_*(\gamma_i'), b_i)$. With this notation we have

$$\begin{aligned} \|C_1 \cdots C_m - D_1 \cdots D_{m-1}\| \leq \\ & \|C_1 \cdots C_{m-1}\| \left\|C_m - D_{m-1}^r\right\| \\ & + \|C_1 \cdots C_{m-2}\| \left\|C_{m-1} - D_{m-2}^r D_{m-1}^l\right\| \left\|D_{m-1}^r\right\| \\ & + \cdots + \|C_1 \cdots C_{s-1}\| \left\|C_s - D_{s-1}^r D_s^l\right\| \|D_s^r D_{s+1} \cdots D_{m-1}\| \\ & + \cdots + \left\|C_1 - D_1^l\right\| \|D_1^r D_2 \cdots D_{m-1}\| . \end{aligned}$$

Then, by Lemma 2.2, there is a positive number M_1, depending on L and M_0, which is an upper bound for the norm of the factors of the form $C_1 \cdots C_s$, $D_s^r D_{s+1} \cdots D_{m-1}$. By Lemma 2.6, there is a positive number M_2, depending on L and M_0, such that each factor of the form $C_s - D_{s-1}^r D_s^l$ has norm bounded by $M_2 r(m) \tilde{\mu}(\tilde{\gamma}_s)$. Then

$$\|C_1 \cdots C_m - D_1 \cdots D_{m-1}\| \quad \leq \quad M_0 M_1^2 M_2 r(m). \tag{1}$$

In the following we want to examine the behaviour of $D_1 \cdots D_{m-1}$ as $m \to \infty$ and as the lamination μ changes. For this we must consider more carefully the leaves of the lamination near x.

By Lemma 2.13, there is an open set $U \subset G(K)$, depending on d, θ and d'/m such that, if δ is any geodesic in $U \cap \operatorname{supp} \mu$, then δ intersects the geodesics

γ and $g(\gamma)$ at a distance less than d'/m from x. Let $\chi\colon G(K) \to [0,1]$ be a continuous function, with $\operatorname{supp}\chi \subset U$ and $\chi|_{G(x)} = 1$. We introduce the notation

$$a' = \int_{[x_0,x_1]} \chi\mu \qquad\qquad a'' = \int'_{[x_0,x_1]} (1-\chi)\mu$$

$$b' = \int_{[x_{m-1},x_m]} (\chi\circ g^{-1})\mu \qquad\qquad b'' = \int'_{[x_{m-1},x_m]} (1-\chi\circ g^{-1})\mu$$

$$P = A(\phi_*(\gamma'_1), a') \qquad\qquad Q = A(\phi_*(\gamma'_1), a'')$$

$$R = A(\phi_*(\gamma'_{m-1}), b'') \qquad\qquad S = A(\phi_*(\gamma'_{m-1}), b'),$$

and we have

$$D_1 = PQD_1^r \qquad\qquad D_{m-1} = D_{m-1}^l RS.$$

Let $\{\mu_n\}$ be a sequence of complex measured geodesic laminations on the surface S_0, converging weakly in $\mathcal{M}(G(K))$ to a measured lamination μ_0. Then, by the Uniform Boundedness Principle, there is a positive number M_0 such that $||\mu_n|| \le M_0$ for all $n \ge 0$.

For each positive integer m, for each $i = 1, \ldots, m-1$, for each $j = 1, \ldots, k$ and for each measured lamination μ_n, $n \ge 0$, we define as above the points $x_{j,m,i}$, the geodesics $\gamma'_{n,j,m,i}$, the functions $\lambda_{j,m,i}$, the quantities $a_{n,,j,m,i}$, $b_{n,j,m,i}$, $a'_{n,j,m}$, $b'_{n,j,m}$ and the isometries $D_{n,j,m,i}$, $P_{n,j,m}$, $Q_{n,j,m}$, $R_{n,j,m}$, $S_{n,j,m}$.

Let $B_{n,j,m} = D_{n,j,m,1} \cdots D_{n,j,m,m-1}$. We want to find a bound for the norm of the difference between $B_{0,j,m}g_j$ and some conjugate of $B_{n,j,m}g_j$.

Lemma 3.1 *With the above notation, there exist positive numbers N_1, N_2 and functions $r\colon \mathbb{N} \to \mathbb{R}$, $\varepsilon\colon \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ such that*

$$\lim_{m\to\infty} r(m) = 0, \qquad \lim_{n\to\infty} \varepsilon(m,n) = 0 \quad \textit{for each } m \in \mathbb{N}$$

and

$$\left\| P_{0,1,m}P_{n,1,m}^{-1}B_{n,j,m}g_jP_{n,1,m}P_{0,1,m}^{-1} - B_{0,j,m}g_j \right\| \le N_1 r(m) + N_2\varepsilon(m,n).$$

Proof To simplify notation, we drop the index m for the time being, and write, for example, $D_{n,j;i}$ for $D_{n,j,m,i}$. We have

$$\begin{aligned}
&\left\|P_{0,1}P_{n,1}^{-1}B_{n,j}g_jP_{n,1}P_{0,1}^{-1}-B_{0,j}g_j\right\| \le \\
&\qquad \left\|P_{0,1}P_{n,1}^{-1}B_{n,j}g_jP_{n,1}P_{0,1}^{-1}-P_{0,j}P_{n,j}^{-1}B_{n,j}g_jP_{n,j}P_{0,j}^{-1}\right\| \\
&\qquad +\left\|P_{0,j}P_{n,j}^{-1}B_{n,j}g_jP_{n,j}P_{0,j}^{-1}g_j^{-1}-P_{0,j}P_{n,j}^{-1}B_{n,j}S_{n,j}^{-1}S_{0,j}\right\|\|g_j\| \\
&\qquad +\left\|P_{0,j}P_{n,j}^{-1}B_{n,j}S_{n,j}^{-1}S_{0,j}-B_{0,j}\right\|\|g_j\|.
\end{aligned} \tag{2}$$

We will find upper bounds for the three terms of the right hand side of the above inequality.

The first term of (2) is bounded above by

$$\begin{aligned}
&\left\|P_{0,1}P_{n,1}^{-1}-P_{0,j}P_{n,j}^{-1}\right\|\left\|B_{n,j}g_jP_{n,1}P_{0,1}^{-1}\right\| \\
&\qquad +\left\|P_{0,j}P_{n,j}^{-1}B_{n,j}g_j\right\|\left\|P_{n,j}P_{0,j}^{-1}-P_{n,j}P_{0,j}^{-1}\right\|.
\end{aligned}$$

By Lemma 2.2, the factors containing g_j are bounded above by M_1. We consider the other factor in each term. Recall that $P_{n,j}=A(\phi_*(\gamma'_{n,j;1}),a'_{n,j})$. We have

$$\begin{aligned}
&\left\|P_{0,j}P_{n,j}^{-1}-P_{0,1}P_{n,1}^{-1}\right\| \le \\
&\qquad \|P_{0,j}\|\left\|P_{n,j}^{-1}-A(\phi_*(\gamma'_{0,j;1}),-a'_{n,j})\right\| \\
&\qquad +\left\|A(\phi_*(\gamma'_{0,j;1}),a'_{0,j}-a'_{n,j})-A(\phi_*(\gamma'_{0,1;1}),a'_{0,1}-a'_{n,1})\right\| \\
&\qquad +\|P_{0,1}\|\left\|A(\phi_*(\gamma'_{0,1;1}),-a'_{n,1})-P_{n,1}^{-1}\right\|.
\end{aligned} \tag{3}$$

By Lemma 2.5, there is a positive constant M' such that the first and the third term of the right hand side of (3) are bounded by $M_0M_1M'r(m)$. To find a bound for the second term we consider two cases.

(1) The segment $[x_0,x_{j;1}]$ intersects the same geodesics in $\operatorname{supp}(\chi\mu_n)$ as does the segment $[x_0,x_{1;1}]$.

(2) The two segments intersect different sets of geodesics in $\operatorname{supp}(\chi\mu_n)$.

Let $z_{n,i}=\int_{[x_0,x_{i;1}]}\chi(\mu_0-\mu_n)=a'_{0,i}-a'_{n,i}$.

In case (1), $z_{n,j}=z_{n,1}$, and the geodesics $\gamma'_{0,j;1},\gamma'_{0,1;1}$ lie in a (2–dimensional) solid cylinder of radius d/m based at x_0. The segments $[x_0,x_{j;1}]$ and $[x_0,x_{1;1}]$

induce concurrent orientations on the geodesics $\gamma'_{0,j;1}$ and $\gamma'_{0,1;1}$ respectively. So, by Lemma 2.5,

$$\left\|A(\phi_*(\gamma'_{0,j;1}), z_{n,j}) - A(\phi_*(\gamma'_{0,1;1}), z_{n,1}\right\| \leq M_0 M' r(m).$$

Note that if μ_n satisfies the conditions of case (1) for large enough n, then μ_0 also satisfies these conditions.

In case (2), the orientations induced by the segments $[x_0, x_{j;1}]$ and $[x_0, x_{1,1}]$ on the geodesics $\gamma'_{0,j;1}$ and $\gamma'_{0,1;1}$ respectively, are not concurrent. Hence, by Lemma 2.5,

$$\left\|A(\phi_*(\gamma'_{0,j;1}), z_{n,j}) - A(\phi_*(\gamma'_{0,1;1}), z_{n,1})\right\| \leq M_0 M' r(m) + M''|z_{n,j} + z_{n,1}|.$$

Note that, in this case,

$$a'_{0,j} + a'_{0,1} = \int_{[x_0, x_{j;1}]} \chi\mu_0 + \int_{[x_0, x_{1;1}]} \chi\mu_0 = \chi\mu_0(G)$$

and similarly for μ_n. Hence $z_{n,j} + z_{n,1} = \chi\mu_0(G) - \chi\mu_n(G)$. Let

$$\varepsilon_0(m,n) = \sup_{s \geq n} |\chi_m \mu_0(G) - \chi_m \mu_s(G)|.$$

Now we turn our attention to the second term of equation (2). This term involves only the generator g_j, so we drop the subscript j from the notation. We have

$$\begin{aligned}\left\|P_0 P_n^{-1} B_n g P_n P_0^{-1} g^{-1} - P_0 P_n^{-1} B_n S_n^{-1} S_0\right\| \leq \\ \left\|P_0 P_n^{-1} B_n\right\| \left\|S_n^{-1}\right\| \left\|S_n g P_n^{-1} g^{-1} - S_0 g P_0 g^{-1}\right\| \left\|g P_0^{-1} g^{-1}\right\|.\end{aligned}$$

We consider the term $S_n g P_n^{-1} g^{-1}$, which is equal to

$$A\left(\phi_*(\gamma'_{n;m-1}), \int_{[x_{;m-1}, x_{;m}]} (\chi \circ g^{-1})\mu_n\right) A\left(\phi_*(g(\gamma'_{n;1}), \int_{[x_0, x_{;1}]} \chi\mu_n\right).$$

Since μ_n is invariant by g, and $x_{;m} = g(x_0)$, we have

$$\int_{[x_{;m}, g(x_{;1})]} (\chi \circ g^{-1})\mu_n = \int_{[x_0, x_{;1}]} \chi\mu_n.$$

We have to consider two cases:

(1) The segments $[x_{;m-1}, x_{;m}]$ and $[x_{;m}, g(x_{;1})]$ intersect the same geodesics in $\operatorname{supp}((\chi \circ g^{-1})\mu_n)$.

(2) The segments $[x_{;m-1}, x_{;m}]$ and $[x_{;m}, g(x_{;1})]$ intersect different sets of geodesics in $\operatorname{supp}((\chi \circ g^{-1})\mu_n)$.

In case (1), we let $z_n = \int_{[x_{;m-1},x_{;m}]}(\chi \circ g^{-1})\mu_n = \int_{[x_{;m},g(x_{;1})]}(\chi \circ g^{-1})\mu_n$. The geodesics $\gamma'_{n;m-1}$ and $g(\gamma'_{n;1})$ lie in a solid cylinder of radius d/m, based at $x_{;m}$, and the orientations induced by the segments $[x_{;m-1}, x_{;m}]$ and $[x_{;m}, g(x_{;1})]$ are not concurrent. Hence, by Lemma 2.6, $\left\|S_n g P_n g^{-1} - I\right\| \le M_0 M_2 r(m)$. As before, if μ_n satisfies the conditions of case (1) for large enough n, then μ_0 also satisfies these conditions. Hence

$$\left\|S_n g P_n g^{-1} - S_0 g P_0 g^{-1}\right\| \le 2M_0 M_2 r(m).$$

In case (2), since μ_n is invariant by g, and $x_{;m} = g(x_0)$, we have

$$\int_{[x_{;m},g(x_{;1})]}(\chi \circ g^{-1})\mu_n + \int_{[x_{;m-1},x_{;m}]}(\chi \circ g^{-1})\mu_n = \chi\mu_n(G)$$

and if n is large enough, the same is true of μ_0. Then

$$\begin{aligned}\left\|S_n g P_n g^{-1} - S_0 g P_0 g^{-1}\right\| \le &\\ &\left\|S_n g P_n g^{-1} - A(\phi_*(\gamma'_{n;m-1}), \chi\mu_n(G))\right\| \\ &+ \left\|A(\phi_*(\gamma'_{n;m-1}), \chi\mu_n(G)) - A(\phi_*(\gamma'_{0;m-1}), \chi\mu_0(G))\right\| \\ &+ \left\|A(\phi_*(\gamma'_{0;m-1}), \chi\mu_0(G)) - S_0 g P_0 g^{-1}\right\|.\end{aligned}$$

By Lemma 2.5 and Lemma 2.6, this is bounded above by $M'r(m) + M''\varepsilon(m,n)$.

The third term of equation (2) is bounded by

$$\|P_0\| \left\|P_n^{-1} B_n S_n^{-1} - P_0^{-1} B_0 S_0^{-1}\right\| \|S_0\| \|g\|.$$

But

$$\begin{aligned}&\left\|P_n^{-1} B_n S_n^{-1} - P_0^{-1} B_0 S_0^{-1}\right\| = \\ &\qquad \left\|Q_n D^r_{n;1} D_{n;2} \cdots D_{n;m-2} D^l_{n;m-1} R_n - Q_0 D^r_{0;1} D_{0;2} \cdots D_{0;m-2} D^l_{0;m-1} R_0\right\|\end{aligned}$$

and by Lemma 2.2, this is bounded by

$$\begin{aligned} M_1^2 \Big(&\left\|D^l_{n;m-1} R_n - D^l_{0;m-1} R_0\right\| + \sum_{i=2}^{m-2} \|D_{n,i} - D_{0,i}\| + \\ &+ \left\|Q_n D^r_{n;1} - Q_0 D^r_{0;1}\right\| \Big). \end{aligned} \tag{4}$$

Note that $Q_n D^r_{n;1} = A\left(\phi_*(\gamma'_{n;1}), \int_{[x_0,x_{;1}]} \lambda_{;1}(1-\chi)\mu_n\right)$ and hence

$$\left\|Q_n D^r_{n;1} - Q_0 D^r_{0;1}\right\| \le M'r(m) + M''\varepsilon_1(m,n)$$

where $\varepsilon_1(m,n) = \sup_{s \ge n} \left|\int_{[x_0,x_{;1}]} \lambda_{;1}(1-\chi_m)(\mu_s - \mu_0)\right|$, and similarly for the other terms of (4), for suitable ε_i, $i = 2, \ldots, m-1$.

To complete the proof of Lemma 3.1 we must show that $r(m)$ and $\varepsilon(m,n) = \sum_{i=0}^{m-1} \varepsilon_i(m,n)$ have the required properties. It is clear that we can choose a sequence $r(m)$, with $\lim_{m\to\infty} r(m) = 0$, such that the pair $r = r(m)$, $\delta = d/m$ satisfy the conditions of Lemma 2.10. Lemma 2.14 implies that, for each m, $\lim_{n\to\infty} \varepsilon(m,n) = 0$. □

We let $E_{n,j,m} = C_{n,j,m,1} \cdots C_{n,j,m,m}$ and $H_{n,m} = P_{0,1,m} P_{n,1,m}^{-1}$. Then, combining the above result with (1), we have

$$\left\| H_{n,m} E_{n,j,m} g_j H_{n,m}^{-1} - E_{0,j,m} g_j \right\| \le M(r(m) + \varepsilon(m,n). \tag{5}$$

If $g_1, \ldots, g_k$ is a set of generators for Γ_0, the space R of homomorphisms $\rho\colon \Gamma_0 \to PSL(2,\mathbb{C})$ with quasi-Fuchsian image is a subspace of $PSL(2,\mathbb{C})^k$, and $Q(S)$ is a subspace of the quotient by the adjoint action on the left, $PSL(2,\mathbb{C})^k \big/ PSL(2,\mathbb{C})$. Let

$$\rho_{n,m} = \left(H_{n,m} E_{n,j,m} g_j H_{n,m}^{-1}, \quad j = 1, \ldots, k \right)$$

$$\rho_{n,m} = (E_{0,j,m} g_j, \quad j = 1, \ldots, k)$$

and let $[\rho_{n,m}]$ denote the equivalence class of $\rho_{n,m}$ in $PSL(2,\mathbb{C})^k \big/ PSL(2,\mathbb{C})$.

Let $n(m)$ be a sequence such that $n(m) \ge m$ and $\varepsilon(n(m), m) \le 1/m$. Then $\lim_{m\to\infty} \rho_{n(m),m} = \rho_{\mu_0}$. As $m \to \infty$, $[\rho_{n,m}]$ converge, uniformly in n, to the bending deformation $[\rho_{\mu_n}]$, [5]. Hence, $\lim_{m\to\infty} [\rho_{n(m),m}] = \lim_{m\to\infty} [\rho_{\mu_{n(m)}}] = \lim_{n\to\infty} [\rho_{\mu_n}]$, and we have

$$\lim_{n\to\infty} [\rho_{\mu_n}] = [\rho_{\mu_0}]. \tag{6}$$

To complete the proof of Theorem 1, it remains to show that the convergence is uniform in compact subsets of $\mathcal{D}$. If $([\rho], t) \in \mathcal{D}$, each bound used in the proof of (6) depends at most linearly on t, while it depends on ρ only in terms of the endpoints of a finite number of geodesics $\phi_*(\gamma)$. The endpoints of the geodesic $\phi_*(\gamma)$ are, for each γ, holomorphic functions of $[\rho]$. Hence each bound can be chosen uniformly on each compact subset of $\mathcal{D}$.

Note that $\mathcal{D}$ contains in its interior the set $Q(S) \times \{0\}$. If the laminations μ_n are real for all but a finite number of n, then $\mathcal{D}$ also contains the set $Q(S) \times \mathbf{R}$, but this is not true in the general case.

To prove Theorem 2 we recall that the bending vector field T_μ is defined by

$$T_\mu([\rho]) = \frac{\partial}{\partial t} B_\mu([\rho], t).$$

The vector fields T_{μ_n} are holomorphic, and $B_{\mu_n}([\rho],t)$ converge to $B_{\mu_0}([\rho],t)$ for $([\rho],t) \in \mathcal{D}$. It follows that T_{μ_n} converge to T_{μ_0}, uniformly on compact subsets of $Q(S)$.

We conclude with the proof of Theorem 3. We consider the subset of $\mathcal{ML}(S)$ consisting of measured laminations with non negative real and imaginary parts, and we denote it by $\mathcal{ML}^{++}(S)$. We identify $\mathcal{ML}^{++}(S)$ with a subset of the set of pairs of positive measured laminations $\mathcal{ML}^+_{\mathbb{R}}(S) \times \mathcal{ML}^+_{\mathbb{R}}(S)$. If $\nu \in \mathcal{ML}^{++}(S)$, then $\operatorname{Re}\nu$ and $\operatorname{Im}\nu$ are in $\mathcal{ML}^+_{\mathbb{R}}(S)$ and they satisfy the condition

$$\operatorname{supp}(\operatorname{Re}\nu) \cup \operatorname{supp}(\operatorname{Im}\nu) \text{ is a geodesic lamination.} \tag{7}$$

Conversely, any pair ν_1, ν_2 of positive measured laminations satisfying (7) define a measure $\nu = \nu_1 + i\nu_2 \in \mathcal{ML}^{++}(S)$. The mapping is a homeomorphism of $\mathcal{ML}^{++}(S)$ onto a subset of $\mathcal{ML}^+_{\mathbb{R}}(S) \times \mathcal{ML}^+_{\mathbb{R}}(S)$. But $\mathcal{ML}^+_{\mathbb{R}}(S)$ is homeomorphic to $\mathbb{R}^{6g-6}$, [6]. Thus $\mathcal{ML}^{++}(S)$ is first countable, and Theorem 2 implies that $\mu \mapsto T_\mu$ is continuous. Theorem 3 then follows by the continuity of the evaluation map.

References

[1] **R D Canary**, **D B A Epstein**, **P Green**, *Notes on notes of Thurston*, from: "Analytical and Geometric Aspects of Hyperbolic Space", D B A Epstein (editor), LMS Lecture Note Series 111, Cambridge University Press (1987) 3–92

[2] **D B A Epstein**, **A Marden**, *Convex Hulls in Hyperbolic Space, a Theorem of Sullivan, and Measured Pleated Surfaces*, from: "Analytical and Geometric Aspects of Hyperbolic Space", D B A Epstein (editor), LMS Lecture Note Series 111, Cambridge University Press (1987) 113–253

[3] **D Johnson**, **J J Millson**, *Deformation spaces associated to compact hyperbolic manifolds*, from: "Discrete Groups in Geometry and Analysis", Progress in Math. 67, Birkhaüser (1987) 48–106

[4] **C Kourouniotis**, *Deformations of hyperbolic structures*, Math. Proc. Camb. Phil. Soc. 98 (1985) 247–261

[5] **C Kourouniotis**, *Bending in the space of quasi-Fuchsian structures*, Glasgow Math. J. 33 (1991) 41–49

[6] **R C Penner**, **J L Harer** *Combinatorics of train tracks*, Annals of Math. Studies 125, Princeton University Press (1992)

Department of Mathematics
University of Crete
Iraklio, Crete, Greece

Email: `chrisk@math.uch.gr`

Received: 15 November 1997

ISSN 1464-8997 **335**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 335–340

Complex projective structures on Kleinian groups

Albert Marden

Abstract Let M^3 be a compact, oriented, irreducible, and boundary incompressible 3–manifold. Assume that its fundamental group is without rank two abelian subgroups and $\partial M^3 \neq \emptyset$. We will show that every homomorphism $\theta\colon \pi_1(M^3) \to PSL(2, \mathbf{C})$ which is not "boundary elementary" is induced by a possibly branched complex projective structure on the boundary of a hyperbolic manifold homeomorphic to M^3.

AMS Classification 30F50; 30F45, 30F60, 30F99, 30C99

Keywords Projective structures on Riemann surfaces, hyperbolic 3–manifolds

1 Introduction

Let M^3 be a compact, oriented, irreducible, and boundary incompressible 3–manifold such that its fundamental group $\pi_1(M^3)$ is without rank two abelian subgroups. Assume that $\partial M^3 = R_1 \cup \ldots \cup R_n$ has $n \geq 1$ components, each a surface necessarily of genus exceeding one.

We will study homomorphisms

$$\theta\colon \pi_1(M^3) \to G \subset PSL(2, \mathbf{C})$$

onto groups G of Möbius transformations. Such a homomorphism is called *elementary* if its image G fixes a point or pair of points in its action on $\mathbf{H}^3 \cup \partial \mathbf{H}^3$, ie on hyperbolic 3–space and its "sphere at infinity". More particularly, the homomorphism θ is called *boundary elementary* if the image $\theta(\pi_1(R_k))$ of some boundary subgroup is an elementary group. (This definition is independent of how the inclusion $\pi_1(R_k) \hookrightarrow \pi_1(M^3)$ is taken as the images of different inclusions of the same boundary group are conjugate in G).

The purpose of this note is to prove:

Theorem 1 *Every homomorphism $\theta\colon \pi_1(M^3) \to PSL(2, \mathbf{C})$ which is not boundary elementary is induced by a possibly branched complex projective structure on the boundary of some Kleinian manifold $\mathbf{H}^3 \cup \Omega(\Gamma)/\Gamma \cong M^3$.*

This result is based on, and generalizes:

Theorem A (Gallo–Kapovich–Marden [1]) *Let R be a compact, oriented surface of genus exceeding one. Every homomorphism $\pi_1(R) \to PSL(2, \mathbf{C})$ which is not elementary is induced by a possibly branched complex projective structure on $\mathbf{H}^2/\Gamma \cong R$ for some Fuchsian group Γ.*

Theorem 1 is related to Theorem A as simultaneous uniformization is related to uniformization. Its application to quasifuchsian manifolds could be called simultaneous projectivization. For Theorem A finds a single surface on which the structure is determined whereas Theorem 1 finds a structure simultaneously on the pair of surfaces arising from some quasifuchsian group.

2 Kleinian groups

Thurston's hyperbolization theorem [3] implies that M^3 has a hyperbolic structure: there is a Kleinian group $\Gamma_0 \cong \pi_1(M^3)$ with regular set $\Omega(\Gamma_0) \subset \partial\mathbf{H}^3$ such that $\mathcal{M}(\Gamma_0) = \mathbf{H}^3 \cup \Omega(\Gamma_0)/\Gamma_0$ is homeomorphic to M^3. The group Γ_0 is not uniquely determined by M^3, rather M^3 determines the deformation space $\mathcal{D}(\Gamma_0)$ (taking a fixed Γ_0 as its origin).

We define $\mathcal{D}^*(\Gamma_0)$ as the set of those isomorphisms $\phi\colon \Gamma_0 \to \Gamma \subset PSL(2, \mathbf{C})$ onto Kleinian groups Γ which are induced by orientation preserving homeomorphisms $\mathcal{M}(\Gamma_0) \to \mathcal{M}(\Gamma)$. Then $\mathcal{D}(\Gamma_0)$ is defined as $\mathcal{D}^*(\Gamma_0)/PSL(2, \mathbf{C})$, since we do not distinguish between elements of a conjugacy class.

Let $\mathcal{V}(\Gamma_0)$ denote the representation space $\mathcal{V}^*(\Gamma_0)/PSL(2, \mathbf{C})$ where $\mathcal{V}^*(\Gamma_0)$ is the space of boundary nonelementary homomorphisms $\theta\colon \Gamma_0 \to PSL(2, \mathbf{C})$.

By Marden [2], $\mathcal{D}(\Gamma_0)$ is a complex manifold of dimension $\sum[3(\text{genus } R_k) - 3]$ and an open subset of the representation variety $\mathcal{V}(\Gamma_0)$. If M^3 is acylindrical, $\mathcal{D}(\Gamma_0)$ is relatively compact in $\mathcal{V}(\Gamma_0)$ (Thurston [4]).

The fact that $\mathcal{D}(\Gamma_0)$ is a manifold depends on a uniqueness theorem (Marden [2]). Namely two isomorphisms $\phi_i\colon \Gamma_0 \to \Gamma_i$, $i = 1, 2$, are conjugate if and only if $\phi_2\phi_1^{-1}\colon \Gamma_1 \to \Gamma_2$ is induced by a homeomorphism $\mathcal{M}(\Gamma_1) \to \mathcal{M}(\Gamma_2)$ which is homotopic to a conformal map.

3 Complex projective structures

For the purposes of this note we will use the following definition (cf [1]). A *complex projective structure* for the Kleinian group Γ is a locally univalent meromorphic function f on $\Omega(\Gamma)$ with the property that

$$f(\gamma z) = \theta(\gamma) f(z),\ z \in \Omega(\Gamma),\ \gamma \in \Gamma,$$

for some homomorphism $\theta\colon \Gamma \to PSL(2, \mathbf{C})$. We are free to replace f by a conjugate AfA^{-1}, for example to normalize f on one component of $\Omega(\Gamma)$.

Such a function f solves a Schwarzian equation

$$S_f(z) = q(z),\quad q(\gamma z)\gamma'(z)^2 = q(z);\ \gamma \in \Gamma,\ z \in \Omega(\Gamma),$$

where $q(z)$ is the lift to $\Omega(\Gamma)$ of a holomorphic quadratic differential defined on each component of $\partial\mathcal{M}(\Gamma)$. Conversely, solutions of the Schwarzian,

$$S_g(z) = q(z),\ z \in \Omega(\Gamma),$$

are determined on each component of $\Omega(\Gamma)$ only up to post composition by any Möbius transformation. The function f has the property that it not only is a solution on each component, but that its restrictions to the various components fit together to determine a homomorphism $\Gamma \to PSL(2, \mathbf{C})$. Automatically (cf [1]), the homomorphism θ induced by f is boundary nonelementary.

When *branched* complex projective structures for a Kleinian group are required, it suffices to work with the simplest ones: $f(z)$ is meromorphic on $\Omega(\Gamma)$, induces a homomorphism $\theta\colon \Gamma \to PSL(2, \mathbf{C})$ (which is automatically boundary nonelementary), and is locally univalent except at most for one point, modulo $\mathrm{Stab}(\Omega_0)$, on each component Ω_0 of $\Omega(\Gamma)$. At an exceptional point, say $z = 0$,

$$f(z) = \alpha z^2(1 + o(z)),\ \alpha \neq 0.$$

Such f are characterized by Schwarzians with local behavior

$$S_f(z) = q(z) = -3/2z^2 + b/z + \Sigma a_i z^i, \qquad b^2 + 2a_0 = 0.$$

At any designated point on a component R_k of $\partial\mathcal{M}(\Gamma)$, there is a quadratic differential with leading term $-3/2z^2$. To be admissible, a differential must be the sum of this and any element of the $(3g_k - 2)$–dimensional space of quadratic differentials with at most a simple pole at the designated point. In addition it must satisfy the relation $b^2 + 2a_0 = 0$. That is, the admissible differentials are parametrized by an algebraic variety of dimension $3g_k - 3$. For details, see [1].

If a branch point needs to be introduced on a component R_k of $\partial\mathcal{M}(\Gamma)$, it is done during a construction. According to [1], a branch point needs to be introduced if and only if the restriction

$$\theta\colon \pi_1(R_k) \to PSL(2, \mathbf{C})$$

does *not* lift to a homomorphism

$$\theta^*\colon \pi_1(R_k) \to SL(2,\mathbf{C}).$$

4 Dimension count

The vector bundle of holomorphic quadratic differentials over the Teichmüller space of the component R_k of $\partial\mathcal{M}(\Gamma_0)$ has dimension $6g_k - 6$. All together these form the vector bundle $\mathcal{Q}(\Gamma_0)$ of quadratic differentials over the Kleinian deformation space $\mathcal{D}(\Gamma_0)$. That is, $\mathcal{Q}(\Gamma_0)$ has *twice* the dimension of $\mathcal{V}(\Gamma_0)$. The count remains the same if there is a branching at a designated point.

For example, if Γ_0 is a quasifuchsian group of genus g, $\mathcal{Q}(\Gamma_0)$ has dimension $12g-12$ whereas $\mathcal{V}(\Gamma_0)$ has dimension $6g-6$. Corresponding to each nonelementary homomorphism $\theta\colon \Gamma_0 \to PSL(2,\mathbf{C})$ that lifts to $SL(2,\mathbf{C})$ is a group Γ in $\mathcal{D}(\Gamma_0)$ and a quadratic differential on the designated component of $\Omega(\Gamma)$. This in turn determines a differential on the other component. There is a solution of the associated Schwarzian equation $S_g(z) = q(z)$ satisfying

$$f(\gamma z) = \theta(\gamma) f(z),\ z \in \Omega(\Gamma),\ \gamma \in \Gamma.$$

Theorem 1 implies that $\mathcal{V}(\Gamma_0)$ has at most 2^n components. For this is the number of combinations of $(+,-)$ that can be assigned to the n–components of $\partial\mathcal{M}(\Gamma_0)$ representing whether or not a given homomorphism lifts. For a quasifuchsian group Γ_0, $\mathcal{V}(\Gamma_0)$ has two components (see [1]).

5 Proof of Theorem 1

We will describe how the construction introduced in [1] also serves in the more general setting here.

By hypothesis, each component Ω_k of $\Omega(\Gamma_0)$ is simply connected and covers a component R_k of $\partial\mathcal{M}(\Gamma_0)$. In addition, the restriction

$$\theta\colon \pi_1(R_k) \cong Stab(\Omega_k) \to G_k \subset PSL(2,\mathbf{C})$$

is a homomorphism to the nonelementary group G_k.

The construction of [1] yields a simply connected Riemann surface $\mathcal{J}_k$ lying over S^2, called a pants configuration, such that:

(i) There is a conformal group Γ_k acting freely in $\mathcal{J}_k$ such that $\mathcal{J}_k/\Gamma_k$ is homeomorphic to R_k.

(ii) The holomorphic projection $\pi\colon \mathcal{J}_k \to S^2$ is locally univalent if θ lifts to a homomorphism $\theta^*\colon \pi_1(R_k) \to SL(2, \mathbf{C})$. Otherwise π is locally univalent except for one branch point of order two, modulo Γ_k.

(iii) There is a quasiconformal map $h_k\colon \Omega_k \to \mathcal{J}_k$ such that

$$\pi h_k(\gamma_z) = \theta(\gamma)\pi h_k(z), \ \gamma \in Stab(\Omega_k), \ z \in \Omega_k.$$

Once h_k is determined for a representative Ω_k for each component R_k of $\partial\mathcal{M}(\Gamma_0)$, we bring in the action of Γ_0 on the components of $\Omega(\Gamma_0)$ and the corresponding action of $\theta(\Gamma_0)$ on the range. By means of this action a quasiconformal map h is determined on all $\Omega(\Gamma_0)$ which satisfies

$$\pi h(\gamma z) = \theta(\gamma)\pi h(z), \ \gamma \in \Gamma_0, \ z \in \Omega(\Gamma_0).$$

The Beltrami differential $\mu(z) = (\pi h)_{\bar{z}}/(\pi h)_z$ satisfies

$$\mu(\gamma z)\bar{\gamma}'(z)/\gamma'(z) = \mu(z), \ \gamma \in \Gamma_0, \ z \in \Omega(\Gamma_0).$$

It may equally be regarded as a form on $\partial\mathcal{M}(\Gamma_0)$. Using the fact that the limit set of Γ_0 has zero area, we can solve the Beltrami equation $g_{\bar{z}} = \mu g_z$ on S^2. It has a solution which is a quasiconformal mapping g and is uniquely determined up to post composition with a Möbius transformation. Furthermore g uniquely determines, up to conjugacy, an isomorphism $\varphi\colon \Gamma_0 \to \Gamma$ to a group Γ in $\mathcal{D}(\Gamma_0)$.

The composition $\pi h g^{-1}$ is a meromorphic function on each component of $\Omega(\Gamma)$. It satisfies

$$(\pi h g^{-1})(\gamma z) = \theta\varphi^{-1}(\gamma)\pi h g^{-1}(z), \quad \gamma \in \Gamma, \ z \in \Omega(\Gamma).$$

The composition is locally univalent except for at most one point on each component of $\Omega(\Gamma)$, modulo its stabilizer in Γ. That is, $\pi \circ h \circ g^{-1}$ is a complex projective structure on Γ that induces the given homomorphism θ, via the identification φ.

6 Open questions

Presumably, a nonelementary homomorphism $\theta\colon \Gamma_0 \to PSL(2, \mathbf{C})$ can be elementary for one, or all, of the $n \geq 1$ components of $\partial\mathcal{M}(\Gamma_0)$. Presumably too, the restrictions to $\partial\mathcal{M}(\Gamma_0)$ of a boundary nonelementary homomorphism can lift to a homomorphism into $SL(2, \mathbf{C})$ without the homomorphism $\Gamma_0 \to PSL(2, \mathbf{C})$ itself lifting. However we have no examples of these phenomena.

According to Theorem 1, there is a subset $\mathcal{P}(\Gamma_0)$ of the vector bundle $\mathcal{Q}(\Gamma_0)$ consisting of those homomorphic differentials giving rise to, say, unbranched complex projective structures on the groups in $\mathcal{D}(\Gamma_0)$. What is the analytic structure of $\mathcal{P}(\Gamma_0)$; is it a nonsingular, properly embedded, analytic subvariety?

When does a given Schwarzian equation $S_f(z) = q(z)$ on $\Omega(\Gamma)$ have a solution which induces a homomorphism of Γ?

References

[1] **D Gallo**, **M Kapovich**, **A Marden**, *The monodromy groups of Schwarzian equations on compact Riemann surfaces*, preprint (1997 revised)

[2] **A Marden**, *The geometry of finitely generated Kleinian groups*, Annals of Math. 99 (1974) 383–462

[3] **W P Thurston**, *Three dimensional manifolds, Kleinian groups, and hyperbolic geometry*, Bull. Amer. Math. Soc. 6 (1982) 357–381

[4] **W P Thurston**, *Hyperbolic structures on 3–manifolds I*, Annals of Math. 124 (1986) 203–246

School of Mathematics, University of Minnesota
Minneapolis, MN 55455, USA

Email: am@geom.unm.edu

Received: 1 June 1998

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 341–364

Coarse extrinsic geometry: a survey

Mahan Mitra

Abstract This paper is a survey of some of the developments in coarse extrinsic geometry since its inception in the work of Gromov. Distortion, as measured by comparing the diameter of balls relative to different metrics, can be regarded as one of the simplist extrinsic notions. Results and examples concerning distorted subgroups, especially in the context of hyperbolic groups and symmetric spaces, are exposed. Other topics considered are quasiconvexity of subgroups; behaviour at infinity, or more precisely continuous extensions of embedding maps to Gromov boundaries in the context of hyperbolic groups acting by isometries on hyperbolic metric spaces; and distortion as measured using various other filling invariants.

AMS Classification 20F32; 57M50

Keywords Coarse geometry, quasi-isometry, hyperbolic groupsx

To David Epstein on his sixtieth birthday

1 Introduction

Extrinsic geometry deals with the study of the geometry of subspaces relative to that of an ambient space. Given a Riemannian manifold M and a submanifold N, classical (differential) extrinsic geometry studies infinitesimal changes in the Riemannian metric on N induced from M. This involves an analysis of the second fundamental form or shape operator [35]. In coarse geometry local or infinitesimal machinery is absent. Thus it does not make sense to speak of tangent spaces or Riemannian metrics. However, the large scale notion of metric continues to make sense. Given a metric space X and a subspace Y one can still compare the intrinsic metric on Y to the metric inherited from X. This is especially useful for finitely generated subgroups of finitely generated groups. To formalize this, Gromov introduced the notion of distortion in his seminal paper [33].

Definition ([33],[22]) If $i\colon \Gamma_H \to \Gamma_G$ is an embedding of the Cayley graph of H into that of G, then the *distortion* function is given by

$$disto(R) = Diam_{\Gamma_H}(\Gamma_H \cap B(R)),$$

where $B(R)$ is the ball of radius R around $1 \in \Gamma_G$.

The definition above differs from the one in [33] by a linear factor and coincides with that in [22].

Note The above definition continues to make sense when Γ_G and Γ_H are replaced by graphs or (more generally) path-metric spaces (see below for definition) X and Y respectively.

Definition A *path-metric space* is a metric space (X,d) such that for all $x,y \in X$ there exists an isometric embedding $f\colon [0,d(x,y)] \to X$ with $f(0) = x$ and $f(d(x,y)) = y$.

If the distortion function is linear we say Γ_H (or Y) is *quasi-isometrically* (often abbreviated to *qi*) embedded in Γ_G (or X). This is equivalent to the following:

Definition A map f from one metric space (Y,d_Y) into another metric space (Z,d_Z) is said to be a (K,ϵ)*–quasi-isometric embedding* if

$$\frac{1}{K}(d_Y(y_1,y_2)) - \epsilon \leq d_Z(f(y_1),f(y_2)) \leq K d_Y(y_1,y_2) + \epsilon.$$

If f is a quasi-isometric embedding, and every point of Z lies at a uniformly bounded distance from some $f(y)$ then f is said to be a *quasi-isometry.* A (K,ϵ)–quasi-isometric embedding that is a quasi-isometry will be called a (K,ϵ)–quasi-isometry.

We collect here a few other closely related notions:

Definition A subset Z of X is said to be k*–quasiconvex* if any geodesic joining $a,b \in Z$ lies in a k–neighborhood of Z. A subset Z is *quasiconvex* if it is k–quasiconvex for some k.

A (K,ϵ)*–quasigeodesic* is a (K,ϵ)–quasi-isometric embedding of a closed interval in $\mathbb{R}$. A $(K,0)$–quasigeodesic will also be called a K–quasigeodesic.

For hyperbolic metric spaces (in the sense of Gromov [34]) the notions of quasiconvexity and qi embeddings coincide. This is because quasigeodesics lie close to geodesics in hyperbolic metric spaces [3], [31], [21].

Distortion can be regarded, in some sense, as the simplest extrinsic notion in coarse geometry. However a complete understanding of distortion is lacking even in special situations like subgroups of hyperbolic groups or discrete (infinite co-volume) subgroups of higher rank semi-simple Lie groups. One of the aims of this survey is to expose some of the issues involved. This is done in Section 2.

A characterisation of quasi-isometric embeddings in terms of group theory is another topic of extrinsic geometry that has received some attention of late. This will be dealt with in Section 3.

A different perspective of coarse extrinsic geometry comes from the asymptotic point of view. The issue here is behavior 'at infinity'. From this perspective it seems possible to introduce and study finer invariants involving distortion along specified directions. Section 4 deals with this in the special context of hyperbolic subgroups of hyperbolic groups.

Finally in Section 5, we discuss some other invariants of extrinsic geometry that have come up in different contexts.

It goes without saying that this survey reflects the author's bias and is far from comprehensive.

Acknowledgements I would like to thank the organizers of David Epstein's birthday fest where the idea of this survey first came up. I would also like to take this opportunity to thank Benson Farb for several inspiring discussions on coarse geometry over the years.

2 Distortion

If a finitely generated subgroup H of a finitely generated group G is qi–embedded we shall refer to it as undistorted. Otherwise H will be said to be distorted. We shall also have occasion to replace the Cayley graph of G by a symmetric space (equipped with its invariant metric) or more generally a path metric space (X, d). In the latter case, distortion will be measured with respect to the metric d on X.

Distorted subgroups of hyperbolic groups or symmetric spaces are somewhat difficult to come by. This has resulted in a limited supply of examples. Brief accounts will be given of some of the known sources of examples.

An aspect that will not be treated in any detail is the connection to algorithmic problems, especially the Magnus problem. See [33] or (for a more detailed account) [22] for a treatment.

Subgroups of hyperbolic groups and $SL_2(\mathbb{C})$

One of the earliest classes of examples of distorted hyperbolic subgroups of hyperbolic groups came from Thurston's work on 3–manifolds fibering over the circle [62]. Let M be a closed hyperbolic 3–manifold fibering over the circle with fiber F. Then $\pi_1(F)$ is a hyperbolic subgroup of the hyperbolic group $\pi_1(M)$. The distortion is easily seen to be exponential.

It follows from work of Bonahon [8] and Thurston [61] that if H is a closed surface subgroup of the fundamental group $\pi_1(M)$ of a closed hyperbolic 3–manifold M then the distortion of H is either linear or exponential. This continues to be true if H is replaced by any freely indecomposable group. In fact exponential distortion of a freely indecomposable group corresponds precisely (up to passing to a finite cover of M) to the case of a hyperbolic 3–manifold fibering over the circle.

The situation is considerably less clear when we come to freely decomposable subgroups of hyperbolic 3–manifolds. The tameness conjecture (attributed to Marden [40]) asserts that the covering of a closed hyperbolic 3–manifold corresponding to a finitely generated subgroup of its fundamental group is topologically tame, ie is homeomorphic to the interior of a compact 3–manifold with boundary. If this conjecture were true, it would follow (using a Theorem of Canary [19]) that any finitely generated subgroup H of the fundamental group $\pi_1(M) = G$ is either quasiconvex in G or is exponentially distorted. Moreover, exponential distortion corresponds precisely (up to passing to a finite cover of M) to the case of a fiber of a hyperbolic 3–manifold fibering over the circle. Much of this theory can be extended to take parabolics into account.

This class of examples can be generalized in two directions. One can ask for distorted discrete subgroups of $SL_2(\mathbb{C})$ or for distorted hyperbolic subgroups of hyperbolic groups (in the sense of Gromov). We look first at discrete subgroups of $SL_2(\mathbb{C})$. A substantial class of examples comes from geometrically tame groups. In fact the simplest surface group, the fundamental group of a punctured torus (the puncture corresponds to a parabolic element), displays much of the exotic extrinsic geometry that may occur. These examples were studied in great detail by Minsky in [45]. The distortion function was calculated in [49].

Let S be a hyperbolic punctured torus so that the two shortest geodesics a and b are orthogonal and of equal length. Let S_0 denote S minus a neighborhood of the cusp. Let $N_\delta(a)$ and $N_\delta(b)$ be regular collar neighborhoods of a and b in S_0. For $n \in \mathbb{N}$, define $\gamma_n = a$ if n is even and equal to b if n is odd. Let T_n be the open solid torus neighborhood of $\gamma_n \times \{n + \frac{1}{2}\}$ in $S_0 \times [0, \infty)$ given by

$$T_n = N_\delta(\gamma_n)\times(n, n+1)$$

and let $M_0 = (S_0)\times[0,\infty)\backslash\bigcup_{n\in\mathbb{N}} T_n$.

Let $a(n)$ be a sequence of positive integers greater than one. Let $\hat{\gamma_n} = \gamma_n\times\{n\}$ and let μ_n be an oriented meridian for ∂T_n with a single positive intersection with $\hat{\gamma_n}$. Let M denote the result of gluing to each ∂T_n a solid torus $\hat{T_n}$, such that the curve $\hat{\gamma_n}^{a(n)}\mu_n$ is glued to a meridian. Let q_{nm} be the mapping class from S_0 to itself obtained by identifying S_0 to $S_0\times m$, pushing through M to $S_0\times n$ and back to S_0. Then $q_{n(n+1)}$ is given by $\Phi_n = D^{a(n)}_{\gamma_n}$, where D^k_c denotes Dehn twist along c, k times. Matrix representations of Φ_n are given by

$$\Phi_{2n} = \begin{pmatrix} 1 & a(2n) \\ 0 & 1 \end{pmatrix}$$

and

$$\Phi_{2n+1} = \begin{pmatrix} 1 & 0 \\ a(2n+1) & 1 \end{pmatrix}.$$

Recall that the metric on M_0 is the restriction of the product metric. The $\hat{T_n}$'s are given hyperbolic metrics such that their boundaries are uniformly quasi-isometric to $\partial T_n \subset M_0$. Then from [45], M is quasi-isometric to the complement of a rank one cusp in the convex core of a hyperbolic manifold $M_1 = \mathbb{H}^3/\Gamma$. Let σ_n denote the shortest path from $S_0\times 1$ to $S_0\times n$. Let $\overline{\sigma_n}$ denote σ_n with reversed orientation. Then $\tau_n = \sigma_n\gamma_n\overline{\sigma_n}$ is a closed path in M of length $2n+1$. Further τ_n is homotopic to a curve $\rho_n = \Phi_1\cdots\Phi_n(\gamma_n)$ on S_0. Then

$$\Pi_{i=1\cdots n} a(i) \le l(\rho_n) \le \Pi_{i=1\cdots n}(a(i)+2)$$

Hence

$$\Pi_{i=1\cdots n} a(i) \le (2n+1)disto(2n+1) \le \Pi_{i=1\cdots n}(a(i)+2)$$

Since M is quasi-isometric to the complement of the cusp of a hyperbolic manifold [45] and γ_n's lie in a complement of the cusp, the distortion function of Γ is of the same order as the distortion function above. In particular, functions of arbitrarily fast growth may be realized. This answers a question posed by Gromov in [33] page 66.

A closely related class of examples (the so called 'drill–holes' examples of which the punctured torus examples above may be regarded as special cases) appears in work of Thurston [62] and Bonahon and Otal [9].

Let us now turn to finitely generated subgroups of hyperbolic groups. If we restrict ourselves to hyperbolic subgroups there is a considerable paucity of examples. The chief ingredient for constructing distorted hyperbolic subgroups of hyperbolic groups is the celebrated combination theorem of Bestvina and Feighn [4]. This theorem was partly motivated by Thurston's hyperbolization theorem for Haken manifolds [43], [62] and continues to be an inevitable first step in constructing any distorted hyperbolic subgroups. The following Proposition summarizes these examples. The proof follows easily from normal forms.

Proposition 2.1 *Let G be a hyperbolic group acting cocompactly on a simplicial tree T such that all vertex and edge stabilizers are hyperbolic. Also suppose that every inclusion of an edge stabilizer in a vertex stabilizer is a quasi-isometric embedding. Let H be the stabilizer of a vertex or edge of T. Then the distortion of H is linear or exponential.*

Based on Bestvina and Feighn's combination theorem and work of Thurston's on stable and unstable foliations of surfaces [23], Mosher [53] constructed a class of examples of normal surface subgroups of hyperbolic groups where the quotient is free of rank strictly greater than one.

This idea was used by Bestvina, Feighn and Handel in [5] to construct similar examples where the normal subgroup is free.

Thus one has examples of exact sequences

$$1 \to N \to G \to Q \to 1$$

of hyperbolic groups where N is a free group or a surface group. Owing to a general theorem of Mosher's regarding the existence of quasi-isometric sections of Q [54] the distortion of any normal hyperbolic subgroup N of infinite index in a hyperbolic group G is exponential.

Further, it follows from work of Rips and Sela [59], [57] that a torsion free normal hyperbolic subgroup of a hyperbolic group is a free product of free groups and surface groups. However, the only known restriction on Q is that it is hyperbolic [54]. It seems natural to wonder if there exist examples where the exact sequence does not split or at least where Q is not virtually free.

We now describe some examples exhibiting higher distortion [49]. Start with a hyperbolic group G such that $1 \to F \to G \to F \to 1$ is exact, where F is free of rank 3.

Let $F_1 \subset G$ denote the normal subgroup. Let $F_2 \subset G$ denote a section of the quotient group. Let $G_1, \cdots, G_n$ be n distinct copies of G. Let F_{i1} and F_{i2} denote copies of F_1 and F_2 respectively in G_i. Let

$$G = G_1 *_{H_1} G_2 * \cdots *_{H_{n-1}} G_n$$

where each H_i is a free group of rank 3, the image of H_i in G_i is F_{i2} and the image of H_i in G_{i+1} is $F_{(i+1)1}$. Then G is hyperbolic.

Let $H = F_{11} \subset G$. Then the distortion of H is superexponential for $n > 1$. In fact, it can be checked inductively that the distortion function is an iterated exponential of height n.

Starting from Bestvina, Feighn and Handel's examples above, one can construct examples with distortion a tower function. Let a_1, a_2, a_3 be generators of F_1 and b_1, b_2, b_3 be generators of F_2. Then

$$G = \{a_1, a_2, a_3, b_1, b_2, b_3 : b_i^{-1} a_j b_i = w_{ij}\}$$

where w_{ij} are words in a_i's. We add a letter c conjugating a_i's to 'sufficiently random' words in b_j's to get G_1. Thus,

$$G_1 = \{a_1, a_2, a_3, b_1, b_2, b_3, c : b_i^{-1} a_j b_i = w_{ij}, c^{-1} a_i c = v_i\},$$

where v_i's are words in b_j's satisfying a small-cancellation type condition to ensure that G_1 is hyperbolic. See [34], page 151 for details on addition of 'random' relations.

It can be checked that these examples have distortion function greater than any iterated exponential.

The above set of examples were motivated largely by examples of distorted cyclic subgroups in [33], page 67 and [28] (these examples will be discussed later in this paper).

So far, there is no satisfactory way of manufacturing examples of hyperbolic subgroups of hyperbolic groups exhibiting arbitrarily high distortion. It is easy to see that a subgroup of sub-exponential distortion is quasiconvex [33]. Not much else is known. One is thus led to the following question:

Question Given any increasing function $f\colon \mathbb{N} \to \mathbb{N}$ does there exist a hyperbolic subgroup H of a hyperbolic group G such that the distortion of H is of the order of $e^{f(n)}$?

Note that the above question has a positive answer if G is replaced by $SL_2(\mathbb{C})$.

If one does not restrict oneself to hyperbolic subgroups of hyperbolic groups, one has a large source of examples coming from finitely generated subgroups of small cancellation groups. These examples are due to Rips [56].

Let $Q = \{g_1, \cdots, g_n : r_1, \cdots, r_m\}$ be any finitely presented group. Construct a small cancellation ($C'(1/6)$) group G with presentation as follows:

$$G = \{g_1, \cdots, g_n, a_1, a_2 : g_i^{-1} a_j g_i = u_{ij}, g_i a_j g_i^{-1} = v_{ij}, r_k = w_k \text{ for } i = 1 \cdots n,\ j = 1, 2 \text{ and } k = 1 \cdots m.\ \}$$

where u_{ij}, v_{ij}, w_k are words in a_1, a_2 satisfying $C'(1/6)$.

Then one has an exact sequence $1 \to H \to G \to Q \to 1$ where H is the subgroup of G generated by a_1, a_2 and Q is the given finitely presented group. The distortion of H can be made to vary by varying Q (one basically needs to vary the complexity of the word problem in Q). However the subgroups H are generally not finitely presented.

A remarkable example of a finitely presented normal subgroup H of a hyperbolic group G has recently been discovered by Brady [15]. This is the first example of a finitely presented non-hyperbolic subgroup of a hyperbolic group. The distortion in this example is exponential as the quotient group is infinite cyclic.

Distortion in symmetric spaces

Now let G be a semi-simple Lie group. Cyclic discrete subgroups generated by unipotent elements are exponentially distorted. This is because discrete subgroups of the nilpotent subgroup N in a KAN decomposition of G is distorted in this way. This is the most well known source of distortion.

Other known examples seem to have their origin in rank 1 phenomena. Given any Lie group G containing $F_2 \times F_2$ as a discrete subgroup one has distorted subgroups coming from a construction due to Mihailova [44], [33], [22] (see below). In some sense these examples are 'reducible'. Truly higher rank phenomena are hard to come by. One has the following basic question:

Question Are there examples of distorted finitely generated discrete subgroups H of irreducible lattices in higher rank semi-simple Lie groups G such that H has no unipotent element? (See [22] also).

Note that Thurston's construction of normal subgroups cannot possibly go through here on account of the following basic theorem of Kazhdan–Margulis:

Theorem 2.2 [41] *Let Γ be an irreducible lattice in a symmetric space of real rank greater than one. Then any normal subgroup Λ of Γ is either finite or the quotient Γ/Λ is finite.*

Another non-distortion theorem has recently been proven by Lubotzky–Moses–Raghunathan [39] answering a question of Kazhdan:

Theorem 2.3 *Any irreducible lattice in a symmetric space X of rank greater than one is undistorted in X.*

The above theorems indicate the difficulty in obtaining distorted subgroups of higher rank Lie groups.

Similar questions may be asked for rank one symmetric spaces also eg for complex hyperbolic, quaternionic hyperbolic and the Cayley hyperbolic planes. Here, too there is a dearth of examples.

In real hyperbolic spaces, the situation is slightly better owing to Thurston's examples of 3–manifolds fibering over the circle. Based partly on Thurston's examples, Bowditch and Mess [13] have described an example of a finitely generated subgroup of a uniform lattice in $SO(4,1)$ that is not finitely presented. Abresch and Schroeder [1] have given an arithmetic construction of this lattice, too. One wonders if this arithmetic description can be used to give similar examples in $SU(4,1)$ or $Sp(4,1)$.

Such infinitely presented subgroups are necessarily distorted. Related examples have also been discovered by Potyagailo and Kapovich [55], [37].

A natural question is whether Thurston's construction goes through in higher dimensions or not:

Question Does there exist a uniform lattice in a rank one symmetric (other than $\mathbb{H}^3$) space containing a finitely presented (or even finitely generated) infinite normal subgroup of infinite index?

One should note that any such normal subgroup cannot be hyperbolic (by [57]).

Distortion in finitely presented groups

There are certain special classes of distorted subgroups of finitely presented groups that do not fall into any of the above categories.

A basic class of examples comes from the Baumslag Solitar groups

$$BS(1,n) = \{a, t : tat^{-1} = a^n\}$$

where the cyclic group generated by a has exponential distortion for $n > 1$.

A class of examples with higher distortion have appeared in work of Gersten [28]. We briefly describe these.

Take $G = \{g_1, \cdots, g_n : {g_{i-1}}^{g_i} = g_{i-1}^2 \text{ for } i = 2 \cdots n\}$. Then the cyclic subgroup generated by g_1 has distortion an iterated exponential function of height n.

Next consider $G = \{a, b, c : a^b = a^2, a^c = b\}$. Then the cyclic group generated by a has distortion greater than any iterated exponential.

Another class of subgroups with distortion a fractional power occurs in work of Bridson [16]:

Let $G_c = \mathbb{Z}^c \rtimes_{\phi_c} \mathbb{Z}$ where $\phi_c \in GL_c\mathbb{Z}$ is the unipotent matrix with ones on the diagonal and superdiagonal and zeroes elsewhere. For $c > 1$, G_c has infinite cyclic center. Given two such groups G_a, G_b amalgamate them along their cyclic center $\langle z \rangle$ to get $G(a, b) = G_a *_{\langle z \rangle} G_b$. Then the distortion function of G_b in $G(a, b)$ is of the form $n^{\frac{a}{b}}$.

A large class of examples of distortion arise from subgroups of nilpotent and solvable groups [33].

Finally we describe a class of examples due to Mihailova [44] which give rise to non-recursive distortion (see also [33] [22]). Let $G = \{g_1, \cdots, g_n : r_1 \cdots r_m\}$ be any finitely presented group with defining presentation $f \colon F_n \to G$. Then $f \times f$ maps $F_n \times F_n$ to $G \times G$. The pull-back H under this map of the 'diagonal subgroup' $\{(g, g) : g \in G\}$ is generated by elements of the form (g_i, g_i), $i = 1 \cdots n$ and $(1, r_j)$, $j = 1 \cdots m$. If G has unsolvable word problem, then the distortion of H in $F_n \times F_n$ is non-recursive.

3 Characterization of quasiconvexity

It was seen in the previous section that construction of distorted subgroups usually involves some amount of work. In fact for subgroups of hyperbolic groups, Gromov [34] describes 'length–angle' relationships between generators that would ensure quasiconvexity of the subgroup. This can be taken as a genericity result. In another setting, one could ask for examples of groups all whose finitely generated subgroups are undistorted. This is known for free groups, surface groups and abelian groups.

However, a general group-theoretic characterization of quasiconvexity seems far off. Gersten has recently described a functional analytic approach to this problem. We briefly describe this. Later we shall discuss a more group-theoretic approach. We shall restrict ourselves to finitely generated subgroups of hyperbolic groups (in the sense of Gromov) in this section.

The following discussion appears in [27], [25], [2]. Let X' be a complex of type $K(G,1)$ with finite $(n+1)$ skeleton $X'^{(n+1)}$ and let X be the universal cover of X'. The vector space of cellular chains $C_i(X,\mathbb{R})$ is equipped with the l_1 norm for a basis of i–cells. Then the boundary maps $\delta_{i+1}\colon C_{i+1}(X,\mathbb{R}) \to C_i(X,\mathbb{R})$ are bounded linear and (owing to the finiteness of the $n+1$–skeleton) one gets quasi-isometry invariant homology groups $H_i^{(1)}(X,\mathbb{R})$ for $i \leq n$. Since these homology groups are quasi-isometry invariant it makes sense to define $H_i^{(1)}(G,\mathbb{R}) = H_i^{(1)}(X,\mathbb{R})$ for $i \leq n$ for any such X. The following Theorem of Gersten's occurs in [27].

Theorem 3.1 *The finitely presented group G is hyperbolic if and only if $H_1^{(1)}(G,\mathbb{R}) = 0$. Moreover, if H is a finitely generated subgroup of G then H is quasiconvex if and only if the map $H_1^{(1)}(H,\mathbb{R}) \to H_1^{(1)}(G,\mathbb{R})$ induced by inclusion is injective.*

Earlier results along these lines had been found in [26], [25], [2].

In a different direction, one would like a purely group-theoretic characterization of quasiconvexity. We start with some definitions.

Definition Let H be a subgroup of a group G. We say that the elements $\{g_i | 1 \leq i \leq n\}$ of G are essentially distinct if $Hg_i \neq Hg_j$ for $i \neq j$. Conjugates of H by essentially distinct elements are called essentially distinct conjugates.

Note that we are abusing notation slightly here, as a conjugate of H by an element belonging to the normalizer of H but not belonging to H is still essentially distinct from H. Thus in this context a conjugate of H records (implicitly) the conjugating element.

Definition We say that the height of an infinite subgroup H in G is n if there exists a collection of n essentially distinct conjugates of H such that the intersection of all the elements of the collection is infinite and n is maximal possible. We define the height of a finite subgroup to be 0.

The main theorem of [32] states:

Theorem 3.2 *If H is a quasiconvex subgroup of a hyperbolic group G,then H has finite height.*

The following question of Swarup was prompted partly by this result:

Question (Swarup) Suppose H is a finitely presented subgroup of a hyperbolic group G. If H has finite height is H quasiconvex in G?

So far only some partial answers have been obtained. The first result is due to Scott and Swarup:

Theorem 3.3 [58] *Let $1 \to H \to G \to \mathbb{Z} \to 1$ be an exact sequence of hyperbolic groups induced by a pseudo Anosov diffeomorphism of a closed surface with fundamental group H. Let H_1 be a finitely generated subgroup of infinite index in H. Then H_1 is quasiconvex in G.*

In [51] an analogous result for free groups was derived. The methods also provide a different proof of Scott and Swarup's theorem above:

Theorem 3.4 [51] *Let $1 \to H \to G \to \mathbb{Z} \to 1$ be an exact sequence of hyperbolic groups induced by a hyperbolic automorphism ϕ of the free group H. Let $H_1 (\subset H)$ be a finitely generated distorted subgroup of G. Then there exist $N > 0$ and a free factor K of H such that the conjugacy class of K is preserved by ϕ^N and H_1 contains a finite index subgroup of a conjugate of K.*

Another special case where one has a positive answer is the following:

Theorem 3.5 [50] *Let G be a hyperbolic group splitting over H (ie $G = G_1 *_H G_2$ or $G = G_1 *_H$) with hyperbolic vertex and edge groups. Further, assume the two inclusions of H are quasi-isometric embeddings. Then H is of finite height in G if and only if it is quasiconvex in G.*

Swarup's question is therefore still open in the following special case, which can be regarded as a next step following the Theorems of [51] and [50] above.

Question Suppose G splits over H satisfying the hypothesis of Theorem 3.5 above and H_1 is a quasiconvex subgroup of H. If H_1 has finite height in G is it quasiconvex in G? More generally, if H_1 is an edge group in a hyperbolic graph of hyperbolic groups satisfying the qi–embedded condition, is H quasiconvex in G if and only if it has finite height in G?

A closely related problem can be formulated in more geometric terms:

Question Let X_G be a finite 2 complex with fundamental group G. Let X_H be a cover of X_G corresponding to the finitely presented subgroup H. Let $I(x)$ be the injectivity radius of X_H at x.

Does $I(x) \to \infty$ as $x \to \infty$ imply that H is quasi-isometrically embedded in G?

A positive answer to this question for G hyperbolic would provide a positive answer to Swarup's question.

The answer to this question is negative if one allows G to be only finitely generated instead of finitely presented as the following example shows:

Example Let $F = \{a, b, c, d\}$ denote the free group on four generators. Let $u_i = a^i b^i$ and $v_i = c^{f(i)} d^{f(i)}$ for some function $f\colon \mathbb{N} \to \mathbb{N}$. Introducing a stable letter t conjugating u_i to v_i one has a finitely generated HNN extension G. The free subgroup generated by a, b provides a negative answer to the question above for suitable choice of f. In fact one only requires that f grows faster than any linear function.

If f is recursive one can embed the resultant G in a finitely presented group by Higman's Embedding Theorem. But then one might lose malnormality of the free subgroup generated by a, b. If one can have some control over the embedding in a finitely presented group, one might look for a counterexample. A closely related example was shown to the author by Steve Gersten.

So far the following question (attributed to Bestvina and Brady) remains open:

Question Let G be a finitely presented group with a finite $K(G, 1)$. Suppose moreover that G does not contain any subgroup isomorphic to $BS(m, n)$. Is G hyperbolic?

A malnormal counterexample to Swarup's question would provide a counterexample for the above question (observed independently by M. Sageev).

4 Boundary extensions

The purpose of this section is to take an asymptotic rather than a coarse point of view and expose some of the problems from this perspective. Since virtually all the work in this area involves actions of hyperbolic groups on hyperbolic metric spaces we restrict ourselves mostly to this.

Roughly speaking, one would like to know what happens 'at infinity'. We put this in the more general context of a hyperbolic group H acting freely and properly discontinuously by isometries on a proper hyperbolic metric space X. Then there is a natural map $i\colon \Gamma_H \to X$, sending the vertex set of Γ_H to the orbit of a point under H, and connecting images of adjacent vertices in Γ_H by geodesics in X. Let $\widehat{X}$ denote the Gromov compactification of X.

The basic question discussed in this section is the following:

Question Does the continuous proper map $i\colon \Gamma_H \to X$ extend to a continuous map $\hat{i}\colon \widehat{\Gamma_H} \to \widehat{X}$?

A measure–theoretic version of this question was asked by Bonahon in [7]. A positive answer to the above would imply a positive answer to Bonahon's question. Related questions in the context of Kleinian groups have been studied by Cannon and Thurston [20], Bonahon [8], Floyd [24] and Minsky [47].

Much of the work around this problem was inspired by a seminal (unpublished) paper of Cannon and Thurston [20]. The main theorem of [20] states:

Theorem 4.1 [20] *Let M be a closed hyperbolic 3–manifold fibering over the circle with fiber F. Let $\widetilde{F}$ and $\widetilde{M}$ denote the universal covers of F and M respectively. Then $\widetilde{F}$ and $\widetilde{M}$ are quasi-isometric to $\mathbb{H}^2$ and $\mathbb{H}^3$ respectively. Let $\mathbb{D}^2 = \mathbb{H}^2 \cup \mathbb{S}^1_\infty$ and $\mathbb{D}^3 = \mathbb{H}^3 \cup \mathbb{S}^2_\infty$ denote the standard compactifications. Then the usual inclusion of $\widetilde{F}$ into $\widetilde{M}$ extends to a continuous map from $\mathbb{D}^2$ to $\mathbb{D}^3$.*

The proof of the above theorem involved the construction of a local 'Sol-like' metric using affine structures on surfaces coming from stable and unstable foliations. Coupled with Thurston's hyperbolization of 3–manifolds fibering over the circle one has a very explicit description of the boundary extension.

Using these (local) methods Minsky [47] generalized this theorem to the following:

Theorem 4.2 [47] *Let Γ be a Kleinian group isomorphic (as a group) to the fundamental group of a closed surface, such that $\mathbb{H}^3/\Gamma = M$ has injectivity radius uniformly bounded below by some $\epsilon > 0$. Then there exists a continuous map from the Gromov boundary of Γ (regarded as an abstract group) to the limit set of Γ in $\mathbb{S}^2_\infty$.*

Finally Klarreich [38] generalized the above theorem to the case of freely indecomposable Kleinian groups. A different proof was given by the author [49] (see below).

Theorem 4.3 ([38],[49]) *Let Γ be a freely indecomposable Kleinian group, such that $\mathbb{H}^3/\Gamma = M$ has injectivity radius uniformly bounded below by some $\epsilon > 0$. Then there exists a continuous map from the Gromov boundary of Γ (regarded as an abstract group) to the limit set of Γ in $\mathbb{S}^2_\infty$.*

Klarreich proved Theorem 4.3 by combining her Theorem 4.4 below with Theorem 4.2 above.

Theorem 4.4 [38] *Let X and Y be proper, geodesic Gromov–hyperbolic spaces, H_α a collection of closed, disjoint path-connected subsets of X, and $h\colon X \to Y$ a quasi–Lipschitz map such that for every H_α, h restricted to H_α extends continuously to the boundary at infinity. Suppose that the following hold:*

1. *The complement in X of the sets H_α is open and path-connected as also the complement of $h(H_\alpha)$ in Y.*
2. *There is some real number $k > 0$ such that the sets H_α are all k–quasiconvex in X and $h(H_\alpha)$'s are k–quasiconvex in Y.*
3. *There is a real number $c > 0$ such that $d(H_\alpha, H_\beta) > c$ and such that $d(h(H_\alpha), h(H_\beta)) > 0$ for all α and β.*

Then if the map h induced on the electric spaces is a quasi-isometry, h extends continuously to a continuous map from the boundary of X to the boundary of Y. Here the electric spaces are the spaces obtained from X and Y by collapsing each space H_α (or $h(H_\alpha)$) to points: they inherit path metrics from X and Y.

One should note that since Cannon and Thurston's Theorem 4.1 deals with asymptotic behavior it might well be regarded as a theorem in coarse geometry. The above Theorems are all of this form. But the proof techniques in [20], [47] are local as they rely on Thurston's theory of singular foliations of surfaces. In [48] and [49] a different approach was described using purely large-scale techniques giving generalized versions of Theorems 4.1 4.2 and 4.3.

Theorem 4.5 [48] *Let G be a hyperbolic group and let H be a hyperbolic subgroup that is normal in G. Let $i\colon \Gamma_H \to \Gamma_G$ be the continuous proper embedding of Γ_H in Γ_G described above. Then i extends to a continuous map $\hat{i}$ from $\widehat{\Gamma_H}$ to $\widehat{\Gamma_G}$.*

A more useful generalization of Theorem 4.1 is:

Theorem 4.6 [49] *Let (X,d) be a tree (T) of hyperbolic metric spaces satisfying the quasi-isometrically embedded condition. Let v be a vertex of T. Let (X_v, d_v) denote the hyperbolic metric space corresponding to v. If X is hyperbolic then the inclusion of X_v in X extends continuously to the boundary.*

A direct consequence of Theorem 4.6 above is the following:

Corollary 4.7 *Let G be a hyperbolic group acting cocompactly on a simplicial tree T such that all vertex and edge stabilizers are hyperbolic. Also suppose that every inclusion of an edge stabilizer in a vertex stabilizer is a quasi-isometric embedding. Let H be the stabilizer of a vertex or edge of T. Then an inclusion of the Cayley graph of H into that of G extends continuously to the boundary.*

In [4], Bestvina and Feighn give sufficient conditions for a graph of hyperbolic groups to be hyperbolic. Vertex and edge subgroups are thus natural examples of hyperbolic subgroups of hyperbolic groups. These examples are covered by the above corollary.

Using Thurston's pleated surfaces technology one then gives a 'coarse' proof of Theorem 4.3. With some further work and using a theorem of Minsky [46], one can give [49] a 'partly coarse' proof of another result of Minsky [47]: Thurston's Ending Lamination Conjecture for geometrically tame manifolds with freely indecomposable fundamental group and a uniform lower bound on injectivity radius.

Theorem 4.8 [47] *Let N_1 and N_2 be homeomorphic hyperbolic 3–manifolds with freely indecomposable fundamental group. Suppose there exists a uniform lower bound $\epsilon > 0$ on the injectivity radii of N_1 and N_2. If the end invariants of corresponding ends of N_1 and N_2 are equal, then N_1 and N_2 are isometric.*

One should note here that the coarse techniques referred to circumvent only the building of a 'model manifold' — a local construction in [47]. It might be worthwhile to obtain a coarse proof of the main theorem of [46]. A positive answer to the following coarse question will do the job (as can be seen from [49]):

Question Let $\sigma\colon \mathbb{N} \to Teich(S)$ be a map. For l a closed curve on S, let l_i denote the length of the shortest curve freely homotopic to l on $\sigma(i)$.

Suppose there exists $\lambda > 1$ such that for all closed curves l on S one has

$$\lambda l_i \leq max(l_{i-1}, l_{i+1}) \text{ for all } i \in \mathbb{N}.$$

Then does σ lie in a bounded neighborhood of a Teichmuller geodesic?

The above question was motivated in part by the 'hallways flare' condition of [4] and a recent relative hyperbolicity result of Masur–Minsky [42].

Since a continuous image of a compact locally connected set is locally connected [36] Theorem 4.3 also shows that the limit sets of freely indecomposable Kleinian groups with a uniform lower bound on the injectivity radius are locally connected. The issue of local connectivity has received a lot of attention lately due to some recent foundational work of Bowditch and Swarup [10], [11], [14], [12], [60] following earlier work by Bestvina and Mess [6].

Theorem 4.9 ([10], [60]) *Let H be a one-ended hyperbolic group. Then its boundary is locally connected. Next assume H does not split over any two-ended group and acts on a proper hyperbolic metric space X with limit set $\Lambda \subset \partial X$. Then Λ is locally connected.*

The existence of continuous boundary extensions in general would thus imply (using Theorem 4.9) local connectivity of limit sets of hyperbolic groups acting on proper hyperbolic metric spaces. One wonders if some kind of a converse exists.

Such speculations are prompted on the one hand by Theorem 4.9 and by the following observation. Let Γ be a simply degenerate Kleinian group isomorphic to a surface group. Further assume Γ has no parabolics. Let Λ be the limit set of Γ, Ω its domain of discontinuity and X the boundary of the convex hull of Λ. Then 'nearest point projections' give a natural homeomorphism between Ω and X. From this it is easy to conclude that a continuous boundary extension exists if and only if a neighborhood of Λ in S^2_∞ deformation retracts onto Λ. In this special case therefore local connectivity is equivalent to continuous boundary extensions.

Before concluding this section it is worth pointing out that one needs finer invariants than distortion to understand asymptotic extrinsic geometry. One way of approaching the problem is to consider extrinsic geometry of rays (starting at $1 \in \Gamma_H$) and describe those which are not quasigeodesics in the ambient space

X. If one looks at bi-infinite geodesics instead of rays one gets 'ending laminations'. For 3–manifolds fibering over the circle with fiber F and monodromy ϕ one can think of these as the stable and unstable foliations of ϕ. Motivated by this, the author gave a more group theoretic description in [52] in the special case of a hyperbolic normal subgroup of a hyperbolic group.

Recall that for a hyperbolic 3–manifold M fibering over the circle with fiber F Cannon and Thurston show in [20] that the usual inclusion of $\widetilde{F}$ into $\widetilde{M}$ extends to a continuous map from $\mathbb{D}^2$ to $\mathbb{D}^3$. An explicit description of this map was also described in [20] in terms of 'ending laminations' [See [61] for definitions]. The explicit description depends on Thurston's theory of stable and unstable laminations for pseudo-anosov diffeomorphisms of surfaces [23]. In the case of normal hyperbolic subgroups of hyperbolic groups, though existence of a continuous extension $\hat{i}\colon \widehat{\Gamma_H} \to \widehat{\Gamma_G}$ was proven in [48], an explicit description was missing. In [52] some parts of Thurston's theory of ending laminations were generalized to the context of normal hyperbolic subgroups of hyperbolic groups. Using this an explicit description of the continuous boundary extension $\hat{i}\colon \widehat{\Gamma_H} \to \widehat{\Gamma_G}$ was given for H a normal hyperbolic subgroup of a hyperbolic group G.

In general, if

$$1 \to H \to G \to Q \to 1$$

is an exact sequence of finitely presented groups where H, G and hence Q (from [54]) are hyperbolic, one has ending laminations naturally parametrized by points in the boundary $\partial\Gamma_Q$ of the quotient group Q.

Corresponding to every element $g \in G$ there exists an automorphism of H taking h to $g^{-1}hg$ for $h \in H$. Such an automorphism induces a bijection ϕ_g of the vertices of Γ_H. This gives rise to a map from Γ_H to itself, sending an edge $[a,b]$ linearly to a shortest edge-path joining $\phi_g(a)$ to $\phi_g(b)$.

Fixing $z \in \partial\Gamma_Q$ for the time being (for notational convenience) we shall define the set of ending laminations corresponding to z.

Let $[1,z)$ be a semi-infinite geodesic ray in Γ_Q starting at the identity 1 and converging to $z \in \partial\Gamma_Q$. Let σ be a single-valued quasi-isometric section of Q into G. Let z_n be the vertex on $[1,z)$ such that $d_Q(1,z_n) = n$ and let $g_n = \sigma(z_n)$.

Given $h \in H$ let Σ_n^h be the (H–invariant) collection of all free homotopy representatives (or shortest representatives in the same conjugacy class) of $\phi_{g_n^{-1}}(h)$

in Γ_H. Identifying equivalent geodesics in Σ^h_n one obtains a subset S^h_n of (unordered) pairs of points in $\widehat{\Gamma_H}$. The intersection with $\partial^2\Gamma_H$ of the union of all subsequential limits (in the Chabauty topology) of $\{S^h_n\}$ will be denoted by Λ^h_z.

Definition The set of *ending laminations corresponding to* $z \in \partial\Gamma_Q$ is given by

$$\Lambda_z = \bigcup_{h\in H} \Lambda^h_z.$$

Definition The set Λ of all *ending laminations* is defined by

$$\Lambda = \bigcup_{z\in\partial\Gamma_Q} \Lambda_z.$$

It was shown in [52] that the continuous boundary extension $\hat{i}$ identifies end-points of leaves of the ending lamination. Further if $\hat{i}$ identifies a pair of points in $\partial\Gamma_H$, then a bi-infinite geodesic having these points as its end-points is a leaf of the ending lamination.

Similar descriptions of laminations have been used by Bestvina, Feighn and Handel for free groups [5]. Using these two descriptions in conjunction gives further information eg about subgroup structure [51].

5 Other invariants in extrinsic geometry

To fix notions consider a finitely generated group H acting on a path-metric space X. As mentioned in the introduction distortion arises out of comparing the intrinsic metric on Γ_H to the metric inherited from the ambient space X. Alternately this can be regarded as arising out of comparing filling functions, where one fills a copy of S^0 in Γ_H and X and compares the sizes of the chains required.

In Chapter 5 of [33] Gromov defines several filling invariants of spaces. Each of these gives rise to a relative version and corresponding distortion functions. Recall some of these from [33].

Given a simplicial n–cycle S in a homotopically (or homologically) n–connected simplicial complex X one constructs fillings of S by $(n+1)$ chains in X.

Definition Filling volume, denoted $FillVol_n(S,X)$ is the infimal simplicial volume of $(n+1)$ chains filling S.

Definition Filling radius, denoted $FillRad_n(S,X)$ is the minimal R such that S bounds in an R–neighborhood $U_R(S) \subset X$.

A host of other filling invariants are defined in [33] but we focus on these two.

We will define relative versions of the above two notions. Since the definitions of these invariants require n–connectedness of the spaces we shall assume that whenever these invariants are defined, the spaces in question are quasi-isometric to (or admit thickenings that are) n–connected. It will be clear that one gets quasi-isometry invariants in the process. Reference to an explicit quasi-isometry may at times be suppressed.

Distortion of $Fillvol_n$ and $FillRad_n$ can be defined in a somewhat more general context. Fix classes $\mathcal{S}_n(X)$ and $\mathcal{S}_n(Y)$ of n–cycles in X, Y respectively (eg one might restrict to connected cycles or images of spheres) such that $\mathcal{S}_n(X) \subset \mathcal{S}_n(Y)$. Let f_n be one be one of the functions $FillVol_n$ or $FillRad_n$. Define

$$\mathcal{S}_n(f_n, m, X) = \{S \in \mathcal{S}_n(X) : f_n(S) \leq m\}.$$

Finally define

$$Disto(f, X, Y, m) = sup(f_n(S, Y))$$
where the sup is taken over $S \in \mathcal{S}_n(f_n, m, X) \cap \mathcal{S}_n(Y)$.

For $n = 0$, $\mathcal{S}-$ the set of maps of the 0–sphere S^0 and $f_0 = FillVol_0$ or $FillRad_0$ we get back the original distortion function. Note that $FillRad_0$ is approximately half of $FillVol_0$.

For $n = 1$, $\mathcal{S}$ the set of maps of S^1 and $f_0 = FillVol_1$ we get *area distortion* in the sense of Gersten [29]. Distortion has been surveyed in Section 2. We give a brief sketch of Gersten's results on area distortion.

Definition An automorphism of a finitely presented group is *tame* if it lifts to an automorphism of the free group on its generators, preserving the normal subgroup generated by relators.

Theorem 5.1 *Let ϕ be a tame automorphism of a one-relator group G. Then area is undistorted for $G \subset G \rtimes_\phi \mathbb{Z}$.*

In [29] Gersten shows that in extensions of $\mathbb{Z}$ by finitely presented groups G area distortion of G is at most an exponential of an isoperimetric function for the extension. Moreover, he describes examples of undistorted (in the usual sense of length) subgroups that exhibit area distortion. He observes further that for torus bundles over the circle with Sol geometry, area in the fiber subgroup is undistorted whereas length is exponentially distorted.

Gersten showed further that area is undistorted for finitely presented subgroups of finitely presented groups of cohomological dimension 2. From this it follows that finitely presented subgroups H of hyperbolic groups G are finitely presented provided G has cohomological dimension 2 or G is a hyperbolic small cancellation group [30].

The remaining distortion functions are yet to be studied systematically. The first class of examples where $Disto(FillVol_n, X, Y, m)$ seem interesting and tractable are examples coming from extensions of $\mathbb{Z}$ by $\mathbb{Z}^n$, ie for $G = \mathbb{Z}^n \rtimes_\phi \mathbb{Z}$ where $\phi \in GL_n\mathbb{Z}$. Such examples have been studied by Bridson [17] and Bridson and Gersten [18].

Much less is known about $Disto(FillRad_n, X, Y, m)$. These functions are related to topology of balls in groups (Chapter 4 of [33]). For a group Γ admitting a uniformly k–connected thickening X (see [33] for definitions) Gromov defines $\overline{R_k}(r)$ to be the infimal radius $R \geq r$ such that the inclusion of balls $B(r) \subset B(R)$ is k–connected.

The following observations are straightforward generalizations of corresponding statements (for $n = 0$) on pages 74–76 of [33]. Fix a group Γ' and a subgroup Γ.

