

# Normal subgroups in the Cremona group

by

SERGE CANTAT

*Université de Rennes I  
Rennes, France*

STÉPHANE LAMY

*University of Warwick  
Coventry, U.K.*

With an appendix by Yves de Cornulier.

## Contents

1. Introduction . . . . .	31
Part I. Hyperbolic geometry and normal subgroups . . . . .	37
2. Gromov hyperbolic spaces . . . . .	37
3. Hyperbolic spaces with constant negative curvature . . . . .	57
Part II. Algebraic geometry and the Cremona group . . . . .	62
4. The Picard–Manin space . . . . .	62
5. Tight birational maps . . . . .	69
6. Complements . . . . .	84
Appendix A. The Cremona group is not an amalgam . . . . .	88
References . . . . .	92

## 1. Introduction

### 1.1. The Cremona group

The Cremona group in  $n$  variables over a field  $\mathbf{k}$  is the group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^n)$  of birational transformations of the projective space  $\mathbb{P}_{\mathbf{k}}^n$  or, equivalently, the group of  $\mathbf{k}$ -automorphisms of the field  $\mathbf{k}(x_1, \dots, x_n)$  of rational functions of  $\mathbb{P}_{\mathbf{k}}^n$ . In dimension  $n=1$ , the Cremona group coincides with the group  $\mathrm{Aut}(\mathbb{P}_{\mathbf{k}}^1)$  of regular automorphisms of  $\mathbb{P}_{\mathbf{k}}^1$ , that is with the group  $\mathrm{PGL}_2(\mathbf{k})$  of linear projective transformations. When  $n \geq 2$ , the group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^n)$  is infinite-dimensional; it contains algebraic groups of arbitrarily large dimensions (see [2] and [52]). In this article, we prove the following theorem.

---

The second author was supported by a Marie Curie Intra-European Fellowship, on leave from the Institut Camille Jordan, Université Lyon 1.

MAIN THEOREM. *If  $\mathbf{k}$  is an algebraically closed field, the Cremona group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  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  $\mathrm{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  is a simple group in the 1960s; Mumford mentioned also this question in the early seventies (see [47]); in fact, one can trace this problem back to 1894, since Enriques already mentions it in [27]: “*l’importante questione se il gruppo Cremona contenga alcun sottogruppo invariante (questione alla quale sembra probabile si debba rispondere negativamente)*”.

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 spaces  $\mathrm{Bir}_d(\mathbb{P}_{\mathbf{k}}^2)$  of birational transformations of degree  $d$  (see [52], [2] and the references therein). Blanc proved in [3] that  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  has only two *closed* normal subgroups for this topology, namely  $\{\mathrm{Id}\}$  and  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  itself. Déserti proved that  $\mathrm{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  is perfect, hopfian and cohopfian, and that its automorphism group is generated by inner automorphisms and the action of automorphisms of the field of complex numbers (see [18]–[20]). In particular, there is no obvious algebraic reason which explains why  $\mathrm{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  is not simple.

(b) An interesting example is given by the group  $\mathrm{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  of polynomial automorphisms of the affine plane  $\mathbb{A}_{\mathbf{k}}^2$  with Jacobian determinant equal to 1. From Jung’s theorem, one knows that this group is the amalgamated product of the group  $\mathrm{SL}_2(\mathbf{k}) \rtimes \mathbf{k}^2$  of affine transformations with the group of elementary transformations

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

over their intersection (see [40], [42] and the references therein); as such,  $\mathrm{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  acts faithfully on a simplicial tree without fixing any edge or vertex. Danilov used this action, together with small cancellation theory, to prove that  $\mathrm{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  is not simple (see [16] and [31]).

(c) The group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  contains a linear group with Kazhdan property (T), namely  $\mathrm{PGL}_3(\mathbf{C})$ , and it cannot be written as a non-trivial free product with amalgamation: see Appendix A of this paper. 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  $\mathbb{P}_{\mathbf{k}}^2$ . Let  $f$  be a birational transformation of  $\mathbb{P}_{\mathbf{k}}^2$ . There are three homogeneous polynomials  $P, Q, R \in \mathbf{k}[x, y, z]$  of the same degree  $d$ , and without common factor of degree  $\geq 1$ , 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  $\text{Bir}_d(\mathbb{P}_{\mathbf{k}}^2)$  of birational transformations of degree  $d$  is a quasi-projective variety. For instance, Cremona transformations of degree 1 correspond to the automorphism group  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2) = \text{PGL}_3(\mathbf{k})$  and Cremona transformations of degree 2 form an irreducible variety of dimension 14 (see [9]).

We define the *de Jonquières group*  $J$  as the group of birational transformations of the plane  $\mathbb{P}_{\mathbf{k}}^2$  that 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  $\text{Bir}_d(\mathbb{P}_{\mathbf{k}}^2)$  whose elements are compositions  $h_1 \circ f \circ h_2$ , where  $h_1$  and  $h_2$  are in  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2)$  and  $f$  is in  $J_d$ . The dimension of  $\text{Bir}_d(\mathbb{P}_{\mathbf{k}}^2)$  is equal to  $4d+6$  and  $V_d$  is its unique irreducible component of maximal dimension (see [48]).

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  $\geq 1$ , and is said to be *general* if it is satisfied on the complement of countably many Zariski closed subsets of  $W$  of codimension  $\geq 1$ .

**THEOREM A.** *There exists a positive integer  $k$  with the following property. Let  $d \geq 2$  be an integer. If  $g$  is a general element of  $\text{Bir}_d(\mathbb{P}_{\mathbf{C}}^2)$ , and  $n$  is an integer with  $n \geq k$ , then  $g^n$  generates a normal subgroup of the Cremona group  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  that does not contain any element  $f$  of  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2) \setminus \{\text{Id}\}$  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  $\text{Bir}_d(\mathbb{P}_{\mathbf{C}}^2)$ . As a corollary of Theorem A, the group  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  contains an uncountable number of normal subgroups (see §6.3).

*Remark 1.2.* (a) Gizatullin proved that any non-trivial element  $g \in \text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  preserving a pencil of lines generates a normal subgroup equal to  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$ ; the same conclusion holds for any  $g \in \text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  with  $\deg(g) \leq 7$  (see [34], or [9] when  $\deg(g) \leq 2$ ).

(b) We provide an explicit upper bound for the best possible constant  $k$  in Theorem A, namely  $k \leq 201021$ . We do so to help the reader follow the proof. According to Gizatullin's result, if  $g$  is an element of  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2) \setminus \{\text{Id}\}$  with  $\deg(g) \leq 7$ , then the smallest

normal subgroup of  $\mathrm{Bir}(\mathbb{P}_{\mathbb{C}}^2)$  containing  $g$  coincides with  $\mathrm{Bir}(\mathbb{P}_{\mathbb{C}}^2)$ . Thus, the best constant  $k$  is larger than 1. Our method gives a rather small constant for large enough degrees, namely  $k \leq 375$ , but does not answer the following question: Does a general element of  $\mathrm{Bir}_d(\mathbb{P}_{\mathbb{C}}^2)$  generate a strict normal subgroup of  $\mathrm{Bir}(\mathbb{P}_{\mathbb{C}}^2)$  for all large enough degrees  $d$ ? Note that  $g^n$  is not an element of the maximal-dimensional component  $V_{d^n}$  if  $n \geq 2$  and  $g \in \mathrm{Bir}_d(\mathbb{P}_{\mathbb{C}}^2)$ .

### 1.3. Automorphisms of Kummer and Coble surfaces

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

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

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 quotients of abelian surfaces. Due to a result by Coble and Dolgachev, for any algebraically closed field  $\mathbf{k}$  we obtain automorphisms of Coble surfaces that generate proper normal subgroups in the Cremona group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ . The main theorem follows from this construction (see Theorem 5.20).

### 1.4. An infinite-dimensional hyperbolic space (see §4)

The Cremona group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  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  $\pi: X \rightarrow \mathbb{P}_{\mathbf{k}}^2$  obtained from the projective plane by successive blow-ups. If  $\pi': X' \rightarrow \mathbb{P}_{\mathbf{k}}^2$  is obtained from  $\pi: X \rightarrow \mathbb{P}_{\mathbf{k}}^2$  by blowing up more points, then there is a natural birational morphism  $\varphi: X' \rightarrow X$ , defined by  $\varphi = \pi^{-1} \circ \pi'$ , and the pull-back operator  $\varphi^*$  embeds the Néron–Severi group  $\mathbf{N}^1(X) \otimes \mathbf{R}$  into  $\mathbf{N}^1(X') \otimes \mathbf{R}$  (note that  $\mathbf{N}^1(X) \otimes \mathbf{R}$  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  $\mathbf{N}^1(X) \otimes \mathbf{R}$  is called the *Picard–Manin space* of  $\mathbb{P}_{\mathbf{k}}^2$ . 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 this quadratic form 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 which contains classes of ample divisors of rational surfaces; this connected component is an infinite hyperbolic space for the distance  $\text{dist}$  defined in terms of the intersection form by

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

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

### 1.5. Normal subgroups in isometry groups

In Part I, we study the general situation of a group  $G$  acting by isometries on a  $\delta$ -hyperbolic space  $\mathcal{H}$ . Let us explain the content of the central result of Part I, namely Theorem 2.10, in the particular case of the Cremona group  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  acting by isometries on the hyperbolic space  $\mathbb{H}_{\bar{\mathbb{Z}}}$ . Isometries of hyperbolic spaces fall into three types: elliptic, parabolic and hyperbolic. A Cremona transformation  $g \in \text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  determines a hyperbolic isometry  $g_*$  of  $\mathbb{H}_{\bar{\mathbb{Z}}}$  if and only if the following equivalent properties are satisfied:

- The sequence of degrees  $\deg(g^n)$  grows exponentially fast:

$$\lambda(g) := \limsup_{n \rightarrow \infty} \deg(g^n)^{1/n} > 1;$$

- There is a  $g_*$ -invariant plane  $V_g$  in the Picard–Manin space that intersects  $\mathbb{H}_{\bar{\mathbb{Z}}}$  on a curve  $\text{Ax}(g_*)$  (a geodesic line) on which  $g_*$  acts by a non-trivial translation; more precisely,  $\text{dist}(x, g_*(x)) = \log \lambda(g)$  for all  $x \in \text{Ax}(g_*)$ .

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

The second property is a rigidity property of  $\text{Ax}(g)$  with respect to isometries  $f_*$ , for  $f \in \text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ . The third property means that the stabilizer of  $\text{Ax}(g)$  coincides with the normalizer of the cyclic group  $g^{\mathbb{Z}}$ . Applied to the Cremona group, Theorem 2.10 gives the following statement.

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

Theorems A and B are then deduced from the fact that general elements of  $\text{Bir}_d(\mathbb{P}_{\mathbb{C}}^2)$  (resp. automorphisms  $g_M$  on rational Kummer surfaces with  $|\text{tr}(M)| \geq 3$ ) are tight elements of the Cremona group (see Theorems 5.14 and 5.20).

*Remark 1.3.* Theorem C will be proved in the context of groups of isometries of Gromov's  $\delta$ -hyperbolic spaces; the strategy is similar to the one used by Delzant in [17] to construct normal subgroups in Gromov hyperbolic groups. As such, Theorem C is part of the so-called small cancellation theory; we shall explain and comment on this in §2.

## 1.6. Description of the paper

The paper starts with the proof of Theorem C in the general context of  $\delta$ -hyperbolic spaces (§2), and explains how this general statement can be used in the case of isometry groups of spaces with constant negative curvature (§3).

In §4, we provide an overview on the Picard–Manin space, the associated hyperbolic space  $\mathbb{H}_{\bar{\mathbb{Z}}}$  and the isometric action of  $\text{Bir}(\mathbb{P}_{\mathbb{C}}^2)$  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 a few remarks and comments. An appendix by de Cornulier proves that the Cremona group cannot be written as a non-trivial amalgamated product.

### 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. Marat Gizatullin and the anonymous referees made important comments on the first version of this article, which helped clarify the proof and the exposition.

## Part I. Hyperbolic geometry and normal subgroups

### 2. Gromov hyperbolic spaces

This section is devoted to the proof of Theorem 2.10, 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  $(\mathcal{H}, d)$  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

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

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

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

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

$$d(w, x) + d(y, z) \leq \max\{d(w, y) + d(x, z), d(w, z) + d(x, y)\} + 2\delta. \quad (2.1)$$

Two fundamental examples of  $\delta$ -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  $\delta$ -hyperbolic for  $\delta = \log 3$  (see [32, Chapter 2] and [13]).

##### 2.1.2. Geodesics

A *geodesic segment* from  $x$  to  $y$  in  $\mathcal{H}$  is an isometry  $\gamma$  from an interval  $[s, t] \subset \mathbf{R}$  into  $\mathcal{H}$  such that  $\gamma(s) = x$  and  $\gamma(t) = y$ . The metric space  $(\mathcal{H}, d)$  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  $\gamma: [0, l] \rightarrow \mathcal{H}$  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  $\gamma$  from  $\mathbf{R}$  to its image  $\Gamma \subset \mathcal{H}$ . If  $x$  is a point of  $\mathcal{H}$  and  $\Gamma$  is (the image of) a geodesic line, a *projection* of  $x$  onto  $\Gamma$  is a point  $z \in \Gamma$  which realizes the distance from  $x$  to  $\Gamma$ . Projections exist but are not necessarily unique (in a tree, or in  $\mathbb{H}^n$ , the projection is unique). If  $\mathcal{H}$  is  $\delta$ -hyperbolic and if  $z$  and  $z'$  are two projections of a point  $x$  on a geodesic  $\Gamma$ , then the distance between  $z$  and  $z'$  is at most  $4\delta$ . In what follows, we shall denote by  $\pi_\Gamma(x)$  any projection of  $x$  on  $\Gamma$ . Since all small errors will be written in terms of  $\theta = 8\delta$ , we note that the projection  $\pi_\Gamma(x)$  is well defined up to an error of  $\frac{1}{2}\theta$ .

## 2.2. Classical results from hyperbolic geometry

In what follows,  $\mathcal{H}$  will be a  $\delta$ -hyperbolic geodesic metric space.

### 2.2.1. Approximation by trees

In a  $\delta$ -hyperbolic space, the geometry of finite subsets is well approximated by finite subsets of metric trees; we refer to [32, Chapter 2], or [13], for a proof of this basic and crucial fact.

