NORMAL SUBGROUPS IN THE CREMONA GROUP (LONG VERSION)

SERGE CANTAT AND STÉPHANE LAMY

Abstract. Let $\mathbf{k}$ be an algebraically closed field. We show that the Cremona group of all birational transformations of the projective plane $\mathbb{P}^2_{\mathbf{k}}$ is not a simple group. The strategy makes use of hyperbolic geometry, geometric group theory, and algebraic geometry to produce elements in the Cremona group which generate non trivial normal subgroups.

Résumé. Soit $\mathbf{k}$ un corps algébriquement clos. Nous montrons que le groupe de Cremona, formé des transformations birationnelles du plan projectif $\mathbb{P}^2_{\mathbf{k}}$, n'est pas un groupe simple. Pour cela nous produisons des éléments engendrant des sous-groupes normaux non triviaux à l'aide de théorie géométrique des groupes, de géométrie hyperbolique et de géométrie algébrique.

Contents

1. Introduction

1.1. **The Cremona group.** The Cremona group in n variables over a field $\mathbf{k}$ is the group $\mathsf{Bir}(\mathbb{P}^n_{\mathbf{k}})$ of birational transformations of the projective space $\mathbb{P}^n_{\mathbf{k}}$ or,

Date: July 2010.

The second author was supported by a Marie Curie Intra European Fellowship.

Main differences with the submitted version [7]: proofs of Corollary 2.2 and Lemmas 2.3, 2.4, 2.5, 2.6, 3.1 are given.

equivalently, the group of $\mathbf{k}$-automorphisms of the field $\mathbf{k}(x_1, \cdots, x_n)$ of rational functions of $\mathbb{P}^n_{\mathbf{k}}$. In dimension $n = 1$, the Cremona group coincides with the group $\mathsf{Aut}(\mathbb{P}^1_{\mathbf{k}})$ of regular automorphisms of $\mathbb{P}^1_{\mathbf{k}}$, that is with the group $\mathsf{PGL}_2(\mathbf{k})$ of linear projective transformations. When $n \geq 2$, the group $\mathsf{Bir}(\mathbb{P}^n_{\mathbf{k}})$ is infinite dimensional; it contains algebraic groups of arbitrarily large dimensions (see [2, 39]). In this article, we prove the following theorem.

Main Theorem. *If* $\mathbf{k}$ *is an algebraically closed field, the Cremona group* $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ *is not a simple group.*

This answers one of the main open problems concerning the algebraic structure of this group: According to Dolgachev, Manin asked whether $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ is a simple group in the sixties; Mumford mentionned also this question in the early seventies (see [36]). The proof of this theorem makes use of geometric group theory and isometric actions on infinite dimensional hyperbolic spaces. It provides stronger results when one works over the field of complex numbers $\mathbf{C}$; before stating these results, a few remarks are in order.

Remark 1.1. a.– There is a natural topology on the Cremona group which induces the Zariski topology on the space $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{k}})$ of birational transformations of degree d (see [39, 2] and references therein). J. Blanc proved in [3] that $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ has only two *closed* normal subgroups for this topology, namely $\{\mathrm{Id}\}$ and $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ itself. J. Déserti proved that $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ is perfect, hopfian, and co-hopfian, and that its automorphism group is generated by inner automorphisms and the action of automorphisms of the field of complex numbers (see [15, 14]). In particular, there is no obvious algebraic reason which explains why $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ is not simple.
1.1. b.– An interesting example is given by the group $\mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ of polynomial automorphisms of the affine plane $\mathbb{A}^2_{\mathbf{k}}$ with Jacobian determinant equal to 1. From Jung's theorem, one knows that this group is the amalgamated product of the group of affine transformations $\mathsf{SL}_2(\mathbf{k}) \ltimes \mathbf{k}^2$ with the group of elementary transformations

$$f(x, y) = (ax + p(y), y/a),\ a \in \mathbf{k}^*,\ p(y) \in \mathbf{k}[y],$$

over their intersection (see [33] and references therein); as such, $\mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ acts faithfully on a simplicial tree without fixing any edge or vertex. Danilov used this action, together with small cancellation theory, to prove that $\mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ is not simple (see [12] and [24]).
1.1. c.– The group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ contains a linear group with Kazhdan property (T), namely $\mathsf{PGL}_3(\mathbf{C})$, and it cannot be written as a non trivial free product with amalgamation. Thus, even if our strategy is similar to Danilov's proof, it has to be more involved (see §6.1.2).

1.2. **General elements.** Let $[x : y : z]$ be homogeneous coordinates for the projective plane $\mathbb{P}^2_{\mathbf{k}}$. Let f be a birational transformation of $\mathbb{P}^2_{\mathbf{k}}$. There are three homogeneous polynomials P, Q, and $R \in \mathbf{k}[x,y,z]$ of the same degree d, and without common factor of degree ≥ 1, such that

$$f[x : y : z] = [P : Q : R].$$

The degree d is called the **degree** of f, and is denoted by $\deg(f)$. The space $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{k}})$ of birational transformations of degree d is a quasi-projective variety. For instance, Cremona transformations of degree 1 correspond to the automorphism group $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}}) = \mathsf{PGL}_3(\mathbf{k})$ and Cremona transformations of degree 2 form an irreducible variety of dimension 14 (see [8]).

We define the **de Jonquières group** J as the group of birational transformations of the plane $\mathbb{P}^2_{\mathbf{k}}$ which preserve the pencil of lines through the point $[1 : 0 : 0]$. Let J_d be the subset of J made of birational transformations of degree d, and let V_d be the subset of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{k}})$ whose elements are compositions $h_1 \circ f \circ h_2$ where h_1 and h_2 are in $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}})$ and f is in J_d. The dimension of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{k}})$ is equal to $4d+6$ and V_d is its unique irreducible component of maximal dimension (see [37]).

On an algebraic variety W, a property is said to be **generic** if it is satisfied on the complement of a Zariski closed subset of W of codimension ≥ 1, and is said to be **general** if it is satisfied on the complement of countably many Zariski closed subsets of W of codimension ≥ 1.

Theorem A. *There exists a positive integer k with the following property. If g is a general element of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$, and n is an integer with $n \geq k$, then g^n generates a normal subgroup of the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ that does not contain any element f of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}}) \setminus \{Id\}$ with $deg(f) < \deg(g)^n$.*

To prove Theorem A we choose g in the unique irreducible component of maximal dimension V_d. Thus, the proof does not provide any information concerning general elements of the remaining components of the variety $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$. As a corollary of Theorem A, *the group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ contains an uncountable number of normal subgroups* (see §6.3).

Remark 1.2. We provide an explicit upper bound for the best possible constant k, namely $k \leq 86611$. We do so to help the reader to follow the proof. Our method gives a rather small constant for large enough degrees, namely $k \leq 375$, but does not answer the following open question: Is the best constant k equal to 1 ? In other words, does a general element of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ generates a strict normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ for all degrees $d \geq 2$? (Note that g^n is not a general element of $\mathsf{Bir}_{d^n}(\mathbb{P}^2_{\mathbf{C}})$ if $n \geq 2$ and $g \in \mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$)

1.3. **Automorphisms of Kummer and Coble surfaces.** If X is a rational surface, then the group of automorphisms $\mathsf{Aut}(X)$ is conjugate to a subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$. In §5.2, we study two classes of examples. The first one is a (generalized) Kummer surface. Let $\mathbf{Z}[i] \subset \mathbf{C}$ be the lattice of Gaussian integers. We start with the abelian surface $Y = E \times E$ where E is the elliptic curve $\mathbf{C}/\mathbf{Z}[i]$. The group $\mathsf{SL}_2(\mathbf{Z})$ acts by linear automorphisms on Y, and commutes with the order 4 homothety $\eta(x, y) = (ix, iy)$. The quotient Y/η is a (singular) rational surface on which $\mathsf{PSL}_2(\mathbf{Z})$ acts by automorphisms. Since Y/η is rational, the conjugacy by any birational map $\phi\colon Y/\eta \dashrightarrow \mathbb{P}^2_{\mathbf{C}}$ provides an isomorphism between $\mathsf{Bir}(Y/\eta)$ and $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$, and therefore an embedding of $\mathsf{PSL}_2(\mathbf{Z})$ into $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$.

Theorem B. *There is an integer $k \geq 1$ with the following property. Let M be an element of $\mathsf{SL}_2(\mathbf{Z})$ with trace $|tr(M)| \geq 3$. Let g_M be the automorphism of the rational Kummer surface Y/η which is induced by M. Then, ${g_M}^k$ generates a non trivial normal subgroup of the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}}) \simeq \mathsf{Bir}(Y/\eta)$.*

Theorems A and B provide examples of normal subgroups coming respectively from general and from highly non generic elements in the Cremona group. In §5.2.3, we also describe automorphisms of Coble surfaces that generate non trivial normal subgroups of the Cremona group; Coble surfaces are quotients of K3 surfaces, while Kummer surfaces are quotient of abelian surfaces. Thanks to a result due to Coble and Dolgachev, for any algebraically closed field $\mathbf{k}$ we obtain automorphisms of Coble surfaces which generate proper normal subgroups in the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$. The Main Theorem follows from this construction (see Theorem 5.20).

1.4. **An infinite dimensional hyperbolic space (see §4).** The Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ acts faithfully on an infinite dimensional hyperbolic space which is the exact analogue of the classical finite dimensional hyperbolic spaces $\mathbb{H}^n$. This action is at the heart of the proofs of Theorems A and B. To describe it, let us consider all rational surfaces $\pi\colon X \to \mathbb{P}^2_{\mathbf{k}}$ obtained from the projective plane by successive blow-ups. If $\pi'\colon X' \to \mathbb{P}^2_{\mathbf{k}}$ is obtained from $\pi\colon X \to \mathbb{P}^2_{\mathbf{k}}$ by blowing up more points, then there is a natural birational morphism $\varphi\colon X' \to X$, defined by $\varphi = \pi^{-1} \circ \pi'$, and the pullback operator φ^* embeds the Néron-Severi group $\mathsf{N}^1(X) \otimes \mathbf{R}$ into $\mathsf{N}^1(X') \otimes \mathbf{R}$ (note that $\mathsf{N}^1(X) \otimes \mathbf{R}$ can be identified with the second cohomology group $H^2(X, \mathbf{R})$ when $\mathbf{k} = \mathbf{C}$). The direct limit of all these groups $\mathsf{N}^1(X) \otimes \mathbf{R}$ is called the **Picard-Manin space** of $\mathbb{P}^2_{\mathbf{k}}$. This infinite dimensional vector space comes together with an intersection form of signature $(1, \infty)$, induced by the intersection form on divisors; we shall denote it by

$$([\alpha], [\beta]) \mapsto [\alpha] \cdot [\beta].$$

We obtain in this way an infinite dimensional Minkowski space. The set of elements $[\alpha]$ in this space with self intersection $[\alpha]\cdot[\alpha]=1$ is a hyperboloid with two sheets, one of them containing classes of ample divisors of rational surfaces; this connected component is an infinite hyperbolic space for the distance dist defined in terms of the intersection form by

$$\cosh(\mathsf{dist}([\alpha],[\beta])) = [\alpha]\cdot[\beta].$$

Taking the completion of this metric space, we get a complete, hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$. The Cremona group acts faithfully on the Picard-Manin space, preserves the intersection form, and acts by isometries on the hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$.

1.5. **Normal subgroups in isometry groups.** In Part A, we study the general situation of a group G acting by isometries on a δ-hyperbolic space $\mathcal{H}$. Let us explain the content of the central result of Part A, namely Theorem 2.9, in the particular case of the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ acting by isometries on the hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$. Isometries of hyperbolic spaces fall into three types: elliptic, parabolic, and hyperbolic. A Cremona transformation $g\in\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ determines a hyperbolic isometry g_* of $\mathbb{H}_{\bar{\mathcal{Z}}}$ if and only if the following equivalent properties are satisfied:

- The sequence of degrees $\deg(g^n)$ grows exponentially fast:
$$\lambda(g) := \limsup_{n\to\infty}\left(\deg(g^n)^{1/n}\right) > 1;$$
- There is a g_*-invariant plane V_g in the Picard-Manin space that intersects $\mathbb{H}_{\bar{\mathcal{Z}}}$ on a curve $\mathrm{Ax}(g_*)$ (a geodesic line) on which g_* acts by a translation; more precisely, $\mathsf{dist}(x,g_*(x))=\log(\lambda(g))$ for all x in $\mathrm{Ax}(g_*)$.

The curve $\mathrm{Ax}(g_*)$ is uniquely determined and is called the **axis** of g_*. We shall say that an element g of the Cremona group is **tight** if it satisfies the following three properties:

- The isometry $g_*\colon\mathbb{H}_{\bar{\mathcal{Z}}}\to\mathbb{H}_{\bar{\mathcal{Z}}}$ is hyperbolic;
- There exists a positive number B such that: If f is an element of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ and $f_*(\mathrm{Ax}(g_*))$ contains two points at distance B which are at distance at most 1 from $\mathrm{Ax}(g_*)$ then $f_*(\mathrm{Ax}(g_*))=\mathrm{Ax}(g_*)$;
- If f is in $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ and $f_*(\mathrm{Ax}(g_*))=\mathrm{Ax}(g_*)$, then $fgf^{-1}=g$ or g^{-1}.

The second property is a rigidity property of $\mathrm{Ax}(g)$ with respect to isometries f_* with f in $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$. The third property means that the stabilizer of $\mathrm{Ax}(g)$ coincides with the normalizer of the cyclic group $g^{\mathbf{Z}}$. Applied to the Cremona group, Theorem 2.9 gives the following statement.

Theorem C. *Let g be an element of the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$. If the isometry $g_*\colon \mathbb{H}_{\bar{\mathcal{Z}}} \to \mathbb{H}_{\bar{\mathcal{Z}}}$ is tight, there exists a positive integer k such that g^k generates a non trivial normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$.*

Theorems A and B can now be rephrased by saying that general elements of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ (resp. automorphisms g_M on rational Kummer surfaces) are tight elements of the Cremona group (see Theorems 5.14 and 5.20).

1.6. **Description of the paper.** The paper starts with the proof of Theorem C in the general context of δ-hyperbolic spaces (section 2), and explains how this general statement can be used in the case of isometry groups of spaces with constant negative curvature (section 3). Section 4 provides an overview on the Picard-Manin space, the associated hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$, and the isometric action of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ on this space. Algebraic geometry and geometric group theory are then put together to prove Theorem A (§5.1), Theorem B, and the Main Theorem (§5.2). At the end of the paper we list some remarks and comments.

1.7. **Acknowledgment.** Thanks to Thomas Delzant, Koji Fujiwara, and Vincent Guirardel for interesting discussions regarding hyperbolic groups and geometric group theory, and to Jérémy Blanc, Dominique Cerveau, Julie Déserti, Igor Dolgachev, and Jean-Pierre Serre for nice discussions on the Cremona group and Coble surfaces.

Part A. Hyperbolic Geometry and Normal Subgroups

2. Gromov hyperbolic spaces

This section is devoted to the proof of Theorem 2.9, a result similar to Theorem C but in the general context of isometries of Gromov hyperbolic spaces.

2.1. Basic definitions.

2.1.1. *Hyperbolicity.* Let $(\mathcal{H}, \mathsf{d})$ be a metric space. If x is a base point of $\mathcal{H}$, the Gromov product of two points $y, z \in \mathcal{H}$ is

$$(y|z)_x = \frac{1}{2}\left\{\mathsf{d}(y,x) + \mathsf{d}(z,x) - \mathsf{d}(y,z)\right\}.$$

The triangle inequality implies that $(y|z)_x$ is a non negative real number. Let δ be a non negative real number. The metric space $(\mathcal{H}, \mathsf{d})$ is **δ-hyperbolic** in the sense of Gromov if

$$(x|z)_w \geq \min\{(x|y)_w,\, (y|z)_w\} - \delta$$

for all x, y, w, and z in $\mathcal{H}$. Equivalently, the space $\mathcal{H}$ is δ-hyperbolic if, for any $w, x, y, z \in \mathcal{H}$, we have

$$\mathsf{d}(w,x) + \mathsf{d}(y,z) \leq \max\{\mathsf{d}(w,y) + \mathsf{d}(x,z), \mathsf{d}(w,z) + \mathsf{d}(x,y)\} + \delta. \tag{2.1}$$

Two fundamental examples of δ-hyperbolic spaces are given by simplicial metric trees, or more generally by $\mathbf{R}$-trees, since they are 0-hyperbolic, and by the n-dimensional hyperbolic space $\mathbb{H}^n$, or more generally CAT(-1) spaces, since both are δ-hyperbolic for $\delta = \log(3)$ (see [25, chapter 2], [9]).

2.1.2. *Geodesics.* A **geodesic segment** from x to y in $\mathcal{H}$ is an isometry γ from an interval $[s,t] \subset \mathbf{R}$ into $\mathcal{H}$ such that $\gamma(s) = x$ and $\gamma(t) = y$. The metric space $(\mathcal{H}, \mathsf{d})$ is a **geodesic space** if any pair of points in $\mathcal{H}$ can be joined by a geodesic segment. Let $[x,y] \subset \mathcal{H}$ be a geodesic segment, and let $\gamma \colon [0,l] \to \mathcal{H}$ be an arc length parametrization of $[x,y]$, with $\gamma(0) = x$ and $\gamma(l) = y$. By convention, for $0 \leq s \leq l - t \leq l$, we denote by $[x+s, y-t]$ the geodesic segment $\gamma([s, l-t])$.

A **geodesic line** is an isometry γ from $\mathbf{R}$ to its image $\Gamma \subset \mathcal{H}$. If x is a point of $\mathcal{H}$ and Γ is (the image of) a geodesic line, a **projection** of x onto Γ is a point $z \in \Gamma$ which realizes the distance from x to Γ. Projections exist but are not necessarily unique (in a tree, or in $\mathbb{H}^n$, the projection is unique). If $\mathcal{H}$ is δ-hyperbolic and if z and z' are two projections of a point x on a geodesic Γ, the distance between z and z' is at most 2δ. In other words, the projection is unique up to an error of magnitude 2δ. In what follows, we shall denote by $\pi_\Gamma(x)$ any projection of x on Γ.

2.2. **Classical results from hyperbolic geometry.** In what follows, $\mathcal{H}$ will be a δ-hyperbolic and geodesic metric space.

2.2.1. *Approximation by trees.* In a δ-hyperbolic space, the geometry of finite subsets is well approximated by finite subsets of metric trees; we refer to [25] chapter 2, or [9], for a proof of this basic and crucial fact.

Lemma 2.1 (Approximation by trees)**.** *Let $\mathcal{H}$ be a δ-hyperbolic space, and $(x_0, x_1, \cdots, x_n)$ a finite list of points in $\mathcal{H}$. Let X be the union of the n geodesic segments $[x_0, x_i]$ from the base point x_0 to each other x_i, $1 \leq i \leq n$. Then there exist a metric tree T and a map $\Phi : X \mapsto T$ such that:*

(1) Φ is an isometry from $[x_0, x_i]$ to $[\Phi(x_0), \Phi(x_i)]$ for all $1 \leq i \leq n$;
(2) for all $x, y \in X$

$$\mathsf{d}(x,y) - 2\delta \log_2(n) \leq \mathsf{d}(\Phi(x), \Phi(y)) \leq \mathsf{d}(x,y).$$

The map $\Phi \colon X \mapsto T$ is called an **approximation tree**. For the sake of simplicity, the distance in the tree is also denoted $\mathsf{d}(\cdot,\cdot)$. However, to avoid any confusion, we stick to the convention that a point in the tree will always be written under the form $\Phi(x)$, with $x \in X$.

Convention. When $n \leq 4$ then $2\log_2(n) \leq 4$; choosing θ in $\mathbf{R}_+$ such that $\theta = 4\delta$, we have

$$\mathsf{d}(x, y) - \theta \leq \mathsf{d}(\Phi(x), \Phi(y)) \leq \mathsf{d}(x, y). \tag{2.2}$$

From now on we fix such a θ, and we always use the approximation lemma in this way for at most 5 points $(x_0, \cdots, x_4)$. The first point in the list will always be taken as the base point and will be denoted by a white circle in the pictures. Since two segments with the same extremities are δ-close, the specific choice of the segments between x_0 and the x_i is not important. We may therefore forget which segments are chosen, and refer to an approximation tree as a pair (Φ, T) associated to $(x_0, \cdots, x_n)$. In general, the choice of the segments between x_0 and the x_i is either clear, or irrelevant.

In the remaining of §2.2, well known facts from hyperbolic geometry are listed; complete proofs can also be found in [25], [9]. First, the following corollary is an immediate consequence of the approximation lemma, and Lemma 2.3 follows from the corollary.

Corollary 2.2. *Let $\Phi\colon X \mapsto T$ be an approximation tree for at most 5 points. Then*

$$(\Phi(x)|\Phi(y))_{\Phi(z)} - \frac{\theta}{2} \leq (x|y)_z \leq (\Phi(x)|\Phi(y))_{\Phi(z)} + \theta$$

for all x, y, and z in X. In particular $(x|y)_z \leq \theta$ as soon as $\Phi(z) \in [\Phi(x), \Phi(y)]$.

Proof. By definition of the Gromov product we have

$$\begin{aligned} 2(x|y)_z &= \mathsf{d}(x, z) + \mathsf{d}(y, z) - \mathsf{d}(x, y) \\ &\geq \mathsf{d}(\Phi(x), \Phi(z)) + \mathsf{d}(\Phi(y), \Phi(z)) - \mathsf{d}(\Phi(x), \Phi(y)) - \theta \\ &= 2(\Phi(x)|\Phi(y))_{\Phi(z)} - \theta. \end{aligned}$$

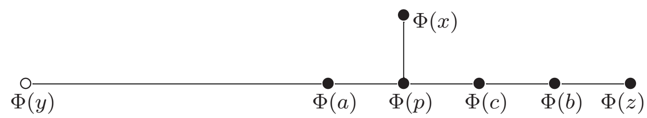
On the other hand

$$\begin{aligned} 2(x|y)_z &= \mathsf{d}(x, z) + \mathsf{d}(y, z) - \mathsf{d}(x, y) \\ &\leq \mathsf{d}(\Phi(x), \Phi(z)) + \theta + \mathsf{d}(\Phi(y), \Phi(z)) + \theta - \mathsf{d}(\Phi(x), \Phi(y)) \\ &= 2(\Phi(x)|\Phi(y))_{\Phi(z)} + 2\theta. \end{aligned}$$

□

Lemma 2.3 (Obtuse angle implies thinness)**.** *Let $\Gamma \subset \mathcal{H}$ be a geodesic line. Let x be a point of $\mathcal{H}$ and $a \in \Gamma$ be a projection of x onto Γ. For all b in Γ and c in the segment $[a, b] \subset \Gamma$, we have $(x|b)_c \leq 2\theta$.*

Proof. Let y and z be two points of Γ such that a is the middle of $[y,z] \subset \Gamma$, b is contained in $[a,z]$, and $\mathsf{d}(y,z) \geq 10\theta$. Let $\Phi\colon X \mapsto T$ be an approximation tree of (y,z,x), where X is the union of the segment $[y,z] \subset \Gamma$ and a segment $[y,x]$. Let p be the point of Γ which is mapped onto the branch point of the tripode T by Φ.



We have

$$\begin{aligned}\mathsf{d}(\Phi(x),\Phi(p)) + \theta &\geq \mathsf{d}(x,p)\\ &\geq \mathsf{d}(x,a)\\ &\geq \mathsf{d}(\Phi(x),\Phi(a))\\ &= \mathsf{d}(\Phi(x),\Phi(p)) + \mathsf{d}(\Phi(p),\Phi(a)),\end{aligned}$$

because $\Phi(a)$ is contained in the segment $\Phi([y,z])$. In particular $\mathsf{d}(\Phi(p),\Phi(a))$ is bounded by θ; from this and Corollary 2.2 we get

$$\begin{aligned}(x|b)_c &\leq (\Phi(x),\Phi(b))_{\Phi(c)} + \theta\\ &\leq \mathsf{d}(\Phi(a),\Phi(p)) + \theta,\end{aligned}$$

so that $(x|b)_c$ is bounded by 2θ. □

2.2.2. *Shortening and weak convexity.* The following lemma says that two geodesic segments are close as soon as there extremities are not too far. Lemma 3.1 makes this statement much more precise in the context of CAT(-1) spaces.