Proposition 5.2 *If $Disto(FillRad_n, \Gamma', \Gamma, m)$ is superexponential in m then the function $\overline{R_k}(m)$ for $(\Gamma, dist_{\Gamma'}|\Gamma)$ grows faster than any linear function Cm.*

Proposition 5.3 *Take two copies of $(\Gamma', \Gamma \subset \Gamma')$ and let $\Gamma_1 = \Gamma' *_\Gamma \Gamma'$ be the double. Then the function $\overline{R_k}(m)$ for Γ_1 is minorized by $\overline{R_{k-1}}(m)$ for $(\Gamma, dist_{\Gamma'}|\Gamma)$.*

This leads to the following

Question Do there exist pairs of groups $H \subset G$ (with n–connected inclusions of thickenings of the Cayley Graph) such that $Disto(FillRad_n, \Gamma_G, \Gamma_H, m)$ is superexponential in m?

A positive answer will furnish (via Proposition 5.3) examples of groups with fast growing $\overline{R_k}(m)$ for $k \geq 2$ (page 80 of [33]). No such example has been found yet.

References

[1] **U Abresch**, **V Schroeder**, *Graph manifolds, ends of negatively curved spaces and the hyperbolic 120–cell space*, J. Differential Geometry 35 (1992) 299–336

[2] **D Allcock**, **S Gersten**, *A Homological Characterization of Hyperbolic Groups*, preprint

[3] **J Alonso**, **T Brady**, **D Cooper**, **V Ferlini**, **M Lustig**, **M Mihalik**, **M Shapiro**, **H Short**, *Notes on word hyperbolic groups*, from: "Group Theory from a Geometrical Viewpoint", E Ghys, A Haefliger, and A Verjovsky (editors), World Scientific (1991) 3–63

[4] **M Bestvina**, **M Feighn**, *A Combination theorem for Negatively Curved Groups*, J. Differential Geometry 35 (1992) 85–101

[5] **M Bestvina**, **M Feighn**, **M Handel**, *Laminations, trees and irreducible automorphisms of free groups*, Geom. and Func. Anal. 7 (1997) 215–244

[6] **M Bestvina**, **G Mess**, *The boundary of negatively curved groups*, J. Amer. Math. Soc. 4 (1991) 469–481

[7] **F Bonahon**, *Geodesic Currents on Negatively Curved Groups*, from: "Arboreal Group Theory", R C Alperin (editor), MSRI Publ. 19 Springer–Verlag (1991) 143–168

[8] **F Bonahon**, *Bouts de varietes hyperboliques de dimension* 3, Annals of Math. 124 (1986) 71–158

[9] **F Bonahon**, **J P Otal**, *Varietes hyperboliques a geodesiques arbitrairement courtes*, Bull. London Math. Soc. 20 (1988) 255–261

[10] **B H Bowditch**, *Connectedness properties of limit sets*, Trans. Amer. Math. Soc. (to appear)

[11] **B H Bowditch**, *Convergence groups and configuration spaces*, from: "Group Theory Down Under", J Cossey, C F Miller, W D Neumann, M Shapiro (editors), de Gruyter (to appear)

[12] **B H Bowditch**, *Treelike structures arising from continua and convergence groups*, Trans. Amer. Math. Soc. (to appear)

[13] **B H Bowditch**, **G Mess**, *A four–dimensional Kleinian group*, Trans. Amer. Math. Soc. 344 (1994) 391–405

[14] **B H Bowditch**, **G A Swarup**, *Cut points in the boundaries of hyperbolic groups*, preprint

[15] **N Brady**, *Branched coverings of cubical complexes and subgroups of hyperbolic groups*, J. London Math. Soc. (to appear)

[16] **M Bridson**, *Fractional Isoperimetric Inequalities and Subgroup Distortion*, preprint

[17] **M Bridson**, *Optimal Isoperimetric Inequalities for abelian–by–free groups*, Topology 34 (1995) 547–564

[18] **M Bridson**, *The optimal isoperimetric inequality for torus bundles over the circle*, Quart. J. Math. 47 (1996) 1–39

[19] **R D Canary**, *Ends of hyperbolic* 3*–manifolds*, J. Amer. Math. Soc. (1993) 1–35

[20] **J Cannon**, **W P Thurston**, *Group Invariant Peano Curves*, preprint

[21] **M Coornaert**, **T Delzant**, **A Papadopoulos**, *Geometrie et theorie des groupes*, Lecture Notes in Math. 1441 Springer–Verlag (1990)

[22] **B Farb**, *The extrinsic geometry of subgroups and the generalized word problem*, Proc. London Math. Soc. (3) 68 (1994) 577–593

[23] **A Fathi**, **M Laudenbach**, **V Poenaru**, *Travaux de Thurston sur les surfaces*, Asterisque 66-67 (1979) 1–284

[24] **W J Floyd**, *Group Completions and Limit Sets of Kleinian Groups*, Invent. Math. 57 (1980) 205–218

[25] **S Gersten**, *A Cohomological Characterization of Hyperbolic Groups*, preprint

[26] **S Gersten**, *Cohomological Lower Bounds on Isoperimetric Functions of Groups*, Topology 37 (1998) 1031–1072

[27] **S Gersten**, *Some Remarks on the l_1 Homology of Finitely Presented Groups*, preprint

[28] **S Gersten**, *Isoperimetric and Isodiametric Functions of Finitely Presented Groups*, from: "Geometric Group Theory I", G A Niblo and M A Robler (editors), LMS Lecture Notes 182, Cambridge University Press (1993)

[29] **S Gersten**, *Preservation and distortion of area*, Geom. and Func. Anal. 6 (1996) 301–345

[30] **S Gersten**, *Subgroups of hyperbolic groups in dimension* 2, J. London Math. Soc. (1996) 261–283

[31] **E Ghys**, **P de la Harpe** (editors), *Sur les groupes hyperboliques d'apres Mikhael Gromov*, Progress in Math. 83, Birkhauser, Boston MA (1990)

[32] **R Gitik**, **M Mitra**, **E Rips**, **M Sageev**, *Widths of Subgroups*, Trans. Amer. Math. Soc. (1997) 321–329

[33] **M Gromov**, *Asymptotic Invariants of Infinite Groups*, from: "Geometric Group Theory II", LMS Lecture Notes 182, Cambridge University Press (1993)

[34] **M Gromov**, *Hyperbolic Groups*, from: "Essays in Group Theory", S Gersten (editor), MSRI Publ. 8, Springer–Verlag (1985) 75–263

[35] **M Gromov**, *Sign and geometric meaning of curvature*, Sem. Mat. e Fis. di Milano LXI ('91) Pavia 9 (1994) 10–123

[36] **J G Hocking**, **G S Young**, *Topology*, Addison–Wesley (1961)

[37] **M Kapovich**, **L Potyagailo**, *On absence of Ahlfors' finiteness theorem for Kleinian groups in dimension* 3, Topology and its Appl. 40 (1991) 83–91

[38] **E Klarreich**, PhD Thesis, SUNY Stonybrook

[39] **A Lubotzky**, **S Mozes**, **M S Raghunathan**, *Cyclic subgroups of exponential growth and metrics on discrete groups*, preprint

[40] **A Marden**, *The geometry of finitely generated kleinian groups*, Annals of Math. 99 (1974) 383–462

[41] **G A Margulis**, *Discrete Subgroups of Semisimple Lie Groups*, Springer–Verlag (1990)

[42] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves I: Hyperbolicity*, preprint

[43] **C McMullen**, *Iteration on Teichmuller Space*, Invent. Math. 9 (1990) 425–454

[44] **K A Mihailova**, *The occurrence problem for direct products of groups*, Dokl. Akad. Nauk. SSSR 119 (1958) 1103–1105

[45] **Y N Minsky**, *The Classification of Punctured Torus Groups*, Annals of Math. (to appear)

[46] **Y N Minsky**, *Teichmuller Geodesics and Ends of 3–Manifolds*, Topology (1992) 1–25

[47] **Y N Minsky**, *On Rigidity, Limit Sets, and End Invariants of Hyperbolic 3–Manifolds*, J. Amer. Math. Soc. 7 (1994) 539–588

[48] **M Mitra**, *Cannon–Thurston Maps for Hyperbolic Group Extensions*, Topology 37 (1998) 527–538

[49] **M Mitra**, *Cannon–Thurston Maps for Trees of Hyperbolic Metric Spaces*, J. Differential Geometry 48 (1998) 135–164

[50] **M Mitra**, *Height in Splittings of Hyperbolic Groups*, preprint

[51] **M Mitra**, *On a theorem of Scott and Swarup*, preprint

[52] **M Mitra**, *Ending Laminations for Hyperbolic Group Extensions*, Geom. and Func. Anal. 7 (1997) 379–402

[53] **L Mosher**, *A hyperbolic–by–hyperbolic hyperbolic group*, Proc. Amer. Math. Soc. (to appear)

[54] **L Mosher**, *Hyperbolic Extensions of Groups*, J. of Pure and Applied Algebra 110 (1996) 305–314

[55] **L Potyagailo**, *Finitely generated Kleinian groups in 3–space and 3–manifolds of infinite topological type*, Trans. Amer. Math. Soc. 344 (1994) 57–79

[56] **E Rips**, *Subgroups of small cancellation groups*, Bull. London Math. Soc. 14 (1982) 45–47

[57] **E Rips**, **Z Sela**, *Structure and rigidity in hyperbolic groups*, Geom. and Func. Anal. 4 (1994) 337–371

[58] **P Scott**, **G Swarup**, *Geometric Finiteness of Certain Kleinian Groups*, Proc. Amer. Math. Soc. 109 (1990) 765–768

[59] **Z Sela**, *Structure and rigidity in (gromov) hyperbolic groups and discrete subgroups in rank 1 lie groups*, Geom. and Func. Anal. (to appear)

[60] **G A Swarup**, *On the cut point conjecture*, ERA–AMS 2 (1996) 98–100

[61] **W P Thurston**, *The Geometry and Topology of 3–Manifolds*, Princeton University Notes

[62] **W P Thurston**, *Hyperbolic Structures on 3–Manifolds II, Surface Groups and 3-Manifolds which Fiber over the Circle*, preprint

Institute of Mathematical Sciences, C.I.T. Campus
Madras (Chennai) – 600113, India

Email: `mitra@imsc.ernet.in`

Received: 20 November 1997

ISSN 1464-8997 **365**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 365–381

Mutants and $SU(3)_q$ invariants

H R Morton
H J Ryder

Abstract Details of quantum knot invariant calculations using a specific $SU(3)_q$–module are given which distinguish the Conway and Kinoshita–Teresaka pair of mutant knots. Features of Kuperberg's skein-theoretic techniques for $SU(3)_q$ invariants in the context of mutant knots are also discussed.

AMS Classification 57M25; 17B37, 22E47

Keywords Mutants, Vassiliev invariants, $SU(3)_q$

1 Introduction

In previous studies of invariants derived from the Homfly polynomial, or equivalently from the unitary quantum groups, it was noted that no invariant given by a module over $SU(3)_q$ was known to distinguish a mutant pair of knots. Indeed, any quantum group module whose tensor square has no repeated summands determines a knot invariant which fails to distinguish mutants [3]. A table of invariants which fail to distinguish mutants was drawn up in [3], using this and other evidence. Direct Homfly polynomial calculations showed that a certain irreducible $SU(N)_q$ invariant, coming from the module with Young diagram $\yng(2,1)$, could distinguish between some mutant pairs for $N \geq 4$, although not for $N = 3$. These calculations also exhibited a Vassiliev invariant of finite type 11 which distinguishes some mutant pairs. The calculations left open the possibility that $SU(3)_q$ invariants might never distinguish mutant pairs.

In this paper we give details of calculations with a specific $SU(3)_q$–module which result in different invariants for the Conway and Kinoshita–Teresaka pair of mutant knots. We also consider some features of Kuperberg's skein-theoretic techniques for $SU(3)_q$ invariants in the context of mutant knots.

Much of this work was carried out in 1994–95, while the second author was supported by EPSRC grant GR/J72332.

1.1 Background

The term *mutant* was coined by Conway, and refers to the following general construction.

Suppose that a knot K can be decomposed into two oriented 2–tangles F and G as shown in figure 1.

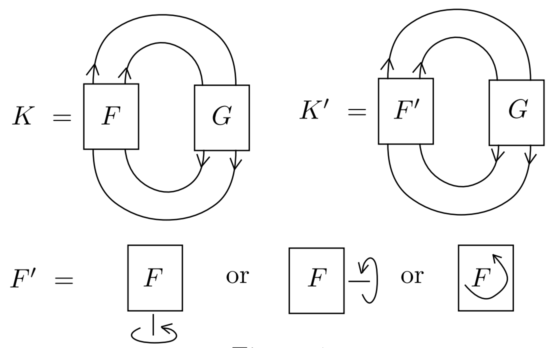


Figure 1

A new knot K' can be formed by replacing the tangle F with the tangle F' given by rotating F through π in one of three ways, reversing its string orientations if necessary. Any of these three knots K' is called a *mutant* of K.

The two 11–crossing knots with trivial Alexander polynomial found by Conway and Kinoshita–Teresaka are the best-known example of mutant knots. They are shown in figure 2.

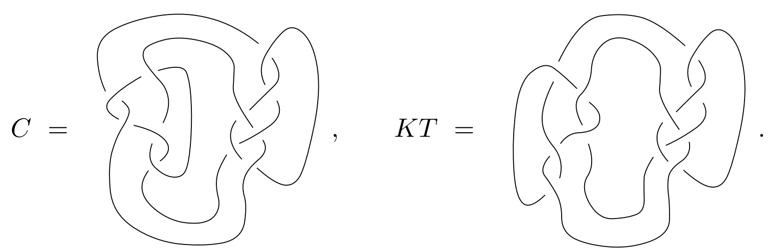


Figure 2

It is clear from figure 2 that the knots C and KT are mutants, and the consituent tangles F and G are both given from a 3–string braid by closing off one of the strings.

The simplest $SU(3)_q$ invariant not previously known to agree on mutant pairs is given by the 15–dimensional irreducible module with Young diagram .

The Homfly polynomial of the 4–parallel with $z = s - s^{-1}$ and $v = s^N$ is a sum of 4–cell invariants for $SU(N)_q$. When $N = 3$ it is known that all 4–cell invariants except that for $\yng(3,1)$ agree on mutants. Thus the Homfly polynomial of the 4–parallel, with the substitution $z = s - s^{-1}$ and $v = s^3$, agrees on mutants if and only if the $SU(3)_q$ invariant for $\yng(3,1)$ agrees on mutants.

Equally, the same substitution in the Homfly polynomial of the satellite consisting of the parallel with 3 strings, two oriented in one direction and one in the reverse direction, gives the sum of certain 4–cell invariants for $SU(3)_q$, because the dual of the fundamental module, used to colour the reverse string, is given by using the Young diagram with a single column of two cells. Then the Homfly polynomial of the 3–parallel with one reverse string, after the substitution $z = s - s^{-1}$, $v = s^3$ agrees on mutants if and only if the $SU(3)_q$ invariant for $\yng(3,1)$ agrees on mutants.

Kuperberg's combinatorial methods for handling $SU(3)_q$ invariants seemed to us for a while to offer a chance that the behaviour of $SU(3)_q$ would follow that of $SU(2)_q$. We explored the $SU(3)_q$ skein of the pair of pants, based on Kuperberg's combinatorial techniques, in the hope of proving this. An analysis of this skein is given later, as it has a geometrically appealing basis, whose first lack of symmetry again pointed the finger at the reversed 3–parallel as the first potential candidate for distinguishing some mutant pairs.

1.2 Choice of calculational method

We did not pursue the Kuperberg skein calculations for these parallels of Conway and Kinoshita–Teresaka. Although we contemplated briefly such an approach it seemed difficult to use computational aids in dealing with combinatorial skein diagrams once the number of crossings to be resolved grew beyond easy blackboard calculations, as no computer implementation of the graphical calculations in this skein was available to us.

While we could, in principle, have calculated the Homfly polynomial of the 3–parallel of Conway's knot with one string reversed there is a considerable problem in computation of Homfly polynomials of links with a large number of crossings. A number of computer programs will calculate the Homfly polynomials of general links. Mostly these rely on implementation of the skein relation, and the time required grows exponentially with the number of crossings. Such programs include those by Ochiai, Millett and Hoste. They will work up to the order of maybe 40 or even 50 crossings but slow down rapidly after that. In the application needed for this paper we have to deal with the 3–string parallel to Conway's knot with two strings in one direction and one in the other, which gives a link with 99 crossings. Even if the calculation is restricted to dealing with terms up to z^{13} only, or some similar bound, these programs are unlikely to make any impact on the calculations.

There does exist a program, developed by Morton and Short [4], which can handle links with a large numbers of crossings, under some circumstances. This is based on the Hecke algebras, but it requires a braid presentation of the link on a restricted number of strings; in practice 9 strings is a working limit, although in favourable circumstances it can be enough to break the link into pieces which meet this bound more locally. Unfortunately the reverse orientation of one string which is needed in the present case means that any braid presentation for the resulting link falls well outside the limitations of this program.

In [3] the Hecke algebra calculations on 3–string parallels with all strings in the same direction could be carried out in terms of 9–string braids, and lent themselves well to an effective truncation to restrict the degree of Vassiliev invariants which had to be calculated. The alternative possibility here of using 4 parallel strings, all with the same orientation, faces the uncomfortable growth of these calculations from 9–string to 12–string braids, entailing a growth in storage from 9! to 12! for a calculation which was already nearing its limit. There are also almost twice as many crossings (11×16), as well as a similar factorial growth in overheads for the calculations.

We consequently did not pursue Homfly calculations any further. Instead we returned to the $SU(3)_q$–module calculations and made explicit computations for the invariants of the knots C and KT when coloured by the 15–dimensional module $V_{\yng(3,1)}$, using the following scheme. This approach has the merit of focussing directly on the key part of the $SU(3)_q$ specialisation, rather than using the full Homfly polynomial on some parallel link. We give further details of the method later.

When each of the knots C and KT is coloured by the $SU(3)_q$–module $V_{\yng(3,1)}$ the two constituent tangles F and G will be represented by an endomorphism of the module $V_{\yng(3,1)} \otimes V_{\yng(3,1)}$. To calculate the invariant of the knot, presented as the closure of the composite of the two 2–tangles, we may compose the endomorphisms for the two 2–tangles, and then calculate the invariant of the closure of the composite tangle in terms of the resulting endomorphism. Let us suppose that $V_{\yng(3,1)} \otimes V_{\yng(3,1)}$ decomposes as a sum $\bigoplus a_\nu V_\nu$ of irreducible modules, where $a_\nu \in \mathbf{N}$ and $a_\nu V_\nu$ denotes the sum of all submodules which are isomorphic to V_ν. Any endomorphism then maps each isotypic piece $a_\nu V_\nu$ to itself. It is convenient to regard each isotypic piece as a vector space of the form $W_\nu \otimes V_\nu$, where W_ν has dimension a_ν, and can be explicitly identified with the space of highest weight vectors for the irreducible module V_ν in $V_{\yng(3,1)} \otimes V_{\yng(3,1)}$. Any endomorphism α of $V_{\yng(3,1)} \otimes V_{\yng(3,1)}$ maps each space W_ν to itself, and is determined by the resulting linear maps $\alpha_\nu\colon W_\nu \to W_\nu$.

Where two endomorphisms α and β of $\bigoplus(W_\nu \otimes V_\nu)$ are composed, the corresponding restrictions to each weight space W_ν compose, to give $(\alpha \circ \beta)_\nu = \alpha_\nu \circ \beta_\nu$. Now the invariant of the closure of a tangle represented by an endomorphism γ of $\bigoplus(W_\nu \otimes V_\nu)$ is known to be $\sum(tr(\gamma_\nu) \times \delta_\nu)$, where $\delta_\nu = J_O(V_\nu)$ is the quantum dimension of the module V_ν. The difference of the invariants for two knots represented respectively by γ and γ' is then given in the same way using $\gamma - \gamma'$ in place of γ.

The invariants for Conway and Kinoshita–Teresaka arise in this way from endomorphisms $\gamma = \alpha \circ \beta$ and $\gamma' = \alpha' \circ \beta$, in which α and α' represent one of the 2–tangles for Conway, and the same tangle turned over for Kinoshita–Teresaka, while the other tangle gives the same β in each case. We can write $\alpha' = R^{-1} \circ \alpha \circ R$ as module endomorphisms, where R is the R–matrix for $V_{\yng(3,1)}$. Clearly, for those ν with dim $W_\nu = 1$ we will have $\alpha'_\nu = \alpha_\nu$, and so $\gamma'_\nu - \gamma_\nu = 0$. (As noted in [3], if this happens for all ν then the invariant cannot distinguish any mutant pair). The final difference of invariants will thus depend only on those ν where the summand V_ν has multiplicity greater than 1. In the case here there are just two such ν and in each case the space W_ν has dimension 2. The calculation then reduces to the determination of the 2×2 matrices representing α_ν, α'_ν and β_ν.

1.3 Result of the explicit calculation

The difference between the values of the invariant on Conway's knot and on the Kinoshita–Teresaka knot is

$$s^{-80}(s^8+1)^2(s^4+1)^4(s+1)^{13}(s-1)^{13}(s^2-s+1)^3(s^2+s+1)^3$$
$$(s^6-s^5+s^4-s^3+s^2-s+1)(s^6+s^5+s^4+s^3+s^2+s+1)$$
$$(s^4-s^3+s^2-s+1)(s^4+s^3+s^2+s+1)(s^4-s^2+1)(s^2+1)^6$$
$$(s^{46}-s^{44}+2\,s^{40}-4\,s^{38}+2\,s^{36}+3\,s^{34}-4\,s^{32}+6\,s^{30}-s^{28}-3\,s^{26}+6\,s^{24}$$
$$-4\,s^{22}+4\,s^{20}+2\,s^{18}-5\,s^{16}+5\,s^{14}-2\,s^{12}-2\,s^{10}+4\,s^8-2\,s^6+s^2-1)$$

up to a power of the variable s.

This may be rewritten to indicate more clearly the appearance of roots of unity as the product of $(s^{46}-s^{44}+2\,s^{40}-4\,s^{38}+2\,s^{36}+3\,s^{34}-4\,s^{32}+6\,s^{30}-s^{28}-3\,s^{26}+$

$6\,s^{24}-4\,s^{22}+4\,s^{20}+2\,s^{18}-5\,s^{16}+5\,s^{14}-2\,s^{12}-2\,s^{10}+4\,s^{8}-2\,s^{6}+s^{2}-1)$ with the factors $(s^8-s^{-8})^2(s^7-s^{-7})(s^6-s^{-6})(s^5-s^{-5})(s^4-s^{-4})^2(s^3-s^{-3})^2(s^2-s^{-2})(s-s^{-1})^3$, and a power of s.

When this is written as a power series in h with $s=e^{h/2}$ the first term becomes $7+O(h)$ and the other factors contribute $ch^{13}+O(h^{14})$, where the coefficient c is $c=8^2.7.6.5.4^2.3^2.2$. The coefficient of h^{13} in the power series expansion of the $SU(3)_q$ invariant for the 15–dimensional irreducible module is thus a Vassiliev invariant of type at most 13 which differs on the two mutant knots.

1.4 Some background to the calculational method

In the following section we give details of the methods used in our calculations. We feel it is important that others can in principle check the calculations, as we were very much aware in setting up our initial data just how much scope there is for error. It can easily cause problems, for example, if some of the data is taken from one source and some from another which has been normalised in a slightly different way. When the goal is to show that some polynomial arising from the calculations is non-zero any mistake is almost bound to result in a non-zero polynomial even if the true polynomial is zero.

In our work here we have been reassured to find that the non-zero difference polynomial above at least has some roots which could be anticipated, since the difference must vanish at certain roots of unity. An error in the calculations would have been likely to give a difference which did not have these roots.

The computations were done in Maple, using its polynomial handling and linear algebra routines. In this way we avoided the need to write explicit Pascal or C programs for matrices and polynomials, although the computations were probably not as fast as with a compiled program. For comparison, a Maple version of the Hecke algebra program in [4] took roughly 50 times as long as the compiled Pascal program to calculate the Homfly polynomial of a variety of links when tested some time ago on the same machine.

1.5 The quantum group $SU(3)_q$

We start from a presentation of the quantum group $SU(3)_q$ as an algebra with six generators, $X_1^\pm$, $X_2^\pm$, H_1, H_2, and a description of the comultiplication and antipode. Let M be any finite-dimensional left module over $SU(3)_q$. The action of any one of these six generators Y will determine a linear endomorphism Y_M

of M. We build up explicit matrices for these endomorphisms on a selection of low-dimensional modules, using the comultiplication to deal with the tensor product of two known modules, and the antipode to construct the action on the linear dual of a known module. We must eventually determine the matrices Y_M for the 15–dimensional module $M = V_{\yng(3,1)}$ above, and find the 225×225 R–matrix, R_{MM} which represents the endomorphism of $M \otimes M$ needed for crossings.

Knowing Y_M we can find the generators Y_{MM} for the module $M \otimes M$, and thus identify the highest-weight vectors for this module. We can follow the effect of each 2–tangle F and G on the highest-weight vectors when we know how to take account of the closure of one of the strings in forming the 2–tangle from the 3–braid. To do this we need the fixed element T of the quantum group, corresponding to Turaev's 'enhancement' [6], which is used in forming the 'quantum trace'.

For the quantum groups coming from the classical Lie algebras there is a simple prescription for $T = \exp(h\rho)$ in terms of a linear form $\rho = \sum \mu_i H_i$, with coefficients determined by the Cartan matrix for the Lie algebra, [1]. In the case of $SU(3)_q$ we have $\rho = H_1 + H_2$. The quantum dimension of any module M is the trace of the matrix T_M representing the action of T on M. More generally, the effect of closing a string which is coloured by M, to convert an endomorphism of $V \otimes M$ into an endomorphism of V, can be realised by acting on M by T and then taking the partial trace of the composite linear endomorphism of $V \otimes M$. The element T is variously written as $u^{\pm 1} v$ or $u^{-1}\theta$ where v is Turaev's 'ribbon element' representing the full twist and u is constructed directly from the universal R–matrix, [7], [1].

We follow Kassel in the basic description of the quantum group from [1], chapter 17, using generators H_1 and H_2 for the Cartan sub-algebra, but with generators $X_i^{\pm}$ in place of X_i and Y_i. We use the notation $K_i = \exp(hH_i/4)$, and set $a = \exp(h/4)$, $s = \exp(h/2) = a^2$ and $q = \exp(h) = s^2$, unlike Kassel. The elements satisfy the commutation relations $[H_i, H_j] = 0$, $[H_i, X_j^{\pm}] = \pm a_{ij} X_j^{\pm}$, $[X_i^+, X_i^-] = (K_i^2 - K_i^{-2})/(s - s^{-1})$, where $(a_{ij}) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ is the Cartan matrix for $SU(3)$, and also the Serre relations of degree 3 between $X_1^{\pm}$ and $X_2^{\pm}$.

Comultiplication is given by

$$\begin{aligned} \Delta(H_i) &= H_i \otimes I + I \otimes H_i, \\ (\text{so } \Delta(K_i) &= K_i \otimes K_i,) \\ \Delta(X_i^{\pm}) &= X_i^{\pm} \otimes K_i + K_i^{-1} \otimes X_i^{\pm}, \end{aligned}$$

and the antipode S by $S(X_i^{\pm}) = -s^{\pm 1}X_i^{\pm}$, $S(H_i) = -H_i$, $S(K_i) = K_i^{-1}$.

The fundamental 3–dimensional module, which we denote by E, has a basis in which the quantum group generators are represented by the matrices Y_E as listed here.

$$X_1^+ = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; X_2^+ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$X_1^- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; X_2^- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; H_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

For calculations we keep track of the elements K_i rather than H_i, represented by

$$K_1 = \begin{pmatrix} a & 0 & 0 \\ 0 & a^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; K_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a^{-1} \end{pmatrix}$$

for the module E.

We can then write down the elements Y_{EE} for the actions of the generators Y on the module $E \otimes E$, from the comultiplication formulae. The R–matrix R_{EE} representing the endomorphism of $E \otimes E$ which is used for the crossing of two strings coloured by E can be given, up to a scalar, by the prescription

$$\begin{aligned} R_{EE}(e_i \otimes e_j) &= e_j \otimes e_i, \text{ if } i > j, \\ &= s\, e_i \otimes e_i, \text{ if } i = j, \\ &= e_j \otimes e_i + (s - s^{-1}) e_i \otimes e_j, \text{ if } i < j, \end{aligned}$$

for basis elements $\{e_i\}$ of E.

We made a quick check with Maple to confirm that the matrices Y_{EE} all commute with R_{EE}, as they should. It can also be checked that R_{EE} has eigenvalues s with multiplicity 6 and $-s^{-1}$ with multiplicity 3, and satisfies the equation $R - R^{-1} = (s - s^{-1})\mathrm{Id}$.

The linear dual M^* of a module M becomes a module when the action of a generator Y on $f \in M^*$ is defined by $< Y_{M^*} f, v > = < f, S(Y_M) v >$, for $v \in M$. For the dual module $F = E^*$ we then have matrices for Y_F, relative to the dual basis, as follows.

$$X_1^+ = \begin{pmatrix} 0 & 0 & 0 \\ -s & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; X_2^+ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -s & 0 \end{pmatrix}$$
$$X_1^- = \begin{pmatrix} 0 & -s^{-1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; X_2^- = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -s^{-1} \\ 0 & 0 & 0 \end{pmatrix}$$
$$K_1 = \begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; K_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a^{-1} & 0 \\ 0 & 0 & a \end{pmatrix}.$$

The most reliable way to work out the R–matrices R_{EF}, R_{FE} and R_{FF} is to combine R_{EE} with module homomorphisms cup_{EF}, cup_{FE}, cap_{EF} and cap_{FE} between the modules $E \otimes F$, $F \otimes E$ and the trivial 1–dimensional module, I, on which $X_i^{\pm}$ acts as zero and K_i as the identity. For example, to represent a homomorphism from I to $E \otimes F$ the matrix for cup_{EF} must satisfy $Y_{EF}\,\mathrm{cup}_{EF} = \mathrm{cup}_{EF}\,Y_I$, which identifies cup_{EF} as a common eigenvector of the matrices Y_{EF}, with eigenvalue 0 or 1 depending on Y. The matrices are determined up to a scalar by such considerations; when one has been chosen the scalar for the others is dictated by diagrammatic considerations. They are quite easy to write down theoretically, although to be careful about compatibility and possible miscopying it is as well to get Maple to find them in this way for itself. Once these matrices have been found they can be combined with the matrix R_{EE}^{-1} to construct the R–matrices R_{EF}, R_{FE}, R_{FF}, using the diagram shown in figure 3, for example, to determine R_{EF}. This gives

$$R_{EF} = 1_F \otimes 1_E \otimes \mathrm{cap}_{EF} \circ 1_F \otimes R_{EE}^{-1} \otimes 1_F \circ \mathrm{cup}_{FE} \otimes 1_E \otimes 1_F.$$

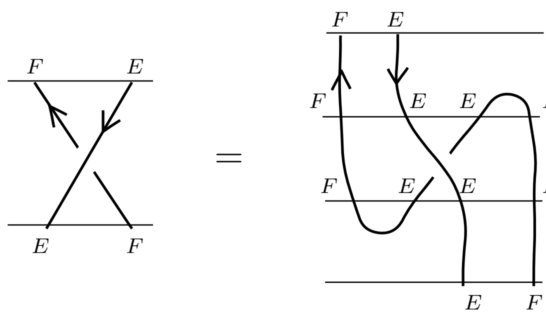


Figure 3

The module structure of $M = V_{\yng(3,1)}$ can be found by identifying M as a 15–dimensional submodule of $E \otimes E \otimes F$. We know that there will be a direct sum decomposition of $E \otimes E \otimes F$ as $M \oplus N$, and indeed that N will decompose further into the sum of two copies of a 3–dimensional module isomorphic to E and one 6–dimensional module with Young diagram ⊞ . The full twist element on the three strings coloured by E, E and F acts by a scalar on each of the irreducible submodules of $E \otimes E \otimes F$. It can be expressed as a 27×27 matrix in terms of the R–matrices above. Maple can then produce a basis for each of the eigenspaces, one of dimension 15 and the other two each of dimension 6. Write P and Q for the 27×15 and 27×12 matrices whose columns are made of these basis vectors. Then P and Q give bases for M and N respectively. The partitioned matrix $(P|Q)$ is invertible. When its inverse, found by Maple, is written in the form $\left(\frac{R}{S}\right)$ we have a 15×27 matrix R which satisfies $RP = I_{15}$ and $RQ = 0$. Regard P as the matrix representing the inclusion of the module M into $E \otimes E \otimes F$. Then R is the matrix, in the same basis, of the projection from $E \otimes E \otimes F$ to M. The module generators Y_M satisfy $Y_M = R Y_{EEF} P$, giving the explicit action of the quantum group on M.

We use the injection and projection further to find the $15^2 \times 15^2$ R–matrix R_{MM}. First include $M \otimes M$ in $(E \otimes E \otimes F) \otimes (E \otimes E \otimes F)$, then construct the R–matrix for $E \otimes E \otimes F$ from the crossing of three strings each coloured with E or F over three others using the various matrices R_{EF} from above, and finally project to $M \otimes M$.

The calculations can be completed in principle from here. Represent the 3–braid in the 2–tangle F by an endomorphism of $M \otimes M \otimes M$, using R_{MM} and its inverse. Then use T_M and the partial trace to close off one string, hence giving the endomorphism F_{MM} of $M \otimes M$ determined by F. A similar calculation gives the endomorphism G_{MM}. The invariant for one of the knots is given by the trace of $T_{MM} F_{MM} G_{MM}$. The other is given by replacing G_{MM} with the conjugate $R^{-1}_{MM} G_{MM} R_{MM}$. Some calculation can be avoided by using $G_{MM} - R^{-1}_{MM} G_{MM} R_{MM}$ in place of G_{MM}, to get the difference of the invariants directly.

A considerable shortcut can be made at this point by concentrating on the effect of F_{MM} and G_{MM} on certain highest weight vectors in $M \otimes M$, rather than considering the whole of the module. A *highest weight vector* v of a module V is a common eigenvector of H_1 and H_2 (or equally K_1 and K_2) which satisfies $X_1^+(v) = X_2^+(v) = 0$. The submodule of V generated by a highest weight vector is irreducible. Its isomorphism type is determined by the eigenvalues of H_1 and H_2, which are non-negative integers. It follows easily from the relations in the quantum group that any module homomorphism $f\colon V \to W$ carries highest weight vectors to highest weight vectors of the same type.

Calculation in Maple determines the linear subspace of $M \otimes M$ which is the common null-space of X_1^+ and X_2^+. This turns out to have dimension 10, spanned by two highest weight vectors of type $(3,1)$, two of type $(1,2)$ and six further highest weight vectors each of a different type. Then the endomorphism F restricts to a linear endomorphism F_ν of the space of highest weight vectors of type ν, for each ν. We remarked earlier that weight spaces of dimension 1 will not contribute to the difference of the invariants on two mutant knots, so we need only calculate the maps F_ν and G_ν for the two 2–dimensional weight spaces $\nu=(3,1)$ and $\nu=(1,2)$. We thus choose two spanning vectors for one of these spaces and follow each of these through the 2–tangle F, taking the tensor product with M and mapping to $M \otimes M \otimes M$ as above (using repeated operations of the 225×225 R–matrix on a vector of length 225×15) before applying the matrix T_M and taking a partial trace to finish in $M \otimes M$. Since the result in each case must be a linear combination of the two chosen weight vectors it is not difficult to find the exact combination. This determines a 2×2 matrix representing F_ν for the weight space of type ν. Similar calculations for the other weight space and for G, along with a quick calculation of the 2×2 matrix representing R_{MM} on each weight type gives enough to find the contribution of each of these weight types to the difference. The final difference comes from multiplying the trace of the 2×2 difference matrix for each type ν by the quantum dimension of the irreducible module of type ν for each of the two types and then adding the results.

Up to the same power of s in each case the contribution from the weight space of type $(3,1)$ was found to be

$$\begin{aligned} t_{31} = {} & (s^8+1)^2(s^2+1)^4(s^4+1)^3(s+1)^{13}(s-1)^{13}s^6(s^2-s+1)(s^2+s+1) \\ & (s^4-s^3+s^2-s+1)(s^4+s^3+s^2+s+1) \\ & (s^6-s^5+s^4-s^3+s^2-s+1)(s^6+s^5+s^4+s^3+s^2+s+1) \\ & (2\,s^{20}+s^{18}+s^{14}-s^{12}+2\,s^8-s^6-1) \\ & (s^{22}-s^{20}+s^{16}-2\,s^{14}+3\,s^{12}+2\,s^{10}-s^8+2\,s^6+2) \\ = {} & (2\,s^{20}+s^{18}+s^{14}-s^{12}+2\,s^8-s^6-1) \\ & (s^{22}-s^{20}+s^{16}-2\,s^{14}+3\,s^{12}+2\,s^{10}-s^8+2\,s^6+2) \\ & \times(s^8-s^{-8})^2(s^7-s^{-7})(s^5-s^{-5})(s^4-s^{-4}) \\ & (s^3-s^{-3})(s^2-s^{-2})(s-s^{-1})^6 s^{49}, \end{aligned}$$

and the contribution from type $(1,2)$ to be

$$
\begin{aligned}
t_{12} = \ & (s^6 - s^5 + s^4 - s^3 + s^2 - s + 1)^2 (s^6 + s^5 + s^4 + s^3 + s^2 + s + 1)^2 \\
& (s^4 - s^2 + 1)(s^8 + 1)^2 (s^4 + 1)^5 (s^2 + 1)^8 \\
& (s^2 + s + 1)(s^2 - s + 1)(s - 1)^{14} (s + 1)^{14} (s^{10} - s^8 + s^4 - s^2 + 1) \\
& (s^{18} - s^{16} - s^{14} + 2\, s^{12} - 2\, s^{10} + 2\, s^6 - 2\, s^4 - s^2 + 1) \\
= \ & (s^{18} - s^{16} - s^{14} + 2\, s^{12} - 2\, s^{10} + 2\, s^6 - 2\, s^4 - s^2 + 1) \\
& (s^{10} - s^8 + s^4 - s^2 + 1) \\
& \times (s^8 - s^{-8})^2 (s^7 - s^{-7})^2 (s^6 - s^{-6})(s^4 - s^{-4})^3 \\
& \times (s^2 - s^{-2})^2 (s - s^{-1})^4 s^{56}.
\end{aligned}
$$

The quantum dimension for the irreducible module of type $(3,1)$, which has Young diagram $\yng(4,1)$, is a product of quantum integers $[6][4] = (s^6 - s^{-6})(s^4 - s^{-4})/(s - s^{-1})^2$. For the module of type $(1,2)$, with Young diagram $\yng(3,2)$, it is $[5][3] = (s^5 - s^{-5})(s^3 - s^{-3})/(s - s^{-1})^2$.

The difference between the $SU(3)_q$ invariants with the module $V_{\yng(3,1)}$ for the Conway and Kinoshita–Teresaka knots is then given, up to a power of $s = e^{h/2}$, by $[5][3]t_{12} + [6][4]t_{31}$. This yields the polynomial quoted earlier.

2 The Kuperberg skein for mutants

Let K and K' be the mutants shown schematically in figure 1. As K and K' are knots, precisely one of F or G must induce the identity permutation on the endpoints by following the strings through the tangle, while the other induces the transposition. We will consider these two cases separately.

In [2] Kuperberg gives a skein-theoretic method for handling the $SU(3)_q$ invariant of a link when coloured by the fundamental module, which he denotes by $<>_{A_2}$. Knot diagrams are extended to allow 3–valent oriented graphs in which any vertex is either a sink or a source. Crossings can be replaced locally in this skein by a linear combination of planar graphs, and any planar circles, 2–gons or 4–gons can be replaced by linear combinations of simpler pieces.

In using skein-based calculations it is helpful when dealing, for example, with satellites to regard the pattern as a diagram in an annulus, and note that it can be replaced by any equivalent linear combination of diagrams in the skein of the annulus. Thus we should consider the Kuperberg skein of the annulus, namely linear combinations of admissibly oriented 3–valent graph diagrams subject to local relations as before. A similar definition can be made for the skein of other surfaces. Notice that the relations ensure that the skein is spanned by oriented graphs lying entirely in the surface, without simple closed curves, 2–gons or 4–gons which bound discs in the surface.

In the case of the annulus this shows that the skein is spanned by unions of oriented simple closed curves parallel to the boundary of the annulus, with orientations in either direction.

When a mutant knot K is made up from two 2–tangles F and G as above then one of F and G, let us suppose G, must be a pure tangle, in the sense that the arcs of G connect the entry point at top left with the exit at bottom left, and top right with bottom right. Then K can be viewed as made from the diagram in the disc P with two holes, shown in figure 4, by embedding the planar surface P so that the two 'ears' are tied around the arcs of G. Turning the diagram in P over along the axis indicated before embedding it in the same way, and reversing all string orientations, will give one of the mutants K' of K. Any satellites of K and K' are related in a similar way, for we can view a satellite of K as constructed by decorating the diagram in P with the required pattern, and then tying the ears of P around G as before. The corresponding satellite of K' is given by turning P over, with the decorated diagram, reversing all strings, and then using the same embedding of P.

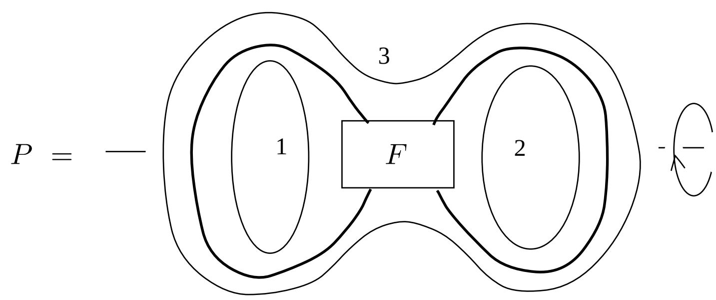


Figure 4

If we could show that the Kuperberg skein of P is spanned by elements which are invariant under turning over and reversing orientation then we could deduce that satellites of mutants such as K and K' would have the same $SU(3)_q$ invariants, by considering the decorated diagram in this skein. A proof for all mutants would need a similar analysis for the skein of the once-punctured torus, to deal with one of the other mutation operations, and the third case would then follow, using a similar argument to [5], where the truth of the corresponding results in the Kauffman bracket skein showed that satellites of mutants have the same $SU(2)_q$ invariants.

We shall now describe a basis for the Kuperberg skein of P, which has some nice symmetry properties, but not enough to give the invariance above. Indeed a diagram coming from a 3–fold parallel with one reversed string will give a linear combination of basis elements in the skein in which all but at most one pair are invariant. (Diagrams from 2–fold parallels of any orientation determine elements of the invariant subspace.)

Theorem 2.1 *The Kuperberg skein of a disc with two holes has a basis of diagrams consisting of the union of simple closed curves parallel to each boundary component and a trivalent graph with a* 2*–gon nearest to each of the three boundary components and* 6*–gons elsewhere.*

Proof Use the skein relations to write any diagram as a linear combination of admissibly oriented trivalent graphs in the surface. We can assume that there are no simple closed curves or 2–gons or 4–gons with null-homotopic boundary. There may be a number of simple closed curves parallel to each of the boundary components. The remaining graph must be connected, otherwise one of its components lies in an annulus inside the surface, and can be reduced further to a linear combination of unions of parallel simple closed curves. Consider the graph as lying in S^2, by filling in the three boundary components of the surface. It dissects S^2 into a number of n–gons, with n even, and $n \geq 6$ except possibly for the three n–gons containing the added discs. Now calculate the Euler characteristic of the resulting sphere S from the dissection by the graph. As vertices are trivalent and each edge now bounds two faces, we can count the Euler characteristic as a sum over the n–gons, in which each vertex contributes $1/3$ and each edge $-1/2$. Therefore each n–gon will contribute $1 - n/6$, so the only positive contribution to $\chi(S)$ can come from 2–gons or 4–gons. These can only arise from the original three boundary components, where the maximum possible total positive contribution is 2 when each boundary component gives a 2–gon. Since the total must be 2 and the only other contributions are negative or zero, we must have three 2–gons forming the original boundary components and 6–gons elsewhere. □

If we start with a 3–parallel of a tangle F inside the planar surface P, with two strands in one direction and one in the other, and write it in the Kuperberg skein we will get a linear combination of graphs as above, each having at most 3 strings around each 'ear'. Some of these will be the union of some simple closed curves around the punctures and trivalent graphs. In figure 5 we show one such trivalent graph which fails to be symmetric under the order 2 operation of turning the surface over (and reversing edge orientations).

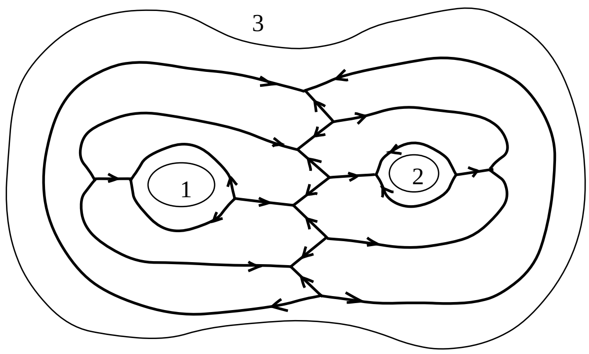


Figure 5

Note however that this graph is symmetric under the operation of order 3 in which the three boundary components are cycled. This is a general feature of the connected trivalent graphs which arise in our construction, as appears from the following description, where we replace P by a 3–punctured sphere.

We call a trivalent graph in the 3–punctured sphere *admissible* if it is oriented so that each vertex is either a sink or a source, and every region not containing a puncture is a hexagon.

Theorem 2.2 *Every admissible graph in the* 3*–punctured sphere is symmetric, up to isotopy avoiding the punctures, under a rotation which cycles the punctures. It can be constructed from the hexagonal tesselation of the plane by choosing an equilateral triangle lattice whose vertices lie at the centres of some of the hexagons and factoring out the translations of the lattice and the rotations of order* 3 *which preserve the lattice.*

Proof Let Γ be the admissible graph. By our Euler characteristic calculations we know that each puncture is contained in a 2–gon. There is a 3–fold branched cover of S^2 by the torus T^2 with three branch points, each cyclic of order 3. The inverse image of Γ in T^2 then consists of hexagonal regions, with three distinguished regions containing the branch points. This inverse image is invariant under the deck transformation of order 3 which leaves each distinguished region invariant. The further inverse image under the regular covering of T^2 by the plane is a tesselation of the plane by hexagons, and the inverse image of the centre of one of the distinguished regions determines a lattice in the plane. We want to show that this is an equilateral triangle lattice, when the hexagonal tesselation is drawn in the usual way. We need only lift the deck transformation to a transformation of the plane keeping the tesselation invariant and fixing one of the lattice points to see that it must lift to a rotation of the tesselation about the centre of a distinguished hexagon. Since the lattice is invariant under this transformation it follows that the lattice must be equilateral. The inverse image of each of the other two branch points will also form an equilateral lattice, invariant under the first rotation, and so their vertices lie in the centres of the triangles; by construction they also lie in the middle of hexagons. Although the equilateral lattice need not lie symmetrically with respect to reflections of the tesselation, as in the example shown below, it does follow that the rotation which permutes the three lattices will also preserve the tesselation. This rotation induces the symmetry of the sphere which cycles the branch points and preserves Γ. □

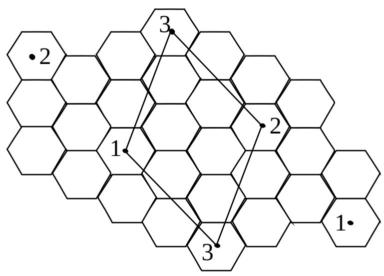


Figure 6

Figure 6 shows such an equilateral triangle lattice superimposed on a hexagon tesselation. The resulting graph in the 3–punctured sphere, whose fundamental domain is indicated, is the graph shown in figure 5 as a non-symmetric skein element in the disk with two holes. The labelling of the puncture points as 1, 2 and 3 corresponds to that of the boundary components. The 3–fold symmetry of the graph in the surface when the boundary components are cycled is evident from this viewpoint.

The Kuperberg skein of the punctured torus does not appear to have such a simple basis. The region around the puncture may be a 2–gon or a 4–gon, giving the following possible combinations: (i) a 2–gon, two 8–gons and 6–gons elsewhere, (ii) a 2–gon, one 10–gon and 6–gons elsewhere, (iii) a 4–gon, one 8–gon and 6–gons elsewhere, (iv) 6–gons only. We did not try to analyse the configurations further, in view of the results of our quantum calculations.

References

[1] **C Kassel**, *Quantum groups*, Graduate Texts in Mathematics, Springer–Verlag (1995)

[2] **G Kuperberg**, *The quantum G_2 invariant*, International J. Math. 5 (1994) 61–85

[3] **H R Morton**, **P R Cromwell**, *Distinguishing mutants by knot polynomials*, J. Knot Theory Ramif. 5 (1996) 225–238

[4] **H R Morton**, **H B Short**, *Calculating the 2–variable polynomial for knots presented as closed braids*, J. Algorithms 11 (1990) 117–131

[5] **H R Morton**, **P T Traczyk**, *The Jones polynomial of satellite links around mutants*, from: "Braids", Joan S Birman and Anatoly Libgober (editors), Contemp. Math. 78, AMS (1988) 587–592

[6] **V G Turaev**, *The Yang–Baxter equation and invariants of links*, Invent. Math. 92 (1988) 527–553

[7] **V G Turaev**, *Quantum invariants of knots and 3–manifolds*, W de Gruyter (1994)

Department of Mathematical Sciences, University of Liverpool
Liverpool L69 3BX, England

Email: h.r.morton@liv.ac.uk

Received: 2 September 1997

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 383–411

Hilbert's 3rd Problem and Invariants of 3–manifolds

Walter D Neumann

Abstract This paper is an expansion of my lecture for David Epstein's birthday, which traced a logical progression from ideas of Euclid on subdividing polygons to some recent research on invariants of hyperbolic 3–manifolds. This "logical progression" makes a good story but distorts history a bit: the ultimate aims of the characters in the story were often far from 3–manifold theory.

We start in section 1 with an exposition of the current state of Hilbert's 3rd problem on scissors congruence for dimension 3. In section 2 we explain the relevance to 3–manifold theory and use this to motivate the Bloch group via a refined "orientation sensitive" version of scissors congruence. This is not the historical motivation for it, which was to study algebraic K–theory of $\mathbb{C}$. Some analogies involved in this "orientation sensitive" scissors congruence are not perfect and motivate a further refinement in section 4. Section 5 ties together various threads and discusses some questions and conjectures.

AMS Classification 57M99; 19E99, 19F27

Keywords Scissors congruence, hyperbolic manifold, Bloch group, dilogarithm, Dehn invariant, Chern–Simons

1 Hilbert's 3rd Problem

It was known to Euclid that two plane polygons of the same area are related by scissors congruence: one can always cut one of them up into polygonal pieces that can be re-assembled to give the other. In the 19th century the analogous result was proved with euclidean geometry replaced by 2–dimensional hyperbolic geometry or 2–dimensional spherical geometry.

The 3rd problem in Hilbert's famous 1900 Congress address [18] posed the analogous question for 3–dimensional euclidean geometry: are two euclidean polytopes of the same volume "scissors congruent," that is, can one be cut into subpolytopes that can be re-assembled to give the other. Hilbert made clear that he expected a negative answer.

One reason for the nineteenth century interest in this question was the interest in a sound foundation for the concepts of area and volume. By "equal area" Euclid *meant* scissors congruent, and the attempt in Euclid's Book XII to provide the same approach for 3–dimensional euclidean volume involved what was called an "exhaustion argument" — essentially a continuity assumption — that mathematicians of the nineteenth century were uncomfortable with (by Hilbert's time mostly for aesthetic reasons).

The negative answer that Hilbert expected to his problem was provided the same year[1] by Max Dehn [7]. Dehn's answer is delighfully simple in modern terms, so we describe it here in full.

Definition 1.1 Consider the free $\mathbb{Z}$–module generated by the set of congruence classes of 3–dimensional polytopes. The *scissors congruence group* $\mathcal{P}(\mathbb{E}^3)$ is the quotient of this module by the relations of scissors congruence. That is, if polytopes $P_1, \dots, P_n$ can be glued along faces to form a polytope P then we set

$$[P] = [P_1] + \cdots + [P_n] \quad \text{in } \mathcal{P}(\mathbb{E}^3).$$

(A *polytope* is a compact domain in $\mathbb{E}^3$ that is bounded by finitely many planar polygonal "faces.")

Volume defines a map

$$\mathrm{vol}\colon \mathcal{P}(\mathbb{E}^3) \to \mathbb{R}$$

and Hilbert's problem asks[2] about injectivity of this map.

Dehn defined a new invariant of scissors congrence, now called the *Dehn invariant*, which can be formulated as a map $\delta\colon \mathcal{P}(\mathbb{E}^3) \to \mathbb{R} \otimes \mathbb{R}/\pi\mathbb{Q}$, where the tensor product is a tensor product of $\mathbb{Z}$–modules (in this case the same as tensor product as $\mathbb{Q}$–vector spaces).

---

[1]In fact, the same answer had been given in 1896 by Bricard, although it was only fully clarified around 1980 that Bricard was answering an equivalent question — see Sah's review 85f:52014 (AMS Mathematical Reviews) of [9] for a concise exposition of this history.

[2]Strictly speaking this is not quite the same question since two polytopes P_1 and P_2 represent the same element of $\mathcal{P}(\mathbb{E}^3)$ if and only if they are *stably scissors congruent* rather than scissors congruent, that is, there exists a polytope Q such that $P_1 + Q$ (disjoint union) is scissors congruent to $P_2 + Q$. But, in fact, stable scissors congruence implies scissors congruence ([47, 48], see [35] for an exposition).

Definition 1.2 If E is an edge of a polytope P we will denote by $\ell(E)$ and $\theta(E)$ the length of E and dihedral angle (in radians) at E. For a polytope P we define the *Dehn invariant* $\delta(P)$ as

$$\delta(P) := \sum_E \ell(E) \otimes \theta(E) \quad \in \quad \mathbb{R} \otimes (\mathbb{R}/\pi\mathbb{Q}), \quad \text{sum over all edges } E \text{ of } P.$$

We then extend this linearly to a homomorphism on $\mathcal{P}(\mathbb{E}^3)$.

It is an easy but instructive exercise to verify that

- δ is well-defined on $\mathcal{P}(\mathbb{E}^3)$, that is, it is compatible with scissors congruence;
- δ and vol are independent on $\mathcal{P}(\mathbb{E}^3)$ in the sense that their kernels generate $\mathcal{P}(\mathbb{E}^3)$ (whence $\mathrm{Im}(\delta|\,\mathrm{Ker}(\mathrm{vol})) = \mathrm{Im}(\delta)$ and $\mathrm{Im}(\mathrm{vol}\,|\,\mathrm{Ker}(\delta)) = \mathbb{R}$);
- the image of δ is uncountable.

In particular, ker(vol) is not just non-trivial, but even uncountable, giving a strong answer to Hilbert's question. To give an explicit example, the regular simplex and cube of equal volume are not scissors congruent: a regular simplex has non-zero Dehn invariant, and the Dehn invariant of a cube is zero.

Of course, this answer to Hilbert's problem is really just a start. It immediately raises other questions:

- Are volume and Dehn invariant sufficient to classify polytopes up to scissors congruence?
- What about other dimensions?
- What about other geometries?

The answer to the first question is "yes." Sydler proved in 1965 that

$$(\mathrm{vol}, \delta)\colon\ \mathcal{P}(\mathbb{E}^3) \to \mathbb{R} \oplus (\mathbb{R} \otimes \mathbb{R}/\pi\mathbb{Q})$$

is injective. Later Jessen [19, 20] simplified his difficult argument somewhat and proved an analogous result for $\mathcal{P}(\mathbb{E}^4)$ and the argument has been further simplified in [13]. Except for these results and the classical results for dimensions ≤ 2 no complete answers are known. In particular, fundamental questions remain open about $\mathcal{P}(\mathbb{H}^3)$ and $\mathcal{P}(\mathbb{S}^3)$.

Note that the definition of Dehn invariant applies with no change to $\mathcal{P}(\mathbb{H}^3)$ and $\mathcal{P}(\mathbb{S}^3)$. The Dehn invariant should be thought of as an "elementary" invariant, since it is defined in terms of 1–dimensional measure. For this reason (and other reasons that will become clear later) we are particularly interested in the kernel of Dehn invariant, so we will abbreviate it: for $\mathbb{X} = \mathbb{E}^3, \mathbb{H}^3, \mathbb{S}^3$

$$\mathcal{D}(\mathbb{X}) := \operatorname{Ker}(\delta\colon\ \mathcal{P}(\mathbb{X}) \to \mathbb{R} \otimes \mathbb{R}/\pi\mathbb{Q})$$

In terms of this notation Sydler's theorem that volume and Dehn invariant classify scissors congruence for $\mathbb{E}^3$ can be reformulated:

$$\operatorname{vol}\colon\ \mathcal{D}(\mathbb{E}^3) \to \mathbb{R} \quad \text{is injective.}$$

It is believed that volume and Dehn invariant classify scissors congruence also for hyperbolic and spherical geometry:

Conjecture 1.3 *Dehn Invariant Sufficiency* $\operatorname{vol}\colon\ \mathcal{D}(\mathbb{H}^3) \to \mathbb{R}$ is injective and $\operatorname{vol}\colon\ \mathcal{D}(\mathbb{S}^3) \to \mathbb{R}$ is injective.

On the other hand $\operatorname{vol}\colon\ \mathcal{D}(\mathbb{E}^3) \to \mathbb{R}$ is also surjective, but this results from the existence of similarity transformations in euclidean space, which do not exist in hyperbolic or spherical geometry. In fact, Dupont [8] proved:

Theorem 1.4 $\operatorname{vol}\colon\ \mathcal{D}(\mathbb{H}^3) \to \mathbb{R}$ *and* $\operatorname{vol}\colon\ \mathcal{D}(\mathbb{S}^3) \to \mathbb{R}$ *have countable image.*

Thus the Dehn invariant sufficiency conjecture would imply:

Conjecture 1.5 *Scissors Congruence Rigidity* $\mathcal{D}(\mathbb{H}^3)$ and $\mathcal{D}(\mathbb{S}^3)$ are countable.

The following collects results of Bökstedt, Brun, Dupont, Parry, Sah and Suslin ([3], [12], [36], [37]).

Theorem 1.6 $\mathcal{P}(\mathbb{H}^3)$ *and* $\mathcal{P}(\mathbb{S}^3)$ *and their subspaces* $\mathcal{D}(\mathbb{H}^3)$ *and* $\mathcal{D}(\mathbb{S}^3)$ *are uniquely divisible groups, so they have the structure of* $\mathbb{Q}$*–vector spaces. As* $\mathbb{Q}$*–vector spaces they have infinite rank. The rigidity conjecture thus says* $\mathcal{D}(\mathbb{H}^3)$ *and* $\mathcal{D}(\mathbb{S}^3)$ *are* $\mathbb{Q}$*–vector spaces of countably infinite rank.*

Corollary 1.7 *The subgroups* $\operatorname{vol}(\mathcal{D}(\mathbb{H}^3))$ *and* $\operatorname{vol}(\mathcal{D}(\mathbb{S}^3))$ *of* $\mathbb{R}$ *are* $\mathbb{Q}$*–vector subspaces of countable dimension.*

1.1 Further comments

Many generalizations of Hilbert's problem have been considered, see eg [35] for an overview. There are generalizations of Dehn invariant to all dimensions and the analog of the Dehn invariant sufficiency conjectures have often been made in greater generality, see eg [35], [12], [16]. The particular Dehn invariant that we are discussing here is a codimension 2 Dehn invariant.

Conjecture 1.3 appears in various other guises in the literature. For example, as we shall see, the $\mathbb{H}^3$ case is equivalent to a conjecture about rational relations among special values of the dilogarithm function which includes as a very special case a conjecture of Milnor [22] about rational linear relations among values of the dilogarithm at roots of unity. Conventional wisdom is that even this very special case is a very difficult conjecture which is unlikely to be resolved in the forseeable future. In fact, Dehn invariant sufficiency would imply the ranks of the vector spaces of volumes in Corollary 1.7 are infinite, but at present these ranks are not even proved to be greater than 1. Even worse: although it is believed that the volumes in question are always irrational, it is not known if a single one of them is!

As we describe later, work of Bloch, Dupont, Parry, Sah, Wagoner, and Suslin connects the Dehn invariant kernels with algebraic K–theory of $\mathbb{C}$, and the above conjectures are then equivalent to standard conjectures in algebraic K–theory. In particular, the scissors congruence rigidity conjectures for H^3 and S^3 are together equivalent to the rigidity conjecture for $K_3(\mathbb{C})$, which can be formulated that $K_3^{ind}(\mathbb{C})$ (indecomposable part of Quillen's K_3) is countable. This conjecture is probably much easier than the Dehn invariant sufficiency conjecture.

The conjecture about rational relations among special values of the dilogarithm has been broadly generalized to polylogarithms of all degrees by Zagier (section 10 of [46]). The connections between scissors congruence and algebraic K–theory have been generalised to higher dimensions, in part conjecturally, by Goncharov [16].

We will return to some of these issues later. We also refer the reader to the very attractive exposition in [14] of these connnections in dimension 3.

I would like to acknowledge the support of the Australian Research Council for this research, as well as the the Max–Planck–Institut für Mathematik in Bonn, where much of this paper was written.

2 Hyperbolic 3–manifolds

Thurston's geometrization conjecture, much of which is proven to be true, asserts that, up to a certain kind of canonical decomposition, 3–manifolds have geometric structures. These geometric structures belong to eight different geometries, but seven of these lead to manifolds that are describable in terms of surface topology and are very easily classified. The eighth geometry is hyperbolic geometry $\mathbb{H}^3$. Thus if one accepts the geometrization conjecture then the central issue in understanding 3–manifolds is to understand hyperbolic 3–manifolds.

Suppose therefore that $M = \mathbb{H}^3/\Gamma$ is a hyperbolic 3–manifold. We will always assume M is oriented and for the moment we will also assume M is compact, though we will be able to relax this assumption later. We can subdivide M into small geodesic tetrahedra, and then the sum of these tetrahedra represents a class $\beta_0(M) \in \mathcal{P}(\mathbb{H}^3)$ which is an invariant of M. We call this the *scissors congruence class of* M.

Note that when we apply the Dehn invariant to $\beta_0(M)$ the contributions coming from each edge E of the triangulation sum to $\ell(E) \otimes 2\pi$ which is zero in $\mathbb{R} \otimes \mathbb{R}/\pi\mathbb{Q}$. Thus

Proposition 2.1 *The scissors congruence class* $\beta_0(M)$ *lies in* $\mathcal{D}(\mathbb{H}^3)$. □

How useful is this invariant of M? We can immediately see that it is non-trivial, since at least it detects volume of M:

$$\mathrm{vol}(M) = \mathrm{vol}(\beta_0(M)).$$

Now it was suggested by Thurston in [42] that the volume of hyperbolic 3–manifolds should have some close relationship with another geometric invariant, the Chern–Simons invariant $\mathrm{CS}(M)$. A precise analytic relationship was then conjectured in [30] and proved in [44] (a new proof follows from the work discussed here, see [24]). We will not discuss the definition of this invariant here (it is an invariant of compact riemmanian manifolds, see [6, 5], which was extended also to non-compact finite volume hyperbolic 3–manifolds by Meyerhoff [21]). It suffices for the present discussion to know that for a finite volume hyperbolic 3–manifold M the Chern–Simons invariant lies in $\mathbb{R}/\pi^2\mathbb{Z}$. Moreover, the combination $\mathrm{vol}(M) + i\,\mathrm{CS}(M) \in \mathbb{C}/\pi^2\mathbb{Z}$ turns out to have good analytic properties and is therefore a natural "complexification" of volume for hyperbolic manifolds. Given this intimate relationship between volume and Chern–Simons invariant, it becomes natural to ask if $CS(M)$ is also detected by $\beta_0(M)$.

The answer, unfortunately, is an easy "no." The point is that $CS(M)$ is an orientation sensitive invariant: $CS(-M) = -CS(M)$, where $-M$ means M with reversed orientation. But, as Gerling pointed out in a letter to Gauss on 15 April 1844: scissors congruence cannot see orientation because any polytope is scissors congruent to its mirror image[3]. Thus $\beta_0(-M) = \beta_0(M)$ and there is no hope of $CS(M)$ being computable from $\beta_0(M)$. This raises the question:

Question 2.2 Is there some way to repair the orientation insensitivity of scissors congruence and thus capture Chern–Simons invariant?

The answer to this question is "yes" and lies in the so called "Bloch group," which was invented for entirely different purposes by Bloch (it was put in final form by Wigner and Suslin). To explain this we start with a result of Dupont and Sah [12] about ideal polytopes — hyperbolic polytopes whose vertices are at infinity (such polytopes exist in hyperbolic geometry, and still have finite volume).

Proposition 2.3 *Ideal hyperbolic tetrahedra represent elements in* $\mathcal{P}(\mathbb{H}^3)$ *and, moreover,* $\mathcal{P}(\mathbb{H}^3)$ *is generated by ideal tetrahedra.*

To help understand this proposition observe that if $ABCD$ is a non-ideal tetrahedron and E is the ideal point at which the extension of edge AD meets infinity then $ABCD$ can be represented as the difference of the two tetrahedra $ABCE$ and $DBCE$, each of which have one ideal vertex. We have thus, in effect, "pushed" one vertex off to infinity. In the same way one can push a second and third vertex off to infinity, ... and the fourth, but this is rather harder. Anyway, we will accept this proposition and discuss its consequence for scissors congruence.

The first consequence is a great gain in convenience: a non-ideal tetrahedron needs six real parameters satisfying complicated inequalities to characterise it up to congruence while an ideal tetrahedron can be neatly characterised by a single complex parameter in the upper half plane.

We shall denote the standard compactification of $\mathbb{H}^3$ by $\overline{\mathbb{H}}^3 = \mathbb{H}^3 \cup \mathbb{CP}^1$. An ideal simplex Δ with vertices $z_1, z_2, z_3, z_4 \in \mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$ is determined up to congruence by the cross-ratio

$$z = [z_1 : z_2 : z_3 : z_4] = \frac{(z_3 - z_2)(z_4 - z_1)}{(z_3 - z_1)(z_4 - z_2)}.$$

[3]Gauss, Werke, Vol. 10, p. 242; the argument for a tetrahedron is to barycentrically subdivide by dropping perpendiculars from the circumcenter to each of the faces; the resulting 24 tetrahedra occur in 12 mirror image pairs.

Permuting the vertices by an even (ie orientation preserving) permutation replaces z by one of

$$z, \quad z' = \frac{1}{1-z}, \quad \text{or} \quad z'' = 1 - \frac{1}{z}.$$

The parameter z lies in the upper half plane of $\mathbb{C}$ if the orientation induced by the given ordering of the vertices agrees with the orientation of $\mathbb{H}^3$.

There is another way of describing the cross-ratio parameter $z = [z_1 : z_2 : z_3 : z_4]$ of a simplex. The group of orientation preserving isometries of $\mathbb{H}^3$ fixing the points z_1 and z_2 is isomorphic to the multiplicative group $\mathbb{C}^*$ of nonzero complex numbers. The element of this $\mathbb{C}^*$ that takes z_4 to z_3 is z. Thus the cross-ratio parameter z is associated with the edge z_1z_2 of the simplex. The parameter associated in this way with the other two edges z_1z_4 and z_1z_3 out of z_1 are z' and z'' respectively, while the edges z_3z_4, z_2z_3, and z_2z_4 have the same parameters z, z', and z'' as their opposite edges. See figure 1. This description makes clear that the dihedral angles at the edges of the simplex are $\arg(z)$, $\arg(z')$, $\arg(z'')$ respectively, with opposite edges having the same angle.

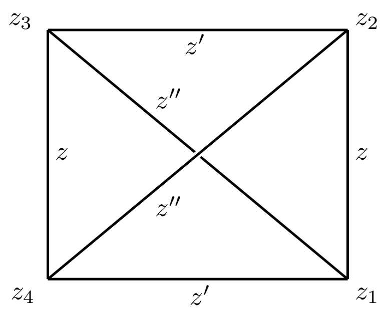


Figure 1

Now suppose we have five points $z_0, z_1, z_2, z_3, z_4 \in \mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$. Any four-tuple of these five points spans an ideal simplex, and the convex hull of these five points decomposes in two ways into such simplices, once into two of them and once into three of them. We thus get a scissors congruence relation equating the two simplices with the three simplices. It is often called the "five-term relation." To express it in terms of the cross-ratio parameters it is convenient first to make an orientation convention.

We allow simplices whose vertex ordering does not agree with the orientation of $\mathbb{H}^3$ (so the cross-ratio parameter is in the lower complex half-plane) but make the convention that this represents the negative element in scissors congruence.

An odd permutation of the vertices of a simplex replaces the cross-ratio parameter z by

$$\frac{1}{z}, \quad \frac{z}{z-1}, \quad \text{or} \quad 1-z,$$

so if we denote by $[z]$ the element in $\mathcal{P}(\mathbb{H}^3)$ represented by an ideal simplex with parameter z, then our orientation rules say:

$$[z] = [1-\frac{1}{z}] = [\frac{1}{1-z}] = -[\frac{1}{z}] = -[\frac{z-1}{z}] = -[1-z]. \tag{1}$$

These orientation rules make the five-term scissors congruence relation described above particularly easy to state:

$$\sum_{i=0}^{4}(-1)^i[z_0 : \ldots : \hat{z}_i : \ldots : z_4] = 0.$$

The cross-ratio parameters occuring in this formula can be expressed in terms of the first two as

$$[z_1:z_2:z_3:z_4] =: x \qquad [z_0:z_2:z_3:z_4] =: y$$
$$[z_0:z_1:z_3:z_4] = \frac{y}{x} \quad [z_0:z_1:z_2:z_4] = \frac{1-x^{-1}}{1-y^{-1}} \quad [z_0:z_1:z_2:z_3] = \frac{1-x}{1-y}$$

so the five-term relation can also be written:

$$[x]-[y]+[\frac{y}{x}]-[\frac{1-x^{-1}}{1-y^{-1}}]+[\frac{1-x}{1-y}] = 0. \tag{2}$$

We lose nothing if we also allow degenerate ideal simplices whose vertices lie in one plane so the parameter z is real (we always require that the vertices are distinct, so the parameter is in $\mathbb{R}-\{0,1\}$), since the five-term relation can be used to express such a "flat" simplex in terms of non-flat ones, and one readily checks no additional relations result. Thus we may take the parameter z of an ideal simplex to lie in $\mathbb{C}-\{0,1\}$ and every such z corresponds to an ideal simplex.

One can show that relations (1) follow from the five-term relation (2), so we consider the quotient

$$\mathcal{P}(\mathbb{C}) := \mathbb{Z}\langle\mathbb{C}-\{0,1\}\rangle/(\text{five-term relations (2)})$$

of the free $\mathbb{Z}$–module on $\mathbb{C}-\{0,1\}$. Proposition 2.3 can be restated that there is a natural surjection $\mathcal{P}(\mathbb{C}) \to \mathcal{P}(\mathbb{H}^3)$. In fact Dupont and Sah (loc. cit.) prove:

Theorem 2.4 *The scissors congruence group $\mathcal{P}(\mathbb{H}^3)$ is the quotient of $\mathcal{P}(\mathbb{C})$ by the relations $[z] = -[\overline{z}]$ which identify each ideal simplex with its mirror image*[4].

Thus $\mathcal{P}(\mathbb{C})$ is a candidate for the orientation sensitive scissors congruence group that we were seeking. Indeed, it turns out to do (almost) exactly what we want.

The analog of the Dehn invariant has a particularly elegant expression in these terms. First note that the above theorem expresses $\mathcal{P}(\mathbb{H}^3)$ as the "imaginary part" $\mathcal{P}(\mathbb{C})^-$ (negative co-eigenspace under conjugation[5]) of $\mathcal{P}(\mathbb{C})$.

Proposition/Definition 2.5 *The Dehn invariant $\delta\colon\ \mathcal{P}(\mathbb{H}^3) \to \mathbb{R}\otimes\mathbb{R}/\pi\mathbb{Q}$ is twice the "imaginary part" of the map*

$$\delta_{\mathbb{C}}\colon\ \mathcal{P}(\mathbb{C}) \to \mathbb{C}^* \wedge \mathbb{C}^*, \quad [z] \mapsto (1-z)\wedge z$$

so we shall call this map the "complex Dehn invariant." We denote the kernel of complex Dehn invariant

$$\mathcal{B}(\mathbb{C}) := \operatorname{Ker}(\delta_{\mathbb{C}}),$$

and call it the "Bloch group of $\mathbb{C}$."

(We shall explain this proposition further in an appendix to this section.)

A hyperbolic 3–manifold M now has an "orientation sensitive scissors congruence class" which lies in this Bloch group and captures both volume and Chern–Simons invariant of M. Namely, there is a map

$$\rho\colon\ \mathcal{B}(\mathbb{C}) \to \mathbb{C}/\pi^2\mathbb{Q}$$

introduced by Bloch and Wigner called the *Bloch regulator map*, whose imaginary part is the volume map on $\mathcal{B}(\mathbb{C})$, and one has:

Theorem 2.6 ([29], [8]) *Let M be a complete oriented hyperbolic 3–manifold of finite volume. Then there is a natural class $\beta(M) \in \mathcal{B}(\mathbb{C})$ associated with M and $\rho(\beta(M)) = \frac{1}{i}(\operatorname{vol}(M) + i\operatorname{CS}(M))$.*

This theorem answers Question 2.2. But there are still two aesthetic problems:

[4]The minus sign in this relation comes from the orientation convention described earlier.

[5]$\mathcal{P}(\mathbb{C})$ turns out to be a $\mathbb{Q}$–vector space and is therefore the sum of its ± 1 eigenspaces, so "co-eigenspace" is the same as "eigenspace."

- The Bloch regulator ρ plays the rôle for orientation sensitive scissors congruence that volume plays for ordinary scissors congruence. But vol is defined on the whole scissors congruence group $\mathcal{P}(\mathbb{H}^3)$ while ρ is only defined on the kernel $\mathcal{B}(\mathbb{C})$ of complex Dehn invariant.
- The Chern–Simons invariant $\mathrm{CS}(M)$ is an invariant in $\mathbb{R}/\pi^2\mathbb{Z}$ but the invariant $\rho(\beta(M))$ only computes it in $\mathbb{R}/\pi^2\mathbb{Q}$.

We resolve both these problems in section 4.

We describe the Bloch regulator map ρ later. It would be a little messy to describe at present, although its imaginary part (volume) has a very nice description in terms of ideal simplices. Indeed, the volume of an ideal simplex with parameter z is $D_2(z)$, where D_2 is the so called "Bloch–Wigner dilogarithm function" given by:

$$D_2(z) = \mathrm{Im}\,\ln_2(z) + \log|z|\arg(1-z), \quad z \in \mathbb{C} - \{0,1\}$$

and $\ln_2(z)$ is the classical dilogarithm function. It follows that $D_2(z)$ satisfies a functional equation corresponding to the five-term relation (see below).

2.1 Further comments

To worry about the second "aesthetic problem" above could be considered rather greedy. After all, $CS(M)$ takes values in $\mathbb{R}/\pi^2\mathbb{Z}$ which is the direct sum of $\mathbb{Q}/\pi^2\mathbb{Z}$ and uncountably many copies of $\mathbb{Q}$, and we have only lost part of the former summand. However, it is not even known if the Chern–Simons invariant takes *any* non-zero values[6] in $\mathbb{R}/\pi^2\mathbb{Q}$. As we shall see, this would be implied by the sufficiency of Dehn invariant for $\mathbb{S}^3$ (Conjecture 1.3).

The analogous conjecture in our current situation is:

Conjecture 2.7 *Complex Dehn Invariant Sufficiency* $\rho\colon \mathcal{B}(\mathbb{C}) \to \mathbb{C}/\pi^2\mathbb{Q}$ is injective.

Again, the following is known by work of Bloch:

Theorem 2.8 $\rho\colon \mathcal{B}(\mathbb{C}) \to \mathbb{C}/\pi^2\mathbb{Q}$ *has countable image.*

Thus the complex Dehn invariant sufficiency conjecture would imply:

[6] According to J Dupont, Jim Simons deserted mathematics in part because he could not resolve this issue!

Conjecture 2.9 *Bloch Rigidity* $\mathcal{B}(\mathbb{C})$ is countable.

Theorem 2.10 ([37, 38]) *$\mathcal{P}(\mathbb{C})$ and its subgroup $\mathcal{B}(\mathbb{C})$ are uniquely divisible groups, so they have the structure of $\mathbb{Q}$–vector spaces. As $\mathbb{Q}$–vector spaces they have infinite rank.*

Note that the Bloch group $\mathcal{B}(\mathbb{C})$ is defined purely algebraically in terms of $\mathbb{C}$, so we can define a Bloch group $\mathcal{B}(k)$ analogously[7] for any field k. This group $\mathcal{B}(k)$ is uniquely divisible whenever k contains an algebraically closed field.

It is not hard to see that the rigidity conjecture 2.9 is equivalent to the conjecture that $\mathcal{B}(\overline{\mathbb{Q}}) \to \mathcal{B}(\mathbb{C})$ is an isomorphism (here $\overline{\mathbb{Q}}$ is the field of algebraic numbers; it is known that $\mathcal{B}(\overline{\mathbb{Q}}) \to \mathcal{B}(\mathbb{C})$ is injective). Suslin has conjectured more generally that $\mathcal{B}(k) \to \mathcal{B}(K)$ is an isomorphism if k is the algebraic closure of the prime field in K. Conjecture 2.7 has been made in greater generality by Ramakrishnan [32] in the context of algebraic K–theory.

Conjectures 2.7 and 2.9 are in fact equivalent to the Dehn invariant sufficiency and rigidity conjectures 1.3 and 1.5 respectively for $\mathbb{H}^3$ and $\mathbb{S}^3$ together. This is because of the following theorem which connects the various Dehn kernels. It also describes the connections with algebraic K–theory and homology of the lie group $\mathrm{SL}(2,\mathbb{C})$ considered as a discrete group. It collates results of of Bloch, Bökstedt, Brun, Dupont, Parry and Sah and Wigner (see [3] and [11]).

Theorem 2.11 *There is a natural exact sequence*

$$0 \to \mathbb{Q}/\mathbb{Z} \to H_3(\mathrm{SL}(2,\mathbb{C})) \to \mathcal{B}(\mathbb{C}) \to 0.$$

Moreover there are natural isomorphisms:

$H_3(\mathrm{SL}(2,\mathbb{C})) \cong K_3^{ind}(\mathbb{C})$,

$H_3(\mathrm{SL}(2,\mathbb{C}))^- \cong \mathcal{B}(\mathbb{C})^- \cong \mathcal{D}(\mathbb{H}^3)$,

$H_3(\mathrm{SL}(2,\mathbb{C}))^+ \cong \mathcal{D}(\mathbb{S}^3)/\mathbb{Z}$ *and* $\mathcal{B}(\mathbb{C})^+ \cong \mathcal{D}(S^3)/\mathbb{Q}$,

where $\mathbb{Z} \subset \mathcal{D}(S^3)$ is generated by the class of the 3–sphere and $\mathbb{Q} \subset \mathcal{D}(S^3)$ is the subgroup generated by suspensions of triangles in $\mathbb{S}^2$ with rational angles.

The Cheeger–Simons map $c_2\colon H_3(\mathrm{SL}(2,\mathbb{C})) \to \mathbb{C}/4\pi^2\mathbb{Z}$ of [5] induces on the one hand the Bloch regulator map $\rho\colon \mathcal{B}(\mathbb{C}) \to \mathbb{C}/\pi^2\mathbb{Q}$ and on the other hand its real and imaginary parts correspond to the volume maps on $\mathcal{D}(\mathbb{S}^3)/\mathbb{Z}$ and $\mathcal{D}(\mathbb{H}^3)$ via the above isomorphisms.

[7] Definitions of $\mathcal{B}(k)$ in the literature vary in ways that can mildly affect its torsion if k is not algebraically closed.

The isomorphisms of the theorem are proved via isomorphisms $H_3(\mathrm{SL}(2,\mathbb{C}))^- \cong H_3(\mathrm{SL}(2,R))$ and $H_3(\mathrm{SL}(2,\mathbb{C}))^+ \cong H_3(\mathrm{SU}(2))$. We have described the geometry of the isomorphism $\mathcal{B}(\mathbb{C})^- \cong \mathcal{D}(\mathbb{H}^3)$ in Theorem 2.4. The geometry of the isomorphism $\mathcal{B}(\mathbb{C})^+ \cong \mathcal{D}(S^3)/\mathbb{Q}$ remains rather mysterious.

The exact sequence and first isomorphism in the above theorem are valid for any algebraically closed field of characteristic 0. Thus Conjecture 2.9 is also equivalent to each of the four:

- Is $K_3^{ind}(\overline{\mathbb{Q}}) \to K_3^{ind}(\mathbb{C})$ an isomorphism? Is $K_3^{ind}(\mathbb{C})$ countable?
- Is $H_3(\mathrm{SL}(2,\overline{\mathbb{Q}})) \to H_3(\mathrm{SL}(2,\mathbb{C}))$ an isomorphism? Is $H_3(\mathrm{SL}(2,\mathbb{C}))$ countable?

The fact that volume of an ideal simplex is given by the Bloch–Wigner dilogarithm function $D_2(z)$ clarifies why the $\mathbb{H}^3$ Dehn invariant sufficiency conjecture 1.3 is equivalent to a statement about rational relations among special values of the dilogarithm function. Don Zagier's conjecture about such rational relations, mentioned earlier, is that any rational linear relation among values of D_2 at algebraic arguments must be a consequence of the relations $D_2(z) = D_2(\overline{z})$ and the five-term functional relation for D_2:

$$D_2(x) - D_2(y) + D_2(\frac{y}{x}) - D_2(\frac{1-x^{-1}}{1-y^{-1}}) + D_2(\frac{1-x}{1-y}) = 0.$$

Differently expressed, he conjectures that the volume map is injective on $\mathcal{P}(\overline{\mathbb{Q}})^-$. If one assumes the scissors congruence rigidity conjecture for $\mathbb{H}^3$ (that $\mathcal{B}(\overline{\mathbb{Q}})^- \cong \mathcal{B}(\mathbb{C})^-$) then the Dehn invariant sufficiency conjecture for $\mathbb{H}^3$ is just that D_2 is injective on the subgroup $\mathcal{B}(\overline{\mathbb{Q}})^- \subset \mathcal{P}(\overline{\mathbb{Q}})^-$, so under this assumption Zagier's conjecture is much stronger. Milnor's conjecture, mentioned earlier, can be formulated that the values of $D_2(\xi)$, as ξ runs through the primitive n-th roots of unity in the upper half plane, are rationally independent for any n. This is equivalent to injectivity modulo torsion of the volume map D_2 on $\mathcal{B}(k_n)$ for the cyclotomic field $k_n = \mathbb{Q}(e^{2\pi i/n})$. For this field $\mathcal{B}(k_n)^- = \mathcal{B}(k_n)$ modulo torsion. This is of finite rank but $\mathcal{P}(k_n)^-$ is of infinite rank, so even when restricted to k_n Zagier's conjecture is much stronger than Milnor's. Zagier himself has expressed doubt that Milnor's conjecture can be resolved in the forseeable future.