LEMMA 2.1. (Approximation by trees) *Let  $\mathcal{H}$  be a  $\delta$ -hyperbolic space and  $(x_0, \dots, x_n)$  be 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  $x_i$ ,  $1 \leq i \leq n$ . Let  $k$  be the smallest integer such that  $2n \leq 2^k + 1$ . Then there exist a metric tree  $T$  and a map  $\Phi: X \rightarrow T$  such that*

- (1)  $\Phi$  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  $x, y \in X$ ,

$$d(x, y) - 2k\delta \leq d(\Phi(x), \Phi(y)) \leq d(x, y).$$

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

*Convention.* When  $n \leq 4$  we can choose  $k = 3$ . Let  $\theta$  be the positive real number defined by

$$\theta = 8\delta.$$



We have

$$d(x, y) - \theta \leq d(\Phi(x), \Phi(y)) \leq d(x, y). \quad (2.2)$$

From now on we fix such a  $\theta$ , and we always use the approximation lemma in this way for at most five points  $(x_0, \dots, x_4)$ . 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  $\delta$ -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  $(\Phi, T)$  associated with  $(x_0, \dots, x_n)$ . In general, the choice of the segments between  $x_0$  and the  $x_i$  is either clear, or irrelevant.

In the rest of §2.2, well-known facts from hyperbolic geometry are listed; complete proofs can be found in [32], [13], or in the preprint version of this paper [8]. First, the following corollary is an immediate consequence of the approximation lemma, and Lemma 2.3 follows from the corollary.

**COROLLARY 2.2.** *Let  $\Phi: X \rightarrow T$  be an approximation tree for at most five points. Then*

$$(\Phi(x) | \Phi(y))_{\Phi(z)} - \frac{1}{2}\theta \leq (x | y)_z \leq (\Phi(x) | \Phi(y))_{\Phi(z)} + \theta$$

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

**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  $\Gamma$ . For all  $b \in \Gamma$  and  $c$  in the segment  $[a, b] \subset \Gamma$ , we have  $(x | b)_c \leq 2\theta$ .*

### 2.2.2. Shortening and weak convexity

The following lemma says that two geodesic segments are close as soon as their extremities are not too far. Lemma 3.1 makes this statement much more precise in the context of  $\text{CAT}(-1)$  spaces.

**LEMMA 2.4.** (A little shorter, much closer. See [17, Proposition 1.3.3]) *Let  $[x, y]$  and  $[x', y']$  be two geodesic segments of  $\mathcal{H}$  such that*

- (i)  $d(x, x') \leq \beta$  and  $d(y, y') \leq \beta$ ;
- (ii)  $d(x, y) \geq 2\beta + 4\theta$ .

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

The following lemma is obvious in a  $\text{CAT}(0)$  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  $d(x, x')$  and  $d(y, y')$  are bounded from above by  $\beta$ . Then  $[x', y']$  is in the  $(\beta+2\theta)$ -neighborhood of  $[x, y]$ .*

### 2.2.3. Canoeing to infinity

The following lemma comes from [17, §1.3.4], and says that hyperbolic canoes never make loops (see [39, §2.3.13]). This lemma is at the heart of our proof of Theorem C.

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

- (i)  $d(y_i, y_{i-1}) \geq 10\theta$  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)  $d(y_0, y_j) \geq d(y_0, y_{j-1}) + 2\theta$  if  $j \geq 1$ ;
- (2)  $d(y_0, y_j) \geq \sum_{i=1}^j (d(y_i, y_{i-1}) - 7\theta)$ ;
- (3)  $y_j$  is  $5\theta$ -close to any geodesic segment  $[y_0, y_n]$ .

## 2.3. Rigidity of axis and non-simplicity criterion

Let  $G$  be a group of isometries of a  $\delta$ -hyperbolic space  $(\mathcal{H}, d)$ , 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

If  $f \in \text{Isom}(\mathcal{H})$  is an isometry of a  $\delta$ -hyperbolic space  $(\mathcal{H}, d)$ , we define the *translation length*  $L(f)$  as the limit

$$\lim_{n \rightarrow \infty} \frac{d(x, f^n(x))}{n},$$

where  $x$  is an arbitrary point of  $\mathcal{H}$  (see [13, p. 117]). We denote by  $\text{Min}(f)$  the set of points  $y \in \mathcal{H}$  such that  $d(y, f(y)) = L(f)$ .

There are three types of isometries  $f \in \text{Isom}(\mathcal{H})$ , termed *elliptic*, *parabolic* and *hyperbolic*. An isometry  $f$  is *elliptic* if it admits a bounded orbit (and hence all orbits are bounded). If the orbits of  $f$  are unbounded, we say that  $f$  is *hyperbolic* if  $L(f) > 0$ , and otherwise that  $f$  is *parabolic*.

Say that  $f$  has an *invariant axis* if there is a geodesic line  $\Gamma$  such that  $f(\Gamma) = \Gamma$ . If  $f$  has an invariant axis, then  $f$  is either elliptic or hyperbolic, since the restriction  $f|_{\Gamma}$  is either the identity, a symmetry or a translation.

Suppose that  $f$  is hyperbolic. If it has an invariant axis  $\Gamma$ , this axis is contained in  $\text{Min}(f)$  and  $f$  acts as a translation of length  $L(f)$  along  $\Gamma$ . If  $\Gamma$  and  $\Gamma'$  are two invariant axes, each of them is in the  $2\theta$ -tubular neighborhood of the other (apply Lemma 2.4). When  $f$  is hyperbolic and has an invariant axis, we denote by  $\text{Ax}(f)$  any invariant geodesic line, even if such a line is a priori not unique.

*Remark 2.7.* If  $(\mathcal{H}, d)$  is complete,  $\text{CAT}(0)$  and  $\delta$ -hyperbolic, every hyperbolic isometry has an invariant axis (see [5, Chapter II, Theorem 6.8]). When  $\mathcal{H}$  is a tree or a hyperbolic space  $\mathbb{H}^n$ , the set  $\text{Min}(g)$  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; 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  $\alpha$*  of  $A$  and  $A'$  is the intersection of the tubular neighborhoods  $\text{Tub}_\alpha(A)$  and  $\text{Tub}_\alpha(A')$ ,

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

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

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

This rigidity property involves both the set  $A$  and the group  $G$ . The set  $A$  is  $\varepsilon$ -rigid if there exists a positive number  $B > 0$  such that  $A$  is  $(\varepsilon, B)$ -rigid. If  $\varepsilon' < \varepsilon$  and  $A \subset \mathcal{H}$  is  $\varepsilon$ -rigid, then  $A$  is also  $\varepsilon'$ -rigid (for the same constant  $B$ ). The converse holds for geodesic lines (or convex sets) when  $\varepsilon'$  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  $\Gamma$  is also  $(\varepsilon, B + 6\varepsilon + 4\theta)$ -rigid.*

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

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

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

### 2.3.3. Tight elements, small cancellation 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  $\text{Ax}(g) \subset \text{Min}(g)$ ;
- the geodesic line  $\text{Ax}(g)$  is  $2\theta$ -rigid;
- for all  $f \in G$ , if  $f(\text{Ax}(g)) = \text{Ax}(g)$  then  $fgf^{-1} = g$  or  $fgf^{-1} = g^{-1}$ .

We shall say that  $g \in G$  satisfies the *small cancellation property* if  $g \in G$  is tight, with a  $(14\theta, B)$ -rigid axis  $\text{Ax}(g)$  and

$$\frac{1}{20}L(g) \geq 60\theta + 2B.$$

Thus, the small cancellation property requires both tightness and a large enough translation length. We shall comment on this definition in §2.3.4 and explain how it is related to classical small cancellation properties.

*Remark 2.9.* If  $g$  is tight, then  $g^n$  is tight for all  $n \geq 1$ , and  $g^n$  satisfies the small cancellation property as soon as

$$nL(g) \geq 1200\theta + 40B.$$

We now state the main theorem of Part I; this result implies Theorem C from the introduction.

**THEOREM 2.10.** (Normal subgroup theorem) *Let  $G$  be a group acting by isometries on a hyperbolic space  $\mathcal{H}$ . Suppose that  $g \in G$  satisfies the small cancellation property. Then any element  $h \neq \text{Id}$  in the normal subgroup  $\langle\langle g \rangle\rangle \subset G$  satisfies the following alternative: Either  $h$  is a conjugate of  $g$ , or  $h$  is a hyperbolic isometry with translation length  $L(h) > L(g)$ . In particular, if  $n \geq 2$ , the normal subgroup  $\langle\langle g^n \rangle\rangle$  does not contain  $g$ .*

### 2.3.4. Small cancellation properties and complements

Assume that  $g$  is a hyperbolic element of  $G$  and  $\text{Ax}(g) \subset \text{Min}(g)$  is an axis of  $g$ . Let  $\lambda \in \mathbf{R}_+^*$  be a positive real number. One says that  $g$  satisfies the small cancellation property with parameter  $\lambda$ , or the  $\lambda$ -small cancellation property, if the following is satisfied: *For all elements  $f \in G$  with  $f(\text{Ax}(g)) \neq \text{Ax}(g)$ , the set of points of  $f(\text{Ax}(g))$  at distance  $\leq 4\theta = 32\delta$  from  $\text{Ax}(g)$  has diameter at most  $\lambda L(g)$ .*

When  $g$  is tight, with a  $(14\theta, B)$ -rigid axis, then  $g$  satisfies this small cancellation property as soon as  $\lambda L(g) \geq B$ . In particular,  $g^n$  satisfies the  $\lambda$ -small cancellation property when  $n \geq B/\lambda L(g)$ . Assume now that  $g$  satisfies the small cancellation property with parameter  $\lambda$ ; then  $g$  is tight with a  $(2\theta, \lambda L(g))$ -rigid axis. By Lemma 2.8, the axis of  $g$  is

$(14\theta, B)$ -rigid with  $B = \lambda L(g) + 88\theta$ . This shows that the less precise cancellation property used in Theorem 2.10 corresponds to the following, more classical, pair of assumptions:

- $g$  satisfies the  $\lambda$ -small cancellation property for some  $0 < \lambda < \frac{1}{40}$ ;
- $L(g) \geq 3640\theta / (1 - 40\lambda)$ .

Such hypotheses are well known in geometric group theory (see [17], [49] or [50] for example).

*Remark 2.11.* (a) Theorem 2.10 is similar to the main result of [17], which concerns the case of a hyperbolic group  $G$  acting on its Cayley graph; our strategy of proof, presented in §2.5, follows the same lines (see [17], and also the recent complementary article [10]). There are several results of this type in the literature but none of them seems to contain Theorem 2.10 as a corollary; in our setting, the space  $\mathcal{H}$  is not assumed to be locally compact, and the action of  $G$  is not assumed to be proper. This is crucial for application to the Cremona group: In this case, the space  $\mathcal{H}$  is locally homeomorphic to a non-separable Hilbert space, and the stabilizers of points in  $G$  are algebraic groups, like  $\mathrm{PGL}_3(\mathbf{k}) = \mathrm{Aut}(\mathbb{P}_{\mathbf{k}}^2)$ .

(b) Theorem 2.10 applies in particular to groups acting by isometries on trees: This includes the situations of a free group acting on its Cayley graph, and of an amalgamated product over two factors acting on its Bass–Serre tree. It is a useful exercise, which is done in [43], to write down the proof in this particular case; the proofs of the technical Lemmas 2.13 and 2.16 become much more transparent in this setting because one does not need approximations by trees.

(c) In the definition of a tight element, we could impose a weaker list of hypotheses. The main point would be to replace  $Ax(g)$  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 it makes the proof less technical; we refer to [17] or to the recent work [15] for more general viewpoints. We refer to §6.3 for extensions and improvements of Theorem 2.10.

We now prove Theorem 2.10. We first need to introduce the notion of an admissible presentation.

#### 2.4. Pieces, neutral segments and admissible presentations

Let  $g \in G$  be a hyperbolic isometry with an invariant axis; let  $Ax(g)$  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  $s(Ax(g))$  is an invariant axis for  $f$ .

### 2.4.1. Axis, pieces and neutral segments

Let  $[x, y] \subset \mathcal{H}$  be an oriented geodesic segment, the length of which is at least  $20\theta$ . The segment  $[x, y]$  is a *piece* if there exists an element  $s \in G$  such that  $[x, y]$  is contained in the tubular neighborhood  $\text{Tub}_{7\theta}(s(\text{Ax}(g)))$ . If  $[x, y]$  is a piece, the conjugates  $f = sgs^{-1}$  of  $g$  and  $f^{-1} = sg^{-1}s^{-1}$  of  $g^{-1}$  have the same invariant axis  $\Gamma = s(\text{Ax}(g))$ , and this axis almost contains  $[x, y]$ . Changing  $f$  to  $f^{-1}$  if necessary, we may assume that  $\pi_\Gamma(y)$  and  $f(\pi_\Gamma(x))$  lie on the same side<sup>(1)</sup> of  $\pi_\Gamma(x)$  in  $\Gamma$ . This assumption made, the isometry  $f$  is called a *support* of  $[x, y]$  (and so  $f^{-1}$  is a support of the piece  $[y, x]$ ). The segment  $[x, y]$  is a *piece of size  $p/q$*  if furthermore  $d(x, y) \geq pL/q$ . We say that  $[x, y]$  *contains a piece of size  $p/q$*  if there is a segment  $[x', y'] \subset [x, y]$  which is such a piece.

A pair of points  $(x, y)$  is *neutral* if none of the geodesic segments  $[x, y]$  between  $x$  and  $y$  contains a piece of size  $\frac{11}{20}$ ; 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  $y$ ). If  $[x, y]$  is neutral and  $f$  is an element of the group  $G$ , then  $f([x, y])$  is also neutral; in other words, being neutral is invariant by translation under isometries  $f \in G$ .

Our choice of working with  $\frac{1}{20}L$  as a unit has no particular significance; we use it for convenience.

### 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  $h = h_k h_{k-1} \dots h_1$ , where each  $h_j$  is a conjugate of  $g$  or its inverse: for all  $1 \leq i \leq k$  there exists  $s_i \in G$  such that

$$h_i = s_i g s_i^{-1} \quad \text{or} \quad s_i g^{-1} s_i^{-1}.$$

By convention, each  $h_i$  comes with  $s_i \in G$  and thus with an invariant axis

$$\text{Ax}(h_i) = s_i(\text{Ax}(g));$$

thus, the choice of  $s_i$  is part of the data ( $s_i$  could be changed into  $s_i t$ , with  $t$  in the centralizer of  $g$ ).

We fix a base point  $x_0 \in \mathcal{H}$ . Let us associate three sequences of points  $\{a_i\}_{i=1}^k$ ,  $\{b_i\}_{i=1}^k$  and  $\{x_i\}_{i=1}^k$  to the given base point  $x_0$  and the factorization of  $h$  into the product of the  $h_i$ . Namely, for all  $1 \leq i \leq k$ , we first set:

- $x_i$  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 line  $\text{Ax}(h_i)$ ;
- $b_i$  is equal to  $h_i(a_i)$ ; in particular, both  $a_i$  and  $b_i$  are on  $\text{Ax}(h_i)$ .