Lemma 2.4 (A little shorter, much closer. See 1.3.3 in [13])**.** *Let* $[x,y]$ *and* $[x',y']$ *be two geodesic segments of* $\mathcal{H}$ *such that*

(i) $\mathsf{d}(x,x') \leq \beta$, *and* $\mathsf{d}(y,y') \leq \beta$

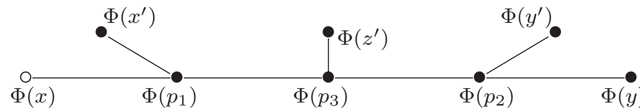
(ii) $\mathsf{d}(x,y) \geq 2\beta + 4\theta$.

Then, the geodesic segment $[x'+\beta+\theta, y'-\beta-\theta]$ *is in the* 2θ*-neighborhood of* $[x,y]$.

Proof. Let z' be a point of $[x',y']$. Consider the approximation tree $\Phi\colon X \mapsto T$ of (x,y,x',z',y'). Note p_1 (resp. p_2, resp. p_3) the points in $[x,y]$ such that $\Phi(p_1)$ (resp. $\Phi(p_2)$, resp. $\Phi(p_3)$) is the branch point of the tripode $(\Phi(x),\Phi(x'),\Phi(y)) \subset T$ (resp. $(\Phi(x),\Phi(y'),\Phi(y))$, resp. $(\Phi(x),\Phi(z'),\Phi(y))$). One easily checks that p_3 is contained in $[p_1,p_2]$ as soon as

$$\min\left(\mathsf{d}(z',x'),\mathsf{d}(z',y')\right) \geq \frac{\theta}{2} + \beta.$$

When z' is in $[x'+\beta+\theta, y'-\beta-\theta]$ this inequality is satisfied. In particular, the tree T looks as follows:



Therefore we have $\mathsf{d}(\Phi(z'),\Phi(p_3)) = (\Phi(x')|\Phi(y'))_{\Phi(z')}$. By Corollary 2.2 we have $(\Phi(x')|\Phi(y'))_{\Phi(z')} \leq \frac{\theta}{2}$. We obtain

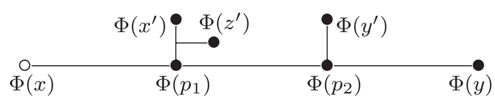
$$\mathsf{d}(z',[x,y]) \leq \mathsf{d}(z',p_3) \leq d(\Phi(z'),\Phi(p_3)) + \theta \leq 2\theta,$$

and $[x'+\beta+\theta, y'-\beta-\theta]$ is in the 2θ-neighborhood of $[x,y]$. □

The following lemma is obvious in a CAT(0) space (by convexity of the distance, see [5, p. 176]).

Lemma 2.5 (Weak convexity)**.** *Let $[x,y]$ and $[x',y']$ be two geodesic segments of $\mathcal{H}$ such that $\mathsf{d}(x,x')$ and $\mathsf{d}(y,y')$ are bounded from above by β. Then $[x',y']$ is in the $(\beta+2\theta)$-neighborhood of $[x,y]$.*

Proof. Pick $z' \in [x',y']$, and consider an approximation tree of (x,y,x',y',z'). Note $p_1 \in [x,y]$ (resp. p_2) the point such that $\Phi(p_1)$ (resp. $\Phi(p_2)$) is the branch point of the tripod $(\Phi(x),\Phi(y),\Phi(x'))$ (resp. $(\Phi(x),\Phi(y),\Phi(y'))$).



We have

$$\mathsf{d}(z',[x,y]) \leq \mathsf{d}(\Phi(z'),\Phi(p_1)) + \theta \leq (\Phi(x')|\Phi(y'))_{\Phi(z')} + \beta + \theta \leq \beta + 2\theta$$

and the lemma is proved. □

2.2.3. *Canoeing to infinity.* The following lemma comes from [13] (see §1.3.4), and says that hyperbolic canoes never make loops (see §2.3.13 in [31]). This lemma is at the heart of our proof of Theorem C.

Lemma 2.6 (Canoeing)**.** *Let $y_0,\cdots,y_n$ be a finite sequence of points in $\mathcal{H}$, such that*

(i) $\mathsf{d}(y_i,y_{i-1}) \geq 10\theta$ *for all* $1 \leq i \leq n$*;*
(ii) $(y_{i+1}|y_{i-1})_{y_i} \leq 3\theta$ *for all* $1 \leq i \leq n-1$.

Then, for all $1 \leq j \leq n$*,*

(1) $\mathsf{d}(y_0,y_j) \geq \mathsf{d}(y_0,y_{j-1}) + 2\theta$ *if* $j \geq 1$*;*
(2) $\mathsf{d}(y_0,y_j) \geq \sum_{i=1}^{j}(\mathsf{d}(y_i,y_{i-1}) - 7\theta)$*;*
(3) y_j *is* 5θ*-close to any geodesic segment* $[y_0,y_n]$*.*

Proof. The second assumption implies

$$\mathsf{d}(y_{i+1}, y_{i-1}) \geq \mathsf{d}(y_{i+1}, y_i) + \mathsf{d}(y_{i-1}, y_i) - 6\theta \tag{2.3}$$

for all $1 \leq i \leq n-1$. Together with the first assumption, we get

$$\mathsf{d}(y_{i+1}, y_{i-1}) \geq \max\{\mathsf{d}(y_i, y_{i-1}), \mathsf{d}(y_{i+1}, y_i)\} + 4\theta. \tag{2.4}$$

When applied to $i = 1$, equation (2.3) implies that

$$\mathsf{d}(y_2, y_0) \geq \mathsf{d}(y_2, y_1) - 7\theta + \mathsf{d}(y_1, y_0) - 7\theta + 8\theta.$$

In particular, properties (1) and (2) are proved for $j = 2$. Now we prove (1) and (2) by induction, assuming that these inequalities have been proved for all indices up to j (included). By property (1) and equation (2.4) (for $i = j$) we have

$$\begin{aligned} \mathsf{d}(y_0, y_j) + \mathsf{d}(y_{j-1}, y_{j+1}) &> \mathsf{d}(y_0, y_{j-1}) + 2\theta + \mathsf{d}(y_j, y_{j+1}) + 4\theta \\ &\geq \mathsf{d}(y_0, y_{j-1}) + \mathsf{d}(y_j, y_{j+1}) + 6\theta. \end{aligned}$$

This, and the hyperbolic inequality (2.1), imply

$$\mathsf{d}(y_0, y_j) + \mathsf{d}(y_{j-1}, y_{j+1}) \leq \mathsf{d}(y_0, y_{j+1}) + \mathsf{d}(y_j, y_{j-1}) + \delta.$$

Hence

$$\mathsf{d}(y_0, y_{j+1}) \geq \mathsf{d}(y_0, y_j) + \mathsf{d}(y_{j-1}, y_{j+1}) - \mathsf{d}(y_j, y_{j-1}) - \delta.$$

From equation (2.4) with $i = j$ and the inequality $4\theta - \delta \geq 2\theta$ we get property (1); then equation (2.3) with $i = j$ implies

$$\mathsf{d}(y_0, y_{j+1}) \geq \mathsf{d}(y_0, y_j) + \mathsf{d}(y_j, y_{j+1}) - (6\theta + \delta).$$

The induction hypothesis concludes the proof of (1) and (2).

We now prove assertion (3). First note that, reversing the labeling of the points, we get

(2') $\mathsf{d}(y_n, y_j) \geq \sum_{i=j+1}^{n} (\mathsf{d}(y_i, y_{i-1}) - 7\theta)$.

Let $[y_0, y_n]$ be a geodesic segment from y_0 to y_n and let $\Phi\colon X \mapsto T$ be an approximation tree for $(y_0, y_{j-1}, y_j, y_{j+1}, y_n)$, where X is the union of $[y_0, y_n]$ and three segments from y_0 to y_{j-1}, y_j, and y_{j+1}. First, we prove that

$$(\Phi(y_n)|\Phi(y_{j+1}))_{\Phi(y_0)} > (\Phi(y_n)|\Phi(y_j))_{\Phi(y_0)}. \tag{2.5}$$

By Corollary 2.2 it is sufficient to prove

$$(y_n|y_{j+1})_{y_0} > (y_n|y_j)_{y_0} + \frac{3\theta}{2}.$$

This inequality is equivalent to

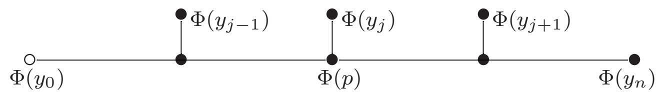
$$\mathsf{d}(y_n, y_0) + \mathsf{d}(y_{j+1}, y_0) - \mathsf{d}(y_n, y_{j+1}) > \mathsf{d}(y_n, y_0) + \mathsf{d}(y_j, y_0) - \mathsf{d}(y_n, y_j) + 3\theta,$$

that is, to

$$\mathsf{d}(y_{j+1}, y_0) + \mathsf{d}(y_n, y_j) > \mathsf{d}(y_j, y_0) + \mathsf{d}(y_n, y_{j+1}) + 3\theta,$$

and therefore follows from property (1) applied to $\mathsf{d}(y_0, y_{j+1})$ and to $\mathsf{d}(y_n, y_j)$ by reversing the labeling.

The inequality (2.5) shows that the approximation tree T has the following form:



Choose $p \in [y_0, y_n]$ such that $\Phi(p)$ is the branch point of the tripod generated by $\Phi(y_0)$, $\Phi(y_n)$, and $\Phi(y_j)$ in T. We have

$$\begin{aligned} \mathsf{d}(y_j - p) &\leq \mathsf{d}(\Phi(y_j), \Phi(p)) + \theta \\ &= (\Phi(y_{j+1})|\Phi(y_{j-1}))_{\Phi(y_j)} + \theta \\ &\leq (y_{j+1}|y_{j-1})_{y_j} + \tfrac{\theta}{2} + \theta \\ &\leq 3\theta + \tfrac{\theta}{2} + \theta. \end{aligned}$$

Thus the point y_j is 5θ-close to the segment $[y_0, y_n]$. □

2.3. **Rigidity of axis and non simplicity criterion.** Let G be a group of isometries of a δ-hyperbolic space $(\mathcal{H}, \mathsf{d})$, and g be an element of G. Our goal is to provide a criterion which implies that the normal subgroup $\langle\langle g \rangle\rangle$ generated by g in G does not coincide with G.

2.3.1. *Isometries.* Let $f \in \mathrm{Isom}(\mathcal{H})$ be an isometry of a δ-hyperbolic space $(\mathcal{H}, \mathsf{d})$. The **translation length** $L(f)$ of f is the infimum of $\mathsf{d}(x, f(x))$ where x runs over all points of $\mathcal{H}$. We denote by $\mathrm{Min}(f)$ the set of points that realize $L(f)$. Following [5, page 229] the isometry f is called

1. **elliptic** if f has a fixed point (in this case $L(f) = 0$ and $\mathrm{Min}(f)$ is the set of fixed points);
2. **hyperbolic** if $\mathsf{d}(x, f(x))$ attains a strictly positive minimum;
3. **parabolic** if the infimum $L(f)$ is not realized by any point in $\mathcal{H}$.

A hyperbolic isometry f has an **invariant axis** if there is a geodesic line Γ such that $f(\Gamma) = \Gamma$. In this case, Γ is contained in $\mathrm{Min}(f)$ and f acts as a translation of length $L(f)$ along Γ. If Γ and Γ' are two invariant axis, each of them is in the 2θ-tubular neighborhood of the other (apply Lemma 2.4). When f has an invariant axis, we denote by $\mathrm{Ax}(f)$ any invariant geodesic line, even if such a line is a priori not unique.

Remark 2.7. If $(\mathcal{H}, \mathsf{d})$ is CAT(0) and δ-hyperbolic, every hyperbolic isometry has an invariant axis (see [5, chap. II, theorem 6.8]). When $\mathcal{H}$ is a tree or a hyperbolic space $\mathbb{H}^n$, the set $\mathrm{Min}(g)$ coincides with the unique geodesic line it contains. If G is a hyperbolic group acting on its Cayley graph, any hyperbolic $g \in G$ admits a periodic geodesic line; [1] thus, there is a non trivial power g^k of g which has an invariant axis.

2.3.2. *Rigidity of geodesic lines.* If A and A' are two subsets of $\mathcal{H}$, the **intersection with precision** α of A and A' is the intersection of the tubular neighborhoods $\mathrm{Tub}_\alpha(A)$ and $\mathrm{Tub}_\alpha(A')$:

$$A \cap_\alpha A' = \left\{x \in \mathcal{H}; \mathsf{d}(x, A) \leq \alpha \text{ and } \mathsf{d}(x, A') \leq \alpha\right\}.$$

Let ε and B be positive real numbers. A subset A of $\mathcal{H}$ is (ε, B)**-rigid** if $f(A) = A$ as soon as $f \in G$ satisfies

$$\mathrm{diam}(A \cap_\varepsilon f(A)) \geq B.$$

This rigidity property involves both A and the group G. The set A is ε**-rigid** is there exists a positive number $B > 0$ such that A is (ε, B)-rigid. If $\varepsilon' < \varepsilon$ and $A \subset \mathcal{H}$ is ε-rigid, then A is also ε'-rigid (for the same constant B). The converse holds for geodesic lines (or convex sets) when ε' is not too small:

Lemma 2.8. *Let $\varepsilon > 2\theta$. If a geodesic line $\Gamma \subset \mathcal{H}$ is $(2\theta, B)$-rigid, then Γ is also $(\varepsilon, B + 6\varepsilon + 4\theta)$-rigid.*

Proof. Let $B' = B + 6\varepsilon + 4\theta$. Suppose that $\mathrm{diam}(\Gamma \cap_\varepsilon f(\Gamma)) \geq B'$. We want to show that $f(\Gamma) = \Gamma$. There exist $x, y \in \Gamma$, $x', y' \in f(\Gamma)$ such that $\mathsf{d}(x, x') \leq 2\varepsilon$, $\mathsf{d}(y, y') \leq 2\varepsilon$, $\mathsf{d}(x, y) \geq B' - 2\varepsilon$ and $\mathsf{d}(x', y') \geq B' - 2\varepsilon$. By Lemma 2.4, the segment

$$[u, v] := [x' + 2\varepsilon + \theta, y' - 2\varepsilon - \theta] \subset \Gamma$$

is 2θ-close to $[x, y]$, and we have $\mathsf{d}(u, v) \geq B$. Thus $\Gamma \cap_{2\theta} f(\Gamma)$ contains the two points u and v and has diameter greater than B. Since Γ is $(2\theta, B)$-rigid, we conclude that $\Gamma = f(\Gamma)$. □

2.3.3. *Tight elements and the normal subgroup theorem.* We shall say that an element $g \in G$ is **tight** if

- g is hyperbolic and admits an invariant axis $\mathrm{Ax}(g) \subset \mathrm{Min}(g)$;
- the geodesic line $\mathrm{Ax}(g)$ is 2θ-rigid;
- for all $f \in G$, if $f(\mathrm{Ax}(g)) = \mathrm{Ax}(g)$ then $fgf^{-1} = g$ or g^{-1}.

Note that if g is tight, then any iterate g^n, $n > 0$, is also tight. We now state the main theorem of Part A.

[1] Sketch of proof: Consider $x, y \in \mathrm{Min}(g)$. If Γ is an infinite geodesic in $\mathrm{Min}(g)$, we have $x, y \in \mathrm{Tub}_{2\theta}(\Gamma)$. By local compacity, there exists $N > 0$ such that all balls of radius 2θ in the Cayley graph $\mathcal{H}$ contains at most N vertices. Thus there is at most N^2 segments from the ball $B(x, 2\theta)$ to the ball $B(y, 2\theta)$. Take $n = (2N)!$, we obtain that g^n fixes all these segments, and therefore also any geodesic $\Gamma \in \mathrm{Min}(g)$.

Theorem 2.9 (Normal subgroup theorem)**.** *Let G be a group acting by isometries on a hyperbolic space $\mathcal{H}$. Suppose that $g \in G$ is tight, with a $(14\theta, B)$-rigid axis $Ax(g)$. Let n be a positive integer with*

$$\frac{nL(g)}{20} \geq 60\theta + 2B.$$

Then any element $h \neq \mathrm{Id}$ in the normal subgroup $\langle\langle g^n \rangle\rangle \subset G$ satisfies the following alternative: Either h is a conjugate of g^n or h is a hyperbolic isometry with translation length $L(h) > nL(g)$. In particular, if $n \geq 2$, the normal subgroup $\langle\langle g^n \rangle\rangle$ does not contain g.

Remark 2.10. Delzant in [13] gives a similar result when G is a hyperbolic group and $\mathcal{H}$ the Cayley graph of G; our strategy of proof, presented in paragraph 2.5, follows the same lines.

In the definition of a tight element, we could impose a weaker list of hypothesis. The main point would be to replace $\mathrm{Ax}(g)$ by a long quasi-geodesic segment (obtained, for example, by taking an orbit of a point x in $\mathrm{Min}(g)$); this would give a similar statement, without assuming that g has an invariant axis. Here, we take this slightly more restrictive definition because it turns out to be sufficient for our purpose, and makes the proof less technical; we refer to [13] or to the work in progress [11] for more general view points.

We now prove Theorem 2.9. First we need to introduce the notion of an admissible presentation.

2.4. **Relators, neutral segments and admissible presentations.** Let $g \in G$ be a hyperbolic isometry with an invariant axis; let $\mathrm{Ax}(g)$ be such an axis (see Remark 2.7), and $L = L(g)$ be the translation length of g. All isometries that are conjugate to g have an invariant axis: If $f = sgs^{-1}$, then $s(\mathrm{Ax}(g))$ is an invariant axis for f.

2.4.1. *Axis, relators, and neutral segments.* Let $[x, y] \subset \mathcal{H}$ be an oriented geodesic segment, the length of which is at least 20θ. The segment $[x, y]$ is **a relator** if there exists an element s in G such that $[x, y]$ is contained in the tubular neighborhood $\mathrm{Tub}_{7\theta}(s(\mathrm{Ax}(g))$. If $[x, y]$ is a relator, the conjugates $f = sgs^{-1}$ of g and $f^{-1} = sg^{-1}s^{-1}$ of g^{-1} have the same invariant axis $\Gamma = s(\mathrm{Ax}(g))$, and this axis almost contains $[x, y]$. Changing f to f^{-1} if necessary, we can assume that $\pi_\Gamma(y)$ and $f(\pi_\Gamma(x))$ lie on the same side[2] of $\pi_\Gamma(x)$ in Γ. This assumption made, the isometry f is called a **support** of $[x, y]$ (and so f^{-1} is the support of the relator $[y, x]$). The segment $[x, y]$ is **a relator of size** p/q if furthermore

[2] Since $\mathsf{d}(x, y) \geq 20\theta$, this assumption is meaningful: It does not depend on the choices of the projections of x and y on $\mathrm{Ax}(f)$.

$\mathsf{d}(x,y) \geq \frac{p}{q}L$. We say that $[x,y]$ **contains a relator of size** p/q if there is a segment $[x',y'] \subset [x,y]$ which is such a relator.

A pair of points (x,y) is **neutral** if none of the geodesic segments $[x,y]$ between x and y contains a relator of size 11/20; by a slight abuse of notation, we also say that the segment $[x,y]$ is neutral if this property holds (even if there are a priori several segments from x to y).

2.4.2. *Admissible presentations.* Let h be an element of the normal subgroup $\langle\langle g\rangle\rangle \subset G$. We can write h as a product $h = h_k h_{k-1} \cdots h_1$ where each h_j is a conjugate of g or its inverse:

$$\forall\, 1 \leq i \leq k, \quad \exists\, s_i \in G, \quad h_i = s_i g s_i^{-1} \text{ or } s_i g^{-1} s_i^{-1}.$$

By convention, each of the h_i comes with $s_i \in G$ and thus with an invariant axis $\mathrm{Ax}(h_i) = s_i(\mathrm{Ax}(g))$.

We fix a base point $x_0 \in \mathcal{H}$, and associate three sequences of points (a_i), (b_i), and (x_i), $1 \leq i \leq k$, to the given base point x_0 and the factorization of h into the product of the $h_i = s_i g^{\pm 1} s_i^{-1}$. The definition requires two steps. First, for all $1 \leq i \leq k$,

- x_i is equal to $h_i(x_{i-1})$; in particular $x_k = h(x_0)$;
- a_i is the projection of x_{i-1} on the geodesic $\mathrm{Ax}(h_i)$;
- b_i is equal to $h_i(a_i)$; in particular, both a_i and b_i are on $\mathrm{Ax}(h_i)$.

This is the first step toward the definition of the three sequences of points.

Remark 2.11. If the translation length L is larger than 480θ then, after a slight modification of the a_i and b_i, the following properties are satisfied:

(1) Each $[a_i, b_i]$ is a subsegment of $\mathrm{Ax}(h_i)$ of length at least $\frac{19}{20}L$;
(2) $(x_{i-1}|b_i)_{a_i} \leq 2\theta$ and $(a_i|x_i)_{b_i} \leq 2\theta$;
(3) all segments $[x_{i-1}, a_i]$ and $[b_i, x_i]$ have length at least 10θ (for $1 \leq i \leq k$).

To obtain property (3) when $[x_{i-1}, a_i]$ is too small, we translate a_i towards b_i on a small distance (at most 12θ), and similarly we translate b_i towards a_i. The first property remains true as long as $\frac{L}{20} \geq 24\theta$. The second one follows from Lemma 2.3.

In a second step, we use this remark to move the points a_i and b_i along the geodesic line $\mathrm{Ax}(h_i)$, and to define new sequences (a_i), (b_i), (x_i) that satisfy the three properties (1), (2), and (3). The definition of (x_i) has not been changed during this second step. The point a_i is no more equal to a projection of x_i on $\mathrm{Ax}(h_i)$, but is at distance at most 12θ from such a projection; the point b_i is not equal to $h_i(a_i)$, but the distance between these two points is less than 24θ.

We say that $h_k \cdots h_1$ is an **admissible presentation** of h (with respect to the base point x_0) if one can construct the three sequences as above with the additional property that all pairs (x_{i-1}, a_i), $1 \leq i \leq k$, are neutral.

Lemma 2.12. *Let x_0 be a base point in $\mathcal{H}$. Let g be an element of G, and h be an element of the normal subgroup generated by g. Assume that $28\theta < \frac{L}{20}$. Then h admits at least one admissible presentation.*

Remark 2.13. Thanks to this lemma, the following property can be added to the first three properties listed in Remark 2.11:

(4) All pairs (x_{i-1}, a_i) and (b_i, x_i) are neutral.

Proof of Lemma 2.12. Start with a decomposition $h = h_k \cdots h_1$ with $h_i = s_i g^{\pm 1} s_i^{-1}$ and construct the sequences of points (a_i), (b_i), and (x_i) as above. Let $\mathcal{I}$ be the set of indices $1 \leq i \leq k$ such that (x_{i-1}, a_i) is not neutral. Suppose $\mathcal{I}$ is not empty (otherwise we are done). Our goal is to modify the construction in order to get a new decomposition of h which is admissible.