Conjecture 2.7 can be similarly formulated as a statement about special values of a different dilogarithm function, the "Rogers dilogarithm," which we will define later.

2.2 Appendix to section 2: Dehn invariant of ideal polytopes

To define the Dehn invariant of an ideal polytope we first cut off each ideal vertex by a horoball based at that vertex. We then have a polytope with some horospherical faces but with all edges finite. We now compute the Dehn invariant using the geodesic edges of this truncated polytope (that is, only the edges that come from the original polytope and not those that bound horospherical faces). This is well defined in that it does not depend on the sizes of the horoballs we used to truncate our polytope. (To see this, note that dihedral angles of the edges incident on an ideal vertex sum to a multiple of π, since they are the angles of the horospherical face created by truncation, which is an euclidean polygon. Changing the size of the horoball used to truncate these edges thus changes the Dehn invariant by a multiple of something of the form $l\otimes\pi$, which is zero in $\mathbb{R}\otimes\mathbb{R}/\pi\mathbb{Q}$.)

Now consider the ideal tetrahedron $\Delta(z)$ with parameter z. We may position its vertices at $0,1,\infty,z$. There is a Klein 4–group of symmetries of this tetrahedron and it is easily verified that it takes the following horoballs to each other:

- At ∞ the horoball $\{(w,t)\in\mathbb{C}\times\mathbb{R}^+ | t\ge a\}$;
- at 0 the horoball of euclidean diameter $|z|/a$;
- at 1 the horoball of euclidean diameter $|1-z|/a$;
- at z the horoball of euclidean diameter $|z(z-1)|/a$.

After truncation, the vertical edges thus have lengths $2\log a-\log|z|$, $2\log a-\log|1-z|$, and $2\log a-\log|z(z-1)|$ respectively, and we have earlier said that their angles are $\arg(z), \arg(1/(1-z)), \arg((z-1)/z)$ respectively. Thus, adding contributions, we find that these three edges contribute $\log|1-z|\otimes\arg(z)-\log|z|\otimes\arg(1-z)$ to the Dehn invariant. By symmetry the other three edges contribute the same, so the Dehn invariant is:

$$\delta(\Delta(z)) = 2\big(\log|1-z|\otimes\arg(z)-\log|z|\otimes\arg(1-z)\big)\in\mathbb{R}\otimes\mathbb{R}/\pi\mathbb{Q}.$$

Proof of Proposition 2.5 To understand the "imaginary part" of $(1-z)\wedge z\in\mathbb{C}^*\wedge\mathbb{C}^*$ we use the isomorphism

$$\mathbb{C}^*\to\mathbb{R}\oplus\mathbb{R}/2\pi\mathbb{Z},\quad z\mapsto\log|z|\oplus\arg z,$$

to represent

$$\begin{aligned}\mathbb{C}^*\wedge\mathbb{C}^* &= (\mathbb{R}\oplus\mathbb{R}/2\pi\mathbb{Z})\wedge(\mathbb{R}\oplus\mathbb{R}/2\pi\mathbb{Z})\\ &= (\mathbb{R}\wedge\mathbb{R})\oplus(\mathbb{R}/2\pi\mathbb{Z}\wedge\mathbb{R}/2\pi\mathbb{Z})\quad\oplus\quad(\mathbb{R}\otimes\mathbb{R}/2\pi\mathbb{Z})\\ &= (\mathbb{R}\wedge\mathbb{R})\oplus(\mathbb{R}/\pi\mathbb{Q}\wedge\mathbb{R}/\pi\mathbb{Q})\quad\oplus\quad(\mathbb{R}\otimes\mathbb{R}/\pi\mathbb{Q}),\end{aligned}$$

(the equality on the third line is because tensoring over $\mathbb{Z}$ with a divisible group is effectively the same as tensoring over $\mathbb{Q}$). Under this isomorphism we have

$$(1-z)\wedge z = \big(\log|1-z| \wedge \log|z| \oplus \arg(1-z)\wedge \arg z\big)$$
$$\oplus \quad \big(\log|1-z|\otimes \arg z - \log|z|\otimes \arg(1-z)\big),$$

confirming the Proposition 2.5. □

3 Computing $\beta(M)$

The scissors congruence invariant $\beta(M)$ turns out to be a very computable invariant. To explain this we must first describe the "invariant trace field" or "field of definition" of a hyperbolic 3–manifold. Suppose therefore that $M = \mathbb{H}^3/\Gamma$ is a hyperbolic manifold, so Γ is a discrete subgroup of the orientation preserving isometry group $\mathrm{PSL}(2,\mathbb{C})$ of $\mathbb{H}^3$.

Definition 3.1 [33] The *invariant trace field* of M is the subfield of $\mathbb{C}$ generated over $\mathbb{Q}$ by the squares of traces of elements of Γ. We will denote it $k(M)$ or $k(\Gamma)$.

This field $k(M)$ is an algebraic number field (finite extension of $\mathbb{Q}$) and is a commensurability invariant, that is, it is unchanged on passing to finite covers of M (finite index subgroups of Γ). Moreover, if M is an arithmetic hyperbolic 3–manifold (that is, Γ is an arithmetic group), then $k(M)$ is the field of definition of this arithmetic group in the usual sense. See [33, 26].

Now if k is an algebraic number field then $\mathcal{B}(k)$ is isomorphic to $\mathbb{Z}^{r_2}\oplus(\text{torsion})$, where r_2 is the number of conjugate pairs of complex embeddings $k\to\mathbb{C}$ of k. Indeed, if these complex embeddings are $\sigma_1,\dots,\sigma_{r_2}$ then a reinterpretation of a theorem of Borel [4] about $K_3(\mathbb{C})$ says:

Theorem 3.2 *The "Borel regulator map"*

$$\mathcal{B}(k)\to\mathbb{R}^{r_2}$$

induced on generators of $\mathcal{P}(k)$ *by* $[z]\mapsto(\mathrm{vol}[\sigma_1(z)],\dots,\mathrm{vol}[\sigma_{r_2}(z)])$ *maps* $\mathcal{B}(k)/(\textit{torsion})$ *isomorphically onto a full lattice in* $\mathbb{R}^{r_2}$.

A corollary of this theorem is that an embedding $\sigma\colon k\to\mathbb{C}$ induces an embedding $\mathcal{B}(k)\otimes\mathbb{Q}\to\mathcal{B}(C)\otimes\mathbb{Q}$. (This is because the theorem implies that an element of $\mathcal{B}(k)$ is determined modulo torsion by the set of volumes of its Galois conjugates, which are invariants defined on $\mathcal{B}(\mathbb{C})$.) Moreover, since $\mathcal{B}(\mathbb{C})$ is a $\mathbb{Q}$–vector space, $\mathcal{B}(\mathbb{C})\otimes\mathbb{Q}=\mathcal{B}(\mathbb{C})$.

Now if M is a hyperbolic manifold then its invariant trace field $k(M)$ comes embedded in $\mathbb{C}$ so we get an explicit embedding $\mathcal{B}(k(M))\otimes\mathbb{Q}\to\mathcal{B}(\mathbb{C})$ whose image, which is isomorphic to $\mathbb{Q}^{r_2}$, we denote by $\mathcal{B}(k(M))_{\mathbb{Q}}$.

Theorem 3.3 ([28, 29]) *The element $\beta(M)$ lies in the subspace $\mathcal{B}(k(M))_{\mathbb{Q}}\subset\mathcal{B}(\mathbb{C})$.*

In fact Neumann and Yang show that $\beta(M)$ is well defined in $\mathcal{B}(K)$ for some explicit multi-quadratic field extension K of $k(M)$, which implies that $2^c\beta(M)$ is actually well defined in $\mathcal{B}(k(M))$ for some c. Moreover, one can always take $c=0$ if M is non-compact, but we do not know if one can for compact M.

In view of this theorem we see that the following data effectively determines $\beta(M)$ modulo torsion:

- The invariant trace field $k(M)$.
- The image of $\beta(M)$ in $\mathbb{R}^{r_2}$ under the Borel regulator map of Theorem 3.2.

To compute $\beta(M)$ we need a collection of ideal simplices that triangulates M in some fashion. If M is compact, this clearly cannot be a triangulation in the usual sense. In [29] it is shown that one can use any "degree one ideal triangulation" to compute $\beta(M)$. This means a finite complex formed of ideal hyperbolic simplices plus a map of it to M that takes each ideal simplex locally isometrically to M and is degree one almost everywhere. These always exist (see [29] for a discussion). Special degree one ideal triangulations have been used extensively in practice, eg in Jeff Weeks' program Snappea [43] for computing with hyperbolic 3–manifolds. Oliver Goodman has written a program Snap [17] (building on Snappea) which finds degree one ideal triangulations using exact arithmetic in number fields and computes the invariant trace field and high precision values for the Borel regulator on $\beta(M)$.

Such calculations can provide numerical evidence for the complex Dehn invariant sufficiency conjecture. Here is a typical result of such calculations.

3.1 Examples

To ensure that the Bloch group has rank >1 we want a field with at least two complex embeddings. One of the simplest is the (unique) quartic field over $\mathbb{Q}$ of discriminant 257. This is the field $k = \mathbb{Q}(x)/(f(x))$ with $f(x) = x^4+x^2-x+1$. This polynomial is irreducible with roots $\tau_1^\pm = 0.54742\ldots \pm 0.58565\ldots i$ and $\tau_2^\pm = -0.54742\ldots \pm 1.12087\ldots i$. The field k thus has two complex embeddings σ_1, σ_2 up to complex conjugation, one with image $\sigma_1(k) = \mathbb{Q}(\tau_1^-)$ and one with image $\sigma_2(k) = \mathbb{Q}(\tau_2^-)$. The Bloch group $\mathcal{B}(k)$ is thus isomorphic to $\mathbb{Z}^2$ modulo torsion.

This field occurs as the invariant trace field for two different hyperbolic knot complements in the standard knot tables up to 8 crossings, the 6–crossing knot 6_1 and the 7–crossing knot 7_7, but the embeddings in $\mathbb{C}$ are different. For 6_1 one gets $\sigma_1(k)$ and for 7_7 one gets $\sigma_2(k)$. The scissors congruence classes are

$$\begin{aligned}\beta(6_1) &=: \beta_1 = 2[\frac{1}{2}(1-\tau^2-\tau^3)] + [1-\tau] + [\frac{1}{2}(1-\tau^2+\tau^3)] \in \mathcal{B}(k)\\ \beta(7_7) &=: \beta_2 = 4[2-\tau-\tau^3] + 4[\tau+\tau^2+\tau^3] \in \mathcal{B}(k)\end{aligned}$$

where τ is the class of x in $k = \mathbb{Q}(x)/(x^4+x^2-x+1)$. These map under the Borel regulator $\mathcal{B}(k) \to \mathbb{R}^2$ (with respect to the embeddings σ_1, σ_2) to

$$\begin{aligned}6_1:&\quad (3.16396322888314398399101 4716..,-1.41510489726556334068950 8587..)\\ 7_7:&\quad (-1.39708816556888143946145 3224..,7.64337517235995547822184 4448..)\end{aligned}$$

In particular, the volumes of these knot complements are 3.1639632288831439.. and 7.6433751723599554.. respectively

Snap has access to a large database of small volume compact manifolds. Searching this database for manifolds whose volumes are small rational linear combinations of $\mathrm{vol}(\sigma_1(\beta_1)) = 3.1639632..$ and $\mathrm{vol}(\sigma_1(\beta_2)) = -1.3970881..$ yielded just eight examples, three with volume 3.16396322888314.., four with volume 4.396672801932495.. and one with volume 5.629382374981847.. . The complex Dehn invariant sufficiency conjecture predicts (under the assumption that the rational dependencies found are exact) that these should all have invariant trace field containing $\sigma_1(k)$.

Checking with Snap confirms that their invariant trace fields equal $\sigma_1(k)$ and their scissors congruence classes in $\mathcal{B}(k) \otimes \mathbb{Q}$ (computed numerically using the Borel regulator) are β_1, $(3/2)\beta_1 + (1/2)\beta_2$, and $2\beta_1 + \beta_2$ respectively.

4 Extended Bloch group

In section 2 we saw that $\mathcal{P}(\mathbb{C})$ and $\mathcal{B}(C)$ play a role of "orientation sensitive" scissors congruence group and kernel of Dehn invariant respectively, and that the analog of the volume map is then the Borel regulator ρ. But, as we described there, this analogy suffers because ρ is defined on the Dehn kernel $\mathcal{B}(\mathbb{C})$ rather than on the whole of $\mathcal{P}(\mathbb{C})$ and moreover, it takes values in $\mathbb{C}/\pi^2\mathbb{Q}$, rather than in $\mathbb{C}/\pi^2\mathbb{Z}$.

The repair turns out to be to use, instead of $\mathbb{C}-\{0,1\}$, a certain disconnected $\mathbb{Z}\times\mathbb{Z}$ cover of $\mathbb{C}-\{0,1\}$ to define "extended versions" of the groups $\mathcal{P}(\mathbb{C})$ and $\mathcal{B}(\mathbb{C})$. This idea developed out of a suggestion by Jun Yang.

We shall denote the relevant cover of $\mathbb{C}-\{0,1\}$ by $\widehat{\mathbb{C}}$. We start with two descriptions of it. The second will be a geometric interpretation in terms of ideal simplices.

Let P be $\mathbb{C}-\{0,1\}$ split along the rays $(-\infty,0)$ and $(1,\infty)$. Thus each real number r outside the interval $[0,1]$ occurs twice in P, once in the upper half plane of $\mathbb{C}$ and once in the lower half plane of $\mathbb{C}$. We denote these two occurences of r by $r+0i$ and $r-0i$. We construct $\widehat{\mathbb{C}}$ as an identification space from $P\times\mathbb{Z}\times\mathbb{Z}$ by identifying

$$\begin{aligned}(x+0i,p,q)&\sim(x-0i,p+2,q) &&\text{for each } x\in(-\infty,0)\\ (x+0i,p,q)&\sim(x-0i,p,q+2) &&\text{for each } x\in(1,\infty).\end{aligned}$$

We will denote the equivalence class of (z,p,q) by $(z;p,q)$. $\widehat{\mathbb{C}}$ has four components:

$$\widehat{\mathbb{C}}=X_{00}\cup X_{01}\cup X_{10}\cup X_{11}$$

where $X_{\epsilon_0\epsilon_1}$ is the set of $(z;p,q)\in\widehat{\mathbb{C}}$ with $p\equiv\epsilon_0$ and $q\equiv\epsilon_1 \pmod 2$.

We may think of X_{00} as the riemann surface for the function $\mathbb{C}-\{0,1\}\to\mathbb{C}^2$ defined by $z\mapsto(\log z,-\log(1-z))$. If for each $p,q\in\mathbb{Z}$ we take the branch $(\log z+2p\pi i,-\log(1-z)+2q\pi i)$ of this function on the portion $P\times\{(2p,2q)\}$ of X_{00} we get an analytic function from X_{00} to $\mathbb{C}^2$. In the same way, we may think of $\widehat{\mathbb{C}}$ as the riemann surface for the collection of all branches of the functions $(\log z+p\pi i,-\log(1-z)+q\pi i)$ on $\mathbb{C}-\{0,1\}$.

We can interpret $\widehat{\mathbb{C}}$ in terms of ideal simplices. Suppose we have an ideal simplex Δ with parameter $z\in\mathbb{C}-\{0,1\}$. Recall that this parameter is associated to an edge of Δ and that other edges of Δ have parameters

$$z'=\frac{1}{1-z},\quad z''=1-\frac{1}{z},$$

with opposite edges of Δ having the same parameter (see figure 1). Note that $zz'z''=-1$, so the sum

$$\log z+\log z'+\log z''$$

is an odd multiple of πi, depending on the branches of log used. In fact, if we use the standard branch of log then this sum is πi or $-\pi i$ depending on whether z is in the upper or lower half plane. This reflects the fact that the three dihedral angles of an ideal simplex sum to π.

Definition 4.1 We shall call any triple of the form

$$\mathbf{w} = (w_0, w_1, w_2) = (\log z + p\pi i, \log z' + q\pi i, \log z'' + r\pi i)$$

with

$$p, q, r \in \mathbb{Z} \quad \text{and} \quad w_0 + w_1 + w_2 = 0$$

a *combinatorial flattening* for our simplex Δ. Thus a combinatorial flattening is an adjustment of each of the three dihedral angles of Δ by a multiple of π so that the resulting angle sum is zero.

Each edge E of Δ is assigned one of the components w_i of $\mathbf{w}$, with opposite edges being assigned the same component. We call w_i the *log-parameter* for the edge E and denote it $l_E(\Delta, \mathbf{w})$.

For $(z; p, q) \in \widehat{\mathbb{C}}$ we define

$$\ell(z; p, q) := (\log z + p\pi i, -\log(1-z) + q\pi i, \log(1-z) - \log z - (p+q)\pi i),$$

and ℓ is then a map of $\widehat{\mathbb{C}}$ to the set of combinatorial flattenings of simplices.

Lemma 4.2 *This map ℓ is a bijection, so $\widehat{\mathbb{C}}$ may be identified with the set of all combinatorial flattenings of ideal tetrahedra.*

Proof We must show that $(w_0, w_1, w_2) = \ell(z; p, q)$ determines $(z; p, q)$. It clearly suffices to recover z. But up to sign z equals e^{w_0} and $1-z$ equals e^{-w_1}, and the knowledge of both z and $1-z$ up to sign determines z. □

4.1 The extended groups

We shall define a group $\widehat{\mathcal{P}}(\mathbb{C})$ as $\mathbb{Z}\langle\widehat{\mathbb{C}}\rangle/(\text{relations})$, where the relations in question are a lift of the five-term relations (2) that define $\mathcal{P}(\mathbb{C})$, plus an extra relation that just eliminates an element of order 2.

We first recall the situation of the five-term relation (2). If $z_0, \dots, z_4$ are five distinct points of $\partial\overline{\mathbb{H}}^3$, then each choice of four of five points $z_0, \dots, z_4$ gives an ideal simplex. We denote the simplex which omits vertex z_i by Δ_i. We denote the cross-ratio parameters of these simplices by $x_i = [z_0 : \dots : \hat{z}_i : \dots : z_4]$. Recall that $(x_0, \dots, x_4)$ can be written in terms of $x = x_0$ and $y = x_1$ as

$$(x_0, \dots, x_4) = \left(x, y, \frac{y}{x}, \frac{1-x^{-1}}{1-y^{-1}}, \frac{1-x}{1-y} \right)$$

The five-term relation was $\sum_{i=0}^{4} (-1)^i [x_i] = 0$, so the lifted five-term relation will have the form

$$\sum_{i=0}^{4} (-1)^i (x_i; p_i, q_i) = 0 \tag{3}$$

with certain relations on the p_i and q_i. We need to describe these relations.

Using the map of Lemma 4.2, each summand in this relation (3) represents a choice $\ell(x_i; p_i, q_i)$ of combinatorial flattening for one of the five ideal simplices. For each edge E connecting two of the points z_i we get a corresponding linear combination

$$\sum_{i=0}^{4} (-1)^i l_E(\Delta_i, \ell(x_i; p_i, q_i)) \tag{4}$$

of log-parameters (Definition 4.1), where we put $l_E(\Delta_i, \ell(x_i; p_i, q_i)) = 0$ if the line E is not an edge of Δ_i. This linear combination has just three non-zero terms corresponding to the three simplices that meet at the edge E. One easily checks that the real part is zero and the imaginary part can be interpreted (with care about orientations) as the sum of the "adjusted angles" of the three flattened simplices meeting at E.

Definition 4.3 We say that the $(x_i; p_i, q_i)$ satisfy the *flattening condition* if each of the above linear combinations (4) of log-parameters is equal to zero. That is, the adjusted angle sum of the three simplices meeting at each edge is zero. In this case relation (3) is an instance of the *lifted five-term relation.*

There are ten edges in question, so the flattening conditions are ten linear relations on the ten integers p_i, q_i. But these equations turn out to be linearly dependant, and the space of solutions is 5–dimensional. For example, if the five parameters $x_0, \dots, x_4$ are all in the upper half-plane (one can check that this means y is in the upper half-plane and x is inside the triangle with vertices $0, 1, y$) then the conditions are equivalent to:

$$p_2 = p_1 - p_0, \quad p_3 = p_1 - p_0 + q_1 - q_0, \quad p_4 = q_1 - q_0$$
$$q_3 = q_2 - q_1, \quad q_4 = q_2 - q_1 - p_0$$

which express p_2, p_3, p_4, q_3, q_4 in terms of p_0, p_1, q_0, q_1, q_2. Thus, in this case the lifted five-term relation becomes:

$$(x_0; p_0, q_0) - (x_1; p_1, q_1) + (x_2; p_1 - p_0, q_2) - \\ - (x_3; p_1 - p_0 + q_1 - q_0, q_2 - q_1) + (x_4; q_1 - q_0, q_2 - q_1 - p_0) = 0$$

This situation corresponds to the configuration of figure 2 for the ideal vertices $z_0, \dots, z_4$, with z_1 and z_3 on opposite sides of the plane of the triangle $z_0z_2z_4$

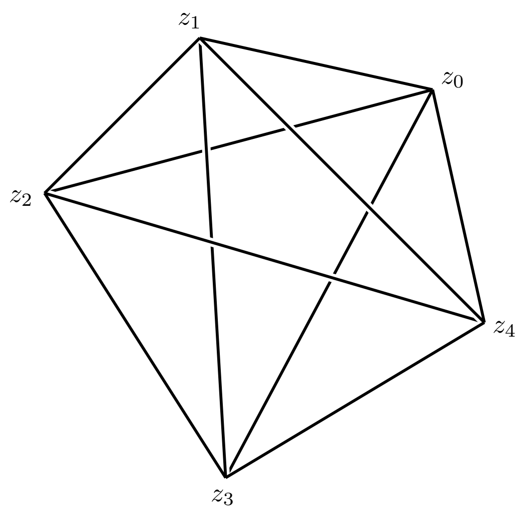


Figure 2

and the line from z_1 to z_3 passing through the interior of this triangle.

Definition 4.4 The extended pre-Bloch group $\widehat{\mathcal{P}}(\mathbb{C})$ is the group

$$\widehat{\mathcal{P}}(\mathbb{C}) := \mathbb{Z}\langle\widehat{\mathbb{C}}\rangle/(\text{lifted five-term relations and the following relation})$$

$$[x; p, q] + [x; p', q'] = [x; p, q'] + [x; p', q]. \tag{5}$$

(We call relation (5) the *transfer relation*; one can show that if one omits it then $\widehat{\mathcal{P}}(\mathbb{C})$ is replaced by $\widehat{\mathcal{P}}(\mathbb{C}) \oplus \mathbb{Z}/2$, where the $\mathbb{Z}/2$ is generated by the element $\kappa := [x, 1, 1] + [x, 0, 0] - [x, 1, 0] - [x, 0, 1]$, which is independant of x.)

The relations we are using are remarkably natural. To explain this we need a beautiful version of the dilogarithm function called the *Rogers dilogarithm*:

$$\mathcal{R}(z) = -\frac{1}{2}\left(\int_0^z \left(\frac{\log t}{1-t} + \frac{\log(1-t)}{t}\right) dt\right) - \frac{\pi^2}{6}.$$

The extra $-\pi^2/6$ is not always included in the definition but it improves the functional equation. $\mathcal{R}(z)$ is singular at 0 and 1 and is not well defined on $\mathbb{C}-\{0,1\}$, but it lifts to an analytic function

$$R:\widehat{\mathbb{C}}\to\mathbb{C}/\pi^2\mathbb{Z}$$
$$R(z;p,q)=\mathcal{R}(z)+\frac{\pi i}{2}(p\log(1-z)+q\log z).$$

We also consider the map

$$\hat{\delta}\colon\ \widehat{\mathbb{C}}\to\mathbb{C}\wedge\mathbb{C},\quad \hat{\delta}(z;p,q)=\big(\log z+p\pi i\big)\wedge\big(-\log(1-z)+q\pi i\big).$$

Relation (5) is clearly a functional equation for both R and $\hat{\delta}$. It turns out that the same is true for the lifted five-term relation. In fact:

Proposition 4.5 *If $(x_i;p_i,q_i)$, $i=0,\dots,4$ satisfy the flattening condition, so*

$$\sum_{i=0}^{4}(-1)^i(x_i;p_i,q_i)=0$$

is an instance of the lifted five-term relation, then

$$\sum_{i=0}^{4}(-1)^iR(x_i;p_i,q_i)=0$$

and

$$\sum_{i=0}^{4}(-1)^i\hat{\delta}(x_i;p_i,q_i)=0.$$

Moreover, each of these equations also characterises the flattening condition.

Thus the flattening condition can be defined either geometrically, as we introduced it, or as the condition that makes the five-term functional equation for either R or $\hat{\delta}$ valid. In any case, we now have:

Theorem 4.6 *R and $\hat{\delta}$ define maps*

$$R\colon\ \widehat{\mathcal{P}}(\mathbb{C})\to\mathbb{C}/\pi^2\mathbb{Z}$$
$$\hat{\delta}\colon\ \widehat{\mathcal{P}}(\mathbb{C})\to\mathbb{C}\wedge\mathbb{C}.$$

We call $\hat{\delta}$ the *extended Dehn invariant* and call its kernel

$$\widehat{\mathcal{B}}(\mathbb{C}):=\ker(\hat{\delta})$$

the *extended Bloch group.* The final step in our path from Hilbert's 3rd problem to invariants of 3–manifolds is given by the following theorem.

Theorem 4.7 *A hyperbolic 3–manifold M has a natural class $\hat\beta(M) \in \widehat{\mathcal{B}}(\mathbb{C})$. Moreover, $R(\hat\beta(M)) = \frac{1}{i}(\mathrm{vol}(M) + i\,\mathrm{CS}(M)) \in \mathbb{C}/\pi^2\mathbb{Z}$.*

To define the class $\hat\beta(M)$ directly from an ideal triangulation one needs to use a more restrictive type of ideal triangulation than the degree one ideal triangulations that suffice for $\beta(M)$. For instance, the triangulations constructed by Epstein and Penner [15] in the non-compact case and by Thurston [41] in the compact case are of the appropriate type. One then chooses flattenings of the ideal simplices of K so that the whole complex K satisfies certain "flatness" conditions. The sum of the flattened ideal simplices then represents $\hat\beta(M)$ up to a $\mathbb{Z}/6$ correction. The main part of the flatness conditions on K are the conditions that adjusted angles around each edge of K sum to zero together with similar conditions on homology classes at the cusps of M. If one just requires these conditions one obtains $\hat\beta(M)$ up to 12–torsion. Additional mod 2 flatness conditions on homology classes determine $\hat\beta(M)$ modulo 6–torsion. The final $\mathbb{Z}/6$ correction is eliminated by appropriately ordering the vertices of the simplices of K. It takes some work to see that all these conditions can be satisfied and that the resulting element of $\widehat{\mathcal{B}}(\mathbb{C})$ is well defined, see [23, 24].

5 Comments and questions

5.1 Relation with the non-extended Bloch group

What really underlies the above Theorem 4.7 is the

Theorem? 5.1 *There is a natural short exact sequence*

$$0 \to \mathbb{Z}/2 \to H_3(\mathrm{PSL}(2,\mathbb{C});\mathbb{Z}) \to \widehat{\mathcal{B}}(\mathbb{C}) \to 0.$$

The reason for the question mark is that, at the time of writing, the proof that the kernel is exactly $\mathbb{Z}/2$ has not yet been written down carefully.

The relationship of our extended groups with the "classical" ones is as follows.

Theorem 5.2 *There is a commutative diagram with exact rows and columns:*

$$
\begin{array}{ccccccccccc}
 & & 0 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \\
0 & \longrightarrow & \mu^* & \longrightarrow & \mathbb{C}^* & \longrightarrow & \mathbb{C}^*/\mu^* & \longrightarrow & 0 & & \\
 & & \downarrow{\scriptstyle \chi|\mu^*} & & \downarrow{\scriptstyle \chi} & & \downarrow{\scriptstyle \xi} & & \downarrow & & \\
0 & \longrightarrow & \widehat{\mathcal{B}}(\mathbb{C}) & \longrightarrow & \widehat{\mathcal{P}}(\mathbb{C}) & \xrightarrow{\hat\delta} & \mathbb{C}\wedge\mathbb{C} & \longrightarrow & K_2(\mathbb{C}) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow{\scriptstyle \epsilon} & & \downarrow{\scriptstyle =} & & \\
0 & \longrightarrow & \mathcal{B}(\mathbb{C}) & \longrightarrow & \mathcal{P}(\mathbb{C}) & \xrightarrow{\delta} & \mathbb{C}^*\wedge\mathbb{C}^* & \longrightarrow & K_2(\mathbb{C}) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & & 0 & &
\end{array}
$$

Here μ^* is the group of roots of unity and the labelled maps that have not yet been defined are:

$$
\begin{aligned}
\chi(z) &= [z,0,1] - [z,0,0] \in \widehat{\mathcal{P}}(\mathbb{C});\\
\xi[z] &= \log z \wedge \pi i;\\
\epsilon(w_1 \wedge w_2) &= (e^{w_1} \wedge e^{w_2}).
\end{aligned}
$$

5.2 Extended extended Bloch

The use of the disconnected cover $\widehat{\mathbb{C}}$ of $\mathbb{C} - \{0,1\}$ rather than the universal abelian cover (the component X_{00} of $\widehat{\mathbb{C}}$) in defining the extended Bloch group may seem unnatural. If one uses X_{00} instead of $\widehat{\mathbb{C}}$ one obtains extended Bloch groups $\mathcal{EP}(\mathbb{C})$ and $\mathcal{EB}(\mathbb{C})$ which are non-trivial $\mathbb{Z}/2$ extensions of $\widehat{\mathcal{P}}(\mathbb{C})$ and $\widehat{\mathcal{B}}(\mathbb{C})$. Theorem 5.1 then implies a natural *isomorphism* $H_3(\mathrm{PSL}(2,\mathbb{C});\mathbb{Z}) \to \mathcal{EB}(\mathbb{C})$. The homomorphism of Theorem 5.1 is given explicitely by "flattening" homology classes in the way sketched after Theorem 4.7, and the isomorphism $H_3(\mathrm{PSL}(2,\mathbb{C});\mathbb{Z}) \to \mathcal{EB}(\mathbb{C})$ presumably has a similar explicit description using "X_{00}–flattenings," but we have not yet proved that these always exist (note that an X_{00}–flattening of a simplex presupposes a choice of a pair of opposite edges of the simplex; changing this choice turns it into a X_{01}– or X_{10}–flattening).

For the same reason, we do not yet have a simplicial description of the class $\hat\beta(M) \in \mathcal{EB}\mathbb{C}$ for a closed hyperbolic manifold M, although this class exists for homological reasons. It is essential here that M be closed — the class

$\hat{\beta}(M) \in \widehat{\mathcal{B}}(\mathbb{C})$ almost certainly has no natural lift to $\mathcal{EB}(\mathbb{C})$ in the non-compact case.

The Rogers dilogarithm induces a natural map $R\colon \mathcal{EB}(\mathbb{C}) \to \mathbb{C}/2\pi^2\mathbb{Z}$, and this is the Cheeger–Simons class $H_3(PSL(2,\mathbb{C}) \to \mathbb{C}/2\pi^2\mathbb{Z}$ via the above isomorphism.

5.3 Computing Chern–Simons invariant

The formula of [23] for $CS(M)$ used in the programs Snappea and Snap uses ideal triangulations that arise in Dehn surgery. These triangulations are not of the type mentioned after Theorem 4.7, but by modifying them one can put them in the desired form and use Theorem 4.7 to compute $\hat{\beta}(M)$, reconfirming the formula of [23]. The formula computes $CS(M)$ up to a constant for the infinite class of manifolds that arise by Dehn surgery on a given manifold. It was conjectured in [23] that this constant is always a multiple of $\pi^2/6$, and this too is confirmed. The theorem also gives an independent proof of the relation of volume and Chern–Simons invariant conjectured in [30] and proved in [44], from which a formula for eta-invariant was also deduced in [25] and [31].

5.4 Realizing elements in the Bloch group and Gromov norm

One way to prove the Bloch group rigidity conjecture 2.9 would be to show that $\mathcal{B}(\mathbb{C})$ is generated by the classes $\beta(M)$ of 3–manifolds. This question is presumably much harder than the rigidity conjecture, although modifications of it have been used in attempts on it. More specifically, one can ask

Question 5.3 For which number fields k is $\mathcal{B}(k)_{\mathbb{Q}}$ generated as a $\mathbb{Q}$ vector space by classes $\beta(M)$ of 3–manifolds with invariant trace field contained in k?

For totally real number fields (ie $r_2 = 0$) the answer is trivially "yes" while for number fields with $r_2 = 1$ the existence of arithmetic manifolds again shows the answer is "yes." But beyond this little is known. In fact it is not even known if for every non-real number field $k \subset \mathbb{C}$ a 3–manifold exists with invariant trace field k. (For a few cases, eg multi-quadratic extensions of $\mathbb{Q}$, the author and A Reid have unpublished constructions to show the answer is "yes.")

Jun Yang has pointed out that "Gromov norm" gives an obstruction to a class $\alpha \in \mathcal{B}(\mathbb{C})$ being realizable as $\beta(M)$ (essentially the same observation also occurs in [34]). We define the *Gromov norm* $\nu(\alpha)$ as

$$\nu(\alpha) = \inf\{\sum|\frac{n_i}{k}| : k\alpha = \sum n_i[z_i], \quad z_i \in \mathbb{C}\},$$

and it is essentially a result of Gromov, with proof given in [40], that:

Theorem 5.4

$$|\operatorname{vol}(\alpha)| \leq V\nu(\alpha),$$

where $V = 1.00149416...$ is the volume of a regular ideal tetrahedron. If $\alpha = \beta(M)$ for some 3–manifold M then

$$\operatorname{vol}(\alpha) = V\nu(\alpha).$$

In particular, since $\nu(\alpha)$ is invariant under the action of Galois, for $\alpha = \beta(M)$ one sees that the $\operatorname{vol}(M)$ component of the Borel regulator is the largest in absolute value and equals $V\nu(\alpha)$. This suggests the question:

Question 5.5 Is it true for any number field k and for any $\alpha \in \mathcal{B}(k)$ that $V\nu(\alpha)$ equals the largest absolute value of a component of the Borel regulator of α?

This question is rather naive, and at this point we have no evidence for or against. Another naive question is the following. For $\alpha \in \mathcal{B}(k)_{\mathbb{Q}}$, where k is a number field, we can define a stricter version of Gromov norm by

$$\nu_k(\alpha) = \inf\Big\{\sum \Big|\frac{n_i}{k}\Big| : k\alpha = \sum n_i[z_i], \quad z_i \in k\Big\}.$$

Question 5.6 Is $\nu_k(\alpha) = \nu(\alpha)$ for $\alpha \in \mathcal{B}(k)_{\mathbb{Q}}$?

If not, then ν_k gives a sharper obstruction to realizing α as $\beta(M)$ since it is easy to show that for $\alpha = \beta(M)$ one has $\operatorname{vol}(\alpha) = V\nu_K(\alpha)$ for some at most quadratic extension K of k.

References

[1] **C Batut**, **D Bernardi**, **H Cohen**, **M Olivier**, *Pari-GP*, the program, available from: `ftp://megrez.ceremab.u-bordeaux.fr/pub/pari/` and `http://pari.home.ml.org`

[2] **S Bloch**, *Higher regulators, algebraic K–theory, and zeta functions of elliptic curves,* Lecture notes UC Irvine (1978)

[3] **M Bökstedt**, **M Brun**, **J L Dupont**, *Homology of $O(n)$ and $O^1(1,n)$ made discrete: an application of edgewise subdivision*, J. Pure Appl. Algebra (to appear)

[4] **A Borel**, *Cohomologie de SL_n et valeurs de fonction zeta aux points entiers*, Ann. Scuola Norm. Sup. Pisa Cl. Sup. (4) 4 (1977) 613–636

[5] **J Cheeger**, **J Simons**, *Differential characters and geometric invariants*, Springer Lecture Notes in Math. 1167 (1985) 50–80

[6] **S Chern**, **J Simons**, *Some cohomology classes in principal fiber bundles and their application to Riemannian geometry*, Proc. Nat. Acad. Sci. USA 68 (1971) 791-794

[7] **M Dehn**, *Über den Rauminhalt*, Math. Ann. 55 (1901) 465–478

[8] **J L Dupont**, *The dilogarithm as a characteristic class for flat bundles*, J. Pure and App. Algebra 44 (1987) 137–164

[9] **J L Dupont**, *Algebra of polytopes and homology of flag complexes*, Osaka J. Math. 19 (1982) 599–641

[10] **J L Dupont**, **F L Kamber**, *Cheeger–Chern–Simons classes of transversally symmetric foliations: dependance relations and eta-invariants*, Math. Ann. 295 (1993) 449–468

[11] **J L Dupont**, **W Parry**, **H Sah**, *Homology of classical Lie groups made discrete II*, J. Algebra 113 (1988) 215–260

[12] **J L Dupont**, **H Sah**, *Scissors congruences II*, J. Pure and Appl. Algebra 25 (1982) 159–195

[13] **J L Dupont**, **H Sah**, *Homology of Euclidean groups of motions made discrete and Euclidean scissors congruences*, Acta Math. 164 (1990) 1–27

[14] **J L Dupont**, **H Sah**, *Three questions about simplices in spherical and hyperbolic 3–space*, preprint (1997)

[15] **D B A Epstein**, **R Penner**, *Euclidean decompositions of non-compact hyperbolic manifolds*, J. Diff. Geom. 27 (1988) 67–80

[16] **A B Goncharov**, *Volumes of hyperbolic manifolds and mixed Tate motives*, preprint MPI/96-10 Max-Planck-Institut für Math. Bonn (1996)

[17] **O Goodman**, *Snap*, the program, (an ARC funded project) available from http://www.ms.unimelb.edu.au/~snap/

[18] **D Hilbert**, *Mathematical Problems*, English translation from: "Mathematical developements arising from Hilbert's problems", Proc. Symp. Pure Math. 28 part 1 Amer. Math. Soc. (1976)

[19] **B Jessen**, *The algebra of polyhedra and the Dehn–Sydler theorem*, Math. Scand. 22 (1968) 241–256

[20] **B Jessen**, *Zur Algebra der Polytope*, Göttinger Nachr. Math. Phys. (1972) 47–53

[21] **R Meyerhoff**, *Hyperbolic 3–manifolds with equal volumes but different Chern–Simons invariants*, from: "Low-dimensional topology and Kleinian groups", D B A Epstein (editor), London Math. Soc. Lecture notes 112 (1986) 209–215

[22] **J Milnor**, *Hyperbolic geometry: the first 150 years*, Bulletin Amer. Math. Soc. 6 (1982) 9–24

[23] **W D Neumann**, *Combinatorics of triangulations and the Chern–Simons invariant for hyperbolic* 3*–manifolds*, from: "Topology 90, Proceedings of the Research Semester in Low Dimensional Topology at Ohio State", Walter de Gruyter Verlag, Berlin–New York (1992) 243–272

[24] **W D Neumann**, *Extended Bloch group and the Chern–Simons class*, in preparation

[25] **R Meyerhoff**, **W Neumann**, *An asymptotic formula for the* η*–invariant of hyperbolic* 3*–manifolds*, Comment. Math. Helvetici 67 (1992) 28–46

[26] **W D Neumann**, **A W Reid**, *Arithmetic of hyperbolic manifolds*, from: "Topology 90, Proceedings of the Research Semester in Low Dimensional Topology at Ohio State" Walter de Gruyter Verlag, Berlin–New York (1992) 273–310

[27] **W D Neumann**, **J Yang**, *Problems for* K*–theory and Chern–Simons Invariants of Hyperbolic* 3*–Manifolds*, L'Enseignement Mathématique 41 (1995) 281–296

[28] **W D Neumann**, **J Yang**, *Invariants from triangulation for hyperbolic* 3*–manifolds*, Electronic Research Announcements of the Amer. Math. Soc. 1 (2) (1995) 72–79

[29] **W D Neumann**, **J Yang**, *Bloch invariants of hyperbolic* 3*–manifolds*, Duke Math. J. (to appear)

[30] **W D Neumann**, **D Zagier**, *Volumes of hyperbolic* 3*–manifolds*, Topology 24 (1985) 307-332

[31] **M Ouyang**, *A simplicial formula for the eta–invariant of hyperbolic* 3*–manifolds*, Topology (to appear)

[32] **D Ramakrishnan**, *Regulators, algebraic cycles, and values of* L*–functions*, Contemp. Math. 83 (1989) 183–310

[33] **A W Reid**, *A note on trace–fields of Kleinian groups*, Bull. London Math. Soc. 22 (1990) 349–352

[34] **A Reznickov**, *Rationality of secondary classes*, J. Diff. Geom. 43 (1996) 674–692

[35] **C S Sah**, *Hilbert's Third Problem: Scissors Congruence*, Res. Notes in Math. 33, Pitman (1979)

[36] **C S Sah**, *Scissors congruences, I, Gauss-Bonnet map*, Math. Scand. 49 (1982) 181–210

[37] **A A Suslin**, *Algebraic* K*–theory of fields*, Proc. Int. Cong. Math. Berkeley 1986, 1 (1987) 222–244

[38] **A A Suslin**, K_3 *of a field and the Bloch Group*, Proc. Steklov Inst. of Math. 183 (1990) English Translation by Amer. Math. Soc. 1991 4 217–239

[39] **J P Sydler**, *Conditions nécessaires et sufficiantes pour l'équivalence des polyédres l'espace Euclideien à trois dimensions*, Comment. Math. Helv. 42 (1965) 43–80

[40] **W P Thurston**, *The geometry and topology of 3–manifolds*, Lecture notes, Princeton University (1977)

[41] **W P Thurston**, *Hyperbolic structures on 3–manifolds I: Deformation of acylindrical manifolds*, Annals of Math. 124 (1986) 203–246

[42] **W P Thurston**, *Three dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bull. Amer. Math. Soc. 6 (1982) 357–381

[43] **J Weeks**, *Snappea*, the program, available from ftp://geom.umn.edu/pub /software/snappea/

[44] **T Yoshida**, *The η–invariant of hyperbolic 3–manifolds*, Invent. Math. 81 (1985) 473–514

[45] **D Zagier**, *The Bloch–Wigner–Ramakrishnan polylogarithm function*, Math. Annalen 286 (1990) 613–624

[46] **D Zagier**, *Polylogarithms, Dedekind zeta functions and the algebraic K–theory of fields*, Arithmetic algebraic geometry, Progr. Math. 89, Birkhäuser Boston, Boston MA, (1991) 391–430

[47] **V B Zylev**, *Equicomposability of equicomplementable polyhedra*, Sov. Math. Doklady 161 (1965) 453–455

[48] **V B Zylev**, *G–composedness and G–complementability*, Sov. Math. Doklady 179 (1968) 403–404

Department of Mathematics, The University of Melbourne
Parkville, Vic 3052, Australia

Email: `neumann@maths.mu.oz.au`

Received: 21 August 1997

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 413–418

The engulfing property for 3–manifolds

Graham A Niblo
Daniel T Wise

Abstract We show that there are Haken 3–manifolds whose fundamental groups do not satisfy the engulfing property. In particular one can construct a π_1–injective immersion of a surface into a graph manifold which does not factor through any proper finite cover of the 3–manifold.

AMS Classification 20E26; 20F34, 57M05, 57M25

Keywords Double coset decompositions, subgroup separability, 3–manifolds, engulfing property

1 Introduction

Definition A subgroup H of a group G is said to be **separable** if it is an intersection of finite index subgroups of G. It is said to be **engulfed** if it is contained in a proper subgroup of finite index in G.

Subgroup separability was first explored as a tool in low dimensional topology by Scott in [7]. He showed that if $f\colon \Sigma \longrightarrow M$ is a π_1–injective immersion of a surface in a 3–manifold and $f_*(\pi_1(\Sigma))$ is a separable subgroup of $\pi_1(M)$ then the immersion factors (up to homotopy) through an embedding in a finite cover of M. This technique has applications to the still open "virtual Haken conjecture" and the "positive virtual first Betti number conjecture".

The virtual Haken conjecture *If M is a compact, irreducible 3–manifold with infinite fundamental group then M is virtually Haken, that is it has a finite cover which contains an embedded, 2–sided, incompressible surface.*

The positive virtual first Betti number conjecture *If M is a compact, irreducible 3–manifold with infinite fundamental group then it has a finite cover with positive first Betti number.*

Unfortunately it is difficult in general to show that a given subgroup is separable, and it is known that not every subgroup of a 3–manifold group need be separable; the first example was given by Burns, Karrass and Solitar, [1]. On the other hand Shalen has shown that if an aspherical 3–manifold admits a π_1–injective immersion of a surface which factors through infinitely many finite covers then the 3–manifold is virtually Haken [2]. In group theoretic terms Shalen's condition says that the surface subgroup is contained in infinitely many finite index subgroups of the fundamental group of the 3–manifold, and this is clearly a weaker requirement than separability.

The engulfing property is apparently weaker still. It was introduced by Long in [3] to study hyperbolic 3–manifolds, and he was able to show that in some circumstances it implies separability. He remarks that "One of the difficulties with the LERF (separability) property is that there often appears to be nowhere to start, that is, it is conceivable that a finitely generated proper subgroup could be contained in no proper subgroups of finite index at all." In this note we show that this can happen for finitely generated subgroups of the fundamental group of a Haken (though not hyperbolic) 3–manifold. We give two examples, both already known not to be subgroup separable. One is derived from the recent work of Rubinstein and Wang, [6], and we consider it in Theorem 1. The other was the first known example of a 3–manifold group which failed to be subgroup separable and was introduced in [1] and further studied in [4] and [5]. Our proof that it fails to satisfy the engulfing property is more elementary than the original proof that it fails subgroup separability, and we hope that it sheds some light on this fact. Both of the examples are graph manifolds so they leave open the question of whether or not hyperbolic 3–manifold groups are subgroup separable or satisfy the engulfing property. In this connection we note that if every surface subgroup of any closed hyperbolic 3–manifold does satisfy the engulfing property then any such subgroup must be contained in infinitely many finite index subgroups, and Shalen's theorem would give a solution to the "virtual Haken conjecture" for closed hyperbolic 3–manifolds containing surface subgroups.

2 The example of Rubinstein and Wang

We will use the following lemma:

Lemma 1 *Let H be a separable subgroup of a group G. Then the index $[G : H]$ is finite if and only if there is a finite subset $F \subset G$ such that $G = HFH$.*

Proof If $[G : H]$ is finite then $G = FH$ for some finite subset $F \subseteq G$, so $G = HFH$ as required.

Now suppose that $G = HFH$ for some finite subset $F \subset G$. For each element $g \in F - (H \cap F)$, we can find a finite index subgroup $H_g \in G$ with $H < H_g$ but $g \notin H_g$. Now let $K = \bigcap_g H_g$. Since F is finite, K has finite index in G, and since $H < K$, K contains every double coset HgH which it intersects non-trivially. It follows that K only intersects a double coset HgH non-trivially if $g \in H$, and so $K = H$. □

Given a subgroup $H < G$ let $\overline{H}$ denote the intersection of the finite index subgroups of G which contain H. ($\overline{H}$ is the closure of H in the profinite topology on G). It is obvious that H is separable if and only if $H = \overline{H}$, and it is engulfed if and only if $G \neq \overline{H}$. If G is a finite union of double cosets of a subgroup H then it is also a finite union of double cosets of $\overline{H}$ and this is clearly a separable subgroup of G so by Lemma 1 it must have finite index. Now if H has infinite index in G and $\overline{H}$ has finite index in G they cannot be equal, and H is not separable. Hence we may interpret a finite double coset decomposition $G = HFH$ as an obstruction to separability for an infinite index subgroup $H < G$.

In [6] Rubinstein and Wang constructed a graph manifold M and a π_1–injective immersion $\phi\colon \Sigma \looparrowright M$ of a surface Σ which does not factor through an embedding into any finite cover of M. It follows from [7] that the surface group $H = \phi_*(\pi_1(\Sigma))$ is not separable in the 3–manifold group $G = \pi_1(M)$. In fact as we shall see G has a finite double coset decomposition $G = HFH$:

Lemma 2 *Let $\phi\colon \Sigma \looparrowright M$ be a π_1–injective immersion of a surface Σ in a 3–manifold M, and let M_H be the cover of M defined by the inclusion $\phi_*(\pi_1(\Sigma)) \hookrightarrow \pi_1(M)$. Let $\tilde{\phi}\colon \mathbb{R}^2 \looparrowright \tilde{M}$ be some lift of ϕ to the universal covers, and $\tilde{\Sigma}$ denote the image of $\tilde{\phi}$. Then the number of H orbits for the action on $G\tilde{\Sigma} = \{g\tilde{\Sigma} \mid g \in G\}$ is precisely the number of distinct double cosets HgH.*

Proof By construction $\tilde{\Sigma}$ is H–invariant, so for each double coset HgH we have $HgH\tilde{\Sigma} = Hg\tilde{\Sigma}$. It follows that if $F = \{g_i \mid i \in I\}$ is a complete family of representatives for the distinct double cosets Hg_iH in G then the G–orbit $G\tilde{\Sigma}$ breaks into $|F|$ H–orbits as required. □

Now in the example in [6] we are told in Corollary 2.5 that the image of each orbit $Hg\tilde{(\Sigma)}$ intersects the image of $H\tilde{\Sigma}$ which by construction of H is compact. Hence there are only finitely many such images, and therefore only finitely many H–orbits for the action of H on the set $G\tilde{\Sigma}$. Hence $G = HFH$ for some finite subset $F \subset G$.

Corollary *The profinite closure of H must have finite index in G, ie there are only finitely many finite index subgroups of G containing H, or, in topological terms, there are only finitely many finite covers of the 3–manifold M to which the surface Σ lifts by degree 1.*

Now as in the proof of Lemma 1, let K denote the intersection of the finite index subgroups of G containing H, and let M_K denote the finite cover of M corresponding to the finite index subgroup $K < G$. Then the immersion of Σ in M lifts to an immersion $\bar{\phi}\colon \Sigma \looparrowright M_K$ which does not lift to any finite cover of M_K. Hence:

Theorem 1 *There is a compact 3–manifold M_K and a π_1–injective immersion $\bar{\phi}\colon \Sigma \looparrowright M_K$ which does not factor through any proper finite cover of M_K.*

3 The example of Burns, Karrass and Solitar

In [1], Burns Karrass and Solitar gave an example of a 3–manifold group with a finitely generated subgroup which is not separable. Their example is a free by $\mathbb{Z}$ group with presentation $\langle \alpha, \beta, y \mid \alpha^y = \alpha\beta, \beta^y = \beta \rangle$. It is easy to show that their example is isomorphic to the group G with presentation $\langle a, b, t \mid [a, b], a^t = b \rangle$, and it is in this form that we shall work with G. Note that here and below we use the notation $x^y = y^{-1}xy$ and $[x, y] = x^{-1}y^1xy$.

In this section we show that G has a proper subgroup $K \subset G$ such that K is not engulfed. In particular, this yields an easier proof that G has non-separable subgroups.

Lemma 3 *Let $J = \langle abb, t \rangle$. Let H be a finite index subgroup of G containing J. Then $G = H\langle a \rangle$.*

Proof We express the argument in terms of covering spaces. Let X denote the standard based 2–complex for the presentation of G. Let T denote the torus subcomplex $\langle a, b \mid [a, b] \rangle$ of X. The complex X is formed from T by the addition of a cylinder C whose top and bottom boundary components are attached to the loops a and b respectively, and C is subdivided by a single edge labeled t which is oriented from the a loop to the b loop.

Let $\hat{X}$ denote the finite based cover of X corresponding to the subgroup H. Let $\hat{T}$ denote the cover of T at the basepoint of $\hat{X}$. Let $\hat{a}$ and $\hat{b}$ denote the covers of the loops a and b at the basepoint.

Since t lifts to a closed path in $\hat{X}$, we see that C has a finite cover $\hat{C}$ which lifts at the basepoint to a cylinder attached at its ends to $\hat{a}$ and $\hat{b}$. Now $\hat{C}$ gives a one-to-one correspondence between 0–cells on $\hat{a}$ and 0–cells on $\hat{b}$. In particular, each t edge of $\hat{C}$ is directed from some 0–cell in $\hat{a}$ to some 0–cell in $\hat{b}$ and therefore $\text{Degree}(\hat{a}) = \text{Degree}(\hat{b})$.

Because $abb \in J \subset H$ and hence $abb \in \pi_1(\hat{T})$, we see that b generates the covering group of the regular cover $\hat{T} \longrightarrow T$, and therefore $\text{Degree}(\hat{b}) = \text{Degree}(\hat{T})$. Thus we have $\text{Degree}(\hat{T}) = \text{Degree}(\hat{b}) = \text{Degree}(\hat{a})$, and because $\text{Degree}(\hat{T})$ is finite, we see that every 0–cell of $\hat{T}$ lies in both $\hat{a}$ and $\hat{b}$.

As above, each 0–cell of $\hat{a}$ has an outgoing t edge in $\hat{C}$ and each 0–cell of $\hat{b}$ has an incoming t edge in $\hat{C}$, and so we see that each 0–cell of $\hat{T} \cup \hat{C}$ has an incoming and outgoing t edge. Since 0–cells of $\hat{T} \cup \hat{C}$ obviously have incoming and outgoing a and b edges in $\hat{T}$, we see that $\hat{X} = \hat{T} \cup \hat{C}$ and in particular, every 0–cell of $\hat{X}$ is contained in $\hat{T}$ and therefore in $\hat{a}$. Thus $\langle a \rangle$ contains a set of right coset representatives for H in G, and consequently $G = H\langle a \rangle$. □

Lemma 4 *Let $K = \langle J \cup a^g \rangle$ for some $g \in G$. Then K is not engulfed.*

Proof Let H be a subgroup of finite index containing K. Since $J \subset H$ we may apply Lemma 3 to conclude that $G = H\langle a \rangle$ and so it is sufficient to show that $a \in H$. Observe that $g^{-1} = ha^n$ for some $h \in H$ and $n \in \mathbb{Z}$. But $a^g = (ha^n)aa^{-n}h^{-1} = hah^{-1}$, and obviously $hah^{-1} \in H$ implies that $a \in H$. □

Theorem 2 *Let K be the subgroup $\langle abb, t, btat^{-1}b^{-1} \rangle$. Then the engulfing property fails for K, that is, $K \neq G$ and the only subgroup of finite index containing K is G.*

Proof Lemma 4 with $g = t^{-1}b^{-1}$ shows that K is not engulfed. To see that $K \neq G$ we observe that the normal form theorem for an HNN extension shows that there is no non-trivial cancellation between the generators of K so it is a rank 3 free group, but G is not free. □

Remark It is not difficult to see that there are many finitely generated subgroups J for which some version of Lemma 3 is true. In addition, one has some freedom to vary the choice of g in theorem 2. Consequently subgroups of G which are not engulfed are numerous.

References

[1] **R G Burns**, **A Karrass**, **D Solitar** *A note on groups with separable finitely generated subgroups*, Bull. Aust. Math. Soc. 36 (1987) 153–160

[2] **G A Niblo**, *Finding splittings of groups and 3–manifolds*, Bull. London Math. Soc. 27 (1995) 567–574

[3] **D D Long**, *Engulfing and subgroup separability for hyperbolic 3–manifolds*, Trans. Amer. Math. Soc. 308 (1988) 849–859

[4] **D D Long**, **G A Niblo**, *Subgroup separability and 3–manifold groups*, Math. Z. 207 (1991) 209–215

[5] **G A Niblo**, **D T Wise**, *Subgroup separability, knot groups, and graph manifolds*, Proc. Amer. Math. Soc. (to appear)

[6] **H Rubinstein**, **S Wang**, π_1*–injective surfaces in graph manifolds*, Comm. Math. Helv. (to appear)

[7] **G P Scott**, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. 17 (1978) 555–565

Faculty of Mathematical Studies, University of Southampton
Highfield, Southampton, SO17 1BJ, UK
and
Department of Mathematics, Cornell University
Ithaca, NY 14853, USA

Email: gan@maths.soton.ac.uk, daniwise@math.cornell.edu

Received: 15 November 1997

ISSN 1464-8997 **419**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 419–450

Divergent sequences of Kleinian groups

Ken'ichi Ohshika

Abstract One of the basic problems in studying topological structures of deformation spaces for Kleinian groups is to find a criterion to distinguish convergent sequences from divergent sequences. In this paper, we shall give a sufficient condition for sequences of Kleinian groups isomorphic to surface groups to diverge in the deformation spaces.

AMS Classification 57M50; 30F40

Keywords Kleinian group, hyperbolic 3–manifold, deformation space

Dedicated to Prof David Epstein on the occasion of his 60th birthday

1 Introduction

The deformation space of a Kleinian group Γ is the space of faithful discrete representations of Γ into $PSL_2\mathbf{C}$ preserving parabolicity modulo conjugacy. It is one of the important aspects of Kleinian group theory to study the structures of deformation spaces. The first thing that was studied among the structures of deformation spaces was that of subspaces called quasi-conformal deformation spaces. By works of Ahlfors, Bers, Kra, Marden and Sullivan among others, the topological types and the parametrization of quasi-conformal deformation spaces are completely determined using the theory of quasi-conformal mappings and the ergodic theory on the sphere ([2], [5], [14], [24]). On the other hand, the total deformation spaces are less understood. A recent work of Minsky [16] makes it possible to determine the topological structure of the total deformation space completely in the case of once-puncture torus groups. The other cases are far from complete understanding. Although very rough topological structures, for instance the connected components of deformation spaces can be understood by virtue of recent works of Anderson–Canary and Anderson–Canary–McCullough, more detailed structures like the frontier of quasi-conformal deformation spaces are not yet known even in the case of surface groups with genus greater than 1.

A first step to understand the topological structure of the deformation space of a Kleinian group Γ is to give a criterion for a sequence $\{\Gamma_i\}$ in the deformation space to converge or diverge. In this paper, we shall consider the simplest case when the group Γ is isomorphic to a hyperbolic surface group $\pi_1(S)$ and has no accidental parabolic elements. In this case, Γ_i is either quasi-Fuchsian or a totally degenerate b–group, or a totally doubly degenerate group. Hence by taking a subsequence, we have only to consider the following three cases: all of the $\{\Gamma_i\}$ are quasi-Fuchsian, or totally degenerate b–groups, or totally doubly degenerate groups. For such groups, some conditions for sequences to converge are given for example in Bers [5], Thurston [28] and Ohshika [18]. Thurston's convergence theorem is called the double limit theorem. The purpose of this paper is to give a sufficient condition for sequences to diverge in the deformation space, which is in some sense complementary to the condition of the double limit theorem.

Before explaining the content of our main theorem, let us recall that a Kleinian group isomorphic to a hyperbolic surface group without accidental parabolic elements has two pieces of information describing the structures near ends as follows. When such a Kleinian group Γ is quasi-Fuchsian, by the Ahlfors–Bers theory, we get a pair of points in the Teichmuüller space $\mathcal{T}(S)$ corresponding to the group. In the case when Γ is a totally degenerate b–group, as there is one end of the non-cuspidal part $(\mathbf{H}^3/\Gamma)_0$ which is geometrically finite, we have a point in the Teichmüller space. In addition, the geometrically infinite end of $(\mathbf{H}^3/\Gamma)_0$ determines an ending lamination which is defined uniquely up to changes of transverse measures. Finally in the case when Γ is a totally doubly degenerate group, $(\mathbf{H}^3/\Gamma)_0$ have two geometrically infinite ends, and we have a pair of measured laminations which are ending laminations of the two ends. We shall define an end invariant of such a group Γ to be a pair (χ, υ) where each factor is either a point of the Teichmüller space or a projective lamination represented by an ending lamination, which gives the information on one of the ends.

The statement of our main theorem is as follows. Suppose that we are given a sequence of Kleinian groups (Γ_i, ϕ_i) in the parabolicity-preserving deformation space $AH_p(S)$ of Kleinian groups isomorphic to $\pi_1(S)$ for a hyperbolic surface S. Suppose moreover that the end invariants (χ_i, υ_i) have the following property: Either in the Thurston compactification or in the projective lamination space, $\{\chi_i\}$ and $\{\upsilon_i\}$ converge to maximal and connected projective laminations with the same support. Then the sequence $\{(\Gamma_i, \phi_i)\}$ does not converge in $AH_p(S)$.

To understand the meaning of this theorem, let us contrast it with Thurston's double limit theorem. For simplicity, we only consider the case when Γ_i is a quasi-Fuchsian group for the time being. By Ahlfors–Bers theory, a sequence of quasi-Fuchsian groups $\{(\Gamma_i, \phi_i)\}$ corresponds to a sequence of pairs of marked hyperbolic structures $\{(m_i, n_i)\}$ on S. Consider the case when both m_i and n_i diverge in the Teichmüller space and their limits in the Thurston compactification are projective laminations $[\mu]$ and $[\nu]$ respectively. The double limit theorem asserts that if μ and ν fill up S, viz., any measured lamination has non-zero intersection number with either μ or ν, then the sequence $\{(\Gamma_i, \phi_i)\}$ converges in the deformation space passing through a subsequence if necessary. The situation of our theorem is at the opposite pole to that of the double limit theorem. We assume in our theorem that μ and ν are equal except for the transverse measures and that they are maximal and connected.

We can see the assumption of maximality is essential by taking look at an example of Anderson–Canary [3]. They constructed an example of quasi-Fuchsian groups converging in $AH_p(S)$ which correspond to pairs of marked hyperbolic structures (m_i, n_i) such that $\{m_i\}$ and $\{n_i\}$ converge to the same point in $\mathcal{PL}(S)$. In this example, the support of the limit projective lamination is a simple closed curve, far from being maximal.

The proof of our theorem is based on an argument sketched in Thurston [26] which was used to prove his theorem stating that sequences of Kleinian groups isomorphic to surface groups which converge algebraically to Kleinian groups without accidental parabolic elements converge strongly. We shall give a detailed proof of this theorem in the last section as an application of our theorem.

The original version of this paper was written during the author's stay in University of Warwick for the symposium "Analytic and geometric aspects of hyperbolic spaces". The author would like to express his gratitude to the organizers of the symposium, Professors David Epstein and Caroline Series for inviting him there and giving him a lot of mathematical stimuli.

2 Preliminaries

Kleinian groups are discrete subgroups of the Lie group $PSL_2\mathbf{C}$ which is the group of conformal automorphisms of the 2–sphere S^2 and the orientation preserving isometry group of the hyperbolic 3–space $\mathbf{H}^3$. A Kleinian group acts conformally on S^2 and discontinuously on $\mathbf{H}^3$ by isometries. In this paper, we always assume that Kleinian groups are torsion free. For a torsion-free Kleinian group Γ, the quotient $\mathbf{H}^3/\Gamma$ is a complete hyperbolic 3–manifold.

Let Γ be a Kleinian group, which is regarded as acting on S^2. The subset of S^2 which is the closure of the set consisting of the fixed points of non-trivial elements in Γ, is called the limit set of Γ, and denoted by Λ_Γ. The limit set Λ_Γ is invariant under the action of Γ. The complement of Λ_Γ is called the region of discontinuity of Γ and denoted by Ω_Γ. The group Γ acts on Ω_Γ properly discontinuously. If Γ is finitely generated, the quotient Ω_Γ/Γ is a Riemann surface of finite type (ie a disjoint union of finitely many connected Riemann surfaces of finite genus with finitely many punctures) by Ahlfors' finiteness theorem [1].

A homeomorphism $\omega\colon S^2 \to S^2$ is said to be quasi-conformal if it has an L^2–distributional derivative and there exists a function $\mu\colon S^2 \to \mathbf{C}$ called a Beltrami coefficient whose essential norm is strictly less than 1, such that $\omega_{\overline{z}} = \mu\omega_z$. If the Beltrami coefficient μ for ω satisfies the condition $\mu \circ \gamma(z)\overline{\gamma}'(z)/\gamma'(z) = \mu(z)$ for every $\gamma \in \Gamma$, then the conjugate $\omega\Gamma\omega^{-1}$ is again a Kleinian group. A Kleinian group obtained by such a fashion from Γ is called a quasi-conformal deformation of Γ. By identifying two quasi-conformal deformations which are conformally conjugate, and giving the topology induced from the representation space, we obtain the quasi-conformal deformation space of Γ, which we shall denote by $QH(\Gamma)$. A quasi-conformal deformation of Ω_Γ/Γ can be extended to that of Γ. This gives rise to a continuous map $\rho\colon \mathcal{T}(\Omega_\Gamma/\Gamma) \to QH(\Gamma)$. By the works of Ahlfors, Bers, Kra, Marden and Sullivan among others, it is known that when Γ is finitely generated, ρ is a covering map, and that especially if Γ is isomorphic to a surface group (or more generally if Γ satisfies the condition $(*)$ introduced by Bonahon [7]), then ρ is a homeomorphism. The inverse of ρ is denoted by Q.

Let Γ be a finitely generated Kleinian group. We shall define the deformation space of Γ. An element γ of $PSL_2\mathbf{C}$ is said to be parabolic if it is conjugate to a parabolic element $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The deformation space of Γ, denoted by $AH_p(\Gamma)$, is the space of faithful discrete representations of Γ into $PSL_2\mathbf{C}$ preserving the parabolicity modulo conjugacy with the quotient topology induced from the representation space. We shall often denote an element (ie an equivalence class of groups) in $AH_p(\Gamma)$ in a form (G, ϕ) where ϕ is a faithful discrete representation with the image G which represents the equivalence class. The quasi-conformal deformation space $QH(\Gamma)$ is regarded as a subspace of $AH_p(\Gamma)$.

Let $C(\Lambda_\Gamma)$ be the intersection of $\mathbf{H}^3$ and the convex hull of the limit set Λ_Γ in the Poincaré ball $\mathbf{H}^3 \cup S^2_\infty$. As $C(\Lambda_\Gamma)$ is Γ–invariant, $C(\Lambda_\Gamma)$ can be taken quotient by Γ and gives rise to a closed convex set $C(\Lambda_\Gamma)/\Gamma$ in $\mathbf{H}^3/\Gamma$, which is called the convex core of $\mathbf{H}^3/\Gamma$. The convex core is the minimal closed convex set of $\mathbf{H}^3/\Gamma$ which is a deformation retract. A Kleinian group Γ is said to be geometrically finite if it is finitely generated and if the convex core of $\mathbf{H}^3/\Gamma$ has finite volume, otherwise it is geometrically infinite. When Γ is geometrically finite, $QH(\Gamma)$ is an open subset of $AH_p(\Gamma)$.

For a sequence $\{\Gamma_i\}$ of Kleinian groups, its geometric limit is defined as follows.

Definition 2.1 A Kleinian group H is called the geometric limit of $\{\Gamma_i\}$ if every element of H is the limit of a sequence $\{\gamma_i\}$ for $\gamma_i \in \Gamma_i$, and the limit of any convergent sequence $\{\gamma_{i_j} \in \Gamma_{i_j}\}$ for a subsequence $\{\Gamma_{i_j}\} \subset \{\Gamma_i\}$ is contained in H.

The geometric limit of non-elementary Kleinian groups is also a Kleinian group. We call a limit in the deformation space an algebraic limit to distinguish it from a geometric limit. We also call the first factor of a limit in the deformation space, ie the Kleinian group which is the image of the limit representation, an algebraic limit. Suppose that $\{(\Gamma_i, \phi_i)\}$ converges in $AH_p(\Gamma)$ to (Γ', ϕ). Then there is a subsequence of $\{\Gamma_i\}$ converging to a Kleinian group H geometrically. Moreover, the algebraic limit Γ' is contained in the geometric limit H. (Refer to Jørgensen–Marden [13] for the proofs of these facts.) When the algebraic limit Γ' coincides with the geometric limit H, we say that the sequence $\{(\Gamma_i, \phi_i)\}$ converges to (Γ', ϕ) strongly.

When $\{\Gamma_i\}$ converges geometrically to H, there exists a framed (K_i, r_i)–approximate isometry defined below between $\mathbf{H}^3/\Gamma_i$ and $\mathbf{H}^3/H$ with base-frames which are the projections of a base-frame on a point in $\mathbf{H}^3$ where $K_i \to 1$ and $r_i \to \infty$ as $i \to \infty$. (See Canary–Epstein–Green [9]).)

Definition 2.2 Let (M_1, e_1) and (M_2, e_2) be two Riemannian 3–manifolds with base-frame whose base-frames are based at $x_1 \in M_1$, and $x_2 \in M_2$ respectively. A (K, r)–approximate isometry between (M_1, e_1) and (M_2, e_2) is a diffeomorphism from (X_1, x_1) to (X_2, x_2) for subsets X_1, X_2 of M_1, M_2 containing the r–balls centred at x_1, x_2 such that $df(e_1) = e_2$ and

$$d_{M_1}(x,y)/K \le d_{M_2}(f(x), f(y)) \le K d_{M_1}(x,y)$$

for any $x, y \in X_1$.

Let $\{(M_i, v_i)\}$ be a sequence of hyperbolic 3–manifolds with base-frame. We say that (M_i, v_i) converges geometrically (in the sense of Gromov) to a hyperbolic 3–manifold with base-frame (N, w) when for any large r and $K > 1$ there exists an integer i_0 such that there exists a (K, r)–approximate isometry between

(M_i, v_i) and (N, w) for $i \geq i_0$. As described above, by choosing base-frames which are the images of a fixed base-frame in $\mathbf{H}^3$, the sequence of $\mathbf{H}^3/\Gamma_i$ with the base-frame converges geometrically to $\mathbf{H}^3/H$ with the base-frame when Γ_i converges to H geometrically.

Let $M = \mathbf{H}^3/\Gamma$ be a complete hyperbolic 3–manifold. A parabolic element of Γ is contained in a maximal parabolic subgroup, which is isomorphic to either $\mathbf{Z}$ or $\mathbf{Z} \times \mathbf{Z}$ and corresponds to a cusp of M. This is derived from Margulis' lemma. By deleting mutually disjoint neighbourhoods of the cusps of M, we obtain a non-cuspidal part of M, which we shall denote by M_0. We delete the cusp neighbourhoods where the injectivity radius is less than ϵ for some universal constant $\epsilon > 0$ so that this procedure of deleting cusp neighbourhoods is consistent among all the hyperbolic 3–manifolds. The non-cuspidal part M_0 is a 3–manifold whose boundary component is either a torus or an open annulus.

By theorems of Scott [22] and McCullough [15], there exists a submanifold $C(M)$ of M_0 such that $(C(M), C(M) \cap \partial M_0)$ is relatively homotopy equivalent to $(M_0, \partial M_0)$ by the inclusion, which is called a core of M. An end of M_0 is said to be geometrically finite if some neighbourhood of the end contains no closed geodesics, otherwise it is called geometrically infinite. A geometrically infinite end e is called geometrically infinite tame (or simply degenerate) if that end faces an incompressible frontier component S of a core and there exists a sequence of simple closed curves $\{\gamma_i\}$ on S such that the closed geodesic in M homotopic to γ_i tends to the end e as $i \to \infty$. (In this paper we use this term only when every component of the frontier of the core is incompressible.) A Kleinian group Γ is geometrically finite if and only if every end of $(\mathbf{H}^3/\Gamma)_0$ is geometrically finite.

In this paper, we shall consider sequences of Kleinian groups isomorphic to surface groups. Let S be a hyperbolic surface of finite area. We call punctures of S cusps. We denote by $AH_p(S)$ the space of Kleinian groups modulo conjugacy which are isomorphic to $\pi_1(S)$ by isomorphisms mapping elements represented by cusps to parabolic elements. We can also identify this space $AH_p(S)$ with the deformation space of a Fuchsian group G such that $\mathbf{H}^2/G = S$. Let (Γ, ϕ) be a class in $AH_p(S)$. We say that a parabolic element $\gamma \in \Gamma$ is accidental parabolic when $\phi^{-1}(\gamma)$ does not correspond to a cusp of S. Assume that (Γ, ϕ) in $AH_p(S)$ has no accidental parabolic element. Then the non-cuspidal part $(\mathbf{H}^3/\Gamma)_0$ has only two ends since one can see that a core is homeomorphic to $S \times I$ and has exactly two frontier components. Therefore in this case, Γ is either (1) a quasi-Fuchsian group, ie geometrically finite and the limit set Λ_Γ is homeomorphic to the circle or (2) a totally degenerate b–group, ie Ω_Γ is connected and simply connected, and $(\mathbf{H}^3/\Gamma)_0$ has one geometrically finite end and one geometrically infinite end, or (3) a totally doubly degenerate group, ie $\Omega_\Gamma = \emptyset$, and $(\mathbf{H}^3/\Gamma)_0$ has two geometrically infinite tame ends. Recall that a Kleinian group is called a b–group when its region of discontinuity has a unique invariant component, which is simply connected.

For a hyperbolic surface $S = \mathbf{H}^2/\Gamma$, we denote the quasi-conformal deformation space of Γ by $QF(S)$. This space consists of quasi-Fuchsian groups isomorphic to $\pi_1(S)$ by isomorphisms taking elements representing cusps to parabolic elements. By the Ahlfors–Bers theory, there is a homeomorphism $Q\colon QF(S) \to \mathcal{T}(S) \times \mathcal{T}(\overline{S})$, which we shall call the Ahlfors–Bers homeomorphism. Here $\mathcal{T}(\overline{S})$ denotes the Teichmüller space of the "complex conjugate" of S. This can be interpreted as the space of marked hyperbolic structures on S such that the complex conjugate of the corresponding complex structure is equal to the structure on the second component of Ω_Γ/Γ. We identify $\mathcal{T}(\overline{\mathcal{S}})$ with $\mathcal{T}(\mathcal{S})$ by the above correspondence from now on. By this correspondence, the Fuchsian representations of $\pi_1(S)$ are mapped onto the diagonal of $\mathcal{T}(S) \times \mathcal{T}(S)$.

Thurston introduced a natural compactification of a Teichmüller space in [27], which is called the Thurston compactification nowadays. Let S be a hyperbolic surface of finite area. Let $\mathcal{S}$ denote the set of free homotopy classes of simple closed curves on S. Let $P\mathbf{R}_+^{\mathcal{S}}$ denote the projective space obtained from the space $\mathbf{R}_+^{\mathcal{S}}$ of non-negative functions on $\mathcal{S}$. We endow $P\mathbf{R}_+^{\mathcal{S}}$ with the quotient topology of the weak topology on $\mathbf{R}_+^{\mathcal{S}} \setminus \{0\}$. The Teichmüller space $\mathcal{T}(S)$ is embedded in $P\mathbf{R}_+^{\mathcal{S}}$ by taking $g \in \mathcal{T}(\mathcal{S})$ to the class represented by a function whose value at $s \in \mathcal{S}$ is the length of the closed geodesic in the homotopy class. The closure of the image of $\mathcal{T}(S)$ in $P\mathbf{R}_+^{\mathcal{S}}$ is homeomorphic to the ball and defined to be the Thurston compactification of $\mathcal{T}(S)$. The boundary of $\mathcal{T}(S)$ corresponds to "the space of projective laminations" in the following way.

A compact subset of S consisting of disjoint simple geodesics is called a geodesic lamination. A geodesic lamination endowed with a transverse measure which is invariant under a homotopy along leaves is called a measured lamination. The subset of a measured lamination λ consisting of the points $x \in \lambda$ such that any arc containing x at the interior has a positive measured with respect to the transverse measure is called the support of λ. We can easily see that the support of a measured lamination λ is a geodesic lamination. The set of measured laminations with the weak topology with respect to measures on finite unions of arcs is called the measured lamination space and denoted by $\mathcal{ML}(S)$. The set of simple closed geodesics with positive weight is dense in $\mathcal{ML}(S)$. For a measured lamination (λ, μ), where μ denotes the transverse measure, and a homotopy class of simple closed curves σ, we define their intersection number $i(\lambda, \sigma)$ to be

$\inf_{s\in\sigma}\mu(s)$. (We also use the notation $i(\lambda,s)$ to denote $i(\lambda,[s])$.) By defining the value at $\sigma\in\mathcal{S}$ to be $i(\lambda,\sigma)$, we can define a map $\iota\colon \mathcal{ML}(S)\to\mathbf{R}_+^{\mathcal{S}}$. By projectivising the both spaces, we have a map $\bar\iota\colon \mathcal{PL}(S)\to P\mathbf{R}_+^{\mathcal{S}}$, where $\mathcal{PL}(S)$ denotes the projectivization of $\mathcal{ML}(S)$, ie $(\mathcal{ML}(S)\setminus\{\emptyset\})/(0,\infty)$. It can be proved that in fact $\bar\iota$ is an embedding and coincides with the boundary of the image of $\mathcal{T}(S)$, that is, the boundary of the Thurston compactification of $\mathcal{T}(S)$. Refer to Fathi et al [11] for further details of these facts.

Let e be a geometrically infinite tame end of the non-cuspidal part of a hyperbolic 3–manifold M, which faces a frontier component Σ of a core. From now on, we always assume that every frontier component of a core is incompressible in M. By the definition of geometrically infinite tame end, there exists a sequence of simple closed curves $\{\gamma_i\}$ on Σ such that the closed geodesic homotopic to γ_i tends to e as $i\to\infty$. Consider the sequence $\{[\gamma_i]\}$ (the projective classes represented by $\{\gamma_i\}$) in $\mathcal{PL}(\Sigma)$. (We identify γ_i with the closed geodesic homotopic to γ_i with respect to some fixed hyperbolic structure on Σ.) Since $\mathcal{PL}(\Sigma)$ is compact, the sequence $\{[\gamma_i]\}$ converges to a projective lamination $[\lambda]\in\mathcal{PL}(\Sigma)$ after taking a subsequence. Such a measured lamination λ is called an ending lamination of e. (The original definition is due to Thurston [26].) An ending lamination is maximal (ie it is not a proper sublamination of another measured lamination), and connected. (Thurston [26], see also Ohshika [17].) If both λ and λ' are ending laminations of an end e, their intersection number $i(\lambda,\lambda')$ is equal to 0 (essentially due to Thurston [26] and Bonahon [7]). We shall give a proof of this fact, based on Bonahon's result in section 3. By the maximality, this implies that $|\lambda|=|\lambda'|$ where $|\lambda|$ denotes the support of λ.

In this paper, we shall deal with a hyperbolic 3–manifold $M=\mathbf{H}^3/\Gamma$ with a homotopy equivalence $\phi\colon S\to M$ preserving cusps. In this case, M has a core which is homeomorphic to $S\times I$. For a homotopy equivalence $\phi\colon S\to M$ and a lamination λ, its image $\phi(\lambda)$ is homotopic to a unique lamination on $S\times\{t\}$ for both $t=0,1$. When the measured lamination homotopic to $\phi(\lambda)$ is an ending lamination, we say that $\phi(\lambda)$ represents an ending lamination. For an end e of M, the *end invariant* of e is defined to be a projective lamination $[\lambda]$ on S such that $\phi(\lambda)$ represent an ending lamination of e when e is geometrically infinite, and the point in the Teichmüller space corresponding to the conformal structure of the component of Ω_Γ/Γ when e is geometrically finite.

Now, let e_1,e_2 be the two ends of M_0 which are contained in the "upper complement" and the "lower complement" of a core respectively with respect to the orientation give on M and S. We define the *end invariant* of $M=\mathbf{H}^3/\Gamma$ to be a pair (χ,υ), where χ is the end invariant of e_1 and υ that of e_2. This means in particular that when Γ is a quasi–Fuchsian group, the end invariant is equal to $Q(\Gamma, \phi) \in \mathcal{T}(S) \times \mathcal{T}(S)$, where $Q\colon QF(S) \to \mathcal{T}(S) \times \mathcal{T}(S)$ is the Ahlfors–Bers map with the second factor $\mathcal{T}(\overline{S})$ identified with $\mathcal{T}(S)$.

Let S be a hyperbolic surface of finite area and M a complete hyperbolic 3–manifold. A pleated surface $f\colon S \to M$ is a continuous map which is totally geodesic in $S - \ell$ for some geodesic lamination ℓ on S such that the path metric induced by f coincides with the hyperbolic metric on S. We say that a sequence of pleated surfaces with base point $\{f_i\colon (S_i, x_i) \to (M_i, y_i)\}$ converges geometrically to a pleated surface with base point $f\colon (S, x) \to (M, y)$ when there are (K_i, r_i)–approximate isometries ρ_i between (M_i, v_i) and (M, v), and $\overline{\rho}_i$ between (S_i, w_i) and (S, w) such that $K_i \to 1$ and $r_i \to \infty$ as $i \to \infty$ and $\{\rho_i \circ f_i \circ \overline{\rho}_i^{-1}\}$ converges to f uniformly on every compact subset of S, where v_i, v, w_i, w are base-frames on x_i, x, y_i, y respectively. The space of pleated surfaces has the following compactness property due to Thurston whose proof can be found in Canary–Epstein–Green [9].

Proposition 2.3 *For any sequence of pleated surfaces with base point $\{f_i\colon (S_i, x_i) \to (M_i, y_i)\}$ such that the injectivity radius at y_i is bounded away from 0 as $i \to \infty$, there exists a subsequence which converges geometrically.*

We say that a (measured or unmeasured) geodesic lamination λ on S is realized by a pleated surface f when λ is mapped totally geodesically by f. A measured lamination λ lying on a component of the frontier of a core of M represents an ending lamination of an end of M_0 if and only if there is no pleated surface (homotopic to the inclusion) realizing λ. (This follows from Proposition 5.1 in Bonahon [7] which we shall cite below as Proposition 2.5.)