---

<sup>(1)</sup> Since  $d(x, y) \geq 20\theta$ , this assumption is meaningful: It does not depend on the choices of the projections of  $x$  and  $y$  on  $\text{Ax}(f)$ .

However, we do not want the distances  $d(x_{i-1}, a_i) = d(b_i, x_i)$  to be too small. Thus, if  $d(x_{i-1}, a_i) \leq 10\theta$ , we translate  $a_i$  towards  $b_i$  at a distance of  $12\theta$ , and similarly we translate  $b_i$  towards  $a_i$ . This defines the three sequences  $\{a_i\}_{i=1}^k$ ,  $\{b_i\}_{i=1}^k$  and  $\{x_i\}_{i=1}^k$ .

LEMMA 2.12. *If the translation length  $L$  of  $g$  is larger than  $480\theta$ , then the following properties hold:*

- (1) *each  $[a_i, b_i]$  is a subsegment of  $\text{Ax}(h_i)$  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  $10\theta$  (for  $1 \leq i \leq k$ ).*

*Proof.* Since we performed some translations when  $[x_{i-1}, a_i]$  was too small, we have property (3). The segment  $[a_i, b_i]$  was initially of length  $L$ , so (1) remains true as long as  $\frac{1}{20}L \geq 24\theta$ , which is our assumption. Property (2) follows from Lemma 2.3.  $\square$

We say that  $h_k \dots h_1$  is an *admissible presentation* of  $h$  (with respect to the base point  $x_0$ ) if the three sequences  $\{a_i\}_{i=1}^k$ ,  $\{b_i\}_{i=1}^k$  and  $\{x_i\}_{i=1}^k$  have the additional property that all pairs  $(x_{i-1}, a_i)$ ,  $1 \leq i \leq k$ , are neutral. It follows that all pairs  $(b_j, x_j)$  are also neutral, because the property of being neutral is stable under translation by isometries.

LEMMA 2.13. *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 the translation length of  $g$  satisfies  $28\theta < \frac{1}{20}L(g)$ . Then  $h$  admits at least one admissible presentation.*

*Remark 2.14.* Due to this lemma, the following property can be added to the first three properties listed in Lemma 2.12:

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

Lemma 2.13 corresponds to [17, Lemma 2.2]. We give a complete proof because it is more involved in our case.

*Proof.* As before, denote by  $L$  the translation length  $L(g)$ . Start with a decomposition  $h = h_k \dots h_1$  with  $h_i = s_i g^{\pm 1} s_i^{-1}$ , and construct the sequences of points  $\{a_i\}_{i=1}^k$ ,  $\{b_i\}_{i=1}^k$  and  $\{x_i\}_{i=1}^k$  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 that  $\mathcal{I}$  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  $f = s g s^{-1}$  of  $g$  such that  $[y, z]$  is  $7\theta$ -close to  $\text{Ax}(f) = s(\text{Ax}(g))$  and  $d(y, z) \geq \frac{11}{20}L$ ; we fix such a segment  $[y, z]$  with maximal length. Let  $y'$  and  $z'$  be the projections of  $y$  and  $z$  on  $\text{Ax}(f)$ ; we have  $d(y, y') \leq 7\theta$  and  $d(z, z') \leq 7\theta$ . If  $y'$  is contained in the segment  $[f(y'), z'] \subset \text{Ax}(f)$ , we change  $f$  into  $f^{-1}$ . Replace  $h_i$  by the product of three conjugates of  $g$  or  $g^{-1}$

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

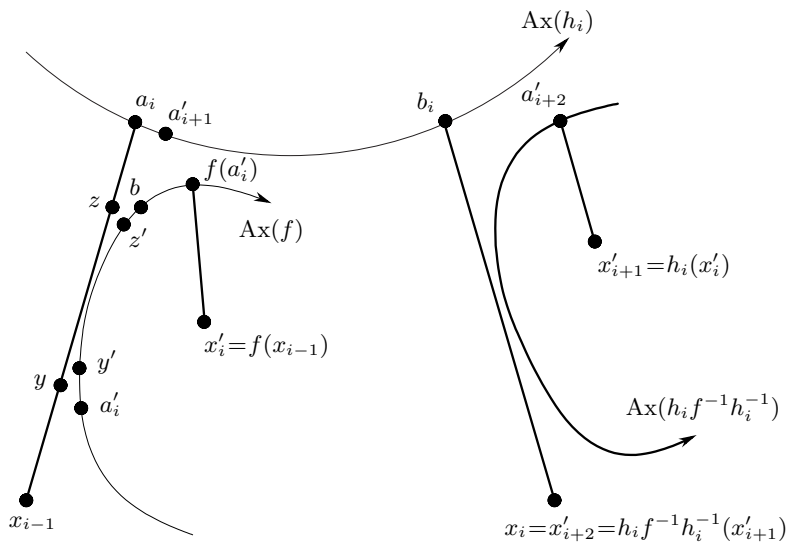


Figure 1.

This gives rise to a new decomposition of  $h$  as a product of conjugates of  $g^{\pm 1}$ , and hence to new sequences of points. Let  $a'_i$  be a projection of  $x_{i-1}$  on  $\text{Ax}(f)$ . Concerning the sequence  $\{x_i\}_{i=1}^k$ , 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 may assume that there was no need to move  $a'_i$ ,  $a'_{i+1}$  or  $a'_{i+2}$  as in Lemma 2.12.

*Two estimates.*

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

To prove it, we write

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

and

$$d(x_{i-1}, a'_i) \leq d(x_{i-1}, y') \leq d(x_{i-1}, y) + d(y, y') \leq d(x_{i-1}, y) + 7\theta; \quad (2.3)$$

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

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



Let us postpone the proof of this second estimate, and deduce the lemma from these two inequalities.

*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  $d(x_{j-1}, a_j)$  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.

*Proof of claim 2.* Denote by  $b$  the projection of  $a_i$  on  $\text{Ax}(f)$ . Similarly as (2.3), we have

$$d(b, a_i) \leq d(z, a_i) + 7\theta. \quad (2.4)$$

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

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

Applying Lemma 2.3 to the triangles  $x_{i-1}, y', a'_i$  and  $a_i, z', b$ , and the inequalities

$$d(y, y'), d(z, z') \leq 7\theta,$$

we get

$$d(x_{i-1}, y) = d(x_{i-1}, a'_i) + d(y, a'_i),$$

up to an error of  $11\theta$ , and similarly

$$d(a_i, z) = d(a_i, b) + d(z, b),$$

up to  $11\theta$ . Hence, we get

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

This concludes the proof of claim 2 in the first case, because  $-L + 22\theta < -\frac{1}{20}L$ .

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

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

and thus

$$d(x'_i, a_i) \leq d(x_{i-1}, a_i) - d(a'_i, b) + \frac{9}{20}L + 14\theta. \quad (2.5)$$

On the other hand, the triangular inequality implies that

$$d(x_{i-1}, a'_i) + d(a'_i, b) + d(b, a_i) \geq d(x_{i-1}, a_i) = d(x_{i-1}, y) + d(y, z) + d(z, a_i),$$

and so

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

Using inequalities (2.3) and (2.4), we obtain

$$d(a'_i, b) \geq \frac{11}{20}L - 14\theta.$$

Finally, inequality (2.5) gives

$$d(x'_i, a'_{i+1}) \leq d(x'_i, a_i) \leq d(x_{i-1}, a_i) - \frac{2}{20}L + 28\theta,$$

and hence claim 2 because  $28\theta < \frac{1}{20}L$ .  $\square$

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

LEMMA 2.15. *Let  $h = h_k \dots h_1$  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

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

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

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

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

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

and one checks that all relevant segments are neutral because they are obtained from neutral segments of the previous presentation by isometric translations (either by Id or by  $h_i$ ).  $\square$

## 2.5. Proof of the normal subgroup theorem

Let  $g$  be an element of  $G$  which satisfies the small cancellation property. By definition,  $g$  is a tight element of  $G$ , its axis is  $(14\theta, B)$ -rigid for some  $B > 0$ , and

$$\frac{1}{20}L(g) \geq 60\theta + 2B.$$

Denote by  $L$  the translation length  $L(g)$ . Let  $h$  be a non-trivial element of the normal subgroup  $\langle\langle g \rangle\rangle$ . Our goal is to prove  $L(h) \geq L$ , with equality if and only if  $h$  is conjugate to  $g$ . Pick a base point  $x_0$  such that  $d(x_0, h(x_0)) \leq L(h) + \theta$ . Lemma 2.13 applied to  $g$  implies the existence of an admissible presentation with respect to the base point  $x_0$ :

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

We assume that  $m$  is minimal among all such choices of base points and admissible presentations. Lemma 2.15 implies that  $h_j$  is different from  $h_i^{-1}$  for all  $1 \leq i < j \leq m$ .

Let  $\{a_i\}_{i=1}^k$ ,  $\{b_i\}_{i=1}^k$  and  $\{x_i\}_{i=1}^k$  be the three sequences of points defined in §2.4; they satisfy properties (1)–(4) listed in Lemma 2.12 and Remark 2.14. Since the constructions below are more natural with segments than pairs 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  $[a_i, b_i] \subset \text{Ax}(h_i)$ .

We now introduce the following definition in order to state the key Lemma 2.16. A sequence of points  $(c_{-1}, c_0, \dots, c_k, c_{k+1})$  in  $\mathcal{H}$ , with some choice 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  $0 \leq i \leq k$  and all  $a \in [c_i, c_{i+1}]$ , we have  $(c_{i-1} | a)_{c_i} \leq 3\theta$ , in particular,

$$(c_{i-1} | c_{i+1})_{c_i} \leq 3\theta;$$

- (iii) for all  $0 \leq i \leq k+1$  we have  $d(c_{i-1}, c_i) \geq 10\theta$ ;
- (iv) for all  $0 \leq i \leq k$  the segment  $[c_i, c_{i+1}]$  is either neutral or a piece, 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 piece of size  $\frac{18}{20}$  if  $[c_1, c_2]$  is neutral (this is always the case when  $k=1$ ), and of size  $\frac{17}{20}$  otherwise;
- (iv-d) for the other pieces  $[c_{i-1}, c_i]$ , with  $i > 1$ , the size is  $\frac{5}{20}$  when  $[c_i, c_{i+1}]$  is neutral and  $\frac{4}{20}$  otherwise;
- (v) for all  $0 \leq i \leq k$ , if  $[c_i, c_{i+1}]$  is a piece, then there is an index  $l$  with  $1 \leq l \leq j$  such that  $h_l$  is the support of the piece  $[c_i, c_{i+1}]$ .

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

The following lemma is inspired by [17, Lemma 2.4]. As mentioned there, this can be seen as a version of the famous Greendlinger lemma in classical small cancellation theory. Recall that this lemma claims the existence of a region with a large external segment in a van Kampen diagram (see [45, Chapter V]). The segment  $[c_0, c_1]$  in the previous definition plays a similar role. One can consult [43] for a simpler proof of Lemma 2.16 and Theorem 2.10 in the case of a group acting on a tree, and for small cancellation theory in the context of  $\text{Aut}[\mathbf{k}^2]$  and its amalgamated product structure.

**LEMMA 2.16.** *For each  $j=1, \dots, m$ , there exists  $k \geq 1$  such that the segment  $[x_0, x_j]$  admits a configuration of order  $k$ . Moreover, if  $j \geq 2$  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 Lemma 2.12 and Remark 2.14; we refer to them as properties (1)–(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.

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

$$(x_0 | b_1)_{a_1} \leq 2\theta \quad \text{and} \quad (a_1 | x_1)_{b_1} \leq 2\theta.$$

By property (1),  $[c_0, c_1]$  is a piece of size  $\frac{19}{20}$  (it is a subsegment of  $Ax(h_1)$ ). The segments  $[x_0, c_0]$  and  $[x_1, c_1]$  are neutral (property (4)) and of length at least  $10\theta$  (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  $(\Phi, T)$  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  $d(a, a_{j+1}) \leq 5\theta$  and  $d(b, b_{j+1}) \leq 5\theta$ . Thus  $d(a, b) \geq \frac{19}{20}L - 10\theta$  and, by Lemma 2.5,  $[a, b]$  is a piece with support  $h_{j+1}$ .

For all  $i \leq k+1$ , note that  $c'_i$  a projection of  $c_i$  on the segment  $[x_0, x_j]$ . Note that, by Lemma 2.6, we have  $d(c_i, c'_i) \leq 5\theta$ , 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  $\Phi$  are two intervals  $I \subset [a, b]$  and  $I' \subset [c'_i, c'_{i+1}]$  such that

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

*Fact 2.17.* If  $[c_i, c_{i+1}]$  is neutral then  $\text{diam}(S_i) \leq \frac{12}{20}L$ .

Since  $\Phi$  is an isometry along the geodesic segments  $[x_j, x_0]$  and  $[x_j, x_{j+1}]$ , we only have to prove that  $\text{diam}(I') \leq \frac{12}{20}L$ . By Lemma 2.5, we know that

$$[c_i, c_{i+1}] \subset \text{Tub}_{7\theta}([c'_i, c'_{i+1}]).$$

Due to the triangular inequality, we can choose  $J \subset [c_i, c_{i+1}]$  such that  $J \subset \text{Tub}_{7\theta}(I')$  and  $\text{diam}(J) \geq \text{diam}(I') - 14\theta$ . The properties of the approximation tree imply that

$$I' \subset \text{Tub}_\theta(I) \quad \text{and} \quad I \subset \text{Tub}_{7\theta}(\text{Ax}(h_{j+1})).$$

Thus  $J \subset \text{Tub}_{15\theta}(\text{Ax}(h_{j+1}))$ . Applying Lemma 2.4 (with  $\beta = 16\theta$ ) and shortening  $J$  by  $32\theta$  ( $16\theta$  at each end), we obtain  $\text{diam}(J) \geq \text{diam}(I') - 46\theta$  and  $J \subset \text{Tub}_{2\theta}(\text{Ax}(h_{j+1}))$ . Thus  $J$  is a piece contained in the neutral segment  $[c_i, c_{i+1}]$ , and hence

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

*Fact 2.18.* If  $[c_i, c_{i+1}]$  is a piece then  $\text{diam}(S_i) \leq B$ .

By property (v), there exists an index  $l$ , with  $1 \leq l \leq j$ , such that

$$[c_i, c_{i+1}] \subset \text{Tub}_{7\theta}(\text{Ax}(h_l)).$$

Since  $I' \subset \text{Tub}_{7\theta}([c_i, c_{i+1}])$ , we get

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

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

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

If  $\text{diam}(I') > B$ , 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 2.15, this contradicts the minimality of the presentation of  $h$ . This proves that  $\text{diam}(I') \leq B$ .