Changing the decomposition.– Pick $i \in \mathcal{I}$. Then there are two geodesic segments $[y, z] \subset [x_{i-1}, a_i]$ (with $y \in [x_{i-1}, z]$) and a conjugate $f = sgs^{-1}$ of g such that $[y, z]$ is 7θ-close to $\mathrm{Ax}(f) = s(\mathrm{Ax}(g))$ and $\mathsf{d}(y, z) \geq \frac{11}{20}L$; we fix such a segment $[y, z]$ with a maximal length. Let y', z' be projections of y and z on $\mathrm{Ax}(f)$; we have $\mathsf{d}(y, y') \leq 7\theta$ and $\mathsf{d}(z, z') \leq 7\theta$. If y' is contained in the segment $[f(y'), z'] \subset \mathrm{Ax}(f)$, we change f into f^{-1}. Replace h_i by the product of three conjugates of g

$$(h_i f^{-1} h_i^{-1}) h_i f.$$

This gives rise to a new decomposition of h as a product of conjugates of $g^{\pm 1}$, hence to new sequences of points. Let a'_i be a projection of x_{i-1} on $\mathrm{Ax}(f)$. Concerning the sequence (x_i), we have two new points $x'_i = f(x_{i-1})$ and $x'_{i+1} = h_i(x'_i)$; the point $x'_{i+2} = h_i f^{-1} h_i^{-1}(x'_{i+1})$ is equal to the old point $x_i = h_i(x_{i-1})$. Thus, the neutral pair (x_{i-1}, a_i) disappears and is replaced by three new pairs (x_{i-1}, a'_i), (x'_i, a'_{i+1}), and (x'_{i+1}, a'_{i+2}) (see Figure 1). Note that if any of these segments is small then it is automatically neutral; so without loss of generality we can assume that there was no need to move a'_i, a'_{i+1} or a'_{i+2} as in Remark 2.11.

Two estimates.– Our first claim is

(1) $\mathsf{d}(x_{i-1}, a'_i) = \mathsf{d}(x'_{i+1}, a'_{i+2}) \leq \mathsf{d}(x_{i-1}, a_i) - \frac{10}{20}L.$

To prove it, we write

$$\mathsf{d}(x_{i-1}, a_i) \geq \mathsf{d}(x_{i-1}, y) + \mathsf{d}(y, z) \geq \mathsf{d}(x_{i-1}, y) + \tfrac{11}{20}L,$$

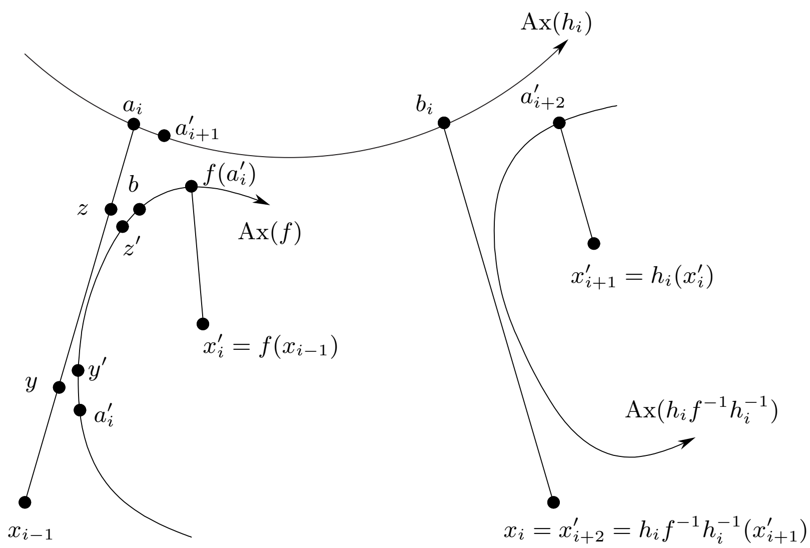


FIGURE 1.

and

$$\mathsf{d}(x_{i-1}, a'_i) \leq \mathsf{d}(x_{i-1}, y') \leq \mathsf{d}(x_{i-1}, y) + \mathsf{d}(y, y') \leq \mathsf{d}(x_{i-1}, y) + 7\theta; \tag{2.6}$$

claim (1) follows because $7\theta < \frac{L}{20}$.

Our second claim is

(2) $\mathsf{d}(x'_i, a'_{i+1}) \leq \mathsf{d}(x_{i-1}, a_i) - \frac{1}{20}L$.

To prove it, note b the projection of a_i on $\mathrm{Ax}(f)$. Similarly as (2.6)we have

$$\mathsf{d}(b, a_i) \leq \mathsf{d}(z, a_i) + 7\theta. \tag{2.7}$$

There are two cases, according to the position of $f(a'_i)$ with respect to the segment $[a'_i, b] \subset \mathrm{Ax}(f)$. If $f(a'_i)$ is in this segment, then

$$\begin{aligned} d(x'_i, a_i) &\leq d(f(x_{i-1}), f(a'_i)) + d(f(a'_i), b) + d(b, a_i) \\ &\leq d(x_{i-1}, a'_i) + d(a'_i, b) - L + d(b, a_i). \end{aligned}$$

Applying Lemma 2.3 for the triangles x_{i-1}, y', a'_i and a_i, z', b, and the inequalities $\mathsf{d}(y, y')$, $\mathsf{d}(z, z') \leq 7\theta$, we get $\mathsf{d}(x_{i-1}, y) = \mathsf{d}(x_{i-1}, a'_i) + \mathsf{d}(y, a'_i)$ up to an error of 11θ, and similarly $\mathsf{d}(a_i, z) = \mathsf{d}(a_i, b) + \mathsf{d}(z, b)$ up to 11θ. Hence, we get

$$\begin{aligned} d(x'_i, a_i) &\leq d(x_{i-1}, y) + d(y, z) + d(z, a_i) - L + 22\theta \\ &\leq d(x_{i-1}, a_i) - L + 22\theta. \end{aligned}$$

This concludes the proof of (2) in the first case, because $-L + 22\theta < -L/20$.

The second case occurs when b is in the segment $[a_i', f(a_i')] \subset \mathrm{Ax}(f)$ (see Figure 1). In this case we have

$$\begin{aligned} \mathsf{d}(x_i', a_i) &\leq \mathsf{d}(f(x_{i-1}), f(a_i')) + \mathsf{d}(f(a_i'), b) + \mathsf{d}(b, a_i) \\ &\leq \mathsf{d}(x_{i-1}, z) - \tfrac{11}{20}L + 7\theta + L - \mathsf{d}(a_i', b) + d(z, a_i) + 7\theta, \end{aligned}$$

and thus

$$\mathsf{d}(x_i', a_i) \leq \mathsf{d}(x_{i-1}, a_i) - \mathsf{d}(a_i', b) + \tfrac{9}{20}L + 14\theta. \tag{2.8}$$

On the other hand, the triangular inequality implies

$$\begin{aligned} \mathsf{d}(x_{i-1}, a_i') + \mathsf{d}(a_i', b) + \mathsf{d}(b, a_i) &\geq \mathsf{d}(x_{i-1}, a_i) \\ &= \mathsf{d}(x_{i-1}, y) + \mathsf{d}(y, z) + \mathsf{d}(z, a_i). \end{aligned}$$

and so

$$\mathsf{d}(a_i', b) \geq \tfrac{11}{20}L + \mathsf{d}(x_{i-1}, y) - \mathsf{d}(x_{i-1}, a_i') + \mathsf{d}(z, a_i) - \mathsf{d}(b, a_i).$$

Using the inequalities (2.6) and (2.7) we obtain

$$\mathsf{d}(a_i', b) \geq \tfrac{11}{20}L - 14\theta.$$

Finally inequality (2.8) gives

$$\mathsf{d}(x_i', a_{i+1}') \leq \mathsf{d}(x_i', a_i) \leq \mathsf{d}(x_{i-1}, a_i) - \tfrac{2}{20}L + 28\theta$$

hence claim (2) because $28\theta < \frac{L}{20}$.

Conclusion.– Since the modification is local, and does not change the points a_j for $j \neq i$, we can perform a similar replacement for all indices i in $\mathcal{I}$. We obtain a new presentation for h and a new list of bad indices $\mathcal{I}'$. Either this new list is empty, or by the two estimates (claims (1) and (2)) the maximum of the lengths $\mathsf{d}(x_{j-1}, a_j)$ over all the non neutral pairs (x_{j-1}, a_j), $j \in \mathcal{I}'$, drops at least by $\frac{1}{20}L$. By induction, after a finite number of such replacements, we obtain an admissible presentation. □

The following lemma provides a useful property of admissible presentations with a minimum number of factors h_i.

Lemma 2.14. *Let $h = h_k \cdots h_1$ be an admissible presentation with base point x_0. If there exist two indices $j > i$ such that $h_j = h_i^{-1}$, then h admits an admissible presentation with base point x_0 and only $k - 2$ factors.*

Proof. We may assume that $j \geq i + 2$, otherwise the simplification is obvious. The decomposition for h is then

$$h = h_k \cdots h_{j+1} h_i^{-1} h_{j-1} \cdots h_{i+1} h_i h_{i-1} \cdots h_1.$$

Note that i can be equal to 1, and j can be equal to k. We have a sequence of triplets (a_i, b_i, x_i), $i = 1, \cdots, k$, associated with this presentation and with the base point x_0 Then we claim that

$$h = h_k \cdots h_{j+1}(h_i^{-1}h_{j-1}h_i)(h_i^{-1}h_{j-2}h_i)\cdots(h_i^{-1}h_{i+1}h_i)h_{i-1}\cdots h_1$$

is another admissible presentation with base point x_0 and with $k-2$ factors. Indeed the sequence of $k-2$ triplets associated with this new presentation is

$$(a_1, b_1, x_1), \cdots, (a_{i-1}, b_{i-1}, x_{i-1}), (h_i^{-1}(a_{i+1}), h_i^{-1}(b_{i+1}), h_i^{-1}(x_{i+1})), \cdots,$$
$$(h_i^{-1}(a_{j-1}), h_i^{-1}(b_{j-1}), h_i^{-1}(x_{j-1})), (a_{j+1}, b_{j+1}, x_{j+1}), \cdots, (a_k, b_k, x_k)$$

and one checks that all relevant segments are neutral because they come from neutral segments of the previous presentation. □

2.5. **Proof of the normal subgroup theorem.** Let g be a tight element of G. Choose $B > 0$ a constant associated with the 14θ-rigidity of the axis of g. Choose $n \geq 1$ such that

$$\frac{n}{20}L(g) \geq 60\theta + 2B.$$

Let $L = nL(g)$ be the translation length of g^n. Note that $g^n \in G$ is tight, with a $(14\theta, B)$-rigid axis $\mathrm{Ax}(g^n) = \mathrm{Ax}(g)$. Let h be a non trivial element of the normal subgroup $\langle\langle g^n \rangle\rangle$. Our goal is to prove $L(h) \geq L$, with equality if and only if h is conjugate to g^n.

Pick a base point x_0 such that $\mathsf{d}(x_0, h(x_0)) \leq L(h) + \theta$. Lemma 2.12 applied to g^n implies the existence of an admissible presentation with respect to the base point x_0:

$$h = h_m \circ \cdots \circ h_1, \quad h_i = s_i g^{\pm n} s_i^{-1}.$$

We assume that m is minimal among all such choices of base points and admissible presentations. Lemma 2.14 implies that h_j is different from h_i^{-1} for all $1 \leq i < j \leq m$.

Let (a_i), (b_i), (x_i) be the three sequences of points defined in §2.4; they satisfy the properties (1) to (4) listed in Remarks 2.11 and 2.13. Since the constructions below are more natural with segments than pair of points, and $\mathcal{H}$ is not assumed to be uniquely geodesic, we choose geodesic segments between the points x_i, as well as geodesic segments $[a_i, b_i] \subset \mathrm{Ax}(h_i)$.

We now introduce the following definition in order to state the key Lemma 2.15 (compare with Lemma 2.4 in [13]). A sequence of points $(c_{-1}, c_0, \cdots, c_k, c_{k+1})$ in $\mathcal{H}$, and of a chain of segments $[c_i, c_{i+1}]$, $-1 \leq i \leq k$, is a **configuration of order** $k \geq 1$ for the segment $[x_0, x_j]$ if

(i) $x_0 = c_{-1}$ and $x_j = c_{k+1}$;

(ii) For all $0 \le i \le k$, and all $a \in [c_i, c_{i+1}]$, we have $(c_{i-1}|a)_{c_i} \le 3\theta$. In particular $(c_{i-1}|c_{i+1})_{c_i} \le 3\theta$;
(iii) For all $0 \le i \le k+1$ we have $\mathsf{d}(c_{i-1}, c_i) \ge 10\theta$;
(iv) For all $0 \le i \le k$ the segment $[c_i, c_{i+1}]$ is either neutral or a relator, with the following rules:
  (iv-a) There are never two consecutive neutral segments;
  (iv-b) The last segment $[c_k, c_{k+1}] = [c_k, x_j]$ is neutral;
  (iv-c) The second segment $[c_0, c_1]$ is a relator of size 18/20 if $[c_1, c_2]$ is neutral (this is always the case when $k = 1$), and of size 17/20 otherwise;
  (iv-d) For the other relators $[c_{i-1}, c_i]$, with $i > 1$, the size is 5/20 when $[c_i, c_{i+1}]$ is neutral and 4/20 otherwise;
(v) For all $0 \le i \le k$, if $[c_i, c_{i+1}]$ is a relator, then there is an index l with $1 \le l \le j$, such that h_l is the support of the relator $[c_i, c_{i+1}]$.

Note that properties (iv) and (v) do not concern the initial segment $[x_0, c_0]$, and that the size p/q of a relator $[c_i, c_{i+1}]$, $0 \le i \le k-1$, is equal to 18/20, 17/20, 5/20, or 4/20; moreover, this size is computed with respect to $L = L(g^n)$, and thus the minimum length of a relator $[c_i, c_{i+1}]$ is bounded from below by $4(60\theta + 2B)$.

Lemma 2.15. *For each $j = 1, \cdots, m$, there exists $k \ge 1$ such that the segment $[x_0, x_j]$ admits a configuration of order k. Moreover if $j \ge 2$ and $k = 1$ then the first segment $[x_0, c_0]$ of the configuration has length at least $\frac{3}{20}L$.*

Proof. The proof is by induction on j, and uses the four properties that are listed in Remarks 2.11 and 2.13; we refer to them as properties (1), (2), (3), and (4). Note that (2) and (3) enable us to apply Lemma 2.6; similarly, properties (ii) and (iii) for a configuration of order k show that Lemma 2.6 can be applied to the sequence of points in such a configuration.

Initialization.– When $j = 1$, we take $c_0 = a_1, c_1 = b_1$ and get a configuration of order 1. Indeed, by property (2), we have

$$(x_0|b_1)_{a_1} \le 2\theta \text{ and } (a_1|x_1)_{b_1} \le 2\theta,$$

and, by property (1), $[c_0, c_1]$ is a relator of size 19/20 (it is a subsegment of $\mathrm{Ax}(h_1)$). The segments $[x_0, c_0]$ and $[x_1, c_1]$ are neutral (property (4)) and of length at least 10θ (property (3)).

Suppose now that $[x_0, x_j]$ admits a configuration of order k. We want to find a configuration of order k' for $[x_0, x_{j+1}]$. As we shall see, the proof provides a configuration of order $k' = 1$ in one case, and of order $k' \leq k+2$ in the other case.

Six preliminary facts.– Consider the approximation tree (Φ, T) of $(x_j,\ x_0,\ x_{j+1})$. We choose $p \in [x_j, x_0]$ such that $\Phi(p) \in T$ is the branch point of the tripod T (with $p = x_j$ if T is degenerate). Let a (resp. b) be a projection of a_{j+1} (resp. b_{j+1}) on the segment $[x_j, x_{j+1}]$. By assertion (3) in Lemma 2.6 we have $\mathsf{d}(a, a_{j+1}) \leq 5\theta$ and $\mathsf{d}(b, b_{j+1}) \leq 5\theta$. Thus, the subsegment $[a, b] \subset [x_j, x_{j+1}]$ is contained in $[x_j + 5\theta, x_{j+1} - 5\theta]$; moreover, by Lemma 2.5, $[a, b]$ is a relator with support h_{j+1}.

For all $i \leq k+1$, note c'_i a projection of c_i on the segment $[x_0, x_j]$. Note that, by Lemma 2.6, we have $\mathsf{d}(c_i, c'_i) \leq 5\theta$, and the points c'_i, $-1 \leq i \leq k+1$, form a monotonic sequence of points from x_0 to x_j along the geodesic segment $[x_0, x_j]$.

Let S_i be the interval of T defined by $S_i = [\Phi(a), \Phi(b)] \cap [\Phi(c'_i), \Phi(c'_{i+1})]$. The preimages of S_i by Φ are two intervals $I \subset [a, b]$ and $I' \subset [c'_i, c'_{i+1}]$ such that

$$\Phi(I) = \Phi(I') = S_i.$$

Fact 2.16. *If* $[c_i, c_{i+1}]$ *is neutral then* $\operatorname{diam}(S_i) \leq \frac{12}{20} L$.

Since Φ is an isometry along the geodesic segments $[x_j, x_0]$ and $[x_j, x_{j+1}]$, we only have to prove $\operatorname{diam}(I') \leq \frac{12}{20} L$. By Lemma 2.5 we know that $[c_i, c_{i+1}] \subset \mathrm{Tub}_{7\theta}([c'_i, c'_{i+1}])$. Thanks to the triangular inequality, we can choose $J \subset [c_i, c_{i+1}]$ such that $J \subset \mathrm{Tub}_{7\theta}(I')$ and $\operatorname{diam}(J) \geq \operatorname{diam}(I') - 14\theta$. The properties of the approximation tree imply $I' \subset \mathrm{Tub}_{\theta}(I)$, and $I \subset \mathrm{Tub}_{7\theta}(\mathrm{Ax}(h_{j+1}))$. Thus $J \subset \mathrm{Tub}_{15\theta}(\mathrm{Ax}(h_{j+1}))$. Applying Lemma 2.4 (with $\beta = 16\theta$) and shortening J by 32θ (16θ at each end), we obtain $\operatorname{diam}(J) \geq \operatorname{diam}(I') - 46\theta$ and $J \subset \mathrm{Tub}_{2\theta}(\mathrm{Ax}(h_{j+1}))$. Thus J is a relator contained in the neutral segment $[c_i, c_{i+1}]$, hence

$$\frac{11}{20} L \geq \operatorname{diam}(J) \geq \operatorname{diam}(I') - 46\theta, \quad \text{and therefore} \quad \frac{12}{20} L \geq \operatorname{diam}(I').$$

Fact 2.17. *If* $[c_i, c_{i+1}]$ *is a relator then* $\operatorname{diam}(S_i) \leq B$.

By property (v), there exists an index l, with $1 \leq l \leq j$, such that $[c_i, c_{i+1}] \subset \mathrm{Tub}_{7\theta}(\mathrm{Ax}(h_l))$. Since $I' \subset \mathrm{Tub}_{7\theta}([c_i, c_{i+1}])$, we get

$$I' \subset \mathrm{Tub}_{14\theta}(\mathrm{Ax}(h_l)).$$

On the other hand $I' \subset \mathrm{Tub}_{\theta}(I)$ and $I \subset \mathrm{Tub}_{7\theta}(\mathrm{Ax}(h_{j+1}))$, thus

$$I' \subset \mathrm{Tub}_{8\theta}(\mathrm{Ax}(h_{j+1})).$$

If $\operatorname{diam}(I') > B$, the rigidity assumption shows that h_{j+1} and h_l share the same axis, with opposite orientations; since g is tight, we obtain $h_{j+1} = h_l^{-1}$. By Lemma 2.14 this contradicts the minimality of the presentation of h. This proves that $\operatorname{diam}(I') \leq B$.

Fact 2.18. *Suppose* $[c_i, c_{i+1}]$ *is a relator. Then*

$$\begin{aligned} \mathrm{diam}([\Phi(x_j), \Phi(a)] \cap [\Phi(c'_i), \Phi(c'_{i+1})]) &\leq \tfrac{12}{20}L. \\ \mathrm{diam}([\Phi(b), \Phi(x_{j+1})] \cap [\Phi(c'_i), \Phi(c'_{i+1})]) &\leq \tfrac{12}{20}L. \end{aligned}$$

The two inequalities are similar; we prove only the first one. By property (v), there exists an index l, with $1 \leq l \leq j$, such that h_l is the support of $[c_i, c_{i+1}]$. Let $K \subset [a, x_j]$ and $K' \subset [c'_i, c'_{i+1}]$ be two intervals such that

$$\Phi(K) = \Phi(K') = [\Phi(x_j), \Phi(a)] \cap [\Phi(c'_i), \Phi(c'_{i+1})].$$

We want to prove $\mathrm{diam}(K) \leq \frac{12}{20}L$. Applying the triangular inequality, we can choose $J \subset [a_{j+1}, x_j]$ such that $J \subset \mathrm{Tub}_{7\theta}(K)$ and $\mathrm{diam}(J) \geq \mathrm{diam}(K) - 14\theta$. Now we have $K \subset \mathrm{Tub}_\theta(K')$, $K' \subset \mathrm{Tub}_{7\theta}([c_i, c_{i+1}])$ and $[c_i, c_{i+1}] \subset \mathrm{Tub}_{7\theta}(\mathrm{Ax}(h_l))$. Thus $J \subset \mathrm{Tub}_{22\theta}(\mathrm{Ax}(h_l))$. Applying Lemma 2.4 with $\beta = 23\theta$, we shorten J by 46θ (23θ on each end) and obtain

$$J \subset \mathrm{Tub}_{2\theta}(\mathrm{Ax}(h_l)) \quad \text{and} \quad \mathrm{diam}(J) \geq \mathrm{diam}(K) - 60\theta.$$

The admissibility condition implies that $J \subset [a_{j+1}, x_j]$ is neutral, and therefore

$$\tfrac{11}{20}L \geq \mathrm{diam}(J) \geq \mathrm{diam}(K) - 60\theta, \quad \text{so that} \quad \tfrac{12}{20}L \geq \mathrm{diam}(K).$$

Fact 2.19. *The segment* $[\Phi(b), \Phi(a)]$ *is not contained in the segment* $[\Phi(c'_0), \Phi(x_j)]$.

Let us prove this by contradiction, assuming $[\Phi(b), \Phi(a)] \subset [\Phi(c'_0), \Phi(x_j)]$. By Property (iv-a) and Fact 2.16, $[\Phi(b), \Phi(a)]$ intersects at least one segment $[\Phi(c'_i), \Phi(c'_{i+1})]$ for which $[c_i, c_{i+1}]$ is a relator; Fact 2.17 implies that $[\Phi(c'_i), \Phi(c'_{i+1})]$ is not contained in $[\Phi(a), \Phi(b)]$ (it must intersect the boundary points of $[\Phi(a), \Phi(b)]$). From this follows that $[\Phi(a), \Phi(b)]$ intersects at most two relators and one neutral segment. Facts 2.16 and 2.17 now give the contradictory inequality $2B + \frac{12}{20}L \geq \frac{19}{20}L - 10\theta$.

Fact 2.20. *The segment* $[\Phi(c'_0), \Phi(c'_1)]$ *is not contained in the segment* $[\Phi(x_{j+1}), \Phi(x_j)]$.

Since $[a_{j+1}, x_j]$ is neutral and $[c_0, c_1]$ is a relator of size $\geq 17/20$, the segment $[\Phi(c'_0), \Phi(c'_1)]$ is not contained in $[\Phi(), \Phi(x_j)]$. Assume that $[\Phi(c'_0), \Phi(c'_1)] \subset [\Phi(x_{j+1}), \Phi(x_j)]$, apply Fact 2.19 and then Fact 2.17 and Fact 2.18; this gives the contradictory inequality $B + \frac{12}{20}L \geq \frac{17}{20}L - 10\theta$.

These last two facts imply $\Phi(b) \in [\Phi(x_{j+1}), \Phi(p)]$ and $\Phi(c'_0) \in [\Phi(x_0), \Phi(p)]$. Moreover, with this new property in mind, the proofs of Facts 2.19 and 2.20 give:

Fact 2.21. *The segment* $[\Phi(b), \Phi(p)]$ *has length at least* $\frac{6}{20}L$*, and the segment* $[\Phi(c_0'), \Phi(p)]$ *has length at least* $\frac{4}{20}L$*.*

In particular the tripod T is not degenerate at $\Phi(x_0)$ nor at $\Phi(x_{j+1})$.

The induction.– We now come back to the proof by induction. We distinguish two cases, with respect to the position of $\Phi(a)$ relatively to the branch point $\Phi(p)$.

• **First case.** $(\Phi(b)|\Phi(x_0))_{\Phi(a)} \leq \frac{1}{20}L - 12\theta$.