We shall use the following two results of Bonahon [7] several times in this paper. The first is Proposition 3.4 in his paper.

Lemma 2.4 (Bonahon) *Let M be a complete hyperbolic 3–manifold. Let S be a properly embedded incompressible surface in the non-cuspidal part M_0. Then there exists a constant C with the following property. Let α^*, β^* be closed geodesics in M which are homotopic to closed curves α, β on S by homotopies coming to the same side of S, and are located at distance at least D from S. Suppose that neither α^* nor β^* intersects a Margulis tube whose axis is not itself, α^* or β^*. Then we have*

$$i(\alpha, \beta) \leq Ce^{-D}\text{length}(\alpha)\text{length}(\beta) + 2.$$

The second is Proposition 5.1 in Bonahon's paper. Before stating the proposition, we need to define some terms used there. A train track on a surface S is a graph with C^1–structure such that all edges coming to a vertex are tangent mutually there. Furthermore we impose the condition that there is no component of the complement which is the interior of a monogon or a bigon or an annulus without angle. We call edges of a train track branches and vertices switches. A regular neighbourhood of a train track τ can be foliated by arcs transverse to τ. Such a neighbourhood is called a tied neighbourhood of τ, and the arcs are called ties. We say that a geodesic lamination λ is carried by a train track τ when a tied neighbourhood of τ can be isotoped to contain λ so that each leaf of λ should be transverse to the ties.

When λ is a measured lamination and carried by a train track τ, the transverse measure induces a weight system on the branches of τ, by defining the weight of a branch to be the measure of ties intersecting the branch. We can easily prove that such a weight system is uniquely determined by λ and τ. Conversely a weight system w on a train track τ satisfying the switch condition that the sum of weights on incoming branches and the sum of those on outgoing branches coincides at each switch, determines a unique measured lamination such that the weight system which it induces on τ is equal to w. Refer to Penner–Harer [23] for more precise definitions and explanations for these facts.

A continuous map f from a surface S to a hyperbolic manifold M is said to be adapted to a tied neighbourhood N_τ of a train track τ on S when each branch of τ is mapped to a geodesic arc in M and each tie of N_τ is mapped to a point. Consider a map f adapted to a tied neighbourhood of a train track τ. For a weight system w on τ, we define the length of $f(\tau, w)$ to be $\sum w_b \text{length}(f(b))$, where the sum is taken over all the branches of τ, and w_b denotes the weight on b assigned by w. For a measured lamination λ carried by τ, if it induces a weight system w on τ, we define the length of $f(\lambda)$ to be the length of $f(\tau, w)$.

For two branches b, b' meeting at a switch σ from opposite directions, the exterior angle $\theta(f(b, b'))$ between b, b' with respect to f is the exterior angle formed by $f(b)$ and $f(b')$ at $f(\sigma)$. The weight system w determines the weight flowing from b to b'. Let $b_1, \ldots, b_p$ and $b'_1, \ldots, b'_q$ be the branches meeting at a switch σ with $b_1, \ldots, b_p$ coming from one direction and $b'_1, \ldots, b'_q$ from the other. The exterior angle at $f(\sigma)$ is the sum of $w_{k,l}\theta(f(b_k, b'_l))$ for all $k = 1, \ldots, p, l = 1, \ldots, q$, where $w_{k,l}$ denotes the weight flowing from b_k to b'_l. The quadratic variation of angle at $f(\sigma)$ is the sum of $w_{k,l}\theta^2 f(b_k, b'_l)$ in the same situation as above. The total curvature of $f(\tau, w)$ is defined to be the sum of the exterior angles at all the images of switches on τ. Similarly, the quadratic variation of angle for $f(\tau, w)$ is defined to be the sum of the quadratic variations of angle at all switches.

Proposition 2.5 (Bonahon) *Let M be a complete hyperbolic 3–manifold and S a hyperbolic surface of finite type. Let $\phi\colon S \to M$ be a continuous incompressible map taking cusps to cusps, and λ a measured lamination on S. Then the one of the following two cases occurs and they are mutually exclusive.*

(1) *For any $\epsilon > 0$, there is a map ϕ_ϵ homotopic to ϕ, which is adapted to a train track carrying λ such that* $\mathrm{length}(\phi_\epsilon(\lambda)) < \epsilon$.

(2) *For any ϵ, there is a map ϕ_ϵ homotopic to ϕ, which is adapted to a train track τ carrying λ by a weight system ω, with the following property: The total curvature and the quadratic variation of angle for $\phi_\epsilon(\tau, w)$ are less than ϵ. Furthermore such a map ϕ_ϵ satisfies the following: There are $\delta > 0, t < 1$ such that $\delta \to 0, t \to 1$ as $\epsilon \to 0$, and for any simple closed curve γ such that $[\gamma]$ is sufficiently close to $[\lambda]$ in $\mathcal{PL}(S)$, the closed geodesic γ^* homotopic to $\phi(\gamma)$ in M has a part of length at least $t\mathrm{length}\phi_\epsilon(\gamma)$ which lies within distance δ from $\phi_\epsilon(\gamma)$.*

We can easily see that the first alternative exactly corresponds to the case when λ represents an ending lamination, and that the second alternative holds if and only if there is a pleated surface realizing λ. Taking this into account, the proposition implies in particular the following. First, in the situation as in the proposition, $\phi(\lambda)$ represents an ending lamination of an end of M_0 if and only if it is not realized by a pleated surface homotopic to ϕ since the two alternatives are exclusive.

Secondly, if λ is an ending lamination, then any measured lamination λ' with the same support as λ is also an ending lamination. This is because a train track carrying λ also carries λ' and if the condition (1) holds for λ, it equally holds for the weight system corresponding to λ'.

There is another proposition which we shall make use of essentially in our proof. The proposition is an application of Thurston's covering theorem which originally appeared in [26] (see also [19] for its proof, and Canary [8] for its generalization).

Proposition 2.6 (Thurston) *Let S be a hyperbolic surface of finite area. Let $\{(\Gamma_i, \phi_i)\}$ be a sequence of Kleinian groups in $AH_p(S)$ converging to (G, ψ). Let Γ_∞ be a geometric limit of $\{\Gamma_i\}$ after taking a subsequence, and let $q\colon \mathbf{H}^3/G \to \mathbf{H}^3/\Gamma_\infty$ be the covering map associated with the inclusion $G \subset \Gamma_\infty$. Suppose that $(\mathbf{H}^3/G)_0$ has a geometrically infinite end e. Then there exists a neighbourhood E of e such that $q|E$ is a proper embedding.*

3 The main theorem

Our main theorem on a sufficient condition for Kleinian groups isomorphic to surface groups to diverge in the deformation spaces is the following.

Theorem 3.1 *Let S be a hyperbolic surface of finite area. Let $\{(\Gamma_i, \phi_i)\}$ be a sequence of Kleinian groups in $AH_p(S)$ with isomorphisms $\phi_i\colon \pi_1(S) \to \Gamma_i$ inducing homotopy equivalences $\phi_i\colon S \to \mathbf{H}^3/\Gamma_i$. Let (χ_i, υ_i) be an end invariant of (Γ_i, ϕ_i). Suppose that $\{\chi_i\}$ and $\{\upsilon_i\}$ converge in either the Thurston compactification of the Teichmüller space $\mathcal{T}(S)$ or the projective lamination space $\mathcal{PL}(S)$ to maximal connected projective laminations $[\mu]$ and $[\nu]$ with the same support. Then $\{(\Gamma_i, \phi_i)\}$ does not converge in $AH_p(S)$.*

Let us briefly sketch the outline of the proof of our main theorem. Note that we can assume by taking a subsequence that all the Γ_i are the same type of the three; quasi-Fuchsian groups or totally degenerate b–groups or totally doubly degenerate groups. We consider here only the case when all the Γ_i are quasi-Fuchsian. The proof is by reductio ad absurdum. Suppose that our sequence $\{(\Gamma_i, \phi_i)\}$ converges in $AH_p(S)$. Then we have the algebraic limit (G, ψ) which is a subgroup of a geometric limit Γ_∞. By applying the continuity of the length function on $AH_p(S) \times \mathcal{ML}(S)$, which will be stated and proved in Lemma 4.2, we shall show that $\psi(\mu)$ represents an ending lamination of an end e_μ in $(\mathbf{H}^3/G)_0$. We shall take a neighbourhood E_μ of e_μ which can be projected homeomorphically by the covering map $q\colon \mathbf{H}^3/G \to \mathbf{H}^3/\Gamma_\infty$ to a neighbourhood of an end of $(\mathbf{H}^3/\Gamma_\infty)_0$ using Proposition 2.6. Let S_0 denote the non-cuspidal part of S. We shall then show that deep inside E_μ there is an embedded surface $f'(S_0)$ homotopic to $\psi|S_0$ such that every pleated surface homotopic to $q \circ \psi$ touching $q \circ f'(S_0)$ is contained in $q(E_\mu)$.

By projecting f' to $\mathbf{H}^3/\Gamma_\infty$ and pulling back by an approximate isometry, we get an embedded surface $f_i\colon S_0 \to \mathbf{H}^3/\Gamma_i$ which is homotopic to ϕ_i converging to an embedded surface $f_\infty\colon S_0 \to \mathbf{H}^3/\Gamma_\infty$ geometrically which is the projection of f'. By using a technique of interpolating pleated surfaces due to Thurston, we shall show that there is a pleated surface $k_i\colon S \to \mathbf{H}^3/\Gamma_i$ homotopic to ϕ_i which intersects $f_i(S_0)$ at an essential simple closed curve. These pleated surfaces converge geometrically to a pleated surface $k_\infty\colon S' \to \mathbf{H}^3/\Gamma_\infty$, where S' is an open incompressible surface on S. The condition that the limit surface k_∞ touches $f_\infty(S)$ forces k_∞ to be a pleated surface from S, and to be lifted to a pleated surface to $\mathbf{H}^3/G$ which realizes a measured lamination with the same support as μ. This will contradict the fact that $\psi(\mu)$ represents an ending lamination.

4 Ending laminations and pleated surfaces

In this section, we shall prove lemmata basically due to Thurston which will be used in the proof of our main theorem.

Throughout this section, $\{(\Gamma_i, \phi_i)\}$ denotes a sequence as in Theorem 3.1. Suppose that $\{(\Gamma_i, \phi_i)\}$ converges to (G, ψ) in $AH_p(S)$ where $\psi\colon \pi_1(S) \to G$ is an isomorphism. (Our proof of Theorem 3.1 is by reductio ad absurdum. Therefore we assumed above the contrary of the conclusion of Theorem 3.1.) We also use this symbol ψ to denote the homotopy equivalence from S to $\mathbf{H}^3/G$ corresponding to the isomorphism. We can assume that ϕ_i converges to ψ as representations by taking conjugates if necessary.

Now let $\tilde{z} \in \mathbf{H}^3$ be a point and $\tilde{v}$ be a frame based on $\tilde{z}$. Then $\tilde{z}, \tilde{v}$ are projected by the universal covering maps to z_i, v_i of $\mathbf{H}^3/\Gamma_i$ and z, v of $\mathbf{H}^3/G$. Since we assumed that $\{\Gamma_i\}$ converges algebraically to G, we can assume by passing through a subsequence that $\{\Gamma_i\}$ converges geometrically to a Kleinian group Γ_∞ which contains G as a subgroup. Let v_∞, z_∞ be the images in $\mathbf{H}^3/\Gamma_\infty$ of $\tilde{v}, \tilde{z}$ by the universal covering map.

The hyperbolic manifolds with base frame $\{(\mathbf{H}^3/\Gamma_i, v_i)\}$ converge in the sense of Gromov to $(\mathbf{H}^3/\Gamma_\infty, v_\infty)$. Let $q\colon \mathbf{H}^3/G \to \mathbf{H}^3/\Gamma_\infty$ be the covering associated with the inclusion $G \subset \Gamma_\infty$. Then $q(z) = z_\infty$ and $dq(v) = v_\infty$.

Consider the case when at least one end e of $(\mathbf{H}^3/\Gamma_i)_0$ is geometrically finite. Let Σ_i be the boundary components of the convex core of $\mathbf{H}^3/\Gamma_i$ facing e which corresponds to a component of the quotient of the region of discontinuity $\Omega^0_{\Gamma_i}/\Gamma_i$. Let $h_i\colon S \to \Sigma_i$ be a homeomorphism homotopic to ϕ_i. Now by the assumption of Theorem 3.1, the marked conformal structures of $\Omega^0_{\Gamma_i}/\Gamma_i$ converge to either $[\mu]$ or $[\nu]$, say $[\mu]$. Then we have the following.

Lemma 4.1 *There exist an essential simple closed curve γ_i on Σ_i, and a sequences of positive real numbers $\{r_i\}$ going to 0 such that $r_i \mathrm{length}_{\Sigma_i}(\gamma_i) \to 0$ and $\{r_i(h_i^{-1}(\gamma_i)) \in \mathcal{ML}(S)\}$ converges to a measured lamination with the same support as the measured lamination μ, where we regard $h_i^{-1}(\gamma_i)$ as an element in $\mathcal{ML}(S)$.*

Proof Let m_i be the point in $\mathcal{T}(S)$ determined by the marked conformal structure on $\Omega^0_{\Gamma_i}/\Gamma_i$. By Sullivan's theorem proved in Epstein–Marden [10], the assumption in Theorem 3.1 that $m_i \to [\mu]$ implies that the marked hyperbolic structures g_i on S induced by h_i from those on Σ_i as subsurfaces in $\mathbf{H}^3/\Gamma_i$ also converge to $[\mu]$ as $i \to \infty$ in the Thurston compactification of $\mathcal{T}(S)$.

Let γ_i be the shortest essential closed curves on Σ_i with respect to the hyperbolic metrics induced from $\mathbf{H}^3/\Gamma_i$. Consider the limit $[\mu_0]$ of $\{[h_i^{-1}(\gamma_i)]\}$ in $\mathcal{PL}(S)$ passing through a subsequence if necessary. Then there are bounded sequences of positive real numbers r_i such that $r_i h_i^{-1}(\gamma_i) \to \mu_0$ in $\mathcal{ML}(S)$. Suppose that $i(\mu, \mu_0) \neq 0$. Then by the "fundamental lemma" 8-II-1 in Fathi–Laudenbach–Poenaru, we should have $\mathrm{length}(r_i h_i^{-1}(\gamma_i)) \to \infty$. On the other hand, since γ_i is the shortest essential closed curve with respect to g_i, we see that $\mathrm{length}_{g_i}(h^{-1}(\gamma_i)) = \mathrm{length}_{\Sigma_i}(\gamma_i)$ is bounded. This implies that $r_i \mathrm{length}(h_i^{-1}(\gamma_i))$ is also bounded as $i \to \infty$, which is a contradiction. Thus we have proved that $i(\mu, \mu_0) = 0$.

As μ is assumed to be maximal and connected, this means that $|\mu| = |\mu_0|$. In particular μ_0 is not a simple closed curve, and we can see the sequences $\{r_i\}$ must go to 0 as $i \to \infty$. □

The next lemma, which asserts the continuity of the lengths of realized measured laminations, appeared in Thurston [28]. The following proof is based on Proposition 2.5 due to Bonahon. Soma previously suggested a possibility of such a proof.

Lemma 4.2 *Let $L\colon AH_p(S) \times \mathcal{ML}(S) \to \mathbf{R}$ be the function such that $L((\Gamma, \phi), \lambda)$ is the length of the realization of λ on a pleated surface homotopic to ϕ when such a pleated surface exists, otherwise set $L((\Gamma, \phi), \lambda) = 0$. Then L is continuous.*

Proof Let $\{(G_i, \psi_i)\} \in AH_p(S)$ be a sequence which converges to $(G', \psi') \in AH_p(S)$, and let $\{\lambda_j\}$ be measured laminations on S converging to λ'. We shall prove that L is continuous at $((G', \psi'), \lambda')$. We can take representatives for elements of the sequence so that the representations $\{\psi_i\}$ converge to ψ'. Fix a base frame $\tilde{v}$ on $\mathbf{H}^3$ and let w_i be the base frame of $\mathbf{H}^3/G_i$ which is the projection of $\tilde{v}$ by the universal covering map. Since G_i converges algebraically, the injectivity radius at the basepoint under w_i is bounded away from 0 as $i \to \infty$. By compactness of geometric topology (see Corollary 3.1.7 in Canary–Epstein–Green [9]) and the diagonal argument, we can see that for any large $r > 0$ and small $\epsilon > 0$, there exists i_0 such that for any $i > i_0$, there exists a Kleinian group H' containing G' and a $((1+\epsilon), r)$–approximate isometry $\rho_i\colon B_r(\mathbf{H}^3/G_i, w_i) \to B_r(\mathbf{H}^3/H', w')$, where B_r denotes an r–ball. (Note that the group H' may depend on i since a geometric limit exists only after taking a subsequence.)

First suppose that λ' can be realized by a pleated surface homotopic to ψ'. Then by Proposition 2.5, for any small $\delta > 0$, there exists a train track τ with a weight system ω carrying λ' and a continuous map $f\colon S \to \mathbf{H}^3/G'$ homotopic to ψ' which is adapted to a tied neighbourhood N_τ of τ such that the total curvature and the quadratic variation of angle for $f(\tau, \omega)$ are less than δ.

For a Kleinian group H' containing G', by composing the covering $q\colon \mathbf{H}^3/G' \to \mathbf{H}^3/H'$ to f, we get a map with the same property homotopic to $q \circ \psi'$. We take r and ϵ so that for any geometric limit H', the r–ball centred at the base point under w' contains the image of $q \circ f$ and so that if we pull back $q \circ f$ by a $((1+\epsilon), r)$–approximate isometry and straighten the images of branches to geodesic arcs, the image of (τ, ω) has the total curvature and the quadratic variation of angle less than 2δ. Then for $i > i_0$, there exists a map $f_i\colon S \to \mathbf{H}^3/G_i$ homotopic to ϕ_i which is adapted to τ such that $f_i(\tau, \omega)$ has total curvature and quadratic variation of angle less than 2δ. Again by Proposition 2.5, this implies that there is a neighbourhood U of λ' in $\mathcal{ML}(S)$ such that for any weighted simple closed curve γ in U, there exist $\nu_U > 0$ depending on U, $\eta_\delta > 0$, and $t_\delta < 1$ depending on δ such that $\nu_U \to 0$ as U gets smaller and $\eta_\delta \to 0, t_\delta \to 1$ as $\delta \to 0$, and the following holds. We can homotope γ so that $N_\tau \cap \gamma$ corresponds the weight system ω' (which may not satisfy the switch condition since γ may not be homotoped into N_τ) whose value at each branch differs from that of ω at most ν_U, and the closed geodesic γ_i^* homotopic to $\psi_i(\gamma)$ has a part with length $t_\delta \mathrm{length}(\gamma_i^*)$ which lies within distance η_δ from $f^i(\tau \cap \gamma)$. The same holds for f and the closed geodesic γ^* homotopic to $\psi(\gamma)$.

It follows that there is a positive real number ζ depending on ϵ, δ, U which goes to 0 as $\epsilon \to 0, \delta \to 0$ and U gets smaller remaining to be a neighbourhood of λ', such that if γ, γ' are weighted simple closed curves in U, then $|\mathrm{length}(\gamma_i^*) - \mathrm{length}(\gamma'^*)| < \zeta$, where γ'^* is the closed geodesic in $\mathbf{H}^3/G'$ homotopic to $\psi(\gamma')$. Since the set of weighted simple closed curves is dense in $\mathcal{ML}(S)$ and any realization of measured lamination can be approximated by realizations of simple closed curves, this implies our lemma in the case when λ' is realizable by a pleated surface homotopic to ψ'.

Next suppose that λ' is not realizable by a pleated surface homotopic to ψ'. This means that λ' is an ending lamination of an end of $(\mathbf{H}^3/G')_0$. By a result of Thurston in [26] (see also Lemma 4.4 in [17]), it follows that λ' is maximal and connected. In this case the alternative (i) of Proposition 2.5 holds. Hence for any small $\epsilon > 0$, there exists a train track τ carrying λ' with weight ω and a continuous map f homotopic to ψ' which is adapted to a tied neighbourhood N_τ of τ, such that λ' can be homotoped so that the length of $f(\tau,\omega)$ is less than δ. Then by the same argument as the last paragraph, there exists i_0 such that if $i > i_0$ there exists a map f_i adapted to N_τ such that $f_i(\tau,\omega)$ has length less than 2δ.

Since $\{\lambda_j\}$ converges to λ' and λ' is maximal, λ_j is carried by τ for sufficiently large j with weight ω_j whose values at branches are close to those of ω. Hence there exists j_0 such that $f_i(\tau,\omega_j)$ is less than 3δ if $j > j_0$. As the length of realization of λ_j by a pleated surface homotopic to ψ_i is less than that of $f_i(\tau,\omega_j)$, this implies our lemma in the case when λ' cannot be realized by a pleated surface homotopic to ψ'. □

The following is a well-known result of Thurston appeared in [26] and also a corollary of Lemma 2.4 due to Bonahon. Nevertheless, as its proof is not so straightforward when sequences of closed geodesics intersect Margulis tubes non-trivially, we shall prove here that Lemma 2.4 implies this lemma.

Lemma 4.3 *Let M be a hyperbolic 3–manifold. Let e be a geometrically infinite tame end of the non-cuspidal part M_0. Let λ, λ' be measured laminations on a frontier component T of a core, which faces e. Suppose that both λ and λ' are ending laminations of the end e. Then the supports of λ and λ' coincide.*

Proof Let s_j and s'_j be simple closed curves on T such that for some positive real numbers x_j and y_j, we have $x_j s_j \to \lambda$, $y_j s'_j \to \mu$ and such that the closed geodesics s_j^* homotopic to s_j and ${s'_j}^*$ homotopic to s'_j tend to the end e as $j \to \infty$. If there exists a constant $\epsilon_0 > 0$ such that neither s_j^* nor ${s'_j}^*$ intersects an ϵ_0–Margulis tube whose axis is not s_j^* or ${s'_j}^*$ itself, then we can apply Lemma 2.4 and the proof is completed.

Next suppose that for at least one of s_j^* and ${s'_j}^*$ (say s_j^*), a constant as ϵ_0 above does not exist. We shall prove that we can replace s_j with another simple closed curve to which we can apply Lemma 2.4. By assumption, there exist closed geodesics ξ_j whose lengths go to 0 and such that s_j^* intersect the ϵ_j–Margulis tube whose axis is ξ_j, where $\epsilon_j \to 0$. Let $h_j\colon (T,\sigma_j) \to M$ be a pleated surface homotopic to the inclusion whose image contains s_j^* as the image of its pleating locus, where σ_j is the hyperbolic structure on T induced by h_j. Put a base point y_i on T which is mapped into s_j^* but outside the ϵ_0–Margulis tubes by h_j. Let $h_\infty\colon ((T',\sigma_\infty),y_\infty) \to (M',y_\infty)$ be the geometric limit of $\{h_j: (T', y_j) \to (M, h_j(y_j))\}$ after taking a subsequence, where T' is an incompressible subsurface in T. We shall first show that T' cannot be the entire of T.

Suppose that $T' = T$ on the contrary. Let l be the geodesic lamination on (T, σ_∞) which is the geometric limit of the closed geodesic on (T, σ_j) corresponding to s_j^* as $j \to \infty$. Since l cannot approach to a cusp (as $T = T'$), it is compact. Therefore we can take a point in the intersection of s_j^* and the ϵ_j–Margulis tube which converges to a point x on $h_\infty(l)$ associated with the geometric convergence of $\{h_j\}$ to h_∞ as $j \to \infty$. Then for any small ϵ, there is an essential closed curve passing x with length less than ϵ which can be obtained by pushing forward by an approximate isometry an essential loop intersecting s_j^* of length less than ϵ_j for sufficiently large j. This is a contradiction.

Thus there is an extra cusp for h_∞. Let c be a simple closed curve on T' representing an extra cusp. Let $\overline{\rho}_j: B_{r_j}((T, \sigma_j), y_j) \to B_{r_j}((T', \sigma_\infty), y_\infty)$ be an approximate isometry associated with the geometric convergence of $\{h_j\}$ to h_∞. Let c_j be a simple closed curve on T which is homotopic to $\overline{\rho}_j^{-1}(c)$. Let l' be a measured lamination to which $\{r_j c_j\}$ converges for some positive real numbers r_j. Let c_j^+ be the closed geodesic on (T, σ_j) homotopic to c_j. Let α be a measured lamination to whose projective class the hyperbolic structures σ_j converge, after passing through a subsequence if necessary. Then as $\mathrm{length}_{\sigma_j}(c_j^+)$ goes to 0 as $j \to \infty$, we have $i(\alpha, l') = 0$ by Lemma 3.4 in [17]. By the same reason, considering $\{s_j\}$, we have $i(\lambda, \alpha) = 0$. Since λ is maximal and connected, these imply that the supports of λ and l' coincide. In particular, l' is an ending lamination of the end for which λ is an ending lamination.

Because the length of the closed geodesic c_j^+ goes to 0 as $j \to \infty$, the closed geodesic homotopic to $h_j(c_j)$, whose length is at most the length of c_j^+, must be the axis of an ϵ_0–Margulis tube for sufficiently large j. Thus we can replace s_j with c_j, and and by the same fashion, we can replace s'_j with another simple closed curve if necessary. We can apply Lemma 2.4 for such simple closed curves. □

5 Proof of the main theorem

We shall complete the proof of Theorem 3.1 in this section. Recall that under the assumption for the reductio ad absurdum, we have (G, ψ) which is the algebraic limit of $\{(\Gamma_i, \phi_i)\}$.

Lemma 5.1 *In the situation of Theorem 3.1, the non-cuspidal part* $(\mathbf{H}^3/G)_0$ *of the hyperbolic 3–manifold* $\mathbf{H}^3/G$ *has a geometrically infinite tame end for which* $\psi(\mu)$ *represents an ending lamination.*

Proof Suppose first that the end e^i of $(\mathbf{H}^3/\Gamma_i)_0$ corresponding to the first factor of the end invariant is geometrically finite. Then by Lemma 4.1, there exists a sequence of weighted simple closed curves $r_i\gamma_i$ on S converging to μ such that for the closed geodesic γ_i^* in $\mathbf{H}^3/\Gamma_i$ homotopic to $\phi_i(\gamma_i)$, we have $r_i\mathrm{length}(\gamma_i^*) \to 0$. By the continuity of length function L on $AH_p(S) \times \mathcal{ML}(S)$ (Lemma 4.2), we have $L((G,\psi),\mu)$ is 0, which means that μ cannot be realized by a pleated surface homotopic to ψ. As we assumed that μ is maximal and connected, there must be a geometrically infinite tame end of $(\mathbf{H}^3/G)_0$ with ending lamination represented by $\psi(\mu)$. This last fact, originally due to Thurston, can be proved using Bonahon's result: by Proposition 2.5, if $L((G,\psi),\mu) = 0$ and μ is maximal and connected, then for any sequence of simple closed curves δ_j on S whose projective classes converge to that of μ, the closed geodesics δ_j^* homotopic to $\psi(\delta_j)$ tend to an end of $(\mathbf{H}^3/G)_0$. This means that $\psi(\mu)$ is an ending lamination for a geometrically tame end of $(\mathbf{H}^3/G)_0$.

Next suppose that the end e^i is geometrically infinite. Then χ_i is represented by a measured lamination μ_i which represents an ending lamination of e^i, hence $L((\Gamma_i,\phi_i),\mu_i) = 0$. We can assume that μ_i lies on the unit ball of $\mathcal{ML}(S)$ with respect to the metric induced from some fixed hyperbolic structure on S. Then μ_i converges to a scalar multiple of μ since we assumed that $\chi_i = [\mu_i]$ converges to $[\mu]$. By the continuity of L, this implies that $L((G,\psi),\mu) = 0$ and that $\psi(\mu)$ represents an ending lamination for $(\mathbf{H}^3/G)_0$. □

We shall denote the end in Lemma 5.1, for which $\psi(\mu)$ represents an ending lamination, by e_μ.

Recall that $q\colon \mathbf{H}^3/G \to \mathbf{H}^3/\Gamma_\infty$ is a covering associated with the inclusion. Now by Proposition 2.6, the end e_μ has a neighbourhood E_μ such that $q|E_\mu$ is a proper embedding. Since μ is maximal and connected, the end e_μ has a neighbourhood homeomorphic to $S_0 \times \mathbf{R}$, where S_0 is the non-cuspidal part of S Hence by refining E_μ, we can assume that E_μ is also homeomorphic to $S_0 \times \mathbf{R}$.

Lemma 5.2 *We can take an embedding* $f'\colon S_0 \to E_\mu$ *homotopic to* $\psi|S_0$ *whose image is contained in the convex core such that for any pleated surface* $g\colon S \to \mathbf{H}^3/\Gamma_\infty$ *homotopic to* $q \circ \psi$ *with non-empty intersection with* $qf'(S_0)$*, we have* $g(S) \cap (\mathbf{H}^3/\Gamma_\infty)_0 \subset q(E_\mu)$*.*

Proof Fix a constant $\epsilon_0 > 0$ less than the Margulis constant. There exists a constant K such that for any hyperbolic metric on S, the diameter of S modulo the ϵ_0–thin part is bounded above by K. (This can be easily seen by considering the moduli of S.)

Note that since the end e_μ is geometrically infinite, it has a neighbourhood contained in the convex core. Take $t \in \mathbf{R}$ large enough so that $S_0 \times \{t\} \subset E_\mu$ is contained in the convex core and the distance from $S_0 \times \{t\}$ to the frontier of E_μ in $(\mathbf{H}^3/G)_0$ modulo the ϵ_0–thin part is greater than $2K$. Choose f' homotopic $\psi|S_0$ so that its image is $S_0 \times \{t\}$. Then the distance between $qf'(S_0)$ and the frontier of $q(E_\mu)$ modulo the ϵ_0–thin part is also greater than $2K$. Suppose that a pleated surface $g\colon S \to \mathbf{H}^3/\Gamma_\infty$ touches $qf'(S_0)$. Then $g(S)$ cannot meet the frontier of $q(E_\mu)$ since the ϵ_0–thin part of S' with respect to the hyperbolic structure induced by g is mapped into the ϵ_0–thin part of $\mathbf{H}^3/\Gamma_\infty$, and any path on $g(S)$ has length less than K modulo the ϵ_0–thin part of $\mathbf{H}^3/\Gamma_\infty$. Also it is impossible for $g(S)$ to go into the cuspidal part of $\mathbf{H}^3/\Gamma_\infty$ and come back to the non-cuspidal part since the intersection of $g(S)$ with the cuspidal part of $\mathbf{H}^3/\Gamma_\infty$ is contained in a neighbourhood of cusps of $g(S)$. This means that without meeting the frontier of $q(E_\mu)$, the pleated surface $g(S)$ cannot go outside $q(E_\mu)$ in $(\mathbf{H}^3/\Gamma_\infty)_0$. Thus the intersection of such a pleated surface with $(\mathbf{H}^3/\Gamma_\infty)_0$ must be contained in $q(E_\mu)$. □

We denote $q \circ f'$ by $f_\infty\colon S_0 \to (\mathbf{H}^3/\Gamma_\infty)_0$. Pulling back this embedding f_∞ by an approximate isometry ρ_i for sufficiently large i, we get an embedding $f_i\colon S_0 \to (\mathbf{H}^3/\Gamma_i)_0$. Since f_∞ comes from the surface homotopic to ψ in the algebraic limit, for sufficiently large i, the surface f_i is homotopic to ϕ_i.

Consider the case when $(\mathbf{H}^3/\Gamma_i)_0$ has a geometrically finite end; that is Γ_i is either quasi-Fuchsian or a totally degenerate b–group. As in the previous section, let Σ_i be a boundary components of the convex core of $\mathbf{H}^3/\Gamma_i$, and let $h_i\colon S \to \Sigma_i$ be a homeomorphism homotopic to ϕ_i. The homeomorphisms h_i can also be regarded as pleated surfaces in $\mathbf{H}^3/\Gamma_i$. Let μ_i be the bending locus of h_i, to which we give transverse measures with full support so that μ_i should converge to measured laminations as $i \to \infty$ after taking subsequences. (Since the unit sphere of the measured lamination space is compact, this is always possible. Also if h_i happens to be totally geodesic, we can set μ_i to be any measured lamination on S.)

Lemma 5.3 *Suppose that Γ_i is either quasi-Fuchsian or a totally degenerate* b*–group as above. The sequence of the measured laminations $\{\mu_i\}$ converges to a measured lamination with the same support as μ after taking subsequences.*

Proof Let μ' be a limit of $\{\mu_i\}$ after taking a subsequence. If $i(\mu, \mu') = 0$, we have nothing to prove any more because μ is maximal and connected. Now assume that $i(\mu', \mu) \neq 0$. Then, by the fact that the marked hyperbolic structure on Σ_i converges to $[\mu]$ as $i \to \infty$ and Lemma 3.4 in [17], we have $\mathrm{length}_{\Sigma_i}(\mu_i) \to \infty$. On the other hand, by the continuity of the length function L on $AH_p(S) \times \mathcal{ML}(S)$ (Lemma 4.2), we have

$$\mathrm{length}_{\mathbf{H}^3/\Gamma_i}(\phi_i(\mu_i)) = L((\Gamma_i, \phi_i), \mu_i) \to L((G, \psi), \mu') = \mathrm{length}_{\mathbf{H}^3/G}(\psi(\mu')) < \infty$$

where $\mathrm{length}_{\mathbf{H}^3/\Gamma_i}(\phi_i(\mu_i))$ denotes the length of the image of μ_i realized by pleated surface homotopic to ϕ_i etc. Since μ_i is mapped by h_i into the bending locus of Σ_i, it is realized by h_i, hence $\mathrm{length}_{\Sigma_i}(\mu_i) = \mathrm{length}_{\mathbf{H}^3/\Gamma_i}(\phi_i(\mu_i))$. This is a contradiction. □

Now we assume that Γ_i is quasi-Fuchsian. Then there are two boundary components Σ_i, Σ_i' of the convex core of $\mathbf{H}^3/\Gamma_i$, and homeomorphisms $h_i\colon S \to \Sigma_i \subset \mathbf{H}^3/\Gamma_i$ and $h_i'\colon S \to \Sigma_i' \subset \mathbf{H}^3/\Gamma_i$ homotopic to ϕ_i which are regarded as pleated surfaces. We have two measured laminations of unit length μ_i and μ_i' whose supports are the bending loci of h_i and h_i'. By Lemma 5.3, the sequence of the measured laminations $\{\mu_i\}$ converges to a measured lamination μ' and $\{\mu_i'\}$ converges to a measured lamination μ'' such that $|\mu'| = |\mu''| = |\mu|$. As the space of transverse measures on a geodesic lamination is connected (or more strongly, convex with respect to the natural PL structure), we can join μ' and μ'' by an arc $\alpha\colon I = [0,1] \to \mathcal{ML}(S)$ such that $|\alpha(t)| = |\mu|$. Join μ_i and μ_i' by an arc $\alpha_i\colon I \to \mathcal{ML}(S)$ which converges to the arc α joining μ' and μ''.

Next suppose that Γ_i is a totally degenerate b–group. We can assume without loss of generality that the first factor χ_i of the end invariant represents an ending lamination and the second υ_i a conformal structure. Then we have a pleated surface $h_i\colon S \to \Sigma_i$ homotopic to ϕ_i whose image is the boundary of the convex core. Let μ_i be a measured lamination of the unit length whose support is equal to that of the bending locus as before. Again by Lemma 5.3, we see that $\{\mu_i\}$ converges to a measured lamination μ' with the same support as μ. Let μ_i' be a measured lamination of the unit length representing the class χ_i. By the assumption of Theorem 3.1, the sequence $\{\mu_i'\}$ converges to a measured lamination μ'' with the same support as μ. As in the case of quasi-Fuchsian group, we join μ' and μ'' by an arc α, and then join μ_i and μ_i' by an arc α_i which does not pass an ending lamination for $\mathbf{H}^3/\Gamma_i$ at the interior so that it will converge to α uniformly.

In the case when Γ_i is a totally doubly degenerate group, both χ_i and υ_i are represented by ending laminations. Let μ_i representing χ_i and μ_i' representing

υ_i be measured laminations of the unit length. Then by assumption, $\{\mu_i\}$ and $\{\mu_i'\}$ converge to measured laminations μ' and μ'' with the same support as μ. As before we join μ' and μ'' by an arc α, and μ_i, μ_i' by α_i which does not pass an ending lamination of $\mathbf{H}^3/\Gamma_i$ at the interior so that $\{\alpha_i\}$ converges to α.

Next we shall consider constructing for each i a homotopy consisting of pleated surfaces and negatively curved surfaces in $\mathbf{H}^3/\Gamma_i$ as in Thurston [26]. What we shall have is a homotopy between h_i and h_i' in the case when Γ_i is quasi-Fuchsian; a half-open homotopy $\hat{H}_i$: $S \times [0,1) \to \mathbf{H}^3/\Gamma_i$ such that $\hat{H}_i(S \times \{t\})$ tends to the unique geometrically infinite end as $r \to 1$ in the case when Γ_i is a totally degenerate b–group; and an open homotopy $\hat{H}_i$: $S \times (0,1) \to \mathbf{H}^3/\Gamma_i$ such that $\hat{H}_i(S \times \{t\})$ tends to one end as $t \to 0$ and to the other as $t \to 1$ in the case when Γ_i is a totally doubly degenerate group. To construct such a homotopy, we need the notion of rational depth for measured laminations due to Thurston. An alternative approach to construct such a homotopy using singular hyperbolic triangulated surfaces can be found in Canary [8].

A train track τ is called birecurrent when the following two conditions are satisfied. (This definition is due to Penner–Harer [23].) (1) The τ supports a weight system which is positive on each branch b of τ. (2) For each branch b of τ, there exists a multiple curve σ (ie a disjoint union of non-homotopic essential simple closed curves) transverse to τ which intersects b such that $S - \tau - \sigma$ has no bigon component.

A birecurrent train track which is not a proper sub-train track of another birecurrent train track is said to be complete.

Any measured lamination is carried by some complete train track. (Refer to Corollary 1.7.6 in [23].) The weight systems on a complete train track gives rise to a coordinate system in the measured lamination space. (See Lemma 3.1.2 in [23].) The rational depth of a measured lamination is defined to be the dimension of the rational vector space of linear functions with rational coefficients vanishing on the measured lamination with respect to a coordinate system associated with a complete train track carrying the measured lamination. This definition is independent of the choice of a coordinate system since functions corresponding to coordinate changes are linear functions with rational coefficients. The set of measured laminations with rational depth n has codimension n locally at regular points. In particular a generic arc in the measured lamination space does not pass a measured lamination with rational depth more than 1.

Now perturb α and α_i to a piece-wise linear path with respect to the PL structure of $\mathcal{ML}(S)$ determined by complete train tracks fixing the endpoints so that for each $t \in I$, the measured lamination $\alpha_i(t)$ is not an ending lamination and has rational depth 0 or 1, and that for each i there exist only countably many values t for which $\alpha_i(t)$ has rational depth 1.

The following lemma was first proved in section 9 in Thurston [26]. A fairly detailed proof can be found there.

Lemma 5.4 *If a measured lamination has rational depth 0, then each component of its complement is either an ideal triangle or a once-punctured monogon except when S is a once-punctured torus. In the case when S is a once-punctured torus, each component of the complement is an ideal once-punctured bigon. A pleated surface $f\colon S \to M$ realizing a measured lamination ζ of rational depth 0 is unique among the maps in the homotopy class, and every sequence of homotopic pleated surfaces realizing measured laminations converging to ζ converges to the pleated surface realizing ζ.*

Proof First we shall show that each complementary region of a measured lamination of rational depth 0 is either an ideal triangle or an ideal once-punctured monogon unless S is a once-punctured torus.

Suppose that S is not a once-punctured torus and that a measured lamination ζ has a complementary region which is neither an ideal triangle nor an ideal once-punctured monogon. Then, we can construct a birecurrent train track τ carrying ζ whose complement has a component which is neither a triangle nor a once-punctured monogon. (Refer to section 1.7 in [23].)

A birecurrent train track is maximal if and only if every component of its complement is either a triangle or a once-punctured monogon, and non-maximal birecurrent train track is a sub-train track of a complete train track. (Theorem 1.3.6 in [23].) Hence there exists a complete train track τ' containing τ as a proper sub-train track. Since there is a branch of τ' through which ζ does not pass after homotoping ζ so that it is carried by τ', it follows that with respect to the coordinate system corresponding to τ', the measured lamination ζ has rational depth at least 1.

In the case when S is a once-punctured torus, again Theorem 1.3.6 in [23] says that a birecurrent train track is maximal if and only if its (unique) complementary region is a once-punctured bigon. Thus the same argument as above also implies our claim in the case of once-punctured torus.

Next we shall show the uniqueness of realization of a measured lamination of rational depth 0. Let f, g be two pleated surfaces realizing a measured lamination ζ of depth 0. The pleated surfaces f, g induce hyperbolic metrics m_1, m_2 respectively on S. (These may differ as we do not know if f and g coincide.) The measured lamination ζ is homotopic to measured geodesic laminations ζ_1 with respect to m_1 and ζ_2 with respect to m_2. Consider the universal covers $p_1\colon \mathbf{H}^2 \to (S, m_1)$ and $p_2\colon \mathbf{H}^2 \to (S, m_2)$. Let $\tilde{\zeta}_1$ be $p_1^{-1}(\zeta_1)$ and let $\tilde{\zeta}_2$ be $p_2^{-1}(\zeta_2)$.

The pleated surfaces f, g are lifted to maps $\tilde{f}, \tilde{g}\colon \mathbf{H}^2 \to \mathbf{H}^3$. Since ζ has compact support, there is a homeomorphism from S to S homotopic to the identity which takes ζ_1 to ζ_2 and is equal to the identity near cusps. Also for a homotopy between f and g, the distance moved by the homotopy on the compact set ζ has an upper bound. These imply that for each leaf l of $\tilde{\zeta}$ the images of the corresponding leaves l_1 of $\tilde{\zeta}_1$ by $\tilde{f}$ and l_2 of $\tilde{\zeta}_2$ by $\tilde{g}$ are within a bounded distance. Since both $\tilde{f}(l_1)$ and $\tilde{g}(l_2)$ are geodesics in $\mathbf{H}^3$ and two geodesics lying in bounded distance coincides in $\mathbf{H}^3$, these two images must coincide. Hence we have a map $q\colon \mathbf{H}^2 \to \mathbf{H}^2$ equivariant with respect to the action of $\pi_1(S)$ with the property $\tilde{f}|\tilde{\zeta}_1 = \tilde{g} \circ q|\tilde{\zeta}_1$ which maps $\tilde{\zeta}_1$ to $\tilde{\zeta}_2$ isometrically.

It remains to prove that q extends to an equivariant isometry $\overline{q}$ of $\mathbf{H}^2$ with the property $\tilde{f} = \tilde{g} \circ \overline{q}$. Since ζ has rational depth 0, each of its complementary regions is either an ideal triangle or an ideal once-punctured monogon unless S is a once-punctured torus. An ideal triangle on S is lifted to that on $\mathbf{H}^2$. Since the three sides of the triangle are mapped to geodesics by $\tilde{f}$ or $\tilde{g}$, the triangle must be mapped totally geodesically. Considering that there is only one isometry type of ideal triangles, we can see that this implies q can be extended to ideal triangle complementary components without problem.

For complementary regions which are ideal once-punctured monogon, or ideal once-punctured bigon in the case when S is a once-punctured torus, we have to use the fact that pleated surfaces are totally geodesics near cusps. (This is proved in Proposition 9.5.5 in Thurston [26].) Once we know this, we can subdivide such regions into ideal triangles by adding geodesics tending to cusps on S, which are mapped to geodesics by f or g. Since each cusp of S is mapped to the same cusp of M by f and g, we can arrange them so that the lifts of these added geodesics should be compatible with q. Hence by extending the map finally to ideal triangles, we get a map $\overline{q}$ as we wanted.

Finally let us prove the last sentence of our lemma. Let ξ_j be measured laminations converging to ζ, and f_j a pleated surface realizing ξ_j. Since ζ can be realized by a pleated surface, the alternative (2) of Proposition 2.5 should be valid for ζ. We shall show that if there is no compact set in M which intersects all the images of f_j, then we can see that the alternative (2) of Proposition 2.5 fails to hold for ζ.

Suppose that the alternative (2) of Proposition 2.5 holds for ζ. Then for any $\delta > 0$ and $t < 1$, there exist a map f_δ: $S \to M$ homotopic to f such that for any simple closed curve γ whose projective class is close to that of ζ, the closed geodesic γ^* homotopic to $f_\delta(\gamma)$ has a part of length at least $t\mathrm{length}(\gamma^*)$ which is contained in the δ–neighbourhood of $f_\delta(\zeta)$. Note that as $\delta \to 0$, this map f_δ converges to a pleated surface realizing ζ, which must be equal to f. (Refer to [20] for a further explanation.) On the other hand, since ξ_j is also realized by a pleated surface homotopic to f, the alternative (2) holds also for ξ_j. Then we have a surface f_j^δ with the same property as f_δ above replacing ζ with ζ_j. Since we assumed that f_j tends to an end of M, we can have surfaces $f_j^{\delta_j}$ going to an end and a simple closed curve γ_j whose projective class is close to that of ζ_j such that a large part of the closed geodesic γ_j^* is contained in the δ_j–neighbourhood of $f_j^{\delta_j}(S)$. This is a contradiction because γ_j^* must also have a large part contained in the δ–neighbourhood of $f_\delta(S)$ which remains in a neighbourhood of $f(S)$.

Thus the surfaces $f_j(S)$ remain to intersect a compact set, hence converge to a pleated surface g homotopic to f uniformly on any compact set of S. (Theorem 5.2.18 in Canary–Epstein–Green [9].) The pleated surface g realizes a geodesic lamination ζ_∞ which is a geometric limit of $\{\zeta_j\}$ regarded as geodesic laminations forgetting the transverse measures. It is known that ζ_∞ contains the support of ζ. (See for example Lemma 5.3.2. in [9].) Thus g also realizes ζ, and by the uniqueness of such pleated surfaces proved above, we see that $f = g$, which means that $\{f_j\}$ converges to f uniformly on any compact set of S. □

Now let H_i: $S \times I, S \times [0,1), S \times (0,1) \to \mathbf{H}^3/\Gamma_i$ (depending on the type of Γ_i; a quasi-Fuchsian group or a totally degenerate b–group or a totally doubly degenerate) be a map such that for each $t \in I, [0,1), (0,1)$, the map $H_i(\ ,t)$: $S \to \mathbf{H}^3/\Gamma_i$ is a pleated surface realizing $\alpha_i(t)$. Then H_i is continuous with respect to t by Lemma 5.4 except at values t where $\alpha(t)$ has rational depth 1, which are countably many. Since we made α_j piece-wise linear we can see that the right and left limits exists even at t where $\alpha_i(t)$ has depth 1. (This can be seen by considering a complete train track giving a coordinate near $\alpha_i(t)$.) As was shown in Thurston [26], (see section 4 in [20] for an explanation), at such a point of discontinuity t, the left limit and the right limit differ only in a complementary region R of $\alpha_i(t)$ which is either an ideal quadrilateral or an ideal once-punctured bigon except when S is a once-punctured torus. Then we can modify H_i to a continuous homotopy $\hat{H}_i$ by interpolating negatively curved surfaces realizing $\alpha_i(t_0)$ between $\lim_{t\to t_0-0} H_i(\ ,t)$ and $\lim_{t\to t_0+0} H_i(\ ,t)$ at each t_0 where $\alpha_i(t_0)$ has rational depth 1 as in Thurston [26]. These negatively curved surface coincide with the left and the right limit outside R where the right and left limit differ.

We need to prove that a family of surfaces thus obtained is continuous with respect to the parameter. The only case that we have to take care of is when the values t_k for which $\alpha_i(t_k)$ has depth 1 accumulates to a point $t_0 \in I$. The negatively curved surfaces interpolated at t_k have the same image as a pleated surface realizing $\alpha(t_k)$ outside a complementary region R_k. The image of R_k by the left limit pleated surface and the right limit surface bound an ideal tetrahedron if R_k is an ideal quadrilateral or a solid torus with cusps if R_k is an ideal once-punctured bigon in the case when S is not a once-punctured torus. The form of R_k gets thinner and thinner as $k \to \infty$ since t_k accumulates. (This can again be seen by considering a coordinate chart given by a complete train track.) This implies that the trajectories of the homotopy between the left limit and the right limit, which are contained in the ideal tetrahedron or the solid torus have length going to 0 as $k \to \infty$. Even in the case when S is a once-punctured torus, a similar argument can work although we need to take more cases into account. Thus we can see that $\hat{H}(\ ,t)$ is continuous with respect t even at the point t_0 to which depth-1 points t_k accumulate.

Lemma 5.5 *For each i, there is a pleated surface $k_i\colon S \to \mathbf{H}^3/G$ homotopic to ϕ_i touching $f_i(S_0)$ which realizes a measured lamination $\overline{\mu}_i$ such that $\{\overline{\mu}_i\}$ converges to a measured lamination $\overline{\mu}$ with the same support as μ after taking a subsequence. Moreover we can choose k_i so that $k_i^{-1}(f_i(S_0))$ contains an essential component relative to cusps.*

Remark 1 Although the last sentence of this lemma is not necessary for our purpose now, it will be used for our forthcoming work in [21]. Also Canary's result on filling a convex core by pleated surfaces in [8] will suffice to prove only the former part of our lemma.

Proof Recall that $f'(S_0)$ is contained in the convex core of $\mathbf{H}^3/G$. Then we can assume that $f_i(S_0)$ is also contained in the convex core of $\mathbf{H}^3/\Gamma_i$. It follows that $f_i(S_0)$ is contained in the image of $\hat{H}_i$.

By perturbing $f_i(S_0)$ if necessary, we can assume that $\hat{H}_i$ is transverse to $f_i(S_0)$ and that $\hat{H}_i^{-1}(f_i(S_0))$ is an embedded surface in $S \times (0,1)$. Let F

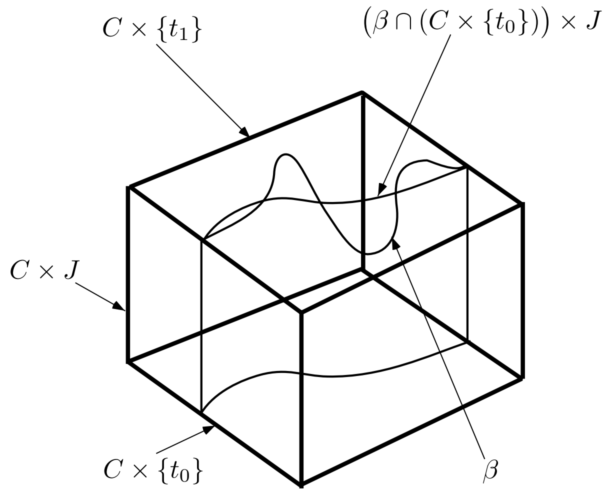
be a component of $\hat{H}_i^{-1}(f_i(S_0))$ which separates $S \times \{0\}$ from $S \times \{1\}$. It is easy to see such a component exists because in the case when Γ_i is quasi-Fuchsian, Σ_i and Σ_i' lie in different components of $\mathbf{H}^3/\Gamma_i - f_i(S_0)$, in the case when Γ_i is a b–group, $f_i(S_0)$ separates a geometrically infinite end from Σ_i, and in the case when Γ_i is doubly degenerate, $f_i(S_0)$ separates two ends. Then $\pi_1(F)$ is mapped onto $\pi_1(S)$ by the homomorphism induced by inclusion, hence $(\hat{H}_i|F)_\#\colon \pi_1(F) \to \pi_1(f_i(S_0))$ is surjective.

We can assume that for each $t \in I$, the intersection $(S \times \{t\}) \cap F$ is at most one dimensional by perturbing $f_i(S_0)$ again if necessary. Then there exists $t_0 \in I$ such that $(S \times \{t_0\}) \cap F$ contains a simple closed curve K which represents a non-trivial element of $\pi_1(S)$ relatively to the punctures of S. If $\hat{H}_i(\ ,t_0)$ is a pleated surface, we simply let k_i be $\hat{H}_i(\ ,t_0)$. In this case, $k_i^{-1}(f_i(S_0))$ contains K, which is essential relatively to the cusps. The pleated surface k_i realizes a measured lamination $\alpha_i^{t_0}$ in the image of α_i, which we let be $\overline{\mu}_i$. The measured lamination $\overline{\mu}_i = \alpha_i^{t_0}$ converges after taking a subsequence to a measured lamination in $\alpha(I)$ hence with the same support as μ.

Suppose that $\hat{H}_i(\ ,t_0)$ is an interpolated negatively curved surface. Let $\alpha_i^{t_0}$ be the measured lamination of rational depth 1 realized by $\hat{H}_i(\ ,t_0)$. We can assume that $\hat{H}_i(\ ,t_0)(\alpha_i^{t_0})$ is transverse to $f_i(S_0)$ again by a perturbation of $f_i(S_0)$ without changing the homotopy class of K. Let $J = [t_0, t_1] \subset I$ be an interval such that $\hat{H}_i(\ ,[t_0,t_1))$ are interpolated negatively curved surfaces and $\hat{H}_i(\ ,t_1)$ is a pleated surface realizing $\alpha_i^{t_0}$. Let C be a component of the complement of $\alpha_i^{t_0}$ which is not an ideal triangle. Since $\alpha_i^{t_0}$ has rational depth 1, such a component is unique and every simple closed curve in C is either represents a cusp or homotopic to $\mathrm{Fr}C$.

On the other hand, by the construction of interpolated surfaces, $\hat{H}_i|\mathrm{FrC} \times J$ is constant with respect to $t \in J$. If $C \times \{t_0\}$ does not intersect K, the pleated surface $\hat{H}_i(\ ,t_1) \cap F$ contains a simple closed curve homotopic to K, and we can let $\hat{H}_i(\ ,t_1)$ be k_i. Suppose that $C \times \{t_0\}$ intersects K.

First consider the case when S is not a once-punctured torus. Then C is either simply connected or an ideal once-punctured monogon. Consider a component β of $(C \times J) \cap F$ intersecting K. Since K does not represent a cusp, each component of $\beta \cap (C \times \{t_0\}) \cap K$ must be an open arc. Since $\hat{H}_i|\mathrm{Fr}C \times J$ is constant with respect to $t \in J$, the component β must be isotopic to $\{\beta \cap (C \times \{t_0\})\} \times J$ fixing $\overline{\beta} \cap (\mathrm{Fr}C \times J)$. This implies that there exists a component K' of $S \times \{t_1\} \cap F$ such that $\hat{H}_i(K', t_1)$ is homotopic in $\mathbf{H}^3/\Gamma_i$ to $\hat{H}_i(K, t_0)$ on $f_i(S_0)$. Hence by letting $\hat{H}_i(\ ,t_1)$ be k_i, we get a surface as we wanted.



Next suppose that S is a once-punctured torus. The only case to which the argument above cannot be applied is one when C is a once-punctured open annulus and K is contained in $C \times \{t_0\}$. By isotoping $f_i(S_0)$ if necessary we can assume that all the components of $(C \times J) \cap F$ are annuli. Still there is a possibility that the component of $(C \times J) \cap F$ containing K is an annulus which is parallel into $C \times \{t_0\}$, and our argument above would break down. If there is another essential (ie incompressible and not boundary-parallel, where we regard $C \times \partial J$ as the boundary,) component of $(C \times J) \cap F$, then we can retake K so that K lies on its boundary and our argument above can be applied. Suppose that all the components are inessential. Then consider another interval $J' = [t_2, t_0] \subset I$, such that $\hat{H}_i(\ ,t)$ is an interpolated surface if $t \in (t_2, t_0]$ and $\hat{H}_i(\ ,t_2)$ is a pleated surface realizing $\alpha_i^{t_0}$. Again we can assume that all the components of $(C \times J') \cap F$ are annuli. Then some component of $(C \times J') \cap F$ is essential because otherwise F cannot be a surface separating $S \times \{0\}$ from $S \times \{1\}$. Hence by the argument as before, retaking K, we can assume that the component of $(S \times J') \cap F$ containing K intersects $S \times \{t_2\}$ by a simple closed curves homotopic to K.

Thus in either case, we can get a pleated surface k_i realizing $\alpha_i^{t_0}$, which is either $\hat{H}_i(\ ,t_1)$ or $\hat{H}_i(\ ,t_2)$, and which intersects $f_i(S_0)$ so that the inverse image of $f_i(S_0)$ has a non-contractible component that is not homotopic to a cusp. □

Proof of Theorem 3.1 Consider a geometric limit $k_\infty\colon S' \to \mathbf{H}^3/\Gamma_\infty$ of the sequence of pleated surfaces $k_i\colon S \to \mathbf{H}^3/\Gamma_i$ constructed above. (Here S' is an open incompressible surface on S.) By construction, $k_\infty(S)$ intersects $f_\infty(S_0)$.

Suppose that S' is not equal to S. Then there is a frontier component c of S' on S which does not represent a cusp of S. Note that we can apply the same argument as Lemma 5.2, and prove that $k_\infty(S)$ does not meet the frontier of $q(E_\mu)$. Now since $k_\infty(c)$ is homotopic to a cusp component of $\mathbf{H}^3/\Gamma_\infty$ which can be reached from $q(E_\mu)$, it is homotopic to the image of a cusp of S by f_∞. By pulling back a homotopy by an approximate isometry, this implies that $k_i(c)$ is homotopic to the image of a cusp by f_i. Since both k_i and f_i are homotopic to ϕ_i, this means that c is homotopic to a cusp of S. This is a contradiction.

Thus S' must be equal to S, and we have a limit pleated surface $k_\infty\colon S \to \mathbf{H}^3/\Gamma_\infty$ touching $f_\infty(S_0)$. By Lemma 5.2, we see that $k_\infty(S) \cap (\mathbf{H}^3/\Gamma_\infty)_0$ is contained in $q(E_\mu)$. Therefore k_∞ can be lifted to a pleated surface $k'\colon S \to \mathbf{H}^3/G$ whose intersection with $(\mathbf{H}^3/G)_0$ is contained in E_μ.

Now since k_i is homotopic to the pull-back of $k_\infty = q \circ k'$ by an approximate isometry for sufficiently large i, and k_i is homotopic to ϕ_i, we see that k' must be homotopic to ψ. As k_i realizes $\overline{\mu}_i$ and $\{\overline{\mu}_i\}$ converges to $\overline{\mu}$, the pleated surfaces k_∞ and k' realize $\overline{\mu}$. As $|\overline{\mu}| = |\mu|$, by changing the transverse measure, μ can also be realized by k'. On the other hand, by Lemma 5.1, $\psi(\mu)$ is an ending lamination hence μ cannot be realized by a pleated surface homotopic to ψ. This is a contradiction. Thus we have completed the proof of Theorem 3.1. □

6 Strong convergence of surface groups

In Theorems 9.2, 9.6.1 in Thurston [26], it is stated and roughly proved that if a sequence of Kleinian groups, which are isomorphic to a freely indecomposable Kleinian group (ie satisfying the condition $(*)$ introduced by Bonahon) without accidental parabolics preserving the parabolicity, converges algebraically to a Kleinian group without accidental parabolic elements, then the convergence is strong (ie the geometric limit coincides with the algebraic limit.) (See also Canary [8].) We gave its detailed proof in Ohshika [19] except for the case when the Kleinian group is algebraically isomorphic to a surface group. The reason why we did not include the case of surface group there is that it would necessitate to prove that for a convergent sequence, the hyperbolic structures on the two boundary components cannot degenerate to the same point in the Thurston boundary. As this is proved in Theorem 3.1, we can give the proof for the case of surface group here. Let (Γ_i, ϕ_i) be a Kleinian group without accidental parabolic elements with isomorphism $\phi_i\colon \pi_1(S) \to \Gamma_i \subset PSL_2\mathbf{C}$ for a hyperbolic surface of finite area S. Thurston's original proof in [26] in this case consists of proving that the projectivized bending laminations of two boundary components of the convex cores of $\mathbf{H}^3/\Gamma_i$ cannot converge to projective lamination with the same support. This is exactly the argument on which our proofs of the main theorems are based.

Corollary 6.1 *Let S be a hyperbolic surface of finite area. Let (Γ_i, ϕ_i) be a Kleinian group without accidental parabolic elements with isomorphism $\phi_i\colon \pi_1(S) \to \Gamma_i \subset PSL_2\mathbf{C}$. Suppose that $\{(\Gamma_i, \phi_i)\}$ converges algebraically to a Kleinian group (G, ψ) without accidental parabolic elements. Then G is also the geometric limit of $\{\Gamma_i\}$. In other words, $\{\Gamma_i\}$ converges strongly to G.*

Proof We have only to prove that every subsequence of $\{(\Gamma_i, \phi_i)\}$ has a subsequence which converges strongly to (G, ψ). Since a subsequence of $\{(\Gamma_i, \phi_i)\}$ satisfies the condition of Corollary 6.1, we only need to show that $\{(\Gamma_i, \phi_i)\}$ in the statement of the corollary has a subsequence strongly converging to (G, ψ).

By taking a subsequence, we can assume that all of the $\{(\Gamma_i, \phi_i)\}$ are either quasi-Fuchsian or totally degenerate groups or totally doubly degenerate groups, and that $\{\Gamma_i\}$ converges geometrically to a Kleinian group Γ_∞. Suppose first that all of the $\{\Gamma_i\}$ are quasi-Fuchsian. Let $(m_i, n_i) \in \mathcal{T}(S) \times \mathcal{T}(S)$ be $Q((\Gamma_i, \phi_i))$. If both $\{m_i\}$ and $\{n_i\}$ converge inside the Teichmüller space (after taking a subsequence), $\{(\Gamma_i, \phi_i)\}$ converges to a quasi-Fuchsian group strongly as is well known. (See for example Jørgensen–Marden [13].) Assume that one of $\{m_i\}$ and $\{n_i\}$, say $\{m_i\}$, does not converge inside the Teichmüller space and converges to a projective lamination $[\lambda]$ in the Thurston compactification of the Teichmüller space, and that the other, say $\{n_i\}$, converges inside the Teichmüller space. Then G is a b–group. By the same argument as the proof of Lemma 5.1, the measured lamination λ cannot be realized by a pleated surface homotopic to ψ. If λ is not maximal and connected, as is shown in Thurston [26] or Lemma 4.4 in [17], G has an accidental parabolic element, which contradicts our assumption. Hence λ is maximal and connected, $\psi(\lambda)$ represents an ending lamination of the geometrically infinite tame end of $(\mathbf{H}^3/G)_0$, and G is a totally degenerate b–group. Let Σ_i be a boundary component of the convex core of $\mathbf{H}^3/\Gamma_i$ corresponding to the ideal boundary component with the structure n_i. Then as is shown in [19], the pleated surface Σ_i converges geometrically to a boundary component Σ_∞ of the convex core of $\mathbf{H}^3/\Gamma_\infty$ which can be lifted to a boundary component Σ of the convex core of $\mathbf{H}^3/G$, which must be the whole boundary of the convex core as G is a totally degenerate b–group. Hence a neighbourhood of the geometrically finite end of $(\mathbf{H}^3/G)_0$ is mapped homeomorphically to that of a geometrically finite end of $(\mathbf{H}^3/\Gamma_\infty)_0$ by the covering projection $q\colon \mathbf{H}^3/G \to \mathbf{H}^3/\Gamma_\infty$. On the other hand, by Proposition 2.6, there is also a neighbourhood of the geometrically tame end of $(\mathbf{H}^3/G)_0$ which is mapped homeomorphically to a neighbourhood of a geometrically infinite tame end of $(\mathbf{H}^3/\Gamma_\infty)_0$ by q. This implies that $G = \Gamma_\infty$.

Next assume that neither $\{m_i\}$ nor $\{n_i\}$ converges inside the Teichmüller space. After taking a subsequence, we can assume that $\{m_i\}$ converges to a projective lamination $[\lambda] \in \mathcal{PL}(S)$ and $\{n_i\}$ converges to a projective lamination $[\mu] \in \mathcal{PL}(S)$ in the Thurston compactification of the Teichmüller space. Since neither λ nor μ can be realized by a pleated surface homotopic to ψ by Lemma 5.1, they must be maximal and connected again by Thurston [26] or Lemma 4.4 in [17]. Then we can apply Theorem 3.1 to our situation and see that the support of λ is different from that of μ. This implies that the end of $(\mathbf{H}^3/G)_0$ with ending lamination represented by $\psi(\mu)$ is different from one with ending lamination represented by $\psi(\lambda)$ by Lemma 4.3, hence G is totally doubly degenerate. Let e_λ and e_μ denote the two distinct ends of $(\mathbf{H}^3/G)_0$ with ending laminations represented by $\psi(\lambda)$ and $\psi(\mu)$ respectively. By Proposition 2.6, there are neighbourhoods E_λ of e_λ and E_μ of e_μ such that $q|E_\lambda$ and $q|E_\mu$ are homeomorphisms to neighbourhoods of ends of $(\mathbf{H}^3/\Gamma_\infty)_0$. As $(\mathbf{H}^3/G)_0$ has only two ends, this can happen only when $\Gamma_\infty = G$ or G is a subgroup of Γ_∞ of index 2. We can see that the latter cannot happen by Lemma 2.3 in [19] (this fact is originally due to Thurston [26]). Thus we have proved our corollary when all of $\{\Gamma_i\}$ are quasi-Fuchsian.

Next assume that all the Γ_i are totally degenerate b–groups. Let m_i be the marked hyperbolic structure on S determined by the conformal structure of $\Omega_{\Gamma_i}/\Gamma_i$, and let λ_i be an ending lamination of unit length of the geometrically infinite tame end of $(\mathbf{H}^3/\Gamma_i)_0$. We can assume that $\{\lambda_i\}$ converges to a measured lamination λ after taking a subsequence. By the same argument as before, λ is maximal and connected, and $\psi(\lambda)$ represents an ending lamination of $(\mathbf{H}^3/G)_0$ by Lemma 5.1. First assume that $\{m_i\}$ converges inside the Teichmüller space. Then as before, the boundary Σ_i of the convex core of $\mathbf{H}^3/\Gamma_i$ converges geometrically to a boundary component Σ_∞ of the convex core of $\mathbf{H}^3/\Gamma_\infty$ which can be lifted to a boundary component Σ of the convex core of $\mathbf{H}^3/G$. Hence G is a totally degenerate b–group, and a neighbourhood of the geometrically finite end of $(\mathbf{H}^3/G)_0$ is mapped homeomorphically by q to a neighbourhood of an end of $\mathbf{H}^3/\Gamma_\infty$. Then as before, using Proposition 2.6, we can conclude that $G = \Gamma_\infty$.

Next assume that $\{m_i\}$ does not converge inside the Teichmüller space. Then after taking a subsequence, we can assume that $\{m_i\}$ converges to a projective lamination $[\mu]$. By the same argument as before, we can see that μ is maximal and connected, and $\psi(\mu)$ represents an ending lamination. Then by Theorem 3.1, we can see that the support of λ is different from that of μ. Hence G is totally doubly degenerate, and by Proposition 2.6, we conclude that $G = \Gamma_\infty$. Finally suppose that all the Γ_i are totally doubly degenerate. Let λ_i and μ_i be measured laminations of the unit length such that $\phi_i(\lambda_i)$ and $\phi_i(\mu_i)$ represent ending laminations of the two geometrically infinite tame ends of $(\mathbf{H}^3/\Gamma_i)_0$. By taking a subsequence, we can assume that $\{\lambda_i\}$ converges to a measured lamination λ and $\{\mu_i\}$ converges to a measured lamination μ in $\mathcal{ML}(S)$. Then as before, both λ and μ are maximal and connected, and $\psi(\lambda)$ and $\psi(\mu)$ represent ending laminations of $(\mathbf{H}^3/G)_0$. By Theorem 3.1, we can see that the support of λ is different from that of μ. Hence the end of $(\mathbf{H}^3/G)_0$ with ending lamination represented by $\psi(\lambda)$ is different from that with ending lamination represented by $\psi(\mu)$ by Lemma 4.3, which implies that G is totally doubly degenerate. Then by Proposition 2.6 again, we conclude that $G = \Gamma_\infty$, and the proof is completed. □

References

[1] **L Ahlfors**, *Finitely generated Kleinian groups*, Amer. J. Math. 86 (1964) 413–423

[2] **L Ahlfors**, *Lectures on quasiconfomal mappings*, Van Nostrand (1966)

[3] **J Anderson**, **R Canary**, *Algebraic limits of Kleinian groups which rearrange the pages of a book*, Inventiones Math. 126 (1996) 205–214

[4] **J Anderson**, **R Canary**, **D McCullough**, *On the topology of deformation space of Kleinian groups*, preprint

[5] **L Bers**, *Spaces of Kleinian groups*, from: “Several Complex Variables, I (Proc. Conf. Univ. of Maryland, College Park, Md. 1970)”, Springer Lecture Notes on Mathematics 155 (1970) 9–34

[6] **L Bers**, *On boundaries of Teichmüller spaces and on kleinian groups: I*, Annals of Math. 91 (1970) 570–600

[7] **F Bonahon**, *Bouts des variétés hyperboliques de dimension 3*, Annals of Math. 124 (1986) 71–158

[8] **R D Canary**, *A covering theorem for hyperbolic 3–manifolds and its applications*, Topology 35 (1996) 751–778

[9] **R D Canary**, **D B A Epstein**, **P Green**, *Notes on notes of Thurston*, from: “Analytical and geometric aspects of hyperbolic spaces”, LMS Lecture Notes Ser. 111, Cambridge Univ. Press (1987) 3–92

[10] **D B A Epstein**, **A Marden**, *Convex hulls in hyperbolic space, a theorem of Sullivan, and measured pleated surfaces*, from: “Analytical and geometric aspects of hyperbolic spaces”, LMS Lecture Notes Ser. 111, Cambridge Univ. Press (1987)113–253

[11] **A Fathi**, **F Laudenbach**, **V Poénaru**, *Travaux de Thurston sur les surfaces*, Astérisque 66–67 (1979)

[12] **M Freedman**, **J Hass**, **P Scott**, *Least area incompressible surfaces in 3–manifolds*, Inventiones Math. 71 (1983) 609–642

[13] **T Jørgensen**, **A Marden**, *Algebraic and geometric convergence of Kleinian groups*, Math. Scand. 66 (1990) 47–72

[14] **A Marden**, *The geometry of finitely generated Kleinian groups*, Annals of Math. 99 (1974) 465–496

[15] **D. McCullough**, *Compact submanifolds of 3–manifolds with boundary*, Quart. J. Math. Oxford 37 (1986) 299–307

[16] **Y Minsky**, *The classification of punctured-torus groups*, to appear in Annals of Math.

[17] **K Ohshika**, *On limits of quasi-conformal deformations of Kleinian groups*, Math. Z. 201 (1989) 167–176

[18] **K Ohshika**, *Limits of geometrically tame Kleinian groups*, Inventiones Math. 99 (1990) 185–203

[19] **K Ohshika**, *Geometric behaviour of Kleinian groups on boundaries for deformation spaces*, Quart. J. Math. Oxford (2) 43 (1992) 97–111

[20] **K Ohshika**, *Kleinian groups which are limits of geometrically finite groups*, preprint, Tokyo Institute of Technology

[21] **K Ohshika**, *Divergent sequences of Kleinian groups II*, in preparation

[22] **G P Scott**, *Compact submanifolds of 3–manifolds*, J. London Math. Soc. 7 (1973) 246–250

[23] **R. Penner, J. Harer**, *Combinatorics of train tracks*, Annals of Math. Studies 125 Princeton Univ. Press (1992)

[24] **D Sullivan**, *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, from: "Riemann surfaces and related topics", Annals of Math. Studies 97 (1981) 465–496

[25] **D. Sullivan**, *Quasiconformal homeomorphisms and dynamics II: Structural stability implies hyperbolicity for Kleinian groups*, Acta Math. 150 (1985) 243–260

[26] **W Thurston**, *The geometry and topology of 3–manifolds*, lecture notes, Princeton Univ. (1980)

[27] **W Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. 19 (1988) 417–431

[28] **W. Thurston**, *Hyperbolic structures on 3–manifolds II: Surface groups and 3–manifolds which fiber over the circle*, preprint, Princeton Univ. (1986)

Graduate School of Mathematical Sciences
University of Tokyo
Komaba, Meguro-ku, Tokyo 153, Japan

Email: `ohshika@ms.u-tokyo.ac.jp`

Received: 13 November 1997 Revised: 15 October 1998

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 451–478

Coordinates for Quasi-Fuchsian Punctured Torus Space

John R Parker
Jouni Parkkonen

Abstract We consider complex Fenchel–Nielsen coordinates on the quasi-Fuchsian space of punctured tori. These coordinates arise from a generalisation of Kra's plumbing construction and are related to earthquakes on Teichmüller space. They also allow us to interpolate between two coordinate systems on Teichmüller space, namely the classical Fuchsian space with Fenchel–Nielsen coordinates and the Maskit embedding. We also show how they relate to the pleating coordinates of Keen and Series.

AMS Classification 20H10; 32G15

Keywords Quasi-Fuchsian space, complex Fenchel-Nielsen coordinates, pleating coordinates

0 Introduction

In this note we study the holomorphic extension of the classical Fenchel–Nielsen coordinates of the Teichmüller space of once-punctured tori to the quasi-conformal deformation space of a Fuchsian group representing two punctured tori, quasi-Fuchsian punctured torus space. A *punctured torus group* $G = \langle S, T \rangle$ is a discrete, marked, free subgroup of $\mathrm{PSL}(2,\mathbb{C})$ with two generators whose commutator $K = T^{-1}S^{-1}TS$ is parabolic. This group acts naturally on the Riemann sphere by conformal transformations. The limit set $\Lambda(G)$ consists of all accumulation points of this action and is the smallest nonempty closed $G-$invariant subset of the Riemann sphere. Its complement is called the ordinary set $\Omega(G)$. The group G is called *quasi-Fuchsian* if its ordinary set $\Omega(G)$ consists of two simply connected components or equivalently if its limit set $\Lambda(G)$ is a topological circle. The space of all quasi-Fuchsian punctured torus groups up to conjugation within $\mathrm{PSL}(2,\mathbb{C})$ is called *quasi-Fuchsian punctured torus space* and will be denoted by $\mathcal{Q}$. The subset of $\mathcal{Q}$ consisting of groups whose limit set is a round circle is the space of all Fuchsian punctured torus groups. We call this *Fuchsian punctured torus space* and we will denoted it by $\mathcal{F}$. It is a copy of the Teichmüller space of the punctured torus.

Our approach to quasi-Fuchsian punctured torus groups is a combination of the classical Fenchel–Nielsen construction of Fuchsian groups and the gluing construction used by Kra in [12] for terminal b-groups. This is rather natural as Fuchsian groups form a real subspace inside the space of quasi-Fuchsian groups, and terminal b-groups form part of the boundary of the same space. We start with a Fuchsian group F of the second kind such that X_0, the quotient of the hyperbolic plane by F, is a sphere with a puncture and two infinite area ends with boundary geodesics of equal lengths. We then extend the group by adding a Möbius transformation that glues together the infinite area ends of the quotient to make a punctured torus. If the resulting group G is Fuchsian, this is the Fenchel–Nielsen construction. The construction is carried out in Section 1 and the Fenchel–Nielsen parameter is connected with the gluing parameter in Proposition 3.2. We can also regard F and G as acting on the Riemann sphere and we allow the Fenchel–Nielsen parameters to be complex. For other allowed values of the gluing parameter the resulting group G is a quasi-Fuchsian group bent along the geodesic in $\mathbb{H}^3$ corresponding to the boundary geodesics of X_0. The analysis of this bending, the associated shear, and their use for parametrising the deformation space of quasi-Fuchsian groups from different points of view is the main goal of the second half of the paper. We show that the resulting complexified Fenchel–Nielsen twist parameter can be interpreted as a complex shear as introduced by Parker and Series in [18] and that it has another natural interpretation as a $zw = t$ plumbing parameter as in Kra [12] The relationship between the various points of view is often easy at a conceptual level but can be hard to make explicit. In this paper we aim to make these connections as explicit as possible. Part of this involves writing down generators for punctured torus groups as matrices depending on parameters. This is useful for making explicit computations which we illustrate by drawing pictures of various slices through $\mathcal{Q}$.

One of the main themes of this paper will be a partial description of Keen–Series pleating invariants in terms of complex Fenchel–Nielsen parameters. For completeness we now give a brief account of pleating invariants [5, 8]. Unlike complex Fenchel–Nielsen coordinates these are not holomorphic coordinates but they do reflect the geometrical structure of the associated 3–manifold as well as the limit set of G. In particular, they may be used to determine the shape of the embedding of $\mathcal{Q}$ into $\mathbb{C}^2$ given by complex Fenchel–Nielsen coordinates. We will illustrate this with pictures of various slices through this embedding. Let G be a punctured torus group that is quasi-Fuchsian but not Fuchsian. We call such a group *strictly quasi-Fuchsian*. Consider $C(G)$, the the hyperbolic convex hull in $\mathbb{H}^3$ of the limit set of G (sometimes called the Nielsen region for G). This is a G–invariant, simply connected, convex subset of $\mathbb{H}^3$. Thus, its quotient $C(G)/G$ is a convex 3–manifold with boundary, whose fundamental group is G. In other words $C(G)/G$ is topologically, the product of a closed interval with a punctured torus. Each boundary component is topologically a punctured torus and naturally inherits a hyperbolic structure from the three manifold (this structure is different from the obvious hyperbolic structure on the corresponding component of $\Omega(G)/G$). This hyperbolic structure makes the boundary component into a pleated surface in the sense of Thurston. That is, it consists of totally geodesic flat pieces joined along a geodesic lamination, called the *pleating locus*, and which carries a natural transverse measure, the *bending measure*. The length l_μ of a measured lamination μ on a surface with a given hyperbolic structure, is the total mass on this surface of the measure given by the product of hyperbolic length along the leaves of μ with the transverse measure μ. For the punctured torus it is well known that measured geodesic laminations are projectively parametrised by the extended real line. If the support of the lamination is drawn on the square flat torus then this parameter is just the gradient. From this we see that the possible types of support that this lamination that can have fall into two categories. First, simple closed curves, sometimes called rational laminations because of their parametrisation by rational slopes on a square torus. The transverse measure is just the δ–measure on these curves. Secondly, laminations whose leaves are unbounded geodesic arcs and which correspond to "infinite words" in G. We refer to these as *infinite laminations*. They correspond to curves of irrational slope on a square torus and so are sometimes referred to as irrational laminations. The measure they carry is called bending measure. We remark that the pleating locus cannot be the same on both components of the convex hull boundary. This is an important observation. Most of the time in this paper, we will be concerned with the case where the pleating locus on one component of $\partial C(G)/G$ is a simple closed geodesic. In this case, there will be a constant angle across this geodesic between the two adjacent flat pieces. In this case, the lamination length is just the length of the geodesic in the hyperbolic structure on the convex hull boundary. Keen and Series show in [8] that a marked punctured torus group is determined by its pleating invariants, namely the projective classes (μ, l_μ), (ν, l_ν) where the supports of μ and ν are the pleating loci on the two components of $\partial C(G)$ and l_μ, l_ν are their lamination lengths.

Suppose that the pleating loci on both components of $\partial C(G)$ are simple closed curves γ, δ. The corresponding group elements necessarily have real trace (though this is not a sufficient condition). The collection of all groups in $\mathcal{Q}$ for which γ, δ are the pleating loci is called the *(rational) pleating plane* $\mathcal{P}_{\gamma,\delta}$. This is a two dimensional non-singular subset of $\mathcal{Q}$ and is parametrised by the lengths of the geodesics γ and δ (which in this case are the lamination lengths), see Theorem 2 of [8]. Keen and Series also define pleating planes for the cases where one or both of the pleating loci are infinite laminations. We will only make passing reference to such pleating planes.

We have been greatly helped by conversations with Linda Keen and Caroline Series. We would like to thank them for their help. We would also like to thank the referee for her/his comments which have improved the paper. The second author was supported by the Academy of Finland and by the foundation Magnus Ernroothin Säätiö of the Finnish Society of Sciences and Letters. Figures 4.1, 5.1 and 6.1 were drawn using a computer program developed by David Wright. The second author would like to thank him for his help in installing and using the program. Both authors would like to thank the Centre Emile Borel at the Institut Henri Poincaré for their hospitality.

1 Real Fenchel–Nielsen coordinates

In this section we show how to write down generators for Fuchsian punctured torus groups in terms of Fenchel–Nielsen coordinates. This section gives a foundation for the subsequent sections: In order to obtain complex Fenchel–Nielsen coordinates we simply keep the same normal form for the generators but make the parameters complex. The material in this section is quite standard, for a more complete discussion of Fenchel–Nielsen coordinates see Buser [2].

Let X be a punctured torus and $\gamma \subset X$ a simple closed geodesic. Then $X_0 = X \setminus \gamma$ is a hyperbolic surface of genus 0 with one puncture and two geodesic boundary components of equal length, say l. X_0 can be realised as a quotient $X_0 = N(G_0)/G_0$, where G_0 is a Fuchsian group of the second kind generated by two hyperbolic transformations with multiplier $\lambda = l/2 \in \mathbb{R}_+$:

$$S = \begin{pmatrix} \cosh(\lambda) & \cosh(\lambda)+1 \\ \cosh(\lambda)-1 & \cosh(\lambda) \end{pmatrix} \quad \text{and} \quad S' = \begin{pmatrix} \cosh(\lambda) & \cosh(\lambda)-1 \\ \cosh(\lambda)+1 & \cosh(\lambda) \end{pmatrix}, \tag{1.1}$$

and $N(G_0)$ is the *Nielsen region* of G_0, that is, the hyperbolic convex hull in $\mathbb{H}$ of the limit set of G_0. For later reference we record that the fixed points of these transformations are $\operatorname{fix} S = \pm\coth(\lambda/2)$ and $\operatorname{fix} S' = \pm\tanh(\lambda/2)$. The transformations S and S' correspond to the boundary geodesics of X_0 and their product $K = S'^{-1}S$ corresponds to the puncture. In other words

$$K = S'^{-1}S = \begin{pmatrix} -1+2\cosh(\lambda) & 2\cosh(\lambda) \\ -2\cosh(\lambda) & -1-2\cosh(\lambda) \end{pmatrix} \tag{1.2}$$

is a parabolic transformation fixing -1.