*Fact 2.19.* Suppose that  $[c_i, c_{i+1}]$  is a piece. Then

$$\text{diam}([\Phi(x_j), \Phi(a)] \cap [\Phi(c'_i), \Phi(c'_{i+1})]) \leq \frac{12}{20}L.$$

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 that  $\text{diam}(K) \leq \frac{12}{20}L$ . Applying the triangular inequality, we can choose  $J \subset [a_{j+1}, x_j]$  such that  $J \subset \text{Tub}_{7\theta}(K)$  and  $\text{diam}(J) \geq \text{diam}(K) - 14\theta$ . Now we have

$$K \subset \text{Tub}_\theta(K'), \quad K' \subset \text{Tub}_{7\theta}([c_i, c_{i+1}]) \quad \text{and} \quad [c_i, c_{i+1}] \subset \text{Tub}_{7\theta}(\text{Ax}(h_l)).$$

Thus  $J \subset \text{Tub}_{22\theta}(\text{Ax}(h_l))$ . Applying Lemma 2.4 with  $\beta = 23\theta$ , we shorten  $J$  by  $46\theta$  ( $23\theta$  at each end) and obtain

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

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

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

*Fact 2.20.* 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 that  $[\Phi(b), \Phi(a)] \subset [\Phi(c'_0), \Phi(x_j)]$ . By (iv-a) and Fact 2.17,  $[\Phi(b), \Phi(a)]$  intersects at least one segment  $[\Phi(c'_i), \Phi(c'_{i+1})]$  for which  $[c_i, c_{i+1}]$  is a piece; Fact 2.18 implies that  $[\Phi(c'_i), \Phi(c'_{i+1})]$  is not contained in  $[\Phi(a), \Phi(b)]$  (it must intersect the boundary points of  $[\Phi(a), \Phi(b)]$ ). It follows from this that  $[\Phi(a), \Phi(b)]$  intersects at most two pieces and one neutral segment. Facts 2.17 and 2.18 now give the contradictory inequality

$$2B + \frac{12}{20}L \geq \frac{19}{20}L - 10\theta.$$

*Fact 2.21.* The segment  $[\Phi(c'_0), \Phi(c'_1)]$  is not contained in  $[\Phi(x_{j+1}), \Phi(x_j)]$ .

Since  $[a_{j+1}, x_j]$  is neutral and  $[c_0, c_1]$  is a piece of size  $\geq \frac{17}{20}$ , the segment  $[\Phi(c'_0), \Phi(c'_1)]$  is not contained in  $[\Phi(a_{j+1}), \Phi(x_j)]$ . Assume that  $[\Phi(c'_0), \Phi(c'_1)] \subset [\Phi(x_{j+1}), \Phi(x_j)]$ , apply Fact 2.20 and then Facts 2.18 and 2.19; 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 2.20 and 2.21 give the following.

*Fact 2.22.* The segment  $[\Phi(b), \Phi(p)]$  has length at least  $\frac{6}{20}L$ , and  $[\Phi(c'_0), \Phi(p)]$  has length at least  $\frac{4}{20}L$ .

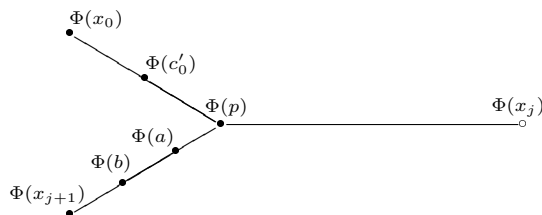


Figure 2. The point  $\Phi(a)$  could also be on  $[\Phi(p), \Phi(x_j)]$ , but not far from  $\Phi(p)$ .

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  $\Phi(a)$  relative to the branch point  $\Phi(p)$ .

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

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 2.22, 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  $5\theta$ , the triangular inequality and Corollary 2.2 imply that

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

and the assumption made in this first case gives

$$(b_{j+1} | x_0)_{a_{j+1}} \leq \frac{1}{20}L - \theta. \tag{2.6}$$

At this point we would like to define a configuration of order 1 for  $[x_0, x_{j+1}]$  by using  $a$  and  $b$ . The problem is that since the property to be a neutral segment is not stable under perturbation, there is no guarantee that  $[b, x_{j+1}]$  is neutral even if  $[b_{j+1}, x_{j+1}]$  is. Other candidates would be  $a_{j+1}$  and  $b_{j+1}$ , but then we would not have estimate (ii) in the definition of a configuration of order  $k$ .

To solve this dilemma we consider another approximation tree  $\Psi: X \rightarrow T'$ , 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).

We then define a new configuration of points  $\hat{c}_{-1} = x_0$ ,  $\hat{c}_0 = q$ ,  $\hat{c}_1 = b_{j+1}$  and  $\hat{c}_2 = x_{j+1}$ , and we show that this defines a configuration of order 1 for  $[x_0, x_{j+1}]$ . Using inequality (2.6), the fact that  $\Psi$  is an isometry along  $[a_{j+1}, b_{j+1}]$ , and Corollary 2.2, we obtain

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

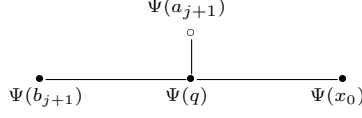


Figure 3.

Thus  $[q, b_{j+1}]$  is a piece of size  $\frac{18}{20}$  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 Lemma 2.12). Thus,  $\{\hat{c}_j\}_{j=-1}^2$  is a configuration of order 1 for the segment  $[x_0, x_{j+1}]$ .

Moreover, we have

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

Fact 2.22 then gives

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

so we obtain the second assertion in Lemma 2.16.

*Case 2.*  $(\Phi(b) | \Phi(x_0))_{\Phi(a)} > \frac{1}{20}L - 12\theta$ .

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 that  $\Phi(a) \in [\Phi(p), \Phi(x_j)]$  and

$$d(\Phi(p), \Phi(a)) > \frac{1}{20}L - 12\theta.$$

By Fact 2.22 again, the situation is similar to Figure 4. We distinguish two subcases according to the nature of  $[c_i, c_{i+1}]$ .

*Subcase 2 (i).*  $[c_i, c_{i+1}]$  is neutral.

Then if  $i < k$  the segment  $[c_{i+1}, c_{i+2}]$  is a piece (if  $i = k$  the following discussion is even easier). If

$$d(\Phi(p), \Phi(c'_{i+1})) \leq 34\theta,$$

then Fact 2.18 implies that  $d(\Phi(p), \Phi(a)) \leq B + 34\theta \leq \frac{1}{20}L - 12\theta$ . This contradicts the assumption of the second case; as a consequence,

$$\min\{d(\Phi(p), \Phi(c'_{i+1})), d(\Phi(p), \Phi(a))\} > 34\theta. \quad (2.7)$$



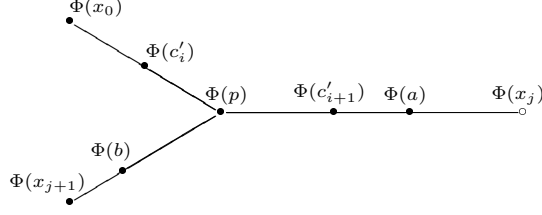


Figure 4. In this picture,  $\Phi(c'_{i+1})$  is between  $\Phi(p)$  and  $\Phi(a)$ ; another possible case would be  $\Phi(a) \in [\Phi(p), \Phi(c'_{i+1})]$ .

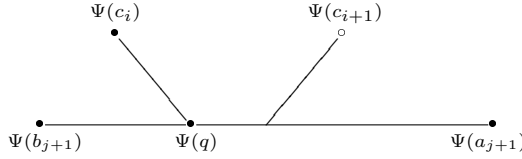


Figure 5.

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

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

and

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

By (2.7) 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.

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

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

Thus, by Fact 2.22, we see that  $[q, b_{j+1}]$  is a piece of size  $\frac{5}{20}$ , and the same conclusion holds even if one moves  $q$  along the segment  $[q, b_{j+1}]$  by at most  $12\theta$ . 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  $\{\hat{c}_l\}_{l=-1}^{i+3}$  by  $\hat{c}_l = c_l$  for  $-1 \leq l \leq i$ ,  $\hat{c}_{i+1} = q$ ,  $\hat{c}_{i+2} = b_{j+1}$  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  $[q, b_{j+1}]$  at a distance less than  $12\theta$ . This does not change the fact that  $[q, b_{j+1}]$  is a piece of size  $\frac{5}{20}$ . Since the new configuration is obtained from the previous one by cutting it after  $c_i$  and then adding a piece  $[q, b_{j+1}]$  and a neutral segment  $[b_{j+1}, x_{j+1}]$ , property (iv) is also satisfied.

*Subcase (2) (ii).*  $[c_i, c_{i+1}]$  is a piece.

By Fact 2.18 the segment  $[\Phi(p), \Phi(c_{i+1})]$  has length at most  $B$ , and  $[c_{i+1}, c_{i+2}]$  must be neutral: otherwise  $d(\Phi(p), \Phi(a)) \leq 2B$ , contradicting the assumption

$$d(\Phi(p), \Phi(a)) > \frac{1}{20}L - 12\theta.$$

Thus the piece  $[c_i, c_{i+1}]$  has size  $\frac{5}{20}$  (or  $\frac{18}{20}$  if  $i=0$ ). We consider the approximation tree  $\Psi: X \rightarrow T$  of  $(c_{i+1}, c_i, c_{i+2}, a_{j+1}, b_{j+1})$ . We obtain one of the situations depicted on Figure 6, where now  $q \in [c_{i+1}, c_i]$  or  $q \in [c_{i+1}, c_{i+2}]$ .

In case (a),  $[c_{i+1}, q]$  is small, and therefore neutral. We are in a case similar to subcase 2 (i) and to Figure 5. We still have

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

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

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

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

As in subcase 2 (i), the estimate for  $d(b_{j+1}, q)$  gives

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

and Fact 2.22 implies that  $[q, b_{j+1}]$  is a piece of size  $\frac{5}{20}$ ; moreover  $(x_{j+1}|p_1)_{b_{j+1}} \leq 3\theta$ .

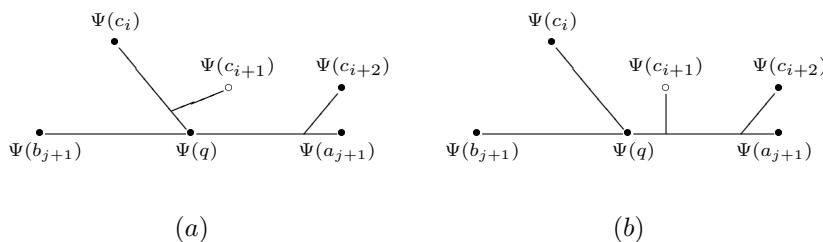


Figure 6.

We define the new configuration  $\{\hat{c}_l\}_{l=-1}^{i+3}$  by taking  $\hat{c}_l = c_l$  for  $-1 \leq l \leq i$ ,  $\hat{c}_{i+1} = q$ ,  $\hat{c}_{i+2} = b_{j+1}$  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)–(v) are satisfied. □

*Proof of Theorem 2.10.* By Lemma 2.16, there exists a configuration  $\{c_i\}_{i=-1}^k$  of order  $k$  for  $[x_0, x_m]$ , where  $x_m = h(x_0)$ . Recall that by our choice of  $x_0$  we have

$$L(h) \geq d(x_0, x_m) - \theta.$$

If  $k \geq 2$ , then we have at least two distinct pieces: The first one,  $[c_0, c_1]$ , has size at least  $\frac{17}{20}$ , and the last one,  $[c_{k-1}, c_k]$ , has size  $\frac{5}{20}$ . 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  $g$ , or we have  $d(x_0, c_0) \geq \frac{3}{20}L$ . On the other hand,  $d(c_0, c_1) \geq \frac{18}{20}L$ , so we obtain

$$L(h) \geq \frac{21}{20}L - 11\theta > L. \quad \square$$

### 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  $\delta$ -hyperbolicity, is a useful property to decide whether the axis  $Ax(g)$  of a hyperbolic 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  $\|x\|_H$  denote the norm of the 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  $\langle \cdot, \cdot \rangle: H \times H \rightarrow \mathbf{R}$  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  $\langle \cdot, \cdot \rangle$ . 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  $\text{dist}: \mathbb{H} \times \mathbb{H} \rightarrow \mathbf{R}_+$  defined by

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

gives a distance on  $\mathbb{H}$ , and  $(\mathbb{H}, \text{dist})$  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  $\text{CAT}(-1)$  space, and therefore a  $\delta$ -hyperbolic space. More precisely,  $\mathbb{H}$  is  $\delta$ -hyperbolic, in any dimension, even infinite, with  $\delta = \log 3$  (see [13, §I.4, p. 11]). In particular, all properties listed in §2.1 are satisfied in  $\mathbb{H}$ .

#### 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  $\langle \cdot, \cdot \rangle$ ; we shall denote it by  $\partial\mathbb{H}$  and call it the *boundary* of  $\mathbb{H}$  (or boundary at infinity).

Let  $\Gamma$  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  $V_\Gamma$  intersects the light cone in two lines, and these lines determine two points of  $\partial\mathbb{H}$ , called the endpoints of  $\Gamma$ . 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  $\text{Vect}(x, y) \cap \mathbb{H}$ .

Let  $x$  be a point of  $\mathbb{H}$  and  $\Gamma$  be a geodesic. The projection  $\pi_\Gamma(x) \in \Gamma$ , i.e. the point  $y \in \Gamma$  which minimizes  $\text{dist}(x, y)$ , 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  $\delta$ -hyperbolic spaces because  $\delta$ -hyperbolicity does not tell anything at scale  $< \delta$ .

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

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

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

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

Thus, when  $B$  is large,  $\nu$  becomes much smaller than  $\varepsilon$ . The constants in this inequality are not optimal and depend on the choice of the ratio

$$\frac{\text{dist}(u, v)}{\text{dist}(x, y)} = \frac{3}{4}.$$

A proof of this lemma can be found in the preprint version of this paper [8].

### 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  $f: H \rightarrow H$  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  $f: H \rightarrow H$ .