In other words we assume that either $\Phi(a) \in [\Phi(b), \Phi(p)]$, or $\Phi(a)$ is close to $\Phi(p)$. Note that this includes the situation where $\Phi(x_j)$ is a degenerate vertex of T. By Fact 2.21, the situation is similar to Figure 2. Since the distance between a_{j+1} (resp. b_{j+1}) and a (resp. b) is at most 5θ, the triangular inequality and Corollary 2.2 imply

$$\begin{aligned}(b_{j+1}|x_0)_{a_{j+1}} &\leq (b|x_0)_a + (10\theta + 5\theta + 5\theta)/2 \\ &\leq (\Phi(b)|\Phi(x_0))_{\Phi(a)} + \theta + 10\theta,\end{aligned}$$

and the assumption made in this first case gives

$$(b_{j+1}|x_0)_{a_{j+1}} \leq \frac{1}{20}L - \theta. \tag{2.9}$$

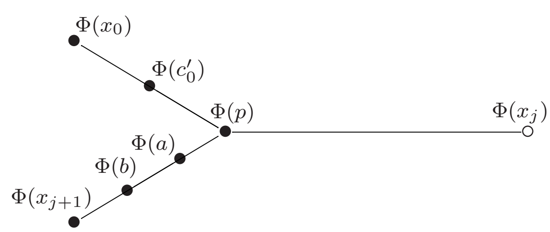


FIGURE 2. The point $\Phi(a)$ could also be on $[\Phi(p), \Phi(x_j)]$, but not far from $\Phi(p)$.

We now consider another approximation tree $\Psi \colon X \mapsto T'$, for the list (a_{j+1}, b_{j+1}, x_0), and choose a point $q \in [a_{j+1}, b_{j+1}]$ such that $\Psi(q)$ is the branch point of T' (see Figure 3).

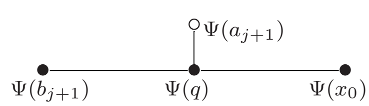


FIGURE 3.

We then define a new configuration of points $\hat{c}_{-1} = x_0$, $\hat{c}_0 = q$, $\hat{c}_1 = b_{j+1}$, $\hat{c}_2 = x_{j+1}$, and we show that this defines a configuration of order 1 for $[x_0, x_{j+1}]$. Using inequality (2.9), the fact that Ψ is an isometry along $[a_{j+1}, b_{j+1}]$, and Corollary 2.2, we obtain

$$\mathsf{d}(q, b_{j+1}) \geq \tfrac{19}{20}L - (b_{j+1}|x_0)_{a_{j+1}} - \theta \geq \tfrac{18}{20}L.$$

Thus $[q, b_{j+1}]$ is a relator of size 18/20 with support h_{j+1}: This gives properties (iv-c) and (v).

By Lemma 2.3 we have $(q|x_{j+1})_{b_{j+1}} \leq 2\theta$, and by Corollary 2.2 we have $(x_0|b_{j+1})_{p_1} \leq \theta$, so we obtain property (ii); property (iii) follows from the definition of b_{j+1} (property (3) in Remark 2.11). Thus, $(\hat{c}_j)_{-1\leq j\leq 2}$ is a configuration of order 1 for the segment $[x_0, x_{j+1}]$.

Moreover, we have

$$\begin{aligned} d(x_0, q) &= (\Psi(a_{j+1})|\Psi(b_{j+1}))_{\Psi(x_0)} \\ &\geq (a_{j+1}|b_{j+1})_{x_0} - \theta \\ &\geq (a|b)_{x_0} - 11\theta \\ &\geq (\Phi(a)|\Phi(b))_{\Phi(x_0)} - 12\theta \\ &\geq \mathsf{d}(\Phi(x_0), \Phi(p)) - 12\theta. \end{aligned}$$

Fact 2.21 then gives

$$\mathsf{d}(x_0, q) \geq \tfrac{4}{20}L - 12\theta \geq \tfrac{3}{20}L$$

so we obtain the second assertion in Lemma 2.15.

• **Second case.** $(\Phi(b)|\Phi(x_0))_{\Phi(a)} > \frac{1}{20}L - 12\theta$.

Let i be the index such that $\Phi(p) \in [\Phi(c'_i), \Phi(c'_{i+1})]$. This index is uniquely defined if we impose $\Phi(c'_i) \neq \Phi(p)$. The assumption implies $\Phi(a) \in [\Phi(p), \Phi(x_j)]$ and $\mathsf{d}(\Phi(p), \Phi(a)) > \frac{1}{20}L - 12\theta$. By Fact 2.21 again, the situation is similar to Figure 4. We distinguish two subcases according to the nature of $[c_i, c_{i+1}]$.

Second case - first subcase. Assume $[c_i, c_{i+1}]$ is neutral. Then if $i < k$ the segment $[c_{i+1}, c_{i+2}]$ is a relator (if $i = k$ the following discussion is even easier). If $\mathsf{d}(\Phi(p), \Phi(c'_{i+1})) \leq 34\theta$ then Fact 2.17 implies $\mathsf{d}(\Phi(p), \Phi(a)) \leq B + 34\theta \leq \frac{1}{20}L - 12\theta$. This contradicts the assumption of the second case; as a consequence,

$$\min\{\mathsf{d}(\Phi(p), \Phi(c'_{i+1})), \mathsf{d}(\Phi(p), \Phi(a))\} > 34\theta. \tag{2.10}$$

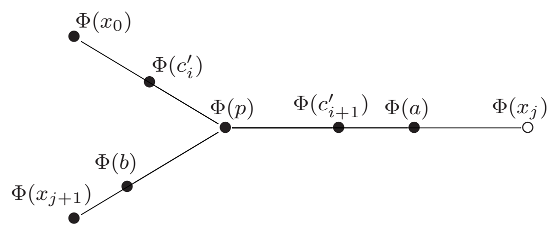


FIGURE 4. On this picture, $\Phi(c'_{i+1})$ is in between $\Phi(p)$ and $\Phi(a)$; another possible case would be $\Phi(a) \in [\Phi(p), \Phi(c'_{i+1})]$.

Consider now the approximation tree $\Psi\colon X \mapsto T'$ of $(c_{i+1},\ c_i,\ a_{j+1},\ b_{j+1})$. We have

$$\begin{aligned}(\Psi(c_{i+1})|\Psi(a_{j+1}))_{\Psi(b_{j+1})} &\geq (c'_{i+1}|a)_b - 15\theta - \theta\\ &\geq (\Phi(c'_{i+1})|\Phi(a))_{\Phi(b)} - 17\theta\\ &\geq \mathsf{d}(\Phi(b), \Phi(p)) - 17\theta\\ &\quad + \min\{\mathsf{d}(\Phi(p), \Phi(c'_{i+1})), \mathsf{d}(\Phi(p), \Phi(a))\}\end{aligned}$$

and

$$\begin{aligned}(\Psi(c_i)|\Psi(a_{j+1}))_{\Psi(b_{j+1})} &\leq (c'_i|a)_b + 15\theta + \theta\\ &\leq (\Phi(c'_i)|\Phi(a))_{\Phi(b)} + 17\theta\\ &\leq \mathsf{d}(\Phi(b), \Phi(p)) + 17\theta.\end{aligned}$$

By (2.10) we get

$$(\Psi(c_{i+1})|\Psi(a_{j+1}))_{\Psi(b_{j+1})} > (\Psi(c_i)|\Psi(a_{j+1}))_{\Psi(b_{j+1})}.$$

Thus we obtain the pattern depicted on Figure 5, where q is a point of $[c_i, c_{i+1}]$ which is mapped to the branch point of T'. Note that $[c_i, q] \subset [c_i, c_{i+1}]$ is again a neutral segment.

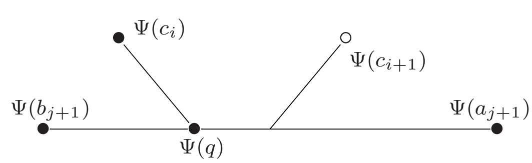


FIGURE 5.

Since the point q is 4θ-close to $\mathrm{Ax}(h_{j+1})$, the segment $[q, b_{j+1}]$ is a relator with support h_{j+1}. We have

$$\begin{aligned} \mathsf{d}(b_{j+1}, q) &\geq \mathsf{d}(\Psi(b_{j+1}), \Psi(q)) \\ &\geq (c_i|a_{j+1})_{b_{j+1}} - \theta \\ &\geq (c_i'|a)_b - 15\theta - \theta \\ &\geq \mathsf{d}(\Phi(b), \Phi(p)) - 17\theta. \end{aligned}$$

Thus by Fact 2.21 we see that $[q, b_{j+1}]$ is a relator of size 5/20, and the same conclusion holds even if one moves q along the segment $[q, b_{j+1}]$ by at most 12θ. The inequalities $(x_{j+1}|a_{j+1})_{b_{j+1}} \leq 2\theta$ and $(a_{j+1}|b_{j+1})_q \leq \theta$ imply that $(x_{j+1}|q)_{b_{j+1}} \leq 3\theta$.

In this first subcase, we define the new configuration $(\hat{c}_l)$ by $\hat{c}_l = c_l$ for l between -1 and i, and by $\hat{c}_{i+1} = q$, $\hat{c}_{i+2} = b_{j+1}$ and $\hat{c}_{i+3} = x_{j+1}$; this defines a configuration of order $i+2$ for the segment $[x_0, x_{j+1}]$.

By construction, properties (i), (ii), and (v) are satisfied, and property (iii) is obtained after translating q along $[q, b_{j+1}]$ on a distance less than 12θ. This does not change the fact that $[q, b_{j+1}]$ is a relator of size 5/20. Since the new configuration is obtained from the previous one by cutting it after c_i and then adding a relator $[q, b_{j+1}]$ and a neutral segment $[b_{j+1}, x_{j+1}]$, property (iv) is also satisfied.

Second case - second subcase. Assume $[c_i, c_{i+1}]$ is a relator. By Fact 2.17 the segment $[\Phi(p), \Phi(c_{i+1})]$ has length at most B, and $[c_{i+1}, c_{i+2}]$ must be neutral (otherwise again $[\Phi(p), \Phi(a)]$ would be too small). Thus the relator $[c_i, c_{i+1}]$ has size 5/20 (or 18/20 if $i = 0$). We consider the approximation tree $\Psi\colon X \mapsto T$ of $(c_{i+1},\ c_i,\ c_{i+2},\ a_{j+1},\ b_{j+1})$. We obtain one of the situations depicted on Fig. 6, where now $q \in [c_{i+1}, c_i]$ or $[c_{i+1}, c_{i+2}]$.

In case (a), $[c_{i+1}, q]$ is small, and therefore neutral. We are in a case similar to the first subcase and to Figure 5. We still have

$$\mathsf{d}(b_{j+1}, q) \geq \mathsf{d}(\Phi(b), \Phi(p)) - 17\theta,$$

and one can define a new sequence $(\hat{c}_l)$ by $\hat{c}_l = c_l$ for l between -1 and $i+1$, and by $\hat{c}_{i+2} = q$, $\hat{c}_{i+3} = b_{j+1}$ and $\hat{c}_{i+4} = x_{j+1}$. This new sequence cuts (c_l) after $l = i+1$, add a neutral segment after the last relator $[c_i, c_{i+1}]$, and then a relator $[q, b_{j+1}]$ and a final neutral segment $[b_{j+1}, x_{j+1}]$. Hence, $(\hat{c}_l)$ is a configuration of order $i+3$ for the segment $[x_0, x_{j+1}]$.

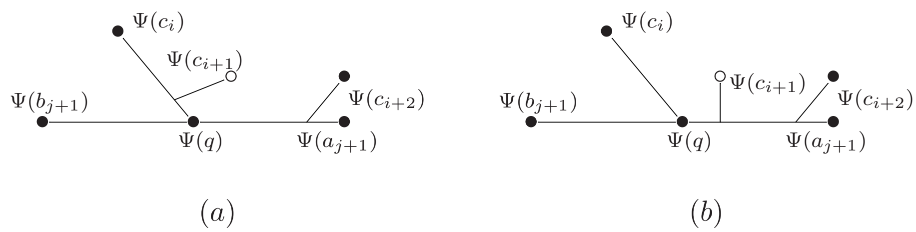


Figure 6.

In case (b), let us check that $\mathsf{d}(c_{i+1}, q) \leq B + 15\theta$, so that $[c_i, q]$ is a relator of size 4/20 (or 17/20 if $i = 0$). We have

$$\begin{aligned}\mathsf{d}(c_{i+1}, q) &\leq (c_i | b_{j+1})_{c_{i+1}} + 2\theta \\ &\leq (\Phi(c'_i) | \Phi(b))_{\Phi(c'_{i+1})} + 15\theta \\ &\leq B + 15\theta.\end{aligned}$$

As in the first subcase, the estimate for $\mathsf{d}(b_{j+1}, q)$ gives

$$\mathsf{d}(b_{j+1}, q) \geq \mathsf{d}(\Phi(b), \Phi(p)) - 17\theta,$$

and Fact 2.21 implies that $[q, b_{j+1}]$ is a relator of size 5/20; moreover $(x_{j+1}|p_1)_{b_{j+1}} \leq 3\theta$.

We define the new configuration $(\hat{c}_l)$ by taking $\hat{c}_l = c_l$ for l between -1 and i, and $\hat{c}_{i+1} = q$, $\hat{c}_{i+2} = b_{j+1}$ and $\hat{c}_{i+3} = x_{j+1}$; this defines a configuration of order $i + 2$ for the segment $[x_0, x_{j+1}]$.

By construction, properties (i), (ii), (iii), (iv) and (v) are satisfied. □

Proof of Theorem 2.9. By Lemma 2.15 there exists (c_i) a configuration of order k for $[x_0, x_m]$, where $x_m = h(x_0)$. Recall that by our choice of x_0 we have $L(h) \geq \mathsf{d}(x_0, x_m) - \theta$.

If $k \geq 2$, then we have at least two distinct relators: the first one, $[c_0, c_1]$, has size at least 17/20, and the last one, $[c_{k-1}, c_k]$, has size 5/20. By Lemma 2.6 we obtain $L(h) \geq \frac{22}{20}L - 11\theta > L$.

If $k = 1$, either $m = 1$ and h is conjugate to g^n, or we have $\mathsf{d}(x_0, c_0) \geq \frac{3}{20}L$. On the other hand $\mathsf{d}(c_0, c_1) \geq \frac{18}{20}L$, so we obtain $L(h) \geq \frac{21}{20}L - 11\theta > L$. □

3. Hyperbolic spaces with constant negative curvature

This section is devoted to the classical hyperbolic space $\mathbb{H}^n$, where the dimension n is allowed to be infinite. As we shall see, constant negative curvature, which is stronger than δ-hyperbolicity, is a useful property to decide whether the axis $\mathrm{Ax}(g)$ of an isometry is rigid.

3.1. **Hyperbolic spaces.**

3.1.1. *Definition.* Let H be a real Hilbert space, with scalar product $(x{\cdot}y)_H$, and let $\|x\|_H$ denotes the norm of any element $x \in H$. Let u be a unit vector in H, and $u^\perp$ be its orthogonal complement; each element $x \in H$ decomposes uniquely into $x = \alpha(x)u + v_x$ with $\alpha(x)$ in $\mathbf{R}$ and v_x in $u^\perp$. Let $\langle\, . \,|\, . \,\rangle \colon H \times H \mapsto \mathbf{R}$ be the symmetric bilinear mapping defined by

$$\langle x|y\rangle = \alpha(x)\alpha(y) - (v_x \cdot v_y)_H.$$

This bilinear mapping is continuous and has signature equal to $(1, \dim(H) - 1)$. The set of points x with $\langle x|x\rangle = 0$ is the light cone of $\langle\, . \,|\, . \,\rangle$. Let $\mathbb{H}$ be the subset of H defined by

$$\mathbb{H} = \{x \in H\, ;\, \langle x|x\rangle = 1 \text{ and } \langle u|x\rangle > 0\}.$$

The space $\mathbb{H}$ is the sheet of the hyperboloid $\langle x|x\rangle = 1$ which contains u.

The function $\mathsf{dist} \colon \mathbb{H} \times \mathbb{H} \mapsto \mathbf{R}_+$ defined by

$$\cosh(\mathsf{dist}(x, y)) = \langle x|y\rangle$$

gives a distance on $\mathbb{H}$, and $(\mathbb{H}, \mathsf{dist})$ is a complete and simply connected riemannian manifold of dimension $\dim(H) - 1$ with constant scalar curvature -1 (this characterizes $\mathbb{H}$ if the dimension is finite). As such, $\mathbb{H}$ is a CAT(-1) space, and therefore a δ-hyperbolic space. More precisely, $\mathbb{H}$ is δ-hyperbolic (in any dimension, even infinite) with $\delta = \log(3)$ (see [9], § I.4 page 11). In particular, all properties listed in §2.1 (or even stronger ones) are satisfied.

3.1.2. *Geodesics and boundary.* The hyperbolic space $\mathbb{H}$ is in one to one correspondence with its projection into the projective space $\mathbb{P}(H)$. The boundary of this subset of $\mathbb{P}(H)$ is the projection of the light cone of $\langle\, . \,|\, . \,\rangle$; we shall denote it by $\partial\mathbb{H}$ and call it the **boundary** of $\mathbb{H}$ (or boundary at infinity).

Let Γ be a geodesic line in $\mathbb{H}$. Then there is a unique plane $V_\Gamma \subset H$ such that $\Gamma = V_\Gamma \cap \mathbb{H}$. The plane V_Γ intersects the light cone on two lines, and these lines determine two points of $\partial\mathbb{H}$, called the endpoints of Γ. If x and y are two distinct points of $\mathbb{H}$, there is a unique geodesic segment $[x, y]$ from x to y; this segment is contained in a unique geodesic line, namely $\mathsf{Vect}(x, y) \cap \mathbb{H}$.

Let x be a point of $\mathbb{H}$ and Γ be a geodesic. The projection $\pi_\Gamma(x) \in \Gamma$, i.e. the point $y \in \Gamma$ which minimizes $\mathsf{dist}(x, y)$, is unique.

3.1.3. *Distances between geodesics.* The following lemma shows that close geodesic segments are indeed exponentially close in a neighborhood of their middle points; this property is not satisfied in general δ-hyperbolic spaces because δ-hyperbolicity does not tell anything at scale $\leq \delta$.

Lemma 3.1. *Let x, x', y, and y' be four points of the hyperbolic space $\mathbb{H}$. Assume that*

(i) $\mathsf{dist}(x,x') \leq \varepsilon$, *and* $\mathsf{dist}(y,y') \leq \varepsilon$,
(ii) $\mathsf{dist}(x,y) \geq B$,
(iii) $B \geq 10\varepsilon$,

where ε *and* B *are positive real numbers. Let* $[u,v] \subset [x,y]$ *be the geodesic segment of length* $\frac{3}{4}\mathsf{dist}(x,y)$ *centered in the middle of* $[x,y]$*. Then* $[u,v]$ *is contained in the tubular neighborhood* $Tub_\nu([x',y'])$ *with*

$$\cosh(\nu) - 1 \leq 5\frac{\cosh(2\varepsilon)-1}{\exp(B/2-4\varepsilon)}.$$

The constants in this inequality are not optimal and depend on the choice of the ratio $\mathsf{dist}(u,v)/\mathsf{dist}(x,y) = 3/4$.

Proof. We use coordinates $(x_0, x_1, ..., x_n, ...)$ in H with

$$\langle x|x\rangle = x_0^2 - \sum_{i\geq 1} x_i^2.$$

We denote by $|x|$ the square root of the absolute value of $\langle x|x\rangle$, and denote by $\|x\|_{euc}$ the square root of $\sum_{i\geq 0} x_i^2$. We have $|\langle u|v\rangle| \leq \|u\|_{euc}\|v\|_{euc}$.

If x and y are two elements of $\mathbb{H}$ the segment $[x,y]$ is parametrized by $t \mapsto u_t/|u_t|$ where $u_t = tx + (1-t)y$.

Let x, x', y, y' satisfy $\mathsf{dist}(x,x') \leq \varepsilon$, $\mathsf{dist}(y,y') \leq \varepsilon$, and $\mathsf{dist}(x,y) = B$. We first apply an isometry of $\mathbb{H}$ to assume that $x = (1,0,...0,..)$ and $y = (\cosh(B), \sinh(B), 0, ...)$.

The sphere of $\mathbb{H}$ centered at x with radius B is the set $\{y+z \in \mathbb{H} \,|\, \langle x|z\rangle = 0\}$.

First Step.— Let us first assume that $\mathsf{dist}(x,y') = B = \mathsf{dist}(y,x')$, and write $x' = x + r$ and $y' = y + s$. We have

- $\langle x'|x'\rangle = 1$, i.e. $2\langle r|x\rangle + \langle r|r\rangle = 0$;
- $\langle x|x+r\rangle \leq \cosh(\varepsilon)$;
- $\langle x+r|y\rangle = \cosh(B)$.

Thus, writing $r = (r_0, r_1, ...)$ we get

$$0 \leq r_0 \leq \cosh(\varepsilon) - 1, \text{ and } 2r_0 + r_0^2 = \sum_{i\geq 1} r_1^2.$$

In particular,

$$\|r\|_{euc}^2 = 2r_0 + 2r_0^2 \leq 2\cosh(\varepsilon)(\cosh(\varepsilon)-1)$$

A similar analysis for s leads to the following estimates

- $s_0 = 0$;
- $-2\sinh(B)s_1 = \sum_{i\geq 1} s_i^2$;
- $\|s\|_{euc}^2 = \sum_{i\geq 1} s_i^2 \leq 2(\cosh(\varepsilon)-1)$.

Let us now parametrize the geodesic segments between x and y and between x' and y'. The first one is parametrized by

$$m_t = u_t/|u_t|, \text{ where } u_t = tx + (1-t)y.$$

The second one by $m'_t = v_t/|v_t|$ where

$$v_t = tx' + (1-t)y' = u_t + tr + (1-t)s = u_t + w_t$$

where $w_t = tr + (1-t)s$. Thus,

$$\langle m_t | m'_t \rangle = \frac{|u_t|}{|v_t|} + \frac{1}{|v_t|} \langle m_t | tr + (1-t)s \rangle.$$

We now restrict this computation to $t \in [t_0, 1-t_0]$ where t_0 is a fixed positive number (we shall need $t_0 = \exp(-B/2)$). We have

$$\begin{aligned} |u_t|^2 = t^2 + (1-t)^2 + 2t(1-t)\langle x|y\rangle &\geq 1/2 + 2t_0(1-t_0)\cosh(B), \\ |v_t|^2 &\geq 1/2 + 2t_0(1-t_0)\cosh(B-2\varepsilon) \\ \langle u_t | w_t \rangle \leq \frac{1}{2}(\cosh(\varepsilon) - 1). & \end{aligned}$$

Expanding $|v_t|^2$ and using that $2\langle u_t|w_t\rangle + \langle w_t|w_t\rangle$ is equal to $2t(1-t)\langle r|s\rangle$

$$\left| \frac{|u_t|}{|v_t|} - 1 \right| \leq 2t_0(1-t_0)\frac{\langle r|s\rangle}{|v_t|^2}$$

All together, one easily gets the estimate

$$\left|\langle m_t|m'_t\rangle - 1\right| \leq \left(2\sqrt{2}t_0(1-t_0)\sqrt{\cosh(\varepsilon)+1} + \frac{1}{2}\right) \frac{\cosh(\varepsilon)-1}{\frac{1}{2} + 2t_0(1-t_0)\cosh(B-2\varepsilon)}.$$

Let us now choose $t_0 = \exp(-B/2)$; with such a choice, the geodesic segments described by m_t (resp. m'_t) contains a segment of length $3B/4$ centered at the middle of $[x, y]$ (resp. $[x', y']$). Since $B \geq 10\varepsilon$ we have

$$\begin{aligned} \left|\langle m_t|m'_t\rangle - 1\right| &\leq (4\exp(-B/2 + \varepsilon/2) + 1/2)) \frac{\cosh(\varepsilon)-1}{\cosh(B/2 - 2\varepsilon)} \\ &\leq 5\frac{\cosh(\varepsilon)-1}{\cosh(B/2-2\varepsilon)} \end{aligned}$$

This upper bound show that the point m'_t is at distance at most ν from $[x, y]$ with

$$\cosh(\nu) - 1 \leq 5\frac{\cosh(\varepsilon)-1}{\cosh(B/2-2\varepsilon)}$$

Second Step.— When x' is not on the sphere of radius B centered at y, we replace x' by the intersection of the geodesic containing $[x', y']$ with the sphere; since $B \geq 10\varepsilon$, the distance between x and the new point x' is at most 2ε. We do the same for y' and the sphere of radius B centered at x, and then apply the first step to conclude. □

3.1.4. *Isometries.* Let f be an isometry of $\mathbb{H}$; then f is the restriction of a unique continuous linear transformation of the Hilbert space H. In particular, f extends to a homeomorphism of the boundary $\partial\mathbb{H}$. The three types of isometries (see § 2.3.1) have the following properties.