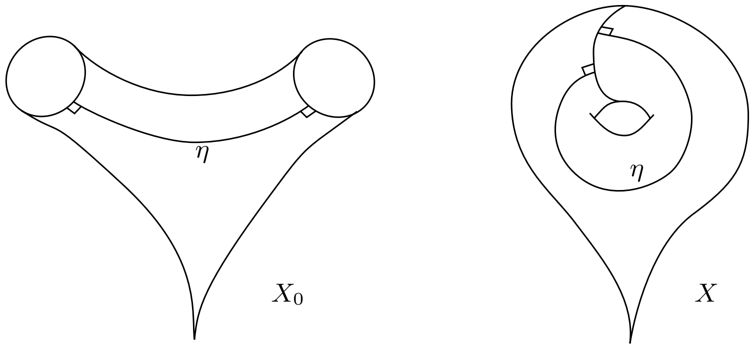


Figure 1.1 The Fenchel–Nielsen construction

The original surface X can be reconstructed by gluing together the geodesic boundary components of X_0. The gluing can be realised by adding to the group a hyperbolic Möbius transformation T that preserves $\mathbb{H}^2$. We form a new Fuchsian group, an HNN extension of G_0:

$$G = \langle G_0, T\rangle = (G_0) *_{\langle T\rangle} .$$

The transformation T is required to conjugate the cyclic subgroups $\langle S\rangle$ and $\langle S'\rangle$ in a manner compatible with the gluing operation:

$$T^{-1}ST = S'.$$

This condition fixes T up to one free parameter $\tau \in \mathbb{R}$, and T can be written in the form

$$T = \begin{pmatrix} \cosh(\tau/2)\coth(\lambda/2) & -\sinh(\tau/2) \\ -\sinh(\tau/2) & \cosh(\tau/2)\tanh(\lambda/2) \end{pmatrix}. \tag{1.3}$$

We recover the original (marked) surface with the correct geometry for exactly one parameter $\tau_0 \in \mathbb{R}$. However, the group G is a Fuchsian group for any real τ, and the parameter has a geometric interpretation: There is a unique simple geodesic arc η on X_0 perpendicular to both geodesic boundary curves. A distinguished lift of this arc to the universal covering $\mathbb{H}^2$ is the segment of the positive imaginary axis connecting $i\tanh(\lambda/2) \in \operatorname{axis}(S')$ and $i\coth(\lambda/2) \in \operatorname{axis}(S)$. Now T maps $i\tanh(\lambda/2)$ to a point on the axis of S, namely

$$T(i\tanh(\lambda/2)) = i\coth(\lambda/2)\big(\operatorname{sech}(\tau) + i\tanh(\tau)\big).$$

The (signed) hyperbolic distance of this point from $i\coth(\lambda/2)$ is exactly τ, the sign of τ is chosen to be positive if moving from $i\coth(\lambda/2)$ to $T\big(i\tanh(\lambda/2)\big)$ takes one in a positive (anti-clockwise) direction around the circle of radius $\coth(\lambda/2)$. The map $G \longmapsto (\lambda, \tau)$ is the *Fenchel–Nielsen coordinate* of the

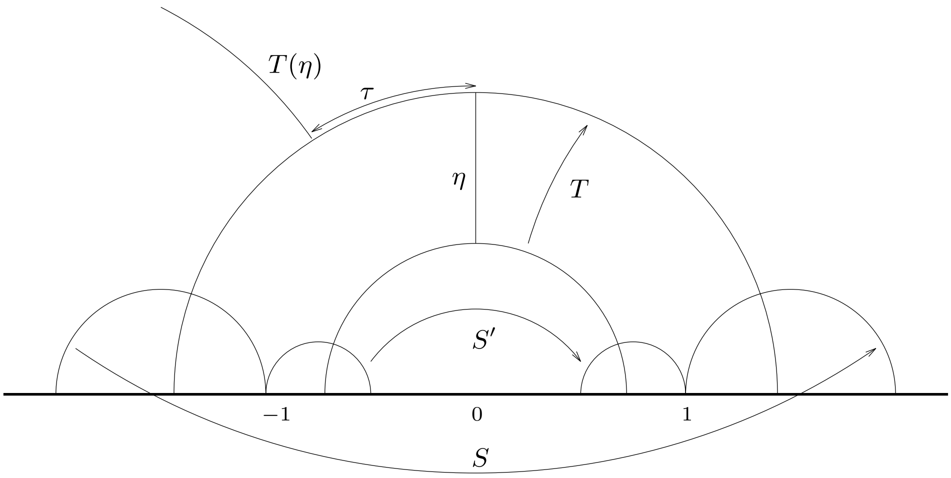


Figure 1.2 The fundamental domain

Teichmüller space of punctured tori. It defines a global real analytic parametrisation and identifies $\mathcal{F}$ with $\mathbb{R}_+ \times \mathbb{R}$ (see Buser [2]). Fenchel–Nielsen coordinates depend on the choice of an ordered pair of (homotopy classes of) simple closed curves on the punctured torus intersecting exactly once, that is a marking. We obtain different coordinates for different choices of marking. These choices are related by elements of the modular group. We investigate this in more detail in the next section. In [21] Waterman and Wolpert give computer pictures for the action of the modular group on Fenchel–Nielsen coordinates. They also give pictures of this action in another set of coordinates which can be easily derived from traces of generating triples.

Varying τ and keeping λ fixed is the Fenchel–Nielsen deformation considered by Wolpert in [22] and [23].

2 Complex Fenchel–Nielsen coordinates

The Teichmüller space of punctured tori seen as the space of Fuchsian groups representing a punctured torus, $\mathcal{F}$, is a natural subspace of the corresponding quasi-Fuchsian space, $\mathcal{Q}$. Kourouniotis [11] and Tan [20] showed that, for compact surfaces, the Fenchel–Nielsen coordinates can be complexified to give a global parametrisation of quasi-Fuchsian space. With this in mind we now suppose that λ and τ are complex. That is $(\lambda, \tau) \in \mathbb{C}_+ \times \mathbb{C}$ where $\mathbb{C}_+$ denotes those complex numbers with positive real part. With such λ and τ we consider groups generated by S and T with the normal forms (1.1) and (1.3). This means that S and T are now in $\mathrm{PSL}(2,\mathbb{C})$ rather than in $\mathrm{PSL}(2,\mathbb{R})$. The group $\langle S,T\rangle$ is not quasi-Fuchsian for all $(\lambda,\tau)\in\mathbb{C}_+\times\mathbb{C}$ but the *complex Fenchel–Nielsen coordinates* (λ,τ) do give global coordinates on $\mathcal{Q}$. We present a short proof of this fact using the stratification method developed by Kra and Maskit in [13].

Proposition 2.1 *The map* $h\colon \mathcal{Q}\to\mathbb{C}^2$ *given by* $h(G)=(\cosh^2(\lambda),e^\tau)$ *is a global complex analytic coordinate map on* $\mathcal{Q}$.

Proof Let $G=\langle A,B\rangle$ be a quasi-Fuchsian group of type $(1,1)$ generated by two loxodromic transformations A and B. Assume that the group is normalised so that 0 is the repelling fixed point, and ∞ is the attracting fixed point of A, and that $B(0)=1$. Let $x_1=B(\infty)$, and $x_2=B(1)$. Note that $x_1,x_2\in\Lambda(G)$.

We claim that G is determined by giving x_1 and x_2: Clearly B is determined, as we know how it maps three points. Also, from the normalisation we know that

$$A=\begin{pmatrix} a & 0\\ 0 & 1/a\end{pmatrix},\qquad B=\begin{pmatrix} x_1(x_2-1) & x_1-x_2\\ x_2-1 & x_1-x_2\end{pmatrix},$$

where $a\in\mathbb{C},\ |a|>1$. Now

$$\mathrm{tr}[A,B]=\frac{2a^2x_1-1-a^4}{a^2(x_1-1)}.$$

As $[A,B]$ is assumed to be a parabolic, solving for a^2 in the equation $\mathrm{tr}[A,B]=-2$ gives $a^2=2x_1-1\pm2\sqrt{x_1(x_1-1)}$. Only one of these solutions satisfies $|a|>1$. This fixes A. (The choice of the branch of the square root $a=\sqrt{a^2}$ does not affect A.)

Let us normalise the group $G=\langle S',T\rangle$ of Section 1 as above: We conjugate G with a transformation (here written as an element of $\mathrm{PGL}(2,\mathbb{C})$)

$$R=\begin{pmatrix} \cosh(\lambda)/(1-\cosh(\lambda)) & -\coth(\lambda)\\ 1/(1-\cosh(\lambda)) & \mathrm{csch}(\lambda)\end{pmatrix}.$$

This gives

$$S_0=RS'R^{-1}=\begin{pmatrix} e^\lambda & 0\\ 0 & e^{-\lambda}\end{pmatrix},$$

where we can assume $|e^\lambda|>1$, and

$$T_0=RTR^{-1}=\begin{pmatrix} \coth(\lambda)e^{-\tau/2} & \coth(\lambda)e^{\tau/2}\\ \mathrm{csch}(\lambda)\,\mathrm{sech}(\lambda)e^{-\tau/2} & \coth(\lambda)e^{\tau/2}\end{pmatrix}.$$

Now

$$x_1=\cosh^2(\lambda),\qquad x_2=\frac{1+e^\tau}{\mathrm{sech}^2(\lambda)+e^\tau}.$$

□

Remark 2.2 The choice $|e^\lambda| > 1$ implies $\lambda \in \mathbb{C}_+$. Unlike real Fenchel–Nielsen coordinates, there is no simple description of which pairs $(\lambda, \tau) \in \mathbb{C}_+ \times \mathbb{C}$ are in $h(\mathcal{Q})$, the image of quasi-Fuchsian space under the coordinate map. Using the pleating invariants of Keen and Series [8] one can determine how $h(\mathcal{Q})$ lies inside $\mathbb{C}^2$. In this paper we carry out part of this construction and illustrate our results by drawing slices through $\mathcal{Q}$ in Figure 5.1.

We now use the fact that $(\cosh^2(\lambda), e^\tau)$ give global coordinates to show that (λ, τ) give global coordinates on quasi-Fuchsian space. Let

$$\widetilde{\mathcal{FN}} = \left\{(\lambda, \tau) \in \mathbb{C}^2 \colon (\cosh^2(\lambda), e^\tau) \in h(\mathcal{Q})\right\},$$

where h is the map of Proposition 2.1. We denote by $\mathcal{FN}$ the component of $\widetilde{\mathcal{FN}}$ containing $\mathbb{R}_+ \times \mathbb{R}$. Our proof that (λ, τ) give global coordinates involves showing that there are no paths in $\widetilde{\mathcal{FN}}$ between two places where the parameters are different but the groups are the same.

Proposition 2.3 *Let $\gamma\colon [0,1] \longrightarrow \mathbb{C}_+ \times \mathbb{C}$ denote any path from $\gamma(0) = (\lambda_0, \tau_0)$ to $\gamma(1) = (\lambda_0 + m\pi i, \tau_0 + 2n\pi i)$ for any $(\lambda_0, \tau_0) \in \widetilde{\mathcal{FN}}$ and integers m and n not both zero. Then $\gamma([0,1])$ is not contained in $\widetilde{\mathcal{FN}}$.*

Proof We begin with the case $m = 1$ and $n = 0$.

Using the normalisation of Proposition 2.1 we have $T_0(\lambda_0, \tau_0) = T_0(\lambda_0 + \pi i, \tau_0)$. Also notice that $S_0(\lambda_0, \tau_0)$ and $S_0(\lambda_0 + \pi i, \tau_0)$ are the same in $\mathrm{PSL}(2,\mathbb{C})$ but differ by $-I$ in $\mathrm{SL}(2,\mathbb{C})$. They correspond to the two choices of square root for a^2 in Proposition 2.1. Thus moving along γ from (λ_0, τ_0) to $(\lambda_0 + \pi i, \tau_0)$ adds $i\pi$ to the multiplier of S_0. For more details of the relationship between multipliers and the different lifts of Möbius transformations in $\mathrm{PSL}(2,\mathbb{C})$ to matrices in $\mathrm{SL}(2,\mathbb{C})$ see the discussion in Section 1 of [18]. Let Π_1 be any hyperplane in $\mathbb{H}^3$ orthogonal to the axis of S_0 and let $\Pi_2 = S_0(\Pi_1)$ be its image under S_0. Because going along γ from (λ_0, τ_0) to $(\lambda_0 + \pi i, \tau_0)$ changes the multiplier of S_0 by πi then also Π_2 is rotated by 2π with respect to Π_1. We can think of going along γ as being the same as doing a Dehn twist of the annulus between $\partial\Pi_1$ and $\partial\Pi_2$ in $\widehat{\mathbb{C}}$.

Specifically we may decompose S_0 into a product of half turns (that is elliptic involutions in $\mathrm{PSL}(2,\mathbb{C})$ of order 2) as follows:

$$S_0 = \iota_1 \iota_2 = \begin{pmatrix} 0 & e^\lambda \\ -e^{-\lambda} & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The geodesic fixed by $\iota_1(\lambda, \tau)$ has end points $\pm i e^\lambda$. Replacing (λ_0, τ_0) by $(\lambda_0 + \pi i, \tau_0)$ interchanges these end points. Equivalently this reverses the orientation of the geodesic. Therefore if Π_1 is hyperplane orthogonal to the axis of S_0 and containing the geodesic with end points $\pm i$ (that is the axis of ι_2) it is clear that its image under the ι_1 is rotated by 2π when we replace λ_0 by $\lambda_0 + \pi i$.

Let ξ_1 be any point of $\partial\Pi_1 \cap \Omega$ and $\xi_2 = S_0(\xi_1)$ be its image under S_0. Let α be any path in Ω joining ξ_1 and ξ_2. Now consider the homotopy H given by following α while (λ,τ) varies along γ. Denote the image of α at time t by α_t.

If the whole of γ were in $\mathcal{Q}$ then the homotopy H would induce an isotopy from $\Omega\big(G(\lambda_0,\tau_0)\big)$ to $\Omega\big(G(\lambda_0,\tau_0+2\pi i)\big)$. At each stage S_0 is loxodromic so Π_1 and Π_2 are disjoint and α_t consists of more than one point. Now α_0 and α_1 are both paths in $\Omega\big(G(\lambda_0,\tau_0)\big) = \Omega\big(G(\lambda_0+\pi i,\tau_0)\big)$ joining ξ_1 and ξ_2. It is clear from the earlier discussion that the path $\alpha_1{\alpha_0}^{-1}$ formed by going along α_1 and then backwards along α_0 winds once around the (closed) annulus between $\partial\Pi_1$ and $\partial\Pi_2$. This it separates the fixed points of S_0. This contradicts the fact that the limit set Λ is connected.

We can adapt this proof to cover the case where λ_0 is sent to $\lambda_0+m\pi i$ for some non-zero integer m. This is done by observing that the path $\alpha_1{\alpha_0}^{-1}$ now winds m times around the annulus between $\partial\Pi_1$ and $\partial\Pi_2$. Moreover this argument does not use the value of τ at each end of the path. It merely uses the fact that $T_0(\lambda_0,\tau_0) = T_0(\lambda_1,\tau_1)$ and so we may take $\tau_1 = \tau_0 + 2n\pi i$ without changing anything.

Thus we have proved the result when m and n are any integers with m not zero. It remains to prove the result when $m = 0$ and n is an integer other than zero. We do this as follows. Observe that, with the normalisation of (1.1) and (1.3), $S(\lambda_0,\tau_0) = S(\lambda_0,\tau_0+2\pi i)$ but $T(\lambda_0,\tau_0+2\pi i)$ and $T(\lambda_0,\tau_0)$ give distinct lifts in $\mathrm{SL}(2,\mathbb{C})$. As before we decompose T into a product of half turns as follows:

$$T = \iota_1\iota_2 = \begin{pmatrix} \sinh(\tau/2) & \cosh(\tau/2)\coth(\lambda/2) \\ -\cosh(\tau/2)\tanh(\lambda/2) & -\sinh(\tau/2) \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The geodesic fixed by $\iota_1(\lambda,\tau)$ has end points

$$\frac{-\sinh(\tau/2) \pm i}{\cosh(\tau/2)\tanh(\lambda/2)}.$$

Replacing (λ_0,τ_0) by $(\lambda_0,\tau_0+2\pi i)$ interchanges these end points. The rest of the argument follows as before. □

The next two results are direct consequences of Propositions 2.1 and 2.3.

Corollary *The functions* $\cosh^2(\lambda)$ *and* e^τ *have well defined inverses in* $h(\mathcal{Q})$ *and so we can regard* (λ,τ) *is a global coordinate system for quasi-Fuchsian space.*

Corollary *The pair* $\big(\cosh(\lambda),\sinh(\tau/2)\big)$ *give global coordinates for quasi-Fuchsian space. In particular, the points where* $\sinh(\lambda)=0$ *or* $\cosh(\tau/2)=0$ *are not in* $\mathcal{FN}$.

Proof The first part follows from the previous corollary. We give a simple justification for the last statement. If $\sinh(\lambda)=0$ then $\cosh(\lambda)=\pm 1$ and S is parabolic. Similarly if $\cosh(\tau/2)=0$ then T is elliptic or else $\coth(\lambda)$ is infinite and S is parabolic as before. □

Complex Fenchel–Nielsen coordinates depend on the choice of a marking for the punctured torus, that is an ordered pair of generators for S. It is intuitively clear that changing this marking gives a biholomorphic change of the coordinates $\big(\cosh(\lambda),\sinh(\tau/2)\big)$. We now make this explicit.

Proposition 2.4 *Let* (S_0,T_0) *and* (S_1,T_1) *be any two generating pairs for a punctured torus group* G. *Let* (λ_0,τ_0) *and* (λ_1,τ_1) *be the corresponding complex Fenchel–Nielsen coordinates on* $\mathcal{Q}$. *Then the map*

$$\big(\cosh(\lambda_0),\sinh(\tau_0/2)\big)\longmapsto\big(\cosh(\lambda_1),\sinh(\tau_1/2)\big)$$

is a biholomorphic homeomorphism of $\mathcal{Q}$ *to itself.*

Proof A classical result of Nielsen [17] states that we can obtain the pair (S_1,T_1) from (S_0,T_0) by a sequence of elementary Nielsen-moves on the generators. As one of our aims is to make things explicit, we list these Nielsen moves and write down the effect that they have on the coordinates $\big(\cosh(\lambda),\sinh(\tau/2)\big)$. From this, it is clear that these changes of coordinate are holomorphic.

First, suppose that $(S',T')=(S,S^{\pm1}T)$. Then

$$\cosh(\lambda')=\cosh(\lambda),\quad \sinh(\tau'/2)=\sinh(\tau/2)\cosh(\lambda)\mp\cosh(\tau/2)\sinh(\lambda).$$

Secondly, suppose that $(S',T')=(S,T^{-1})$. Then

$$\cosh(\lambda')=\cosh(\lambda),\quad \sinh(\tau'/2)=-\sinh(\tau).$$

Finally, suppose that $(S',T')=(T,S)$

$$\cosh(\lambda')=\frac{\cosh(\lambda)\cosh(\tau/2)}{\sinh(\lambda)},\quad \sinh(\tau'/2)=\frac{-\sinh(\tau/2)\sinh(\lambda)}{\cosh(\tau/2)}.$$

□

3 Plumbing and earthquakes

In this section we show how the Fenchel–Nielsen construction is related to two standard constructions in Teichmüller theory, namely the $zw = t$ plumbing construction and to quake-bends. In particular, the Fenchel–Nielsen twist parameter is a special case of the quake-bend parameter and we show how to express the plumbing parameter in terms of Fenchel–Nielsen parameters.

Consider Teichmüller space of the punctured torus $\mathcal{F}$ with Fenchel–Nielsen coordinates as in Section 1. The motion through Teichmüller space obtained by fixing the length parameter λ but varying the shear τ is the Fenchel–Nielsen deformation (see [22]) which is the simplest example of an earthquake (see Waterman and Wolpert [21] and McMullen [16] for some other earthquakes). One may think of this as cutting along $\mathrm{Ax}(S)$ twisting and then regluing.

If we reglue so that along $\mathrm{Ax}(S)$ the two sides make a constant angle then we have an example of a *quake-bend* (see Epstein and Marden [4]). We can say that the group $G(\lambda, \tau)$ is obtained from $G(\lambda, 0)$ by doing a quake-bend along S with parameter τ. That is, for $\lambda \in \mathbb{R}_+$, we take the Fuchsian group $G(\lambda, 0)$ with generators

$$S = \begin{pmatrix} \cosh(\lambda) & \cosh(\lambda)+1 \\ \cosh(\lambda)-1 & \cosh(\lambda) \end{pmatrix}, \quad T = \begin{pmatrix} \coth(\lambda/2) & 0 \\ 0 & \tanh(\lambda/2) \end{pmatrix}.$$

This group has a fundamental domain rather like the one shown in Figure 1.2 except with $\tau = 0$ (the copy of the hyperbolic plane in question is the hyperplane in $\mathbb{H}^3$ whose boundary is the extended real axis). Let $Q(\tau)$ be a loxodromic map with the same fixed points as S and trace $2\cosh(\tau/2)$. Apply $Q(\tau)$ to that part of $\mathbb{H}^2$ lying above $\mathrm{Ax}(S)$, ie those points with $|z| > \coth(\lambda/2)$. What we have done is essentially cut along $\mathrm{Ax}(S)$ and reglued after performing a shear and a bend. Now repeat this construction along the axis of every conjugate of S. This is a quake-bend. For more details and a precise definition of what is involved, see [4]. A discussion of quake-bends and complex Fenchel–Nielsen coordinates in given in Section 5.3 of [7].

One can perform this construction for irrational measured laminations. In this case the new measure is obtained by multiplying the initial bending measure by the quake-bend parameter. This gives a way of generalising the Fenchel–Nielsen twist parameter τ analogous to the way lamination length generalises the hyperbolic length of a simple closed curve.

We now relate these ideas by extending the $zw = t$*–plumbing construction* to this situation. Essentially the same construction was used by Earle and Marden [3] and Kra [12] in the case of punctured surfaces and it was extended by Arés [1] and Parkkonen [19] for surfaces with elliptic cone points.

Let X_0 be a punctured cylinder (as in Section 1). Assume that the boundary geodesics γ_1 and γ_2 corresponding to boundary components b_1 and b_2 have equal length $l = 2\lambda > 0$. Let U_1 and U_2 be neighbourhoods of, respectively, the ends of X_0 corresponding to γ_1 and γ_2. Let γ_{12} be the shortest geodesic arc connecting the two boundary components, and let

$$\mathcal{A}_\lambda = \{\zeta \in \mathbb{C} \mid e^{-\pi^2/\lambda} < |\zeta| < 1\}$$

with its hyperbolic metric of constant curvature -1. The curve $\{|z| = e^{-\pi^2/2\lambda}\}$ is the unique geodesic in $\mathcal{A}_\lambda$ with this metric.

We define local coordinates at the ends of X_0 by

$$z\colon U_1 \to \mathcal{A}_\lambda \quad \text{and} \quad w\colon U_2 \to \mathcal{A}_\lambda$$

by requiring that the maps are isometries and that the segments $\gamma_{12} \cap U_1$ and $\gamma_{12} \cap U_2$ are mapped into $\mathcal{A}_\lambda \cap \mathbb{R}_+$. These conditions define the maps z and w uniquely.

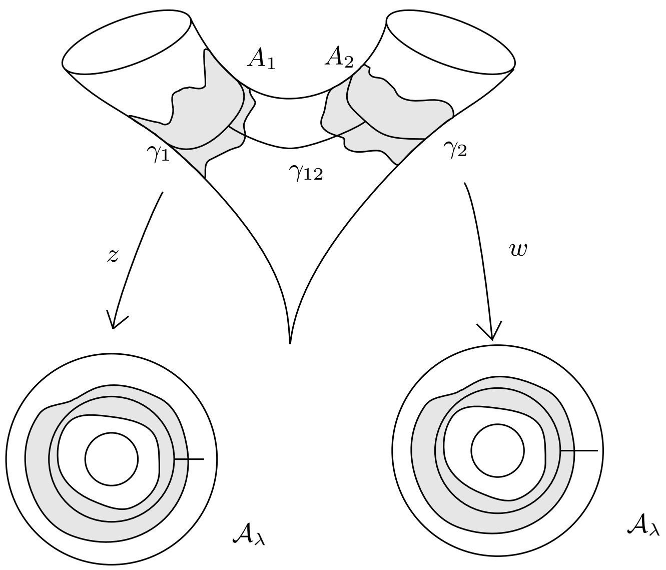


Figure 3.1 The $zw = t$ plumbing construction

If $A \subset X_0$ is an annulus homotopic to a boundary component b of X_0, we call the component of ∂A separating the other component of ∂A from b, the outer boundary of A. The remaining component of ∂A is the inner boundary of A. Assume there are annuli $A_i \subset U_i$ and a holomorphic homeomorphism $f\colon A_1 \to A_2$ so that

$$z(x)w(f(x)) = t$$

for some constant $t \in \mathbb{C}$ and f maps the outer boundary of A_1 to the inner boundary of A_2. The outer boundaries bound annuli on X_0. Remove these annuli to form a new Riemann surface X_{trunc}. Define

$$X_t := X_{\text{trunc}}/\sim,$$

where the equivalence is defined by setting

$$x \sim y \Longleftrightarrow z(x)w(y) = t.$$

We say that X_t was obtained from X_0 by the *$zw = t$ plumbing construction with plumbing or gluing parameter t*. If the annuli A_i can be chosen to be collar neighbourhoods of the boundary geodesics γ_i, we say that the plumbing is *tame*.

Next we show that the Fenchel–Nielsen twist parameter is naturally associated with a plumbing parameter:

Lemma 3.1 *If G is in $\mathcal{Q}$ with $\lambda \in \mathbb{R}_+$ then $t = e^{-\pi^2/\lambda}\, e^{-\pi i\tau/\lambda} = e^{i\pi\mu}$ where $\mu = (i\pi - \tau)/\lambda$.*

Proof Let $\Pi\colon \mathbb{H}^2 \to \mathbb{H}^2/G_0$ be the canonical projection. Let $\widetilde{\gamma_1}$ be the geodesic in $\mathbb{H}^2$ connecting the fixed points of S and $\widetilde{\gamma_2}$ the geodesic connecting the fixed points of S' . Now the boundary geodesics for which the gluing will be done are $\gamma_i = \Pi(\widetilde{\gamma_i})$. The local coordinates are given by

$$z(P) = \exp\left(\frac{\pi i}{\lambda}\log\left(\frac{\Pi^{-1}(P)\sinh(\lambda/2) + \cosh(\lambda/2)}{-\Pi^{-1}(P)\sinh(\lambda/2) + \cosh(\lambda/2)}\right)\right),$$

and

$$w(Q) = \exp\left(\frac{\pi i}{\lambda}\log\left(\frac{\Pi^{-1}(Q)\cosh(\lambda/2) - \sinh(\lambda/2)}{\Pi^{-1}(Q)\cosh(\lambda/2) + \sinh(\lambda/2)}\right)\right).$$

Substituting for T we see, after simplifying, that

$$z\big(T(Q)\big) = \exp\left(\frac{\pi i}{\lambda}\log\left(\frac{e^{-\tau}\,\Pi^{-1}(Q)\cosh(\lambda/2) + \sinh(\lambda/2)}{-\Pi^{-1}(Q)\cosh(\lambda/2) + \sinh(\lambda/2)}\right)\right).$$

Thus $z(T(Q))\; w(Q) = \exp(-\pi^2/\lambda - \pi i\tau/\lambda)$ as claimed. □

The same proof also yields the following:

Proposition 3.2 *The classical Fenchel–Nielsen construction is a $zw = t$ plumbing construction for a parameter t of modulus $e^{-\pi^2/\lambda}$.*

4 λ–slices

In this section we keep λ real but allow τ to be complex. When $\theta = \mathrm{Im}(\tau)$ is in the interval $(0, \pi]$ we will show that the axis of S is the pleating locus on one component of the convex hull boundary and when $\theta \in [-\pi, 0)$ then it is the pleating locus on the other component. We will show that τ has an interpretation as a *complex shear* along the pleating locus, $\mathrm{Ax}(S)$, see Parker and Series [18]. The complex shear σ is defined as follows. The imaginary part of σ, which we require to be in the interval $(-\pi, \pi)$, is the bending angle on the convex hull boundary across $\mathrm{Ax}(S)$. The real part of σ defined as follows. Let η be the unique simple geodesic arc in the convex hull boundary from $\mathrm{Ax}(S)$ to itself and orthogonal to $\mathrm{Ax}(S)$ at both ends. Then we form a curve in the convex hull boundary in the homotopy class specified by T by going along η and then along $\mathrm{Ax}(S)$. The real part of the complex shear is the signed distance we go along $\mathrm{Ax}(S)$. This definition is made precise on page 172 of [18]. The theorems of this section should be compared with the constructions found in [12] and section 2.2 of [5]. We also note that one may use the formulae of [18] to show that, when λ is real, the imaginary part of σ cannot be $\pm\pi$, Proposition 7.1 of [8].

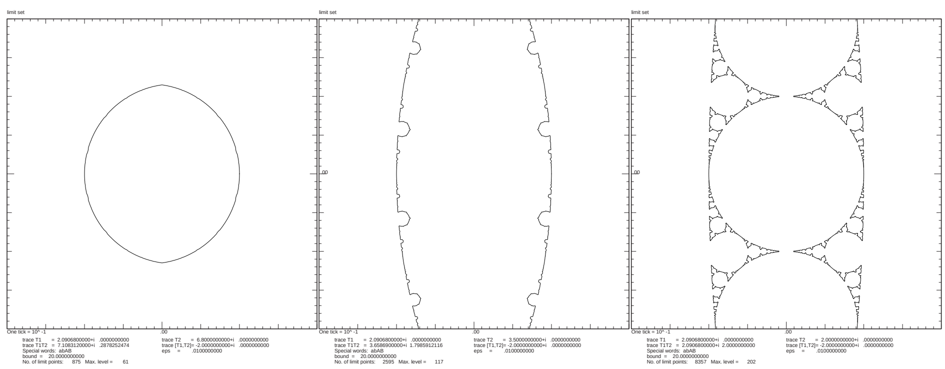


Figure 4.1 Limit sets of groups in a λ–slice

Let us fix $\lambda > 0$. Consider the set

$$\{\tau \in \mathbb{C} \mid (\lambda, \tau) \in \mathcal{FN}\}$$

The λ*–slice* $\mathcal{Q}_\lambda$ is defined to be the component of this set containing the points where $\tau \in \mathbb{R}$ (compare with the quake-bend planes of [8]). We wish to obtain an estimate for the allowed values of τ for each λ. In order to do this we will construct pleating coordinates on each λ–slice. A first approximation can be achieved by estimating the values of $\theta = \mathrm{Im}(\tau)$ that correspond to tame plumbing constructions. The following theorem is an explicit version of Theorem 6.1 of [7]. Specifically, we show that the constant ϵ of that theorem can be taken as $\theta_0 = 2\arccos\big(\tanh(\lambda)\big)$ (compare Section 6 of [9]). Because the point $(\lambda, i\theta_0)$ is on the boundary of quasi-Fuchsian space, there can be no larger uniform bound on $\mathrm{Im}(\tau)$ that ensures discreteness. The fact that θ is the imaginary part of the quake-bend will follow from Theorem 4.2.

Theorem 4.1 *Let* $\theta_0 \in (0, \pi)$ *be defined by the equation* $\cos(\theta_0/2) = \tanh(\lambda)$. *Then for* $\mathrm{Im}(\tau) = \theta \in (-\theta_0, \theta_0)$ *the group* G *is a quasi-Fuchsian punctured torus group.*

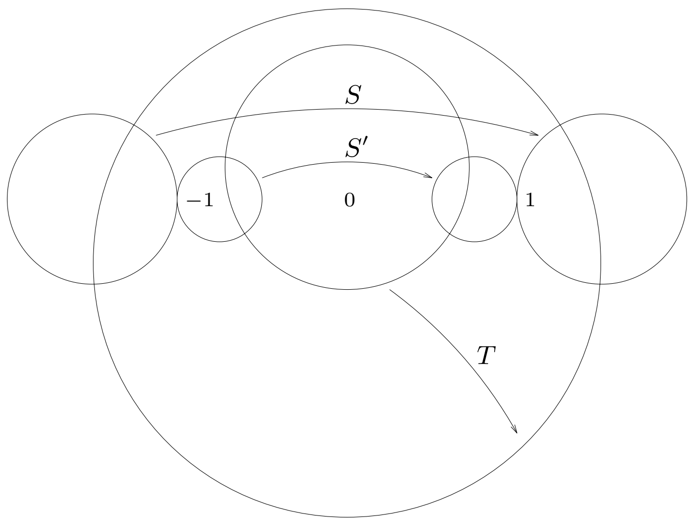


Figure 4.2 The construction for the combination theorem

Proof It is easy to check that the circle with centre at $i\tanh(\lambda/2)\tan(\theta/2)$ and radius $\tanh(\lambda/2)\sec(\theta/2)$ is mapped by T to the circle with centre at $-i\coth(\lambda/2)\tan(\theta/2)$ and radius $\coth(\lambda/2)\sec(\theta/2)$. Moreover these circles are mapped to themselves under $\langle S'\rangle$ and $\langle S\rangle$ respectively (the circles pass through the fixed points of S' and S). Providing the two circles are disjoint then the annulus between them is a fundamental domain for $\langle T\rangle$. It is easy to check that the circles are disjoint if and only if $\cos(\theta/2) > \tanh(\lambda)$, that is $\theta \in (-\theta_0, \theta_0)$. When this happens we can use Maskit's second combination theorem [14, 15] to show that G is discrete, has a fundamental domain with two components each of which glues up to give a punctured torus and G is quasi-Fuchsian. □

For a positive real number λ, suppose that G is a quasi-Fuchsian punctured torus group. The ordinary set of G has two components. There is an obvious way to label these as the "top" and "bottom" components so that, for the case when G is Fuchsian, the upper half plane is the "top"component. In what follows, we give a result that enables us to make this definition precise. Namely in Lemmas 4.3 and 4.4, we show that either the "top" component contains the upper half plane or the "bottom" component contains the lower half plane (or both, in which case the group would be Fuchsian). When G is strictly quasi-Fuchsian there are two components to the convex hull boundary facing these two components of the ordinary set. We label them "top" and "bottom" as well (this notation is also used by Keen and Series on page 370 of [7]). Both of these components is a pleated surface and so we may speak of the pleating locus on the "top" and "bottom". The following theorem may be thought of as a generalisation of Proposition 6.2 of [18].

Theorem 4.2 *For any parameter in a λ slice ($\lambda \in \mathbb{R}$) with $\theta \in (0,\pi)$ (respectively $\theta \in (-\pi,0)$) the pleating locus on the "bottom" (respectively "top") surface is S and τ (respectively $-\tau$) is the complex shear along S with respect to the curve T as defined in [18].*

Intuitively this should be clear as we are keeping λ real and bending away from $\mathrm{Ax}(S)$. As we are only bending along one curve the result is convex. In the general case we could not expect a Fenchel–Nielsen complex twist to always be the complex shear on the convex hull boundary as we may bend along different curves in different directions. In what follows we only consider the case $\theta > 0$. By symmetry this is sufficient. The proof will be by way of several lemmas.

Lemma 4.3 *If $\theta \in (0,\theta_0)$ then the lower half plane $\mathbb{L}$ is contained in $\Omega(G)$.*

Proof We will consider the lower half plane $\mathbb{L}$ with its Poincaré metric. We then use plane hyperbolic geometry to prove the result.

Let D^* be the fundamental region for the action of $F = \langle S, S' \rangle$ on $\mathbb{L}$ formed by the intersection of $\mathbb{L}$ with the exterior of the isometric circles for S and S'. That is

$$D^* = \{z \in \mathbb{L} \colon |(\cosh(\lambda)+\varepsilon_1)z + \varepsilon_2\cosh(\lambda)| \geq 1 \text{ for all choices of } \varepsilon_1, \varepsilon_2 = \pm 1\}.$$

We are now going to consider various hypercycles (that is arcs of circles) with endpoints at the fixed points of S and S'. To begin with, let c_0 and c_0' be the semicircles centred at 0 of radius $\coth(\lambda/2)$ and $\tanh(\lambda/2)$. Clearly these are the Poincaré geodesics joining the fixed points of S and S' respectively. Let D_0 be the subset of $\mathbb{L}$ between these two semi-circles:

$$D_0 = \{z \in \mathbb{L} \colon \tanh(\lambda/2) \leq |z| \leq \coth(\lambda/2)\}\,.$$

The Nielsen region $N(F)$ of $F = \langle S, S' \rangle$, that is the hyperbolic convex hull of $\Lambda(F)$ in $\mathbb{L}$, is

$$N(F) = \bigcup_{g \in F} g(D^* \cap D_0).$$

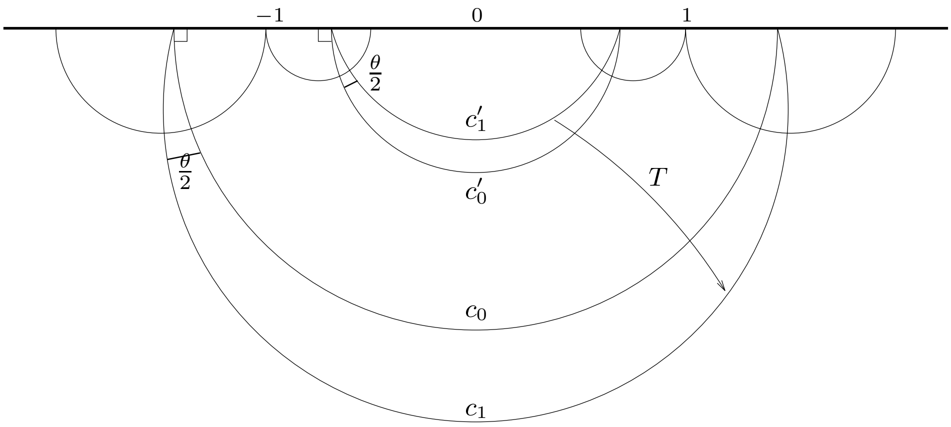


Figure 4.3 The construction in the lower half plane

Now consider the circular arcs c_1 and c'_1 in $\mathbb{L} - D_0$ with endpoints at $\pm\coth(\lambda/2)$ and $\pm\tanh(\lambda/2)$ which make an angle $\theta/2$ with c_0 and c'_0 respectively. In other words c_1 is the arc of the circle centred at $-i\coth(\lambda/2)\tan(\theta/2)$ with radius $\coth(\lambda/2)\sec(\theta/2)$ lying in the lower half plane. Similarly c'_1 is the intersection of $\mathbb{L}$ with the circle centred at $i\tanh(\lambda/2)\tan(\theta/2)$ with radius $\tanh(\lambda/2)\sec(\theta/2)$. Figure 4.3 shows c_1 and c'_1. Observe that c_1 and c'_1 are a constant distance $d(\theta)$ from c_0 and c'_0 where

$$d(\theta) = \log\bigl(\sec(\theta/2) + \tan(\theta/2)\bigr).$$

Denote the lune between c_0 and c_1 by $B(\theta)$ and the lune between c'_0 and c'_1 by $B'(\theta)$. Let D_1 be the subset of the lower half plane lying between c_1 and c'_1. Now D_1 is just the intersection of $\mathbb{L}$ with the fundamental region for T considered in Theorem 4.1. One of the consequences of Maskit's combination theorem is that $D^* \cap D_1$ is contained in $\Omega(G)$. (It is at this point that we have used $\theta < \theta_0$.) Let $N(\theta)$ be the union of all F translates of $D^* \cap D_1$:

$$N(\theta) = \bigcup_{g \in F} g(D^* \cap D_1).$$

It is clear that $N(\theta)$ is just the $d(\theta)$ neighbourhood of $N(F)$. Since $D^* \cap D_1$ is contained in $\Omega(G)$ then so is $N(\theta)$.

We are going to mimic this construction with more arcs. For each n with $n\theta < \pi$, let c_n and c'_n be the circular arcs in $\mathbb{L} - D_0$ with endpoints at $\pm\coth(\lambda/2)$ and $\pm\tanh(\lambda/2)$ making an angle of $n\theta/2$ with c_0 and c'_0 respectively. That is c_n is the arc of a circle with centre at $-i\coth(\lambda/2)\tan(n\theta/2)$ and radius $\coth(\lambda/2)\sec(n\theta/2)$ and c'_n is the arc of a circle with centre at $i\tanh(\lambda/2)\tan(n\theta/2)$ and radius $\tanh(\lambda/2)\sec(n\theta/2)$. As before, c_n is a constant distance $d(n\theta)$ from c_0 and c'_n is the same distance from c'_0. We define D_n, the subset of $\mathbb{L}$ between c_n and c'_n, and the lunes $B(n\theta)$ and $B'(n\theta)$ as before. Let

$$N(n\theta) = \bigcup_{g\in F} g(D^* \cap D_n).$$

Again $N(n\theta)$ is the $d(n\theta)$ neighbourhood of $N(F)$.

Furthermore, let n_0 be the integer with $(n_0 - 1)\theta < \pi \leq n_0\theta$. We define arcs c_{n_0} and c'_{n_0} which are now in the closed upper half plane. We also define $B(n_0\theta)$, $B'(n_0\theta)$ and $N(n_0\theta)$ geometrically but remark that these no longer have any metrical properties. An important observation is that $\mathbb{L}$ is contained in $N(n_0\theta)$.

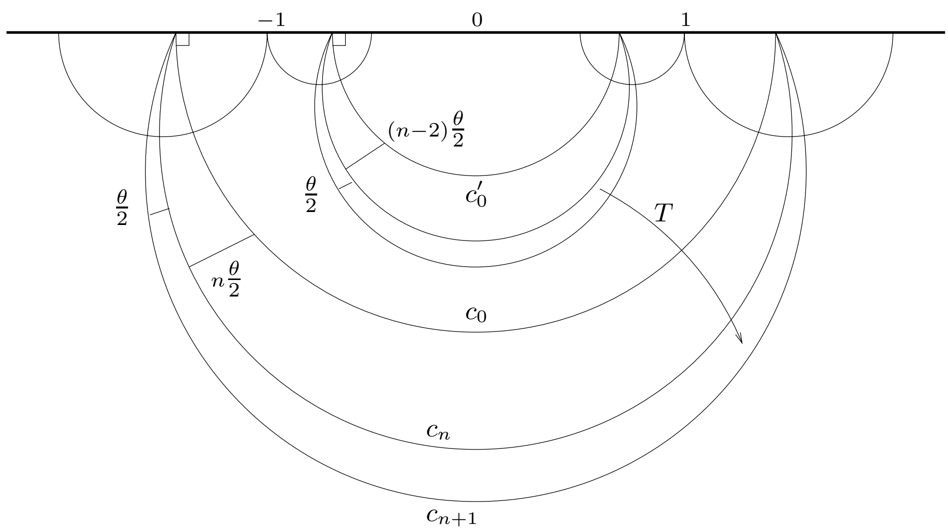


Figure 4.4 The inductive step

The rest of the proof follows by an induction from $n = 1$ up to $n = n_0$. We claim that, for $1 \leq n < n_0$ that if $B(n\theta)$ and $B'(n\theta)$ are in $\Omega(G)$ then so are

$B\big((n+1)\theta\big)$ and $B'\big((n+1)\theta\big)$. This in turn means that $N\big((n+1)\theta\big)$ is in $\Omega(G)$. In particular $N(n_0\theta)$, which contains $\mathbb{L}$, is in $\Omega(G)$.

Thus all we have do is prove the claim, which we now do. Since $B(n\theta)$ and $B'(n\theta)$ are contained in $\Omega(G)$ then so is $N(n\theta)$. Consider $T^{-1}B\big((n+1)\theta\big)$. Since c_{n+1} makes an angle of $n\theta/2$ with c_1 and T acts conformally on $\widehat{\mathbb{C}}$ we see that $T^{-1}(c_{n+1})$ makes an angle of $n\theta/2$ with $T^{-1}(c_1) = c_1'$, see Figure 4.4. In other words $T^{-1}(c_{n+1})$ is a hypercycle a constant distance $d\big((n-1)\theta\big)$ from c_0' (also it is not c_{n-1}'). This means that $T^{-1}(c_{n+1})$, and hence also $T^{-1}B\big((n+1)\theta\big)$, is contained within the $d(n\theta)$ neighbourhood of $N(F)$, that is $N(n\theta)$. Since $N(n\theta)$ was assumed to be in $\Omega(G)$, we see that $T^{-1}B\big((n+1)\theta\big)$ and hence also $B\big((n+1)\theta\big)$ is contained in $\Omega(G)$, as claimed. We remark that if $n > n_0$ then $T^{-1}(c_{n+1})$ lies in the closed upper half plane and the argument breaks down. A similar argument shows that $B'\big((n+1)\theta\big)$ is also contained in $\Omega(G)$. This completes the proof. □

Lemma 4.4 *If $\tau \in \mathcal{Q}_\lambda$ and $\theta \in (0,\pi)$ then the pleating locus on the "bottom" surface is S.*

Proof Suppose first that $\theta \in (0,\theta_0)$. From Lemma 4.3 we see that $\mathbb{L}$ is contained in $\Omega(G)$. Thus the geodesic plane in $\mathbb{H}^3$ with boundary the real axis is a support plane for $\partial C(G)$. Moreover the image of this plane under T must also be a support plane for $\partial C(G)$. As the intersection of these two planes is the axis of S we have the result.

Now consider $\tau = t+i\theta \in \mathcal{Q}_\lambda$ and $\theta \in [\theta_0,\pi)$. We proceed as in Proposition 5.4 of [5]. Suppose that S is not the pleating locus for the bottom surface. Consider a path α in $\mathcal{Q}_\lambda$ joining τ with $\tau' = t' + i\theta'$ where $\theta' \in (0,\theta_0)$. Without loss of generality, suppose that if $\tau \in \alpha$ then $\mathrm{Im}(\tau) \geq \theta' > 0$. We know that at τ' the pleating locus on the bottom surface is S. Using the standard identification of projective measured laminations on the punctured torus with the extended real line (with the topology given by stereographic projection of the usual topology on the circle) then Keen and Series show that the pleating locus is continuous with respect to paths in $\mathcal{Q}$ [6]. Therefore there are points on the path α for which the pleating locus is a projective measured lamination arbitrarily close to γ_∞. In particular there are points where the pleating locus is γ_m for $m \in \mathbb{Z}$ which corresponds to $W_m = S^{-m}T \in G$ (in the next section we will give more details of how to associate words with simple closed curves). In particular, this group element must have real trace. In other words there is a point of α where $\mathrm{tr}(S^{-m}T) = 2\cosh(\tau/2 + m\lambda)\coth(\lambda)$ is real, and so

$$0 = \sinh(t/2 + m\lambda)\sin(\theta/2).$$

As $\theta \in [\theta', \pi)$ we see that $\sin(\theta/2) \neq 0$. Thus $t/2 + m\lambda = 0$ and $\mathrm{tr}(S^{-m}T) = 2\cos(\theta/2)$. This means $S^{-m}T$ is elliptic and so τ is not in $\mathcal{Q}_\lambda$ after all. □

Lemma 4.5 *With S and T as in the theorem and $\theta \in (0, \pi)$ (respectively $\theta \in (-\pi, 0)$) the complex shear σ along S with respect to T is $\sigma = \tau$ (respectively $\sigma = -\tau$).*

Proof The trace of T is

$$\cosh(\tau/2)\big(\coth(\lambda/2) + \tanh(\lambda/2)\big) = 2\cosh(\tau/2)\coth(\lambda).$$

Writing $\mathrm{tr}(T) = 2\cosh\big(\lambda(T)\big)$ and $\mathrm{tr}(S) = 2\cosh\big(\lambda(S)\big)$ the formula (I) of [18] gives the complex shear along S with respect to T as σ where

$$\begin{aligned}\cosh(\sigma/2) &= \cosh\big(\lambda(T)\big)\tanh\big(\lambda(S)\big)\\ &= \cosh(\tau/2)\coth(\lambda)\tanh(\lambda)\\ &= \cosh(\tau/2).\end{aligned}$$

Thus σ and τ agree up to sign and addition of multiples of $2\pi i$. Since $\mathrm{Im}(\sigma)$ is in $(0, \pi)$ we find that $\sigma = \tau$ when $\theta = \mathrm{Im}(\tau) > 0$ and $\sigma = -\tau$ when $\theta < 0$. □

5 Pleating rays on λ–slices

We have shown that on a λ–slice the pleating locus on one component of the convex hull boundary is γ_∞ which corresponds to S. We now investigate the intersection of each λ–slice with the rational pleating plane associated to the simple closed curves γ_∞ and $\gamma_{p/q}$. We call this intersection a *pleating ray*. Part of the this section will be a justification of this name.

In order to obtain pleating rays on each λ–slice, we follow the arguments in [5], many of which are inherently two-dimensional in nature. These arguments have been superseded by more general arguments in [8]. We give these arguments to help the reader interpret Figure 5.1 and Figure 6.1 without having to refer to [5] or [8]. But, since these arguments are not new, we shall not give all the details. Furthermore, we indicate how one may use pleating rays on λ–slices to obtain the rational pleating planes. This is the simplest part of the construction of pleating coordinates. The more complicated parts are treated at length in [8].

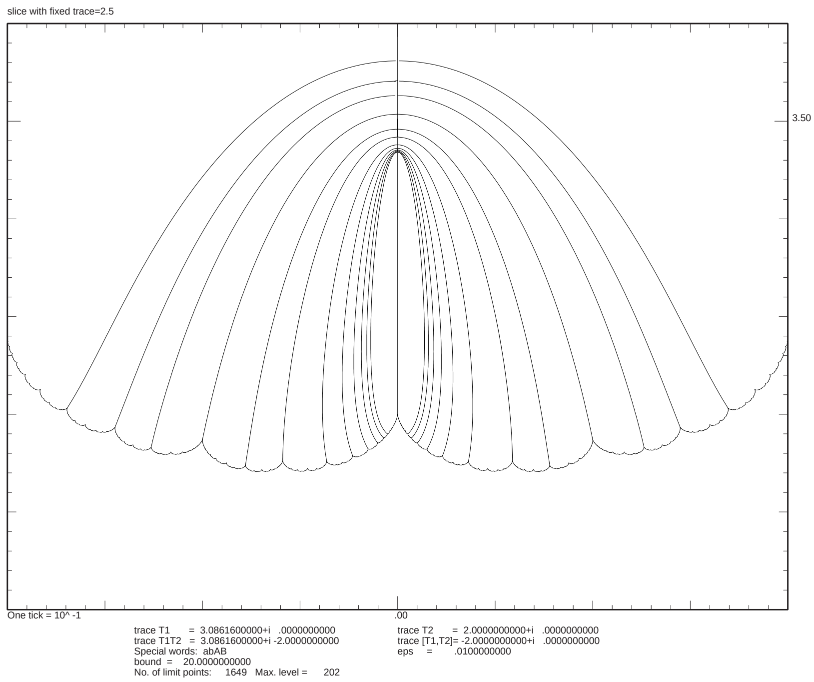


Figure 5.1 Part of a slice through $\mathcal{Q}$ with λ held to be real and fixed. In this case $\cosh(\lambda) = 5/4$. This figure shows the image of the slice under the 2 to 1 map $\tau \longmapsto i\,\mathrm{tr}\,T = 2i\cosh(\tau/2)\coth(\lambda) = \frac{10}{3}i\cosh(\tau/2)$. The figure shows pleating rays for this slice, see [8] or Section 6. The vertical line from $10i/3$ upwards represents Fuchsian space (which has been folded onto itself at the point corresponding to a rectangular torus). Observe that the pleating rays meet Fuchsian space orthogonally.

In what follows, we assume that the pleating locus on one component of the convex hull boundary is γ_∞, represented by S, and the pleating locus the other is also a simple closed curve, $\gamma_{p/q}$ for some $p/q \in \mathbb{Q}$. There is a special word $W_{p/q} \in G = \langle S, T\rangle$ corresponding to the homotopy class of simple closed curves $[\gamma_{p/q}]$. These words are defined recursively in [24] (see also Section 3.1 of [5]) but of course, we now need to use the generators S and T defined (1.1) and (1.3). First, $W_\infty = S^{-1}$, $W_m = S^{-m}T$ for $m \in \mathbb{Z}$. If $qr - ps = 1$ then we inductively define $W_{(p+r)/(q+s)} = W_{r/s}W_{p/q}$.

For each $\gamma_{p/q}$ the *p/q–pleating ray* $\mathcal{P}^\lambda_{p/q,\infty}$ on $\mathcal{Q}_\lambda$ is defined to be the those points of $\mathcal{Q}_\lambda$ for which the pleating locus is $\gamma_{p/q}$ on the "top" and γ_∞ on the "bottom". Thus these points have $\mathrm{Im}(\tau) \in (0,\pi)$, Theorem 4.2. Likewise $\mathcal{P}^\lambda_{\infty,p/q}$ consists of those points in $\mathcal{Q}_\lambda$ where the pleating locus on the "top" surface is γ_∞ and that on the "bottom" is $\gamma_{p/q}$. Such points have $\mathrm{Im}(\tau) \in$

$(-\pi, 0)$. This discussion may be summarised in the following result which should be compared to Theorem 5.1 of [5].

Proposition 5.1 *On each λ–slice $\mathcal{Q}_\lambda$ and for $p/q \in \mathbb{Q}$ the pleating rays $\mathcal{P}^\lambda_{p/q,\infty}$ and $\mathcal{P}^\lambda_{\infty,p/q}$ each consist of a non-empty, connected, non-singular arc on which $\mathrm{tr}(W_{p/q})$ is real and which meet $\mathcal{F}$ orthogonally at the same point from the opposite side. Their other end-points lie on the boundary of $\mathcal{Q}_\lambda$ and at these points $|\,\mathrm{tr}(W_{p/q})| = 2$.*

Some rational pleating rays are shown in the pictures Figures 5.1 and 6.1. It can be observed that the pleating rays are non-singular connected arcs that meet Fuchsian space orthogonally.

Sketch proof This is an adaptation of ideas in [5] and [8]. First we fix a particular λ–slice $\mathcal{Q}_\lambda$. In Theorem 4.2 we showed that γ_∞, represented by S, is the pleating locus on one component of the convex hull boundary. For definiteness we take this to be the "bottom" component. By symmetry all our arguments go through when the pleating loci are the other way round.

It was shown in Corollary 6.4 of [18] that, when the complex shear is purely imaginary, the pleating locus on the "top" component is T (that is γ_0). Using a change of generators (marking) as in Proposition 2.4, it follows that, when the real part of the complex shear is $-2m\lambda$, for an integer m, then the pleating locus on the "top" component is $S^{-m}T$ (that is γ_m). Consider the line where $\mathrm{Im}(\tau) = \theta_0/2$. Such groups are all quasi-Fuchsian (Theorem 4.1) and at $\tau = -2m\lambda + i\theta_0/2$ the pleating locus is γ_m for $m \in \mathbb{Z}$. Thus, by the continuity of the pleating locus, see [6], as we move along this line we find points whose pleating locus is given by any real parameter. This shows that any real pleating ray on $\mathcal{Q}_\lambda$ is non-empty.

It is clear that $\mathcal{P}^\lambda_{p/q,\infty}$ is contained in the real locus of $\mathrm{tr}(W_{p/q})$. We now investigate how this real locus meets Fuchsian space. Any brach of the real locus of $\mathrm{tr}(W_{p/q})$ contained in $\mathcal{Q}_\lambda - \mathcal{F}$ meets $\mathcal{F}$ in a singularity of $\mathrm{tr}(W_{p/q})$. A result of Wolpert, page 226 of [23], says that the second derivative of $\left|\mathrm{tr}(W_{p/q})\right|$ with respect to τ along Fuchsian space is strictly positive. (We have used here that $\gamma_{p/q}$ and γ_∞ are both simple and they intersect.) Thus $\mathrm{tr}(W_{p/q})$ has a unique singularity in $\mathcal{F}$ and this singularity is quadratic. Hence the branches of its real locus must meet orthogonally. In particular there is one branch meeting $\mathcal{F}$ at this point on which $\mathrm{Im}(\tau) > 0$ and one brach where $\mathrm{Im}(\tau) < 0$.

For $0 < p/q < 1$ the pleating ray $\mathcal{P}^\lambda_{p/q,\infty}$ (which is non-empty) must be contained in the open set bounded by $\mathcal{F}$, that is $\mathrm{Im}(\tau) = 0$; the pleating rays

$\mathcal{P}^{\lambda}_{0,\infty}$, that is $\mathrm{Re}(\tau) = 0$, and $\mathcal{P}^{\lambda}_{1,\infty}$, that is $\mathrm{Re}(\tau) = -2\lambda$; and the boundary of $\mathcal{Q}_\lambda$. The pleating ray must be a union of connected components of the intersection of this set with the real locus of $\mathrm{tr}(W_{p/q})$. The proof of this statement follows Proposition 5.4 of [5]. A similar argument has been used in Lemma 4.4 so we will not repeat it. It is clear that if the pleating locus on the "top" is $\gamma_{p/q}$ and if $|\,\mathrm{tr}(W_{p/q})| > 2$ then the group is in the interior of $\mathcal{Q}$. Thus, moving along $\mathcal{P}^{\lambda}_{p/q,\infty}$ in the direction of increasing $|\,\mathrm{tr}(W_{p/q})|$ we cannot reach the boundary of $\mathcal{Q}$ and so we must reach $\mathcal{F}$. It follows that $\mathcal{P}^{\lambda}_{p/q,\infty}$ is connected and non-singular. If not, there would be at least two branches of $\mathcal{P}^{\lambda}_{p/q,\infty}$ on which $|\,\mathrm{tr}(W_{p/q})|$ is increasing. But there is only one branch that meets $\mathcal{F}$, a contradiction. A similar analysis takes care of other p/q.

Finally, when $|\,\mathrm{tr}(W_{p/q})| = 2$ the pleating ray reaches the boundary of $\mathcal{Q}_\lambda$ and the curve $\gamma_{p/q}$ has become parabolic. This completes our sketch proof of Proposition 5.1. □

In order to obtain the pleating planes associated to the pairs γ_∞, $\gamma_{p/q}$ we must vary λ. As we do this, the pleating rays on each λ–slice now sweep out the whole pleating plane. Keen and Series prove that this gives a connected, non-singular two dimensional subset of $\mathcal{Q}$. In order to obtain pleating planes associated to other pairs of curves we use the change of coordinates given in Proposition 2.4. Specifically, if the pleating loci we are interested in are $\gamma_{a/b}$ and $\gamma_{c/d}$ which intersect $q = ad - bc \neq 0$ times then there is a sequence of Nielsen moves taking the pair $(\gamma_\infty, \gamma_{p/q})$ to the pair $(\gamma_{a/b}, \gamma_{c/d})$. Associated to these Nielsen moves is a biholomorphic change of coordinates on $\mathcal{Q}$ and the pleating plane associated to $\gamma_{a/b}$ and $\gamma_{c/d}$ is the image under this change of coordinates of the pleating plane associated to γ_∞ and $\gamma_{p/q}$.

We conclude this section with a discussion of how one may take data associated to one component of the convex hull boundary and find information about the other component. At first sight it does not seem clear how this could be done. But, at least when the pleating locus on one component is a simple closed curve, this follows from the relationship between complex Fenchel–Nielsen coordinates and Keen–Series pleating invariants. Let G be a strictly quasi-Fuchsian punctured torus group. Suppose that the pleating locus on one component of the convex hull boundary is a simple closed curve γ of length λ. Then we can construct Fenchel–Nielsen coordinates relative to a generating pair S, T where γ is represented by S. The complex Fenchel–Nielsen coordinates are given purely in terms of data associated to the component of the convex hull boundary on which γ is the pleating locus. By considering the associated λ–slice $\mathcal{Q}_\lambda$, we can find the Keen–Series pleating invariants for G in terms of the complex Fenchel–Nielsen coordinates. We have not mentioned lamination length on $\mathcal{Q}_\lambda$

in the above discussion. It suffices to remark that when the lamination on the other component of the convex hull boundary is also a simple closed curve given by $W \in G$, then the lamination length can be easily found from $\mathrm{tr}(W)$. For irrational pleating rays, we just use a continuity argument. In particular, we can determine information about the pleating on the other component of the convex hull boundary (this generalises Corollary 6.4 of [18], where it is shown that if the pleating locus on one component of $\partial C(G)/G$ is S and the complex shear is purely imaginary then the pleating locus on the other component is T). Moreover, if the pleating locus on the other component of the convex hull boundary is also a simple closed curve, we can use a sequence of Nielsen moves (see Proposition 2.4) to determine the Fenchel–Nielsen coordinates with respect to δ. In fact this is very straightforward.

On the other hand, suppose the pleating locus is an infinite measured lamination μ with lamination length l_μ. The projective class (μ, l_μ) (see [8]) generalises the choice of simple closed curve with δ–measure and the hyperbolic length of that curve. It follows from the work of Epstein–Marden, [4], that the group is completely determined by (μ, l_μ) and the quake-bend parameter τ (see [7, 8] for a discussion of the quake-bend parameter for quasi-Fuchsian punctured torus groups). These generalise the Fenchel–Nielsen coordinates for an infinite lamination. However, it does not seem that there is a straightforward way to go explicitly from these parameters to the pleating invariants or to the corresponding parameters on the other component of the convex hull boundary.

6 Degeneration to the Maskit embedding

In the previous sections we have considered what happens when λ is a fixed real positive number. In this section, we consider what happens when $\lambda = 0$. We should expect the complex shear to tend to $i\pi$ as λ tends to 0 (compare Theorem 4.1(i) of [18], see Proposition 6.1 below). This means that complex Fenchel–Nielsen coordinates degenerate. In this section we show that by using the plumbing parameter instead, we obtain the *Maskit embedding* of Teichmüller space, denoted $\mathcal{M}$ (see [24, 5]). This is defined to be the space of free Kleinian groups G on two generators S, T up to conjugation, such that each group has the following properties. First, the generator S and the commutator $K = T^{-1}S^{-1}TS$ are both parabolic. Secondly, the components of the ordinary set are of two kinds. Namely, a simply connected, G–invariant component whose quotient is a punctured torus; and also infinitely many round discs whose stabilisers are thrice punctured sphere groups, all conjugate within G. In other words these groups are terminal b-groups. This space is a holomorphically parametrised copy of the Teichmüller space of a punctured torus.

There is a standard normal form for the generators in terms of a parameter μ, see [24, 5], which is

$$S_0 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad T_0 = \begin{pmatrix} -i\mu & -i \\ -i & 0 \end{pmatrix}. \tag{6.1}$$

The goal of this section is to show that as we let λ tend to zero, the normal form for S and T given in (1.1), (1.3) degenerate to generators of groups in the Maskit embedding (6.1). Moreover, the λ slices $\mathcal{Q}_\lambda$ with their pleating rays tend to the Maskit embedding with its pleating rays. We illustrate this with a series of pictures which should be compared to Figure 1 of [5]. There is a discussion of how the Maskit embedding lies on the boundary of quasi-Fuchsian space on page 190 of [18].

Consider the limit of S as λ tends to zero:

$$S_0 = \lim_{\lambda\to 0} \begin{pmatrix} \cosh(\lambda) & \cosh(\lambda)+1 \\ \cosh(\lambda)-1 & \cosh(\lambda) \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

Similarly the limit of $S' = T^{-1}ST$ as λ tends to zero is:

$$S_0' = \lim_{\lambda\to 0} \begin{pmatrix} \cosh(\lambda) & \cosh(\lambda)-1 \\ \cosh(\lambda)+1 & \cosh(\lambda) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

The parabolic transformations S_0 and S_0' generate the level 2 principal congruence subgroup of $\mathrm{PSL}(2,\mathbb{Z})$, a torsion-free triangle group. A comparison of the plumbing parameter calculated in Lemma 3.1 with the corresponding result for terminal b-groups (see Kra [12; Section 6.4]) suggests that, in order to study the degeneration of quasi-Fuchsian groups in $\bigcup_{\lambda>0}\mathcal{Q}_\lambda$ as $\lambda \longrightarrow 0$, it is useful to make a change of parameters

$$\mu = \frac{i\pi - \tau}{\lambda}.$$

We refer to μ as the *plumbing parameter*. In terms of this parameter the matrix T can be written as

$$T = \begin{pmatrix} -i\sinh(\lambda\mu/2)\coth(\lambda/2) & -i\cosh(\lambda\mu/2) \\ -i\cosh(\lambda\mu/2) & -i\sinh(\lambda\mu/2)\tanh(\lambda/2) \end{pmatrix}. \tag{6.2}$$

Using Lemma 3.1, we see that (1.1) and (6.2) give a parametrisation of the generators of G in terms of a length parameter and a plumbing parameter. The following result on the limit groups, which should be compared to Theorem 4.1(i) of [18], now follows rather easily:

Proposition 6.1 *Consider a sequence of groups where λ tends to zero but μ remains fixed. Then the complex shear along S tends to $i\pi$.*

Proof The conclusion is immediate from the definition of μ: $\tau = i\pi - \mu\lambda \to i\pi$ as $\lambda \to 0$. □

We now show that when λ tends to zero with μ being kept fixed we obtain the standard form for group generators in the Maskit embedding.

Proposition 6.2 *Assume that $\mu \in \mathcal{Q}_\lambda$ for small λ. As λ tends to zero the group with parameter (λ, μ) tends to the terminal b-group representing punctured torus on its invariant component with parameter μ.*

Proof We have already seen that S_0 and S_0' have the correct form.

Let μ be fixed. For small λ we have

$$\sinh(\lambda\mu/2)\coth(\lambda/2) = \big(\lambda\mu/2 + O(\lambda^2)\big)\big(2/\lambda + O(1)\big) = \mu + O(\lambda).$$

Therefore we have

$$\lim_{\lambda\to 0}\big(\sinh(\lambda\mu/2)\coth(\lambda/2)\big) = \mu.$$

This means that the limit as λ tends to zero of T is

$$\begin{aligned} T_0 &= \lim_{\lambda\to 0}\begin{pmatrix} -i\sinh(\lambda\mu/2)\coth(\lambda/2) & -i\cosh(\lambda\mu/2) \\ -i\cosh(\lambda\mu/2) & -i\sinh(\lambda\mu/2)\tanh(\lambda/2) \end{pmatrix} \\ &= \begin{pmatrix} -i\mu & -i \\ -i & 0 \end{pmatrix}. \end{aligned}$$

The limiting matrices S_0 and T_0 are just the usual group generators of terminal b-groups in the Maskit embedding $\mathcal{M}$ of Teichmüller space of the punctured torus. □

The convergence of λ–slices to $\mathcal{M}$ is illustrated in Figure 6.1.

Remarks 6.3 (a) The plumbing construction is tame when $\mathrm{Im}(\tau) = \theta \in (0, \theta_0)$ or equivalently $\mathrm{Im}(\mu) \in \big((\pi - \theta_0)/\lambda, \pi/\lambda\big)$. For small λ we have $\theta_0 = \pi - 2\lambda + O(\lambda)^2$. As λ tends to zero this interval tends to $(2, \infty)$, which is the condition for tame plumbing in the Maskit slice, Section 6.2 of [12] or Proposition 2.3 of [24].

(b) In the (λ, μ) parameters, Fuchsian space corresponds to the union of the lines $\mathrm{Im}(\mu) = \pi/\lambda$. When $\lambda \to 0$, $\mathrm{Im}\,\mu \to \infty$, that is, the closure of Fuchsian space touches $\mathcal{M}$ at the boundary point corresponding to the parameter $\mu = \infty$ (see page 191 of [18]).

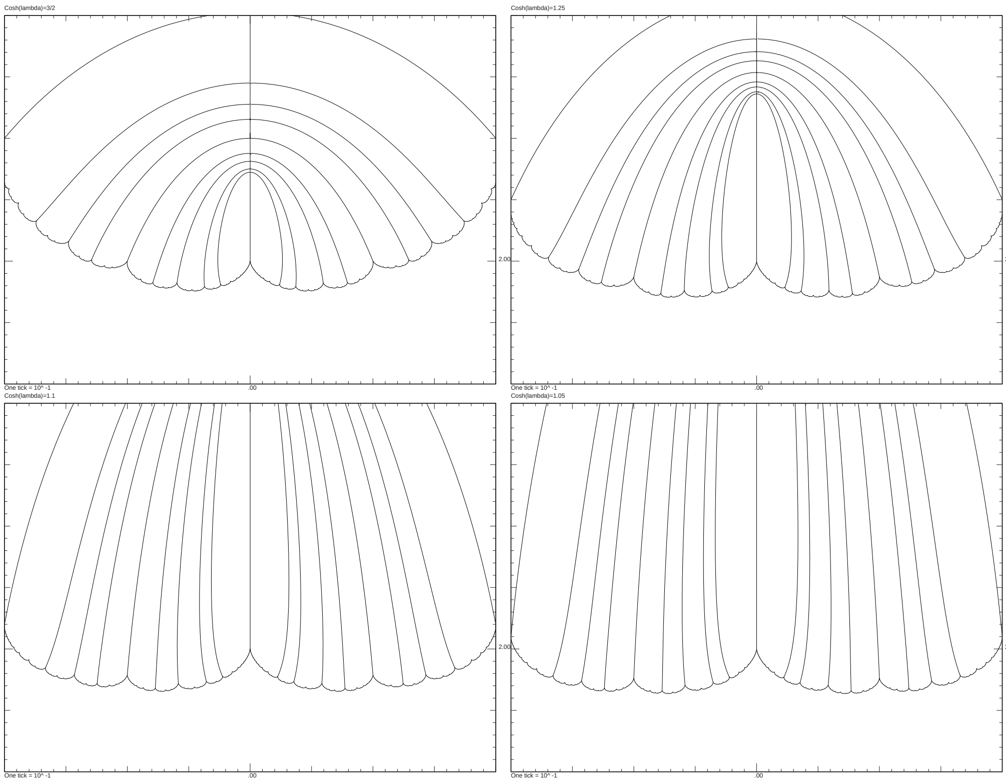


Figure 6.1 λ–slices for $\mathrm{tr}(T) = 3,\ 2.5,\ 2.2$ and 2.1 drawn with a collection of rational pleating rays

References

[1] **P Arés**, *Coordinates for Teichmüller spaces of b-groups with torsion*, Ann. Acad. Sci. Fenn. 20 (1995) 279–300

[2] **P Buser**, *Geometry and Spectra of Compact Riemann Surfaces*, Birkhauser (1992)

[3] **C J Earle**, **A Marden**, *Geometric complex coordinates for Teichmüller space*, in preparation

[4] **D B A Epstein**, **A Marden**, *Convex hulls in hyperbolic space, a theorem of Sullivan and measured pleated surfaces*, from: "Analytic and geometric aspects of hyperbolic space", D B A Epstein (editor), LMS Lecture Notes 111, Cambridge University Press, (1987) 112–253

[5] **L Keen**, **C Series**, *Pleating coordinates for the Maskit embedding of the Teichmüller space of punctured tori*, Topology 32 (1993) 719–749

[6] **L Keen**, **C Series**, *Continuity of convex hull boundaries*, Pacific J. Math. 168 (1995) 183–206

[7] **L Keen**, **C Series**, *How to bend pairs of punctured tori*, from: "Lipa's legacy", J Dodziuk and L Keen (editors), Contemporary Mathematics 211 (1997) 359–387

[8] **L Keen**, **C Series**, *Pleating invariants for punctured torus groups*, *Warwick preprint 10/1998*

[9] **C Kourouniotis**, *Deformations of hyperbolic structures* Math. Proc. Cambridge Phil. Soc. 98 (1985) 247–261

[10] **C Kourouniotis**, *The geometry of bending quasi-Fuchsian groups*, from: "Discrete groups and geometry", W J Harvey and C Maclachlan (editors), LMS Lecture Notes 173, Cambridge University Press, (1992) 148–164

[11] **C Kourouniotis**, *Complex length coordinates for quasi-Fuchsian groups*, Mathematika 41 (1994) 173–188

[12] **I Kra**, *Horocyclic coordinates for Riemann surfaces and moduli spaces I: Teichmüller and Riemann spaces of Kleinian groups*, Journal Amer. Math. Soc. 3 (1990) 500–578

[13] **I Kra**, **B Maskit**, *The deformation space of a Kleinian group*, American J. Math. 103 (1980) 1065–1102

[14] **B Maskit**, *On Klein's combination theorem II*, Trans. AMS 131 (1968) 32–39

[15] **B Maskit**, *Kleinian Groups*, Springer–Verlag, (1987)

[16] **C McMullen**, *Complex earthquakes and Teichmüller theory*, J. Amer. Math. Soc. 11 (1998) 283–320

[17] **J Nielsen**, *Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden*, Math. Ann. 78 (1918) 385–397

[18] **J R Parker**, **C Series**, *Bending formulae for convex hull boundaries*, J. d'Analyse Math. 67 (1995) 165–198

[19] **J Parkkonen**, *Geometric complex analytic coordinates for deformation spaces of Koebe groups*, Ann. Acad. Sci. Fenn. Math. Diss. 102 (1995) 1–50

[20] **S P Tan**, *Complex Fenchel-Nielsen coordinates for quasi-Fuchsian structures*, International J. Math. 5 (1994) 239–251

[21] **P L Waterman**, **S A Wolpert**, *Earthquakes and tessellations of Teichmüller space*, Trans. AMS 278 (1983) 157–167

[22] **S A Wolpert**, *The Fenchel–Nielsen deformation*, Annals of Math. 115 (1982) 501–528

[23] **S A Wolpert**, *On the symplectic geometry of deformations of a hyperbolic surface*, Annals of Math. 117 (1983) 207–234

[24] **D J Wright**, *The shape of the boundary of Maskit's embedding of the Teichmüller space of once punctured tori*, preprint

Department of Mathematical Sciences, University of Durham
Durham DH1 3LE, UK

Department of Mathematics, University of Jyväskylä
P.O.Box 35, FIN-40351 Jyväskylä, Finland

Email: `J.R.Parker@durham.ac.uk, parkkone@math.jyu.fi`

Received: 12 November 1997 Revised: 12 August 1998

ISSN 1464-8997 **479**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 479–492

The boundary of the deformation space of the fundamental group of some hyperbolic 3–manifolds fibering over the circle

Leonid Potyagailo

Abstract By using Thurston's bending construction we obtain a sequence of faithful discrete representations ρ_n of the fundamental group of a closed hyperbolic 3–manifold fibering over the circle into the isometry group $Iso\,\mathbf{H}^4$ of the hyperbolic space $\mathbf{H}^4$. The algebraic limit of ρ_n contains a finitely generated subgroup F whose 3–dimensional quotient $\Omega(F)/F$ has infinitely generated fundamental group, where $\Omega(F)$ is the discontinuity domain of F acting on the sphere at infinity $S^3_\infty = \partial\mathbf{H}^4$. Moreover F is isomorphic to the fundamental group of a closed surface and contains infinitely many conjugacy classes of maximal parabolic subgroups.

AMS Classification 57M10, 30F40, 20H10; 57S30, 57M05, 30F10, 30F35

Keywords Discrete (Kleinian) subgroups, deformation spaces, hyperbolic 4–manifolds, conformally flat 3–manifolds, surface bundles over the circle

1 Introduction and statement of results

By a Kleinian (discontinuous) group G we mean a subgroup of the group $\mathrm{Conf}(\mathbf{S}^n) \cong SO_+(1, n+1)$ of conformal transformations of $\overline{R}^n = S^n = R^n \cup \{\infty\}$ which acts discontinuously on a non-empty set $\Omega(G) \subset S^n$ called its domain of discontinuity. It may be connected or not; we will say that G is a function group if there is a connected component $\Omega_G \subset \Omega(G)$ that is invariant under the action of the whole group: $G\Omega_G = \Omega_G$. The quotient spaces $M_G = \Omega_G/G$ and $M(G) = \Omega(G)/G$ are n–manifolds in the case in which G is torsion-free. The complement $\Lambda(G) = (S^n \backslash \Omega(G)) \subset \partial\mathbf{H}^{n+1}$ is called the limit set of G.

A finitely generated Kleinian group G is called geometrically finite if for some $\varepsilon > 0$ there exists an ε–neighbourhood of H_G/G in $\mathbf{H}^{n+1}/G$ which is of finite hyperbolic volume. Here $H_G \subset \mathbf{H}^{n+1}$ is the convex hull of $\Lambda(G)$.

Let us consider for $n = 3$ a hyperbolic 3–manifold $M = H^3/\Gamma$ $(\Gamma \subset PSL_2\mathbf{C})$ fibering over the circle S^1 with fiber a closed surface σ. The notation is $M = \sigma\tilde{\times}S^1$. A representation $\rho\colon \pi_1(M) \to \mathrm{Conf}(\mathbf{S}^3)$ is called admissible if the following conditions are satisfied.

(1) $\rho\colon \Gamma \to \mathrm{Conf}(\mathbf{S}^3)$ is faithful and $\rho(\Gamma) = \Gamma_0$ is Kleinian.

(2) ρ preserves the type of each element, ie $\rho(\gamma)$ is loxodromic for all $\gamma \in \Gamma$.

(3) ρ is induced by a homeomorphism $f_\rho\colon \Omega(\Gamma) \to \Omega(\Gamma_0)$, namely $f_\rho \gamma f_\rho^{-1} = \rho(\gamma)$, $\gamma \in \Gamma$.

The set of all admissible representations modulo conjugation in $\mathrm{Conf}(\mathbf{S}^3)$ is called the deformation space $\mathrm{Def}(\Gamma)$ of the group Γ.

The set $\mathrm{Def}(\Gamma)$ inherits the topology of convergence on generators of Γ on compact subsets in $\mathbf{S}^3$ because $\mathrm{Def}(\Gamma) \subset \left(\mathrm{Conf}(\mathbf{S}^3)\right)^k/\!\sim$, $k \in \mathbf{N}$ ($\sim$ is conjugation in $\mathrm{Conf}(\mathbf{S}^3)$). As $\mathrm{Def}(\Gamma)$ is a bounded domain [13] two questions have arisen. The first is to describe the cases when $\mathrm{Def}(\Gamma)$ is non-trivial and the second is to study the boundary $\partial\,\mathrm{Def}(\Gamma)$, as was done for the classical Teichmüller space [2], [10]. The answer to the first question is still unknown even in the case when M is Haken. We will consider the case when M contains many totally geodesic surfaces. Each of them produces a curve in $\mathrm{Def}(\Gamma)$ by Thurston's "bending" construction [19]. Our main interest is in groups which appear on the boundary $\partial\,\mathrm{Def}(\Gamma)$. These are higher dimensional analogs of B–groups which arise as the limits of sequences of quasifuchsian groups in classical Teichmüller space.

One of the most fundamental questions is to describe the topological type of the orbifold $M(\Gamma) = \Omega(\Gamma)/\Gamma$ (a manifold in the case when Γ is torsion-free), in particular, when Γ is a function group it is important to know when the fundamental group $\pi_1(M_G = \Omega_\Gamma/\Gamma)$ turns out to be finitely generated, or even more generally when it has finite homotopy type.

In dimension 2 the famous theorem of Ahlfors [1] says that a finitely generated non-elementary Kleinian group $G \subset \mathrm{Conf}(\mathbf{R}^2)$ has a factor-space $\Omega(G)/G$ consisting of a finite number of Riemann surfaces $S_1, \ldots, S_n$ each having a finite hyperbolic area.

We discovered in [7] that the weakest topological version of Ahlfors' theorem does not hold starting already with dimension 3. Namely we constructed a finitely generated function group $F \subset \mathrm{Conf}(\mathbf{S}^3)$ such that the group $\pi_1(\Omega_F/F)$ is not finitely generated. Afterwards it was pointed out in [15] that this group is in fact not finitely presented.

It has also been shown that there exists a finitely generated Kleinian group with infinitely many conjugacy classes of parabolics [6].

In [14] we constructed a finitely generated group F_1 such that $\pi_1(\Omega_{F_1}/F_1)$ is not finitely generated and having infinitely many non-conjugate elliptic elements; moreover F_1 appears as an infinitely presented subgroup of a geometrically finite Kleinian group in $\mathbf{H}^4$ without parabolic elements. On the other hand, it was shown in [4] that a finitely generated but infinitely presented group can also appear as a subgroup of a cocompact group in $SO(1,4)$.

Theorem 1 *Let $\Gamma = \pi_1(M)$ be the fundamental group of a hyperbolic $3-$ manifold M fibering over the circle with fiber a closed surface σ. Suppose that Γ is commensurable with the reflection group R determined by the faces of a right-angular polyhedron $D \subset \mathbf{H}^3$. Then there exists a finite-index subgroup $L \subset \Gamma$ and a path β_t: $[0,1[\mapsto \mathrm{Def}(\Gamma)$ such that β_t converges to a faithful representation $\beta_1 \in \partial\mathrm{Def}(\Gamma)$ (as $t \to 1$) and the following hold:*

(1) *$\beta_1(F_L)$ contains infinitely many conjugacy classes of maximal parabolic subgroups,*

(2) *$\pi_1(\Omega_{\beta_1(F_L)})/\beta_1(F_L)$ is infinitely generated,*

where $F_L = L \cap \pi_1\sigma$ is isomorphic to the fundamental group of a closed hyperbolic surface which finitely covers σ and $\beta_1(F_L)$ acts discontinuously on an invariant component $\Omega_{\beta_1(F_L)} \subset \mathbf{S}^3$.

Remark Groups satisfying all the conditions of Theorem 1 do exist. An example of Thurston, of the reflection group in the faces of the right-angular dodecahedron, which is commensurable with a group of a closed surface bundle, is given in [18].

Acknowledgement This paper was prepared several years ago while the author had a Humboldt Fellowship at the Rühr-Universität in Bochum. The author is deeply grateful to Heiner Zieschang and to the Humboldt Foundation for this opportunity. I would also like to thank Nicolaas Kuiper (who died recently) for reading a preliminary version of the manuscript and to express my gratitude to the referee for many useful remarks and corrections.

2 Outline of the proof

Before giving a formal proof of the Theorem let us describe it informally.