(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  $x \in \mathbb{H}$  tends to  $\alpha(f)$  as  $k \rightarrow -\infty$  and tends to  $\omega(f)$  as  $k \rightarrow \infty$ . The set  $\text{Min}(f)$  coincides with the geodesic line from  $\alpha(f)$  to  $\omega(f)$ . In particular,  $\text{Min}(f)$  coincides with the unique  $f$ -invariant axis

$$\text{Ax}(f) = \text{Vect}(\alpha(f), \omega(f)) \cap \mathbb{H}.$$

The points  $\alpha(f)$  and  $\omega(f)$  correspond to eigenlines of the linear extension  $f: H \rightarrow H$  with eigenvalues  $\lambda^{-1}$  and  $\lambda$ , where  $\lambda > 1$ ; the translation length  $L(f)$  is equal to  $\log \lambda$  (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  $(\varepsilon', B')$ -rigid, then  $A$  is also  $(\varepsilon, B)$ -rigid with*

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

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

*Proof.* By assumption,  $B$  satisfies the following three properties:

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

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  $2\varepsilon$ -close. Lemma 3.1 and the inequalities satisfied by  $B - 2\varepsilon$  show that  $I$  and  $J$  contain subsegments of length  $\frac{3}{4}(B - 2\varepsilon)$  which are  $\varepsilon'$ -close. Since  $A$  is  $(\varepsilon', B')$ -rigid,  $f(A)$  coincides with  $A$ , and therefore  $A$  is  $(\varepsilon, B)$ -rigid.  $\square$

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,  $p \in \text{Ax}(g)$  and  $\eta > 0$ . If  $\text{Ax}(g)$  is not rigid, there exists an element  $f$  of  $G$  such that*

$$f(\text{Ax}(g)) \neq \text{Ax}(g) \quad \text{and} \quad \text{dist}(f(x), x) \leq \eta \quad \text{for all } x \in [g^{-n}(p), g^n(p)].$$

*Proof.* Since  $\text{Ax}(g)$  is not rigid, in particular  $\text{Ax}(g)$  is not  $(\varepsilon, B)$ -rigid for

$$B = (3n + 4)L(g) \quad \text{and} \quad \varepsilon = \frac{1}{2}\eta.$$

Then there exists an isometry  $h \in G$  such that  $h(\text{Ax}(g))$  is different from  $\text{Ax}(g)$ , but  $\text{Ax}(g)$  contains a segment  $J$  of length  $B$  which is mapped by  $h$  into the tubular neighborhood  $\text{Tub}_\varepsilon(\text{Ax}(g))$ .

Changing  $h$  into  $g^j \circ h$ , we may 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 may 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  $\text{Ax}(g)$  that satisfy the following properties:

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

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

We now change  $h_1$  into the commutator  $h_2 = h_1^{-1}g^{-1}h_1g$ . We still have

$$h_2(\text{Ax}(g)) \neq \text{Ax}(g),$$

because otherwise  $h_1^{-1}g^{-1}h_1$  fixes  $\text{Ax}(g)$  and by uniqueness of the axis of a hyperbolic isometry of  $\mathbb{H}$  we would have  $h_1^{-1}(\text{Ax}(g)) = \text{Ax}(g)$ . Moreover, property (2) above is replaced by

$$\text{dist}(x, h_2(x)) \leq 2\varepsilon < \eta \quad \text{and} \quad \text{dist}(y, h_2(y)) \leq \eta$$

in case (a), and by

$$\text{dist}(g^2(x), h_2(x)) \leq \eta \quad \text{and} \quad \text{dist}(g^2(y), h_2(y)) \leq \eta$$

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 may assume that

$$\text{dist}(x, h_3(x)) \leq \eta \quad \text{and} \quad \text{dist}(y, h_3(y)) \leq \eta.$$

Consider the arc length parametrization

$$m: t \in [-\infty, \infty] \mapsto \text{Ax}(g)$$

such that  $m(0) = p$  and  $g(m(t)) = m(t + L(g))$ . Since  $p$  is in the interval  $[x + \frac{1}{3}B_1, y - \frac{1}{3}B_1]$ , we obtain  $\text{dist}(m(t), h_3(m(t))) \leq \eta$  for all  $t \in [-\frac{1}{3}B_1, \frac{1}{3}B_1]$ . Defining  $f = h_3$ , we get

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

because  $nL(g) \leq \frac{1}{3}B_1$ . □

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  $\text{Ax}(g)$ . Let  $\eta$  be a positive real number. If there is no  $f$  in  $G \setminus \{\text{Id}\}$  such that  $\text{d}(f(x), x) \leq \eta$  for all  $x \in [g^{-1}(p), g(p)]$ , then  $\text{Ax}(g)$  is  $(\frac{1}{2}\eta, 7L(g))$ -rigid.*

## Part II. 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  $\mathbf{N}^1(X)$  is the group of Cartier divisors modulo numerical equivalence. When the field of definition is the field of complex numbers,  $\mathbf{N}^1(X)$  coincides with the space of Chern classes of holomorphic line bundles of  $X$  (see [44]), and thus

$$\mathbf{N}^1(X) = H^2(X(\mathbf{C}), \mathbf{Z})_{\text{tf}} \cap H^{1,1}(X, \mathbf{R}),$$

where  $H^2(X(\mathbf{C}), \mathbf{Z})_{\text{tf}}$  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  $\mathbf{N}^1(X)$ . The intersection form defines an integral quadratic form

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

on  $\mathbf{N}^1(X)$ , the signature of which is equal to  $(1, \rho(X) - 1)$  by the Hodge index theorem (see [37, §V.1]). We also denote  $\mathbf{N}^1(X) \otimes \mathbf{R}$  by  $\mathbf{N}^1(X)_{\mathbf{R}}$ .

If  $\pi: X \rightarrow Y$  is a birational morphism, then the pull-back morphism

$$\pi^*: \mathbf{N}^1(Y) \longrightarrow \mathbf{N}^1(X)$$

is injective and preserves the intersection form. For example, if  $\pi$  is just the blow-up of a point with exceptional divisor  $E \subset X$ , then  $\mathbf{N}^1(X)$  is isomorphic to  $\pi^*(\mathbf{N}^1(Y)) \oplus \mathbf{Z}[E]$ , 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 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  $[H], [E_1], \dots, [E_n]$  such that  $[H]^2 = 1$  and  $[E_i]^2 = -1$  for all  $1 \leq i \leq n$ .

#### 4.2. Indeterminacies

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

Let  $H$  be an ample line bundle on  $Y$ . Consider the pull-back 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  $\text{Ind}(f)$ , but it can of course include infinitely near points (that is, points on surfaces that dominate  $\mathbb{P}^2$ ). We call it the *base locus* of  $f$ . To resolve the indeterminacies of  $f$ , one blows up the base locus (see [44] for base loci, base ideal and their blow-ups); in other words, one blows up  $\text{Ind}(f)$ , obtaining  $\pi: X' \rightarrow X$ , then one blows up  $\text{Ind}(f \circ \pi)$ , and so on; the process stops in a finite number of steps (see [37, §V.5]).

*Remark 4.2.* If  $f$  is a birational transformation of a projective surface with Picard number 1, then  $\text{Ind}(f)$  is contained in  $\text{Exc}(f)$ . 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  $\text{Exc}(f)$  corresponds to a  $(-1)$ -curve at some point in the sequence of blow-downs, but is also the strict transform of a curve of positive self-intersection from the source (this is where we use  $\rho(X)=1$ ).

### 4.3. Dynamical degrees

The rational map  $f: X \dashrightarrow Y$  determines a linear map  $f^*: \mathbf{N}^1(Y) \rightarrow \mathbf{N}^1(X)$ . For complex surfaces,  $f$  determines a linear map  $f^*: H^2(Y, \mathbf{Z}) \rightarrow H^2(X, \mathbf{Z})$  which preserves the Hodge decomposition: The action of  $f^*$  on  $\mathbf{N}^1(X)$  coincides with the action by pull-back on  $H^2(X(\mathbf{C}), \mathbf{Z})_{\text{tf}} \cap H^{1,1}(X, \mathbf{R})$  (see [21] for example).

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

$$\lambda(f) = \lim_{n \rightarrow \infty} \|(f^n)^*\|^{1/n},$$

where  $\|\cdot\|$  is any operator norm on  $\text{End}(\mathbf{N}^1(X)_{\mathbf{R}})$ ; the limit exists because the sequence  $\|(f^n)^*\|$  is submultiplicative (see [21]). The number  $\lambda(f)$  is invariant under conjugacy:  $\lambda(f) = \lambda(gfg^{-1})$  if  $g: X \dashrightarrow Y$  is a birational map.

*Example 4.3.* Let  $[x:y:z]$  be homogeneous coordinates for the projective plane  $\mathbb{P}_{\mathbf{k}}^2$ . Let  $f$  be an element of  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ . Then there are three homogeneous polynomials

$$P, Q, R \in \mathbf{k}[x, y, z]$$

of the same degree  $d$ , and without common factor of degree  $\geq 1$ , such that

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

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

#### 4.4. Picard–Manin classes

We follow the presentations given in [4], [6] and [29] which, in turn, are inspired by Chapter 5 of Manin’s book [46].

##### 4.4.1. Models and Picard–Manin space

A *model* of  $X$  is a smooth projective surface  $X'$  with a birational morphism  $X' \rightarrow X$ . If  $\pi_1: X_1 \rightarrow X$  and  $\pi_2: X_2 \rightarrow X$  are two models, we say that  $X_2$  *dominates*  $X_1$  if the induced birational map  $\pi_1^{-1} \circ \pi_2: X_2 \dashrightarrow X_1$  is a morphism. In this case,  $\pi_1^{-1} \circ \pi_2$  contracts a finite number of exceptional divisors and induces an injective map

$$(\pi_1^{-1} \circ \pi_2)^*: \mathbf{N}^1(X_1)_{\mathbf{R}} \hookrightarrow \mathbf{N}^1(X_2)_{\mathbf{R}}.$$

Let  $\mathcal{B}_X$  be the set of models that dominate  $X$ . If  $X_1, X_2 \in \mathcal{B}_X$  then, 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* is the direct limit

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

The Néron–Severi group  $\mathbf{N}^1(X')_{\mathbf{R}}$  of any model  $X' \rightarrow X$  embeds in  $\mathcal{Z}(X)$  and can be identified with 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  $\mathbf{N}^1(X')$  (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  $\pi$ , the pull-back operator  $\pi^*$  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  $\pi: Y \rightarrow X$  together with marked points  $p \in Y$ , we introduce the following equivalence relation:  $(p, Y) \sim (p', Y')$  if the induced birational map  $(\pi')^{-1} \circ \pi: Y \dashrightarrow Y'$  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  $\pi: Y \rightarrow X$  with  $p \in Y$ .

Let  $(p, Y)$  be an element of  $\mathcal{V}_X$ . Consider the blow-up  $\bar{Y} \rightarrow Y$  of  $p$  and the exceptional divisor  $E_p \subset \bar{Y}$ ; 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  $(Y, \pi)$ , see [46]). The classes  $[E_p]$ ,  $p \in \mathcal{V}_X$ , have self-intersection  $-1$ , are mutually orthogonal, and are orthogonal to  $\mathbf{N}^1(X)_{\mathbf{R}}$ . Moreover,

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

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  $\{[H_i]\}_{i=1}^{g(X)}$  of  $\mathbf{N}^1(X)_{\mathbf{R}}$ , 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] and [6] for details); in other words

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

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  $\mathbb{P}_{\mathbf{k}}^2$ , the Néron–Severi group  $\mathbf{N}^1(\mathbb{P}_{\mathbf{k}}^2)$  is isomorphic to  $\mathbf{Z}[H]$ , where  $H$  is a line; elements of  $\bar{\mathcal{Z}}(X)$  are then given by sums

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

with  $\sum_{p \in \mathcal{V}_{\mathbb{P}_{\mathbf{k}}^2}} a_p^2 < \infty$ . We shall call this space the *Picard–Manin space* without further reference to  $\mathbb{P}_{\mathbf{k}}^2$  or to the completion.

### 4.5. Action of $\text{Bir}(X)$ on $\bar{\mathcal{Z}}(X)$

If  $\pi: X' \rightarrow X$  is a morphism, then  $\pi$  induces an isomorphism  $\pi^*: \mathcal{Z}(X) \rightarrow \mathcal{Z}(X')$ . Let us describe this fact when  $\pi$  is the (inverse of the) blow-up of a point  $q \in X$ . In this case we have

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

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  $\mathbf{N}^1(X') \subset \mathcal{Z}(X')$ ; the isomorphism  $\pi^*$  corresponds to this bijection of bases. Note that  $\pi^*$  extends uniquely as a continuous isomorphism  $\pi^*: \bar{\mathcal{Z}}(X') \rightarrow \bar{\mathcal{Z}}(X)$  that preserves the intersection form. Since any birational morphism  $\pi$  is a composition of blow-downs of exceptional curves of the first kind, this proves that  $\pi^*$  is an isometry from  $\mathcal{Z}(X)$  (resp.  $\bar{\mathcal{Z}}(X)$ ) to  $\mathcal{Z}(X')$  (resp.  $\bar{\mathcal{Z}}(X')$ ).

Now consider a birational map  $f: X \dashrightarrow X'$ . There exist a surface  $Y$  and two morphisms  $\pi: Y \rightarrow X$  and  $\sigma: Y \rightarrow X'$  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

$$f \longmapsto f_*$$

of  $\text{Bir}(X)$  into the orthogonal group of  $\mathcal{Z}(X)$  (resp.  $\bar{\mathcal{Z}}(X)$ ) 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.

In what follows, we restrict the study to the projective plane  $\mathbb{P}_{\mathbf{k}}^2$  and the Cremona group  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ .

## 4.6. Action on an infinite-dimensional hyperbolic space

### 4.6.1. The hyperbolic space $\mathbb{H}_{\bar{\mathcal{Z}}}$

We define

$$\mathbb{H}_{\bar{\mathcal{Z}}} = \{[D] \in \bar{\mathcal{Z}}(\mathbb{P}_{\mathbf{k}}^2) : [D]^2 = 1 \text{ and } [H] \cdot [D] > 0\},$$

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

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

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

### 4.6.2. Cremona isometries

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

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

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  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ . The isometry  $f_*$  of  $\mathbb{H}_{\bar{\mathbb{Z}}}$  is hyperbolic if and only if the dynamical degree  $\lambda(f)$  is strictly larger than 1.*

As a consequence, when  $\lambda(f) > 1$ , we have that  $f_*$  preserves a unique geodesic line  $\text{Ax}(f) \subset \mathbb{H}_{\bar{\mathbb{Z}}}$ ; this line is the intersection of  $\mathbb{H}_{\bar{\mathbb{Z}}}$  with a plane  $V_f \subset \bar{\mathbb{Z}}(\mathbb{P}_{\mathbf{k}}^2)$  which intersects the isotropic cone of  $\bar{\mathbb{Z}}(\mathbb{P}_{\mathbf{k}}^2)$  in 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  $\alpha(f)$ , respectively, in the notation of §3.1.4).