(1) If f is elliptic there is a point x in $\mathbb{H}$ with $f(x) = x$, and f acts as a rotation centered at x. Fixed points of f are eigenvectors of the linear extension $f\colon H \to H$ corresponding to the eigenvalue 1.
(2) If f is parabolic there is a unique fixed point of f on the boundary $\partial\mathbb{H}$; this point corresponds to a line of eigenvectors with eigenvalue 1 for the linear extension $f\colon H \to H$. The orbit $(f^k(x))$ of all points x in $\mathbb{H}$ converges toward this fixed point in $\mathbb{P}(H)$ when k goes to $+\infty$ and to $-\infty$.
(3) If f is hyperbolic, then f has exactly two fixed points $\alpha(f)$ and $\omega(f)$ on the boundary $\partial\mathbb{H}$ and the orbit $f^k(x)$ of every point $x \in \mathbb{H}$ goes to $\alpha(f)$ as k goes to $-\infty$ and $\omega(f)$ as k goes to $+\infty$. The set $\mathrm{Min}(f)$ coincides with the geodesic line from $\alpha(f)$ to $\omega(f)$. In particular, $\mathrm{Min}(f)$ coincides with the unique f-invariant axis $\mathrm{Ax}(f) = \mathsf{Vect}(\alpha(f), \omega(f)) \cap \mathbb{H}$. The points $\alpha(f)$ and $\omega(f)$ correspond to eigenlines of the linear extension $f\colon H \to H$ with eigenvalues λ^{-1} and λ, where $\lambda > 1$; the translation length $L(f)$ is equal to the logarithm of λ (see Remark 4.5 below).

3.2. **Rigidity of axis for hyperbolic spaces.** Let G be a group of isometries of the hyperbolic space $\mathbb{H}$, and g be a hyperbolic element of G. The following strong version of Lemma 2.8 is a direct consequence of Lemma 3.1.

Lemma 3.2. *Let $\varepsilon' < \varepsilon$ be positive real numbers. If the segment $A \subset \mathbb{H}$ is (ε', B')-rigid, then A is also (ε, B)-rigid with*

$$B = \max\left\{22\varepsilon, 18\varepsilon + 2\log 5\left(\frac{\cosh(4\varepsilon)-1}{\cosh(\varepsilon')-1}\right), 4\frac{B'}{3} + 2\varepsilon\right\}.$$

In the context of the hyperbolic spaces $\mathbb{H}$, Lemma 3.2 enables us to drop the ε in the notation for ε-rigidity; we simply say that $A \subset \mathbb{H}$ is rigid.

Proof. By assumption, B satisfies the following three properties

- $B - 2\varepsilon \geq 10(2\varepsilon) = 20\varepsilon$;
- $\cosh(\varepsilon') - 1 \geq (\cosh(4\varepsilon) - 1)/\exp(B/2 - 9\varepsilon)$;
- $(3/4)(B - 2\varepsilon) \geq B'$.

Let f be an element of the group G such that $A \cap_\varepsilon f(A)$ contains two points at distance B. Then A and $f(A)$ contain two segments I and J of length $B - 2\varepsilon$ which are 2ε-close. Lemma 3.1 and the inequalities satisfied by $(B - 2\varepsilon)$ show that I and J contain subsegments of length $3(B-2\varepsilon)/4$ which are ε'-close. Since A is (ε', B')-rigid, $f(A)$ coincides with A, and therefore A is (ε, B)-rigid. □

Proposition 3.3. *Let G be a group of isometries of $\mathbb{H}$, and $g \in G$ be a hyperbolic isometry. Let n be a positive integer, $p \in Ax(g)$, and $\eta > 0$. If $Ax(g)$ is not rigid, there exists an element f of G such that $f(Ax(g)) \neq Ax(g)$ and $\mathsf{dist}(f(x), x) \leq \eta$ for all $x \in [g^{-n}(p), g^n(p)]$.*

Proof. Since $\mathrm{Ax}(g)$ is not rigid, in particular $\mathrm{Ax}(g)$ is not (ε, B)-rigid for $B = (3n+4)L(g)$ and $\varepsilon = \eta/2$. Then there exists an isometry $h \in G$ such that $h(\mathrm{Ax}(g))$ is different from $\mathrm{Ax}(g)$, but $\mathrm{Ax}(g)$ contains a segment J of length B which is mapped by h into the tubular neighborhood $\mathrm{Tub}_\varepsilon(\mathrm{Ax}(g))$.

Changing h into $g^j \circ h$, we can assume that the point p is near the middle of the segment $h(J)$; and changing h into $h \circ g^k$ moves J to $g^{-k}(J)$. We can thus change h into $h_1 = h^j \circ h \circ g^k$ for some $j, k \in \mathbf{Z}$, B into $B_1 \geq B - 4L(g) \geq 3nL(g)$, and find two points x and y on $\mathrm{Ax}(g)$ that satisfy

(1) $\mathsf{dist}(x, y) \geq B_1$, $p \in [x, y]$, and $\mathsf{dist}(p, x) \geq B_1/3$, $\mathsf{dist}(p, y) \geq B_1/3$;
(2) either (a) $h_1(x) \in \mathrm{Tub}_\varepsilon([g^{-1}(x), x])$ and $h_1(y) \in \mathrm{Tub}_\varepsilon([g^{-1}(y), y])$,
or (b) $h_1(y) \in \mathrm{Tub}_\varepsilon([g^{-1}(x), x])$ and $h_1(x) \in \mathrm{Tub}_\varepsilon([g^{-1}(y), y])$;
(3) $g^i(x)$ and $g^i(y)$, with $-2 \leq i \leq 2$, are at distance at most ε from $h(\mathrm{Ax}(g))$.

This does not change the axis $h(\mathrm{Ax}(g))$ and the value of ε.

We now change h_1 into the commutator $h_2 = h_1^{-1} g^{-1} h_1 g$. We still have $h_2(\mathrm{Ax}(g)) \neq \mathrm{Ax}(g)$, because otherwise $h_1^{-1} g^{-1} h_1$ fixes $\mathrm{Ax}(g)$ and by uniqueness of the axis of a hyperbolic isometry of $\mathbb{H}$ we would have $h_1^{-1}(\mathrm{Ax}(g)) = \mathrm{Ax}(g)$. Moreover, property (2) above is replaced by $\mathsf{dist}(x, h_2(x)) \leq 2\varepsilon < \eta$ and $\mathsf{dist}(y, h_2(y)) \leq \eta$ in case (a), and by $\mathsf{dist}(g^2(x), h_2(x)) \leq \eta$ and $\mathsf{dist}(g^2(y), h_2(y)) \leq \eta$ in case (b); similar properties are then satisfied by the points $g^i(x)$ and $g^i(y)$, $-2 \leq i \leq 2$, in place of x and y.

Changing, once again, h_2 into $h_3 = h_2 \circ g^{-2}$ if necessary, we can assume that $\mathsf{dist}(x, h_3(x)) \leq \eta$ and $\mathsf{dist}(y, h_3(y)) \leq \eta$.

Consider the arc length parametrization

$$m \colon t \in [-\infty, +\infty] \mapsto \mathrm{Ax}(g)$$

such that $m(0) = p$ and $g(m(t)) = m(t + L(g))$. Since p is in the interval $[x + B_1/3, y - B_1/3]$, we obtain $\mathsf{dist}(m(t), h_3(m(t))) \leq \eta$ for all t in $[-B_1/3, B_1/3]$. Defining $f = h_3$ we get

$$\mathsf{dist}(z, f(z)) \leq \eta, \quad \forall z \in [g^{-n}(p), g^n(p)]$$

because $nL(g) \leq B_1/3$. □

In particular, the proof for $n = 1$ gives the following corollary.

Corollary 3.4. *Let G be a group of isometries of $\mathbb{H}$. Let g be a hyperbolic element of G and p be a point of $Ax(g)$. Let η be a positive real number. If there* *is no* f *in* $G\setminus\{\mathrm{Id}\}$ *such that* $\mathsf{d}(f(x),x)\leq\eta$ *for all* $x\in[g^{-1}(p),g(p)]$*, then* $Ax(g)$ *is* $(\eta/2,7L(g))$*-rigid.*

Part B. Algebraic Geometry and the Cremona Group

4. The Picard-Manin space

4.1. **Néron-Severi groups and rational morphisms.** Let X be a smooth projective surface defined over an algebraically closed field $\mathbf{k}$. The Néron-Severi group $\mathsf{N}^1(X)$ is the group of Cartier divisors modulo numerical equivalence. When the field of definition is the field of complex numbers, $\mathsf{N}^1(X)$ coincides with the space of Chern classes of holomorphic line bundles of X (see [34]), and thus

$$\mathsf{N}^1(X)=H^2(X(\mathbf{C}),\mathbf{Z})_{t.f.}\cap H^{1,1}(X,\mathbf{R})$$

where $H^2(X(\mathbf{C}),\mathbf{Z})_{t.f.}$ is the torsion free part of $H^2(X(\mathbf{C}),\mathbf{Z})$ (the torsion part being killed when one takes the image of $H^2(X(\mathbf{C}),\mathbf{Z})$ into $H^2(X(\mathbf{C}),\mathbf{R})$). The rank $\rho(X)$ of this abelian group is called the Picard number of X. If D is a divisor, we denote by $[D]$ its class in $\mathsf{N}^1(X)$. The intersection form defines an integral quadratic form

$$([D_1],[D_2])\mapsto[D_1]\cdot[D_2]$$

on $\mathsf{N}^1(X)$, the signature of which is equal to $(1,\rho(X)-1)$ by Hodge index theorem (see [29], §V.1). We also note $\mathsf{N}^1(X)_{\mathbf{R}}=\mathsf{N}^1(X)\otimes\mathbf{R}$.

If $\pi\colon X\to Y$ is a birational morphism, then the pullback morphism

$$\pi^*\colon\mathsf{N}^1(Y)\to\mathsf{N}^1(X)$$

is injective and preserves the intersection form. For example, if π is just the blow-up of a point with exceptional divisor $E\subset X$, then $\mathsf{N}^1(X)$ is isomorphic to $\pi^*(\mathsf{N}^1(Y))\oplus\mathbf{Z}[E]$, this sum is orthogonal with respect to the intersection form, and $[E]\cdot[E]=-1$.

Example 4.1. The Néron-Severi group of the plane is isomorphic to $\mathbf{Z}[H]$ where $[H]$ is the class of a line H and $[H]\cdot[H]=1$. After n blow-ups of points, the Néron-Severi group is isomorphic to $\mathbf{Z}^{n+1}$ with a basis of orthogonal vectors $[H]$, $[E_1],\cdots,[E_n]$ such that $[H]^2=1$ and $[E_i]^2=-1$ for all $1\leq i\leq n$.

4.2. **Indeterminacies.** Let $f\colon X\dashrightarrow Y$ be a rational map between smooth projective surfaces. The indeterminacy set $\mathrm{Ind}(f)$ is finite, and the curves which are blown down by f form a codimension 1 analytic subset of X, called the exceptional set $\mathrm{Exc}(f)$.

Let H be an ample line bundle on Y. Consider the pullback by f of the linear system of divisors which are linearly equivalent to H; when $X=Y=\mathbb{P}^2$ and H

is a line, this linear system is called the homaloidal net of f. The base locus of this linear system is supported on $\mathrm{Ind}(f)$, but it can of course include infinitely near points. We shall call it the **base locus** of f. To resolve the indeterminacies of f, one blows up the base locus (see [34] for base loci, base ideal, and their blow ups); in other words, one blows up $\mathrm{Ind}(f)$, obtaining $\pi\colon X' \to X$, then one blows up $\mathrm{Ind}(f\circ\pi)$, and so on; the process stops in a finite number of steps (see [29], §V.5).

Remark 4.2. If f is a birational transformation of a projective surface with Picard number one, then $\mathrm{Ind}(f)$ is contained in $\mathrm{Exc}(f)$.[3]

4.3. **Dynamical degrees.** The rational map $f\colon X \dashrightarrow Y$ determines a linear map $f^*\colon \mathsf{N}^1(Y) \to \mathsf{N}^1(X)$. For complex surfaces, f determines a linear map $f^*\colon H^2(Y,\mathbf{Z}) \to H^2(X,\mathbf{Z})$ which preserves the Hodge decomposition: The action of f^* on $\mathsf{N}^1(X)$ coincides with the action by pull-back on $H^2(X(\mathbf{C}),\mathbf{Z})_{t.f.} \cap H^{1,1}(X,\mathbf{R})$ (see [16] for example).

Assume now that f is a birational selfmap of X. The **dynamical degree** $\lambda(f)$ of f is the spectral radius of the sequence of linear maps $((f^n)^*)_{n\geq 0}$:

$$\lambda(f) = \lim_{n\to+\infty} \left(\|(f^n)^*\|^{1/n} \right)$$

where $\|\,.\,\|$ is an operator norm on $\mathrm{End}(\mathsf{N}^1(X)_{\mathbf{R}})$; the limit exists because the sequence $\|(f^n)^*\|$ is submultiplicative (see [16]). The number $\lambda(f)$ is invariant under conjugacy: $\lambda(f) = \lambda(gfg^{-1})$ if $g\colon X \dashrightarrow Y$ is a birational map.

Example 4.3. Let $[x : y : z]$ be homogeneous coordinates for the projective plane $\mathbb{P}^2_{\mathbf{k}}$. Let f be an element of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$; there are three homogeneous polynomials P, Q, and $R \in \mathbf{k}[x,y,z]$ of the same degree d, and without common factor of degree ≥ 1, such that

$$f[x : y : z] = [P : Q : R].$$

The degree d is called the degree of f, and is denoted by $\deg(f)$. The action f^* of f on $\mathsf{N}^1(\mathbb{P}^2_{\mathbf{k}})$ is the multiplication by $\deg(f)$. The dynamical degree of f is thus equal to the limit of $\deg(f^n)^{1/n}$.

4.4. **Picard-Manin classes.** We follow the presentation given in [4, 6, 22] which, in turn, is inspired by the fifth chapter of Manin's book [35].

[3] One can prove this by considering the factorization of f as a sequence of blow-ups followed by a sequence of blow-downs. Any curve in $\mathrm{Exc}(f)$ corresponds to a (-1)-curve at some point in the sequence of blow-down, but is also the strict transform of a curve of positive self-intersection from the source (this is where we use $\rho(X) = 1$).

4.4.1. *Models and Picard-Manin space.* A model of $\mathbb{P}^2_{\mathbf{k}}$ is a smooth projective surface X with a birational morphism $X \to \mathbb{P}^2_{\mathbf{k}}$. Let $\mathcal{B}$ be the set of all models of $\mathbb{P}^2_{\mathbf{k}}$. If $\pi\colon X \to \mathbb{P}^2_{\mathbf{k}}$, and $\pi'\colon X' \to \mathbb{P}^2_{\mathbf{k}}$ are two models, we say that X' dominates X if the induced birational map $\pi^{-1} \circ \pi'\colon X' \dashrightarrow X$ is a morphism. In this case, $\pi^{-1} \circ \pi'$ contracts a finite number of exceptional divisors and induces an injective map $(\pi^{-1} \circ \pi')^*\colon \mathsf{N}^1(X)_{\mathbf{R}} \hookrightarrow \mathsf{N}^1(X')_{\mathbf{R}}$.

Let $\mathcal{B}_X \subset \mathcal{B}$ be the set of models that dominate X. Note that if $X_1, X_2 \in \mathcal{B}_X$, by resolving the indeterminacies of the induced birational map $X_1 \dashrightarrow X_2$ we obtain $X_3 \in \mathcal{B}_X$ which dominates both X_1 and X_2. The space $\mathcal{Z}(X)$ of **(finite) Picard-Manin classes** [4] is the direct limit

$$\mathcal{Z}(X) = \lim_{\to \mathcal{B}_X} \mathsf{N}^1(X')_{\mathbf{R}}.$$

The Néron-Severi group $\mathsf{N}^1(X')_{\mathbf{R}}$ of any model $X' \to X$ embeds in $\mathcal{Z}(X)$ and can be identified to its image into $\mathcal{Z}(X)$. Thus, a Picard-Manin class is just a (real) Néron-Severi class of some model dominating X. The Picard-Manin class of a divisor D is still denoted by $[D]$, as for Néron-Severi classes. Note that $\mathcal{Z}(X)$ contains the direct limit of the lattices $\mathsf{N}^1(X')$ (with integer coefficients). This provides an integral structure for $\mathcal{Z}(X)$. In the following paragraph we construct a basis of $\mathcal{Z}(X)$ made of integral points.

For all birational morphisms π, the pull-back operator π^* preserves the intersection form and maps nef classes to nef classes; as a consequence, the limit space $\mathcal{Z}(X)$ is endowed with an intersection form (of signature $(1, \infty)$) and a nef cone.

4.4.2. *Basis of $\mathcal{Z}(X)$.* A basis of the real vector space $\mathcal{Z}(X)$ is constructed as follows. On the set of models $\pi\colon Y \to X$ together with marked points $p \in Y$, we introduce the equivalence relation: $(p, Y) \sim (p', Y')$ if the induced birational map $\pi'^{-1} \circ \pi\colon Y \dashrightarrow Y'$ is an isomorphism in a neighborhood of p that maps p onto p'. Let $\mathcal{V}_X$ be the quotient space; to denote points of $\mathcal{V}_X$ we just write $p \in \mathcal{V}_X$, without any further reference to a model $\pi : Y \to X$ with $p \in Y$.

Let (p, Y) be an element of $\mathcal{V}_X$. Consider $\bar{Y} \to Y$ the blow-up of p and $E_p \subset \bar{Y}$ the exceptional divisor; the Néron-Severi class $[E_p]$ determines a Picard-Manin class and one easily verifies that this class depends only on the class $p \in \mathcal{V}_X$ (not on the model (Y, π), see [35]). The classes $[E_p]$, $p \in \mathcal{V}_X$, have self-intersection -1, are mutually orthogonal, and are orthogonal to $\mathsf{N}^1(X)_{\mathbf{R}}$. Moreover,

$$\mathcal{Z}(X) = \mathsf{N}^1(X)_{\mathbf{R}} \oplus \mathsf{Vect}([E_p];\, p \in \mathcal{V}_X)$$

[4] Picard-Manin classes can be thought of as classes of divisors on the so-called Riemann-Zariski space; one can also define Weil classes by mean of a projective limit. These spaces are not useful for our purposes.

and this sum is orthogonal with respect to the intersection form on $\mathcal{Z}(X)$. To get a basis of $\mathcal{Z}(X)$, we then fix a basis $([H_i])_{1\leq i\leq \rho(X)}$ of $\mathsf{N}^1(X)_{\mathbf{R}}$, where the H_i are Cartier divisors, and complete it with the family $([E_p])_{p\in\mathcal{V}_X}$.

4.4.3. *Completion.* The **(completed) Picard-Manin space** $\bar{\mathcal{Z}}(X)$ of X is the L^2-completion of $\mathcal{Z}(X)$ (see [4, 6] for details); in other words

$$\bar{\mathcal{Z}}(X) = \left\{ [D] + \sum_{p\in\mathcal{V}_X} a_p[E_p];\ [D]\in \mathsf{N}^1(X)_{\mathbf{R}},\ a_p\in\mathbf{R} \text{ and } \sum a_p^2 < +\infty \right\}$$

whereas $\mathcal{Z}(X)$ corresponds to the case where the a_p vanish for all but a finite number of $p\in\mathcal{V}_X$. For the projective plane $\mathbb{P}^2_{\mathbf{k}}$, the Néron-Severi group $\mathsf{N}^1(\mathbb{P}^2_{\mathbf{k}})$ is isomorphic to $\mathbf{Z}[H]$, where H is a line; elements of $\bar{\mathcal{Z}}(X)$ are then given by sums

$$a_0[H] + \sum_{p\in\mathcal{V}_{\mathbb{P}^2_{\mathbf{C}}}} a_p[E_p]$$

with $\sum a_p^2 < +\infty$. We shall call this space the Picard-Manin space without further reference to $\mathbb{P}^2_{\mathbf{k}}$ or to the completion.

4.5. **Action of $\mathsf{Bir}(X)$ on $\bar{\mathcal{Z}}(X)$.** If $\pi\colon X'\to X$ is a morphism, then π induces an isomorphism $\pi^*\colon \mathcal{Z}(X)\to\mathcal{Z}(X')$. Let us describe this fact when π is the (inverse of the) blow-up of a point $q\in X$. In this case we have

$$\mathsf{N}^1(X') = \pi^*(\mathsf{N}^1(X))\oplus\mathbf{Z}[E_q] \quad \text{and} \quad \mathcal{V}_X = \mathcal{V}_{X'}\cup\{(q,X)\}$$

where $(q,X)\in\mathcal{V}_X$ denotes the point of $\mathcal{V}_X$ given by $q\in X$. Thus the bases of $\mathcal{Z}(X)$ and $\mathcal{Z}(X')$ are in bijection, the only difference being that $[E_q]$ is first viewed as the class of an exceptional divisor in $\mathcal{Z}(X)$, and then as an element of $\mathsf{N}^1(X')\subset\mathcal{Z}(X')$; the isomorphism π^* corresponds to this bijection of basis. Note that π^* extends uniquely as a continuous isomorphism $\pi^*\colon \bar{\mathcal{Z}}(X')\to\bar{\mathcal{Z}}(X)$ that preserves the intersection form. Since any birational morphism π is a composition of blow-downs of exceptional curves of the first kind, this proves that π acts by isometries on both $\mathcal{Z}(X)$ and $\bar{\mathcal{Z}}(X)$.

Now consider $f\colon\mathbb{P}^2_{\mathbf{k}}\dashrightarrow\mathbb{P}^2_{\mathbf{k}}$ a birational map. There exist a surface X and two morphisms $\pi\colon X\to\mathbb{P}^2_{\mathbf{k}}$, $\sigma\colon X\to\mathbb{P}^2_{\mathbf{k}}$, such that $f=\sigma\circ\pi^{-1}$. Defining f^* by $f^*=(\pi^*)^{-1}\circ\sigma^*$, and f_* by $f_*=(f^{-1})^*$, we get a representation

$$f\mapsto f_*$$

of the Cremona group in the orthogonal group of $\mathcal{Z}(\mathbb{P}^2_{\mathbf{k}})$ (resp. $\bar{\mathcal{Z}}(\mathbb{P}^2_{\mathbf{k}})$) with respect to the intersection form. This representation is faithful, because $f_*[E_p]=[E_{f(p)}]$ for all points p in the domain of definition of f; it preserves the integral structure of $\mathcal{Z}(X)$ and the nef cone.

4.6. **Action on an infinite dimensional hyperbolic space.**

4.6.1. *The hyperbolic space* $\mathbb{H}_{\bar{\mathcal{Z}}}$. We define

$$\mathbb{H}_{\bar{\mathcal{Z}}} = \left\{[D] \in \bar{\mathcal{Z}};\ [D]^2 = 1, [H] \cdot [D] > 0\right\}$$

and a distance dist on $\mathbb{H}_{\bar{\mathcal{Z}}}$ by the following formula

$$\cosh(\mathsf{dist}([D_1],[D_2])) = [D_1] \cdot [D_2].$$

Since the intersection form is of Minkowski type, this gives to $\mathbb{H}_{\bar{\mathcal{Z}}}$ the structure of an infinite dimensional hyperbolic space, as in §3.1.