Our construction is inspired essentially by papers [6], [8] and [14]. In the first two a free Kleinian group of finite rank satisfying the conclusion (2) was produced, whereas now we give an example of a closed surface group with this property. Our present construction is essentially easier than that of [14]. Also, we produce a curve in the deformation space whose limit point is the group in question.

Step 1 We start with an uniform lattice $\Gamma \subset PSL_2\mathbf{C}$ commensurable with the reflection group R whose limit set is the Euclidean 2–sphere ∂B_1 – the boundary of the ball $B_1 \subset \mathbf{S^3}$. There exists a Fuchsian subgroup $H_2 \subset \Gamma$ leaving invariant a vertical plane π whose intersection with B_1 is a round circle, its limit set $\Lambda(H_2)$ (see figure 1). The group H_2 also leaves invariant a geodesic plane $w_2 \subset B_1$. Consider the action of the group Γ in the outside ball $B_1^* = \mathbf{S^3} \setminus B_1$. For some finite-index subgroup Γ_1 of Γ we construct a new group G_1 obtained by Maskit's Combination theorem from Γ_1 and $\tau_\pi \Gamma_1 \tau_\pi$ combined along the common subgroup $H_2 = \text{Stab}\, w_2$, where τ_π is the reflection in π. The new group G_1 is still isomorphic to some subgroup $G^* \subset R$ of finite index essentially because the same construction can be done inside B_1 by reflecting the picture along the geodesic plane w_2. Thus G_1 belongs to the deformation space $\text{Def}(G_1^*)$. One can obtain a fundamental domain $R(G_1) \subset B_1^*$ of G_1 which is situated in a small neighbourhood of the spheres ∂B_1 and $\tau_\pi(\partial B_1)$.

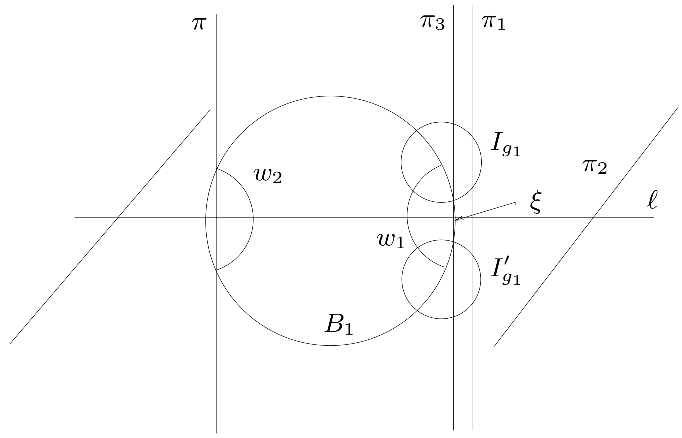


Figure 1

Step 2 There is another geodesic plane $w_1 \subset B_1$ disjoint from w_2 whose stabilizer in Γ_1 is H_1 (see figure 2). Denote by B_2 the ball $\tau_\pi(B_1)$. Take a sphere $\Sigma \subset B_1^*$ passing through the circle $w_3 \cap B_2$ – the limit set of the group $\tau_\pi H_1 \tau_\pi$ – and tangent to the isometric spheres of some element $g_1 \in \Gamma_1$, where H_1 is a subgroup of Γ_1 stabilizing w_1. We now construct a family of Euclidean spheres Σ_t $(0 \leq t \leq 1,\ \Sigma_1 = \Sigma)$ and corresponding groups $\mathcal{G}_t$ obtained as before from G_1 and $\tau_{\Sigma_t} G_1 \tau_{\Sigma_t}$ by using the combination method along common closed surface subgroups. We prove then that there is a path $\beta_t\colon t \in [0,1[\ \mapsto \beta \in \text{Def}(L')$ such that $\beta_0 = L'$, $\beta_t = \mathcal{G}_t$ where L' is some finite-index subgroup of R. One can equally say that β_t is obtained by using Thurston's bending deformation. The main point is now to prove that the limit

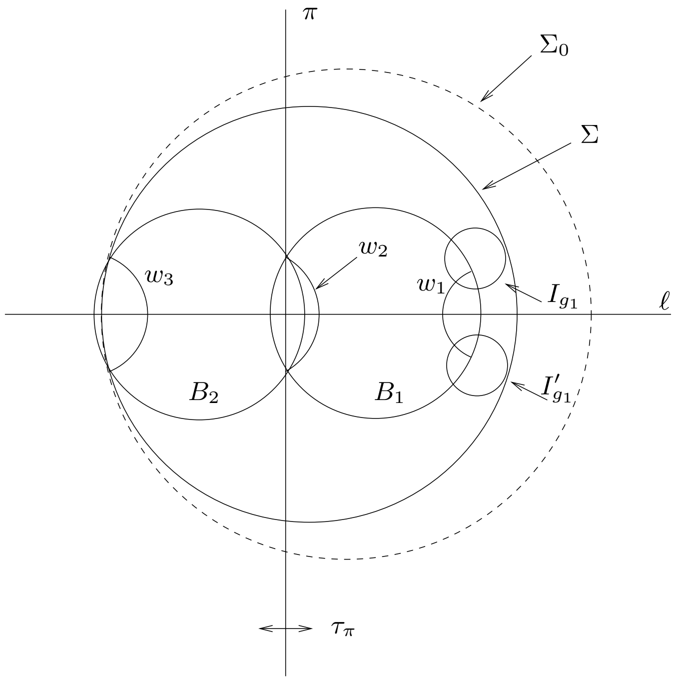


Figure 2

group $\mathcal{G}_1 = \lim_{t\to 1} \beta_t(L')$ is discontinuous and has a fundamental domain obtained from the part of $R(G_1)$ by doubling along the sphere Σ. The group $\mathcal{G}_1$ is also isomorphic to L' and so contains a fundamental group $\mathcal{N}$ of a closed surface bundle over the circle which is isomorphic to the group $L = \Gamma \cap L'$. Let $\mathcal{F}$ be the fundamental group of the fiber given by $\beta_1(F_L = F \cap L)$. Since two isometric spheres of the element $g_1 \in \Gamma_1$ are tangent to Σ, we get a new accidental parabolic element $g = g_1 \cdot g_2$, $g_2 = \tau_\Sigma g_1 \tau_\Sigma$ in the group $\mathcal{G}_1$. By a choice of g_1 made from the very beginning we assure that $g \in \mathcal{F}$, so we have a pseudo-Anosov action of some element $t \in \mathcal{N} \setminus \mathcal{F}$ such that the orbit $t^n \cdot g \cdot t^{-n}$ $(n \in \mathbf{Z})$ gives us infinitely many conjugacy classes of maximal parabolic subgroups of $\mathcal{F}$. Now Scott's compact core theorem implies that $\pi_1(\Omega_{\mathcal{F}})/\mathcal{F}$ is not finitely generated. *End of outline*

3 Preliminaries

We will consider the Poincaré model of hyperbolic space $\mathbf{H}^3$ in the unit ball B_1 equipped with the hyperbolic metric ρ. By a right-anguled polyhedron $D \subset \mathbf{H}^3$ we mean a polyhedron all of whose dihedral angles are $\pi/2$.

Consider the tesselation of $\mathbf{H^3}$ by images of D under the reflection group R from Theorem 1. Denote by $W \subset \mathbf{H}^3$ the collection of geodesic planes w such that there exists $r \in R$, for which $r(w) \cap \partial D$ is a face of D.

It is easy to see that if σ_1 and σ_2 are two faces of D with $\sigma_1 \cap \sigma_2 = \emptyset$, then also the geodesic planes $\tilde{\sigma}_1 \supset \sigma_1$ and $\tilde{\sigma}_2 \supset \sigma_2$ have no point in common. One can easily show that the distance between σ_1 and σ_2, as well as that of $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$, is realized by a common perpendicular ℓ for which $\ell \cap \mathrm{int} D \neq \emptyset$.

Let $\Gamma_0 = R \cap \Gamma$ which is a subgroup of a finite index in both groups R and Γ. By passing to a subgroup of a finite index and preserving notation, we may assume that Γ_0 is a normal subgroup in R, $|R : \Gamma_0| < \infty$. For a plane $w \in W$ we write $H_w = \mathrm{Stab}(w, \Gamma_0) = \{g \in \Gamma_0,\ gw = w\}$. It is not hard to see that H_w is a Fuchsian group of the first kind commensurable with the reflection group determined by the edges of some face of the polyhedron $r(D_1),\ r \in R$.

Let us now fix two disjoint planes w_1 and w_2 from W containing opposite faces of D and let ℓ be their common perpendicular; up to conjugation in Isom $\mathbf{H}^3$ we can assume that ℓ is a Euclidean diameter of B_1. Denote $B_1^* = \mathbf{S}^3 \backslash cl(B_1)$ as well (where $cl(\cdot)$ is the closure of a set). We have the following:

Lemma 1 *For every horosphere π_3 in B_1^* centered at the point $\xi \in \ell \cap \partial B_1$ (see figure 1) there exists $\varepsilon_0 > 0$ such that for every ε–close sphere $\pi_1 \subset B_1^*$ to π_3 ($\varepsilon < \varepsilon_0$) orthogonal to the plane π_2 there exists a geodesic plane w and an element $g_1 \in [H_w, H_w]$ (commutator subgroup) such that:*

$$\begin{gathered} I_{g_1} \cap \pi_1 \neq \emptyset \ \text{ and } \ g_1(I_{g_1} \cap \pi_1) = I'_{g_1} \cap \pi_1, \\ \text{where} \quad I_{g_1},\ I'_{g_1} = I_{g_1^{-1}} \quad \textit{are isometric spheres of } g_1. \end{gathered} \tag{1}$$

Proof Up to further conjugation in Isom B_1 preserving ℓ we may assume that π_3 is the vertical plane tangent to ∂B_1 at $\xi \in \ell \cap \partial B_1$. Take $w = w_1$ and let $g_1 \in [H_{w_1}, H_{w_1}]$ be any primitive element corresponding to a simple dividing loop on the surface w_1/H_{w_1}.

Suppose first that $I_{g_1} \cap \pi_3 = \emptyset$. In this case we proceed as follows. Put $\chi = \tau_{w_1} \circ \tau_{w_2} \in R$, where τ_{w_i} denotes the reflection in plane w_i $(i = 1, 2)$. Then χ is a hyperbolic element whose invariant axis is ℓ. Consider the sequence of planes $\chi^n(w_1)$. We claim that, for some n, $\chi^n(I_{g_1}) \cap \pi_3 \neq \emptyset$. In fact this follows directly from the fact that the fixed point ξ of the hyperbolic element χ is a conical limit point of Γ_0, and so the approximating sequence $\chi^n(I_{g_1})$ should intersect a fixed horosphere (or equivalently by sending ξ to the infinity and passing to the half-space model one can see that χ becomes now a dilation $z \mapsto \lambda z$ $(\lambda > 0)$ which implies that the translations of the image of I_{g_1} by powers of the dilation will intersect a fixed horosphere at infinity). Since Γ_0 is normal in R it now follows that $\chi^n g_1 \chi^{-n} \in [H_{\chi^n(w_1)}, H_{\chi^n(w_1)}] \subset \Gamma_0$ and $\chi^n(I_{g_1}) = I_{\chi^n g_1 \chi^{-n}}$. The latter is true since χ preserves each Euclidean plane passing through $B_1 \cap \ell$ and, hence $(\chi^n g_1 \chi^{-n})|_{\chi^n(I_{g_1})}$ is an Euclidean isometry. So up to replacing w_1 by $\chi^n(w_1)$ and g_1 by $\chi^n g_1 \chi^{-n}$ if needed, we may assume that $I_{g_1} \cap \pi_3 \neq \emptyset$. The same conclusion is then obviously true for a plane $\pi_1 \subset B_1^*$ sufficiently close to π_3.

For $\ell_1 = I_{g_1} \cap \pi_1$ we now claim that $g_1(\ell_1) = \ell_2 = I'_{g_1} \cap \pi_1$. Indeed, $g_1 = \tau_{\pi_2} \cdot \tau_{I_{g_1}}$ where π_2 is orthogonal to π_1 and contains ℓ (figure 1). Evidently

$$g_1(\ell_1) = \tau_{\pi_2}(I_{g_1} \cap \pi_1) = \tau_{\pi_2}(I_{g_1}) \cap \pi_1 = I'_{g_1} \cap \pi_1 \tag{2}$$

since $\tau_{\pi_2}(\pi_1) = \pi_1$. The lemma is proved. □

So we can suppose that $w_1 \in W$ is chosen satisfying all the conclusions of Lemma 1. Let $w_2 \in W$ be a geodesic plane disjoint from w_1 and let ℓ be their common perpendicular passing through the origin of B_1. Now consider the Euclidean plane π orthogonal to ℓ (figure 2) such that

$$\pi \cap \partial B_1 = \pi \cap w_2 \ .$$

It is not hard to see that $\mathrm{Stab}(\pi, \Gamma) = \mathrm{Stab}(w_2, \Gamma) = H_{w_2}$. Reflecting our picture in the plane π we get

$$B_2 = \tau_\pi(B_1) \ , \quad w_3 = \tau_\pi(w_2) \quad \text{and}$$
$$H_{w_3} = \tau_\pi H_{w_1} \tau_\pi \ .$$

By Lemma 1 we can now find a Euclidean sphere Σ centered on ℓ which goes through the circle $w_3 \cap \partial B_2$ and is tangent to I_{g_1} (figure 2). Moreover, by Lemma 1, Σ is tangent also to I'_{g_1}.

Denote $\Sigma' = \tau_\pi^{-1}(\Sigma)$.

Lemma 2 *There exists a subgroup $\Gamma_1 \subset \Gamma_0$ of finite index such that the following conditions hold:*

(a) *The boundary of the isometric fundamental domain $\mathcal{P}(\Gamma_1) \subset B_1^*$ lies in a regular ε–neighbourhood of ∂B_1^** $\big(B_1^* = \mathbf{S}^3 \backslash cl(B_1),\ \varepsilon > 0\big)$.

(b) $\Sigma \cap I_\gamma = \emptyset \ , \quad \gamma \in \Gamma_1 \backslash \{g_1, g_1^{-1}\}$.

(c) *For subgroups $H_1 = \Gamma_1 \cap H_{w_1}, H_2 = \Gamma_1 \cap H_{w_2}$ there exists another fundamental domain $R(\Gamma_1) \subset B_1^*$ of Γ_1 such that*

$$R(\Gamma_1) \cap (\pi \cup \Sigma') = \mathcal{P}(H) \cap (\pi \cup \Sigma'),$$

where $\mathcal{P}(H)$ is an isometric fundamental domain for the group $H = \langle H_1, H_2 \rangle$.

(d) $g_1 \in \Gamma_1 \cap [H_1, H_1]$.

Proof This Lemma can be obtained by repeating the arguments of [14, Main Lemma]. We just sketch these considerations. First, we choose a subgroup $\tilde{\Gamma} \subset \Gamma_0$ of a finite index satisfying conditions (a) and (b) such that $g_1 \in \tilde{\Gamma}$ by using the property of separability of infinite cyclic subgroups in Γ_0 [9].

To obtain (c) we will find Γ_1 by using Scott's *LERF*–property of the group Γ_0 with respect to its geometrically finite subgroups (see [16], [17]). To this end we proceed as follows: the group H is geometrically finite as a result of Klein–Maskit free combination from H_1 and H_2, which are both geometrically finite subgroups of Γ_0. The *LERF* property now says that for the element g_1 there exists a subgroup of Γ_0 of finite index which contains H and does not contain g_1. Call this subgroup Γ_1. Evidently, $g_1 \in [H_1, H_1] \subset \Gamma_1$ by construction. For the complete proof, see [14, Main Lemma]. □

Let us introduce the following notation: $\Omega_1^- = B_1^* \backslash \bigcup_{\gamma \in \Gamma_1} \gamma(\pi^-)$ where π^- is the component of $\mathbf{S}^3 \backslash \pi$ for which $w_3 \in \pi^-$. Let $\Gamma'_1 = \mathrm{Stab}(\Omega_1^-, \Gamma_1)$.

The complete proof of the following assertion can be also found in [14, Lemma 3].

Lemma 3 *The group* $G_1 = \langle \Gamma'_1,\ \tau_\pi \Gamma'_1 \tau_\pi \rangle$ *is discontinuous and*

(1) $G_1 \cong \Gamma'_1 *_{H_2} (\tau_\pi \Gamma'_1 \tau_\pi)$.

(2) G_1 *is isomorphic to a subgroup* $G_1^* \subset R$ *of finite index.*

Sketch of proof (1) This follows from the fact that the plane π is strongly invariant under H_2 in Γ'_1 by [14, Lemma 3.c], which means $H_2\pi = \pi$ and $\gamma\pi \cap \pi = \emptyset$, $\gamma \in \Gamma'_1 \backslash H_2$. One can now get assertion (1) from Maskit's First Combination theorem [11].

(2) Consider the reflection τ_{w_2} in the geodesic plane $w_2 \subset B_1$. We claim that the group $G_1^* = \langle \Gamma'_1, \tau_{w_2} \Gamma'_1 \tau_{w_2} \rangle$ is isomorphic to G_1. Indeed, w_2 is also strongly invariant under H_2 in Γ'_1 and we again observe that $G_1^* = \Gamma'_1 *_{H_2} (\tau_{w_2} \Gamma'_1 \tau_{w_2}) \cong G_1$ because $\tau_{w_2} \mid_{w_2} = \tau_\pi \mid_\pi = id$.

Now $\tau_{w_2} \in R$. Therefore, $G_1^* \subset R$ and G_1^* has a compact fundamental domain $R(G_1^*) = R(\Gamma'_1) \cap \tau_{w_2}(R(\Gamma'_1))$. The covering $\mathbf{H}^3 / (G_1^* \cap \Gamma_0) \to \mathbf{H}^3 / G_1^*$ is finite since $|R : \Gamma_0| < \infty$ and, hence, the manifold $M\,(G_1^* \cap \Gamma_0) = \mathbf{H}^3 / (G_1^* \cap \Gamma_0)$ is compact. Thus, the covering $M(G_1^* \cap \Gamma_0) \to M(\Gamma_0)$ is finite as well and so $|\Gamma_0 : G_1^* \cap \Gamma_0 | < \infty$. □

Corollary 4 *There exists a path* α_t: $[0,1] \to Def(G_1^*)$ *such that* $\alpha_0 = G_1^*$ *and* $\alpha_1 = G_1$.

Proof By choosing a continuous family of spheres μ_t for which $\mu_t \cap \pi = w_2 \cap \pi = \Lambda(H_2)$, $\mu_0 \supset w_2$, $\mu_1 = \pi$, $t \in [0,1)$, we construct the family of groups $G_t = \langle \Gamma_1', \tau_{\mu_t} \Gamma_1' \tau_{\mu_t} \rangle$ by the arguments of Lemma 3. Consider now the action of Γ_1' in B_1^* where $p_1\colon B_1^* \to B_1^*/\Gamma_1$ is the covering map. The surfaces $p_1(\mu_t)$ are all embedded and parallel due to condition (b). If now Ω_{G_t} is the component of G_1 containing ∞ then the manifold $M_{G_t} = \Omega_{G_t}/G_t$ is homeomorphic to the double of the manifold $M_1^- = \Omega_1^-/\Gamma_1'$ along the boundary $p_1(\pi)$. Thus, for all $t \in [0,1]$, M_{G_t} are all homeomorphic and there exists a continuous family of homeomorphisms $f_t\colon \Omega(G_1^*) \to \Omega(G_t)$ such that $G_t = f_t G_1^* f_t^{-1}$, $G_1 = f_1 G_1^* f_1^{-1}$. □

By construction the domain $R(G_1) = R(\Gamma_1') \cap \tau_\pi(R(\Gamma_1'))$ is fundamental for the action of G_1 in Ω_{G_1}.

Claim 5 $R(G_1) \cap \Sigma = \left(\mathcal{P}(H_3) \cup I_{g_1} \cup I_{g_1}'\right) \cap \Sigma$.

Proof Recall that $\pi^+(\pi^-)$ means the right (left) component of $\mathbf{S}^3 \backslash \pi$ $(I_{g_1} \in \pi^+)$. Then $\pi^+ \cap \Sigma \cap R(\Gamma_1') = \mathcal{P}(H_1) \cap \Sigma = \left(I_{g_1} \cup I_{g_1}'\right) \cap \Sigma$ by (b) and (c) of Lemma 2.

Also, $\tau_\pi(\pi^- \cap \Sigma \cap \tau_\pi(R(\Gamma_1'))) = \pi^+ \cap \tau_\pi(\Sigma) \cap R(\Gamma_1') \subset \mathcal{P}(H_1) \cap \Sigma'$, so $\pi^- \cap \Sigma \cap R(G_1) = \tau_\pi(\mathcal{P}(H_1)) \cap \Sigma = \mathcal{P}(H_3) \cap \Sigma$. □

Let us consider now the family of spheres Σ_t centered on the y–axis (figure 2) such that $\Sigma_t \cap w_3 = \Sigma \cap w_3$, $\sigma_1 = \Sigma$, $\sigma_0 = \Sigma_0$, $t \in [0,1]$, where $\Sigma_t \cap \mathrm{ext}(B_1) \cap \mathrm{ext}(B_2) \subset \mathrm{ext}(\Sigma) \cap \mathrm{ext}(B_1) \cap \mathrm{ext}(B_2)$ (recall $\mathrm{ext}(\cdot)$ is the exterior of a set in $\overline{R}^3$), $\Sigma_t \cap I_{g_1} = \emptyset$ $(t > 0)$. Denote by τ_{Σ_t} the corresponding reflections. As before take the domain $\Omega^* = \Omega_{G_1} \backslash G_1(\Sigma_0^-)$ and the group $G_1' = \mathrm{Stab}(\Omega^*, G_1)$, where $\Sigma_0^- = \mathrm{ext}\ (\Sigma_0)$ is the unbounded component of $\overline{R}^3 \backslash \Sigma_0$.

Denote $\mathcal{G}_t = \langle G_1',\ \tau_{\Sigma_t} G_1' \tau_{\Sigma_t} \rangle$. Evidently, $\mathcal{G}_1 = \lim_{t \to 1} \mathcal{G}_t$.

Lemma 6 *The groups $\mathcal{G}_t$ are discontinuous, $t \in [0,1]$.*

Proof First, let us prove the lemma for $t \neq 1$. By Claim 5 we have now that $R(G_1) \cap \Sigma_t = \mathcal{P}(H_3) \cap \Sigma_t$. Moreover we claim also that

$$\begin{aligned} g\Sigma_t \cap \Sigma_t = \emptyset,\ g \in G_1 \backslash H_3,\ H_3 \Sigma_t = \Sigma_t, \\ \text{where} \quad H_3 = \tau_\pi H_1 \tau_\pi\ . \end{aligned} \tag{3}$$

To prove (3) we only need to show that $g(\Sigma_t \cap \Lambda(H_3)) \cap (\Sigma_t \cap \Lambda(H_3)) = \emptyset$, but this can be shown from the fact that each point of $\Lambda(H_3)$ is a point of approximation (see [14, Claim 1]).

All conditions of Maskit's First Combination theorem are now satisfied for the groups G_1' and $\tau_{\Sigma_t} G_1' \tau_{\Sigma_t}$ $(t \neq 1)$ [11] and we obtain also

$$\mathcal{G}_t \cong G_1' *_{H_3} (\tau_{\Sigma_t} G_1' \tau_{\Sigma_t}) \tag{4}$$

where the $\mathcal{G}_t$ are all discontinuous, $t \in [0,1)$.

Let us now consider the group $\mathcal{G}_1$ and the domain $R(\mathcal{G}_1) = R(G_1) \cap \tau_\Sigma(R(G_1))$. Our goal now is to show that $R(\mathcal{G}_1)$ is a fundamental domain for the action of $\mathcal{G}_1$ in $\Omega_{\mathcal{G}_1}$ $(\infty \in \Omega_{\mathcal{G}_1})$. If now $\langle g_1, \gamma_1, \ldots, \gamma_\ell \rangle$ is a set of generators of G_1' then $S = \langle g_1, \gamma_1, \ldots, \gamma_\ell,\ g_2, \gamma_1', \ldots, \gamma_\ell' \rangle$ are generators of $\mathcal{G}_1$, where $\gamma_i' = \tau_\Sigma \cdot \gamma_i \cdot \tau_\Sigma$ and $g_2 = \tau_\Sigma \cdot g_1 \cdot \tau_\Sigma$. Observe that the element g_1 is included in S because some of its isometric spheres belong to the boundary $\partial R(G_1')$

We want to apply the Poincaré Polyhedron theorem [12]. Indeed, an arbitrary cycle of edges in $\partial R(\mathcal{G}_1)$ consists either of edges situated in $\partial(R(G_1)) \cap \mathrm{int}(\Sigma)$, and $\partial(\tau_\Sigma(R(G_1))) \cap \mathrm{ext}(\Sigma)$, or is an edge cycle $\ell_1 = I_{g_1} \cap I_{g_2}$, $\ell_2 = I_{g_1}' \cap I_{g_2}'$, where I_{g_k}, I_{g_k}' are the isometric spheres of g_k and $g_k^{-1} (k = 1,2)$. The sum of angles in any cycle of the first type is 2π because $R(G_1)$ is a fundamental domain [12].

We now claim that the element $g = g_2^{-1} \cdot g_1$ is parabolic with a fixed point $d = I_{g_1} \cap I_{g_2}$. Indeed, $g_2^{-1} \cdot g_1 = \left(\tau_\Sigma \cdot \tau_{I_{g_1}}\right)^2$ because $g_1 = \tau_{\pi_2} \cdot \tau_{I_{g_1}}$ and π_2 is orthogonal to Σ (figure 2). Now it is easy to check that $g(d) = d$, $gI_{g_1} \subset \mathrm{int}(I_{g_2})$ and $g(\mathrm{int}(I_{g_1})) = \mathrm{ext}\ (g(I_{g_1}))$, therefore the elements g and $g' = g_1 \cdot g \cdot g_1^{-1}$ are parabolics.

All conditions of the Maskit–Poincaré theorem are valid at the edges ℓ_i also and, hence, $\mathcal{G}_1$ is discontinuous. Lemma 6 is proved. □

Lemma 7 *The group $\mathcal{G}_0$ is isomorphic to a subgroup $L' \subset R$ of a finite index.*

Proof We repeat our construction of $\mathcal{G}_0$ by modelling it in $\mathbf{H}^3$ so as to get the required isomorphism.

Recall that we started from the group $\Gamma_1' \subset \mathrm{Isom}(\mathbf{H}^3)$ and showed that $G_1 = \langle \Gamma_1', \tau_\pi \Gamma_1' \tau_\pi \rangle \cong G_1^* = \langle \Gamma_1', \tau_{w_2} \Gamma_1' \tau_{w_2} \rangle$ (see Lemma 4). Next we constructed $\mathcal{G}_0$ by using reflection in $\sigma_0 = \Sigma_0$ such that $\sigma_0 \cap w_3 = \Lambda(H_3)$, $\sigma_0 \cap B_1 = \emptyset$, $w_3 = \tau_\pi(w_1)$.

Let $\eta = \tau_{w_2}(w_1) \subset \mathbf{H}^3$, $\eta \in W$. Again let us take the subgroup G_1^{**} of G_1^* which is $G_1^{**} = \mathrm{Stab}(\mathbf{H}^3 \backslash G_1^*(\eta^-),\ G_1^*)$, where η^- is a subspace $\mathbf{H}^3 \backslash \eta$ not containing w_2.

By construction the fundamental domain $R(G_1^*) = R(\Gamma'_1) \cap \tau_{w_2}(R(\Gamma'_1))$ of the group G_1^* satisfies $R(G_1^*) \cap \eta = \mathcal{P}(H'_3 = \mathrm{Stab}(\eta, G_1^*))$. Again by Maskit's First Combination theorem we have a group L':

$$L' = G_1^{**} *_{H'_3} (\tau_\eta G_1^{**} \tau_\eta) \tag{5}$$

We constructed an isomorphism $\varphi_1\colon G_1^* \to G_1$ in Lemma 4 such that $\tau_\pi \cdot \varphi_1 \cdot \tau_{w_2} = \varphi_1$, therefore $\varphi_1(H'_3) = H_3$ and $\varphi_1(G_1^{**}) = G'_1$. It follows now from (4) and (5) that the map $\varphi_1\big|_{G_1^{**}}$ can be extended to an isomorphism $\varphi\colon L' \to \mathcal{G}_0$.

Index $|R : L'|$ is finite because L' has a compact fundamental domain. The Lemma is proved. □

Recall that we identify $[\rho] \in \mathrm{Def}(L')$ with $\rho(L')$.

Lemma 8 *There exists a path* $\beta_t\colon [0,1] \to cl(\mathrm{Def}(L'))$ *such that* $\beta_0 = L'$, $\beta_1 = \mathcal{G}_1 \in \partial\,\mathrm{Def}(L')$, $\beta_t([0,1)) \subset \mathrm{Def}(L')$.

Proof We have constructed a path $\alpha_t\colon [0,1] \to \mathrm{Def}(G_1^*)$ in Corollary 4 such that $\alpha_0 = G_1^*$, $\alpha_1 = G_1$ and α_t is a family of admissible representations. Let further $\alpha_t\big|_{G_1^{**}} = \alpha'_t$. Obviously, the representations α'_t are also admissible and $\alpha'_1(G_1^{**}) = G'_1$. We can easily extend our family α'_t to a family of admissible representations $\theta_t\colon L' \to \mathrm{Def}(L')$ by the formula $\theta_t = \tau_{\mu_t}\alpha'_t\tau_{\mu_t}$, where μ_t are the spheres constructed in Corollary 4.

Observe that $\mu_1 = \pi$ and now take a new continuous family of spheres ν_t for which $\nu_t \cap w_3 = \Lambda(H_s) = w_3 \cap B_2$ and $\nu_1 = \tilde{w}_3$, $\nu_2 = \Sigma_0$ where $\tilde{w}_3$ is the sphere containing w_3 $(t \in [0,1])$.

Again we have a path $\theta'_t(L') = \langle G'_1, \tau_{\nu_t} G'_1 \tau_{\nu_t} \rangle$. Composing the path θ_t with θ'_t and with the path corresponding to spheres Σ_t connecting Σ_0 with Σ_1 we get required path β_t. The Lemma is proved. □

4 Proof of Theorem 1

(1) Denote by $F = \pi_1\sigma$ a fixed fiber group of our initial manifold M, and let also $F_0 = \Gamma_0 \cap F$.

By Jørgensen's theorem [5] the limit $\beta_1 = \lim\limits_{t\to 1} \beta_t$ is an isomorphism $\beta_1\colon L' \to \mathcal{G}_1$. Let us consider the subgroup $L = L' \cap \Gamma_0$, $|\Gamma_0 : L| < \infty$. Put also $F_L = L \cap F_0$ for its normal subgroup. We have also the curve $\beta_t(L) \subset \mathrm{Def}(L)$. Let $\mathcal{N} = \beta_1(L)$, $\mathcal{F} = \beta_1(F_L)$. Let us show that $g = g_2^{-1} \cdot g_1 \in \mathcal{F}$. To this end let us recall that the element g_1 was chosen from the very beginning being in $[H_{w_1}, H_{w_1}]$ (Lemma 1). Recalling also that $\beta_1^{-1}(g_1) = g_1$ and denoting $\beta_1^{-1}(g_2) = g_2'$, by construction we get $g_2' = \tau_\eta \cdot g_1 \cdot \tau_\eta$, $\eta = \tau_{w_2}(w_1)$, $g_1 \in [H_{w_1}, H_{w_1}] \subset [F_0, F_0]$ (see Lemma 1). The group Γ_0 was chosen to be normal in the reflection group R, and since $[\Gamma_0, \Gamma_0] \subset F$, it is straightforward to see that

$$r[F_0, F_0]r^{-1} \subset F_0, \quad r \in R\ .$$

Hence, $g_2' \in F_0$, and for the element $g' = (g_2')^{-1} \cdot g_1$ we immediately obtain $g' \in F_L = F_0 \cap L'$. It follows that $\beta_1(g') = g = g_2^{-1} \cdot g_1 \in F_0 \cap \mathcal{G}_1 = \mathcal{F}$ as was promised.

We have that $\mathcal{N}$ is isomorphic to the semi-direct product of $\mathcal{F}$ and the infinite cyclic group $\mathbf{Z}$, so taking the element $t \in \mathcal{N} \backslash \mathcal{F}$ projecting to the generator of $\mathcal{N}/\mathcal{F}$, we observe that the elements

$$g_n = t^n g t^{-n} \in \mathcal{F}\ , \quad g \in \mathcal{F}, \quad n \in \mathbf{Z} \tag{6}$$

are all parabolics. Since $\mathcal{N}$ contains no abelian subgroups of rank bigger than 1 and $t^n \notin \mathcal{F}$ $(n \in \mathbf{Z})$ one can easily see that the elements (6) are also non-conjugate in $\mathcal{F}$. We have proved (1) of the Theorem.

(2) By the construction, the fundamental polyhedron $R(\mathcal{G}_1)$ of the group $\mathcal{G}_1$ contains only one conjugacy class of parabolic elements g of rank 1. There is a strongly invariant cusp neighborhood $B_g \cong [0,1] \times R^1 \times [0, \infty)$ which comes from the construction of $R(\mathcal{G}_1)$. So each parabolic g_n of type (6) gives rise to submanifold

$$B_{g_n} \big/ \langle g_n \rangle \cong T_n \times [0, \infty),\ T_n \cong S^1 \times S^1 \tag{7}$$

in the manifold $M(\mathcal{F}) = \Omega_N \big/ \mathcal{F}$. Therefore $M(\mathcal{F})$ contains infinitely many parabolic ends (7) bounded by tori T_n. They all are non-parallel in $M(\mathcal{F})$ and therefore by Scott's "core" theorem the group $\pi_1(M(\mathcal{F}))$ is not finitely generated [16]. □

Remark By using the argument of [14] one can prove:

Theorem 2 *There is a (non-faithful) represention $\beta_{1+\varepsilon}$ which is ε–close to β_1 for some small $\varepsilon > 0$ such that the group $\beta_{1+\varepsilon}(F_L)$ is infinitely generated, has infinitely many non-conjugate elliptic elements. Moreover, $\beta_{1+\varepsilon}(F_L)$ is a normal infinitely presented subgroup of a geometrically finite group $\beta_{1+\varepsilon}(L)$ without parabolics.*

To prove the theorem one can continue to deform the group for $1 < t \leq 1+\varepsilon$ (these representations will no longer be faithful) in order to get an elliptic element g_t whose isometric spheres form an angle $\theta(t)$ instead of being tangent. To do this in our Lemma 2, instead of the sphere Σ tangent to the isometric spheres of g_1, one needs to consider a nearby sphere $\Sigma_{1+\varepsilon}$ forming angle $\theta(\varepsilon)$ with them. If $\theta(\varepsilon) = \frac{\pi}{2n}$ and $n > 0$ is large enough the group $\beta_{1+\varepsilon}(F_L)$ is Kleinian, has infinitely many non-conjugate elliptic elements of the order n (obtained as above as an orbit of $g_{1+\varepsilon}$ by a pseudo-Anosov automorphism of the $\beta_{1+\varepsilon}(F_L)$). The construction gives us that $\beta_{1+\varepsilon}(F_L)$ is a normal and finitely generated but infinitely presented subgroup of the geometrically finite group $\beta_{1+\varepsilon}(L)$ without parabolic elements. In particular $\beta_{1+\varepsilon}(L)$ is a Gromov hyperbolic group (see [14, Lemmas 5–7]).

References

[1] **L V Ahlfors**, *Finitely generated Kleinian groups*, Amer. J. Math. 86 (1964) 413–429; 87 (1965) 759

[2] **L Bers**, *On boundaries of Teichmüller spaces and on Kleinian groups I*, Annals of Math. 91 (1970) 570–600

[3] **B Bowditch**, *Geometrical finiteness of hyperbolic groups*, preprint, University of Melbourne

[4] **B Bowditch**, **G Mess**, *A 4–Dimensional Kleinian Group*, Trans. Amer. Math. Soc. 344 (1994) 391–405

[5] **T Jørgensen**, *Compact* 3*–manifolds of constant negative curvature fibering over the circle*, Annals of Math. 106 (1977) 61–72

[6] **M Kapovich**, *On Absence of Sullivan's cusp finiteness theorem in higher dimensions*, preprint

[7] **M Kapovich**, **L Potyagailo**, *On absence of Ahlfors' finiteness theorem for Kleinian groups in dimension* 3, Topology and its Applications, 40 (1991) 83–91

[8] **M Kapovich**, **L Potyagailo**, *On absence of Ahlfors' and Sullivan's finiteness theorems for Kleinian groups in higher dimensions*, Siberian Math. Journal 32 (1992) 61–73

[9] **D Long**, *Immersions and embeddings of totally geodesic surfaces*, Bull. London Math. Soc. 19 (1987) 481–484

[10] **B Maskit**, *On boundaries of Teichmüller spaces and on Kleinian groups, II*, Annals of Math. 91 (1970) 608–638

[11] **B Maskit**, *On Klein's Combination theorem III*, from: "Advances in the theory of Riemann Surfaces", Princeton Univ. Press (1971) 297–310

[12] **B Maskit**, *Kleinian groups*, Springer–Verlag (1988)

[13] **J Morgan**, *Group action on trees and the compactification of the space of conjugacy classes of* $SO(n,1)$ *–representations*, Topology 25 (1986) 1–33

[14] **L Potyagailo**, *Finitely generated Kleinian groups in 3–space and 3–manifolds of infinite homotopy type*, Trans. Amer. Math. Soc. 344 (1994) 57–77

[15] **L Potyagailo**, *The problem of finiteness for Kleinian groups in* 3 *–space*, from: "Proceedings of International Conference, Knots-90", Osaka (1992)

[16] **P Scott**, *Finitely generated* 3 *–manifold groups are finitely presented*, J. London Math. Soc. 6 (1973) 437–440

[17] **P Scott**, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. 17 (1978) 555–565; Correction ibid 32 (1985) 217–220

[18] **D Sullivan**, *Travaux de Thurston sur les groupes quasi-fuchsiens et les varietes hyperboliques de dimension* 3 *fibres sur* S^1, Lecture Notes in Math, 842, Springer–Verlag, Berlin–New York (1981) 196–214

[19] **W Thurston**, *The geometry and topology of* 3 *–manifolds*, Princeton University Lecture Notes (1978)

Département de Mathématiques
Université de Lille 1
59655 Villeneuve d'Ascq, France

Email: `potyag@gat.univ-lille1.fr`

Received: 20 November 1997 Revised: 7 November 1998

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 493–509

Hairdressing in groups: a survey of combings and formal languages

Sarah Rees

Abstract A group is combable if it can be represented by a language of words satisfying a fellow traveller property; an automatic group has a synchronous combing which is a regular language. This article surveys results for combable groups, in particular in the case where the combing is a formal language.

AMS Classification 20F10, 20-04, 68Q40; 03D40

Keywords Combings, formal languages, fellow travellers, automatic groups

Dedicated to David Epstein on the occasion of his 60th birthday

1 Introduction

The aim of this article is to survey work generalising the notion of an automatic group, in particular to classes of groups associated with various classes of formal languages in the same way that automatic groups are associated with regular languages.

The family of automatic groups, originally defined by Thurston in an attempt to abstract certain finiteness properties of the fundamental groups of hyperbolic manifolds recognised by Cannon in [12], has been of interest for some time. The defining properties of the family give a geometrical viewpoint on the groups and facilitate computation with them; to such a group is associated a set of paths in the Cayley graph of the group (a 'language' for the group) which both satisfies a geometrical 'fellow traveller condition' and, when viewed as a set of words, lies in the formal language class of regular languages. (A formal definition is given in section 2.) Epstein et al.'s book [15] gives a full account; the papers [3] and [16] are also useful references (in particular, [16] is very readable and non-technical).

The axioms of an automatic group are satisfied by all finite groups, all finitely generated free and abelian groups, word hyperbolic groups, the fundamental groups of compact Euclidean manifolds, and of compact or geometrically finite hyperbolic manifolds [15, 26], Coxeter groups [10], braid groups, many Artin groups [13, 14, 28, 24], many mapping class groups [27], and groups satisfying various small cancellation conditions [18]. However some very interesting groups are not automatic; the family of automatic groups fails to contain the fundamental groups of compact 3–manifolds based on the *Nil* or *Sol* geometries, and, more generally, fails to contain any nilpotent group (probably also any soluble group) which is not virtually abelian. This may be surprising since nilpotent groups have very natural languages, with which computation is very straightforward.

A family of groups which contains the fundamental groups of all compact, geometrisable 3–manifolds was defined by Bridson and Gilman in [9], through a weakening of both the fellow traveller condition and the formal language requirement of regularity for automatic groups. The fellow traveller condition was replaced by an asynchronous condition of the same type, and the regularity condition by a requirement that the language be in the wider class of 'indexed languages'. The class of groups they defined can easily be seen to contain a range of nilpotent and soluble groups.

Bridson and Gilman's work suggests that it is sensible to examine other families of groups, defined in a similar way to automatic groups with respect to other formal language classes. This paper surveys work on this theme. It attempts to be self contained, providing basic definitions and results, but referring the reader elsewhere for fuller details and proofs. Automatic groups are defined, and their basic properties described in section 2; the more general notion of combings is then explained in section 3. A basic introduction to formal languages is given in section 4 for the sake of the curious reader with limited experience in this area. (This section is included to set the results of the paper into context, but all or part of it could easily be omitted on a first reading.) Section 5 describes the closure properties of various classes of combable groups, and section 6 gives examples (and non-examples) of groups with combings in the classes of regular, context-free, indexed and real-time languages.

Acknowledgment The author would like to thank the Fakultät für Mathematik of the Universität Bielefeld for its warm hospitality while this work was carried out, and the Deutscher Akademischer Austauschdienst for financial support.

2 Automatic groups

Let G be a finitely generated group, and X a finite generating set for G, and define X^{-1} to be the set of inverses of the elements of X. We define a *language* for G over X to be a set of *words* over X (that is, products in the free monoid over $X \cup X^{-1}$) which maps onto G under the natural homomorphism; such a language is called *bijective* if the natural map is bijective.

The group G is automatic if it possesses a language satisfying two essentially independent conditions, one a geometric 'fellow traveller condition', relating to the Cayley graph Γ for G over X, the other a restriction on the computational complexity of the language in terms of the formal language class in which the language lives. Before a precise definition of automaticity can be given, the fellow traveller condition needs to be explained.

Figure 1 gives an informal definition of fellow travelling; we give a more formal definition below. In the figure, the two pairs of paths labelled 1 and 2, and

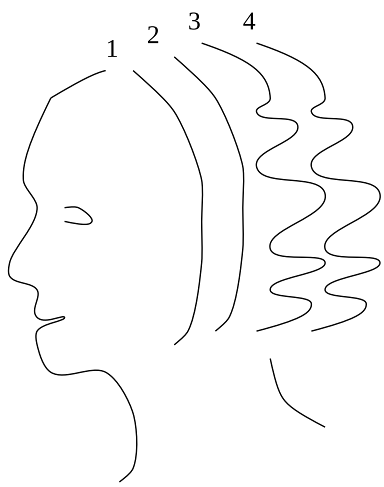


Figure 1: Fellow travellers

3 and 4 synchronously fellow travel at a distance approximately equal to the length of the woman's nose; the pair of paths labelled 2 and 3 asynchronously fellow travel at roughly the same distance. Particles moving at the same speeds along 1 and 2, or along 3 and 4, keep abreast; but a particle on 3 must move much faster than a particle on 2 to keep close to it.

More formally let Γ be the Cayley graph for G over X. (The vertices of Γ correspond to the elements of G, and an edge labelled by x leads from g to gx, for each $g \in G, x \in X$). A word w over X is naturally associated with the finite path γ_w labelled by it and starting at the identity in Γ. The path γ_w can be parametrised by continuously extending the graph distance function d_Γ (which gives edges length 1); where $|w| = d_\Gamma(1, w)$ is the string length of w, for $t \leq |w|$, we define $\gamma_w(t)$ to be a point distance t along γ_w from the identity vertex, and, for $t \geq |w|$, $\gamma_w(t)$ to be the endpoint of γ_w. Two paths γ_1 and γ_2 of Γ are said to *synchronously K–fellow travel* if, for all $t \geq 0$, $d_\Gamma(\gamma_1(t), \gamma_2(t)) \leq K$, and *asynchronously K–fellow travel* if a strictly increasing positive valued function $h = h_{\gamma_1,\gamma_2}$ can be defined on the positive real numbers, mapping $[0, l(\gamma_1) + 1]$ onto $[0, l(\gamma_2) + 1]$, so that, for all $t \geq 0$, $d_\Gamma(\gamma_1(t), \gamma_2(h(t))) \leq K$.

Precisely, G is *automatic* if, for some generating set X, G has a language L over X satisfying the following two conditions. Firstly, for some K, and for any $w, v \in L$ for which γ_v and γ_w lead either to the same vertex or to neighbouring vertices of Γ, γ_v and γ_w synchronously K–fellow travel. Secondly L is regular. A language is defined to be regular if it is the set of words accepted by a finite state automaton, that is, the most basic form of theoretical computer; the reader is referred to section 4 for a crash course on automata theory and formal languages. The regularity of L ensures that computation with L is easy; the fellow traveller property ensures that the language behaves well under multiplication by a generator. Although this is not immediately obvious, the definition of automaticity is in fact independent of the generating set for G; that is, if G has a regular language over some generating set satisfying the necessary fellow traveller condition, it has such a language over every generating set.

If G is automatic, then G is finitely presented and has quadratic isoperimetric inequality (that is, for some constant A, any loop of length n in the Cayley graph Γ can be divided into at most An^2 loops which are labelled by relators). It follows that G has soluble word problem, and in fact there is a straightforward quadratic time algorithm to solve that.

If G is automatic, then so is any subgroup of finite index in G, or quotient of G by a finite normal subgroup, as well as any group in which G is a subgroup of finite index, or of which G is a quotient by a finite normal subgroup. The family of automatic groups is also closed under the taking of direct products, free products (with finite amalgamation), and HNN extensions (over finite subgroups), but not under passage to arbitrary subgroups, or under more general products or extensions.

3 Combings

In an attempt to find a family of groups which has many of the good properties of automatic groups, while also including the examples which are most clearly missing from that family, we define *combable* groups, using a variant of the first axiom for automatic groups.

Let $G = \langle X \rangle$ be a finitely generated group with associated Cayley graph Γ. We define an *asynchronous combing*, or *combing* for G to be a language L for G with the property that for some K, and for any $w, v \in L$ for which γ_v and γ_w lead either to the same vertex or to neighbouring vertices of Γ, γ_v and γ_w asynchronously K–fellow travel; if G has a combing, we say that G is combable. Similarly, we define a *synchronous combing* to be a language for which an analogous synchronous fellow traveller condition holds; hence automatic groups have synchronous combings. Of course, every synchronous combing is also an asynchronous combing.

In the above definitions, we have no requirement of bijectivity, no condition on the length of words in L relative to geodesic words, and no language theoretic restriction. In fact, the term 'combing' has been widely used in the literature, with various different meanings, and some definitions require some of these properties. Many authors require combings to be bijective; in [15] words in the language are required to be quasigeodesic, and in [17] combings are assumed to be synchronous.

The term 'bicombing' is also fairly widely used in the literature, and so, although we shall not be specifically interested in bicombability here, we give a definition for the sake of completeness. Briefly a bicombing is a combing for which words in the language related by left multiplication by a generator also satisfy a fellow traveller property. Specifically, a combing L is a (synchronous, or asynchronous) *bicombing* if paths of the form γ_v and $x\gamma_w$ (synchronously, or asynchronously) fellow travel, whenever $\gamma_v, \gamma_w \in L$, $x \in X$, and $v =_G xw$, and where $x\gamma_w$ is defined to be the concatenation of x and a path from x to xw following edges labelled by the symbols of the word γ_w. A group is *biautomatic* if it has a synchronous bicombing which is a regular language.

Most known examples of combings for non-automatic groups are not known to be synchronous; certainly this is true of the combings for the non-automatic groups of compact, geometrisable 3–manifolds found by Bridson and Gilman. However, in recent and as yet unpublished work, Bestvina and N. Brady have constructed a synchronous, quasigeodesic (in fact linear) combing for a non-automatic group. By contrast, Burillo, in [11], has shown that none of the

Heisenberg groups

$$\begin{aligned} H_{2n+1} &= \langle x_1, \dots x_n, y_1, \dots y_n, z \mid [x_i, y_i] = z, \forall i, \\ & [x_i, x_j] = [y_i, y_j] = [x_i, y_j] = 1, \forall i, j, i \neq j \rangle \end{aligned}$$

or the groups $U_n(\mathbf{Z})$ of n by n unipotent upper-triangular integer matrices can admit synchronous combings by quasigeodesics (all of these groups are asynchronously combable). Burillo's result was proved by consideration of higher-dimensional isoperimetric inequalities; the case of H_3 had been previously dealt with in [15].

Let G be a combable group. Then, by [7] theorem 3.1, G is finitely presented, and, by [7] theorems 4.1 and 4.2, G has an exponential isoperimetric inequality; hence G has soluble word problem (see [15], theorem 2.2.5). By [17], if G has a synchronous, 'prefix closed' combing (that is, all prefixes of words in the language are in the language), then G must actually have a quadratic isoperimetric inequality. Note that, by [25] (or see [4]), there are finitely presented class 3 soluble groups which have insoluble word problem, and so certainly cannot be combable.

For a combing to be of practical use, it must at least be recognisable. It is therefore natural to consider combings which lie in some formal language class, or rather, which can be defined by some theoretical model of computation. Automatic groups are associated with the most basic such model, that is, with finite state automata and regular languages. In general, where $\mathcal{F}$ is a class of formal languages we shall say that a group is $\mathcal{F}$–combable if it has a combing which is a language in $\mathcal{F}$. Relevant formal languages are discussed in section 4.

An alternative generalisation of automatic groups is discussed in [5]. This approach recognises that the fellow traveller condition for a group with language L implies the regularity of the language L' of pairs of words in L which are equal in the group or related by right multiplication by a generator, and examines what happens when both L and L' are allowed to lie in a wider language class (in this particular case languages are considered which are intersections of context-free languages, and hence defined by series of pushdown automata). Some of the consequences of such a generalisation are quite different from those of the case of combings; for example, such groups need not be finitely presented.

4 Hierarchy of computational machines and formal languages

Let A be a finite set of symbols, which we shall call an *alphabet*. We define a *language* L over A to be a set of finite strings (words) over A, that is a subset of $A^* = \cup_{i\in\mathbf{N}} A^i$. We define a *computational machine* M for L to be a device which can be used to recognise the words in L, as follows. Words w over A can be input to M one at a time for processing. If w is in L, then the processing of w terminates after some finite time, and M identifies w as being in L; if w is not in L, then either M recognises this after some time, or M continues processing w indefinitely. We define L to be a *formal language* if it can be recognised by a computational machine; machines of varying complexity define various families of formal languages.

We shall consider various different types of computational machines. Each one can be described in terms of two basic components, namely a finite set S of *states*, between which M fluctuates, and (for all but the simplest machines) a possibly infinite *memory* mechanism. Of the states of S, one is identified as a *start state* and some are identified as *accept states*. Initially (that is, before a word is read) M is always in the start state; the accept states are used by M to help it in its decision process, possibly (depending on the type of the machine) in conjunction with information retrieved from the memory.

We illustrate the above description with a couple of examples of formal languages over the alphabet $A = \{-1, 1\}$, and machines which recognise them.

We define L_1 to be the language over A consisting of all strings containing an even number of 1's. This language is recognised by a very simple machine M_1 with two states and no additional memory. S is the set $\{even,\ odd\}$; *even* is the start state and only accept state. M_1 reads each word w from left to right, and switches state each time a 1 is read. The word w is accepted if M_1 is in the state *even* when it finishes reading w. M_1 is an example of a (deterministic) finite state automaton.

We define L_2 to be the language over A consisting of all strings containing an equal number of 1's and -1's. This language is recognised by a machine M_2 which reads an input word w from left to right, and keeps a record at each stage of the sum of the digits so far read; w is accepted if when the machine finishes reading w this sum is equal to 0. For this machine the memory is the crucial component (or rather, the start state is the only state). The language L_2 cannot be recognised by a machine without memory. M_2 is an example of a pushdown automaton.

A range of machines and formal language families, ranging from the simplest finite state automata and associated regular (sometimes known as rational) languages to the Turing machines and recursively enumerable languages, is described in [23]; a treatment directed towards geometrical group theorists is provided by [19]. One-way nested stack automata and real-time Turing machines (associated with indexed languages and real-time languages respectively) are also of interest to us in this article, and are discussed in [1, 2] and in [29, 33]. We refer the reader to those papers for details, but below we try to give an informal overview of relevant machines and formal languages.

Figure 2 shows known inclusions between the formal language classes which we shall describe.

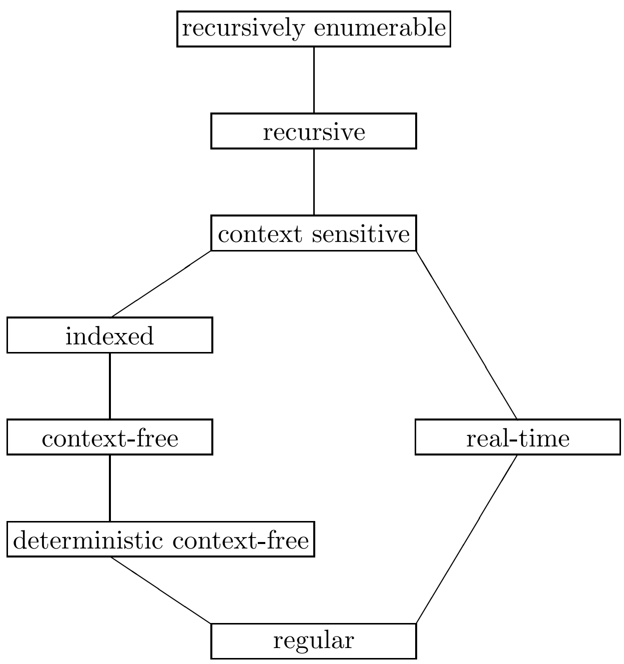


Figure 2: Inclusions between formal language classes

We continue with descriptions of various formal language classes; these might be passed over on a first reading.

4.1 Finite state automata and regular languages

A set of words over a finite alphabet is defined to be a *regular* language precisely if it is the language defined by a finite state automaton. A *finite state automaton* is a machine without memory, which moves through the states of S as it reads words over A from left to right. The simplest examples are the so-called *deterministic* finite state automata. For these a transition function $\tau\colon S \times A \to S$ determines passage between states; a word $w = a_1 \ldots a_n$ $(a_i \in A)$ is accepted if for some sequence of states $s_1, \ldots s_n$, of which s_n is an accept state, for each i, $\tau(s_{i-1}, a_i) = s_i$. Such a machine is probably best understood when viewed as a finite, directed, edge-labelled graph (possibly with loops and multiple edges), of which the states are vertices. The transition $\tau(s, a) = s'$ is then represented by an edge labelled by a from the vertex s to the vertex s'. At most one edge with any particular label leads from any given vertex (but since dead-end non-accept states can easily be ignored, there may be less that $|A|$ edges out of a vertex, and further, several edges with distinct labels might connect the same pair of vertices). A word w is accepted if it labels a path through the graph from the start vertex/state s_0 to a vertex which is marked as an accept state. Figure 3 gives such a graphical description for the machine M_1 described at the beginning of section 4. In such a figure, it is customary to ring the vertices which represent accept states, and to point at the start state with a free arrow, hence the state *even* is recognisable in this figure as the start state and sole accept state.

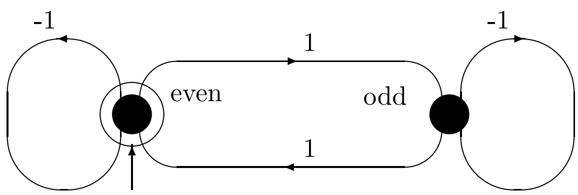


Figure 3: The finite state automaton M_1

A *non-deterministic* finite state automaton is defined in the same way as a deterministic finite state automaton except that the transition function τ is allowed to be multivalued. A word w is accepted if some (but not necessarily all) sequence of transitions following the symbols of w leads to an accept state. The graphical representation of a non-deterministic machine may have any finite number of edges with a given label from each vertex. In addition, further edges labelled by a special symbol ϵ may allow the machine to leap, without reading from the input string, from one state to another, in a so-called ϵ–move.

Given any finite state automaton, possibly with multiple edges from a vertex with the same label, possible with ϵ–edges, a finite state automaton defining the same language can be constructed in which neither of these possibilities occur. Hence, at the level of finite state automata, there is no distinction between the deterministic and non-deterministic models. However, for other classes of machines (such as for pushdown automata, described below) non-determinism increases the power of a machine.

4.2 Turing machines and recursively enumerable languages

The *Turing machines*, associated with the *recursively enumerable* languages, lie at the other end of the computational spectrum from finite state automata, and are accepted as providing a formal definition of computability. In one of the simplest models (there are many equivalent models) of a Turing machine, we consider the input word to be written on a section of a doubly-infinite tape, which is read through a movable *tape-head*. The tape also serves as a memory device. Initially the tape contains only the input word w, the tape-head points at the left hand symbol of that word, and the machine is in the start state s_0. Subsequently, the tape-head may move both right and left along the tape (which remains stationary). At any stage, the tape-head either reads the symbol from the section of tape at which it currently points or observes that no symbol is written there. Depending on the state it is currently in, and what it observes on the tape, the machine changes state, writes a new symbol (possibly from A, but possibly one of finitely many other symbols, or blank) onto the tape, and either halts, or moves its tape-head right or left one position. The input word w is accepted if the machine eventually halts in an accept state; it is possible that the machine may not halt on all input.

Non-deterministic models, where the machine may have a choice of moves in some situations (and accepts a word if some allowable sequence of moves from the obvious initial situation leads it to halt in an accept state), and models with any finite number of extra tapes and tape-heads, are all seen to be equivalent to the above description, in the sense that they also define the recursively enumerable languages.

4.3 Halting Turing machine and recursive languages

A *halting Turing machine* is a Turing machine which halts on all input; thus both the language of the machine and its complement are recursively enumerable. A language accepted by such a machine is defined to be a *recursive language.*

4.4 Linear bounded automaton and context sensitive languages

A *linear bounded automaton* is a non-deterministic Turing machine whose tape-head is only allowed to move through the piece of tape which initally contains the input word; special symbols, which cannot be overwritten, mark the two ends of the tape. Equivalently (and hence the name), the machine is restricted to a piece of tape whose length is a linear function of the length of the input word. A language accepted by such a machine is defined to be a *context sensitive language.*

4.5 Real-time Turing machines and real-time languages

A *real-time Turing machine* is most easily described as a deterministic Turing machine with any finite number of doubly-infinite tapes (one of which initially contains the input, and the others of which are initially empty), which halts as it finishes reading its input. Hence such a machine processes its input in 'real time'.

A 'move' for this machine consists of an operation of each of the tape heads, together with a state change, as follows. On the input tape, the tape-head reads the symbol to which it currently points, and then moves one place to the right. On any other tape, the tape-head reads the symbol (if any) to which it currently points, prints a new symbol (or nothing), and then either moves right, or left, or stays still. The machine changes to a new state, which depends on its current state, and the symbols read from the tapes. When the tape-head on the input head has read the last symbol of the input, the whole machine halts, and the input word is accepted if the machine is in an accept state.

A language accepted by such a machine is defined to be a *real-time language.* $\{a^n b^n c^n : n \in \mathbf{N}\}$ is an example [33]. Examples are descibed in [33] both of real-times languages which do not lie in the class of context-free languages (described below), and of (even deterministic) context-free languages which are not real-time.

4.6 Pushdown automata and context-free languages

A *pushdown automaton* can be described as a Turing machine with a particularly restricted operation on its tape, but it is probably easier to visualise as a machine formed by adding an infinite stack (commonly viewed as a spring-loaded pile of plates in a canteen) to a (possibly non-deterministic) finite state automaton. Initially the stack contains a single start symbol. Only the top symbol of the stack can be accessed at any time, and information can only be appended to the top of the stack. The input word w is read from left to right. During each move, the top symbol of the stack is removed from the stack, and a symbol from w may be read, or may not. Based on the symbols read, and the current state of the machine, the machine moves into a new state, and a string of symbols (possibly empty) from a finite alphabet is appended to the top of the stack. The word w is accepted if after reading it the machine may be in an accept state. The language accepted by a pushdown automaton is defined to be a *context-free language*.

The machine M_2 described towards the beginning of this section can be seen to be a pushdown automaton as follows. The 'sum so far' is held in memory as either a sequence of $+1$'s or as a sequence of -1's with the appropropriate sum. When the top symbol on the stack is $+1$ and a -1 is read from the input tape, the top stack symbol is removed, and nothing is added to the stack. When the top symbol on the stack is -1 and a $+1$ is read from the input tape, the top stack symbol is removed, and nothing is added to the stack. Otherwise, the top stack symbol is replaced, and then the input symbol is added to the stack. Hence the language L_2 recognised by M_2 is seen to be context-free. Similarly so is the language $\{a^n b^n : n \in N\}$ over the alphabet $\{a, b\}$. Neither language is regular. For symbols a, b, c, the language $\{a^n b^n c^n : n \in N\}$ is not context-free.

A pushdown automaton is deterministic if each input word w defines a unique sequence of moves through the machine. This does not in fact mean that a symbol of w must be read on each move, but rather that the decision to read a symbol from w at any stage is determined by the symbol read from the stack and the current state of the machine. The class of deterministic context-free languages forms a proper subclass of the class of context-free languages, which contains both the examples of context-free languages given above. The language consisting of all words of the form ww^R over some alphabet A (where w^R is the reverse of w) is non-deterministic context-free [23], but is not deterministic context-free.

4.7 One-way nested stack automata and indexed languages

A *one-way nested stack automaton* is probably most easily viewed as a generalisation of a pushdown automaton, that is, as a non-deterministic finite state automaton with an attached nest of stacks, rather than a single stack. The input word is read from left to right (as implied by the term 'one-way'). In contrast to a pushdown automaton, the read/write tape-head of this machine is allowed some movement through the system of stacks. At any point of any stack to which the tape-head has access it can read, and a new nested stack can be created; while at the top of any stack it can also write, and delete. The tape-head can move down through any stack, but its upward movement is restricted; basically it is not allowed to move upwards out of a non-empty stack.

The language accepted by a one-way nested stack automaton is defined to be an *indexed language*. For symbols a, b, c, the languages $\{a^n b^n c^n : n \in \mathbf{N}\}$, $\{a^{n^2} : n \geq 1\}, \{a^{2^n} : n \geq 1\}$ and $\{a^n b^{n^2} : n \geq 1\}$ are indexed [23], but $\{a^{n!} : n \geq 1\}$ is not [22], nor is $\{(ab^n)^n : n \geq 1\}$ [20, 22].

5 From one $\mathcal{F}$–combing to another

Many of the closure properties of the family of automatic groups also hold for other classes of combable groups, often for synchronous as well as asynchronous combings.

In the list below we assume that $\mathcal{F}$ is either the set of all languages over a finite alphabet, or is one of the classes of formal languages described in section 4, that is that $\mathcal{F}$ is one of the regular languages, context-free languages, indexed languages, context-sensitive languages, real-time languages, recursive languages, or recursively enumerable languages. (These results for all but real-time languages are proved in [9] and [31], and for real-time languages in [21].) Then just as for automatic groups, we have all the following results:

- If G has a synchronous or asynchronous $\mathcal{F}$–combing then it has such a combing over any generating set.
- Where N is a finite, normal subgroup of G, and G is finitely generated, then G is synchronously or asynchronously $\mathcal{F}$–combable if and only the same is true of G/N.
- Where J is a finite index subgroup of G, then G is synchronously or asynchronously $\mathcal{F}$–combable if and only if the same is true of J.
- If G and H are both asynchronously $\mathcal{F}$–combable then so are both $G \times H$ and $G * H$.

A crucial step in the construction of combings for 3–manifold groups in [9] is a construction of Bridson in [8]; combings for N and H can be put together to give an asynchronous combing for a split-extension of the form $N \rtimes H$ provided that N has a combing which is particularly stable under the action of H. The set of all geodesics in a word hyperbolic group has that stability, and is a regular language; hence, for any of the language classes $\mathcal{F}$ considered in this section, any split extension of a word hyperbolic group by an $\mathcal{F}$–combable group is $\mathcal{F}$–combable. The free abelian group $\mathbf{Z}^n$ also possesses a combing with the necessary stability; hence all split extensions of $\mathbf{Z}^n$ by combable groups are asynchronously combable. It remains only to ask in which language class these combings lie.

Stable combings for $\mathbf{Z}^n$ are constructed by Bridson in [8] as follows. $\mathbf{Z}^n$ is seen embedded as a lattice in $\mathbf{R}^n$, and the group element g is then represented by a word which, as a path through the lattice, lies closest to the real line joining the point 0 to the point representing g. For some group elements there is a selection of such paths; a systematic choice can clearly be made. It was proved in [9] that $\mathbf{Z}^2$ has a combing of this type which is an indexed language; hence all split extensions of the form $\mathbf{Z}^2 \rtimes \mathbf{Z}$ were seen to be indexed combable. It followed from this that the fundamental groups of all compact, geometrisable 3–manifolds were indexed combable; for these are all commensurable with free products of groups which are either automatic or finite extensions of $\mathbf{Z}^2 \rtimes \mathbf{Z}$.

It is unclear whether or not the corresponding combing for $\mathbf{Z}^n$ is also an indexed language when $n > 2$. Certainly it is a real-time language [21]. Hence many split extensions of the form $\mathbf{Z}^n \rtimes H$ are seen to have asynchronous combings which are real-time languages. We give some examples in the final section.

6 Combing up the language hierarchy

6.1 Regular languages

A group with a synchronous regular combing is, by definition, automatic. More generally, a group with a regular combing is called *asynchronously automatic* [15]. It is proved in [15] that the asynchronicity of an asynchronously automatic group is bounded; that is the relative speed at which particles must move along two fellow-travelling words in order to keep apace can be kept within bounds. The Baumslag–Solitar groups

$$G_{p,q} = \langle a, b \mid ba^p = a^q b \rangle$$

are asynchronously automatic, but not automatic, for $p \neq \pm q$ (see [15, 30]), and automatic for $p = \pm q$.

It is proved in [15] that a nilpotent group which is not abelian-by-finite cannot be asynchronously automatic. From this it follows that the fundamental groups of compact manifolds based on the Nil geometry cannot be asynchronously automatic; N. Brady proved that the same is true of groups of the compact manifolds based on the Sol geometry [6].

6.2 Context-free languages

No examples are currently known of non-automatic groups with context-free combings. It is proved in [9] that a nilpotent group which is not abelian-by-finite cannot have a bijective context-free combing; however it remains open whether a context-free combing with more that one representative for some group elements might be possible.

6.3 Indexed languages

Bridson and Gilman proved that the fundamental group of every compact geometrisable 3–manifold (or orbifold) is indexed combable. By the results of [6, 15, 9] described above for regular and context-free combings, this result must be close to being best possible.

It follows immediately from Bridson and Gilman's results that a split extension of $\mathbf{Z}^2$ by an indexed combable (and so, certainly by an automatic) group is again indexed combable.

6.4 Real-time languages

Since the stable combing of $\mathbf{R}^n$ described in section 5 is a real-time language [21], it follows that any split extension over $\mathbf{Z}^n$ of a real-time combable group is real-time combable. Hence (see [21]), any finitely generated class 2 nilpotent group with cyclic commutator subgroup is real-time combable, and also any 3–generated class 2 nilpotent group. Further the free class 2 nilpotent groups, with presentation,

$$\langle x_1, \ldots x_k \mid [[x_i, x_j], x_k], \forall i, j, k \rangle,$$

as well as the n–dimensional Heisenberg groups and the groups of n–dimensional, unipotent upper-triangular integer matrices, can all be expressed as split extensions over free abelian groups, and hence are real-time combable. It follows that any polycyclic-by-finite group (and so, in particular, any finitely generated nilpotent group) embeds as a subgroup in a real-time combable group.

Torsion-free polycyclic metabelian groups with centre disjoint from their commutator subgroup are far from being nilpotent, but are also real-time combable (see [21]). Such groups split over their commutator subgroup, by a theorem of [32]. An example is provided by the group

$$\langle x, y, z \mid yz = zy, y^x = yz, z^x = y^2 z \rangle$$

which is certainly not automatic (it has exponential isoperimetric inequality). In fact this group is also indexed combable, since it is of the form $\mathbf{Z}^2 \rtimes \mathbf{Z}$.

References

[1] **Alfred V Aho**, *Indexed grammars – an extension of context-free grammars*, J. Assoc. Comp. Mach. 15 (1968) 647–671

[2] **Alfred V Aho**, *Nested stack automata*, J. Assoc. Comp. Mach. 16 (1969) 383–406

[3] **G Baumslag**, **S M Gersten**, **M Shapiro**, **H Short**, *Automatic groups and amalgams*, Journal of Pure and Applied Algebra 76 (1991) 229–316

[4] **G Baumslag**, **D Gildenhuys**, **R Strebel**, *Algorithmically insoluble problems about finitely presented solvable groups, Lie and associative algebras. I*, Journal of Pure and Applied Algebra 39 (1986) 53–94

[5] **Gilbert Baumslag**, **Michael Shapiro**, **Hamish Short**, *Parallel poly pushdown groups*, Jorunal of Pure and Applied Algebra, to appear

[6] **Noel Brady**, *The geometry of asynchronous automatic structures on groups*, PhD thesis, UC Berkeley (1993)

[7] **Martin R Bridson**, *On the geometry of normal forms in discrete groups*, Proc. London Math. Soc. (3) 67 (1993) 596–616

[8] **Martin R Bridson**, *Combings of semidirect products and 3–manifold groups*, Geometric and Functional Analysis 3 (1993) 263–278

[9] **M R Bridson**, **R H Gilman**, *Formal language theory and the geometry of 3–manifolds*, Commentarii Math. Helv. 71 (1996) 525–555

[10] **Brigitte Brink**, **Robert Howlett**, *A finiteness property of Coxeter groups*, Math. Ann. 296 (1993) 179–190

[11] **José Burillo**, *Lower bounds of isoperimetric functions for nilpotent groups*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. (25) 1–8

[12] **J W Cannon**, *The combinatorial structure of cocompact discrete hyperbolic groups*, Geom. Dedicata 16 (1984) 123–148

[13] **Ruth Charney**, *Artin groups of finite type are biautomatic*, Math. Ann. 292 (1992) 671–683

[14] **Ruth Charney**, *Geodesic automation and growth functions of Artin groups*, Math. Ann. 301 (1995) 307–324

[15] **David B A Epstein**, **J W Cannon**, **D F Holt**, **S Levy**, **M S Patterson**, **W Thurston**, *Word processing in groups*, Jones and Bartlett, (1992)

[16] **Benson Farb**, *Automatic groups: a guided tour*, Enseign. Math. (2) 38 (1992) 291–313

[17] **S M Gersten**, *Bounded cocycles and combings of groups*, Internat. J. Algebra Comput. 2 (1992) 307–326

[18] **S M Gersten**, **H B Short**, *Small cancellation theory and automatic groups*, Inv. Math. 102 (1990) 305–334

[19] **Robert H Gilman**, *Formal languages and infinite groups*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 25 AMS (1996) 27–51

[20] **Robert H Gilman**, *A shrinking lemma for indexed languages*, Theoret. Comput. Sci. 163 (1996) 277–281

[21] **Robert H Gilman**, **Derek F Holt**, **Sarah Rees**, *Combing nilpotent and polycylic groups*, Internat. J. Algebra Comput. to appear

[22] **T Hayashi**, *On derivation trees of indexed grammars*, Publ. RIMS Kyoto Univ. 9 (1973) 61–92

[23] **John E Hopcroft**, **Jeffrey D Ullman**, *Introduction to automata theory, languages and computation*, Addison–Wesley, (1979)

[24] **Arye Juhász**, *Large Artin groups are biautomatic*, preprint

[25] **O G Kharlampovich**, *A finitely presented group with unsolvable word problem (in Russian)*, Izv. Akad. Nauk SSSR Ser. Mat. 45 (1981) 852–873

[26] **U Lang**, *Quasigeodesics outside horoballs*, Geom. Dedicata 63 (1996) 205–215

[27] **Lee Mosher**, *Mapping class groups are automatic*, Annals of Math. 142 (1995) 303–384

[28] **David Peifer**, *Artin groups of extra-large type are automatic*, J. Pure Appl. Alg. 110 (1996) 15–56

[29] **Michael O Rabin**, *Real time computation*, Israel J. Math. 1 (1963) 203–211

[30] **Sarah Rees**, *Automatic groups associated with word orders other than shortlex*, Int. J. Alg. Comp. (to appear)

[31] **Sarah Rees**, *A language theoretic analysis of combings*, preprint

[32] **Derek J S Robinson**, *Splitting theorems for infinite groups*, Symposia Mathematica 17, Convegni del Novembre e del Dicembre 1973, Academic Press, (1976) 441–470

[33] **A Rosenberg**, *Real-Time Definable Languages*, J. Assoc. Comput. Mach. 14 (1967) 645-662

University of Newcastle, Newcastle NE1 7RU, UK

Email: `Sarah.Rees@ncl.ac.uk`

Received: 15 November 1997

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 511–549

Shapes of polyhedra and triangulations of the sphere

William P Thurston

Abstract The space of shapes of a polyhedron with given total angles less than 2π at each of its n vertices has a Kähler metric, locally isometric to complex hyperbolic space $\mathbb{CH}^{n-3}$. The metric is not complete: collisions between vertices take place a finite distance from a nonsingular point. The metric completion is a complex hyperbolic cone-manifold. In some interesting special cases, the metric completion is an orbifold. The concrete description of these spaces of shapes gives information about the combinatorial classification of triangulations of the sphere with no more than 6 triangles at a vertex.

AMS Classification 51M20; 51F15, 20H15, 57M50

Keywords Polyhedra, triangulations, configuration spaces, braid groups, complex hyperbolic orbifolds

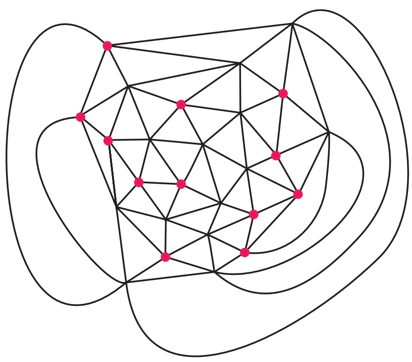


Figure 1: The twelve marked vertices of this triangulation of S^2 have five triangles while all other vertices have six. Theorem 0.1 implies that the possible triangulations satisfying this condition are parametrized, up to isomorphism, by 20–tuples of integers up to the action of a group of integer linear transformations.

Introduction

There are only three completely symmetric triangulations of the sphere: the tetrahedron, the octahedron and the icosahedron. However, finer triangulations with good geometric properties are often encountered or desired for mathematical, scientific or technological reasons, for example, the kinds of triangulations popularized in modern times by Buckminster Fuller and used for geodesic domes and chemical 'Buckyballs'.

There are procedures to refine and modify any triangulation of a surface until every vertex has either 5, 6 or 7 triangles around it, or with more effort, so that there are only 5 or 6 triangles if the surface has positive Euler characteristic, only 6 triangles if the surface has zero Euler characteristic, or only 6 or 7 triangles if the surface has negative Euler characteristic. These conditions on triangulations are combinatorial analogues of metrics of positive, zero or negative curvature. How systematically can they be understood?

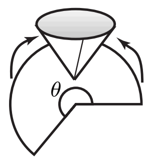
In this paper, we will develop a global theory to describe all triangulations of the S^2 such that each vertex has 6 or fewer triangles at any vertex. Here is one description:

Theorem 0.1 (Polyhedra are lattice points) *There is a lattice L in complex Lorentz space $C^{(1,9)}$ and a group Γ of automorphisms, such that triangulations of non-negative combinatorial curvature are elements of L_+/Γ, where L_+ is the set of lattice points of positive square-norm. The projective action of Γ on complex hyperbolic space $\mathbb{CH}^9$ (the unit ball in $\mathbb{C}^9 \subset \mathbb{CP}^9$) has quotient of finite volume. The square of the norm of a lattice point is the number of triangles in the triangulation.*

A triangulation is *non-negatively curved* if there are never more than six triangles at a vertex. The theorem can be interpreted as describing certain concrete cut-and-glue constructions for creating triangulations of non-negative curvature, starting from simple and easily-classified examples. The constructions are parametrized by choices of integers, subject to certain geometric constraints. The fact that Γ is a discrete group means that it is possible to dispense with most of the constraints, except for an algebraic condition that a certain quadratic form is positive: any choice of integer parameters can be transformed by Γ to satisfy the geometric conditions, and the resulting triangulation is unique. Thus, the collection of all triangulations can be described either as a quotient space, in which identifications of the parameters are made algebraically, or as a fundamental domain (see section 7).

We will study combinatorial types of triangulations by using a metric where each triangle is a Euclidean equilateral triangle with sides of unit length. This metric is locally Euclidean everywhere except near vertices that have fewer than 6 triangles.

It is helpful to consider these metrics as a special case of metrics on the sphere which are locally Euclidean except at a finite number of points, which have neighborhoods locally modelled on cones. A *cone of cone-angle* θ is a metric space that can be formed, if $\theta \le 2\pi$, from a sector of the Euclidean plane between two rays that make an angle θ, by gluing the two rays together. More generally, a cone of angle θ can be formed by taking the universal cover of the plane minus 0, reinserting 0, and then identifying modulo a transformation that “rotates” by angle θ. The *apex curvature* of a cone of cone-angle θ is $2\pi - \theta$.



Cone angle θ

A Euclidean cone metric on a surface satisfies the Gauss–Bonnet theorem, that is, the sum of the apex curvatures is 2π times the Euler characteristic. This fact can be derived from basic Euclidean geometry by subdividing the surface into triangles and looking at the sum of angles of all triangles grouped in two different ways, by triangle or by vertex. It can also be derived from the usual smooth Gauss–Bonnet formula by rounding off the points, replacing a tiny neighborhood of each cone point by a smooth surface (for example part of a small sphere).

Theorem 0.2 (Cone metrics form cone manifold) *Let* $k_1, k_2, \dots, k_n$ $[n > 3]$ *be a collection of real numbers in the interval* $(0, 2\pi)$ *whose sum is* 4π. *Then the set of Euclidean cone metrics on the sphere with cone points of curvature* k_i *and of total area* 1 *forms a complex hyperbolic manifold, whose metric completion is a complex hyperbolic cone manifold of finite volume. This cone manifold is an orbifold (that is, the quotient space of a discrete group) if and only if for any pair* k_i, k_j *whose sum* $s = k_i + k_j$ *that is less than* 2π, *either*

(i) $(2\pi - s)$ *divides* 2π, *or*

(ii) $k_i = k_j$, *and* $(2\pi - s)/2 = (\pi - k_i)$ *divides* 2π.

The definition of “cone-manifold” in dimensions bigger than 2 will be given later.

This turns out to be closely related to work of Picard ([6], [7]) and Mostow and Deligne ([2], [3], [5]). Picard discovered many of the orbifolds; his student LeVavasseur enumerated the class of groups Picard discovered, and they were further analyzed by Deligne and Mostow. Mostow discovered that condition (i) is not always required to obtain an orbifold and that (ii) is sufficient when $k_i = k_j$. However, the geometric interpretations were not apparent in these papers. It is possible to understand the quotient cone-manifolds quite concretely in terms of shapes of polyhedra.

A version of this paper has circulated for a number of years as a preprint, which for a time was circulated as a Geometry Center preprint, and later revised as part of the xxx mathematics archive. In view of this history, some time warp is inevitable: for some parts of this paper, others may have have done further work that is not here taken into account. I would like to thank Derek Holt, Igor Rivin, Chih-Han Sah and Rich Schwarz for mathematical comments and corrections that I hope I have taken into account.

1 Triangulations of a hexagon

Let E be the standard equilateral triangulation of $\mathbb{C}$ by triangles of unit side length, where 0 and 1 are both vertices. The set **Eis** of vertices of E are complex numbers of the form $m + p\omega$, where $\omega = 1/2 + \sqrt{-3}/2$ is a primitive 6th root of unity. These lattice points form a subring of $\mathbb{C}$, called the Eisenstein integers, the ring of algebraic integers in the quadratic imaginary field $\mathbb{Q}(\sqrt{-3})$.

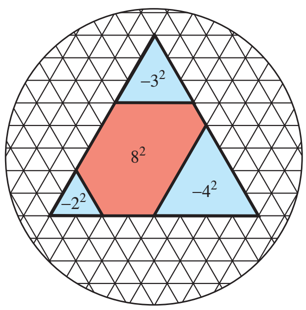


Figure 2: An Eisenstein lattice hexagon has the form of a large equilateral triangle of sidelength n, minus three equilateral triangles that fit inside it of sidelengths p_1, p_2 and p_3. An equilateral triangle of sidelength n contains n^2 unit equilateral triangles, so the hexagon has $n^2 - p_1^2 - p_2^2 - p_3^2$ triangles.

To warm up, we'll analyze all possible shapes of Eisenstein lattice hexagons, with vertices in **Eis** and sides parallel (in order) with the sides of a standard hexagon. Note that any such hexagon with m triangles determines a non-negatively curved triangulation of the sphere with $2m$ triangles, formed by making a hexagonal envelope from two copies of the hexagon glued along the boundary.

If we circumscribe a lattice triangle T about our lattice hexagon H, this gives a description

$$H = T \setminus (S_1 \cup S_2 \cup S_3),$$

where the S_i are smaller equilateral triangles. If T has sidelength n and S_i has sidelength p_i, then H contains

$$m = n^2 - p_1^2 - p_2^2 - p_3^2 \tag{1}$$

triangles.

All such hexagons are described by four integer parameters, subject to the 6 inequalities

$$p_1 \geq 0 \quad p_2 \geq 0 \quad p_3 \geq 0$$
$$p_1 + p_2 \leq n \quad p_2 + p_3 \leq n \quad p_3 + p_1 \leq n,$$

where strict inequalities give non-degenerate hexagons; if one or more inequality becomes an equality then one or more sides of the hexagon shrinks to length 0 and the 'hexagon' becomes a pentagon, quadrilateral or triangle.