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

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

*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  $\varepsilon: \mathbb{P}_{\mathbf{C}}^2 \dashrightarrow X$  such that  $\varepsilon \circ f^k \circ \varepsilon^{-1}$  is an element of  $\text{Aut}(X)^0$  (the connected component of the identity in  $\text{Aut}(X)$ );  $f_*$  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  $f \in \text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  is conjugate, via a birational transformation  $\varphi$ , to an automorphism  $g$  of a smooth rational surface  $X$ :

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

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

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

where  $\mathbf{N}^1(X)_{\mathbf{R}}^{\perp}$  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

$$\varphi_* \mathbf{N}^1(X)_{\mathbf{R}} \subset \bar{\mathbb{Z}}(\mathbb{P}_{\mathbf{k}}^2).$$

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

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

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

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

(3) *if  $f_*$  is hyperbolic then, modulo  $\varphi_*$ -conjugacy, the plane  $V_f$  corresponds to  $V_g$ , which is contained in  $\mathbf{N}^1(X)_{\mathbf{R}}$ , and  $\text{Ax}(f_*)$  corresponds to  $\text{Ax}(g_*)$ .*

#### 4.6.4. Example: quadratic mappings (see [9])

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

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

with  $[H]$  being the class of a line in  $\mathbb{P}_{\mathbf{C}}^2$ . If  $f$  is an isomorphism on a neighborhood of  $p$ , and  $f(p) = q$ , then  $f_*([E_p]) = [E_q]$ . These formulas are correct even when there are 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 to  $q_1$ , and  $q_3$  is infinitely near to  $q_2$ , but the formula for the image of  $[H]$  is still the same.

A Zariski open subset of  $\text{Bir}_2(\mathbb{P}_{\mathbf{C}}^2)$  is made of birational transformations  $f = h_1 \circ \sigma \circ h_2$ , with  $h_i \in \text{Aut}(\mathbb{P}_{\mathbf{C}}^2)$ ,  $i = 1, 2$ , and  $\sigma$  being the standard quadratic involution

$$\sigma[x : y : z] = [yz : zx : xy].$$

For such maps,  $\{p_1, p_2, p_3\}$  is the image of  $\text{Ind}(\sigma) = \{[1:0:0]; [0:1:0]; [0:0:1]\}$  by  $h_2^{-1}$ , and  $\{q_1, q_2, q_3\} = h_1(\text{Ind}(\sigma))$ . Moreover, the exceptional set  $\text{Exc}(f^{-1})$  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 to  $p_1$  by  $f^{-1}$ , then

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

If  $f$  is a general quadratic map, then  $\lambda(f)$  is equal to 2 and  $f_*$  induces a hyperbolic isometry on  $\mathbb{H}_{\bar{\mathbb{Z}}}$ ; 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  $[D] \in \mathcal{Z}(\mathbb{P}_{\mathbb{C}}^2)$ .

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

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

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

where  $A \in \text{PGL}_2(\mathbb{C})$  and  $B_y \in \text{PGL}_2(\mathbb{C}(y))$ . Clearing denominators, we may assume that  $B_y$  is given by a function  $B: y \mapsto B(y)$  with

$$B(y) = \begin{pmatrix} a(y) & b(y) \\ c(y) & d(y) \end{pmatrix} \in \text{GL}_2(\mathbb{C}(y)),$$

where the coefficients  $a$ ,  $b$ ,  $c$  and  $d$  are polynomials of respective degrees  $d-1$ ,  $d$ ,  $d-2$  and  $d-1$ . 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  $x_i \in \mathbb{P}(\mathbf{C}^2)$ . 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  $p_1, \dots, p_{2d-2}$  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  $p_1, \dots, p_{2d-2}$ .

Any set of three distinct points  $\{p_0, p_1, p_2\}$  such that  $p_0 = [1:0:0]$ , and  $p_0$ ,  $p_1$  and  $p_2$  are not collinear, is the indeterminacy set of a de Jonquières transformation of degree 2. All sets of distinct points  $\{p_0, p_1, \dots, p_4\}$  such that  $p_0 = [1:0:0]$ , no three of them are on a line through  $p_0$ , and there is no line containing  $p_1$ ,  $p_2$ ,  $p_3$  and  $p_4$ , can be obtained as the indeterminacy set of a de Jonquières transformation of degree 3. More generally, the indeterminacy sets of de Jonquières transformations of degree  $d$  form a non-empty Zariski open subset in the product  $\{p_0\} \times S^{2d-2}(\mathbb{P}_{\mathbf{C}}^2)$ , where  $S^{2d-2}(\mathbb{P}_{\mathbf{C}}^2)$  is the symmetric product of  $2d-2$  copies of the projective plane.

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

### 5.1.2. Tightness is a general property in $\text{Bir}_d(\mathbb{P}_{\mathbf{C}}^2)$

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  $g \in V_d$ :*

- (1)  *$g$  is a tight Cremona transformation;*
- (2) *if  $n \geq k$ , then  $g^n$  generates a normal subgroup of  $\text{Bir}(\mathbb{P}_{\mathbb{C}}^2)$  whose non-trivial elements  $f$  satisfy  $\lambda(f) \geq d^n$ .*

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

### 5.1.3. Strong algebraic stability is a general property

Given a surface  $X$ , a birational transformation  $f \in \text{Bir}(X)$  is said to be *algebraically stable* if one of the following equivalent properties hold:

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

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

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

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  $\text{Ind}(f) \subset \text{Exc}(f)$ ).

**LEMMA 5.3.** *If  $g$  is a general element of  $V_d$ , then*

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

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  $m \in \text{Ind}(g^{-1})$  such that  $g^k(m) \in \text{Exc}(g)$ , and hence to  $\text{Jac}_g(g^k(m)) = 0$ , where  $\text{Jac}_g$  is the equation of  $\text{Exc}(g)$ . This is an algebraic condition, which defines an algebraic subvariety  $I_k$  in  $\text{Bir}_d(\mathbb{P}_{\mathbb{C}}^2)$ .

**Fact 5.4.** There exists  $g \in V_d$  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 proving Fact 5.4. For this, let  $f$  be the Hénon mapping defined by

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

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

$$U = \left\{ [x:1:z] : |x| < \frac{1}{10} \text{ and } |z| < \frac{1}{10} \right\}.$$

Then,  $f(U)$  is contained in

$$\left\{ [x:1:z] : |x| < \frac{1}{90} \text{ and } |z| < \frac{1}{90} \right\}.$$

Let  $h$  be the linear transformation of the plane which, in the affine coordinates  $(x, z)$ , is the translation

$$(x, z) \mapsto \left( x + \frac{1}{20}, z + \frac{1}{20} \right).$$

We have  $h(f(U)) \subset U \setminus \{[x:y:0]\}$  and  $f(U) \cap h(f(U)) = \emptyset$ . We now take  $g = h \circ f$ . Its exceptional set is the line at infinity  $\{[x:y:0]\}$ ; the unique indeterminacy point of  $g^{-1}$  is  $h(q) = [\frac{1}{20}:1:\frac{1}{20}]$ . By construction, for all  $k \geq 0$ ,  $g^k(\text{Ind}(g^{-1}))$  is a point of  $U \setminus \text{Exc}(g)$ . This proves that  $g$  is not in  $I_k$  for any  $k \geq 0$ .  $\square$

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  $g^k(\text{Ind}(g^{-1}))$ ,  $k \geq 0$ ; similarly,  $g^{-1}$  is well defined along the backward orbit of  $\text{Ind}(g)$ .

LEMMA 5.5. *Let  $g$  be strongly algebraically stable. Then we have*

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

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

Property (2) is equivalent to property (1), replacing  $g$  by  $g^{-1}$ .

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

### 5.1.4. Rigidity is a general property

Let  $g$  be an element of  $V_d$ . Consider the isometry  $g_*$  of  $\bar{Z}(\mathbb{P}_{\mathbf{k}}^2)$ . 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  $H \subset \mathbb{P}_{\mathbb{C}}^2$  and use the fact that

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

as  $n \rightarrow \infty$  (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  $g^{-1}$ . Denote by  $[E^+]$  (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  $\text{Ind}(g)$  (resp.  $\text{Ind}(g^{-1})$ ). 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  $\text{Ind}(g)$  (resp.  $\text{Ind}(g^{-1})$ ). 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  $\mathbb{H}_{\bar{Z}}$ . All points on  $\text{Ax}(g_*)$  are linear combinations  $u[\alpha] + v[\omega]$  with the condition

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

$\cosh a_i$	$a_i$
3	$a_1 \simeq 1.76274$
$\sqrt{2}$	$a_2 = \frac{1}{2}a_1 \simeq 0.88137$
$3/\sqrt{2}$	$a_3 \simeq 1.38432$
$5/2\sqrt{2}$	$a_4 \simeq 1.17108$
4	$a_5 \simeq 2.06343$

Table 1. Distances and hyperbolic cosines.

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

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

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  $\mathbb{P}_{\mathbb{C}}^2$  (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  $\text{dist}(f_*[Q], [Q]) \leq \varepsilon_0$  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 in Table 1. Using this table, we see that

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

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

*Proof.* We proceed in two steps.

*Step 1.* We show that, if  $\text{dist}(f_*[P], [P]) \leq \varepsilon_0$ , then  $f$  is linear.

By the triangular inequality,

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

Recall that  $\cosh \operatorname{dist}([D_1], [D_2]) = [D_1] \cdot [D_2]$  for all pairs of points  $[D_1], [D_2] \in \mathbb{H}_{\overline{\mathbb{Z}}}$  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  $2 \operatorname{dist}([P], [H]) = 2a_2 = a_1$ . 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  $\deg(f) = 3$  and  $\deg(f) = 2$ . The following remark will be used twice: Since  $[H] \cdot [P] = \sqrt{2}$ , we have  $\operatorname{dist}([H], [P]) = a_2$ ; thus, if  $f_*$  is an isometry and  $\operatorname{dist}(f_*[P], [P]) \leq \varepsilon_0$ , applying hyperbolic cosines to the triangular inequality  $\operatorname{dist}(f_*[H], [P]) \leq \operatorname{dist}(f_*[H], f_*[P]) + \operatorname{dist}(f_*[P], [P])$ , 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  $\mathbb{P}_{\mathbb{C}}^2$  (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 the point  $[D]$  of 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

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

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  $\varepsilon_0$ .

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  $\mathbb{P}_{\mathbb{C}}^2$ . The product  $f_*[H] \cdot [P]$  is given by

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

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  $[E_1 + E_2 + E_3] \cdot [R] \geq -1 - 1/d \geq -\frac{3}{2}$ . As a consequence,

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

in contradiction with the choice of  $\varepsilon_0$ .

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

*Step 2.* We now show that, if  $f$  is in  $\text{Aut}(\mathbb{P}_{\mathbb{C}}^2)$  and  $\text{dist}(f_*[Q], [Q]) \leq \varepsilon_0$  for every  $[Q] \in \{g_*^{-1}[P], g_*[P]\}$ , then  $f$  is the identity map.

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

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

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

**COROLLARY 5.8.** *If  $g$  is a general element of  $V_d$ , then  $\text{Ax}(g_*)$  is rigid.*

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

*Proof of Theorem 5.2.* Let  $g$  be a general element of  $V_d$ , with  $d \geq 2$ . Suppose that  $f_*(\text{Ax}(g_*)) = \text{Ax}(g_*)$ . If  $f_*$  preserves the orientation on  $\text{Ax}(g_*)$ , then  $(fgf^{-1}g^{-1})_*$  fixes each point in  $\text{Ax}(g_*)$ , 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  $\text{Ax}(g_*)$  is rigid, we obtain that  $g_*$  is tight, and hence, by Theorem 2.10,  $g_*^k$  generates a proper subgroup of  $\text{Bir}(\mathbb{P}_{\mathbb{C}}^2)$  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  $\varepsilon_0$  is independent of  $d$ , so, by Corollary 3.4, we obtain that  $\text{Ax}(g_*)$  is  $(\frac{1}{2}\varepsilon_0, 7L(g_*))$ -rigid. Recall that  $\theta$  was defined in §2.2.1 by  $\theta = 8\delta$  and that  $\delta = \log 3$  works for the hyperbolic space  $\mathbb{H}_{\mathbb{Z}}$ ; thus we can choose  $\theta = 8 \log 3$ , so that  $14\theta = 112 \log 3 = 123.4 \dots \leq 124$ . By Lemma 3.2, we obtain that  $\text{Ax}(g_*)$  is also  $(14\theta, B)$ -rigid for

$$\begin{aligned} B &\geq \max \left\{ 2728, 2232 + 2 \log \left( 5 \frac{\cosh 56\theta - 1}{\cosh \frac{1}{2}\varepsilon_0 - 1} \right), \frac{28}{3} L(g_*) + 248 \right\} \\ &= \max \left\{ 3220, \frac{28}{3} L(g_*) + 248 \right\}. \end{aligned}$$

Then, Theorem 2.10 requires  $kL(g_*) \geq 40B + 1200\theta$ . Thus, we see that

$$k = \max \left\{ \frac{139347}{L(g_*)}, 374 + \frac{10795}{L(g_*)} \right\} \leq 201021$$

is sufficient to conclude that  $g^k$  generates a proper normal subgroup of  $\text{Bir}(\mathbb{P}_{\mathbb{C}}^2)$ . Asymptotically, for degrees  $d$  with  $d \geq e^{10796}$ , we can take  $k=375$ .  $\square$

## 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  $\varphi: X \dashrightarrow \mathbb{P}_{\mathbf{k}}^2$  is a birational map, conjugation by  $\varphi$  provides an isomorphism between  $\text{Bir}(X)$  and  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ , and  $\varphi_*$  conjugates the actions of  $\text{Bir}(X)$  on  $\overline{\mathcal{Z}}(X)$  and  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  on  $\overline{\mathcal{Z}}(\mathbb{P}_{\mathbf{k}}^2)$ . We thus identify  $\text{Bir}(X)$  with  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  and  $\overline{\mathcal{Z}}(X)$  with  $\overline{\mathcal{Z}}(\mathbb{P}_{\mathbf{k}}^2)$  without further reference to the choice of  $\varphi$ . This paragraph provides a simple criterion to check whether  $g$  is a tight element of the Cremona group  $\text{Bir}(X)$ .