4.6.2. *Cremona isometries.* By paragraph 4.5 the action of the Cremona group on $\bar{\mathcal{Z}}$ preserves the two-sheeted hyperboloid $\left\{[D] \in \bar{\mathcal{Z}}(\mathbb{P}^2_{\mathbf{k}}); [D]^2 = 1\right\}$ and since the action also preserves the nef cone, we obtain a faithful representation of the Cremona group into the group of isometries of $\mathbb{H}_{\bar{\mathcal{Z}}}$:

$$\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}}) \hookrightarrow \mathrm{Isom}(\mathbb{H}_{\bar{\mathcal{Z}}}).$$

In the context of the Cremona group, the classification of isometries into three types (see § 3.1.4) has an algebraic-geometric meaning.

Theorem 4.4 ([6])**.** *Let f be an element of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$. The isometry f_* of $\mathbb{H}_{\bar{\mathcal{Z}}}$ is hyperbolic if and only if the dynamical degree $\lambda(f)$ is > 1.*

As a consequence, when $\lambda(f) > 1$ then f_* preserves a unique geodesic line $\mathrm{Ax}(f) \subset \mathbb{H}_{\bar{\mathcal{Z}}}$; this line is the intersection of $\mathbb{H}_{\bar{\mathcal{Z}}}$ with a plane $V_f \subset \bar{\mathcal{Z}}$ which intersects the isotropic cone of $\bar{\mathcal{Z}}$ on two lines R_f^+ and R_f^- such that

$$f_*(a) = \lambda(f)^{\pm 1} a$$

for all $a \in R_f^{\pm}$ (the lines R_f^+ and R_f^- correspond to $\omega(f)$ and $\alpha(f)$ in the notation of § 3.1.4).

Remark 4.5. The translation length $L(f_*)$ is therefore equal to $\log(\lambda(f))$. Indeed, take $[\alpha] \in R_f^-$, $[\omega] \in R_f^+$ normalized such that $[\alpha] \cdot [\omega] = 1$. The point $[P] = \frac{1}{\sqrt{2}}([\alpha] + [\omega])$ is on the axis $\mathrm{Ax}(f_*)$. Since $f_*[P] = \frac{1}{\sqrt{2}}(\lambda(f)^{-1}[\alpha] + \lambda(f)[\omega])$, we get

$$e^{L(f_*)} + \frac{1}{e^{L(f_*)}} = 2\cosh\big(\mathsf{dist}([P], f_*[P])\big) = 2([P] \cdot f_*[P]) = \lambda(f) + \frac{1}{\lambda(f)}.$$

Remark 4.6. Over the field of complex numbers $\mathbf{C}$, [6] proves that: f_* is elliptic if, and only if, there exists a positive iterate f^k of f and a birational map $\varepsilon\colon \mathbb{P}^2_{\mathbf{C}} \to X$ such that $\varepsilon \circ f^k \circ \varepsilon^{-1}$ is an element of $\mathsf{Aut}(X)^0$ (the connected component of the identity in $\mathsf{Aut}(X)$); f_* is parabolic if, and only if, f preserves a pencil of elliptic curves and $\deg(f^n)$ grows quadratically with n, or f preserves a pencil of rational curves and $\deg(f^n)$ grows linearly with n.

4.6.3. *Automorphisms.* Assume that $f \in \mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ is conjugated, via a birational transformation φ, to an automorphism g of a smooth rational surface X:

$$\begin{array}{ccc} X & \xrightarrow{\quad g \quad} & X \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi} \\ \mathbb{P}^2_{\mathbf{k}} & \overset{f}{\dashrightarrow} & \mathbb{P}^2_{\mathbf{k}} \end{array}$$

Then we have an isomorphism $\varphi_*\colon \bar{\mathcal{Z}}(X) \to \bar{\mathcal{Z}}(\mathbb{P}^2_{\mathbf{k}})$ and an orthogonal decomposition

$$\bar{\mathcal{Z}}(X) = \mathsf{N}^1(X)_{\mathbf{R}} \oplus \mathsf{N}^1(X)_{\mathbf{R}}{}^{\perp}$$

where $\mathsf{N}^1(X)_{\mathbf{R}}{}^{\perp}$ is spanned by the classes $[E_p]$, $p \in \mathcal{V}_X$. This orthogonal decomposition is g_*-invariant. In particular, f_* preserves the finite dimensional subspace $\varphi_*\mathsf{N}^1(X)_{\mathbf{R}} \subset \bar{\mathcal{Z}}(\mathbb{P}^2_{\mathbf{k}})$.

By Hodge index theorem, the intersection form has signature $(1, \rho(X)-1)$ on $\mathsf{N}^1(X)$, so that $\varphi_*\mathsf{N}^1(X)_{\mathbf{R}}$ intersects $\mathbb{H}_{\bar{\mathcal{Z}}}$ on a f_*-invariant hyperbolic subspace of dimension $\rho(X)-1$. This proves the following lemma.

Lemma 4.7. *If f is conjugate to an automorphism $g \in \mathsf{Aut}(X)$ by a birational map $\varphi\colon X \dashrightarrow \mathbb{P}^2_{\mathbf{k}}$, then:*

(1) The isometry $f_*\colon \mathbb{H}_{\bar{\mathcal{Z}}} \to \mathbb{H}_{\bar{\mathcal{Z}}}$ is hyperbolic (resp. parabolic, resp. elliptic) if and only if the isometry $g_*\colon \mathsf{N}^1(X)_{\mathbf{R}} \to \mathsf{N}^1(X)_{\mathbf{R}}$ is hyperbolic (resp. parabolic, resp. elliptic) for the intersection form on $\mathsf{N}^1(X)_{\mathbf{R}}$;

(2) the translation length of f_* is equal to the translation length of g_*;

(3) if f_* is hyperbolic then, modulo φ_*-conjugacy, the plane V_f corresponds to V_g, which is contained in $NS(X)_{\mathbf{R}}$, and $Ax(f_*)$ corresponds to $Ax(g_*)$.

4.6.4. *Example: quadratic mappings (see [8]).* The set of birational quadratic maps $\mathsf{Bir}_2(\mathbb{P}^2_{\mathbf{C}})$ is an irreducible algebraic variety of dimension 14. Let $f\colon \mathbb{P}^2_{\mathbf{C}} \dashrightarrow \mathbb{P}^2_{\mathbf{C}}$ be a quadratic birational map. The base locus of f (resp. f^{-1}) is made of three points p_1, p_2, p_3 (resp. q_1, q_2, q_3), where infinitely near points are allowed. We have

$$f_*([H]) = 2[H] - [E_{q_1}] - [E_{q_2}] - [E_{q_3}]$$

with $[H]$ the class of a line in $\mathbb{P}^2_{\mathbf{C}}$. If f is an isomorphism on a neighborhood of p, and $f(p) = q$, then $f_*([E_p]) = [E_q]$. Note that these formulas are correct even when there are some infinitely near base points. For instance if f is the Hénon map

$$[x : y : z] \dashrightarrow [yz : y^2 - xz : z^2]$$

then q_2 is infinitely near q_1, and q_3 is infinitely near q_2, but nevertheless the formula for the image of H is still the same.

A Zariski open subset of $\mathsf{Bir}_2(\mathbb{P}^2_{\mathbf{C}})$ is made of birational transformations $f = h_1 \circ \sigma \circ h_2$, with $h_i \in \mathsf{Aut}(\mathbb{P}^2_{\mathbf{C}})$, $i = 1, 2$, and σ the standard quadratic involution $\sigma[x : y : z] = [yz : zx : xy]$. For such maps, $\{p_1, p_2, p_3\}$ is the image of $\mathrm{Ind}(\sigma) = \{[1 : 0 : 0]; [0 : 1 : 0]; [0 : 0 : 1]\}$ by h_2^{-1}, and $\{q_1, q_2, q_3\} = h_1(\mathrm{Ind}(\sigma))$. Moreover, the exceptional set $\mathrm{Exc}(f^{-1})$ is the union of the three lines through the pairs of points (q_i, q_j), $i \neq j$. Assume, for example, that the line through q_1 and q_2 is contracted on p_1 by f^{-1}, then we have

$$f_*([E_{p_1}]) = [H] - [E_{q_2}] - [E_{q_3}].$$

If f is a general quadratic map, then $\lambda(f)$ is equal to 2 and f_* induces a hyperbolic isometry on $\mathbb{H}_{\bar{\mathcal{Z}}}$; we shall see that it is possible to compute explicitly the points on the axis of f (see §5.1.4 for precise statements) and that $[H]$ is *not* on the axis. In fact, in this situation the axis of f does not contain any finite class $[D] \in \mathcal{Z}(\mathbb{P}^2_{\mathbf{C}})$.

5. Tight birational maps

5.1. **General Cremona transformations.** In this section we prove Theorem A concerning normal subgroups generated by iterates of general Cremona transformations.

5.1.1. *De Jonquières transformations.* Let d be a positive integer. As mentioned in the introduction, the set $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ of plane birational transformations of degree d is quasi-projective: It is a Zariski open subset in a subvariety of the projective space made of triples of homogeneous polynomials modulo scalar multiplication.

Recall that J_d denotes the set of de Jonquières transformations of degree d, defined as the birational transformations of degree d of $\mathbb{P}^2_{\mathbf{C}}$ that preserve the pencil of lines through $q_0 = [1 : 0 : 0]$. Then we define V_d as the image of the composition map

$$(h_1, f, h_2) \mapsto h_1 \circ f \circ h_2$$

where (h_1, f, h_2) describes $\mathsf{PGL}_3(\mathbf{C}) \times \mathsf{J}_d \times \mathsf{PGL}_3(\mathbf{C})$. As the image of an irreducible algebraic set by a regular map, V_d is an irreducible subvariety of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$. The dimension of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ is equal to $4d + 6$ and V_d is the unique irreducible component of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ of maximal dimension (in that sense, generic elements of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ are contained in V_d). In degree 2, i.e. for quadratic Cremona transformations, V_2 coincides with a Zariski open subset of $\mathsf{Bir}_2(\mathbb{P}^2_{\mathbf{C}})$.

Let f be an element of J_d. In affine coordinates,

$$f(x, y) = (B_y(x), A(y))$$

where A is in $\mathsf{PGL}_2(\mathbf{C})$ and B_y in $\mathsf{PGL}_2(\mathbf{C}(y))$. Clearing denominators we can assume that B_y is given by a function $B\colon y \mapsto B(y)$ with:

$$B(y) = \left(\begin{array}{cc} a(y) & b(y) \\ c(y) & d(y) \end{array} \right) \in \mathsf{GL}_2(\mathbf{C}(y))$$

where the coefficients a, c and b, d are polynomials of degree $d-1$ and d respectively. The degree of the function $\det(B(y))$ is equal to $2d-2$; if B is generic, $\det(B(y))$ has $2d-2$ roots y_i, $1 \leq i \leq 2d-2$, and $B(y_i)$ is a rank 1 complex matrix for each of these roots. The image of $B(y_i)$ is a line, and this line corresponds to a point x_i in $\mathbb{P}^2(\mathbf{C}^2)$. The birational transformation f contracts each horizontal line corresponding to a root y_i onto a point $q_i = (x_i, A(y_i))$. This provides $2d-2$ points of indeterminacy for f^{-1}; again, if B is generic, the $2d-2$ points q_i are distinct, generic points of the plane. The same conclusion holds if we change f into its inverse, and gives rise to $2d-2$ indeterminacy points $p_1, \cdots, p_{2d-2}$ for f. One more indeterminacy point (for f and f^{-1}) coincides with $p_0 = q_0 = [1:0:0]$.

An easy computation shows that the base locus of f is made of

(1) the point p_0 itself, with multiplicity $d-1$;
(2) the $2d-2$ single points $p_1, \cdots, p_{2d-2}$.

Moreover, all sets of distinct points $\{p_0, p_1, \cdots, p_{2d-2}\}$ with $p_0 = [1:0:0]$ and no three of them on a line through p_0 can be obtained as the indeterminacy set of a de Jonquières transformation of degree d. In particular, on the complement of a strict Zariski closed subset of J_d, the points p_i form a set of $2d-1$ distinct points in the plane: There are no infinitely near points in the list. Thus, for a generic element of V_d, we have

$$f_*[H] = d[H] - (d-1)[E_{p_0}] - \sum_{i=1}^{2d-2} [E_{p_i}]$$

where the p_i are generic distinct points of the plane.

Remark 5.1. If $\Sigma \subset \mathbb{P}^2_{\mathbf{C}}$ is a generic set of cardinal k, and h is an automorphism of $\mathbb{P}^2_{\mathbf{C}}$, then h is the identity map as soon as $h(\Sigma) \cap \Sigma$ contains five points. Applied to $\mathrm{Ind}(f)$, we obtain the following: Let g be a generic element of V_d, and h be an automorphism of $\mathbb{P}^2_{\mathbf{C}}$; if h is not the identity map, then $h(\mathrm{Ind}(g)) \cap \mathrm{Ind}(g)$ contains at most four points.

5.1.2. *Tightness is a general property in* $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$. Our goal is to prove the following statement.

Theorem 5.2. *There exists a positive integer k such that for all integers $d \geq 2$ the following properties are satisfied by a general element $g \in \mathsf{V}_d$:*

(1) g is a tight Cremona transformation;

(2) If $n \geq k$, then g^n generates a normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ whose non trivial elements f satisfy $\lambda(f) \geq d^n$.

Since V_d is irreducible, tight Cremona transformations are dense in V_d; since V_d is the unique component of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$ of maximal dimension, properties (1) and (2) are also generally satisfied in $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$. Thus Theorem 5.2 implies Theorem A.

5.1.3. *Strong algebraic stability is a general property.* Given a surface X, a birational transformation $f \in \mathsf{Bir}(X)$ is said to be **algebraically stable** if one of the following equivalent properties hold:

(1) if x is a point of $\mathrm{Ind}(f)$, then $f^k(x) \notin \mathrm{Ind}(f^{-1})$ for all $k \leq 0$;
(2) if y is a point of $\mathrm{Ind}(f^{-1})$ then $f^k(y) \notin \mathrm{Ind}(f)$ for all $k \geq 0$;
(3) the action f^* of f on the Néron-Severi group $\mathsf{N}^1(X)$ satisfies $(f^k)^* = (f^*)^k$ for all $k \in \mathbf{Z}$.

If f is an element of $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$, algebraic stability is equivalent to $\lambda(f) = \deg(f)$ (if f is not algebraically stable, property (3) implies $\lambda(f) < \deg(f)$). Condition (2) can be rephrased by saying that f is algebraically stable if

$$f^k(\mathrm{Ind}(f^{-1})) \cap \mathrm{Ind}(f) = \emptyset$$

for all $k \geq 0$. We now prove that general elements of V_d satisfy a property which is stronger than algebraic stability (recall from Remark 4.2 that $\mathrm{Ind}(f) \subset \mathrm{Exc}(f)$).

Lemma 5.3. *If g is a general element of V_d then*

$$g^k(\mathrm{Ind}(g^{-1})) \cap \mathrm{Exc}(g) \neq \emptyset \tag{5.1}$$

for all $k \geq 0$. In particular g is algebraically stable.

Proof. For a fixed integer $k \geq 0$, condition (5.1) is equivalent to the existence of a point $m \in \mathrm{Ind}(g^{-1})$ such that $g^k(m) \in \mathrm{Exc}(g)$, and hence to $\mathrm{Jac}_g(g^k(m)) = 0$, where Jac_g, the Jacobian determinant of g, is the equation of $\mathrm{Exc}(g)$. This is an algebraic condition, that defines an algebraic subvariety I_k in $\mathsf{Bir}_d(\mathbb{P}^2_{\mathbf{C}})$.

Fact 5.4. *There exists $g \in \mathsf{V}_d$ such that $g \notin I_k$ for any $k \geq 0$.*

Since V_d is irreducible, this fact implies that the codimension of I_k is positive and shows that equation (5.1) is satisfied on the intersection of countably many Zariski dense open subsets of V_d. We are therefore reduced to prove Fact 5.4. For this, let f be the Hénon mapping defined by

$$f([x:y:z]) = \left[\frac{yz^{d-1}}{10} : y^d + xz^{d-1} : z^d\right].$$

The affine plane $\{z \neq 0\}$ is f-invariant, and f restricts to a polynomial automorphism of this plane. The exceptional set $\mathrm{Exc}(f)$ is the line at infinity $\{z = 0\}$, and its image $\mathrm{Ind}(f^{-1})$ is the point $q = [0:1:0]$. This point is fixed by f: In the affine chart $\{y \neq 0\}$, with affine coordinates $(x, z) = [x : 1 : z]$ around q, the map f is given by

$$(x, z) \mapsto \left(\frac{z^{d-1}}{10(1 + xz^{d-1})}, \frac{z^d}{1 + xz^{d-1}} \right).$$

In particular, q is an attracting fixed point. Let U be the neighborhood of q defined by

$$U = \{[x : 1 : z]\, ; \, |x| < 1/10, |z| < 1/10\}.$$

Then, $f(U)$ is contained in $\{[x : 1 : z]\, ; \, |x| < 1/90, |z| < 1/90\}$. Let h be the linear transformation of the plane which, in the affine coordinates (x, z), is the translation $(x, z) \mapsto (x + 1/20, z + 1/20)$. We have $h(f(U)) \subset U \setminus \{z = 0\}$, and $f(U) \cap h(f(U)) = \emptyset$. We now take $g = h \circ f$. Its exceptional set is the line at infinity $\{z = 0\}$; the unique indeterminacy point of g^{-1} is $h(q) = [1/20 : 1 : 1/20]$. By construction, for all $k \geq 0$, $g^k(\mathrm{Ind}(g^{-1}))$ is a point of $U \setminus \mathrm{Exc}(g)$. This proves that g is not in I_k, for any $k \geq 0$. □

We say that a Cremona transformation g is **strongly algebraically stable** if g and g^{-1} satisfy property (5.1) stated in Lemma 5.3. This lemma shows that general elements of V_d are strongly algebraically stable.

Recall that algebraic stability implies that g is well defined along the forward orbit $g^k(\mathrm{Ind}(g^{-1}))$, $k \geq 0$; similarly, g^{-1} is well defined along the backward orbit of $\mathrm{Ind}(g)$.

Lemma 5.5. *Let g be strongly algebraically stable. Then we have:*

(1) *For all $k > j \geq 0$, $g^k(\mathrm{Ind}(g^{-1})) \cap g^j(\mathrm{Ind}(g^{-1})) = \emptyset$;*
(2) *For all $k > j \geq 0$, $g^{-k}(\mathrm{Ind}(g)) \cap g^{-j}(\mathrm{Ind}(g)) = \emptyset$;*
(3) *For all $k \geq 0$ and $j \geq 0$, $g^k(\mathrm{Ind}(g^{-1})) \cap g^{-j}(\mathrm{Ind}(g)) = \emptyset$.*

Proof. Suppose that there is a point $q \in g^k(\mathrm{Ind}(g^{-1})) \cap g^j(\mathrm{Ind}(g^{-1}))$. Suppose first that $j = 0$. This means that there exist $p, q \in \mathrm{Ind}(g^{-1})$ such that $g^k(p) = q$. But then $g^{k-1}(p) \in \mathrm{Exc}(g)$, contradicting the assumption. Now assume $j > 0$. By the first step, we know that g^{-1} is well defined along the positive orbit of $\mathrm{Ind}(g^{-1})$. Thus we can apply g^{-j}, which brings us back to the case $j = 0$.

Property (2) is proven similarly, replacing g by g^{-1}.

Suppose (3) is false. Then there exists $p \in \mathrm{Ind}(g^{-1})$, $q \in \mathrm{Ind}(g)$, $k, j \geq 0$ such that $g^k(p) = g^{-j}(q)$. By (2) we can apply g^j to the right hand side of this equality, thus $q \in g^{k+j}(\mathrm{Ind}(g^{-1})) \cap \mathrm{Ind}(g)$. This contradicts the algebraic stability of g. □

5.1.4. *Rigidity is a general property.* Let g be an element of V_d. Consider the isometry g_* of $\bar{\mathcal{Z}}$. If g is algebraically stable, its dynamical degree is equal to d and the translation length of g_* is equal to $\log(d)$. The Picard-Manin classes $[\alpha]$

and $[\omega]$ corresponding to the end points of the axis of g_* satisfy $g_*[\alpha] = [\alpha]/d$ and $g_*[\omega] = d[\omega]$. To compute explicitly such classes, we start with the class $[H]$ of a line in $\mathbb{P}^2_{\mathbf{C}}$, and use the fact that

$$\frac{1}{d^n} g_*^n[H] \to c^{ste}[\omega]$$

when n goes to $+\infty$ (see § 4.6.2).

Assuming that g is a general element of V_d, its base locus is made of one point p_0 of multiplicity $d-1$ and $2d-2$ points p_i, $1 \leq i \leq 2d-2$, of multiplicity 1, and similarly for the base locus of g^{-1}. Note $[E^+]$ (resp. $[E^-]$) the sum of the classes of the exceptional divisors, with multiplicity $d-1$ for the first one, obtained by blowing-up the $2d-1$ distinct points in $\mathrm{Ind}(g)$ (resp. $\mathrm{Ind}(g^{-1})$). We have

$$g_*[H] = d[H] - [E^-], \quad g_*^2[H] = d^2[H] - d[E^-] - g_*[E^-], \quad \text{etc.}$$

Thus, the lines R_g^- and R_g^+ of the Picard-Manin space generated by

$$[\alpha] = [H] - \sum_{i=1}^{\infty} \frac{g_*^{-i+1}[E^+]}{d^i} \quad \text{and} \quad [\omega] = [H] - \sum_{i=1}^{\infty} \frac{g_*^{i-1}[E^-]}{d^i}$$

correspond to the end points of the axis of g_*. By Lemma 5.5, both infinite sums appearing in these formulas are sums of classes of the exceptional divisors obtained by blowing up the backward (resp. forward) orbit of $\mathrm{Ind}(g)$ (resp. $\mathrm{Ind}(g^{-1})$). By construction, $[\alpha]$ and $[\omega]$ satisfy $[\alpha]\cdot[\omega] = 1$ and $[\alpha]^2 = [\omega]^2 = 0$, because both of them are on the boundary of $\mathbb{H}_{\bar{\mathcal{Z}}}$. All points on $\mathrm{Ax}(g_*)$ are linear combinations $u[\alpha] + v[\omega]$ with the condition

$$1 = (u[\alpha] + v[\omega])^2 = 2uv.$$

The intersection of $[H]$ with a point on $\mathrm{Ax}(g_*)$ is minimal for $u = v = \frac{1}{\sqrt{2}}$ and is then equal to $\sqrt{2}$ (independently on d); denote by $[P] = \frac{1}{\sqrt{2}}(\alpha + \omega)$ the point which realizes the minimum. We have

$$[P] = \sqrt{2}[H] - \frac{1}{\sqrt{2}}[R] \quad \text{with} \quad [R] = \frac{[E^+] + [E^-]}{d} + \frac{g_*^{-1}[E^+] + g_*[E^-]}{d^2} + \cdots$$

Once again, Lemma 5.5 implies the following fact.

Fact 5.6. *The class $[R]$ is a sum of classes of exceptional divisors obtained by blowing up distinct points of $\mathbb{P}^2_{\mathbf{C}}$ (there is no blow-up of infinitely near points).*

Proposition 5.7. *Let $\varepsilon_0 = 0.289$. Let $d \geq 2$ be an integer and g be a general element of V_d. Let $[P]$ be the Picard-Manin class defined above. If f is a birational transformation of the plane such that $\mathsf{dist}(f_*[Q], [Q]) \leq \varepsilon_0$ for all $[Q]$ in $\{g_*^{-1}[P], [P], g_*[P]\}$, then f is the identity map.*

The proof uses explicit values for distances and hyperbolic cosines that are recorded on Table 1. Using this table, we see that

$$\cosh(a_1 + \varepsilon) < 4 \text{ and } a_2 + \varepsilon < a_4 < a_3$$

as soon as $\varepsilon \leq \varepsilon_0 = 0.289$.

$\cosh(a_i)$	a_i
3	$a_1 \simeq 1.76274$
$\sqrt{2}$	$a_2 \simeq 0.88137 = a_1/2$
$3/\sqrt{2}$	$a_3 \simeq 1.38432$
$5/(2\sqrt{2})$	$a_4 \simeq 1.17108$
4	$a_5 \simeq 2.06343$

TABLE 1. Distances and hyperbolic cosines

Proof. We proceed in two steps.