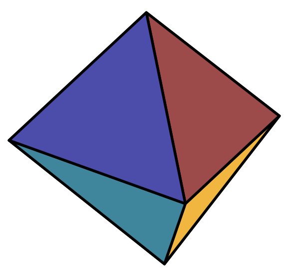


Figure 3: The space of shapes of hexagons is described by this polyhedron in hyperbolic 3–space; the faces represent hexagons degenerated to pentagons, and the edges represent degeneration to quadrilaterals. All dihedral angles are $\pi/2$. The three mid-level vertices are ideal vertices at infinity, and represent the three ways that hexagons can become arbitrarily long and skinny, while the top and bottom are finite vertices, representing the two ways that hexagons can degenerate to equilateral triangles. The polyhedron has hyperbolic volume .91596559417 . . . .

The solutions are elements of the integer lattice inside a convex cone $C \subset \mathbb{R}^4$. This description can be extended to non-integer parameters, which then determine a size and shape for the hexagon, but not a triangulation. Equation (1) expresses the area, measured in triangles, as a quadratic form of signature $(1,3)$. The isometry group of any such a form is $C_2 \times \mathrm{Isom}(\mathbb{H}^3)$ (where C_2 denotes the cyclic group of order 2).

The possible shapes of lattice hexagons (where rescaling is allowed) are parametrized by a convex polyhedron $H \subset \mathbb{H}^3$ which is the projective image of the convex cone $C \subset \mathbb{R}^{(1,3)}$. This polyhedron has three ideal vertices at infinity, which represent the three directions in which shapes of hexagons can tend toward infinity, by becoming long and skinny along one of three axes. In addition, there are two finite vertices (top and bottom), representing the two ways that a hexagon can degenerate to an equilateral triangle. All dihedral angles of this hyperbolic polyhedron are $\pi/2$. Four edges meet at each ideal vertex, while three edges meet at the finite vertices. Triangulations with m triangles are represented by a discrete set $H_m \subset H$. Figure 4 plots the count of how many of these lattice hexagons there are with each possible area up to $1,000$. One indication of the relevance of hyperbolic geometry is that the average number

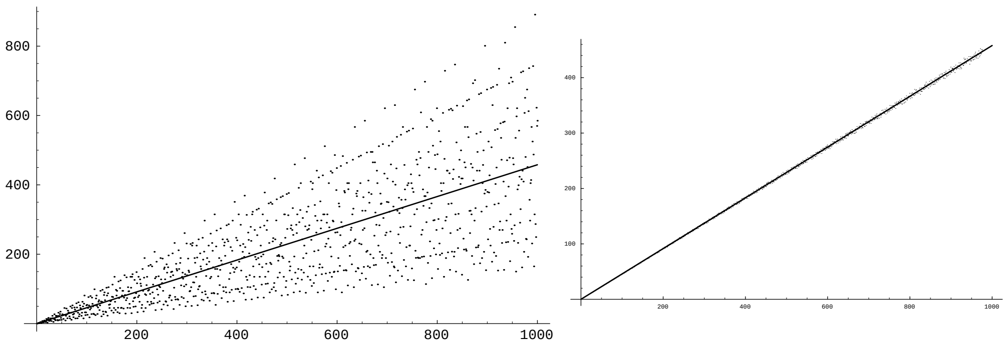


Figure 4: *Left* the weighted count of Eisenstein lattice hexagons containing up to 1000 triangles, using orbifold weights $1/2^k$ where k is the number of sides of a hexagon of length 0. The parameter space of shapes (figure 3) has hyperbolic volume $.91596559417\ldots$ (1/4 that of the Whitehead Link complement), so the number of hexagons containing m triangles should grow on the average as the volume of the intersection of $C/2$ with the shell in $\mathbb{E}^{(1,3)}$ between radius $\sqrt{m}$ and $\sqrt{m+1}$, $.45798279709 \cdots * m$, as indicated. *Right* The same data averaged over windows of size 49.

of hexagons of a given area is well estimated by the hyperbolic volume of the parameter space.

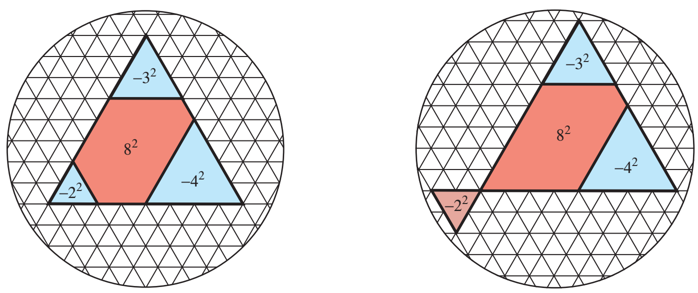


Figure 5: A butterfly operation moves one edge of a hexagon (*left*) across a butterfly-shaped quadrilateral of 0 area, yielding a new hexagon (*right*) of the same area. The set of butterfly moves generate a discrete group of isometries of $\mathbb{H}^3$, generated by reflections in the faces of the polyhedron H.

It's interesting to note that H is the fundamental polyhedron for a discrete group of isometries of $\mathbb{H}^3$, since all dihedral angles equal $\pi/2$. This group can be interpreted in terms of not necessarily simple hexagons in the Eisenstein lattice whose sides are parallel, in order, to those of the standard hexagon. A non-simple lattice hexagon wraps with integer degree around each triangle in the plane; its total area, using these integer weights, is given by the same quadratic form $n^2 - \sum p_i^2$.

Reflection in a face of the polyhedron corresponds to a 'butterfly move', which is described numerically by reversing the sign of the length of one of the edges of the hexagon, and adjusting the two neighboring lengths so that the result is a closed curve. Geometrically, the hexagon moves across a quadrilateral reminiscent of a butterfly, resulting in a new hexagon that algebraically encloses the same area as the original. Note that this operation fixes any hexagon where the given side has degenerated to have length 0—this is one of the faces of the polyhedron H. The operations for two sides of the hexagon that do not meet commute with each other, and fix any shapes of hexagons where both these sides have length 0. These shapes describe an edge of H, and since the reflections in adjacent faces commute, the angle must be $\pi/2$. Two adjacent sides of the hexagon cannot both have 0 length at once, so the 9 non-adjacent pairs of sides of the hexagon correspond 1–1 to the 9 edges of H.

Any solution to the equation $0 < m = n^2 - p_1^2 - p_2^2 - p_3^2$ determines a not necessarily simple hexagon of area m, which projects to a point in $\mathbb{H}^3$. By a sequence of butterfly moves, this point can be transformed to be inside the fundamental domain H. The resulting point inside H is uniquely determined by the initial solution and does not depend on what sequence of butterfly moves were used to get it there, since H is the quotient space (quotient orbifold) for the group action as well as being its fundamental domain.

2 Triangulations of the sphere

Let $P(n; k_1, k_2, \ldots, k_s)$ denote the set of isomorphism classes of "triangulations" of the sphere having exactly $2n$ triangles, where for each i there is one vertex incident to $6-k_i$ triangles, and all remaining vertices are incident to 6 triangles. This paper will be limited to the non-negatively curved cases that $0 < k_i \leq 5$. For there to be any actual triangulations we must have $\sum_i k_i = 12$. We will use the term "triangulation" throughout to refer to a space obtained by gluing together triangles by a pairing of their edges; thus, in the case $k_i = 5$, two edges of a triangle are folded together to form a vertex incident to a single triangle. Every triangulation of the sphere has an even number of triangles.

If $T \in P(n; k_1, \ldots, k_s)$, then there is a *developing map* D_T from the universal cover $\tilde{T}$ of T minus its singular vertices into E. Choose any triangle of $\tilde{T}$, and map it to the triangle $\Delta(0, 1, \omega)$. The developing map D_T is now determined by a form of analytic continuation, so that it is a local isometry, mapping triangles to triangles.

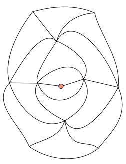
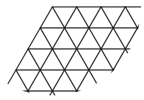
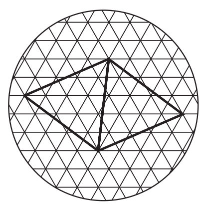


Figure 6: A vertex with 2 triangles (*left*) comes by folding up a 120° angle (*right*). Equivalently, it has a neighborhood whose developing map rolls around a vertex in the plane one time for every three revolutions of the cone. Similar phenomena occur for vertices with 3 triangles or 1 triangle.

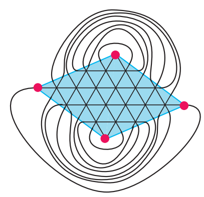


A particularly nice phenomenon happens for any vertices that have only 1, 2, or 3 triangles. Consider a component N_v of the inverse image in $\tilde{T}$ of a small neighborhood of any such vertex v. It develops into the vicinity of some vertex w in $\mathbf{Eis}$. In these cases, the number of triangles around v is a divisor of 6, so the developing map repeats itself when it first wraps around the vertex w, along a path in $\tilde{T}$ which maps to a curve in T wrapping respectively 6, 3, or 2 times around the v. Therefore, the developing map is defined from a smaller covering of T minus its singular vertices, which can be obtained as a certain quotient space $S(T)$ of $\tilde{T}$. In $S(T)$, each component of the preimage of a small neighborhood of v only intersects six triangles. In fact, $S(T)$ is isomorphic to E. Therefore T is a quotient space of a discrete group $\Gamma(T)$ acting on E such that only elements of $\mathbf{Eis}$ are fixed points of elements of $\Gamma(T)$.



The examples where every vertex has 1, 2, 3 or 6 triangles are $P(n;4,4,4)$, $P(n;3,4,5)$ and $P(n;3,3,3,3)$. For $P(n;4,4,4)$ or $P(n;3,4,5)$, the group $\Gamma(T)$ is a triangle group. A fundamental domain can be chosen as the union of two equilateral triangles in the first case and $30°, 60°, 90°$ triangles of opposite orientation in the second. We may arrange that one of the vertices is at the origin.

Let α be a singular vertex closest to the origin.



In the case $T \in P(n;4,4,4)$, the other singular vertices are $\mathbf{Eis}*\alpha$. Clearly this set determines the group, and any $\alpha \neq 0$ will work. The value of n is the ratio $\alpha\bar{\alpha}$ of a fundamental parallelogram $0, \alpha, \alpha(1+\omega), \alpha\omega$ to the area of a primitive lattice parallelogram $0, 1, 1+\omega, \omega$. The possible numbers of triangles are numbers expressible in the form $n = a^2 + 3b^2$.

There is some ambiguity in this description: if we replace α by any of the other 5 numbers $\omega^k\alpha$, we obtain an isomorphic triangulation. Thus, triangulations of this type are in one-to-one correspondence with lattice points on the cone $\mathbb{C}/\langle\omega\rangle$, where $\langle\omega\rangle$ refers to the multiplicative subgroup of order 6 generated by ω.

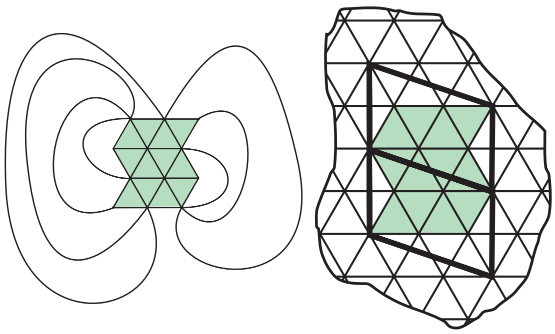


Figure 7: Developing a triangulation with 3 or 6 triangles at each vertex.

Similarly, in the case $T \in P(n;3,4,5)$, the vertices are of the form $(m + p\sqrt{-3})\alpha$, and $n = 2\alpha\bar{\alpha}$. As before, α is well-defined only up to multiplication by powers of ω. In this case, if we replace α by $\omega^k\alpha$, where k is odd, we get a different triangle group, but it has an isomorphic quotient space.

The case $P(n;3,3,3,3)$ allows somewhat more variation. For a singular vertex x in **Eis**, let $\gamma_x \in \Gamma(T)$ be the rotation of order 2 about x. Then for any two elements x and y, the product $\gamma_x\gamma_0\gamma_y$ is a 180° rotation about $x+y$. Therefore, the singular vertices form an additive subgroup of **Eis**. Any additive subgroup will work. The subgroup is determined if we specify the sides α and β of a fundamental parallelogram. If we express α and β as linear combinations of the generators 1 and ω for **Eis**, then the value of n is twice the determinant of the resulting two by two matrix. Every even number is achievable. Of course, α and β are well-defined only up to change of basis for the lattice and up to multiplication by 6th roots of unity. Note that multiplication by $\omega^3 = -1$ is also represented by a change of generators. A nice picture can be formed by considering the shape parameter $z = \beta/\alpha$. The action of the group $SL(2,\mathbb{Z})$ on the set of shape parameters is the usual action by fractional linear transformations on the upper half plane. Figure 8 illustrates the set of shapes obtainable for $n = 246$.

Let us now skip to a more complicated case, that of $P(n;2,2,2,2,2,2)$, which includes the regular octahedron. We have already encountered a special case: the hexagonal envelopes of section 1 are examples of octahedra of this sort.

Just as a hexagon can be described by removing three small triangles from a large triangle, there is a way to describe any element $T \in P(n;2,2,2,2,2,2)$ by modifying an element $\overline{T} \in P(m;3,3,3)$, for some m.

Suppose T is any triangulation of the sphere with 6 vertices incident to four triangles, and the rest incident to 6.

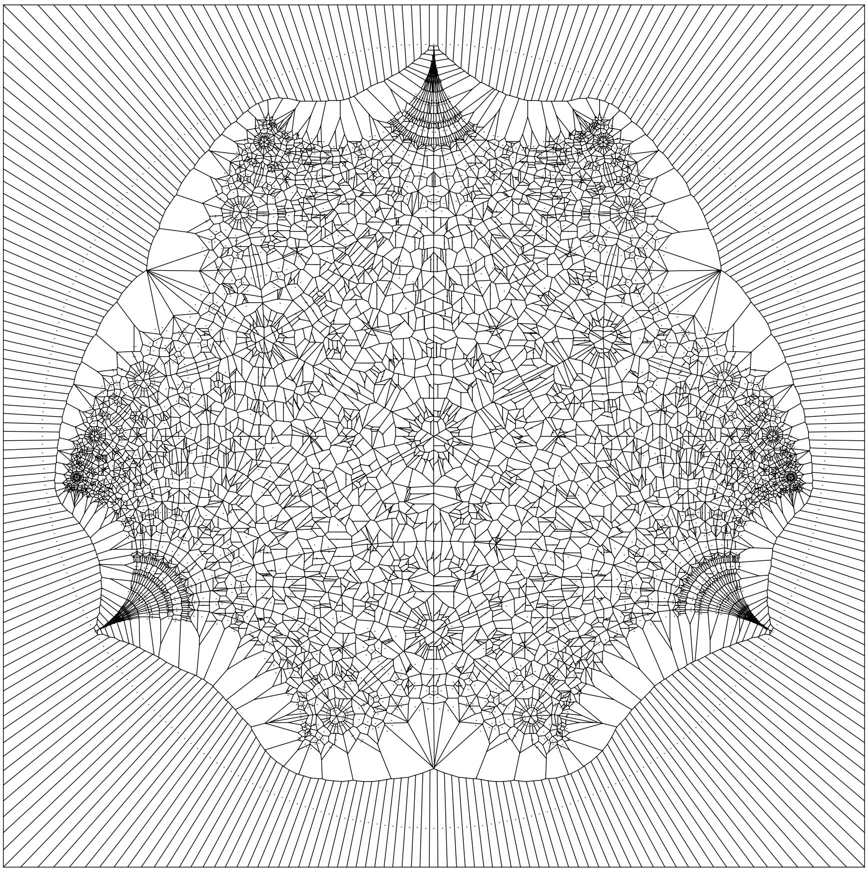


Figure 8: This is $P(246; 3, 3, 3, 3)$, plotted in the Poincaré disk model of $\mathbb{H}^2$. The elements of $P(246; 3, 3, 3, 3)$ are small dots; the Voronoi diagram for these dots is shown, with one small dot inside each region. The position of the dot in $\mathbb{H}^2$ determines the shape of a tetrahedron triangulated by 246 equilateral triangles. Two dots which differ by $\mathrm{PSL}(2, \mathbb{Z})$ represent the same shape. The shape does not always completely determine the triangulation—one also needs an angle for edges, that is, a lifting of the point to a certain line bundle over $\mathbb{H}^2$.

Consider the associated cone metric C. We claim there is at least one way to join the 6 cone points in pairs by three disjoint geodesic segments. To construct such a pairing, first observe that any pair of cone points are joined by at least one geodesic: the shortest path between them is a geodesic. Note that geodesics can never pass through cone points with positive curvature, except at their endpoints. We see that there is a collection of three not-necessarily disjoint geodesic segments joining the 6 points in pairs. Let $\{e, f, g\}$ be such a collection

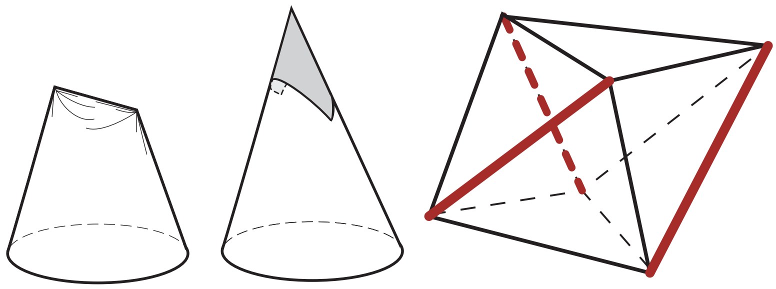


Figure 9: *Left* If a Euclidean cone manifold is cut along a geodesic arc joining the two cone points of curvature α and β, the resulting figure is isometric to a region in a Euclidean cone manifold with a new cone point whose curvature is $\alpha+\beta$ (*middle*). This gives a recursive procedure to reduce the construction of compact Euclidean cone manifolds of non-positive curvature to ones having only three cone points. *Right* An element $T \in P(n;2,2,2,2,2,2)$ can be reduced to $T' \in P(n';3,3,3)$ by slitting 3 arcs, then extending.

of shortest possible length. In particular, e, f and g are shortest paths with their given endpoints. No pair of these edges can intersect: if they did, then by cutting and pasting, one would find that the four endpoints involved could be joined in an alternate way by shorter paths.

Cut C along the three edges e, f and g, and consider the developing map for the resulting surface C'. At an endpoint of say e, the developing image subtends an angle of $120°$; a curve which wraps three times around e in a small neighborhood develops to a curve wrapping once around the outside of a regular hexagon H_e in the plane. Let C_e be H_e modulo a rotation of order 3. If we glue C_e and the similarly constructed cones C_f and C_g to the cuts, we obtain a new cone-manifold C'', with three cone points of order 3. The hexagon H_e has its vertices on lattice points of $\mathbf{Eis}$, so its center is also a lattice point of $\mathbf{Eis}$. Therefore, $C'' \in P(m;4,4,4)$ for some m. Consequently, a general element of $P(n;2,2,2,2,2,2)$ is obtained by choosing some m bigger than n, choosing an element of $P(m;4,4,4)$, and choosing three types of hexagons whose area in triangles adds to $6(m-n)$ such that when they are placed around the three classes of order 3 points in the plane, all their images are disjoint. Cut all these hexagons out of the plane, divide by the $(3,3,3)$ triangle group, and glue together the pair of edges coming from each hexagon. We can express this as a choice of four elements $\alpha_i \in \mathbf{Eis}$, such that

$$\alpha_1\bar{\alpha}_1 - \alpha_2\bar{\alpha}_2 - \alpha_3\bar{\alpha}_3 - \alpha_4\bar{\alpha}_4 :$$

α_1 is used to construct the original $(3,3,3)$ triangle group, and the other α_i's are vectors from the centers of the each of the hexagons to one of the vertices,

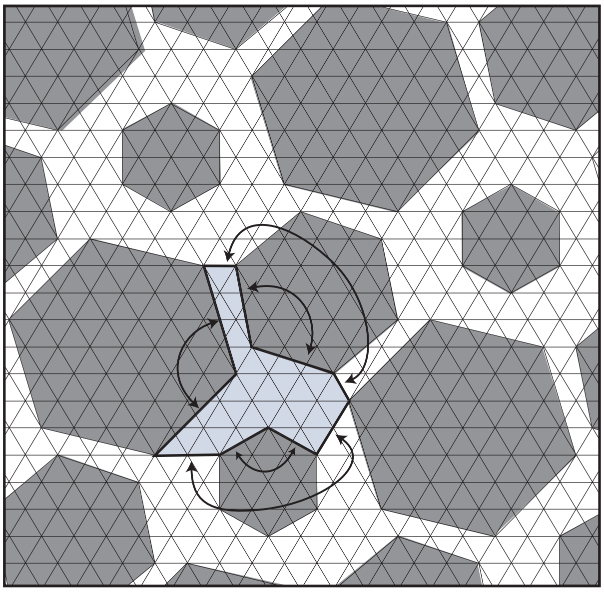


Figure 10: This is an illustration of the construction of a generalized octahedron, that is, an element of $P(n;2,2,2,2,2,2)$. First, choose a $3,3,3$ group acting in the plane with the fixed points of the elements of order 3 on lattice points of **Eis**. Then choose three families of lattice hexagons invariant by the group, centered at the fixed points of elements of order 3. Remove the hexagons, form the quotient by the group, and glue the edges of the resulting slits together. Equivalently, you can glue the boundary of a fundamental domain as illustrated.

yielding a triangulation

$$T(\alpha_1,\alpha_2,\alpha_3,\alpha_4) \in P(n;2,2,2,2,2,2).$$

The α_i's are subject to an additional geometric condition, that the hexagons they define be embedded. The coordinates are only defined up to a geometrically-defined equivalence relation, having to do with the multiplicity of choices for e, f, and g. The easy observation is that when any of the α_i are multiplied by powers of ω, we obtain the same T. These coordinates make it easy to automatically enumerate all examples, although it is somewhat harder to weed out repetitions. The geometric conditions can be nearly determined from the norms: if $|\alpha_i|+|\alpha_j| < |\alpha_1|$, for $i \neq j \in \{2,3,4\}$, then the hexagons are disjoint; if this sum is greater than $(2/\sqrt{3})|\alpha_1| = 1.1547\ldots|\alpha_1|$, then two hexagons intersect; otherwise, one needs to consider the picture. If $|\alpha_i| < |\alpha_1|/3$ for $i > 1$, then the three edges e, f and g are clearly the three shortest possible edges; in general, the question is more complicated. The standard octahedron

$O \in P(4; 2, 2, 2, 2, 2, 2)$, for example, has an infinite number of descriptions, for example $O = T(2k + 1 + (-k + 2)\omega, k + \omega, k + \omega, k + \omega)$ for every $k \geq 0$.

Another construction will be given in section 7 that can be used to search all possibilities while weeding out repetition fairly efficiently.

3 Shapes of polyhedra

Any collection of n–dimensional Euclidean polyhedra whose $(n-1)$–dimensional faces are glued together isometrically in pairs yields an example of a *cone-manifold* and gives a pretty good flavor for the singular behavior that can occur. However, polyhedra are not a suitable substrate for a definition in the context we need, since we will be working with metrics whose local geometry has no concept of polyhedra comparable to the Euclidean case: they have no totally geodesic hypersurfaces.

In general, a cone-manifold is a kind of singular Riemannian metric; in our case, we will work with spaces modelled after a complete Riemannian n–manifold X together with a group G of isometries of X, called an (X, G)–manifold. If G acts transitively, this would be called a *homogeneous space*, but G *does not necessarily act transitively.* Moreover, the group G is part of the structure. It is not necessarily the full group of isometries of X: for instance, we might have $X = \mathbb{E}^2$ and G the group of isometries that preserve the $\mathbf{Eis}$.

An (X, G)*–manifold* is a space equipped with a covering by open sets with homeomorphisms into X, such that the transition maps on the overlap of any two sets is in G.

The concept of an (X, G)–cone-manifold is defined inductively by dimension, as follows:

If X is 1–dimensional, an (X, G)–cone-manifold is just an (X, G)–manifold.

Suppose X is k–dimensional, where $k > 1$. For any point $p \in X$, let G_p be the stabilizer of p, and let X_p be the set of tangent rays through p. Then (X_p, G_p) is a model space of one lower dimension. If Y is any (X_p, G_p)–cone-manifold, there is associated to it a fairly intuitive construction, the *radius* r *cone* of Y, $C_r(Y)$ for any $r > 0$ such that the exponential map at p is an embedding on the ball of radius r in $T_p(X)$, constructed from the geodesic rays from p in X assembled in the same way that Y is. That is, for each subset of X_p, there is associated a cone in the tangent space at p, and to this is associated (via the exponential map) its radius r cone in X. These are glued together, using local coordinates in Y, to form $C_r(Y)$.

An (X,G)–cone-manifold is a space such that each point has a neighborhood modelled on the cone of a compact, connected (X_p,G_p)–manifold.

One reason for considering inhomogeneous model spaces (X,G) is that even if we start with an example as homogeneous as $(\mathbb{CP}^n, U(n))$, during the inductive examination of tangent cones we soon encounter model spaces (X,G) where G is not transitive.

If C is an n–dimensional (X,G)–cone-manifold, then a point $p \in C$ is a *regular* point if p has a neighborhood equivalent as an (X,G)–space to a neighborhood in X, otherwise it is *singular*. It follows by induction that regular points are dense, and that C is the metric completion of its set of regular points. The distinction between regular points and singular points can be refined to give the concept of the *codimension* of a point $p \in C$. If the only cone type neighborhood that a point p belongs to is the neighborhood centered at p, then p has codimension n. Otherwise, there is some cone neighborhood centered at a different point q that p belongs to, and the codimension of p is defined inductively to be the codimension of the ray through p in (X_q,G_q).

By induction, it follows that each point p of codimension k is on an $(n-k)$–dimensional stratum of C which is locally isometric to a totally geodesic subspace $E_p \subset X$ — this stratum is an $(E_p, G(E_p))$–space, where $G(E_p)$ is the subgroup of G sending E_p to itself.

An oriented Euclidean, hyperbolic, or elliptic cone-manifold of dimension n is a space obtained from a collection of totally geodesic simplices via a 2 to 1 isometric identification of their faces.

Suppose that n numbers α_i are specified, all less than 1, such that $\sum \alpha_i = 2$. Let $C(\alpha_1, \alpha_2, \ldots, \alpha_n)$ be the space of Euclidean cone-manifold structures on the sphere with n cone singularities of curvature α_i (cone angles $2\pi(1-\alpha_i)$), up to equivalence by orientation-preserving similarity. We do not specify any homotopy class of map relative to the cone points, nor any labelling of the cone-points in these equivalences. Let $P(A;\alpha_1,\ldots,\alpha_n)$ be the finite-sheeted covering in which the cone points can be consistently labelled. Note that the fundamental group of $P(A;\alpha_1,\ldots,\alpha_n)$ is the pure braid group of the sphere, and the fundamental group of $C(\alpha_1,\ldots,\alpha_n)$ is contained in the full braid group of the sphere and contains the pure braid group. The exact group depends on the collection of angles, since only cone points with equal angles can be interchanged.

How can we understand these spaces? We will first construct a local coordinate system for the space of shapes of such cone-metrics, in a neighborhood of a given metric g.

Proposition 3.1 (Cone-metrics have triangulations) *Let C be any metric on the sphere which is locally Euclidean except at isolated cone-points of positive curvature. Then C admits a triangulation in the sense of a subdivision of C by images of geodesic Euclidean triangles, possibly with identifications of vertices and/or edges, with vertex set the set of cone points.*

Proof Associated to each cone point v of C is the open *Voronoi region* for v, consisting of those points $x \in C$ which are closer to v than to any other cone point, and furthermore, have a unique shortest geodesic arc connecting x to v. A *Voronoi edge* consists of points x that have exactly two shortest geodesic arcs to cone points. Each Voronoi edge is a geodesic segment. It can happen that a Voronoi edge has the same Voronoi region on both sides if C has a fairly long, skinny region with a cone point v far from other cone points. Take any point x on a Voronoi edge, and let D be the largest metric ball centered at x whose interior contains no cone points. Then D is the image of an isometric immersion of a Euclidean disk D' with exactly two points $v_1, v_2 \in \partial D'$ that map to cone points of C. The chord $\overline{v_1 v_2}$ of D' maps to an arc in C. The collection of all such arcs have disjoint interiors, for if not, one could lift the situation to $\mathbb{E}^2$: whenever two chords of two distinct disks in $\mathbb{E}^2$ cross, at least one of the four endpoints is in the interior of at least one of the two disks.

The *Voronoi vertices* are those points that have three or more shortest arcs to cone points. The largest metric disk about a Voronoi vertex with no cone points in the interior is the image of an isometrically immersed Euclidean disk. The convex hull of the set of points on the boundary of the Euclidean disk that map to cone points is a convex polygon mapping to C with boundary mapping to the edges previously constructed. Subdivide each of these polygons into triangles by adjoining diagonals. The result is a geodesic triangulation of C in the sense of the proposition whose vertex set is the set of cone points. □

Let T be any geodesic triangulation of the cone-manifold C; it might or might not be obtained by this construction. Choose one of the edges of T, and map it isometrically into $\mathbb{C}$, with one endpoint at the origin. This map extends to an isometric developing map $D\colon \tilde{C} \to \mathbb{C}$, where $\tilde{C}$ is the universal cover of the complement C_0 of the vertices of C. Associated with each directed edge e of the triangulation $\tilde{T}$ of $\tilde{C}$ is a complex number $Z(e)$ (really a vector), the difference between its endpoints. These vectors satisfy the cocycle condition, that the sum of the vectors associated to the oriented boundary of a triangle is 0. Let $H\colon \pi_1(C_0) \to \mathrm{isom}(\mathbb{E}^2)$ be the holonomy of the Euclidean structure, and let $H_0\colon \pi_1(C_0) \to S^1 \subset \mathbb{C}$ be its orthogonal part. If τ_γ is the covering transformation of $\tilde{C}$ over C_0 associated with the element $\gamma \in \pi_1(C_0)$, then $Z(\tau_\gamma(e)) = H_0(\gamma)Z(e)$. In other words, it is a cocycle with twisted coefficients — the coefficient bundle is the tangent space of C_0. Euclidean structures near C, up to scaling, are parametrized by cocycles near Z, up to multiplicative complex numbers, since any nearby cocycle determines a collection of shapes of triangles which can be glued together to form a cone-manifold with the same set of cone angles.

It is clear that change of coordinates, from those given by T to those given by a triangulation T', is a linear map, since the developing map for the edges of T' can be computed as a linear function of a cocycle expressed in terms of T.

Proposition 3.2 (Dimension is $n-2$) *The complex dimension of the space of cocycles, as described above, is $n-2$, where n is the number of vertices.*

See [8] for various computations related to this.

Proof We will describe a concrete construction for a basis for the cocycles, which amounts to making a gluing diagram to construct C from a polygonal region on a cone.[1]

We will divide the set of edges into *leaders* (the basis elements) and *followers*. Begin by picking any vertex v_{last} of T, and designate all edges leading into that vertex as followers. Now pick a tree in the 1–skeleton connecting all vertices except v_{last}: these will be leaders. The remaining edges are additional followers. There is a dual tree, in the dual 1–skeleton of the cell-division formed by removing the followers touching v_{last}, consisting of the 2–cells and the remaining followers.

Suppose the value of a 1–cocycle is specified on each of the leaders. We can then calculate it on each of the followers, as follows. Inductively, if the current dual tree of undetermined values is bigger than a single point, pick a leaf of the tree. This is a follower which is part of a triangle whose other two sides have determined values; from them, we determine the value for the follower to

[1] In the general, complicated cases, this would likely be an immersed polygonal region on a cone.

satisfy the coboundary condition on the given triangle. What remains is still a tree.

Finally, we are left with everything determined, except for v_{last} and its remnant cluster of followers. At this point, we have enough information to determine the affine holonomy around v_{last}. The orthogonal part is a non-trivial rotation, so that it has a unique fixed point. The values of the cocyle for the remaining followers are determined by pointing them toward the fixed point.

A spanning tree for the $n-1$ vertices excluding the last has $n-2$ edges, so the space of cocycles is $\mathbb{C}^{n-2}$. The projective space then has dimension $n-3$. □

The area of a cone-manifold structure defines a hermitian form on the space of cocyles: that is, given a cocycle Z, $A(Z) = \frac{1}{4}\sum_{\text{triangles}} ie_1\bar{e}_2 - ie_2\bar{e}_1$ where in local coordinates e_1 and e_2 are successive edges of the triangle proceeding counterclockwise. Obviously $A(Z)$ is independent of choice of local coordinates.

Proposition 3.3 (Signature $(1, n-3)$) *If each of the* $\alpha_i > 0$, *then* A *is a hermitian form of signature* $(1, n-3)$.

Proof We have seen this illustrated in several examples already. There is a general procedure for diagonalizing the expression for area. If C has only three vertices, then the vector space is only one dimensional, so A is necessarily positive definite: it is proportional to the square of the length of any of the edges of T.

We have already seen the special case that there are four cone angles all equal to π, under the guise of $P(n; 3, 3, 3, 3)$. The expression for area is the determinant of a 2×2 real matrix, made of the real and imaginary parts of two of the values $Z(e)$. Since determinants can be positive or negative, this is a hermitian form of signature $(1, 1)$.

In every other case, there are at least two cone angles whose curvatures have sum less than 2π. Construct any geodesic path e between them, slit C open, and glue a portion of a cone with curvature the sum of the two curvatures to obtain a cone-manifold C' with one fewer singular points (figure 9). The area of C is the area of C' minus a constant times the square of the length of e. This gives an inductive procedure for diagonalizing A, inductively showing that the signature of the area is $(1, n-3)$. □

The set of positive vectors in a Hermitian form of signature $(1, n-3)$ up to multiplication by scalars, is biholomorphic to the interior of the unit ball in

$\mathbb{C}^{n-3}$, and is known as *complex hyperbolic space* $\mathbb{CH}^{n-3}$. A metric of negative curvature is induced from the Hermitian form; as a Riemannian metric, its sectional curvatures are pinched between -4 and -1. Therefore, $C(A; \alpha_1, \ldots, \alpha_n)$ is a complex hyperbolic manifold.

It is not metrically complete, however. Any two singular points of a c whose curvature adds to less than 2π can collide as the cone-metric changes a finite amount, measured in the complex hyperbolic metric. We will next examine how to adjoin to $C(\alpha_1, \ldots, \alpha_n)$ the degenerate cases where one or more of the cone points collide, to obtain a space $\bar{C}(\alpha_1, \ldots, \alpha_n)$ which is the metric completion of $C(\alpha_1, \ldots, \alpha_n)$.

Each element c of $\bar{C}(\alpha_1, \ldots, \alpha_n)$ is associated with some partition P of the angles α_i; c is a Euclidean cone-manifold where each cone point is associated with a partition element $p \in P$ and has curvature equal to the sum of the elements of p. We regard two partitions as equivalent if one can be transformed to the other by a permutation of the index set which preserves the values of the α_i. A limit of a sequence of cone-manifolds associated with some partition will be associated with a coarser partition, if distances between some of the cone points in the sequence tend to zero.

Theorem 3.4 (Completion is cone-manifold) *The metric completion of* $C(A, \alpha_1, \ldots, \alpha_n)$ *is* $\bar{C}(\alpha_1, \ldots, \alpha_n)$, *which is a complex hyperbolic cone-manifold.*

Proof There is a very natural way to describe regular neighborhoods for the stratum S_P corresponding to a partition P of the set of curvatures concentrated at cone points.

Consider an element $c \in C(\alpha_1, \ldots, \alpha_n)$ such that the cone points are clustered in accordance with P. We may assume that the diameter of each cluster is less than the minimum distance from the cluster to any cone point not in the cluster, and less than some small constant ϵ.

The holonomy for a curve which goes around any cluster D is a rotation by the total curvature of D, unless the total curvature is 2π. When the total curvature of D is 2π, the holonomy is a translation. If the holonomy is actually a rotation, it leaves invariant each of a family of circles; with our assumption that the cluster is isolated from other cone points, the encircling curve is isotopic to one of these circles.

If the total curvature of $K(D)$ is less than 2π, the surface of c near such a circle isometrically matches a cone with apex on the same side as the cluster, with cone point of curvature $K(S)$. In this case, we can define a new cone-manifold $p(c)$ by cutting out each cluster, and replacing it by a portion of this cone. In local coordinates, this gives a local orthogonal projection from a neighborhood of c to S_P. The distance from the singular stratum is $\sqrt{\operatorname{area} p(c) - \operatorname{area} c}$. Note that the normal fibers for strata corresponding to subclusters of a cluster are contained in normal fibers for the larger cluster.

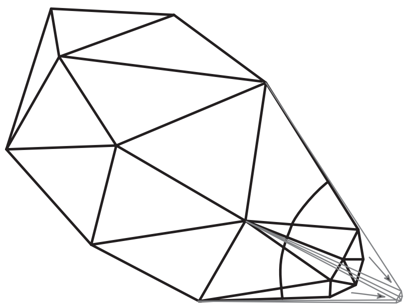


Figure 11: Any cluster of cone points close together compared to the distance to other cone points can be shrunk to a single cone point. This process gives a radial structure to a neighborhood of a singular point in the space of cone-metrics with designated curvatures on a sphere.

The total curvature cannot be greater than 2π, if ϵ is chosen properly: in that case, c would match the surface of a cone with apex on the opposite side from the cluster. The area of C is less than the area of the portion of cone, plus the area within the cluster, so that if ϵ is small compared to θ/A, where θ is the minimum value by which a curvature sum can exceed 2π, this cannot occur.

A cluster of arbitrarily small diameter with total curvature 2π can occur, but this forces the diameter of c to be large: in this case, c matches the surface of a cylinder outside a neighborhood of the cluster, and there is a complementary cluster at the other end of the cylinder. As c moves a finite distance in the complex hyperbolic metric, its diameter cannot go to infinity, so no such cluster goes to 0 in diameter in the metric completion of $C(\alpha_1, \dots, \alpha_n)$.

Within any bounded set of $C(\alpha_1, \dots, \alpha_n)$, we are left only with the case of small diameter clusters whose total curvature is less than 2π.

It is now easy to see that $\bar{C}(\alpha_1, \dots, \alpha_n)$ is the metric completion of $C(\alpha_1, \dots, \alpha_n)$ and that it is a complex hyperbolic cone-manifold. □

Of particular importance are the strata of complex codimension 1 or real codimension 2. These strata correspond to the cases when only two cone points of c have collided. What are the cone angles around these strata?

Proposition 3.5 (Cone angles around collisions) *Let S be a stratum of $C(\alpha_1, \dots, \alpha_n)$ where two cone points with curvature α_i and α_j collide.*

If $\alpha_i = \alpha_j$, the cone angle $\gamma(S)$ around S is $\pi - \alpha_i$, otherwise it is $2\pi - \alpha_i - \alpha_j$.

In other words, the cone angle in parameter space is the same as the physical angle two nearby cone points go through, as measured from the apex of the cone that would be formed by their collapse, when they revolve about each other until they return to their original arrangement.

Proof When cone points x_i and x_j with these two angles are close together on a cone manifold c, we can think of c as constructed from $p(c)$ by replacing a small neighborhood of the cone by a portion of a modified cone $D(\alpha_i, \alpha_j)$ with two cone points. The shape of $D(\alpha_i, \alpha_j)$ is uniquely determined by α_i and α_j up to similarity. Thus, the shape of c is determined by selecting the point x_i on $p(c)$, and may be represented by $p(c)$ together with the vector V from the combined cone point of $p(c)$ to x_i. In local inhomogeneous coordinates coming from a choice of a triangulation, V is a locally linear function, described by a single complex number.

If $\alpha_i = \alpha_j$, then when the argument of V is increased by half the cone angle, or $\pi - \alpha_i$, x_i and x_j are interchanged, and the resulting configuration is indistinguishable. Therefore, $\pi - \alpha_i$ is the cone angle along S, (and $\pi + \alpha_i$ is the curvature concentrated at S). If $\alpha_i \neq \alpha_j$, the argument of V must be increased by the cone angle, $2\pi - \alpha_i - \alpha_j$, before the same configuration is obtained again. In this case, $2\pi - \alpha_i - \alpha_j$ is the cone angle along S, and $\alpha_i + \alpha_j$ is the curvature concentrated along S. □

More generally, if S is a stratum of complex codimension j representing the collapse of a cluster of $j+1$ cone points, each normal fiber to S is a union of 'complex rays', swept out by an ordinary real ray by rotating it the direction i times the radial direction. The space of complex rays is the complex link of the stratum, a complex cone-manifold whose complex dimension is one lower. The real link is a Seifert fiber space over the complex link, with generic fiber a circle of length α which we can call the *scalar cone angle* $\gamma(S)$ at S. We define the *real link fraction* of S to be the ratio of the volume of the real link of S to the volume of S^{2j-1} (the real link in the non-singular case), and similarly the *complex link fraction* is the ratio of the volume of the complex link to the volume of $\mathbb{CP}^{j-1}$.

Proposition 3.6 (Cone angles for multi-collisions) *Let S be a stratum of complex codimension j where $j+1$ cone points of curvature $\kappa_1, \ldots, \kappa_j$ collapse.*

Let N be the order of the subgroup of the symmetric group S_j that preserves these numbers. Then:

a) *The scalar cone angle is*

$$\gamma(S) = 2\pi - \sum_i \kappa_i.$$

b) *The complex link fraction is*

$$\frac{(\gamma(S)/2\pi)^{j-1}}{N}.$$

c) *The real link fraction is*

$$\frac{(\gamma(S)/2\pi)^{j}}{N}.$$

Proof The proof of part (a) is the same as above, with the observation that a cluster of 3 or more cone points can always be slightly perturbed to make it asymmetrical, so in the generic fiber of the Seifert fibration (obtained by rotating the cluster of cone points) no permutations of the cone points occur.

For (b), think first about the case that all cone angles are different, so as to avoid a symmetry group at first. A neighborhood of S is then a manifold, isomorphic to the limiting case when $\kappa_i \to 0$, the space of $(j+1)$–tuples in the plane up to affine transformations. The complex link is a complex cone-manifold structure on $\mathbb{CP}^{j-1}$. If ω is a closed 2–form on $\mathbb{CP}^{j-1}$ that integrates to 1 over $\mathbb{CP}^1$, then ω^{j-1} gives the fundamental class for $\mathbb{CP}^{j-1}$. (This calculus works readily for cone metrics with differential forms that are suitably continuous.) We conveniently obtain such a form as some constant multiple α of the Kähler form of the model geometry $\mathbb{CP}^{j-1}$ of the link. One way to determine α is to reduce to the case $j = 2$ by clustering the cone points into three groups which are collapsed along a codimension 2 stratum limiting to S. In the case $j = 2$, the complex link is S^2 with cone points of curvature $\kappa_1 + \kappa_2$, $\kappa_2 + \kappa_3$ and $\kappa_3 + \kappa_1$. This uses up $2\sum \kappa_i$ out of the total curvature 4π of S^2, so the area of a constant curvature metric is reduced by a factor of $\gamma(S)$.

Part (c) follows from (a) and (b), since the real link fraction is the product of the complex link fraction with $\gamma(S)/2\pi$.

The case with symmetry follows by dividing the asymmetric configuration space by the symmetry. □

4 Orbifolds

An *orbifold* is a space locally modelled on $\mathbb{R}^n$ modulo finite groups; the groups vary from point to point. For an exposition of the basic theory of orbifolds, see

[11]. Our orbifolds will be (X,G)–orbifolds, locally modelled on a homogeneous space X with a Lie Group G of isometries. It is easily seen by induction on dimension that an orientable (X,G)–orbifold has an induced metric which makes it into a cone-manifold. (Use the naturality of the exponential map.)

Here is a basic fact about the relation between cone-manifolds and orbifolds, which essentially is a rephrasing of Poincaré's theorem on fundamental domains:

Theorem 4.1 (Codimension 2 conditions suffice) *Let C be an (X,G)–cone-manifold. Then C is a "weakening" of the structure of an orbifold if and only if all the codimension 2 strata of C have cone angles which that are integral divisors of 2π.*

Proof An orientation-preserving group of isometries whose fixed point set has codimension 2 is a subgroup of $SO(2)$, and the only possibilities are $\mathbb{Z}/n$. The cone angle along such a stratum in an orbifold is therefore an integral divisor of 2π, and the condition is necessary.

The converse can be proved by induction on the codimension of the singular strata of C. Clearly, it works for strata of codimension 2. Suppose that we have proven that C has an orbifold structure in the neighborhood of all strata up through codimension k. Let S be a singular stratum of codimension $k+1$, and consider the neighborhood U of a point $x \in S$. This neighborhood can be taken to have the form of a bundle over a neighborhood of x in S, with fiber the cone on a k–dimensional cone-manifold N, the normal sphere to S. The normal sphere N is modelled on (S^k, G), where $G \subset SO(n)$. By induction, N is an orbifold; its universal cover must be S^k, since for $k \geq 2$ the sphere is simply-connected. Therefore the cone on N is the quotient of B^{k+1} by the group of covering transformations of S^k over N, and therefore U is also the quotient space of a neighborhood in X by the same group. Thus, C is an orbifold. □

To illustrate, let's look at some of the local orbifold structures that arise in multi-way collisions. When k cone points of equal curvature $2\pi\alpha$ collide, the order of the local group $\Gamma(S)$ for a stratum S is the reciprocal of the real volume fraction, so from 3.6, setting $\alpha = \gamma(S)/2\pi$ we have

$$\#\Gamma(S) = \frac{k!}{(1-k\alpha)^{k-1}} \qquad \left[\frac{1}{1/2-\alpha} \in \mathbb{Z} \quad \& \quad 0 < \alpha < 1/k\right]$$

The only three cases satisfying the condition for three colliding equal angle cone points are when α is $1/6$, $1/4$ and $3/10$. The complex links in these three cases are the quotient orbifolds of the sphere by the oriented symmetries of one of the regular polyhedra: $(2,3,3)$, $(2,3,4)$ or $(2,3,5)$. The real link is S^3 with a cone axis along the trefoil knot of order 3, 4 or 5. The formula gives orders for these groups of 24, 96 and 600. (This can be quickly confirmed by an automated check using the 3–dimensional topology program `Snappea` to obtain presentations for the orbifold fundamental groups and feeding then to a group theory program such as `Magnus`.)

An interesting example of a collision of cone points of unequal curvatures is $(19\pi/30, 11\pi/30, 29\pi/30)$. The real link is an orbifold with $(2,3,5)$ cone axes on the 3–component Hopf link. In this case, $\alpha = 1/60$ and $\#(\Gamma(S)) = (60)^2 = 3600$.

The biggest possible multiple collision is when 5 points of curvature $\pi/3$ collide. The local group for this collision has order $6^4 5! = 155,520$.

Infinitely many of the modular spaces for cone-metrics with 4 cone points are orbifolds of complex dimension 1, but for higher dimensional modular spaces, only 94 are orbifolds. These are tabulated in the appendix.

5 Proof of main theorem

Proof of Theorem 0.2 Most of this theorem follows formally from Theorem 3.4, Proposition 3.5, and Theorem 4.1. What still remains is a discussion of the volume of $\bar{C}(\alpha_1, \ldots, \alpha_k)$.

The only case in which $X = \bar{C}(\alpha_1, \ldots, \alpha_k)$ is not compact is where there are cone-manifolds $x \in X$ whose diameters tend to infinity. In such a case, if we normalize so that the area of x is 1, there must be subsets of x with large diameter and small area, free from cone points. This implies that x has subsets which are approximately isometric to a thin Euclidean cylinder. If $\gamma \subset x$ is a short curve going around such an approximate cylinder, then the angle of rotation for γ must be a sum of a subset of the $\{\alpha_i\}$. There are only a finite number of possibilities, so if the diameter is large enough, a neighborhood of γ of large diameter is actually a cylinder. Once γ is determined, the shapes of the two pieces of x cut by γ can be specified independently, and a scale factor $\mathrm{length}(\gamma)^2/\mathrm{area}$ (less than some constant ϵ) together with an angle of rotation can also be specified independently.

It will follow that the ends of x are in 1–1 correspondence with partitions Q of the set of curvatures into two subsets each summing to 2π, if we verify two points:

(i) for any such partition Q, there exists an $x \in X$ with a geodesic γ separating the cone points according to Q, and

(ii) the subspace $X_{\gamma,\epsilon}$ consisting of cone-manifolds in X with a geodesic γ of length ϵ which separates the cone points according to Q is connected.

Actually, the proof does not logically depend on either point, and it is a slight digression to prove them, but it seems worth doing anyway.

An easy demonstration of (i) is to construct a polygon with angles $\pi - \alpha_i/2$. It is easy to find a very thin polygon realizing Q. Doubling such a polygon gives a suitable cone-manifold x.

We will describe an explicit construction for (ii). Let us begin with the special case of $c \in X_{\gamma,\epsilon}$ which are obtained by doubling a convex Euclidean polygon whose angles are half the cone angles for X. It is easy to connect any two convex polygons with the same sequence of angles by a family of polygons having the same angles. If we allow degenerate cases as well, where two angles coincide, the order is irrelevant. Therefore, this special subspace of $X_{\gamma,\epsilon}$ is connected.

Therefore, it suffices to connect any $c \in X_{\gamma,\epsilon}$ to something obtained by doubling a convex polygon. Construct a maximal cylindrical neighborhood N_1 of γ with geodesic boundary. There is at least one cone point on each boundary component of N_1. Let β be one of the boundary components, and $x_1 \in \beta$ a cone point, with curvature α. If c is cut along β, the portion on the other side of β from N_1 has boundary consisting of a geodesic with a convex angle of $\pi - \alpha$ at x_1, and possibly additional angles if it contains other cone points. There is a circular arc β' through x_1, contained in N, which appears to have a convex angle of $\pi - \alpha$ from within N, but appears to be smooth at x_1 when viewed from the outside. Let U_1 be the "outside" component obtained by cutting along β'. Its boundary is now locally isometric to a circle, and a neighborhood, like on a cone, is foliated by parallel circles.

Deform c, by shrinking the "interesting part" of U_1 relative to the rest of c, so that the next cone point in U_1 is not close to β'. Let N_2 be a maximal neighborhood of β' which is foliated by parallel circles, and let x_2 be a point on its boundary. Adjust by a rotation of U_1 until the geodesic through x_1 perpendicular to the foliation by circles hits at x_2. Draw a circular arc through x_2, within U_1, which appears smooth from the outside neighborhood U_2.

This process can be continued, in the same manner, until the last neighborhood U_k is a cone. The geodesic through x_{k-1} automatically hits the cone point. Now do the same process on the other side of N_1, first adjusting by a rotation so a geodesic through x_1 perpendicular to the foliation of N_1 by parallel circles hits at a cone point.

After this sequence of deformations, we have a cone-manifold with a geodesic Hamiltonian path through all the cone points, such that at cone points internal to it the two outgoing geodesics bisect the cone angle. The path can be completed to a curve by one additional geodesic (this is easy to see if you draw the figure in the plane obtained by cutting along the path; it is made of two convex arcs, and has bilateral symmetry).

Note that a similar process works for a general cone-manifold: we do not really need γ for this construction, we can begin at any cone point, and work outward from it.

We call the ends of X *cusps*, in accordance with terminology for manifolds and orbifolds. To justify this word, note that each cusp is foliated by complex geodesics with respect to the Hermitian metric, obtained by rotating the two ends of c with respect to each other and by scaling. The complex geodesics are locally isometric to the hyperbolic plane. The pure scaling, which may be thought of as inserting extra lengths of cylinder between the two ends, generates a real geodesic. These real geodesics converge, as the shrinking increases. The convergence is exponential, so the total volume of each cusp is finite. □

6 The icosahedron and other polyhedra

Let A be the subgroup of isometries of $\mathbb{C}$ which take **Eis** into itself. We may think of the classes of triangulations $P(n; k_1, \ldots, k_m)$ as the space of $(\mathbb{E}^2, A)$–cone-manifolds of area n (measured in double triangles) and cone angles $k_i\pi/3$. They consist of elements of $C(k_1\pi/3, \ldots, k_n\pi/3)$ equipped with a reduction of the $(\mathbb{E}^2, \mathrm{isom}(\mathbb{E}^2))$ structure to $(\mathbb{E}^2, A)$. In more concrete terms, a triangulation is given by a cocycle whose coefficients are elements of **Eis**.

Euclidean cone-manifolds sometimes admit several inequivalent reductions to $(\mathbb{E}^2, A)$—in other words, there are some cone-manifolds that can be subdivided into unit equilateral triangles in more than one way. In complex Lorentz space $\mathbb{C}^{(m-3,1)}$, the set of cocyles with a certain total area form a sheet of a hyperboloid. The hyperboloid fibers over complex hyperbolic space, with fiber a circle (corresponding to multiplication of the cocycle by elements of the unit circle). The set of triangulations are lattice points in $\mathbb{C}^{(m-3,1)}$, and the value of the Hermitian form counts the number of triangles—multiple unit equilateral triangulations of a Euclidean manifold correspond to fibers that intersect more than lattice points. (All lattice points come in groups of 6 whose ratios are units in the ring **Eis**.)

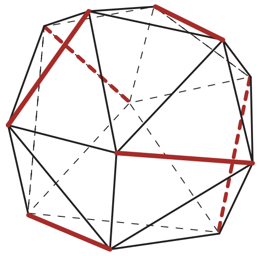


Figure 12: If an icosahedron is slit along 6 disjoint arcs joining its vertices in pairs, conical caps can be inserted to turn it into an octahedron.

The "biggest" of the classes of triangulations is

$$P(n;1,1,\ldots,1)\subset J=C(\pi/3,\pi/3,\ldots,\pi/3),$$

the one which contains the icosahedraon. The "completion" $\bar{P}(n;1,\ldots,1)\subset\bar{J}$ which includes degenerate cases contains all the other classes of triangulations.

By theorem 1.2, $\bar{J}$ is a complex hyperbolic orbifold of dimension 9. The cone angles around the complex codimension 1 singular strata are $2\pi/3$.

There is a concrete construction to describe an arbitrary element of $\bar{J}$ or of $\bar{P}(1,\ldots,1)$, as follows. Suppose first that $x\in J$ is an arbitrary Euclidean cone-metric on the sphere with all cone points having curvature $\pi/3$. Choose a collection of 6 disjoint geodesic arcs with endpoints on the cone points. Slit along each of these arcs.

Locally near the endpoints of the arcs, the developing map maps the slit surface to the complement of a $60°$ angle. A neighborhood of the slit develops to a region outside an equilateral triangle in the plane; when you go once around the slit, the developing image goes $2/3$ of the way around the triangle.

For each slit, take $2/3$ of an equilateral triangle with side equal to the length of the slit, fold it together to form a cone point in the center of the original triangle with curvature $2\pi/3$ and glue it into the slit. The result is a cone-manifold $f(x)$ like the octahedron, in $C(2\pi/3,2\pi/3,2\pi/3,2\pi/3,2\pi/3,2\pi/3)$.

As in section 2, we can analyze the shape of $f(x)$ by joining its cone points in pairs by disjoint geodesic segments, slitting open, and extending to give an element of $C(4\pi/3,4\pi/3,4\pi/3)$ (which is a single point).

If $x\in\bar{J}-J$, the analysis still works: treat the cone points as cone points with multiplicity, and use zero-length slits as much as possible at cone points with curvature greater than $\pi/3$. At the first step, the slits of positive length pair the cone points with curvature an odd multiple of $\pi/3$. When the slits are filled in, the curvature at each of the endpoints is decreased by $\pi/3$, and the resulting cone-manifold has all curvatures an even multiple of $\pi/3$. For the second step, note that no cone point can have curvature $6\pi/3$ or bigger. In this case, the slits of positive length join cone points with curvature $2\pi/3$.

An arbitrary $x \in \bar{J}$ can be reconstructed by reversing this procedure.

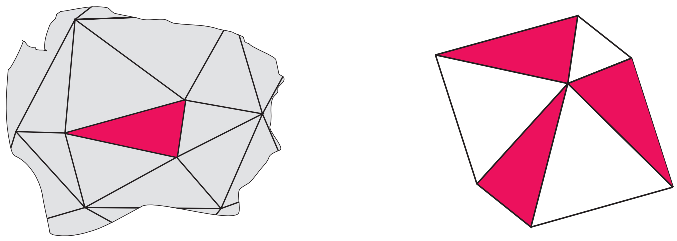


Figure 13: A geodesic triangle whose vertices are on cone points of curvature $\pi/3$ has a deleted neighborhood that develops to the deleted neighborhood of a Napoleon hexagon, formed from three copies of the triangle and three equilateral triangles. This process, applied to the 12 vertices of an icosahedron-like cone-manifold grouped into 3's, recursively reduces it to a tetrahedron-like cone-manifold

There are many alternative coordinate systems for J. For example, another construction is to group the cone points in 3's, by constructing 4 disjoint geodesic triangles with vertices at cone points. If these triangles are cut out, then the developing image of what is left is discrete; it comes from a $2,2,2,2$ group acting in the Euclidean plane. The developing image is the complement of a certain union of hexagons about the lattice of elliptic points. The hexagons are not arbitrary, however—the hexagons $H(T)$ that arise are hexagons that come from Napoleon's theorem, constructed as follows: Suppose T has sides a, b, and c, in counterclockwise order. We will construct 6 triangles around the vertex v of T between a and b. First construct an equilateral triangle on side a. Construct another triangle T_1 congruent to T on the free side of the equilateral triangle which is incident to v. Side c of T_1 also touches v; on this, construct a second equilateral triangle. Continue alternating copies of T and equilateral triangles until it closes, yielding $H(T)$.

Note that $H(T)$ has sides a, b, c, a, b, c in counterclockwise order. The complement of $H(T)$ modulo a rotation of 180° has boundary which matches the boundary of T; when it is glued in, three cone points of curvature $\pi/3$ are obtained at the vertices of T.

A general $x \in \bar{J}$ can be obtained by choosing first a $2,2,2,2$ group, and then choosing four hexagons $H(T_i)$ centered about the four classes of vertices. Form the quotient of the complement of the hexagons by the group, and glue in the triangles T_i. If the hexagons are disjoint and nondegenerate, $x \in J$.

From this concrete point of view, what is amazing is that these coordinate systems have a global meaning, since $\bar{J}$ is an orbifold: even if one chooses a collection of hexagons $H(T_i)$ which overlap, they determine a unique Euclidean cone-manifold, provided the net area (computed formally) is positive.

Using these constructions, it is not hard to show that $P(n; 1, \ldots, 1)$ contains 1 or more elements for all values of n starting with 10, with 11 as the sole exception. If there were an element T of $P(11, 1, \ldots, 1)$, it would have 13 vertices and 22 triangles. One could then construct a spherical cone-manifold by using equilateral spherical triangles with angles $2\pi/5$. This cone-manifold would have only one cone point — which is manifestly impossible, since the holonomy for a curve going around the cone point is a rotation of order 5, but at the same time the holonomy is trivial since the curve is the boundary of a disk having a spherical structure.

From the picture in $C^{(1,9)}$, it follows that the number of non-negatively curved triangulations having up to $2n$ triangles is roughly proportional to the volume of the intersection of some cone with the ball of radius $\sqrt{n}$ in this indefinite metric. The cone in question is neither compact nor convex, but since it comes from a fundamental domain for the group action, its intersection with the ball of norm less than any constant has finite 10–real-dimensional volume. Therefore, the number of triangulations with up to $2n$ triangles is $O(n^{10})$.

7 An explicit construction and fundamental domain

Another method for constructing, manipulating and analyzing non-negatively curved cone structures goes as follows:

Given $k + 1$ real numbers $\alpha_0, \alpha_1, \ldots, \alpha_k \geq 0$ whose sum is 4π.

To Construct Euclidean cone-metrics with the α_i as curvatures.

Choose a k–gon P in the plane, with edges $e_1 \ldots, e_k$.

Construct $(i = 1, \ldots k)$: An isosceles triangles T_i with base on e_i, apex v_i, apex angle α_i, pointing inward if $\alpha_i < \pi$, pointing outward if $\alpha_i > \pi$.

Condition A the triangles T_i are disjoint from each other and disjoint from P except along e_i.

Let Q be thefilled polygon obtained from P by replacing each e_i by the other two sides f_i and g_i of T_i.

Glue f_i to g_i to obtain a cone manifold. The vertex v_i becomes a cone point of curvature α_i. The other k vertices of P all join to form a cone point of cone angle α_0.

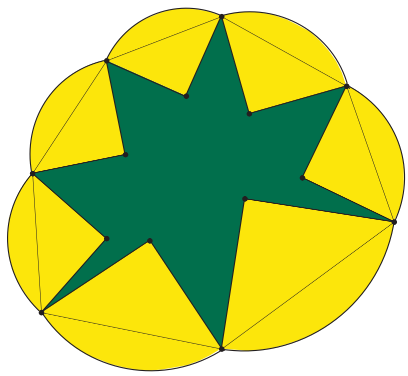


Figure 14: A cube-like cone-metric (8 cone-angles of curvature $\pi/2$) can be constructed by removing isosceles right triangles from the sides of a heptagon and gluing the resulting pairs of equal sides. The seven sharp angles all come together to form the eigth cone point. This illustration (along with the others in this section) was constructed with the program *Geometer's Sketchpad*, where the shape can be varied while preserving the correct geometric relations.

As examples, see figure 14 for a cube-like cone-manifold, or figure 15 for a triangulation of S^2 with 23 vertices and 42 triangles constructed from an icosahedral-like cone-manifold.

Here is the inverse construction. Given a cone-metric with n cone points on S^2:

Choose one of the cone points v_0.

Find for each other cone point v_i a shortest path a_i from v_0 to v_i. The a_i are necessarily simple and disjoint, except at v_0.

Cut along all these paths, to obtain a disk equipped with a Euclidean metric whose boundary is composed of $2(n-1)$ straight segments, each paired to an adjacent segment of the same length and forming an angle equal to the corresponding cone angle. (See figure 16.)

We will show that if P is a cone-metric on the sphere with positive curvature at each vertex, and if $S(P)$ (S because it resembles a star) is the metric on D^2 obtained by cutting P open as above, then $S(P)$ can be flattened out into the plane, that is, it is isometric to the metric of a filled simple polygon.

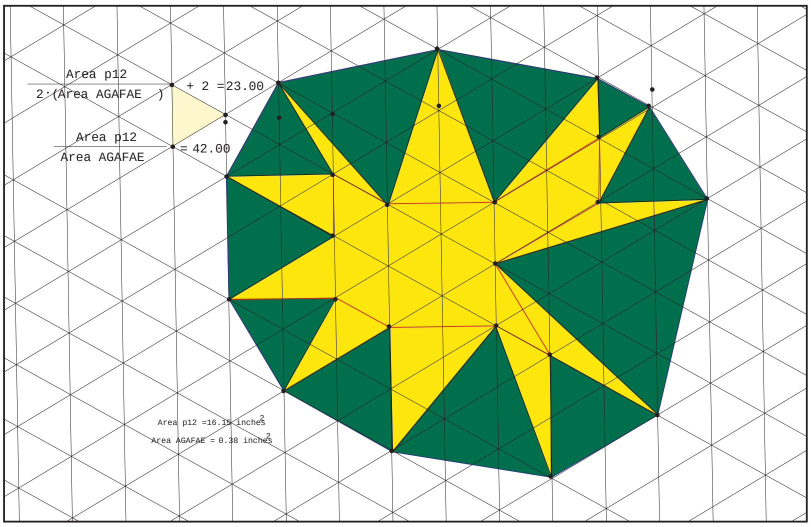


Figure 15: This is a diagram for a triangulation of S^2 with 42 triangles, having 12 vertices of order 5 and 11 vertices of order 6. Eleven–gons whose vertices lie in an ideal of index 3 (generated by $1 \pm \omega$) in the Eisenstein integers determine non-negatively-curved triangulations of S^2. Each valley between star-tips is folded together to form the triangulation; the star-tips come together at the base vertex. If the inner vertices of the 11–pointed star are closer to the two adjacent star-tips than to any other star-tips, this is the canonical 11–gon for the triangulation based at the given vertex.

Actually, we will enlarge $S(P)$ to a surface $F(P)$ (resembling a flower) by adjoining sectors of circles of with apex at each vertex v_i $(i > 0)$ of $S(P)$ and angle equal to the curvature at v_i in P, so that the resulting surface is locally Euclidean everywhere in its interior (as in figure 14).

The minimum distance within $S(P)$ of any point in $S(P)$ from one star-points that assemble at v_0 is equal to the distance of its image in P from v_0. Let $Q \subset S(P) \subset F(P)$ be the set of points whose minimum distance to $\partial F(P)$ is attained at 3 or more points on $\partial F(P)$. Then Q includes $\{v_1, \ldots, v_n\}$, as well as the vertices for the Voronoi diagram of the star tips within $S(P)$. Let R be the collection of open segments consisting of points whose minimum distance to $\partial F(P)$ is attained at two points of $\partial F(P)$; they are the edges of a tree, whose vertex set is Q. We decompose $S(P)$ into dart-shaped quadrilaterals, consisting of union of the two minimum-length arcs from points in an edge $\alpha \in R$ to $\partial F(P)$ (see figure 16). We'll call this dart $D(\alpha)$. Let $\theta(e)$ be the angle of $D(e)$ at either of its two wingtips (vertices that are not vertics of e). Note that

$$\sum_{e \in R} \theta(e) = 1/2\,(2\pi - \kappa(v_0))\,,$$

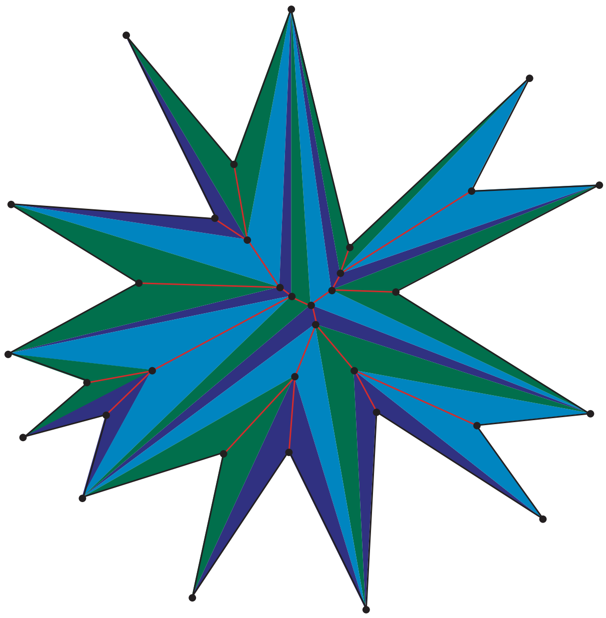


Figure 16: An irregular icosahedron sliced and flattened. A regular point ★ was chosen on a Euclidean cone-metric for S^2 having 12 cone points of curvature $\pi/3$. The surface has been cut along shortest geodesics from ★ to each of the cone points, and flattened into the plane to form a 24–gon resembling a star. The polygon has been subdivided into 45 dart-shaped quadrilaterals. Each quadrilateral is obtained from an edge of the cut locus of the original icosahedron (= the Voronoi diagram for the ★–tips, after cutting) by suspending to its two closest ★–tips.

that is, half the cone angle at the base vertex.

For any vertex $q \in Q$, let $D(q) \subset F(P)$ be the maximal disk in $F(P)$ centered about q_i. If q_1 and q_2 are the endpoints of $e \in R$, then the angle between the bounding circles of $D(q_1)$ and $D(q_2)$ is $\theta(e)$.

Proposition 7.1 *$F(P)$ has an isometric embedding in the plane $\mathbb{E}^2$.*

Proof The developing map $f\colon F(P) \to \mathbb{E}^2$ into the plane is an isometric immersion. To show that it is an embedding, it will suffice to establish that f restricted to the boundary $\partial F(P)$ is an embedding.

The boundary $\partial F(P)$ is composed of inward-curving circle arcs that meet at outward-bending angles. For each edge $e \in R$, there is a pair of these angles, where $\partial F(P)$ turns by an angle $\theta(e)$. For any two points $x, y \in \partial F(P)$, there is at least one path along $\partial F(P)$ where these bending angles sum to no more than π.

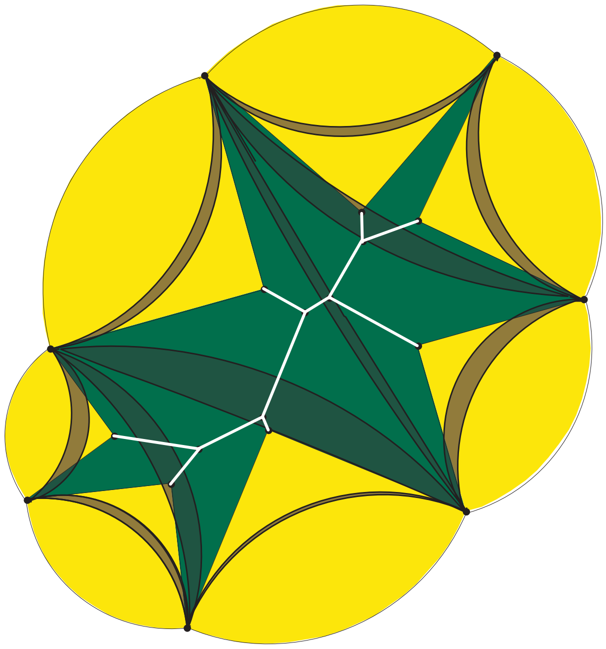


Figure 17: A hyperbolic view associated with the cut-open polyhedron. From the point of view of 3–dimensional hyperbolic geometry, if this figure is interpreted as lying on the boundary of upper half-space, the convex hull of its complement is the union of the hemispherical bubbles which rest on it. The boundary of the convex hull (with geometry induced from the upper half-space metric $ds^2 = 1/z^2(dx^2 + dy^2 + dz^2)$) is isometric with the hyperbolic plane, bent into hyperbolic 3–space. The sum of all bending angles is one half the cone angle at the base point (assembled from the tips of stars). Any immersion of the hyperbolic plane which has total bending measure less than π is an embedding, so this plane is embedded. There are immersed planes with total bending any number greater than π which are not embedded.

For an immersion of a disk in $\mathbb{E}^2$ to fail to be an embedding, any innermost arc whose endpoints are identified by the immersion must have total curvature at least $-\pi$ when orientations are chosen so that the total curvature of the entire boundary is 2π. This is clearly impossible in our situation, so f is actually an embedding. □

Remark This proposition can be rephrased in terms of 3–dimensional hyperbolic geometry: any pleated immersion $f\colon \mathbb{H}^2 \to \mathbb{H}^3$ with positive bending measure whose integral on any geodesic is no greater than π is an embedding. This is related to the inequality of Sullivan analyzed and refined by Epstein and Marden in [4], and also to the global characterization of bending for convex polyhedra of Rivin and Hodgson [10], see Rivin [9] for a related inequality for convex hyperbolic polyhedra.

The n–dimensional *associahedron* is a polyhedron whose vertices are labelled by triangulations of an $(n+3)$–gon using only vertices of the polygon, and whose k–cells are labelled by subdivisions of the $(n+3)$–gon obtained by removing k edges from a triangulation. They can be thought of as describing all ways to parenthesize or associate a product of $n+2$ symbols. The associahedron is a convex polyhedron in n–space that arises in a variety of mathematical context including the theory of loop spaces, Teichmüller theory and numerous combinatorial settings. The numbers of triangulations are called Catalan numbers.

If cone angles at the n points of P are fixed, the angles $\theta(e)$ that can occur for our dart quadrilaterals can be described by mapping the $n-1$ star tips of $S(P)$ to the vertices of a regular $(n-1)$–gon, mapping each dart quadrilateral $D(e)$ to an edge or a chord of this polygon, and labelling each edge by the angle $\theta(e)$. (In terms of hyperbolic geometry, this is an element of the measured lamination space for the ideal polygonal orbifold $(\infty, \infty, \dots, \infty)$.) This determines a point in a polyhedron $\mathcal{F}n$ closely related to an associahedron, namely, the join of the boundary of the dual of the $(n-4)$–dimensional associahedron with the $(n-2)$–simplex. (When the measure on the boundary of the polygon is zero, we get a point on the boundary of the dual of the $(n-4)$–dimensional associahedron. Measures on the polygon itself with fixed total weight form an $(n-2)$–simplex.)

The set of all possible functions $\theta(e)$ (which we refer to as measures, after the usage in hyperbolic geometry and Teichmüller theory) can be described globally as a convex polyhedron using dual train track coordinates, as follows: rotate a copy of the regular $(n-1)$–gon $1/(2n-2)$th of a revolution so it is out of phase with itself. Choose any triangulation of this rotated polygon, using only its vertices. For each edge f of this triangulation, let $m(f)$ be the sum of $\theta(e)$ where e intersects f. For any triangle with sides f, g, h, the quantities $m(f)$, $m(g)$ and $m(h)$ satisfy the three triangle inequalities $m(f)+m(g) \geq m(h)$ etc. These measures are subject to one linear constraint, namely, the sum of $m(f)$ where f ranges over the edges of the rotated polygon adds to the cone angle at the base vertex v_0.

For any set of numbers $\{m(f)\}$ satisfying the linear equation and linear inequalities, a measured lamination having total measure $\pi - \alpha_0/2$, where α_0 is the cone angle at v_0, can be reconstructed by a simple method familiar in the theory of measured foliations or normal curves on surfaces, by first solving for the picture in each triangle of the rotated polygon, then gluing the triangles together. From this measured lamination and from the specification of cone angles (in order) at $v_1, \dots v_{n-1}$, a star polygon in the plane can be constructed recursively, using the principle that the shape of any dart quadrilateral $D(e)$ is determined from $\theta(e)$ together with either of its other two angles. This star polygon is determined up to similarity. When glued together it forms a cone manifold with specified cone-angles.

If all cone angles are equal, and if we are not distinguishing shapes that are the same up to permutation of the labels of cone points $v_1, \dots v_{n-1}$, then we must divide $\mathcal{F}$ by action of the group of order $n-1$ rotations. The faces of $\mathcal{F}$ correspond to measures θ where one of the edges of the $(n-1)$–gon has measure 0. Geometrically, this means that one of the cone points v_i, $i > 0$ has two or more shortest paths on P to v_0. We could cut P open along either of these shortest paths. In $S(P)$, this means one of the "inside" vertices of the star has three or more shortest paths to the tip vertices: two are sides of $S(P)$, and at least one is interior to $S(P)$. You can cut $S(P)$ along such an edge, and rotate one resulting chunk with respect to the other, to obtain a new shape $S'(P)$ with vertices in a permuted order.

To also insist on allowing change of base point requires a further much less direct equivalence relation. If the cone angles $\alpha_1, \dots, \alpha_{n-1}$ are not all the same, then to get all possible cone-metrics, we need one copy of $\mathcal{F}$ for each ordering of the cone angles up to cyclic permutation.

8 Teichmüller space interpretation

Each element of $C(\alpha_1, \dots, \alpha_n)$ determines a point in a certain finite sheeted covering of the modular orbifold for the n–punctured sphere. (The covering corresponds to the subgroup of the modular group for the n–punctured sphere which preserves the cone angles): the map consists of forgetting the metric, and remembering only the conformal structure.

By the uniformization theorem, each of these metrics is equivalent to a metric obtained by deleting n points from the Riemann sphere $\hat{\mathbb{C}}$. The resulting configuration of n points in $\hat{\mathbb{C}}$ is unique up to Möbius transformations.

Proposition 8.1 *The map from $C(\alpha_1, \dots, \alpha_n)$ is a homeomorphism.*

Proof In fact, there is an explicit formula for the inverse map, going from a configuration of n points on $\hat{\mathbb{C}}$ together with the curvatures at those points to a Euclidean cone-manifold with the given conformal structure. The formula is essentially the same as the Schwarz–Christoffel formula for uniformizing a rectilinear polygon. (See [12] for an analysis of these and other cone-manifold structures.)

The idea is to think of the construction of a Euclidean cone metric on $\hat{\mathbb{C}}$ in terms of its developing map h. Consider a collection $\{y_i\}$ of points in $\hat{\mathbb{C}}$, with desired curvatures $\{\alpha_i\}$. Let P be the punctured Riemann sphere $\hat{\mathbb{C}} - \{y_i\}$. The developing map h is not uniquely determined on P, and it is only defined on the universal cover $\tilde{P}$, but any two choices differ by a complex affine transformation. Therefore, the *pre-Schwarzian* of h, that is, $S = h''/h'$, is uniquely determined by the metric, and it is defined on P, not just on the universal cover of P. The Euclidean metric can be easily reconstructed if we are given S, because once we choose an initial value and derivative for the developing map h at one point on $\tilde{P}$, we can integrate the differential equation $h'' = Sh'$ to determine it everywhere else.

How can we determine S? Consider a cone, with curvature α at the its apex. If a cone is conformally mapped to $\mathbb{C}$ with its apex going to the origin, the developing map is $z \mapsto z^{1-\frac{\alpha}{2\pi}}$. The pre-Schwarzian for this map is $-\frac{\alpha}{2\pi}z^{-1}$. It follows that the pre-Schwarzian for the developing map of any Euclidean cone-metric with a cone point having curvature α will have a pole at the cone point, with residue $-\frac{\alpha}{2\pi}z^{-1}$. Conversely, if the pre-Schwarzian of some function h has a pole of this type at any point in $\hat{\mathbb{C}}$, then h will locally be the developing map for a Euclidean structure with a cone point of angle α. (To see this, observe that the analytic continuation of h around the pole differs by post-composition with an affine transformation. Using this information, one can make a local conformal change of coordinates in the domain so that h has the form $z \mapsto z^{1-\frac{\alpha}{2\pi}}$, where α is not necessarily real. From this, one sees that the pre-Schwarzian has a pole at the singularity with residue $\alpha/2\pi$.)

We may as well assume that the the $\{y_i\}$ are in the finite part of $\hat{\mathbb{C}}$. Define

$$S = \sum_i -\frac{\alpha}{2\pi}(z - y^{-1}).$$

Computation shows that in a coordinate patch $w = z^{-1}$ for a neighborhood of ∞, the pre-Schwarzian in terms of the variable w is holomorphic if and only if S behaves asymptotically like $-2z^{-1}$. This is satisfied in our case, since the sum of the α_i is 4π. The condition that S is holomorphic on P, and that it has the given behaviour at the cone points and at ∞, uniquely determines S.

S determines a complex affine structure on P. Since the fundamental group of P is generated by loops going around the punctures, and since the holonomy around these loops is isometric, the affine structure is compatible with a Euclidean structure, well-defined up to scaling. □

Thus, we may think of $C(\alpha_1, \ldots, \alpha_n)$ as a certain geometric interpretation of modular space. The completions $\bar{C}(\alpha_1, \ldots, \alpha_n)$ have a topology which depends on the comparisons of sums of subsets of the α_i with 2π. It is almost never agrees with the standard compactification of the modular space. However, there are only a finite number of possible possibilities for the topology — it is curious that we thus obtain several parameter families of complex hyperbolic structures on the Teichmüller space, and several parameter families of complex hyperbolic cone-manifolds on the various $\bar{C}(\alpha_1, \ldots, \alpha_n)$, with varying cone angles.

Is there any similar phenomenon for the Teichmüller spaces of other surfaces, particularly closed surfaces? The surface of genus 2 has the same modular space as the six–punctured sphere, so for that particular case, the construction carries over. I don't know how to extend it to surfaces of higher genus.

References

[1] **D Cooper**, **W P Thurston**, *Triangulating* 3*–manifolds using* 5 *vertex links*, Topology, 27 (1988) 23–25

[2] **P Deligne**, **G D Mostow**, *Monodromy of hypergeometric functions and non-lattice integral monodromy*, Inst. Hautes Études Sci. Publ. Math. 63 (1986) 5–89

[3] **P Deligne**, **G D Mostow**, *Commensurabilities among lattices in* $\mathrm{PU}(1, n)$, Annals of Mathematics Studies, 132, Princeton University Press, Princeton, NJ (1993)

[4] **D B A Epstein**, **A Marden**, *Convex hulls in hyperbolic space, a theorem of Sullivan, and measured pleated surfaces*, from: "Analytical and geometric aspects of hyperbolic space (Coventry/Durham, 1984)", Cambridge Univ. Press, Cambridge (1987) 113–253

[5] **G D Mostow**, *Generalized Picard lattices arising from half-integral conditions* Inst. Hautes Études Sci. Publ. Math. 63 (1986) 91–106

[6] **E Picard**, *Sur les fonctions hyperfuchsiennes provenant des séries hypergéometriques de deux variables*, Ann. ENS III, 2 (1885) 357–384

[7] **E Picard**, *Sur une extension aux fonctions de deux variables du problème de Riemann relatif aux fonctions hypergéometriques*, Bull. Soc. Math. Fr. 15 (1887) 148–152

[8] **Igor Rivin**, *Euclidean structures on simplicial surfaces and hyperbolic volume*, Annals of Math. 139 (1994) 553–580

[9] **Igor Rivin**, *A characterization of ideal polyhedra in hyperbolic* 3*–space*, Annals of Math. 143 (1996) 51–70

[10] **Igor Rivin**, **Craig D Hodgson**, *A characterization of compact convex polyhedra in hyperbolic* 3*–space*, Invent. Math. 111 (1993) 77–111

[11] **William P Thurston**, *Geometry and Topology of Three–Manifolds*, Princeton lecture notes (1979) `http://www.msri.org/publications/books/gt3m`

[12] **Marc Troyanov**, *Prescribing curvature on compact surfaces with conical singularities*, Trans. Amer. Math. Soc. 324 (1991) 793–821

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

Email: `wpt@math.ucdavis.edu`

Received: 15 November 1997 Revised: 27 November 1998

Appendix: 94 orbifolds

We give below a list of the examples of the spaces $C(\alpha_1, \ldots, \alpha_n)$ which are orbifolds, for $n \geq 5$. When $n = 3$ there is only one example for each feasible triple of cone angles, and for $n = 4$ there are infinitely many examples. In fact, every triangle group in the hyperbolic plane can be interpreted as the modular space for families of tetrahedra. In general, the α_i are of the form $\frac{2\pi p}{q}$, for p and q integers. For each example, we list the least denominator q and the sequence of numerators p_i. We also list the degree of the number field containing the roots of unity $\exp(\frac{2\pi p_i}{q})$ (that is, the number of integers less than q relatively prime to q). We list also three additional bits of information:

(arithmetic) Is the orbifold arithmetic (AR) or non-arithmetic (NR)?