As explained in §4.6.3, the Néron–Severi group  $\mathbf{N}^1(X)$  embeds into the Picard–Manin space  $\overline{\mathcal{Z}}(\mathbb{P}_{\mathbf{k}}^2)$  and  $g_*$  preserves the orthogonal decomposition

$$\overline{\mathcal{Z}}(\mathbb{P}_{\mathbf{k}}^2) = \mathbf{N}^1(X)_{\mathbf{R}} \oplus \mathbf{N}^1(X)_{\mathbf{R}}^{\perp},$$

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

*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  $\mathbf{N}^1(X)_{\mathbf{R}}$ ;
- (ii) the action of  $\text{Aut}(X)$  on  $\mathbf{N}^1(X)$  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  $\mathbf{N}^1(X)$ . 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  $\mathbf{N}^1(X)$ , 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  $\mathbf{N}^1(X)$ . Once again, assumption (ii) implies that  $h$  conjugates  $g$  to its inverse.  $\square$

*Remark 5.11.* The plane  $V_g$  is a subspace of  $\mathbf{N}^1(X)_\mathbf{R}$  and it may very well happen that this plane does not intersect the lattice  $\mathbf{N}^1(X)$  (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,  $V_g \cap \mathbf{N}^1(X)$  is a rank-2 lattice in the plane  $V_g$ . This lattice is  $g_*$ -invariant, and the spectral values of the linear transformation  $g_* \in \mathrm{GL}(V_g)$  are quadratic integers. From this it follows that  $\lambda(g)$  is a quadratic integer. On the other hand, 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  $V_g^\perp \cap \mathbf{N}^1(X)$  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  $[D'] \in \mathbf{N}^1(X) \cap V_g$ . In that case,  $[D']$  and  $g_*[D']$  generate a rank-2 subgroup of  $\mathbf{N}^1(X) \cap V_g$ , and Remark 5.11 implies that  $\lambda(g)$  is a quadratic integer. In what follows, we shall denote by  $[D]$  the class

$$[D] = \frac{1}{\sqrt{[D'] \cdot [D']}} [D']. \quad (5.2)$$

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

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

$$\cosh \operatorname{dist}(h_*[D], [D]) < 1 + \frac{1}{[D'] \cdot [D']},$$

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

*Proof.* Write  $h_*[D] = [D] + [F] + [R]$ , where  $[F]$  is in  $\mathbf{N}^1(X)_\mathbf{R}$  and  $[R]$  is in the subspace  $\mathbf{N}^1(X)_\mathbf{R}^\perp$  of  $\bar{\mathbf{Z}}(X)$ . More precisely,  $F$  is an element of  $\mathbf{N}^1(X)$  divided by the square root of the self-intersection  $[D']^2$ , and  $[R]$  is a sum  $\sum_i m_i [E_i]$  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  $h_*[D]$  is zero, and  $[D]$  is fixed. Since  $[D]$  is ample,  $h$  is an automorphism of  $X$ .  $\square$

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

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

*Then  $\text{Ax}(g_*)$  is rigid. Assume furthermore that*

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

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

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

*Proof.* Let  $[D] \in \text{Ax}(g_*)$  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,  $f_*(\text{Ax}(g_*)) \neq \text{Ax}(g_*)$ . Lemma 5.12 implies that  $f$  is an automorphism of  $X$  fixing both  $[D]$  and  $g_*[D]$ . This contradicts  $f_*(\text{Ax}(g_*)) \neq \text{Ax}(g_*)$  and shows that  $\text{Ax}(g_*)$  is rigid.

Assume now that  $h \in \text{Bir}(X)$  preserves the axis of  $g_*$ . Then  $h_*[D']$  is an ample class, and hence  $h$  is an automorphism of  $X$ . Property (iii) implies that  $g$  is a tight element of  $\text{Bir}(X)$ . The conclusion follows from Theorem 2.10.  $\square$

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. The first example is defined over the field of complex numbers  $\mathbb{C}$ , while the second works for any algebraically closed field  $\mathbf{k}$ .

### 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  $\mathbf{C}/\mathbf{Z}[i] \times \mathbf{C}/\mathbf{Z}[i]$  (in this subsection  $i$  denotes the imaginary unit). The group  $\text{GL}_2(\mathbf{Z}[i])$  acts by linear transformations on  $\mathbf{C}^2$ , preserving the lattice  $\mathbf{Z}[i] \times \mathbf{Z}[i]$ ; this provides an embedding  $\text{GL}_2(\mathbf{Z}[i]) \rightarrow \text{Aut}(Y)$ . 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 quotients of tori by  $(x, y) \mapsto (-x, -y)$ , and are not rational (these surfaces are examples of K3-surfaces).

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

- (i) the trace  $\mathrm{tr}(M)$  of  $M$  is at least 3;
- (ii)  $M$  is in the level-2 congruence subgroup of  $\mathrm{SL}_2(\mathbf{Z})$ , 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  $g: X \rightarrow X$  satisfies properties (i)–(iii) of Proposition 5.13. In particular,  $g$  determines a tight element of  $\mathrm{Bir}(\mathbb{P}_{\mathbf{C}}^2)$ ; hence, if  $k$  is large enough,  $g^k$  generates a non-trivial normal subgroup of  $\mathrm{Bir}(\mathbb{P}_{\mathbf{C}}^2)$ .*

*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  $\{(z, iz) \in Y : z \in \mathbf{C}/\mathbf{Z}[i]\}$ .

The vector space  $H^2(Y, \mathbf{R})$  is isomorphic to the space of bilinear alternating 2-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

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

where  $a = \mathrm{tr}(M)^2 - 2 > 2$ . Thus, the plane  $V_{\hat{g}}$  intersects  $\mathbf{N}^1(Y)$  on a lattice. Let  $[F]$  be an element of  $V_{\hat{g}} \cap \mathbf{N}^1(X)$  with  $[F]^2 > 0$  (see Remark 5.11). Since  $Y$  is an abelian variety,  $[F]$  (or its opposite) is ample. Since  $M$  has integer coefficients, the linear map  $\hat{g}_*$  preserves the 3-dimensional subspace  $W$  of  $\mathbf{N}^1(Y)$  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  $V_{\hat{g}}^{\perp} \subset \mathbf{N}^1(Y)$ .

Transporting this picture in  $\mathbf{N}^1(X)$ , we obtain that the dynamical degree of  $g$  is equal to the dynamical degree of  $\hat{g}$  (see [35], for more general results), the plane  $V_{\hat{g}}$  surjects onto  $V_g$ , the image of  $[F]$  is an ample class  $[D']$  contained in  $V_g \cap \mathbf{N}^1(X)$ , and  $g_*$  is the identity on  $V_g^{\perp}$ .

Automorphisms of  $X$  permute the ten singularities of  $X$ . The fundamental group  $\Gamma$  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  $\eta$ . The

abelian group  $\mathbf{Z}[i] \times \mathbf{Z}[i]$  is the unique maximal free abelian subgroup of rank 4 in  $\Gamma$  and, as such, is invariant under all automorphisms of  $\Gamma$ . This implies that all automorphisms of  $X$  lift to (affine) automorphisms of  $Y$ .

Let  $h$  be an automorphism of  $X$  which preserves the axis  $\text{Ax}(g_*)$ . 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  $N \in \text{GL}_2(\mathbf{C})$  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

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

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  $\eta$ . Thus  $a$  and  $b$  are in  $\frac{1}{2}\mathbf{Z}[i]$ . Since  $M$  is the identity modulo 2, we have  $M(a, b) = (a, b)$  modulo  $\mathbf{Z}[i] \times \mathbf{Z}[i]$ . Hence  $\hat{h} \circ \hat{g} \circ \hat{h}^{-1} = \hat{g}^{\pm 1}$  and, coming back to  $X$ ,  $h$  conjugates  $g$  to  $g$  itself or its inverse.  $\square$

*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  $\eta$  by  $\eta(x, y) = (jx, jy)$ . This leads to a second rational Kummer surface with an action of the group  $\text{PSL}_2(\mathbf{Z})$ ; 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  $S \subset \mathbb{P}_{\mathbf{k}}^2$  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  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2)$ , they depend on nine parameters (see [36], the appendix of [33], or [14]). 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  $\pi: X \rightarrow \mathbb{P}_{\mathbf{k}}^2$  be the natural projection and  $E_i$ ,  $1 \leq i \leq 10$ , be the exceptional divisors of  $\pi$ . The canonical class of  $X$  is

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

where  $[H]$  is the pull-back by  $\pi$  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  $\Omega$  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 [26]). Our definition is more restrictive.

The self-intersection of  $S'$  is  $-4$ , and  $S'$  can be blown down: This provides a birational morphism  $q: X \rightarrow X_0$ ; the surface  $X_0$  has a unique singularity, at  $m=q(S')$ . The section  $\Omega$  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  $\text{Aut}(X)$  acts linearly on the space of sections of  $-2K_X$ , and preserves its base locus  $S'$ . It follows that  $q$  conjugates  $\text{Aut}(X)$  and  $\text{Aut}(X_0)$ .

The Néron–Severi group  $\mathbf{N}^1(X)$  has rank 11. Let  $W$  be the orthogonal complement of  $[K_X]$  with respect to the intersection form. The linear map  $q^*: \mathbf{N}^1(X_0) \rightarrow \mathbf{N}^1(X)$  provides an isomorphism between  $\mathbf{N}^1(X_0)$  and its image  $W = [K_X]^\perp \subset \mathbf{N}^1(X)$ . Let  $O'(\mathbf{N}^1(X))$  be the group of isometries of the lattice  $\mathbf{N}^1(X)$  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 [23] and [14, p. 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 points  $m_i$  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) *Let  $\mathbf{k}$  be an algebraically closed field. The set of special sextics is a proper Zariski closed subset of the space of rational sextic curves  $S \subset \mathbb{P}_{\mathbf{k}}^2$ . If  $S$  is not special and  $X$  is the associated Coble surface, then the morphism*

$$\begin{aligned} \text{Aut}(X) &\longrightarrow O'(\mathbf{N}^1(X)), \\ f &\longmapsto f_*, \end{aligned}$$

*is injective and its image is the level-2 congruence subgroup of  $O'(\mathbf{N}^1(X))$ .*

This theorem is due to Coble, and can be found in [12]. A proof is sketched in [25] and [24], and a complete, characteristic-free, proof is available in [7]; an analogous result and proof for generic Enriques surfaces can be found in [22]. The main steps are the following. The automorphism group of  $X$  can be identified with a normal subgroup of  $O'(\mathbf{N}^1(X))$ ; this group is generated, as a normal subgroup, by an explicit involution. To realize this involution by an automorphism, one constructs a two-to-one morphism from  $X$  to a del Pezzo surface and takes the involution of this cover.

*Remark 5.18.* Let  $\text{Amp}(W)$  be the set of ample classes in  $\mathbf{N}^1(X_0) \otimes \mathbf{R} \simeq W \otimes \mathbf{R}$ . This convex cone is invariant under the action of  $\text{Aut}(X)$ , and hence under the action of a finite-index subgroup of the orthogonal group  $O(W)$ . As such,  $\text{Amp}(W)$  is equal to

$$\{[D] \in W : [D]^2 > 0 \text{ and } [D] \cdot [H] > 0\}.$$

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  $\psi$  be an isometry of the lattice  $W$ ; if  $\psi$  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  $X$ .

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  $\phi$  in the orthogonal group  $O_Q(\mathbf{Z})$  with eigenvalues  $\lambda(Q) > 1 > 1/\lambda(Q) > 0$ .*

*Sketch of proof.* Isometries in  $O_Q(\mathbf{Z})$  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 solving a Pell–Fermat equation (see [30, §V.1]).  $\square$

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

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

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  $\phi$  of  $V$  with an eigenvalue  $\lambda(\phi) > 1$ . In fact, an explicit computation shows that the group of isometries of  $Q$  is the semi-direct product of the group  $\mathbf{Z}/2\mathbf{Z}$  generated by the involution which permutes  $[D_1]$  and  $[D_2]$ , and the cyclic group  $\mathbf{Z}$  generated by the isometry  $\phi$  defined by

$$\phi([D_1]) = 4[D_1] + [D_2] \quad \text{and} \quad \phi([D_2]) = -[D_1].$$

The second iterate of  $\phi$  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  $\phi^2$  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)–(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  $\mathrm{Bir}(X_{\mathbf{k}})=\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ .*

As a corollary, the Cremona group  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  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  $\mathbb{P}_{\mathbb{C}}^2$ , were both sources of inspiration for the results in this paper. We now use these groups to construct hyperbolic elements  $g$  of  $\mathrm{Bir}(\mathbb{P}_{\mathbb{C}}^2)$  for which  $\langle\langle g \rangle\rangle$  coincides with  $\mathrm{Bir}(\mathbb{P}_{\mathbb{C}}^2)$ .

#### 6.1.1. Monomial transformations

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

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

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  $\mathrm{PSL}_2(\mathbf{Z})$  as a subgroup of  $\mathrm{PSL}_2(\mathbf{R}) \simeq \mathrm{Isom}(\mathbb{H}^2)$ , it is an interesting exercise to check that all hyperbolic matrices of  $\mathrm{PSL}_2(\mathbf{Z})$  are tight elements of  $\mathrm{PSL}_2(\mathbf{Z})$  (see [43]). However, when we see  $\mathrm{GL}_2(\mathbf{Z})$  as a subgroup of the Cremona group, we obtain the following striking remark.

**PROPOSITION 6.1.** *Let  $g: \mathbb{P}_{\mathbf{k}}^2 \rightarrow \mathbb{P}_{\mathbf{k}}^2$  be a non-trivial monomial transformation. The normal subgroup of  $\mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  generated by  $g$  is not proper:  $\langle\langle g \rangle\rangle = \mathrm{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ .*

*Remark 6.2.* (Gizatullin and Noether) If  $N$  is a normal subgroup of  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  containing a non-trivial automorphism of  $\mathbb{P}_{\mathbf{k}}^2$ , then  $N$  coincides with  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ . The proof is as follows (see [34] and [9]). Since  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2)$  is the simple group  $\text{PGL}_3(\mathbf{k})$  and  $N$  is normal,  $N$  contains  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2)$ . 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  $\sigma$  satisfies  $\sigma = (h\sigma)h(h\sigma)^{-1}$ ; hence,  $\sigma$  is conjugate to  $h$ , and  $\sigma$  is contained in  $N$ . The conclusion follows from Noether's theorem, which states that  $\sigma$  and  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2)$  generate  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  (see [41, §2.5]).

*Proof.* Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be any non-trivial monomial map in  $\text{Bir}(\mathbb{P}_{\mathbb{C}}^2)$ . The commutator of  $g$  with the diagonal map  $f(x, y) = (\alpha x, \beta y)$  is the diagonal map

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

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

### 6.1.2. Polynomial automorphisms

As mentioned in the introduction, Danilov proved that the group  $\text{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  of polynomial automorphisms of the affine plane  $\mathbb{A}_{\mathbf{k}}^2$  with Jacobian determinant 1 is not simple. Danilov's proof uses an action on a tree. Since  $\text{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  is a subgroup of  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ , we also have the action of  $\text{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  on the hyperbolic space  $\mathbb{H}_{\bar{\mathbf{k}}}$ . It is a nice observation that  $g \in \text{Aut}[\mathbb{A}_{\mathbf{k}}^2]_1$  determines a hyperbolic isometry of the tree if and only if it determines a hyperbolic isometry of  $\mathbb{H}_{\bar{\mathbf{k}}}$ : In both cases, hyperbolicity corresponds to 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  $\mathbb{A}_{\mathbf{k}}^2$  defined by

$$h(x, y) = (y, p(y) - ax).$$

This automorphism determines an algebraically stable birational transformation of  $\mathbb{P}_{\mathbf{k}}^2$ , namely

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

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

**PROPOSITION 6.3.** (See also [9]) *For all integers  $d \geq 2$ , equation (6.2) defines a subset  $H_d \subset \mathcal{V}_d$  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  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ .*

*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  $\mathcal{V}_d$ . 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 (and 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  $\text{Aut}(\mathbb{P}_{\mathbf{k}}^2)$  non-trivially. The conclusion follows from Remark 6.2.  $\square$

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

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

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  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ ?