First step.– We show that if $\mathsf{dist}(f_*[P], [P]) \leq \varepsilon_0$, then f is linear.

By the triangular inequality

$$\begin{aligned} \mathsf{dist}(f_*[H], [H]) &\leq \mathsf{dist}(f_*[H], f_*[P]) + \mathsf{dist}(f_*[P], [P]) + \mathsf{dist}([P], [H]) \\ &\leq 2\mathsf{dist}([P], [H]) + \varepsilon_0. \end{aligned}$$

Recall that $\cosh(\mathsf{dist}([D_1], [D_2])) = [D_1] \cdot [D_2]$ for all pairs of points $[D_1]$, $[D_2]$ in $\mathbb{H}_{\bar{\mathcal{Z}}}$ and that the degree of f is given by $\deg(f) = f_*[H] \cdot [H]$. Using Table 1 we see that $[P] \cdot [H] = \sqrt{2}$ implies $2\mathsf{dist}([P], [H]) = 2a_2 = a_1$. Taking hyperbolic cosines we get

$$\deg(f) \leq \cosh(a_1 + \varepsilon_0) < 4.$$

We conclude that $\deg(f) \leq 3$.

Now we want to exclude the cases $\deg(f) = 3$ or 2. We will use the following remark twice: Since $[H] \cdot [P] = \sqrt{2}$, we have $\mathsf{dist}([H], [P]) = a_2$; thus, if f_* is an isometry and $\mathsf{dist}(f_*[P], [P]) \leq \varepsilon_0$, applying hyperbolic cosines to the triangular inequality $\mathsf{dist}(f_*[H], [P]) \leq \mathsf{dist}(f_*[H], f_*[P]) + \mathsf{dist}(f_*[P], [P])$, we get $f_*[H] \cdot [P] \leq \cosh(a_2 + \varepsilon_0)$.

Suppose that $\deg(f) = 3$. Then

$$f_*[H] = 3[H] - 2[E_1] - [E_2] - [E_3] - [E_4] - [E_5]$$

for some exceptional divisors E_i above $\mathbb{P}^2_{\mathbf{C}}$ (they may come from blow-ups of infinitely near points). By Fact 5.6, all exceptional classes in the infinite sum defining $[R]$ come from blow-ups of distinct points; hence

$$[2E_1 + E_2 + E_3 + E_4 + E_5] \cdot [R] \geq \frac{1}{d}(-2(d-1) - 1 - 1 - 1 - 1) \geq -3.$$

Consider $[D]$ the point on the Picard-Manin space such that $[P]$ is the middle point of the geodesic segment from $[H]$ to $[D]$; explicitly

$$[D] = 2\sqrt{2}[P] - [H] = 3[H] - 2[R].$$

We obtain

$$\begin{aligned} f_*[H] \cdot [D] &= f_*[H] \cdot (3[H] - 2[R]) \\ &= 9 + 2[2E_1 + E_2 + E_3 + E_4 + E_5] \cdot [R] \\ &\geq 3. \end{aligned}$$

On the other hand, $f_*[H] \cdot [H] = 3$ because f has degree 3. Since $2\sqrt{2}[P] = [H] + [D]$ we obtain

$$\cosh(a_2 + \varepsilon_0) \geq f_*[H] \cdot [P] \geq \frac{3+3}{2\sqrt{2}} = \frac{3}{\sqrt{2}} = \cosh(a_3).$$

This contradicts the choice of ε_0.

Suppose that $\deg(f) = 2$. We have $f_*[H] = 2[H] - [E_1] - [E_2] - [E_3]$ where the $[E_i]$ are classes of exceptional divisors above $\mathbb{P}^2_{\mathbf{C}}$. The product $f_*[H] \cdot [P]$ is given by

$$\begin{aligned} f_*[H] \cdot [P] &= [2H - E_1 - E_2 - E_3] \cdot \left(\sqrt{2}[H] - \frac{1}{\sqrt{2}}[R] \right) \\ &= 2\sqrt{2} + \frac{1}{\sqrt{2}}[E_1 + E_2 + E_3] \cdot [R] \\ &\geq 2\sqrt{2} + \frac{1}{\sqrt{2}}(-1 - \frac{1}{d}) \\ &\geq 2\sqrt{2} - \frac{3}{2\sqrt{2}} = \frac{5}{2\sqrt{2}} = \cosh(a_4), \end{aligned}$$

where the inequality follows from Fact 5.6: The curves E_1, E_2, and E_3 appear at most once in $[R]$, so that $[E_1 + E_2 + E_3] \cdot [R] \geq -1 - \frac{1}{d} \geq -\frac{3}{2}$. As a consequence,

$$\cosh(a_2 + \varepsilon_0) \geq \cosh(a_4),$$

in contradiction with the choice of ε_0.

Thus, we are reduced to the case where f is an element of $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{C}})$.

Second step.– Now we show that if f is in $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{C}})$ and $\mathsf{dist}(f_*[Q],[Q]) \le \varepsilon_0$ for $[Q]$ in $\{g_*^{-1}[P], g_*[P]\}$, then f is the identity map.

Applying the assumption to $[Q] = g_*^{-1}[P]$, we get

$$\mathsf{dist}((gfg^{-1})_*[P],[P]) \le \varepsilon_0$$

and the first step shows that gfg^{-1} must be linear. This implies that $f(\mathrm{Ind}(g))$ coincides with $\mathrm{Ind}(g)$. The same argument with $[Q] = g_*[P]$ gives $f(\mathrm{Ind}(g^{-1})) = \mathrm{Ind}(g^{-1})$. If $\deg(g) \ge 3$, the set $\mathrm{Ind}(g)$ is a generic set of points in $\mathbb{P}^2_{\mathbf{C}}$ with cardinal at least 5, so $f(\mathrm{Ind}(g)) = \mathrm{Ind}(g)$ implies that f is the identity map (see remark 5.1). Finally, if $\deg(g) = 2$, the set $\mathrm{Ind}(g) \cup \mathrm{Ind}(g^{-1})$ is a generic set of 6 points in the plane, so if $f(\mathrm{Ind}(g)) = \mathrm{Ind}(g)$ and $f(\mathrm{Ind}(g^{-1})) = \mathrm{Ind}(g^{-1})$ then again f is the identity map. □

Corollary 5.8. *If g is a general element of V_d, then $Ax(g_*)$ is rigid.*

Proof. Suppose $\mathrm{Ax}(g_*)$ is not rigid. Choose $\eta > 0$. By Proposition 3.3 there exists f which does not preserve $\mathrm{Ax}(g_*)$ such that $d(f_*[Q],[Q]) < \eta$ for all $[Q]$ in the segment $\left[g_*^{-1}[P], g_*[P]\right]$. For $\eta < \varepsilon_0$, this contradicts Proposition 5.7. □

Proof of Theorem 5.2. Let g be a general element of V_d, with $d \ge 2$. Suppose that $f_*(\mathrm{Ax}(g_*)) = \mathrm{Ax}(g_*)$. If f_* preserves the orientation on $\mathrm{Ax}(g_*)$, then $(fgf^{-1}g^{-1})_*$ fixes each point in $\mathrm{Ax}(g)$, and Proposition 5.7 gives $fgf^{-1} = g$. Similarly, if f_* reverses the orientation, considering $(fgf^{-1}g)_*$ we obtain $fgf^{-1} = g^{-1}$.

Since we know by Corollary 5.8 that $\mathrm{Ax}(g_*)$ is rigid, we obtain that g_* is tight, hence by Theorem 2.9 g_*^k generates a proper subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ for large enough k.

We can be more precise on k by reconsidering the proof of Corollary 5.8. The rigidity of the axis of g_* follows from Proposition 5.7, where the condition on ε_0 is independent of d, so by Corollary 3.4 we obtain that $\mathrm{Ax}(g_*)$ is $(\varepsilon_0/2, 7L(g_*))$-rigid. Recall that θ is defined in section 2.2.1 by $\theta = 4\delta$ and that $\delta = \log(3)$ works for the hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$; thus we can choose $\theta = 4\log(3)$, so that $14\theta = 56\log(3) = 61.52... \le 62$. By Lemma 3.2 we obtain that $\mathrm{Ax}(g_*)$ is also $(14\theta, B)$-rigid, for

$$B \ge \max(1369, (28/3)L(g_*) + 124).$$

Then, Theorem 2.9 requires $kL(g_*) \ge 40B + 1200\theta$. Thus, we see that

$$k = \max\left(\frac{60034}{L(g_*)}, 374 + \frac{10234}{L(g_*)}\right) \le 86611$$

is sufficient to conclude that g^k generates a proper normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$. Asymptotically, for degrees d with $d \geq \exp(10234)$, we can take $k = 375$. When $d \geq 9$, $k = 30017$ works, etc. □

5.2. **Automorphisms of rational surfaces.**

5.2.1. *Preliminaries.* Let $\mathbf{k}$ be an algebraically closed field, and X be a rational surface defined over $\mathbf{k}$. Let g be an automorphism of X. If $\varphi\colon X \dashrightarrow \mathbb{P}^2_{\mathbf{k}}$ is a birational map, conjugation by φ provides an isomorphism between $\mathsf{Bir}(X)$ and $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ and φ_* conjugates the actions of $\mathsf{Bir}(X)$ on $\bar{\mathcal{Z}}(X)$ and $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ on $\bar{\mathcal{Z}} = \bar{\mathcal{Z}}(\mathbb{P}^2_{\mathbf{k}})$. We thus identify $\mathsf{Bir}(X)$ to $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ and $\bar{\mathcal{Z}}(X)$ to $\bar{\mathcal{Z}}(\mathbb{P}^2_{\mathbf{k}})$ without further reference to the choice of φ. This paragraph provides a simple criterion to check whether g is a tight element of the Cremona group $\mathsf{Bir}(X)$.

As explained in §4.6.3, the Néron-Severi group $\mathsf{N}^1(X)$ embeds into the Picard-Manin space $\bar{\mathcal{Z}}$ and g_* preserves the orthogonal decomposition

$$\bar{\mathcal{Z}} = \mathsf{N}^1(X)_{\mathbf{R}} \oplus \mathsf{N}^1(X)_{\mathbf{R}}{}^{\perp},$$

where $\mathsf{N}^1(X)_{\mathbf{R}}{}^{\perp}$ is the orthogonal complement with respect to the intersection form. Assume that g_* is hyperbolic, with axis $\mathrm{Ax}(g_*)$ and translation length $\log(\lambda(g))$. By Lemma 4.7, the plane V_g such that $\mathrm{Ax}(g_*) = \mathbb{H}_{\bar{\mathcal{Z}}} \cap V_g$ is contained in $\mathsf{N}^1(X)_{\mathbf{R}}$.

Remark 5.9. Let h be a birational transformation of X such that h_* preserves V_g. The restriction of h_* to V_g satisfies one of the following two properties.

(1) h_* and g_* commute on V_g;
(2) h_* is an involution of V_g and h_* conjugates g_* to its inverse on V_g.

Indeed, the group of isometries of a hyperbolic quadratic form in two variables is isomorphic to the semi-direct product $\mathbf{R} \rtimes \mathbf{Z}/2\mathbf{Z}$.

Lemma 5.10. *Let g be a hyperbolic automorphism of a rational surface X. Assume that*

(i) g_* is the identity on the orthogonal complement of V_g in $\mathsf{N}^1(X)_{\mathbf{R}}$;
(ii) the action of $\mathsf{Aut}(X)$ on $\mathsf{N}^1(X)$ is faithful.

Let h be an automorphism of X such that h_* preserves V_g. Then hgh^{-1} is equal to g or g^{-1}.

Proof. Let us study the action of h_* and g_* on $\mathsf{N}^1(X)$. Since h_* preserves V_g, it preserves its orthogonal complement $V_g^{\perp}$ and, by assumption (i), commutes to g_* on $V_g^{\perp}$. If h_* commutes to g_* on V_g, then h_* and g_* commute on $\mathsf{N}^1(X)$, and the conclusion follows from the second assumption. If h_* does not commute to g, Remark 5.9 implies that h_* is an involution on V_g, and that $h_*g_*(h_*)^{-1} = (g_*)^{-1}$

on V_g and therefore also on $\mathsf{N}^1(X)$. Once again, assumption (*ii*) implies that h conjugates g to its inverse. □

Remark 5.11. The plane V_g is a subspace of $\mathsf{N}^1(X)_{\mathbf{R}}$ and it may very well happen that this plane does not intersect the lattice $\mathsf{N}^1(X)$ (except at the origin). But, if g_* is the identity on $V_g^{\perp}$, then $V_g^{\perp}$ and V_g are defined over $\mathbf{Z}$. In that case, $V_g \cap \mathsf{N}^1(X)$ is a rank 2 lattice in the plane V_g. This lattice is g_*-invariant, and the spectral values of the linear transformation $g_* \in \mathsf{GL}(V_g)$ are quadratic integers. From this follows that $\lambda(g)$ is a quadratic integer. In the opposite direction, if $\lambda(g)$ is a quadratic integer, then V_g is defined over the integers, and so is $V_g^{\perp}$. The restriction of g_* to $V_g^{\perp}$ preserves the lattice $V_g^{\perp} \cap \mathsf{N}^1(X)$ and a negative definite quadratic form. This implies that a positive iterate of g_* is the identity on $V_g^{\perp}$.

Let us now assume that there exists an ample class $[D'] \in \mathsf{N}^1(X) \cap V_g$. In that case, $[D']$ and $g_*[D']$ generate a rank 2 subgroup of $\mathsf{N}^1(X) \cap V_g$ and Remark 5.11 implies that $\lambda(g)$ is a quadratic integer. In what follows, we shall denote by $[D]$ the class

$$[D] = \frac{1}{\sqrt{[D'] \cdot [D']}} [D']. \tag{5.2}$$

This is an ample class with real coefficients that determines a point $[D]$ in $\mathbb{H}_{\bar{\mathcal{Z}}}$.

Lemma 5.12. *Let h be a birational transformation of a projective surface X. Let $[D'] \in \mathsf{N}^1(X)$ be an ample class, and $[D] = [D'] / \sqrt{[D'] \cdot [D']} \in \mathbb{H}_{\bar{\mathcal{Z}}}$. If*

$$\cosh(\mathsf{dist}(h_*[D], [D])) < 1 + ([D'] \cdot [D'])^{-1}$$

then h_* fixes $[D']$ and is an automorphism of X.

Proof. Write $h_*[D] = [D] + [F] + [R]$ where $[F]$ is in $\mathsf{N}^1(X)_{\mathbf{R}}$ and $[R]$ is in $\mathsf{N}^1(X)_{\mathbf{R}}{}^{\perp}$. More precisely, F is an element of $\mathsf{N}^1(X)$ divided by the square root of the self intersection $[D']^2$, and $[R]$ is a sum $\sum m_i [E_i]$ of exceptional divisors obtained by blowing up points of X, coming from indeterminacy points of h^{-1}; the m_i are integers divided by the square root of $[D']^2$. We get

$$1 \leq [D] \cdot h_*[D] = [D] \cdot ([D] + [F] + [R]) = 1 + [D] \cdot [F]$$

because $[D]$ does not intersect the $[E_i]$. The number $[D] \cdot [F]$ is a non negative integer divided by $[D']^2$. By assumption, this number is less than $([D'] \cdot [D'])^{-1}$ and so it must be zero. In other words, the distance between $[D]$ and $h_*[D]$ vanishes, and $[D]$ is fixed. Since $[D]$ is ample, h is an automorphism of X. □

Proposition 5.13. *Let g be a hyperbolic automorphism of a rational surface X. Assume that*

(*i*) *V_g contains an ample class $[D']$ and* *(ii)* g_* *is the identity on* $V_g^{\perp} \cap \mathsf{N}^1(X)$.

Then $Ax(g_*)$ *is rigid. Assume furthermore that*

(iii) if $h \in \mathsf{Aut}(X)$ *satifies* $h_*(Ax(g_*)) = Ax(g_*)$ *then* $hgh^{-1} = g$ *or* g^{-1}.

Then any $h \in \mathsf{Bir}(X)$ *which preserves* $Ax(g_*)$ *is an automorphism of* X*, and* g *is a tight element of* $\mathsf{Bir}(X)$*. Thus, for sufficiently large* k *the iterate* g^k *generates a non trivial normal subgroup in the Cremona group* $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}}) = \mathsf{Bir}(X)$.

Note that if the action of $\mathsf{Aut}(X)$ on $\mathsf{N}^1(X)$ is faithful then by Lemma 5.10 condition (iii) is automatically satisfied.

Proof. Let $[D] \in \mathrm{Ax}(g_*)$ be the ample class defined by equation (5.2). If the axis of g_* is not rigid, Proposition 3.3 provides a birational transformation f of X such that the distances between $f_*[D]$ and $[D]$ and between $f_*(g_*[D])$ and $g_*[D]$ are bounded by $([D']^2)^{-1}$, and, moreover, $f_*(\mathrm{Ax}(g_*)) \neq \mathrm{Ax}(g_*)$. Lemma 5.12 implies that f is an automorphism of X fixing both $[D]$ and $g_*[D]$. This contradicts $f_*(\mathrm{Ax}(g_*)) \neq \mathrm{Ax}(g_*)$ and shows that $\mathrm{Ax}(g_*)$ is rigid.

Assume now that $h \in \mathsf{Bir}(X)$ preserves the axis of g_*. Then $h_*[D']$ is an ample class, hence h is an automorphism of X. Property (iii) implies that g is a tight element of $\mathsf{Bir}(X)$ and the conclusion follows from Theorem 2.9. □

In the following paragraphs we construct two families of examples which satisfy the assumption of Proposition 5.13. Note that the surfaces X that we shall consider have quotient singularities; if we blow-up X to get a smooth surface, then the class $[D']$ is big and nef but is no longer ample. For the first example, we work on the field of complex numbers $\mathbf{C}$, while for the second we work on any algebraically closed field.

5.2.2. *Generalized Kummer surfaces.* Consider $\mathbf{Z}[i] \subset \mathbf{C}$ the lattice of Gaussian integers, and let Y be the abelian surface $\mathbf{C}/\mathbf{Z}[i] \times \mathbf{C}/\mathbf{Z}[i]$. The group $\mathsf{GL}_2(\mathbf{Z}[i])$ acts by linear transformations on $\mathbf{C}^2$, preserving the lattice $\mathbf{Z}[i] \times \mathbf{Z}[i]$; this provides an embedding $\mathsf{GL}_2(\mathbf{Z}[i]) \to \mathsf{Aut}(Y)$. Let X be the quotient of Y by the action of the group of order 4 generated by $\eta(x, y) = (ix, iy)$. This surface is rational, with ten singularities, and all of them are quotient singularities that can be resolved by a single blow-up. Such a surface is a so called (generalized) **Kummer surface**; classical Kummer surfaces are quotient of tori by $(x, y) \mapsto (-x, -y)$, and are not rational (these surfaces are examples of K3-surfaces).

The linear map η generates the center of the group $\mathsf{GL}_2(\mathbf{Z}[i])$. As a consequence, $\mathsf{GL}_2(\mathbf{Z}[i])$, or more precisely $\mathsf{PGL}_2(\mathbf{Z}[i])$, acts by automorphisms on X. Let M be an element of $\mathsf{SL}_2(\mathbf{Z})$ such that

(i) the trace $\mathrm{tr}(M)$ of M is at least 3;

(ii) M is in the level 2 congruence subgroup of $\mathsf{SL}_2(\mathbf{Z})$, i.e. M is equal to the identity modulo 2.

Let $\hat{g}$ be the automorphism of Y defined by M and g be the automorphism of X induced by $\hat{g}$.

Theorem 5.14. *The automorphism $g\colon X \to X$ satisfies properties (i), (ii), and (iii) of Proposition 5.13. In particular, g determines a tight element of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$; hence, if k is large enough, g^k generates a non trivial normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$.*

Proof. The Néron-Severi group of Y has rank 4, and is generated by the following classes: The class $[A]$ of horizontal curves $\mathbf{C}/\mathbf{Z}[i] \times \{*\}$, the class $[B]$ of vertical curves $\{*\} \times \mathbf{C}/\mathbf{Z}[i]$, the class $[\Delta]$ of the diagonal $\{(z,z) \in Y\}$, and the class $[\Delta_i]$ of the graph $\{(z,iz) \in Y \mid z \in \mathbf{C}/\mathbf{Z}[i]\}$.

The vector space $H^2(Y,\mathbf{R})$ is isomorphic to the space of bilinear alternating two forms on the 4-dimensional real vector space $\mathbf{C}^2$. The action of g_* is given by the action of M^{-1} on this space. The dynamical degree of $\hat{g}$ is equal to the square of the spectral radius of M, i.e. to the quadratic integer

$$\lambda(\hat{g}) = \frac{1}{2}\left(a + \sqrt{a^2-4}\right)$$

where $a = \mathrm{tr}(M)^2 - 2 > 2$. Thus, the plane $V_{\hat{g}}$ intersects $\mathsf{N}^1(Y)$ on a lattice. Let $[F]$ be an element of $V_{\hat{g}} \cap \mathsf{N}^1(X)$ with $[F]^2 > 0$. Since Y is an abelian variety, then $[F]$ (or its opposite) is ample. Since M has integer coefficients, the linear map $\hat{g}_*$ preserves the three dimensional subspace W of $\mathsf{N}^1(Y)$ generated by $[A]$, $[B]$ and $[\Delta]$. The orthogonal complement of $V_{\hat{g}}$ intersects W on a line, on which $\hat{g}_*$ must be the identity, because $\det(M) = 1$. The orthogonal complement of W is also a line, so that $\hat{g}_*$ is the identity on $V_{\hat{g}}^{\perp} \subset \mathsf{N}^1(Y)$.

Transporting this picture in $\mathsf{N}^1(X)$, we obtain: The dynamical degree of g is equal to the dynamical degree of $\hat{g}$ (see [27], for more general results), the plane $V_{\hat{g}}$ surjects onto V_g, the image of $[F]$ is an ample class $[D']$ contained in $V_g \cap \mathsf{N}^1(X)$, and g_* is the identity on $V_g^{\perp}$.

Automorphisms of X permute the ten singularities of X. The fundamental group Γ of $X \setminus \mathrm{Sing}(X)$ is the affine group $\mathbf{Z}/4\mathbf{Z} \ltimes (\mathbf{Z}[i] \times \mathbf{Z}[i])$ where $\mathbf{Z}/4\mathbf{Z}$ is generated by η. The abelian group $\mathbf{Z}[i] \times \mathbf{Z}[i]$ is the unique maximal free abelian subgroup of rank 4 in Γ and, as such, is invariant under all automorphisms of Γ. This implies that all automorphisms of X lift to (affine) automorphisms of Y.

Let h be an automorphism of X which preserves the axis $\mathrm{Ax}(g_*)$. Then h_* conjugates g_* to g_* or $(g_*)^{-1}$ (Remark 5.9), and we must show that h conjugates g to g or g^{-1}. Let $\hat{h}$ be a lift of h to Y. There exists a linear transformation $N \in \mathsf{GL}_2(\mathbf{C})$ and a point $(a,b) \in Y$ such that

$$\hat{h}(x,y) = N(x,y) + (a,b).$$

The lattice $\mathbf{Z}[i] \times \mathbf{Z}[i]$ is N-invariant, and N conjugates M to M or its inverse M^{-1}, because h_* conjugates g_* to g_* or its inverse. Then

$$\hat{h} \circ \hat{g} \circ \hat{h}^{-1} = M^{\pm}(x, y) + (\mathrm{Id} - M^{\pm})(a, b).$$

On the other hand, since $\hat{h}$ is a lift of an automorphism of X, the translation $t : (x, y) \mapsto (x, y) + (a, b)$ is normalized by the cyclic group generated by η. Thus a and b are in $(1/2)\mathbf{Z}[i]$. Since M is the identity modulo 2, we have $M(a, b) = (a, b)$ modulo $\mathbf{Z}[i] \times \mathbf{Z}[i]$. Hence $\hat{h} \circ \hat{g} \circ \hat{h}^{-1} = \hat{g}^{\pm}$ and, coming back to X, h conjugates g to g or its inverse. □

Remark 5.15. The lattice of Gaussian integers can be replaced by the lattice of Eisenstein integers $\mathbf{Z}[j] \subset \mathbf{C}$, with $j^3 = 1$, $j \neq 1$, and the homothety η by $\eta(x, y) = (jx, jy)$. This leads to a second rational Kummer surface with an action of the group $\mathsf{PSL}_2(\mathbf{Z})$, and a statement similar to Theorem B can be proved for this example.