(pure) Is the completion of the covering of the modular space whose fundamental group is the pure braid group an orbifold (P), or are some interchanges of pairs of cone points needed to make the orbifold (I)?

(compact) Is the orbifold compact (C) or non-compact (N)?

The question of arithmeticity hinges on the signatures of the Hermitian forms obtained when we conjugate the curvatures at the cone points (considered as roots of unity) by the Galois automorphisms. If all the other signatures are negative definite or positive definite, the group is arithmetic; otherwise not. The other two questions are more obvious.

These examples were enumerated by a routine computer program, which checks all possibilities having a given least common denominator q. The enumeration was not rigorously verified (even though it should not be hard to do so and search more 'intelligently' at the same time) but was a simple check of all denominators through 999 in a few minutes of computer time. Mostow has rigorously enumerated examples by hand, so this table can be regarded as just a check.

	Denominator	Numerators												degree	arithmetic?	pure?	compact?
1.	3	1	1	1	1	1	1							2	AR	P	N
2.	3	2	1	1	1	1								2	AR	P	N
3.	4	1	1	1	1	1	1	1	1					2	AR	P	N
4.	4	2	1	1	1	1	1	1						2	AR	P	N
5.	4	3	1	1	1	1	1							2	AR	P	N
6.	4	2	2	1	1	1	1							2	AR	P	N
7.	4	3	2	1	1	1								2	AR	P	N
8.	4	2	2	2	1	1								2	AR	P	N
9.	5	2	2	2	2	2								4	AR	P	C
10.	6	1	1	1	1	1	1	1	1	1	1	1	1	2	AR	I	N
11.	6	2	1	1	1	1	1	1	1	1	1	1		2	AR	I	N
12.	6	3	1	1	1	1	1	1	1	1	1			2	AR	I	N
13.	6	2	2	1	1	1	1	1	1	1	1			2	AR	I	N
14.	6	4	1	1	1	1	1	1	1	1				2	AR	I	N
15.	6	3	2	1	1	1	1	1	1	1				2	AR	I	N
16.	6	5	1	1	1	1	1	1	1					2	AR	I	N
17.	6	2	2	2	1	1	1	1	1	1				2	AR	I	N
18.	6	4	2	1	1	1	1	1	1					2	AR	I	N
19.	6	3	3	1	1	1	1	1	1					2	AR	I	N
20.	6	3	2	2	1	1	1	1	1					2	AR	I	N
21.	6	5	2	1	1	1	1	1						2	AR	I	N
22.	6	4	3	1	1	1	1	1						2	AR	I	N
23.	6	2	2	2	2	1	1	1	1					2	AR	I	N
24.	6	4	2	2	1	1	1	1						2	AR	I	N
25.	6	3	3	2	1	1	1	1						2	AR	I	N
26.	6	5	3	1	1	1	1							2	AR	I	N
27.	6	4	4	1	1	1	1							2	AR	I	N
28.	6	3	2	2	2	1	1	1						2	AR	I	N
29.	6	5	2	2	1	1	1							2	AR	I	N
30.	6	4	3	2	1	1	1							2	AR	I	N
31.	6	3	3	3	1	1	1							2	AR	I	N
32.	6	5	4	1	1	1								2	AR	I	N
33.	6	2	2	2	2	2	1	1						2	AR	I	N
34.	6	4	2	2	2	1	1							2	AR	I	N
35.	6	3	3	2	2	1	1							2	AR	I	N
36.	6	5	3	2	1	1								2	AR	I	N
37.	6	4	4	2	1	1								2	AR	I	N
38.	6	4	3	3	1	1								2	AR	I	N
39.	6	3	2	2	2	2	1							2	AR	P	N
40.	6	5	2	2	2	1								2	AR	P	N
41.	6	4	3	2	2	1								2	AR	P	N
42.	6	3	3	3	2	1								2	AR	P	N
43.	6	3	3	2	2	2								2	AR	P	N
44.	8	3	3	3	3	3	1							4	AR	P	C
45.	8	6	3	3	3	1								4	AR	P	C
46.	8	5	5	2	2	2								4	AR	P	C
47.	8	4	3	3	3	3								4	AR	P	C

	Denominator	Numerators							degree	arithmetic?	pure?	compact?
48.	9	4	4	4	4	2			6	AR	P	C
49.	10	7	4	4	4	1			4	AR	P	C
50.	10	3	3	3	3	3	3	2	4	AR	I	C
51.	10	6	3	3	3	3	2		4	AR	I	C
52.	10	9	3	3	3	2			4	AR	I	C
53.	10	6	6	3	3	2			4	AR	I	C
54.	10	5	3	3	3	3	3		4	AR	I	C
55.	10	8	3	3	3	3			4	AR	I	C
56.	10	6	5	3	3	3			4	AR	I	C
57.	12	8	5	5	5	1			4	AR	P	C
58.	12	7	7	2	2	2	2	2	4	AR	I	C
59.	12	9	7	2	2	2	2		4	AR	I	C
60.	12	7	7	4	2	2	2		4	AR	I	C
61.	12	11	7	2	2	2			4	AR	I	C
62.	12	9	9	2	2	2			4	AR	I	C
63.	12	9	7	4	2	2			4	AR	I	C
64.	12	7	7	6	2	2			4	AR	I	C
65.	12	7	7	4	4	2			4	AR	P	C
66.	12	7	5	3	3	3	3		4	NA	P	N
67.	12	5	5	5	3	3	3		4	AR	P	C
68.	12	10	5	3	3	3			4	AR	P	C
69.	12	8	7	3	3	3			4	NA	P	C
70.	12	8	5	5	3	3			4	AR	P	C
71.	12	7	6	5	3	3			4	NA	P	N
72.	12	6	5	5	5	3			4	AR	P	C
73.	12	7	5	4	4	4			4	NA	P	N
74.	12	6	5	5	4	4			4	NA	P	C
75.	12	5	5	5	5	4			4	AR	P	C
76.	14	11	5	5	5	2			6	AR	I	C
77.	14	8	5	5	5	5			6	AR	I	C
78.	15	8	6	6	6	4			8	NA	P	C
79.	18	11	8	8	8	1			6	AR	P	C
80.	18	13	7	7	7	2			6	NA	I	C
81.	18	10	10	7	7	2			6	AR	I	C
82.	18	14	13	3	3	3			6	AR	I	C
83.	18	10	7	7	7	5			6	AR	I	C
84.	18	8	7	7	7	7			6	NA	I	C
85.	20	14	11	5	5	5			8	NA	P	C
86.	20	13	9	6	6	6			8	NA	I	C
87.	20	10	9	9	6	6			8	NA	I	C
88.	24	19	17	4	4	4			8	NA	I	C
89.	24	14	9	9	9	7			8	NA	P	C
90.	30	26	19	5	5	5			8	AR	I	C
91.	30	23	22	5	5	5			8	NA	I	C
92.	30	22	11	9	9	9			8	AR	I	C
93.	42	34	29	7	7	7			12	NA	I	C
94.	42	26	15	15	15	13			12	NA	I	C

Geometry & Topology is a fully refereed international journal dealing with all aspects of geometry and topology. It is intended to provide free electronic dissemination of high quality research. Paper copy is also available. The *Geometry & Topology Monographs* series is intended to provide a similar forum for conference proceedings and research monographs.

Submission must be made electronically. For full instructions visit our web site at:
`http://www.maths.warwick.ac.uk/gt/`
or choose your nearest mirror site from the EMIS network: `http://www.emis.de/`
or use anonymous ftp to: `ftp.maths.warwick.ac.uk`

Geometry and Topology
Mathematics Institute
University of Warwick
Coventry, CV4 7AL, UK

email: `gt@maths.warwick.ac.uk` fax: +44-1203-524182

ISSN 1464-8997 **551**

Geometry & Topology Monographs
Volume 1: The Epstein Birthday Schrift
Pages 551–576

Sur les espaces-temps homogènes

Abdelghani Zeghib

Abstract Here, we classify Lie groups acting isometrically on compact Lorentz manifolds, and in particular we describe the geometric structure of compact homogeneous Lorentz manifolds.

AMS Classification 57B30; 57S20

Keywords Lorentz manifold, twisted Heisenberg group, condition $(*)$

1 Introduction

Une variété homogène M est par définition munie d'une action transitive d'un groupe de Lie G, de telle façon que M s'identifie à un quotient G/H où H est le groupe d'isotropie (d'un certain point). Dans la suite on supposera toujours que l'action de G est fidèle.

En général, l'action de G préserve une certaine structure géométrique "rigide" [7]. Les plus belles de ces structures sont certainement les métriques pseudo-riemanniennes. Parmi ces dernières, on distingue "dans l'ordre" le cas riemannien et ensuite le cas lorentzien (i.e. une métrique pseudo-riemannienne de type $(1, n-1)$).

Lorsque $M = G/H$ est une variété riemannienne homogène compacte, G est nécessairement compact (on avait supposé l'action fidèle!). Quant à H, il peut être n'importe quel sous-groupe fermé (pas nécessairement discret) de G.

Il n'en est rien, lorsque M est de type lorentzien (et toujours supposée compacte). Le groupe G peut bien être non-compact, et de plus étant donné G, il n'est pas évident quels sous groupes fermés H peuvent figurer.

Notre but ici est d'essayer de comprendre, comme c'est le cas des métriques riemanniennes, la structure des variétés lorentziennes homogènes, ayant un volume fini.

1.1 Exemples

1.1.1 Cas semi-simple: $SL(2,\mathbf{R})$

Pour G semi-simple, sa forme de Killing détermine une métrique pseudo-riemannienne bi-invariante. Ainsi, elle passe aux quotients G/Γ, pour Γ discret, qui seront de plus munis d'une action à gauche isométrique de G. Cette métrique est lorentzienne exactement lorsque G est localement isomorphe à $SL(2,\mathbf{R})$.

1.1.2 Cas résoluble: Groupes de Heisenberg tordus

La discussion concernant les exemples qui suivent, se trouve en grande partie dans [9]. Il en a été également question dans [7] et [16].

L'algèbre de Heisenberg $\mathcal{HE}_d$ de dimension $2d+1$ est identifiée en tant qu'espace vectoriel à $\mathbf{R} \bigoplus \mathbf{C}^d$. Si Z (resp. $\{e_1, \dots, e_d\}$) est la base canonique de $\mathbf{R}$ (resp. $\mathbf{C}^d$), alors les crochets non nuls sont donnés par: $[e_k, ie_k] = Z$. En d'autres termes, si ω est la forme symplectique canonique sur $\mathbf{C}^d$, $\omega(X,Y) = \langle X, iY\rangle_0$, où $\langle\ ,\ \rangle_0$ est le produit hermitien canonique, alors $[X,Y] = \omega(X,Y)Z$.

Considérons l'algèbre résoluble $\mathcal{HE}_d^t$ (algèbre de Heisenberg tordue canonique) définie en ajoutant un élément extérieur t, vérifiant $[t, e_k] = ie_k, [t, ie_k] = -e_k$, pour $1 \leq k \leq d$ et $[t, Z] = 0$.

Définissons sur $\mathcal{HE}_d^t$, un produit scalaire $\langle\ ,\ \rangle$, par les lois suivantes: $\mathbf{C}^d$ est muni du produit scalaire induit par son produit hermitien canonique $\langle\ ,\ \rangle_0$ et est orthogonal au 2–plan engendré par t et Z. De plus $\langle t, t\rangle = \langle Z, Z\rangle = 0$ et $\langle t, Z\rangle = 1$.

Il est remarquable que ceci est un produit lorentzien (en particulier non dégénéré), qui est $Ad(\mathcal{HE}_d^t)$–invariant! (i.e. pour tout générateur u, ad_u est antisymétrique au sens de $\langle\ ,\ \rangle$).

Notons $\tilde{He}_d^t$ le groupe simplement connexe déterminé par $\mathcal{HE}_d^t$. On remarquera dans la suite qu'il admet bien des réseaux co-compacts. Comme dans le cas semi-simple, les variétés lorentziennes quotients qu'ils déterminent sont donc homogènes, et leurs groupes d'isométries contiennent des quotients de $\tilde{He}_d^t$.

En fait, on le constatera au long de ce texte, ce n'est jamais le groupe $\tilde{He}_d^t$ qui agit (fidèlement), mais un quotient, par un réseau de son centre. Pour l'expliciter, notons $\tilde{He}_d$ le groupe de Heisenberg simplement connexe et He_d son quotient par un réseau (isomorphe à $\mathbf{Z}$) de son centre (ce quotient est unique à conjugaison près). Maintenant quotienter He_d^t par un réseau central, revient à quotienter He_d par le groupe engendré par une puissance entière de $\exp(t)$. On notera He_d^t le quotient obtenu à l'aide du groupe engendré par $\exp(t)$. Tous les autres quotients sont des extensions de He_d^t par des groupes cycliques finis.

En fait on peut définir ces quotients comme produit semi-direct du cercle S^1 par He_d. Le cercle agit par rotation sur le facteur $\mathbf{C}^d$ et trivialement sur le centre $\mathbf{R}$. Le cas de He_d^t correspond à celui où l'action de S^1 est fidèle.

Considérons en général une action par automorphismes de S^1 sur l'algèbre de Heisenberg $\mathcal{HE}_d$. Soit $\exp(s2\pi R)$ le groupe à un paramètre d'automorphismes ainsi détérminé sur le quotient de $\mathcal{HE}_d$ par son centre, identifié à $\mathbf{C}^d$. Il préserve la forme symplectique canonique ω sur $\mathbf{C}^d$. Mais un groupe compact de transformations symplectiques de $\mathbf{C}^d$ est conjugué à un sous-groupe de $U(d)$. Il s'ensuit que (après conjugaison) R est une application $\mathbf{C}$–linéaire diagonale (dans une base orthonormée) à valeurs propres $\lambda_1 i, \ldots, \lambda_k i$, où les λ_j sont des nombres entiers (car $\exp(2\pi R) = 1$).

Definition 1.1 *Groupes de Heisenberg tordus* On appelera groupe de Heisenberg tordu tout produit semi-direct du cercle S^1 par He_d tel les entiers λ_j soient tous non nuls et de même signe (en d'autres termes les produits de valeurs propres de R sont tous non nulles et de même signe. IL est également équivalent à dire que l'application $\mathbf{C}$ linéaire symétrique iR admet des valeurs propres (réelles) non nulles de même signe).

Evidemment pour $d = 1$, on n'obtient rien d'autre que les extensions cycliques finis de He_1^t. Ces groupes peuvent en fait se définir autrement comme extensions centrales non triviales du groupe des déplacements directs du plan euclidien (appelé parfois groupe d'Euclide) par le cercle S^1.

Remarque terminologique 1.2 La terminologie ci-dessus n'est certainement pas idéale. En effet il existe, au moins pour $d = 1$, des terminologies concurentes. Par exemple, en physique, un groupe de Heisenberg tordu (pour $d = 1$) est dit groupe oscillateur [11], et dans un autre domaine d'intérêt, il s'appele groupe diamant. Apparemment, le terme, groupe de Heisenberg tordu, contient plus d'informations mathématiques.

Une variété d'exemples de variétés lorentziennes homogènes compactes s'obtient à partir de:

Proposition 1.3 (i) *Un groupe de Heisenberg tordu admet une métrique lorentzienne bi-invariante. Réciproquement si une algèbre de Lie obtenu comme produit semi-direct de* S^1 *par* $\mathcal{HE}_d$*, admet une métrique lorentzienne bi-invariante, alors cette algèbre est l'algèbre de Lie d'un groupe de Heisenberg tordu.*

(ii) *A indice fini près, il y a équivalence entre les réseaux d'un groupe de Heisenberg tordu de dimension* $2d+2$ *et ceux du sous-groupe* He_d*, ainsi que ceux de* $\tilde{He}_d$ *(le groupe de Heisenberg simplement connexe de dimension* $2d+1$*).*

Preuve (i) Cherchons les conditions que doit vérifier une telle métrique $\langle\ ,\ \rangle$. D'abord la $Ad(\mathcal{HE}_d)$ invariance de $\langle\ ,\ \rangle$ restreinte à $\mathcal{HE}_d$ entraîne que cette restriction est positive, à noyau exactement le centre.

Les conditions de $Ad(\mathcal{HE}_d)$ invariance de $\langle\ ,\ \rangle$ elle même (i.e. non restreinte) sont beaucoup plus fortes. En effet, on peut supposer que $R = ad_t$ préserve $\mathbf{C}^d$ et considérons X, Y deux éléments de $\mathbf{C}^d$. Ecrivons la condition d'antisymétrie: $\langle ad_X t, Y\rangle + \langle t, ad_X Y\rangle = 0$. Donc: $\langle RX, Y\rangle = \langle t, Z\rangle\omega(X, Y)$ (où Z engendre le centre). Nécessairement, $\langle t, Z\rangle \neq 0$, car sinon $\langle\ ,\ \rangle$ admettra un noyau non trivial contenant Z.

On voit ainsi apparaître la condition sur les valeurs propres de R car la restriction de $\langle\ ,\ \rangle$ à $\mathbf{C}^d$ est définie positive. Si elle est satisfaite, on définira la métrique sur $\mathbf{C}^d$ par $\langle X, Y\rangle = \alpha\omega(X, R^{-1}Y)$, où $\alpha = \langle t, Z\rangle$ est une constante non nulle assurant que la métrique ainsi obtenue est positive (sur $\mathbf{C}^d$).

On vérifie alors que R restreinte à $\mathbf{C}^d$ est antisymétrique. Pour que R (non restreinte) soit antisymétrique, il suffit que la condition suivante se réalise: $\langle ad_t t, X\rangle + \langle t, ad_T X\rangle = 0$, i.e. $\langle t, RX\rangle = 0$ pour tout $X \in \mathbf{C}^d$. Il résulte de la bijectivité de R sur $\mathbf{C}^d$ que t est orthogonal à $\mathbf{C}^d$. Enfin, on choisit: $\langle t, t\rangle = \beta$, un nombre réel quelconque. La métrique est ainsi complétement définie, avec deux paramètres de choix, α et β.

(ii) Soit G un groupe de Heisenberg tordu, obtenu comme produit semi-direct de S^1 par He_d. Ainsi He_d est co-compact dans G, en particulier un réseau de He_d est aussi un réseau dans G. La proposition signifie que réciproquement un réseau de G coupe He_d en un réseau et aussi qu'un réseau de $\tilde{He}_d$ se projette sur un réseau de He_d. Ce sont deux faits standard de la théorie des goupes discrets dont on peut extraire une preuve de [10] (par exemple le premier fait découle d'un énoncé général affirmant qu'un réseau d'un groupe de Lie résoluble coupe le nilradical en un réseau de ce dernier). □

Ainsi, concrètement, comme dans le cas précédent de $SL(2, \mathbf{R})$, les réseaux des groupes de Heisenberg (simplement connexes), qu'on comprend parfaitement, permettent de construire des variétés lorentziennes compactes homogènes dont le groupe d'isométries est (essentiellement) un groupe de Heisenberg tordu.

Remarquons cependant que si l'on quotiente un groupe de Heisenberg tordu par un réseau Γ contenu (pas seulement à indice fini près) dans He_d, alors on n'aura besoin que de l' $Ad(\Gamma)$–invariance de $\langle\ ,\ \rangle$. Par Zariski densité des réseaux de He_d, ceci équivaut au fait que $\langle\ ,\ \rangle$ est $ad(\mathcal{HE}_d)$–invariante.

Definition 1.4 On dira qu'une métrique lorentzienne sur l'algèbre de Lie d'un groupe de Heisenberg tordu, est essentiellement bi-invariante, si elle est $ad(\mathcal{HE}_d)$–invariante.

Remarque 1.5 D'après la preuve ci-dessus, une métrique essentiellement bi-invariante vérifie les mêmes conditions qu'une métrique bi-invariante, sauf celle de l'othogonalité de t à $\mathbf{C}^d$. Une telle métrique dépend donc des deux paramètres, α et β, ainsi que $2d$ autres paramètres donnant le produit de t avec les élements d'une ($\mathbf{R}$–) base de $\mathbf{C}^d$.

1.2 Classification

Notons que malgré son importance (du moins mathématique), en dehors des exemples de [9] signalés ci-dessus, le seul résultat sensible connu au sujet des variétés lorentziennes homogènes, est celui de [8], affirmant que les variétés lorentziennes homogènes compactes (ou plus généralement pseudo-riemanniennes) sont géodésiquement complètes. On peut aussi noter la classification par [12] des variétés lorentziennes homogènes à courbure constante, mais pas nécessiarement compactes, ainsi que le résultat de [5] affirmant (entre autres) qu'une variété lorentzienne homogène compacte et simplement connexe est de type riemannien. (On reviendra plus loin au cas non homogène, où on citera surtout [16] et [7]).

Le but de cet article est de montrer que les exemples précédents sont essentiellement les seuls:

Théorème 1.6 *Un espace-temps homogène, de volume fini, qui n'est pas de type riemannien, admet un sous-groupe normal co-compact dans son groupe d'isométries général, qui est soit un revêtement fini de $PSL(2,\mathbf{R})$, soit un groupe de Heisenberg tordu. L'algèbre de Lie de ce sous-groupe est en fait un facteur direct dans l'algèbre de tous les champs de Killing. De plus ce sous-groupe agit localement librement.*

Ce résultat nous permet entre autres de répondre à la question qu'on s'était posée précédemment: si $M = G/H$, alors quels sous-groupes d'isotropie H peuvent figurer? Il découle du théorème précèdent que H est essentiellement discret au sens que sa composante neutre est compacte. Le résultat suivant explicite complètement la structure topologique et géométrique des variétés lorentziennes homogènes.

Théorème 1.7 (Classification) *Soit M une variété lorentzienne homogène de volume fini. Supposons que M n'est pas de type riemannien (i.e. à groupe d'isométries compact). Alors:*

(i) *ou bien $Isom(M)$ contient un revêtement fini de $PSL(2, R)$. Dans ce cas M admet un revêtement isométrique $\tilde{M}$ qui est un produit métrique de $\widetilde{SL(2, \mathbf{R})}$ muni de sa forme de Killing, par une variété riemannienne homogène compacte.*

(ii) *ou bien $Isom(M)$ contient S un groupe de Heisenberg tordu. Dans ce cas M admet un revêtement $\tilde{M}$ qui se construit de la façon suivante. Il existe une variété riemannienne homogène compacte (L, r), munie d'une action isométrique localement libre de S^1. Le cercle S^1 isomorphe au centre de S, y agit par translation et agit par suite diagonalement sur $S \times L$, muni de la métrique produit de celle de L par une métrique lorentzienne essentiellement bi-invariante sur S. Alors le revêtement $\tilde{M}$ est le quotient $S \times_{S^1} L$ de cette action. Il est muni de la métrique déduite par projection, de la métrique induite sur $TS \oplus \mathcal{O}$, où $\mathcal{O}$ est la distribution orthogonale à l'action de S^1 sur L.*

En fait $M = \tilde{M}/\Gamma$, où Γ est le graphe d'un homomorphisme ρ d'un réseau co-compact Γ_0 de S dans $Isom_{S^1}(L)$, le groupe d'isométries de L respectant l'action de S^1. De plus le centralisateur de $\rho(\Gamma_0)$ dans $Isom_{S^1}(L)$ agit transitivement sur L.

On peut par exemple prendre pour L la sphère S^3 munie d'une fibration de Hopf. Le groupe d'isométries qui la préserve est isomorphe à $S^1 \times S^3$. On prendra pour ρ un homomorphisme d'un réseau de S à valeurs dans S^1 (ce qui assurera que le centralisateur de l'image de ρ agit transitivement sur S^3). Le groupe d'isométries de la variété lorentzienne homogène compacte ainsi obtenue, sera essentiellement $S^3 \times S^1 \times S$.

Remarque 1.8 On déduit du théorème de structure ci-dessus qu'on peut changer la métrique tout en la gardant homogène, de telle façon que la métrique sur S soit bi-invariante.

Il est bien connu que sur $SL(2, \mathbf{R})$, la forme de Killing est à une constante près, la seule métrique bi-invariante. Sur un groupe de Heisenberg tordu S, il y en a beaucoup, mais elles sont toutes isométriques (pas seulement conformes!) par automorphismes dans le revêtement universel $\tilde{S}$. Ceci est lié au fait (remarquable) qu'une structure Lorentzienne bi-invariante donnée sur $\tilde{S}$, admet des transformations conformes non triviales. Elles sont en fait des homothéties, i.e. à distorsion constante.

Ainsi sur S toutes les métriques bi-invariantes sont localement isométriques. Cependant il y a un module de dimension 2 (les paramètres α et β de la preuve précédente) de telles structures globales (voir 5.6).

1.3 Ingrédients de la preuve

La finitude du volume sera utilisée, comme d'habitude, pour en déduire des propriétés de récurrence de l'action de G. Mais le plus grand intérêt de cette hypothèse pour nous ici, c'est de permettre de construire un produit scalaire L^2, sur l'algèbre de Lie de G, ayant la propriété élémentaire mais remarquable d'être bi-invariant.

En effet, plus généralement, si G est un groupe de Lie agissant sur une variété M (pas nécessairement transitivement) en préservant une métrique pseudo-riemannienne $\langle\ ,\ \rangle$, alors la forme $\kappa(X,Y) = \int_M \langle X(x), Y(x)\rangle dx$ détermine une forme bilinéaire sur l'algèbre de Lie $\mathcal{G}$, qui est bi-invariante. Pour le voir, il suffit de remarquer que si ϕ^t est un groupe à paramètre de G (identifié au flot correspondant de M) et X est un élément $\mathcal{G}$ (identifié au champ de vecteurs correspondant sur M), alors $\phi^t_* X = Ad(\phi^t) X$.

Notons qu'il est aussi possible de considérer des formes du type $\kappa(X,Y) = \int_U \langle X(x), Y(x)\rangle dx$, où $U \subset M$ est un sous-ensemble G–invariant quelconque, ou plus généralement en intégrant par rapport à une mesure G–invariante quelconque. Remarquons aussi que la même construction marche lorsque G préserve un tenseur quelconque sur M, et permet ainsi de construire un tenseur "analogue" bi-invariant sur $\mathcal{G}$.

Cependant, le résultat obtenu est généralement trivial (même nul!). Ainsi, lorsque G est simple, la forme obtenue est un multiple (souvent nul) de la forme de Killing. On peut par exemple prendre $M = G$, qu'on munit d'une structure pseudo-riemannienne invariante à gauche (elle s'obtient simplement d'un produit scalaire de même signature sur $\mathcal{G}$). Ainsi G agit sur M en respectant cette structure. Lorsque G est compact la forme κ construite sur $\mathcal{G}$

sera définie positive, définie négative ou nulle, quelle que soit la signature de la structure pseudo-riemannienne de départ.

1.4 Cas lorentzien

Dans notre cas lorentzien, la forme κ sera suffisamment non triviale dès qu'il existe des champs $X \in G$ tels que $\langle X(x), X(x)\rangle$ garde un signe constant. Il se trouve, comme cela était établi dans [13] que c'est effectivement le cas pour tout champ X engendrant un groupe à paramètre ϕ^t non précompact, i.e. la fermeture dans G de $\{\phi^t / t \in \mathbf{R}\}$ n'est pas compact. C'est à ce niveau là qu'on utilise l'aspect dynamique de la finitude du volume. En fait ona:

Proposition fondamentale 1.9 [13] *Soit $(M, \langle\ ,\ \rangle)$ une variété lorentzienne de volume fini. Soit X un champ de Killing sur M, déterminant un flot non précompact, alors: $\langle X(x), X(x)\rangle \geq 0$ pour tout $x \in M$. On dira que X est (partout) non temporel.*

On en déduit ce fait, qui n'entraîne pas tout à fait que κ est lorentzienne, exactement comme $\langle\ ,\ \rangle$, mais en borne la dégénérescence:

Proposition 1.10 Condition $(*)$ *Soit $\mathcal{P}$ un sous-espace vectoriel de champs de Killing tel pour l'ensemble des éléments $X \in \mathcal{P}$ déterminant des flots non précompacts, est dense dans $\mathcal{P}$. Alors la forme κ est positive sur $\mathcal{P}$ et son noyau est au plus de dimension 1 (ou en d'autres termes, l'ensemble des vecteurs isotropes de $\mathcal{P}$ est un sous-espace vectoriel de dimension ≤ 1).*

1.5 Un résultat algébrique

Il se trouve que les données algébriques, fournies par κ, vérifiant la proposition précédente, suffisent largement pour comprendre le groupe G:

Théorème algébrique 1.11 *Soit G un groupe de Lie non compact dont l'algèbre de Lie $\mathcal{G}$ est munie d'une forme bi-invariante κ, vérifiant l'hypothèse de non dégénérescence $(*)$ suivante:*

Condition $(*)$ *Si $\mathcal{P}$ est un sous-espace vectoriel de $\mathcal{G}$, tel que l'ensemble des éléments $X \in \mathcal{P}$ déterminant des groupes à paramètre non précompacts est dense dans $\mathcal{P}$; alors la forme κ est positive sur $\mathcal{P}$ et son noyau est au plus de dimension 1.*

Alors $\mathcal{G}$ s'écrit comme somme directe orthogonale d'algèbres: $\mathcal{G} = \mathcal{K}+\mathcal{A}+\mathcal{S}$, où: $\mathcal{K}$ est une algèbre compacte (i.e. l'algèbre de Lie d'un groupe de Lie semi-simple compact), $\mathcal{A}$ est une algèbre abélienne, et $\mathcal{S}$ est soit triviale, soit $sl(2, \mathbf{R})$, soit l'algèbre de Lie du groupe affine (des transformations de la droite), soit une algèbre de Heisenberg $\mathcal{HE}_d$, soit une algèbre de Heisenberg tordue. On a:

(i) *Lorsque $\mathcal{S}$ est triviale, κ est positive à noyau de dimension ≤ 1. Lorsque $\mathcal{S}$ est non triviale, κ est définie positive sur $\mathcal{K}$ et $\mathcal{A}$.*

(i) *Lorsque $\mathcal{S}$ est l'algèbre de Lie du groupe affine, κ est positive dégénérée sur $\mathcal{S}$ et admet pour noyau exactement l'idéal déterminé par les translations.*

(i) *Lorsque $\mathcal{S}$ est une algèbre de Heisenberg, κ est positive dégénérée sur $\mathcal{S}$ et admet pour noyau exactement le centre.*

(iv) *Lorsque $\mathcal{S}$ est une algèbre de Heisenberg tordue, la forme κ est lorentzienne sur $\mathcal{S}$. Le sous-groupe de G déterminé par $\mathcal{S}$ est un groupe de Heisenberg tordu. De plus le sous-groupe abelien déterminé par $\mathcal{A}+\mathcal{Z}$, où $\mathcal{Z}$ est le centre de $\mathcal{S}$, est compact.*

(v) *Lorsque $\mathcal{S} = sl(2, \mathbf{R})$, la forme κ sur $\mathcal{S}$ est lorentzienne et le sous-groupe déterminé par $\mathcal{S}$ est un revêtement fini de $PSL(2, \mathbf{R})$. De plus le sous-groupe déterminé par $\mathcal{A}$ est compact.*

1.6 Un résultat géométrique

Le théorème algébrique s'applique en particulier aux groupes de Lie connexes non compact agissant isométriquement sur une variété Lorentzienne $(M, \langle\ ,\ \rangle)$ de volume fini. Certaines parties de ce théorème sont dues dans ce cas à [16] et ensuite [7]. Plus précisément, la structure algébrique de $\mathcal{G}$ est élucidée dans [16] lorsque $\mathcal{G}$ contient $SL(2, \mathbf{R})$. Il y a été également démontré que le nilradical est de degré de nilpotence ≤ 2.

Dans [7], il a été question d'améliorations géométriques de résultats de [16] (surtout dans le cas analytique). En effet, on peut, en général, améliorer le théorème algébrique précédent, par un résultat géométrique, ponctuel. Il exprime essentiellement le fait que si un champ de Killing X est non temporel au sens de κ (i.e. $\kappa(X, X) \geq 0$), c'est qu'il l'est ponctuellement au sens de $\langle\ ,\ \rangle$ (i.e. $\langle X(x), X(x)\rangle \geq 0$, pour tout $x \in M$).

Tout ce qui concerne $SL(2, \mathbf{R})$ dans les résultats suivants est démontré par [7]. Notre approche ici permet de les redémontrer.

Théorème géométrique 1.12 *Soit G un groupe de Lie connexe non compact agissant isométriquement sur une variété lorentzienne $(M, \langle\ ,\ \rangle)$ de volume fini. Notons κ la forme ainsi définie sur $\mathcal{G}$.*

1) *Supposons que κ est positive, alors les orbites de G sont non temporelles (i.e. la restriction de $\langle\ ,\ \rangle$ à ces orbites est ≥ 0). Le noyau de κ, s'il n'est pas trivial est un champ de Killing (partout) de type lumière (au sens de $\langle\ ,\ \rangle$) et à orbites géodésiques.*

2) *Supposons que κ n'est pas positive, donc $\mathcal{G}$ contient un facteur direct $\mathcal{S}$, isomorphe à $sl(2, \mathbf{R})$ ou algèbre de Heisenberg tordue. Alors l'action de $\mathcal{S}$ est partout localement libre.*

Le résultat de [7] pour $sl(2, \mathbf{R})$ est plus précis. Il affirme davantage que la distribution orthogonale (aux orbites) est intégrable (et aussi géodésique). Il s'ensuit qu' un certain revêtement est un produit tordu d'une variété riemannienne par $\widetilde{SL(2, \mathbf{R})}$.

En fait lorsqu'un groupe isomorphe à $\widetilde{SL(2, \mathbf{R})}$ ou à un groupe de Heisenberg tordu agit isométriquement sur une variété lorentzienne de volume fini, alors celle ci s'obtient pratiquement de la même façon que dans le cas homogène, explicité au théorème 1.7, à ceci près que L ne sera supposée ni homogène ni compacte:

Théorème 1.13 [7] *Une variété lorentzienne de volume fini munie d'une action isométrique d'un groupe localement isomorphe à $SL(2, \mathbf{R})$ est revêtue par un produit de $\widetilde{SL(2, \mathbf{R})}$ par une variété riemannienne (L, r), muni d'une métrique tordue $h_{(g,x)} = f(x) k \otimes r_x$, où f est une fonction positive sur L et k est la forme de Killing de $\widetilde{SL(2, \mathbf{R})}$.*

Ici on a un résultat de structure, un peu plus compliqué, dans le cas d'un groupe de Heisenberg tordu G , du fait que la distribution orthogonale n'est pas nécessairement intégrable. C'est en fait sa saturée par le centre de G qui l'est.

La construction est la suivante. Soit (L, r) une variété riemannienne munie d'une action isométrique localement libre de S^1. Notons $\mathcal{O}$ la distribution orthogonale aux orbites.

Soit $\mathcal{M}$ l'espace des métriques lorentziennes essentiellement bi-invariantes sur $\mathcal{G}$ (1.4). Soit ϕ: $L \to \mathcal{M}$ une application (de même classe de régularité que toutes les données). Munissons le produit $G \times L$ de la métrique tordue définie par ϕ: $h_{(g,x)} = m_x \otimes r_x$ (l'espace tangent au facteur G étant partout identifié à $\mathcal{G}$).

Le centre de G, isomorphe à S^1, y agit isométriquement par translation. On a donc une action isométrique diagonale de S^1 sur le produit $G \times L$. Notons $G \times_{S^1} L$ le quotient et munissons le de la métrique déduite par projection, de la métrique induite sur l'horisontal $\mathcal{G} \oplus \mathcal{O}$

Soit Γ un réseau de $G \times Isom_{S^1}(L)$ où $Isom_{S^1}(L)$ désigne le groupe d'isométries de L préservant l'action de S^1. On suppose que Γ agit sans point fixe sur $G \times_{S^1} L$ (ce qui sera toujours vrai pour un sous-groupe d'indice fini). Le quotient $M = \Gamma \setminus G \times_{S^1} L$ est une variété lorentzienne de volume fini munie d'une action isométrique de G.

Théorème 1.14 *Toute variété lorentzienne de volume fini, munie d'une action isométrique d'un groupe de Heisenberg tordu G est construite de la façon précédente.*

Exemple 1.15 On peut prendre pour L le groupe G lui même muni d'une métrique riemannienne invariante à droite, et de l'action de son centre. On voit sur cet exemple que $\mathcal{O}$ peut bien être non intégrable. En jouant sur Γ, qui est un réseau de $G \times G$, on peut réaliser diverses propriétés de densité des orbites de G.

La classification des algèbres de Lie de groupes agissant isométriquement sur des variétés lorentziennes compactes, a été démontrée indépendemment par S Adams et G Stuck [1]. Le présent article, ainsi que [1] sont parus simultanément (sous forme de preprints) en Mai 1995. D'autres résultats complémentaires qui précisent cette classification ont été ensuite démontrés dans [2] et [14].

2 La condition $(*)$

Rappelons brièvement dans ce qui suit les éléments de la preuve de 1.9 [13]. Le premier point est que dans un groupe de Lie la fermeture L d'un groupe à un paramètre ϕ^t est soit $\mathbf{R}$ soit un tore (compact). En effet L est un groupe abélien possédant un groupe à un paramètre dense.

Il en découle que si une sous-suite ϕ^{t_i} est précompacte (i.e. équicontinue) alors le flot ϕ^t lui même est précompact.

Le second point est un phénomène d'uniformité valable pour des suites de transformations f_i préservant une connexion. Il stipule que si la suite des dérivés $D_x f_i$ en un point x donné est équicontinue, alors la suite f_i elle même est équicontinue. Ceci découle de la définition même de la structure différentiable du groupe G d'isométries de la connexion. En effet cette structure est caractérisée par le fait que pour tout repère r_x en x, l'évaluation e: $G \to Rep(M)$, $e(f) = f^*(r_x)$ est un plongement propre.

Le dernier point est qu'au voisinage d'un point x, qu'on peut supposer récurrent, où le champ de Killing X générateur infinitésimal de ϕ^t est de type temps, les applications de retour, ont leurs dérivées équicontinues en x. En effet, ces dérivées respectent la métrique riemannienne (définie au voisinage de x) obtenue canoniquement à partir de la métrique lorentzienne, juste en changeant le signe le long de X. □

La Proposition 1.10 découle du fait suivant:

Lemme 2.1 *Soit $\mathcal{P}$ un sous-espace vectoriel de champs de Killing tel que pour tout $X \in \mathcal{P}$ et $x \in M$, $\langle X(x), X(x)\rangle \geq 0$. Alors la forme κ est positive sur $\mathcal{P}$ et son noyau est au plus de dimension 1.*

Preuve Il découle de l'hypothèse que si $X \in \mathcal{P}$ est isotrope au sens de κ, alors $X(x)$ est isotrope au sens de $\langle\ ,\ \rangle_x$ pour tout x. Donc si $\mathcal{A}$ est un sous-espace isotrope de $\mathcal{P}$, alors: $\mathcal{A}_x = \{X(x), X \in \mathcal{A}\}$ est un sous-espace isotrope de $(T_xM, \langle\ ,\ \rangle_x)$. Il s'ensuit que: $\dim(\mathcal{A}_x) \leq 1$ pour tout x car la métrique $\langle\ ,\ \rangle$ est lorentzienne.

La preuve du lemme sera achevée si l'on montre que deux champs de Killing (partout) colinéaires, sont tels que l'un est multiple constant de l'autre.

En effet soit X et Y deux tels champs et écrivons (localement) $Y = fX$ où f est une certaine fonction. Notons ∇ la connection de Levi-Civita. Alors, un Champ de Killing tel que X vérifie que: pour tout x, l'application $u \in T_xM \to \nabla_u X \in T_xM$ est antisymétrique. Ainsi $0 = \langle \nabla_u(fX), u\rangle = (u.f)\langle X, u\rangle + f\langle \nabla_u X, u\rangle = (u.f)\langle X, u\rangle$, car X et $Y = fX$ sont, tous les deux, deschamps de Killing. Il en découle que f est constante. □

3 Début de la preuve du théorème algébrique

Sans le mentionner, on utilisera parfois, l'affirmation suivante, qui contient des faits classiques standards:

Fait 3.1 *Soit $\mathcal{G}$ une algèbre de Lie muni d'une forme bi-invariante k. On a:*

(i) *Le noyau de k est un idéal de $\mathcal{G}$.*

(ii) *Si k est définie positive, alors $\mathcal{G}$ est somme directe k–orthogonale d'une algèbre abelienne et d'une algèbre compacte (i.e. l'algèbre de Lie d'un groupe de Lie semi-simple compact).*

(iii) *Si $\mathcal{G}$ est compacte, alors k est multiple de sa forme de Killing.*

Soit maintenant $\mathcal{G}$ une algèbre de Lie munie d'une forme κ comme dans le theéorème algébrique.

Lemme 3.2 *Soit $\mathcal{P}$ une sous algèbre abélienne de $\mathcal{G}$ ayant un élément X déterminant un flot non précompact. Alors la forme κ est positive sur $\mathcal{P}$ et son noyau est au plus de dimension 1.*

Preuve On applique la condition $(*)$ sachant que la fermeture du groupe déterminé par $\mathcal{P}$ est un produit d'un tore par un espace vectoriel non trivial. Tous les groupes à un paramètre sont non précompacts sauf ceux tangents au facteur torique. □

Le nilradical Le lemme 3.2 s'étend en fait aux groupes nilpotents grâce à la:

Proposition 3.3 *L'ensemble des groupes à un paramètre non précompacts d'un groupe de Lie nilpotent non compact, est dense. C'est en fait le complémentaire d'un tore maximal (qui est par ailleurs central et unique).*

Preuve Soit N un tel groupe, $\tilde{N}$ son groupe revêtement universel, et Γ le groupe fondamental de N. Alors Γ est central dans $\tilde{N}$. De plus, c'est un réseau dans un unique sous-groupe de Lie (connexe) $\tilde{L}$, également central (pour définir $\tilde{L}$, on se ramène au cas abélien, en remarquant simplement que le centre de $\tilde{N}$ est connexe, car si un élément est central, alors le groupe à paramètre (unique) qui le contient est central). La projection de $\tilde{L}$ dans N est un tore (maximal).

Ainsi N se projette sur $N/L = \tilde{N}/\tilde{L}$, qui est simplement connexe, donc ayant tous ses groupes à un paramètre non précompacts. Il s'ensuit que les groupes à un paramètre de N qui sont précompacts, sont tangents à l'algèbre de Lie de L. □

Notons N le nilradical de G, i.e. le plus grand sous groupe de Lie (connexe) normal nilpotent. On supposera dans cette section qu'il est non compact.

Corollaire 3.4 *Si le nilradical N est non compact, alors la restriction de κ à $\mathcal{N}$ est une forme positive, dont le noyau est un idéal $\mathcal{I}$ de dimension ≤ 1. De plus $\mathcal{N}$ est isomorphe à une somme directe orthogonale d'algèbres $\mathcal{N} = \mathcal{A} + \mathcal{HE}_d$, où $\mathcal{A}$ est abélienne et $\mathcal{HE}_d$ est l'algèbre de Heisenberg de dimension $2d+1$.*

L'action adjointe de G sur $\mathcal{N}/\mathcal{I}$ est à image compacte (car elle préserve une forme définie positive).

Lorsque le facteur correspondant à l'algèbre de Heisenberg est non trivial, le noyau $\mathcal{I}$ de κ est exactement son centre $\mathcal{Z}$.

Preuve On utilise 3.2 pour en déduire que κ est positive sur $\mathcal{N}$ et que $\dim \mathcal{I} \leq 1$. On remarque ensuite que l'algèbre $\mathcal{N}/\mathcal{I}$ est abélienne, car elle est nilpotente et admet une métrique définie positive bi-invariante. □

Remarque 3.5 $\mathcal{A}$ n'est pas canoniquement définie, mais la somme $\mathcal{A}+\mathcal{Z}$ et le facteur de type Heisenberg $\mathcal{HE}_d$ le sont.

Proposition 3.6 (i) *Le centre $\mathcal{Z}$ de $\mathcal{HE}_d \subset \mathcal{N}$ est en fait central dans $\mathcal{G}$.*

(ii) *Tout $X \in \mathcal{HE}_d \subset \mathcal{N}$ non central, engendre un groupe à un paramètre non précompact.*

(iii) *Si Y est un élément non trivial de $\mathcal{G}$ qui commute avec un élément non central de $\mathcal{N}$, alors $\kappa(Y,Y) > 0$.*

Preuve (i) Soit A un automorphisme de $\mathcal{HE}_d$ respectant κ. Supposons que A induit sur $\mathcal{Z}$ une homothétie non triviale. Alors $\mathcal{Z}$ sera le seul sous espace propre associé à une valeur propre de module $\neq 1$, car κ est définie positive sur $\mathcal{HE}_d/\mathcal{Z}$. Il existera donc un supplémentaire $\mathcal{T}$ de $\mathcal{Z}$, sur lequel A respecte une métrique définie positive (et en particulier à valeurs propres de module égale à 1). On obtient une contradiction en considérant deux éléments, X et Y de $\mathcal{T}$, vérifiant $[X,Y] = Z \in \mathcal{Z}$.

(ii) découle du fait qu'alors ad_X est nilpotent (et non trivial) et donc le groupe à un paramètre $exp(tad_X)$ est non précompact.

(iii) En effet si Y commute avec un élément non central $X \in \mathcal{HE}_d$, alors Y, X et $\mathcal{Z}$ déterminent une sous-algèbre abélienne de dimension ≥ 2, vérifiant 3.2. Il s'ensuit que $\mathcal{Z}$ est le seul espace κ isotrope de cette sous-algèbre. □

Proposition 3.7 *Soit $\mathcal{L} \subset \mathcal{G}$ une sous-algèbre semi-simple. Alors la somme $\mathcal{G}' = \mathcal{L} + \mathcal{N}$ est orthogonale (au sens de κ) et directe (au sens d'algèbres)*

Preuve Soit $\mathcal{I} \subset \mathcal{N}$ le noyau de la restriction de κ à $\mathcal{N}$. C'est un idéal de $\mathcal{G}'$ de dimension ≤ 1. Par semi-simplicité, $\mathcal{L}$ centralise $\mathcal{I}$. On peut appliquer le Fait 3.8 pour voir que $\mathcal{I}$ est orthogonale à $\mathcal{L}$ et par suite à $\mathcal{G}'$.

On peut donc en passant au quotient $\mathcal{G}'/\mathcal{I} = \mathcal{L} + \mathcal{N}/\mathcal{I}$, supposer que la forme κ est définie positive sur $\mathcal{N}$.

La proposition est bien connue lorsque κ est définie positive sur $\mathcal{G}'$ (voir 3.1. On va essayer donc de se ramener à cette situation. Par 3.4, l'action adjointe de $\mathcal{L}$ sur $\mathcal{N}$ est à image compacte. On peut donc supposer que $\mathcal{L}$ est compacte. De plus, quitte à traîter facteur par facteur, on peut supposer que $\mathcal{L}$ est simple. Soit k la forme de Killing de $\mathcal{G}'$. Elle est triviale sur $\mathcal{N}$ car c'est le nilradical et sa restriction à $\mathcal{L}$ est un multiple non nul de la forme de Killing de $\mathcal{L}$. Ainsi, sur $\mathcal{L}$, κ est multiple de k. Il s'ensuit qu'il existe un choix d'un réel α tel que $\kappa + \alpha k$ soit définie positive sur $\mathcal{G}'$. Ainsi, on s'est ramené au cas où κ est définie positive sur $\mathcal{G}'$.

Pour montrer l'orthogonalité de la somme $\mathcal{L} + \mathcal{N}$, on utilise le fait général suivant, dont la preuve découle de la bi-invariance de κ. □

Fait 3.8 *Soit $\mathcal{L}$ une sous-algèbre de $\mathcal{G}$, et Y un élément de $\mathcal{G}$ centralisant $\mathcal{L}$. Alors l'application $X \in \mathcal{L} \to \kappa(X, Y) \in \mathbf{R}$ est un homomorphisme, i.e. $\kappa([X, X'], Y) = 0$, pour X, X' dans $\mathcal{L}$.*

Le radical Soit R le radical de G (i.e. le plus grand sous groupe de Lie normal résoluble) et $\mathcal{R}$ son algèbre de Lie. On supposera dans cette section qu'il est non compact. On a d'abord la constatation suivante:

Fait 3.9 *Si R est non compact, alors le nilradical N l'est également.*

Preuve En effet, s' il est compact, N sera central dans G et en particulier dans R. Ainsi l'avant dernier groupe dérivé de R contient strictement N et est nilpotent. Par naturalité, il est normal dans G, ce qui contredit la définition de N. □

La proposition 3.7 se généralise à $\mathcal{R}$:

Proposition 3.10 *Soit $\mathcal{L} \subset \mathcal{G}$ une sous-algèbre semi-simple. Alors la somme $\mathcal{G}' = \mathcal{L} + \mathcal{R}$ est orthogonale (au sens de κ) et directe (au sens d'algèbres)*

Preuve Cela découle de 3.7 et du lemme suivant. □

Lemme 3.11 *Soit A un automorphisme semi-simple de R, trivial sur $\mathcal{N}$, alors A est trivial.*

Preuve Soit $\mathcal{E} \subset \mathcal{R}$ un sous-espace vectoriel supplémentaire de $\mathcal{N}$ invariant par A. Soit $X \in \mathcal{E}$, $Y \in \mathcal{N}$, alors $[X, Y] \in \mathcal{N}$. Donc $[X, Y] = A[X, Y] = [AX, Y]$. Autrement dit $X - AX$ centralise $\mathcal{N}$. Par maximalité de $\mathcal{N}$ en tant que sous-algèbre normale nilpotente, on déduit que l'application $X \in \mathcal{E} \to X - AX \in \mathcal{E}$ est nulle (car son image est contenue dans $\mathcal{E}$). □

Facteur semi-simple

Fait 3.12 *Supposons que R est non compact. Alors on a une décomposition directe et orthogonale $\mathcal{G} = \mathcal{K} + \mathcal{R}$ où $\mathcal{K}$ est une sous-algèbre semi-simple compacte.*

Preuve D'après ce qui précède, il suffit simplement de montrer que le facteur semi simple $\mathcal{K}$ est compacte. Il suffit pour cela d'observer que la restriction de κ à chaque facteur de $\mathcal{K}$ est positive et non triviale. Pour cela on applique la condition $(*)$ à l'algèbre $\mathcal{K}' = \mathcal{K} + \mathbf{R}X$, où X est un élément de $\mathcal{R}$ qui détermine un groupe à un paramètre non précompact. En effet tous les groupes à un paramètre de $\mathcal{K}'$ non tangents à $\mathcal{K}$ sont non précompacts. Ainsi κ est positive sur $\mathcal{K}'$ et à noyau de dimension ≤ 1. Ce noyau intersecte trivialement $\mathcal{K}$, car sinon, il sera un idéal de dimension 1 de $\mathcal{K}$, ce qui contredit son caractère semi-simple. □

4 Preuve du théorème algébrique

Ce qui précède nous amène à distinguer le cas où le radical R est compact du cas où il ne l'est pas.

4.1 Cas où le radical est compact

Le radical étant compact, il est donc abélien et on a une décomposition directe: $\mathcal{G} = \mathcal{L} + \mathcal{R}$, où $\mathcal{L}$ est semi-simple. Une application comme dans la preuve précédente de la condition $(*)$, permet de montrer que κ est définie positive sur $\mathcal{R}$.

Puisque G est non compact, $\mathcal{L}$ contient un facteur (direct) semi-simple $\mathcal{S}$ de type non compact. Ainsi tout facteur de $\mathcal{S}$ contient des vecteurs qui déterminent des groupes à un paramètre non précompacts. Soit $\mathcal{S}_1$ un tel facteur. Alors, une application comme dans la preuve précédente de la condition $(*)$, à tous les autres facteurs de $\mathcal{G}$ (qui centralisent $\mathcal{S}_1$) , permet de montrer que κ est positive, sur chacun de ces facteurs. Il s'ensuit qu'ils sont tous compacts et en particulier, par définition, que $\mathcal{S}$ est simple.

Notons $\mathcal{K}$ le facteur semi-simple compact de $\mathcal{G}$. Le fait 3.8 permet de montrer que la décomposition $\mathcal{G} = \mathcal{S} + \mathcal{K} + \mathcal{R}$ est orthogonale.

Montrons à présent que l'algèbre simple de type non compact $\mathcal{S}$ est isomorphe à $sl(2, \mathbf{R})$.

Il est connu que dans tous les cas $\mathcal{S}$ contient une algèbre $\mathcal{S}'$ isomorphe à $sl(2, \mathbf{R})$. Notons $\mathcal{E}$ l'orthogonal à $\mathcal{S}'$. C'est un supplémentaire de $\mathcal{S}'$ (car ce dernier n'est pas dégénéré) qui est $ad(\mathcal{S}')$–invariant (par bi-invariance de κ).

Il est aussi connu (par algébricité des representations d'algèbres semi-simples) que pour $X \in \mathcal{S}'$, si ad_X est semi-simple (resp. nilpotent) sur $\mathcal{S}'$, alors il en va de même pour ad_X agissant sur $\mathcal{E}$. Il est facile de se convaincre que si tout élément hyperbolique (i.e. semi-simple à valeurs propres réels) $X \in \mathcal{S}'$ agit trivialement sur $\mathcal{E}$, alors toute l'action est triviale, et $\mathcal{S}'$ sera un facteur direct de $\mathcal{S}$, ce qui contredit la simplicité de $\mathcal{S}$.

Par l'absurde, supposons qu'il existe X, un élément hyperbolique agissant non trivialement sur $\mathcal{E}$. Il existe donc un vecteur propre $Z \in \mathcal{E}$ tel que $[X, Z] = \lambda Z$ et $\lambda \neq 0$. Il en découle que Z détermine un groupe à un paramètre non précompact (car sinon ad_Z serait semi-simple à valeurs propres imaginaires pures). Or, il existe $Y \in \mathcal{S}'$ nilpotent vérifiant $[X, Y] = Y$. On en déduit que Z est aussi vecteur propre, nécessairement trivial par nilpotence de ad_Y: $[Y, Z] = 0$. Donc Y et Z engendrent un groupe abélien contenant au moins 2 groupes à un paramètre (différents) non précompacts. Il y en a donc un ensemble dense. Ceci contredit l'hypothèse $(*)$ car Y et Z sont orthogonaux et simultanément isotropes. Ce dernier fait se voit facilement, car $\exp(ad_X)$ induit une homothétie non triviale sur chacune des directions de Y et Z.

Il ne reste à montrer du théorème algébrique dans notre cas (i.e. lorsque R est compact) que le fait que l'action se factorise en l'action, d'un revêtement fini de $PSL(2, \mathbf{R})$, ou de manière équivalente un quotient central infini de $\widetilde{SL(2, \mathbf{R})}$. Il suffit pour cela de remarquer que $\widetilde{SL(2, \mathbf{R})}$ ainsi que ses quotients finis, ont tous leurs groupes à un paramètres non précompacts. Ce qui impliquerait que κ est positive! □

4.2 Cas où R n'est pas compact

On a alors d'après 3.12 une décomposition directe orthogonale $\mathcal{G} = \mathcal{K} + \mathcal{R}$, où $\mathcal{K}$ est compacte. Il suffit donc de montrer que $\mathcal{R}$ se décompose comme énoncé. On peut ainsi à présent oublier $\mathcal{K}$ en supposant que $\mathcal{G}$ est résoluble.

Le nilradical $\mathcal{N}$ est non compact. Considérons la décomposition: $\mathcal{N} = \mathcal{A} + \mathcal{HE}_d$ et notons $\mathcal{Z}$ le centre de $\mathcal{HE}_d$. Rappelons (3.5) que c'est la somme $\mathcal{A} + \mathcal{Z}$ (mais pas $\mathcal{A}$) qui est canoniquement définie.

4.2.1 Cas où κ n'est pas positive. Groupes de Heisenberg tordus

Soit t un élément de $\mathcal{G}$ tel que $\kappa(t,t) < 0$. Il engendre un groupe à un paramètre non précompact, que l'on peut supposer (après approximation) périodique, i.e. engendrant un groupe isomorphe au cercle S^1.

Fait 4.1 *t centralise $\mathcal{A} + \mathcal{Z}$, qui par suite engendre un groupe (abélien) compact, qui est donc en plus central dans $\mathcal{G}$.*

Preuve Soit $T^s = \exp(sad_t)$ le groupe à un paramètre défini par t. Il agit sur $\mathcal{A} + \mathcal{Z}$ par transformations orthogonales (à cause de la précompacité), en particulier semi-simples, à valeurs propres de module égale à 1. Pour montrer que t centralise $\mathcal{A} + \mathcal{Z}$, il suffit de montrer que toute puissance non triviale T^s n'a pas de sous-espace propre de dimension 2. Supposons par l'absurde que $\mathcal{P}$ est un tel sous-espace. C'est en particulier une sous-algèbre de $\mathcal{G}$ car $\mathcal{A} + \mathcal{Z}$ est abélienne. L'algèbre $\mathcal{L}$ engendrée par t et $\mathcal{P}$ est isomorphe à l'algèbre de Lie du groupe des déplacements eucilidien d'un plan (engendrant le groupe des translations-rotations du plan).

Tous les éléments de $\mathcal{P}$ sont nécessairement non précompacts, et donc d'après la condition $(*)$, κ est positive, non triviale sur $\mathcal{P}$. Elle est donc non dégénérée, car son noyau est un idéal propre, qui ne pourrait être que $\mathcal{P}$. En fait κ est une forme lorentzienne bi-invariante sur $\mathcal{L}$ (car on sait déjà que $\kappa(t,t) < 0$).

Il suffit maintenant de remarquer qu'une telle forme, ne peut pas exister. En effet tout groupe à un paramètre défini par un vecteur non tangent à $\mathcal{P}$ (i.e. qui ne soit pas un groupe à un paramètre de translations du plan) est conjugué à celui défini par t, car c'est un groupe de rotation autour d'un certain point. Il s'ensuit que κ est négative en dehors de $\mathcal{P}$, ce qui contredit son caractère lorentzien.

On déduit de 3.2 que $\mathcal{A} + \mathcal{Z}$ détermine un groupe compact. □

En fait, toujours d'après 3.2, le groupe à un paramètre déterminé par t ne commute avec aucun élément non central de $\mathcal{HE}_d$. De plus le groupe engendré par le centre de $\mathcal{HE}_d$ est compact, faute de quoi, toujours d'après 3.2, on aura $\kappa(t,t) \geq 0$.

Notons $\mathcal{S}$ l'algèbre engendrée par t et $\mathcal{HE}_d$ et S le groupe qu'elle détermine.

Un raisonnement élémentaire permet de voir que κ est lorentzienne sur $\mathcal{S}$. On commence par constater que κ est non dégénérée, car son noyau ne pourrait être que le centre, et en quotientant par ce dernier, on trouve une forme lorentzienne bi-invariante sur le produit semi-direct de S^1 agissant, sans vecteur fixe, sur $\mathbf{C}^d$. La preuve qu'on vient de donner ci-dessus, pour $d = 1$, de l'inexistence d'une telle forme, se généralise en toute dimension.

Ce qui précéde montre bien que S est un groupe de Heisenberg tordu.

Considérons l'orthogonal $\mathcal{S}^\perp$. C'est bien un supplémentaire de $\mathcal{S}$. Par bi-invariance de κ, $[X,Y] \in \mathcal{S}^\perp$ dès que $X \in \mathcal{S}$ et $Y \in \mathcal{S}^\perp$. En d'autres termes, $\mathcal{S}$ centralise le sous-espace vectoriel $\mathcal{S}^\perp$. Il en résulte, puisque $\mathcal{HE}_d$ est un idéal de $\mathcal{G}$, que $[X,Y] = 0$ dès que $X \in \mathcal{HE}_d$ et $Y \in \mathcal{S}^\perp$. Autrement dit $\mathcal{S}^\perp$ centralise $\mathcal{HE}_d$.

Soit $X \in \mathcal{S}^\perp$. Il centralise $\mathcal{N} = \mathcal{A} + \mathcal{HE}_d$, car d'après le fait ci-dessus $\mathcal{A}$ est central. Il en découle que $\mathbf{R}X + \mathcal{N}$ est une algèbre nilpotente. C'est en fait un idéal de $\mathcal{G}$, car il est connu que $[\mathcal{G},\mathcal{G}] \subset \mathcal{N}$ (on avait supposé que $\mathcal{G}$ est résoluble). Par maximalité de $\mathcal{N}$, en tant qu' idéal nilpotent, on a: $X \in \mathcal{N}$.

Ainsi $\mathcal{S}^\perp$ est contenue dans le nilradical $\mathcal{N}$. On en déduit pour des raisons de dimension que $\mathcal{N} = \mathcal{S}^\perp + \mathcal{HE}_d$. Ainsi on peut prendre $\mathcal{A} = \mathcal{S}^\perp$. Ce qui achèvera la décomposition dans ce cas. □

4.2.2 Cas où κ est positive

Supposons que κ est positive sur $\mathcal{G}$ (supposée résoluble). Elle admettra un noyau non trivial $\mathcal{I}$, sauf si $\mathcal{G}$ est abélienne. Supposons donc dans la suite que $\mathcal{I}$ est non trivial.

D'après la condition $(*)$, si $\dim(\mathcal{I}) > 1$, alors le sous-groupe I de G qu'il détermine est précompact, i.e. $\bar{I}$ est un tore, nécessairement central. En particulier $\mathcal{I} \subset \mathcal{N}$. Ce qui contredit le fait (3.4) que, sur $\mathcal{N}$, la dimension du noyau de κ est ≤ 1.

Montrons que: $\mathcal{I} \subset \mathcal{N}$. En effet sinon, $\mathcal{I} \cap \mathcal{N} = 0$. Comme $\mathcal{I}$ et $\mathcal{N}$ sont des idéaux, il s'ensuit qu'ils se centraliseent l'un l'autre. En particulier $\mathcal{I} + \mathcal{N}$ est aussi un idéal nilpotent. Ce qui contredit la définition de $\mathcal{N}$. Maintenant, si $\mathcal{G}$ est nilpotente, elle se décompose comme dans 3.4. Ce qui démontre le théorème algébrique dans ce cas. Supposons donc que $\mathcal{G}$ n'est pas nilpotente. L'algèbre quotient est abélienne car elle admet une forme définie positive bi-invariante. Il s'ensuit que $[\mathcal{G}, \mathcal{G}] \subset \mathcal{I}$, mais $\mathcal{I}$ n'est pas central, car sinon $\mathcal{G}$ sera nilpotente. On en déduit que si Y est un générateur de $\mathcal{I}$, alors le noyau de l'application $u \to [u, Y]$ admet un noyau $\mathcal{L}$ de codimension 1. Il existe X orthogonal à $\mathcal{L}$ vérifiant $[X, Y] \neq 0$. On peut en fait supposer quitte à prendre un multiple de X que: $[X, Y] = Y$. Soit $\mathcal{A} \subset \mathcal{L}$ le noyau de $T \in \mathcal{L} \to [X, T] \in \mathcal{I}$. Ainsi X et Y engendre l'algèbre de Lie du groupe affine $\mathcal{GA}$. Pour achever la preuve du théorème dans le présent cas, il suffit de montrer que $\mathcal{A}$ est une algèbre centrale (elle sera alors immédiatement un facteur direct). Comme par construction $\mathcal{A}$ est centralisé par X et Y et s'injecte dans le quotient abélien $\mathcal{G}/\mathcal{I}$, il suffit juste de montrer que $\mathcal{A}$ est bien une algèbre. Soit donc T et T' deux éléments de $\mathcal{A}$. Par l'identité de Jacobi $[X, [T, T']] = 0$. Donc $[T, T']$ est certainement un multiple trivial de Y. □

5 Preuve des Théorèmes géométriques

5.1 Caractère causal de l'action lorsque $\mathcal{G}$ ne contient ni $sl(2, \mathbf{R})$ ni une algèbre de Heisenberg tordue

Pour montrer que lorsque $\mathcal{G}$ ne contient pas $sl(2, \mathbf{R})$ ou une algèbre de Heisenberg tordue, les orbites sont non temporelles, il suffit d'appliquer 1.9, en remarquant que dans ce cas, d'après le théorème algébrique, les groupes à un paramètre non précompact sont denses.

Il s'ensuit que si X est un champ isotrope au sens de κ, alors $X(x)$ est isotrope (au sens de $\langle\ ,\ \rangle_x$) pour tout $x \in M$. Pour montrer que les orbites de X sont géodésiques, on applique le fait suivant:

Lemme 5.1 *Soit X un champ de Killing à norme constante: $\langle X(x), X(x)\rangle$ ne dépend pas de x. Alors les orbites de X sont des géodésiques affinement paramétrées: $\nabla_X X(x) = 0$, pour tout x.*

Preuve En tant que champ de Killing, X vérifie: $\langle \nabla_Y X, X\rangle + \langle Y, \nabla_X X\rangle = 0$, pour tout champ Y. Mais la constance de la norme entraîne: $\langle \nabla_Y X, X\rangle = 0$. Par conséquent: $\nabla_X X = 0$. □

5.2 Caractère localement libre des actions des groupes de Heisenberg tordus

On supposera dans la présente section et la suivante que M est compacte. En effet, on aura affaire dans les démonstrations suivantes à des parties fermées invariantes de M. La comapcité de M assurera l'existence de mesures invariantes supportées par ces parties. La finitude du volume de M ne l'entraîne à priori pas. Cependant, un peu plus d'analyse de notre situation particulière (voir [15]), dont on se permet de se passer pour ne pas encombrer davantage le texte, permet de traîter ce cas là.

Notre approche ressemble à ce niveau à celle de [7].

Soit S un groupe de Heisenberg tordu, produit semi-direct de S^1 par He_d, et soit $Z = \{\phi^s, s \in [0, \pi]\}$ son centre. Il est facile de tirer du fait que (d'après ce qui précède) les orbites du groupes de Heisenberg sont non temporelles, que les orbites de Z sont isotropes. Elles sont ainsi de plus géodésiques d'après le lemme 5.1. Il s'ensuit que Z n'admet pas de point fixe. En effet au voisinage d'un tel point, il y aura des géodésiques fermées arbitrairement petites (ce qui contredit la convexité locale des variétés munies de connexions affines).

Nous allons maintenant montrer par l'absurde que l'action de S est localement libre et ce en montrant que sinon l'action de Z ne l'est pas. En effet soit F le fermé de M des points ayant un stabilisateur S_x non discret. Notons $\mathcal{S}_x$ son algèbre de Lie. On se restreint au fermé F_k où la dimension de $\mathcal{S}_x$ est maximale égale à k (certainement $k < \dim(S)$ car sinon en particulier Z aura un point fixe). L'action de S sur F_k préserve une mesure finie μ car F_k est compact et S est résoluble. La méthode de preuve suivante est standard (voir par exemple [6]). Considérons l'application de Gauss: $Ga\colon F_k \to Gr^k(\mathcal{S})$ qui à $x \in F_k$ associe $\mathcal{S}_x$ l'algèbre de Lie de son stabilisateur. Elle est équivariante par rapport aux actions de S. Ainsi $Ga^*(\mu)$ est une mesure sur $Gr^k(\mathcal{S})$ invariante par l'action de S.

Le lemme de Furstenberg [6], s'applique aux actions des groupes algébriques. Considérons donc la restriction de l'action précédente au groupe de Heisenberg $He_d \subset S$. D'après Furstenberg, cette action se factorise sur le support de la mesure, en l'action d'un groupe compact. Mais He_d n'a aucun groupe quotient compact non trivial. Il s'ensuit que pour μ presque tout x, $\mathcal{S}_x$ est normalisée par He_d. Si $\mathcal{S}_x \cap \mathcal{HE}_d$ est non triviale, on aura un idéal non trivial de $\mathcal{HE}_d$. Il contiendra obligatoirement le centre. Lorsque $\mathcal{S}_x$ intersecte trivialement $\mathcal{HE}_d$, elle en sera un supplémentaire pour des raisons de dimension. Ainsi $\mathcal{S}_x$ sera normalisée par toute l'algèbre $\mathcal{S}$, c'est-à-dire que $\mathcal{S}_x$ est un idéal. Ceci est impossible. □

5.3 Caractère lorentzien des orbites de S

Il découle du fait que l'action est localement libre et du fait que les orbites de He_d sont non temporelles, qu'en tout point x, et pour tout X tangent à He_d, les vecteurs $X(x)$ sont de type espace, sauf exactement celui correspondant au centre, qui est isotrope. Pour montrer que les orbites sont lorentziennes, il suffit donc de montrer qu'elles sont non dégénérées. Or dans ce cas, le noyau de la métrique sera exactement le centre (car l'action est localement libre). L'ensemble des points à orbite dégénérée est un fermé invariant. Il supporte donc une mesure finie invariante μ. La forme L^2 associée, i.e. $\kappa(X,Y) = \int \langle X(x), Y(x)\rangle d\mu(x)$ est une forme bi-invariante sur $\mathcal{S}$, positive et à noyau exactement le centre. Ainsi le quotient de $\mathcal{S}$ par son centre admettra une métrique définie positive bi-invariante. Mais ceci n'arrive pour un groupe résoluble que s'il est abélien. □

5.4 L'orthogonal

Notons $\mathcal{O}$ la distribution orthogonale aux orbites de $\mathcal{S}$. On va montrer que $\mathcal{O}+\mathcal{Z}$ est intégrable (où $\mathcal{Z}$ est le champ de directions déterminé par le centre Z). En tout point x, on a une forme antisymétrique: $\omega : \mathcal{O}_x \times \mathcal{O}_x \to \mathcal{S}_x$, $\omega(A,B) =$ la partie normale du crochet $[A,B]$. L'identification canonique de $\mathcal{S}_x$ à l'algèbre de Lie $\mathcal{S}$ permet d'identifier ω à une forme à valeurs dans $\mathcal{S}$. Elle vérifie la relation d'équivariance évidente: $\omega(gA, gB) = Ad(g)\omega(A,B)$. Or la métrique sur $\mathcal{O}$ est riemannienne, et par suite, pour tous A, B vecteurs de $\mathcal{O}$, l'orbite $\{(gA, gB)/, g \in S\}$ est précompacte dans $\mathcal{O}\times\mathcal{O}$. Il en découle que l'orbite de $\omega(A,B)$ par l'action adjointe de S est précompacte. On vérifie facilement que ceci n'est le cas que du centre Donc ω est à valeurs dans $\mathcal{Z}$. ce qui veut exactement dire que $\mathcal{O}+\mathcal{Z}$ est intégrable. □

Pour ce qui précède ainsi que ce qui suit, on peut consulter respectivement [4] et [3], où l'on traîte de situations semblables mais plus délicates.

5.5 Structure

On va transformer "canoniquement" la métrique lorentzienne $\langle\ ,\ \rangle$ de M en une métrique riemannienne $(\ ,\ )$ (qui ne sera aucunement invariante par l'action de S). On décréte que $\mathcal{O}$ reste orthogonale aux orbites et reste équipée de la même métrique. On définit la métrique sur $\mathcal{S}_x$ par $(X(x), Y(x)) = b(X,Y)$, où b est un produit scalaire défini positif quelconque (loin d'être bi-invariant) sur

$\mathcal{S}$. On prendra par exemple: $(X_i(x), X_j(x)) = \delta_{ij}$ pour une certaine base $\{X_i\}$ de $\mathcal{S}$. Le groupe S sera également équipé de la métrique invariante à droite déterminée par b. Ainsi, pour tout x, le revêtement $S \to Sx$ est isométrique (cela ne veut en aucun cas dire que S agit isométriquement sur l'orbite Sx au sens de la nouvelle métrique riemannienne).

Soit L la feuille du feuilletage $\mathcal{O} + \mathcal{Z}$ passant par un certain point x_0 munie de la métrique induite de $(\, , \,)$. Le centre Z y agit isométriquement.

On a une application: p: $S \times L \to M$, $p(g, x) = gx$. On vérifie que p est une submersion riemannienne dont l'espace horizontal est $\mathcal{S} + \mathcal{O}$. Plus précisément, considérons $S \times_{S^1} L$, le quotient de $S \times L$ par l'action diagonale, et munissons le de la métrique projetée de celle de $\mathcal{S} + \mathcal{O}$. Alors l'application induite π: $S \times_{S^1} L \to M$ est localement isométrique. Par un résultat bien connu sur les applications localement isométriques, π est un revêtement, car la métrique sur $S \times_{S^1} L$ est évidemment complète.

Ainsi on a: $M = \Gamma \setminus S \times_{S^1} L$, où Γ est un réseau de $S \times Isom_{S^1}(L)$. Il ne reste donc du théorème de structure 1.14, dans le cas des groupes de Heisenberg tordus, qu'à expliciter la métrique lorentzienne sur $S \times_{S^1} L$. Plus précisément il s'agit de montrer que la métrique sur S est essentiellement bi-invariante (1.4), ce qui fera l'objet de la section suivante. .

5.6 Métriques lorentziennes sur S

On voit d'après ce qui précède que la métrique lorentzienne, notons la m, le long des orbites, qui est par hypothèse invariante par l'action à gauche de S, doit également être invariante à droite par Γ_0, la projection de Γ sur S. Cette projection n'est pas nécessairement discrète, mais elle est à covolume fini, au sens qu'il existe un sous-ensemble de volume fini dont les itérés par Γ_0 couvrent S. Considérons la fermeture topologique $\bar{\Gamma_0}$. C'est un sous-groupe unimodulaire (car tous les élements de $Ad(\mathcal{S})$ sont à valeurs propres de module égale à 1). On peut facilement voir que la mesure de Haar passe en une mesure finie sur $S/\bar{\Gamma_0}$.

Elle détérmine une mesure finie sur l'orbite de la métrique m, invariante par l'action adjointe de S. Donc, d'après le lemme de Furstenberg, l'action restreinte à He_d se factorise en l'action d'un groupe compact. Comme ci-dessus, ceci entraîne que m est $Ad(He_d)$–invariante.

A titre de complément, on a le fait suivant qui montre qu'il n' y a qu'une seule géométrie lorentzienne locale sur un groupe de Heisenberg tordu S. Elle mérite certainement d' être mieux comprise.

Proposition 5.2 *Deux métriques lorentziennes bi-invariantes quelconques sur $\mathcal{S}$ sont équivalentes par un automorphisme.*

Preuve Reprenons les notations de lapreuve de 1.3. Remarquons d'abord, qu'on peut supposer, après automorphisme, que $\beta = 0$. Il suffit pour cela d'appliquer un automorphisme trivial sur $\mathcal{HE}_d$ et envoyer t sur $t + \delta Z$ pour un δ convenable.

Pour normaliser le paramètre α, on applique le groupe à paramètre d'homothéties célèbres de l'algèbre de Heisenberg $\mathcal{HE}_d$. Il commute avec tous les automorphismes et donc se prolonge trivialement au produit semi-direct $\mathcal{S}$. Il se définit ainsi: $t \to t$, $Z \to \exp(2t)Z$ et $X \to \exp(t)X$, pour $X \in \mathbf{C}^d$. (Ceci induit des homothéties de $\tilde{S}$ muni de la métrique donnée initialement). □

5.7 Cas de $sl(2, \mathbf{R})$

D'après le théorème algébrique, l'action de $sl(2, \mathbf{R})$ s'intégre en une action d'un revêtement fini $PSL_k(2, \mathbf{R})$ de $PSL(2, \mathbf{R})$. Montrons brièvement dans ce qui suit le théorème de structure 1.13 dû à [7].

Soit κ la forme de Killing de $sl(2, \mathbf{R})$. Montrons que si $Y \in sl(2, \mathbf{R})$ est isotrope au sens de κ, alors $Y(x)$ est isotrope au sens de $\langle\ ,\ \rangle_x$ pour tout x. En effet, il est connu qu'un tel Y est caractérisé par le fait que ad_X est nilpotent (ou de manière équivalente que la matrice 2×2 et à trace 0, correspondante, est nilpotente). Il est également connu, qu'alors il existe $X \in sl(2, \mathbf{R})$ tel que $[X, Y] = -Y$. En d'autres termes si ϕ^t est le flot de X, alors $\phi^t Y = \exp(-t)Y$. En particulier la fonction $\langle Y, Y\rangle$ décroît (exponentiellement) le long des orbites de X. Cette fonction est donc constamment nulle, car ϕ^t préserve le volume.

Il en découle que pour tout x, la métrique restreinte à l'orbite de x est proportionnelle à κ: $\langle X(x), Y(x)\rangle = f(x)\kappa(X, Y)$ pour tous X, Y et x.

Il s'ensuit en particulier qu'une orbite singulière est isotrope. Elle est en particulier de dimension 1 ou 0. Si elle est de dimension 1, elle sera d'après le lemme 5.1, une géodésique (isotrope). L'action de $sl(2, \mathbf{R})$ préserve sa structure affine, ce qu'on peut facilement voir être impossible. L'orbite singulière est donc de dimension 0, i.e. un point fixe x_0 de $PSL_k(2, \mathbf{R})$. Soit ϕ^t ($t \in [0, 2\pi]$) un groupe à un paramètre de rotation de $PSL_k(2, \mathbf{R})$. Les orbites par ϕ^t des points proches de x_0 sont des courbes également proches de x_0. On montrera dans la suite que ces courbes sont de type temps, i.e. à l'intérieur du cône de lumière. Ceci est impossible, car une courbe dirigée par un champ de cônes ne peut pas se refermer localement.

Le point x_0 peut être approché par des points non singuliers car l'ensemble de ces derniers points est de mesure totale (comme pour toute action fidèle préservant le volume d'un groupe semi-simple (voir par exemple [6]). La métrique sur l'orbite d'un point non singulier x est lorentzienne, car sinon l'orbite sera isotrope, ce qui est impossible car elle est de dimension 3. De plus $X(x)$ a le même caractère causal que X, pour tout $X \in sl(2, \mathbf{R})$. Comme tout X engendrant un flot non précompact est partout non temporel, il en découle que les champs de type temps sont exactement ceux qui déterminent des flots compacts.

Enfin la même méthode de preuve que pour les groupes de Heisenberg tordus permet de conclure que l'orthogonal est cette fois intégrable. □

5.8 Variétés homogènes

Soit $(M, \langle\ ,\ \rangle)$ une variété lorentzienne homogène de volume fini. Son algèbre de champs de Killing agit dessus localement transitivement. En particulier en tout point, il y a des champs de Killing ayant un caractère causal quelconque. Ceci exclut la situation décrite en 5.1. En d'autres termes, le groupe d'isométries G contient un groupe S qui est soit localement isomorphe à $SL(2, \mathbf{R})$, soit isomorphe à un groupe de Heisenberg tordu. Dans chacun des ces deux cas, d'après ce qui précède, M admet un revêtement qui est un produit tordu de S par une variété riemannienne L. Notre hypothèse d' homogénéité nous permet de choisir L compacte. En effet comme dans les preuves précédentes, en désignant comme toujours par $\mathcal{O}$ l'orthogonal aux orbites, on prendra pour L soit une feuille de $\mathcal{O}$ lorsque S est localement isomorphe à $SL(2, \mathbf{R})$, soit une feuille de $\mathcal{O} + \mathcal{Z}$ lorsque S est un groupe de Heisenberg tordu. Soit H la composante neutre du sous-groupe de G fixant (globalement) L. On déduit aisément du théorème algébrique que H est compact. Il en va de même pour L, car H agit transitivement dessus (à cause de l' homogénéité de M).

Enfin pour voir que le groupe du revêtement Γ est le graphe d'un homomorphisme d'un réseau de S, on remarque simplement que par compacité de L, Γ se projette sur un groupe discret de S. Le noyau de la projection de Γ sur S est un sous-groupe fini d'isométries de L, qu'on peut supposer trivial en passant à un quotient fini de L. □

Bibliographie

[1] **S Adams**, **G Stuck**, *The isometry group of a compact Lorentz manifold, I*, Invent. Math. 129 (1997) 239–261

[2] **S Adams**, **G Stuck**, *The isometry group of a compact Lorentz manifold, II*, Invent. Math. 129 (1997) 263–287

[3] **G Cairns**, **E Ghys**, *Totally geodesic foliations on* 4*–manifolds*, J. Diff. Geom. 23 (1986) 241–254

[4] **Y Carrière**, **E Ghys**, *Feuilletages totalement géodésiques*, An. Acad. Brasil. Ciênc 53 (1981) 427–432

[5] **G D'Ambra**, *Isometry groups of Lorentz manifolds*, Invent. Math. 92 (1988) 555–565

[6] **G D'Ambra**, **M Gromov**, *Lectures on transformation groups: geometry and dynamics*, Surveys in Differential Geometry (Supplement to the Journal of Differential Geometry), 1 (1991) 19–111

[7] **M Gromov**, *Rigid transformation groups*, dans: "Géometrie différentielle ", D Bernard et Choquet-Bruhat (editeurs), Travaux encours 33, Paris, Hermann (1988)

[8] **J Marsden**, *On completeness of homogeneous pseudo-riemannian manifolds*, Ind. Univ. Math. J. 22 (1973) 1065–1066

[9] **A Medina**, **Ph Revoy**, *Les groupes oscillateures et leurs réseaux*, Manuscripta Math. 52 (1985) 81–95

[10] **M S Raghunathan**, *Discrete subgroups of Lie groups*, Springer–Verlag (1972)

[11] **R F Streater**, *The representations of the oscillator group*, Commun. Math. Phys. 4 (1967) 217–236

[12] **J Wolf**, *Homogeneous manifolds of constant curvature*, Comment. Math. Helv. 36 (1961) 112–147

[13] **A Zeghib**, *Killing fields in compact Lorentz* 3*–manifolds*, J. Diff. Geom. 43 (1996) 859–894

[14] **A Zeghib**, *The identity component of the isometry group of a compact Lorentz manifold*, Duke Math. Journal, (à paraître)

[15] **A Zeghib**, *Isometry groups and geodesic foliations of Lorentz manifolds. Part I: Foundations of Lorentz dynamics*, preprint

[16] **R Zimmer**, *On the automorphism group of a compact Lorentz manifold and other geometric manifolds*, Invent. Math. 83 (1986) 411–426

Ecole Normale Superieure de Lyon
UMPA, UMR 128 CNRS
46 Allee d'Italie, 69364 Lyon, FRANCE

Email: zeghib@umpa.ens-lyon.fr

Received: 15 November 1997