5.2.3. *Coble surfaces.* Let $\mathbf{k}$ be an algebraically closed field. Let $S \subset \mathbb{P}^2_{\mathbf{k}}$ be a rational sextic curve, with ten double points m_i, $1 \leq i \leq 10$; such sextic curves exist and, modulo the action of $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}})$, they depend on 9 parameters (see [28], the appendix of [26], or [10]). Let X be the surface obtained by blowing up the ten double points of S: By definition X is the **Coble surface** defined by S.

Let $\pi \colon X \to \mathbb{P}^2_k$ be the natural projection and E_i, $1 \leq i \leq 10$, be the exceptional divisors of π. The canonical class of X is

$$[K_X] = -3[H] + \sum_{i=1}^{10} [E_i]$$

where $[H]$ is the pullback by π of the class of a line. The strict transform S' of S is an irreducible divisor of X, and its class $[S']$ coincides with $-2[K_X]$; more precisely, there is a rational section Ω of $2K_X$ that does not vanish and has simple poles along S'.

Remark 5.16. Another definition of Coble surfaces requires X to be a smooth rational surface with a non zero regular section of $-2K_X$: Such a definition includes our Coble surfaces (the section vanishes along S') but it includes also the Kummer surfaces from the previous paragraph (see [21]). Our definition is more restrictive.

The self-intersection of S' is -4, and S' can be blown down: This provides a birational morphism $q \colon X \to X_0$; the surface X_0 has a unique singularity, at $m = q(S')$. The section Ω defines a holomorphic section of $K_{X_0}^{\otimes 2}$ that trivializes $2K_{X_0}$ in the complement of m; in particular, $H^0(X, -2K_X)$ has dimension 1, and the base locus of $-2K_X$ coincides with S'. The automorphism group $\mathsf{Aut}(X)$ acts linearly on the space of sections of $-2K_X$, and preserves its base locus S'. It follows that q conjugates $\mathsf{Aut}(X)$ and $\mathsf{Aut}(X_0)$.

The rank of the Néron-Severi group $\mathsf{N}^1(X)$ is equal to 11. Let W be the orthogonal complement of $[K_X]$ with respect to the intersection form. The linear map $q^*\colon \mathsf{N}^1(X_0) \to \mathsf{N}^1(X)$ provides an isomorphism between $\mathsf{N}^1(X_0)$ and its image $W = [K_X]^\perp \subset \mathsf{N}^1(X)$. Let $O'(\mathsf{N}^1(X))$ be the group of isometries of the lattice $\mathsf{N}^1(X)$ which preserve the canonical class $[K_X]$.

We shall say that S (resp. X) is **special** when at least one of the following properties occurs (see [18], [10] page 147):

(1) three of the points m_i are colinear;
(2) six of the points m_i lie on a conic;
(3) eight of the m_i lie on a cubic curve with a double point at one of them;
(4) the points m_i lie on a quartic curve with a triple point at one of them.

Theorem 5.17 (Coble's theorem, see [19, 20])**.** *Let* $\mathbf{k}$ *be an algebraically closed field. There is a non empty Zariski open subset* U *of the space of rational sextic curves* $S \subset \mathbb{P}^2_k$ *with the following property: If* S *is a point of* U *and* X *is the associated Coble surface then the morphism*

$$\begin{array}{ccc} \mathsf{Aut}(X) & \mapsto & O'(\mathsf{N}^1(X)) \\ f & \to & f_* \end{array}$$

is injective and its image is the level 2 *congruence subgroup of* $O'(\mathsf{N}^1(X))$*. When the characteristic of* $\mathbf{k}$ *is different from* 2*, the set* U *coincides with the set of sextic curves with ten double points which are not special.*[5]

Remark 5.18. Let $\mathrm{Amp}(W)$ be the set of ample classes in $\mathsf{N}^1(X_0)\otimes\mathbf{R} \simeq W\otimes\mathbf{R}$. This convex cone is invariant under the action of $\mathsf{Aut}(X)$, hence under the action of a finite index subgroup of the orthogonal group $O(W)$. As such, $\mathrm{Amp}(W)$ is equal to $\{[D] \in W \mid [D]^2 > 0,\ [D]\cdot[H] > 0\}$.

Let S be a generic rational sextic, and X be its associated Coble surface. Theorem 5.17 gives a recipe to construct automorphisms of X: Let ψ be an isometry of the lattice W; if ψ is equal to the identity modulo 2, and $\psi[H]\cdot[H] > 0$, then $\psi = g_*$ for a unique automorphism of X. Let us apply this idea to cook up a tight automorphism of X.

[5] The proof is similar to an analogous result concerning generic Enriques surfaces, and follows from [17], [10], and [19]. The automorphism group of X can be identified with a normal subgroup of $O'(\mathsf{N}^1(X))$ which is generated, as a normal subgroup, by an explicit involution. The main point is to realize this involution by an automorphism; to do it one constructs a 2 to 1 morphism from X to a Del Pezzo surface and takes the involution of this cover. For this purpose, one has to excludes all special sextics and, in characteristic 2, one needs to check that the cover is not purely inseparable (this adds one extra condition).

Lemma 5.19 (Pell -Fermat equation)**.** *Let $Q(u,v) = au^2 + buv + cv^2$ be a quadratic binary form with integer coefficients. Assume that Q is non degenerate, indefinite, and does not represent 0. Then there exists an isometry ϕ in the orthogonal group $\mathrm{O}_Q(\mathbf{Z})$ with eigenvalues $\lambda(Q) > 1 > 1/\lambda(Q) > 0$.*

Sketch of the proof. Isometries in $\mathrm{O}_Q(\mathbf{Z})$ correspond to units in the quadratic field defined by the polynomial $Q(t,1) \in \mathbf{Z}[t]$; finding units, or isometries, is a special case of Dirichlet's units theorem and amounts to solve a Pell-Fermat equation (see [23], chapter V.1). □

Let $[D_1]$ and $[D_2]$ be the elements of $\mathsf{N}^1(X)$ defined by

$$\begin{aligned}[D_1] &= 6[H] - [E_9] - [E_{10}] - \sum_{i=1}^{8} 2[E_i],\\ [D_2] &= 6[H] - [E_7] - [E_8] - \sum_{i\neq 7,8} 2[E_i].\end{aligned}$$

Both of them have self intersection 2, are contained in W, and intersect $[H]$ positively; as explained in Remark 5.18, $[D_1]$ and $[D_2]$ are ample classes of X_0. Let V be the plane containing $[D_1]$ and $[D_2]$, and Q be the restriction of the intersection form to V. If u and v are integers, then

$$Q(u[D_1] + v[D_2]) = (u[D_1] + v[D_2]) \cdot (u[D_1] + v[D_2]) = 2u^2 + 8uv + 2v^2$$

because $[D_1] \cdot [D_2] = 4$. This quadratic form does not represent 0, because its discriminant is not a perfect square. From Lemma 5.19, there is an isometry ϕ of V with an eigenvalue $\lambda(\phi) > 1$. An explicit computation shows that the group of isometries of Q is the semi-direct product of a cyclic group $\mathbf{Z}$, generated by ϕ, with

$$\phi([D_1]) = 4[D_1] + [D_2], \quad \phi([D_2]) = -[D_1],$$

and the group $\mathbf{Z}/2\mathbf{Z}$ generated by the involution which permutes $[D_1]$ and $[D_2]$. The second iterate of ϕ is equal to the identity modulo 2. Coble's theorem (Theorem 5.17) now implies that there exists a unique automorphism g of X such that

(1) g_* coincides with ϕ^2 on V;
(2) g_* is the identity on the orthogonal complement $V^\perp$.

The dynamical degree of g is the square of $\lambda(\phi)$, and is equal to $7 + 4\sqrt{3}$. Properties (i), (ii), and (iii) of Proposition 5.13 are satisfied. Thus, large powers of g generate non trivial normal subgroups of the Cremona group:

Theorem 5.20. *Let $\mathbf{k}$ be an algebraically closed field. Let X be a generic Coble surface defined over $\mathbf{k}$. There are hyperbolic automorphisms of X that generate non trivial normal subgroups of the Cremona group $\mathsf{Bir}(X_{\mathbf{k}}) = \mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$.*

As a corollary, the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ is not simple if $\mathbf{k}$ is algebraically closed, as announced in the Introduction.

6. Complements

6.1. **Polynomial automorphisms and monomial transformations.** The group of polynomial automorphisms of the affine plane, and the group of monomial transformations of $\mathbb{P}^2_{\mathbf{C}}$ were both sources of inspiration for the results in this paper. We now use these groups to construct hyperbolic elements g of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ for which $\langle\langle g\rangle\rangle$ coincides with $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$.

6.1.1. *Monomial transformations.* Consider the group of monomial transformations of $\mathbb{P}^2_{\mathbf{k}}$. By definition, this group is isomorphic to $\mathsf{GL}_2(\mathbf{Z})$, acting by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : (x,y) \mapsto (x^a y^b, x^c y^d)$$

in affine coordinates (x, y). The matrix $-\mathrm{Id}$ corresponds to the standard quadratic involution $\sigma(x, y) = (1/x, 1/y)$.

If one considers $\mathsf{PSL}_2(\mathbf{Z})$ as a subgroup of $\mathsf{PSL}_2(\mathbf{R}) \simeq \mathrm{Isom}(\mathbb{H}^2)$, it is an interesting exercise to check that all hyperbolic matrices $\mathsf{PSL}_2(\mathbf{Z})$ are tight elements of $\mathsf{PSL}_2(\mathbf{Z})$. However, when we see $\mathsf{GL}_2(\mathbf{Z})$ as a subgroup of the Cremona group, we obtain the following striking remark.

Proposition 6.1. *The normal subgroup of* $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ *generated by any monomial transformation is never proper: If g is not the identity, then* $\langle\langle g\rangle\rangle = \mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$.

Remark 6.2 (Gizatullin and Noether)**.** If N is a normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ containing a non trivial automorphism of $\mathbb{P}^2_{\mathbf{k}}$, then N coincides with $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$. The proof is as follows. Since $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}})$ is the simple group $\mathsf{PGL}_3(\mathbf{k})$ and N is normal, N contains $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}})$. In particular, N contains the automorphism h defined by

$$h(x, y) = (1 - x, 1 - y)$$

in affine coordinates. An easy calculation shows that the standard quadratic involution σ satisfies $\sigma = (h\sigma)h(h\sigma)^{-1}$; hence, σ is conjugate to h, and σ is contained in N. The conclusion follows from Noether's theorem, which states that σ and $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}})$ generate $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$ (see [32], §2.5).

Proof of Proposition 6.1. Let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be any non trivial monomial map in $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$. The commutator of g with the diagonal map $f(x, y) = (\alpha x, \beta y)$ is the diagonal map

$$g^{-1}f^{-1}gf : (x,y) \mapsto (\alpha^{1-d}\beta^b x, \alpha^c \beta^{1-a} y). \tag{6.1}$$

Thus, the normal subgroup $\langle\langle g\rangle\rangle$ contains an element of $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{C}}) \setminus \{\mathrm{Id}\}$ and Remark 6.2 concludes the proof. □

6.1.2. *Polynomial automorphisms.* As mentioned in the Introduction, Danilov proved that the group $\mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ of polynomial automorphisms of the affine plane $\mathbb{A}^2_{\mathbf{k}}$ with Jacobian determinant one is not simple. Danilov's proof uses an action on a tree. Since $\mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ is a subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$, we also have the action of $\mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ on the hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$. It is a nice observation that $g \in \mathsf{Aut}[\mathbb{A}^2_{\mathbf{k}}]_1$ determines a hyperbolic isometry of the tree if and only if it determines a hyperbolic isometry of $\mathbb{H}_{\bar{\mathcal{Z}}}$: Hyperbolicity corresponds to an exponential growth of the sequence of degrees $\deg(g^n)$.

Fix a number $a \in \mathbf{k}^*$ and a polynomial $p \in \mathbf{k}[y]$ of degree $d \geq 2$, and consider the automorphism of $\mathbb{A}^2_{\mathbf{k}}$ defined by $h(x, y) = (y, p(y) - ax)$. This automorphism determines an algebraically stable birational transformation of $\mathbb{P}^2_{\mathbf{k}}$, namely

$$h[x : y : z] = [yz^{d-1} : P(y, z) - axz^{d-1} : z^d], \tag{6.2}$$

where $P(y, z) = p(y/z)z^d$. There is a unique indeterminacy point $\mathrm{Ind}(h) = \{[1 : 0 : 0]\}$, and a unique indeterminacy point for the inverse, $\mathrm{Ind}(h^{-1}) = \{[0 : 1 : 0]\}$. This Cremona transformation is hyperbolic, with translation length $L(h_*) = \log(d)$; in particular, the translation length goes to infinity with d.

Proposition 6.3. *For all integers $d \geq 2$, equation (6.2) defines a subset $H_d \subset \mathsf{V}_d$ which depends on $d{+}2$ parameters and satisfies: For all h in H_d, h is a hyperbolic, algebraically stable Cremona transformation, but the normal subgroup generated by h coincides with $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$.*

Proof. The automorphism h is the composition of the de Jonquières transformation $(x, y) \mapsto (P(y) - ax, y)$ and the linear map $(x, y) \mapsto (y, x)$. As such, h is an element of V_d. If f denotes the automorphism $f(x, y) = (x, y + 1)$, then

$$(h^{-1} \circ f \circ h)(x, y) = (x - a^{-1}, y)$$

is linear (thus, the second step in the proof of proposition 5.7 does not work for h). As a consequence, the commutator $f^{-1}h^{-1}fh$ is linear and $\langle\langle h \rangle\rangle$ intersects $\mathsf{Aut}(\mathbb{P}^2_{\mathbf{k}})$ non trivially. The conclusion follows from Remark 6.2. □

Note that for h in H_d and large integers n, we expect $\langle\langle h^n \rangle\rangle$ to be a proper normal subgroup of the Cremona group:

Question 6.4. Let $\mathbf{k}$ be any field. Consider the polynomial automorphism

$$g\colon (x, y) \mapsto (y, y^2 + x).$$

Does there exist an integer $n > 0$ (independent of $\mathbf{k}$) such that $\langle\langle g^n \rangle\rangle$ is a proper normal subgroup of $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{k}})$?

The main point would be to adapt Step 2 in the proof of Proposition 5.7.

6.2. **Projective surfaces.** The reason why we focused on the group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ comes from the fact that $\mathsf{Bir}(X)$ is small compared to $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ when X is an irrational complex projective surface. The proof of the following proposition illustrates this property.

Proposition 6.5. *Let X be a complex projective surface. If the group $\mathsf{Bir}(X)$ is infinite and simple, then X is birationally equivalent to $C \times \mathbb{P}^1(\mathbf{C})$ where C is a curve with trivial automorphism group.*

Sketch of the proof. Assume, first, that the Kodaira dimension of X is non negative. Replace X by its unique minimal model, and identify $\mathsf{Bir}(X)$ with $\mathsf{Aut}(X)$. The group $\mathsf{Aut}(X)$ acts on the homology of X, and the kernel is equal to its connected component $\mathsf{Aut}(X)^0$ up to a finite index group. The action on the homology group provides a morphism to $\mathsf{GL}_n(\mathbf{Z})$ for some $n \geq 1$. Reducing modulo p for large primes p, one sees that $\mathsf{GL}_n(\mathbf{Z})$ is residually finite. Since $\mathsf{Aut}(X)$ is assumed to be simple, this implies that $\mathsf{Aut}(X)$ coincides with $\mathsf{Aut}(X)^0$. But $\mathsf{Aut}(X)^0$ is abelian for surfaces with non negative Kodaira dimension (see [1]). Thus $\mathsf{Bir}(X) = \mathsf{Aut}(X)$ is not both infinite and simple when the Kodaira dimension of X is ≥ 0. Assume now that X is ruled and not rational. Up to a birational change of coordinates, X is a product $\mathbb{P}^1_{\mathbf{C}} \times C$ where C is a smooth curve of genus $g(C) \geq 1$. The group $\mathsf{Bir}(X)$ projects surjectively onto $\mathsf{Aut}(C)$. By simplicity, $\mathsf{Aut}(C)$ must be trivial. In that case, $\mathsf{Bir}(X)$ coincides with the infinite simple group $\mathsf{PGL}_2(\mathcal{M}(C))$ where $\mathcal{M}(C)$ is the field of meromorphic functions of C. The remaining case is when X is rational, and Theorem A concludes the proof. □

6.3. **SQ-universality and the number of quotients.** A group is said to be **SQ-universal** (or SubQuotient-universal) if any countable group can be embedded into one of its quotients. For example, the pioneering work [30] proves that the free group over two generators is SQ-universal. If G is a non elementary hyperbolic group, then G is SQ-universal. This result has been obtained by Delzant and Olshanskii in [13] and [38]. It seems reasonable to expect that the Cremona group $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ is also SQ-universal, with a proof similar to Delzant and Olshanskii's proofs, but for $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ acting on $\mathbb{H}_{\bar{\mathcal{Z}}}$ instead of a hyperbolic group acting on itself. Dahmani and Guirardel [11] have announced a general result which, thanks to what is proved in our paper, should easily imply the SQ-universality of the Cremona group. We thus refer to a forthcoming paper by Dahmani and Guirardel, together with Osin, for SQ-universality. Note that SQ-universality implies the existence of an uncountable number of non isomorphic quotients. The following weaker result is a direct consequence of Theorem A.

Proposition 6.6. *The Cremona group* $\mathsf{Bir}(\mathbb{P}^2_{\mathbf{C}})$ *has an uncountable number of distinct normal subgroups.*

References

[1] Dmitri N. Akhiezer. *Lie group actions in complex analysis*. Friedr. Vieweg & Sohn, 1995.

[2] Jérémy Blanc. Sous-groupes algébriques du groupe de Cremona. *Transform. Groups*, 14(2):249–285, 2009.

[3] Jérémy Blanc. Groupes de Cremona, connexité et simplicité. *To appear in Ann. Sci. ENS*, 2010.

[4] Sébastien Boucksom, Charles Favre, and Mattias Jonsson. Degree growth of meromorphic surface maps. *Duke Math. J.*, 141(3):519–538, 2008.

[5] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer-Verlag, Berlin, 1999.

[6] Serge Cantat. Sur les groupes de transformations birationnelles des surfaces. *Annals of Math., to appear*, 2007.

[7] Serge Cantat and Stéphane Lamy. Normal subgroups of the Cremona group. *Preprint arXiv*, pages 1–53, 2010.

[8] Dominique Cerveau and Julie Déserti. Transformations birationnelles de petit degré. *Preprint*, 2009.

[9] M. Coornaert, T. Delzant, and A. Papadopoulos. *Géométrie et théorie des groupes*, volume 1441 of *Lecture Notes in Mathematics*. Springer-Verlag, 1990.

[10] François R. Cossec and Igor V. Dolgachev. *Enriques surfaces. I*, volume 76 of *Progress in Mathematics*. Birkhäuser, 1989.

[11] F. Dahmani and V. Guirardel. *Preprint ???*, 2010.

[12] V. I. Danilov. Non-simplicity of the group of unimodular automorphisms of an affine plane. *Mat. Zametki*, 15:289–293, 1974.

[13] Thomas Delzant. Sous-groupes distingués et quotients des groupes hyperboliques. *Duke Math. J.*, 83(3):661–682, 1996.

[14] Julie Déserti. Groupe de Cremona et dynamique complexe: une approche de la conjecture de Zimmer. *Int. Math. Res. Not.*, pages Art. ID 71701, 27, 2006.

[15] Julie Déserti. Sur les automorphismes du groupe de Cremona. *Compos. Math.*, 142(6):1459–1478, 2006.

[16] J. Diller and C. Favre. Dynamics of bimeromorphic maps of surfaces. *Amer. J. Math.*, 123(6):1135–1169, 2001.

[17] I. Dolgachev. On automorphisms of Enriques surfaces. *Invent. Math.*, 76(1):163–177, 1984.

[18] I. Dolgachev. Infinite Coxeter groups and automorphisms of algebraic surfaces. In *The Lefschetz centennial conference, Part I (Mexico City, 1984)*, volume 58 of *Contemp. Math.*, pages 91–106. Amer. Math. Soc., 1986.

[19] Igor Dolgachev and David Ortland. Point sets in projective spaces and theta functions. *Astérisque*, (165), 1988.

[20] Igor V. Dolgachev. Reflection groups in algebraic geometry. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):1–60, 2008.

[21] Igor V. Dolgachev and De-Qi Zhang. Coble rational surfaces. *Amer. J. Math.*, 123(1):79–114, 2001.

[22] Charles Favre. Sous-groupes de type fini du groupe de Cremona. *Exposé au séminaire Bourbaki*, 2008.

[23] A. Fröhlich and M. J. Taylor. *Algebraic number theory*. Cambridge University Press, 1993.

[24] Jean-Philippe Furter and Stéphane Lamy. Normal subgroup generated by a plane polynomial automorphism. *To appear in Transformation Groups*, 2010.

[25] Étienne Ghys and Pierre de la Harpe. *Sur les groupes hyperboliques d'après Mikhael Gromov*, volume 83 of *Progr. Math.* Birkhäuser Boston, 1990.

[26] M. H. Gizatullin. Rational G-surfaces. *Izv. Akad. Nauk SSSR Ser. Mat.*, 44(1):110–144, 239, 1980.
[27] Vincent Guedj. Propriétés ergodiques des applications rationnelles. *Panorama de la Société mathématique deFrance, to appear*, 30:1–130, 2007.
[28] G. Halphen. Sur les courbes planes du sixième degré à neuf points doubles. *Bull. Soc. Math. France*, 10:162–172, 1882.
[29] Robin Hartshorne. *Algebraic geometry.* Springer-Verlag, 1977. Graduate Texts in Mathematics, No. 52.
[30] Graham Higman, B. H. Neumann, and Hanna Neumann. Embedding theorems for groups. *J. London Math. Soc.*, 24:247–254, 1949.
[31] John Hamal Hubbard. *Teichmüller theory and applications to geometry, topology, and dynamics. Vol. 1.* Matrix Editions, Ithaca, NY, 2006.
[32] K. Kollá, J. Smith and A. Corti. *Rational and Nearly Rational Varieties.* Cambridge University Press, 2004.
[33] Stéphane Lamy. Une preuve géométrique du théorème de Jung. *Enseign. Math. (2)*, 48(3-4):291–315, 2002.
[34] Robert Lazarsfeld. *Positivity in algebraic geometry. I*, volume 48. Springer-Verlag, Berlin, 2004. Classical setting: line bundles and linear series.
[35] Yu. I. Manin. *Cubic forms*, volume 4 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam, second edition, 1986.
[36] David Mumford. Algebraic geometry in mathematical developments arising from hilbert problems. In *(Proc. Sympos. Pure Math., Vol. XXVIII, Northern Illinois Univ., De Kalb, Ill., 1974)*, pages 431–444. Amer. Math. Soc., Providence, R. I., 1976.
[37] Dat Dang Nguyen. Composantes irréductibles des transformations de Cremona de degré *d. Thèse de Mathématique, Univ. Nice, France*, 2009.
[38] A. Yu. Ol′shanskiĭ. SQ-universality of hyperbolic groups. *Mat. Sb.*, 186(8):119–132, 1995.
[39] Jean-Pierre Serre. Le groupe de Cremona et ses groupes finis. *Exposé au séminaire Bourbaki*, 2008.

Département de mathématiques, Université de Rennes, Rennes, France
E-mail address: serge.cantat@univ-rennes1.fr

Mathematics Institute, University of Warwick, Coventry, United Kingdom
E-mail address: s.lamy@warwick.ac.uk

(on leave from Institut Camille Jordan, Université Lyon 1, France
E-mail address: lamy@math.univ-lyon1.fr)