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  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  comes from the fact that  $\text{Bir}(X)$  is small compared to  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$ , 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  $\text{Bir}(X)$  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 proof.* Assume, first, that the Kodaira dimension of  $X$  is non-negative. Replace  $X$  by its unique minimal model, and identify  $\text{Bir}(X)$  with  $\text{Aut}(X)$ . The group  $\text{Aut}(X)$  acts on the homology of  $X$ , and the kernel is equal to its connected component  $\text{Aut}(X)^0$  up to finite index. The action on the homology group provides a morphism to  $\text{GL}_n(\mathbf{Z})$  for some  $n \geq 1$ . Reducing modulo  $p$  for large primes  $p$ , one sees that  $\text{GL}_n(\mathbf{Z})$  is residually finite. Since  $\text{Aut}(X)$  is assumed to be simple, this implies that  $\text{Aut}(X)$  coincides with  $\text{Aut}(X)^0$ . But  $\text{Aut}(X)^0$  is abelian for surfaces with non-negative Kodaira dimension (see [1]). Thus  $\text{Bir}(X) = \text{Aut}(X)$  is not both infinite and simple when the Kodaira dimension of  $X$  is  $\geq 0$ . Assume now that  $X$  is ruled and not rational. Up to a birational change of coordinates,  $X$  is a product  $\mathbb{P}_{\mathbf{C}}^1 \times C$  where  $C$  is a smooth curve of genus  $g(C) \geq 1$ . The group  $\text{Bir}(X)$  projects surjectively onto  $\text{Aut}(C)$ . By simplicity,  $\text{Aut}(C)$  must be trivial. In that case,  $\text{Bir}(X)$  coincides with the infinite simple group  $\text{PGL}_2(\mathcal{M}(C))$ , 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.  $\square$

### 6.3. SQ-universality and the number of quotients

As a direct consequence of Theorem 2.10 and the existence of tight elements in  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$ , the Cremona group  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  has an uncountable number of distinct normal subgroups. Recently, Dahmani, Guirardel and Osin obtained a better, much more powerful version of Theorem 2.10, which applies to the Cremona group in the same way as Theorem C implies Theorem A, because the existence of tight elements in  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  is sufficient to apply Dahmani, Guirardel and Osin's theorems [15]. We only describe one consequence of their results that strengthen the above mentioned fact that  $\text{Bir}(\mathbb{P}_{\mathbf{C}}^2)$  has uncountably many normal subgroups, and refer to [15] for other statements.

A group is said to be *SQ-universal* (or *SubQuotient-universal*) if every countable group can be embedded into one of its quotients. For example, the pioneering work [38] 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 Ol'shanskii in [17] and [49].

**THEOREM 6.6.** (See [15]) *If  $\mathbf{k}$  is an algebraically closed field, then the Cremona group  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  is SQ-universal.*

Note that SQ-universality implies the existence of uncountably many non-isomorphic quotients.

## Appendix A. The Cremona group is not an amalgam

by Yves de Cornulier

Let  $\mathbf{k}$  be a field. The *Cremona group*  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^d)$  of  $\mathbf{k}$  in dimension  $d$  is defined as the group of birational transformations of the  $d$ -dimensional  $\mathbf{k}$ -affine space. It can also be described as the group of  $\mathbf{k}$ -automorphisms of the field of rational functions  $\mathbf{k}(t_1, \dots, t_d)$ . We endow it with the discrete topology.

Let us say that a group has *property*  $(\mathbf{FR})_{\infty}$  if it satisfies the following condition:

(A1) for every isometric action on a complete real tree, every element has a fixed point.

Here we prove the following result.

**THEOREM A.1.** *If  $\mathbf{k}$  is an algebraically closed field, then  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  has property  $(\mathbf{FR})_{\infty}$ .*

**COROLLARY A.2.** *The Cremona group does not decompose as a non-trivial amalgam.*

Recall that a *real tree* can be defined in the following equivalent ways (see [11]):

- a geodesic metric space which is 0-hyperbolic in the sense of Gromov;
- a uniquely geodesic metric space for which  $[ac] \subset [ab] \cup [bc]$  for all  $a, b$  and  $c$ ;
- a geodesic metric space with no subspace homeomorphic to the circle.

In a real tree, a *ray* is a geodesic embedding of the half-line. An *end* is an equivalence class of rays modulo being at bounded distance. For a group of isometries of a real tree, to *stably fix* an end means to pointwise stabilize a ray modulo eventual coincidence (this means that it fixes the end as well as the corresponding Busemann function).

For a group  $\Gamma$ , property  $(\mathbf{FR})_{\infty}$  has the following equivalent characterizations:

(A2) for every isometric action of  $\Gamma$  on a complete real tree, every finitely generated subgroup has a fixed point;

(A3) every isometric action of  $\Gamma$  on a complete real tree either has a fixed point, or stably fixes a point at infinity (in the sense above).

The equivalence between these three properties is justified in Lemma A.9. Similarly, we can define the weaker *property*  $(\mathbf{FA})_{\infty}$ , replacing complete real trees by ordinary trees (and allowing fixed points to be middle of edges), and the three corresponding equivalent properties are equivalent [51] to the following fourth property: the group is not a non-trivial amalgam and has no homomorphism onto the group of integers. In particular, Corollary A.2 follows from Theorem A.1.

*Remark A.3.* (a) Note that the statement for actions on real trees (rather than trees) is strictly stronger. Indeed, unless  $\mathbf{k}$  is algebraic over a finite field, the group  $\text{PGL}_2(\mathbf{k}) = \text{Bir}(\mathbb{P}_{\mathbf{k}}^1)$  does act isometrically on a real tree with a hyperbolic element (this

uses the existence of a non-trivial real-valued valuation on  $\mathbf{k}$ , but does not have such an action on a discrete tree (see Proposition A.8).

(b) Note that  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  always has an action on a discrete tree with no fixed point (i.e. no fixed point on the 1-skeleton) when  $\mathbf{k}$  is algebraically closed, and more generally whenever  $\mathbf{k}$  is an infinitely generated field: write, with the help of a transcendence basis,  $\mathbf{k}$  as the union of an increasing sequence of proper subfields  $\mathbf{k} = \bigcup_{n=1}^{\infty} \mathbf{k}_n$ , then  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  is the increasing union of its proper subgroups  $\text{Bir}(\mathbb{P}_{\mathbf{k}_n}^2)$ , and thus acts on the disjoint union of the coset spaces  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)/\text{Bir}(\mathbb{P}_{\mathbf{k}_n}^2)$ , which is in a natural way the vertex set of a tree on which  $\text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$  acts with no fixed point (this is a classical construction of Serre [51, Chapter I, §6.1]).

(c) Theorem A.1 could be stated, with a similar proof, for actions on  $\Lambda$ -trees when  $\Lambda$  is an arbitrary ordered abelian group (see [11] for an introduction to  $\Lambda$ -metric spaces and  $\Lambda$ -trees).

In the following,  $\mathcal{T}$  is a complete real tree; all actions on  $\mathcal{T}$  are assumed to be isometric. We begin by a few lemmas.

LEMMA A.4. *Let  $x_0, \dots, x_k$  be points in a real tree  $\mathcal{T}$  and  $s \geq 0$ . Assume that*

$$d(x_i, x_j) = s|i - j|$$

*holds for all  $i$  and  $j$  such that  $|i - j| \leq 2$ . Then it holds for all  $i$  and  $j$ .*

*Proof.* This is an induction; for  $k \leq 2$  there is nothing to prove. Suppose  $k \geq 3$  and the result known up to  $k - 1$ , so that the formula holds except maybe for  $\{i, j\} = \{0, k\}$ . Join  $x_i$  to  $x_{i+1}$  by segments. By the induction, the  $k - 1$  first segments, and the  $k - 1$  last segments, concatenate to geodesic segments. But the first and the last of these  $k$  segments are also disjoint, otherwise by picking the “smallest” point in the last segment that also belongs to the first one, we find an injective loop, contradicting that  $\mathcal{T}$  is a real tree. Therefore the  $k$  segments concatenate to a geodesic segment and  $d(x_0, x_k) = sk$ .  $\square$

LEMMA A.5. *If  $\mathbf{k}$  is any field and  $d \geq 3$ , then  $\Gamma = \text{SL}_d(\mathbf{k})$  has property  $(\text{FR})_{\infty}$ . In particular, if  $\mathbf{k}$  is algebraically closed, then  $\text{PGL}_d(\mathbf{k})$  has property  $(\text{FR})_{\infty}$ .*

*Proof.* Let  $\Gamma$  act on  $\mathcal{T}$ . Let  $F$  be a finite subset of  $\Gamma$ . Every element of  $F$  can be written as a product of elementary matrices. Let  $A$  be the (finitely generated) subring of  $\mathbf{k}$  generated by all entries of those matrices. Then  $F \subset \text{EL}_d(A)$ , the subgroup of  $\text{SL}_d(A)$  generated by elementary matrices. By the Shalom–Vaserstein theorem (see [28]),  $\text{EL}_d(A)$  has Kazhdan’s property (T) and in particular has a fixed point in  $\mathcal{T}$ , so  $F$  has a fixed point in  $\mathcal{T}$ . (There certainly exists a more elementary proof, but this one also shows that for every isometric action of  $\text{SL}_d(\mathbf{k})$  on a Hilbert space, every finitely generated subgroup fixes a point.)  $\square$

Fix the following notation:  $G = \text{Bir}(\mathbb{P}_{\mathbf{k}}^2)$ ,  $H = \text{PGL}_3(\mathbf{k}) = \text{Aut}(\mathbb{P}_{\mathbf{k}}^2) \subset G$  and  $\sigma$  is the Cremona involution, acting in affine coordinates by  $\sigma(x, y) = (x^{-1}, y^{-1})$ . The Noether–Castelnuovo theorem is that  $G = \langle H, \sigma \rangle$ . Let  $C$  be the standard Cartan subgroup of  $H$ , that is, the semi-direct product of the diagonal matrices by the Weyl group (of order 6). Let  $\mu \in H$  be the involution given in affine coordinates by  $\mu(x, y) = (1-x, 1-y)$ .

LEMMA A.6. *We have  $\langle C, \mu \rangle = H$ .*

*Sketch of proof.* We only give a sketch, the details being left to the reader. In  $\text{GL}_3$ ,  $\mu$  can be written as the matrix

$$\begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Multiply  $\mu$  by its conjugate by a suitable diagonal matrix to obtain an elementary matrix; conjugating by elements of  $C$  provide all elementary matrices and thus we obtain all matrices with determinant 1; since  $C$  also contains diagonal matrices, we are done.  $\square$

LEMMA A.7. *Let  $G$  act on  $\mathcal{T}$  so that  $H$  has no fixed point and has a (unique) stably fixed end. Then  $G$  stably fixes this unique end.*

*Proof.* Let  $\omega$  be the unique end stably fixed by  $H$  (recall that, if it is represented by a ray  $(x_t)_{t \geq 0}$ , this means that for every  $h \in H$  there exists  $t_0 = t_0(h)$  such that  $h$  fixes  $x_t$  for all  $t \geq t_0$ ). Then  $\sigma H \sigma^{-1}$  stably fixes  $\sigma \omega$ . In particular, since  $\sigma C \sigma^{-1} = C$ , the end  $\sigma \omega$  is also stably fixed by  $C$ . If  $\sigma \omega = \omega$ , then  $\omega$  is stably fixed by  $\sigma$  and then by the Noether–Castelnuovo theorem,  $\omega$  is stably fixed by  $G$ . Otherwise, let  $D$  be the line joining  $\omega$  and  $\sigma \omega \neq \omega$ . Since both ends of  $D$  are stably fixed by  $C$ , the line  $D$  is pointwise fixed by  $C$ . Also,  $\mu$  stably fixes the end  $\omega$  and therefore, for some  $t$ ,  $x_t$  is fixed by  $\mu$  and hence, by Lemma A.6, is fixed by all of  $H$ , contradicting the assumption.  $\square$

*Proof of Theorem A.1.* Note that  $\mu \in H$  and  $\mu \sigma$  has order 3. It follows that

$$\sigma = (\mu \sigma) \mu (\mu \sigma)^{-1}.$$

Using the Noether–Castelnuovo theorem, it follows that  $H_1 = H$  and  $H_2 = \sigma H \sigma^{-1}$  generate  $G$ .

Consider an action of  $G$  on  $\mathcal{T}$ . By Lemmas A.5 and A.7, we only have to consider the case when  $H$  has a fixed point; in this case, let us show that  $G$  has a fixed point. Assume the contrary. Let  $\mathcal{T}_i$  be the set of fixed points of  $H_i$ ,  $i=1, 2$ ; they are exchanged by  $\sigma$  and, since  $\langle H_1, H_2 \rangle = G$ , we see that the two trees  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are disjoint. Let  $\mathcal{S} = [x_1, x_2]$  be the minimal segment joining the two trees ( $x_i \in \mathcal{T}_i$ ), and  $s > 0$  be its length. Then  $\mathcal{S}$  is pointwise fixed by  $C \subset H_1 \cap H_2$  and reversed by  $\sigma$ .

*Claim.* For all  $k \geq 1$ , the distance between  $x_1$  and  $(\sigma\mu)^k x_1$  is exactly  $sk$ .

The claim is clearly a contradiction since  $(\sigma\mu)^3 = 1$ . To check the claim, let us apply Lemma A.4 to the sequence  $\{(\sigma\mu)^k x_1\}_{k=0}^\infty$ : namely to check that

$$d((\sigma\mu)^k x_1, (\sigma\mu)^l x_1) = |k-l|s$$

for all  $k$  and  $l$ , it is enough to check it for  $|k-l| \leq 2$ ; by translation, it is enough to check it for  $k=1, 2$  and  $l=0$ . For  $k=1$ ,  $d(\sigma\mu x_1, x_1) = d(\sigma x_1, x_1) = d(x_2, x_1) = s$ . Since  $\langle C, \mu \rangle = H$ , by Lemma A.6, the image of  $[x_1, x_2]$  by  $\mu$  is a segment  $[x_1, \mu x_2]$  intersecting the segment  $[x_1, x_2]$  only at  $x_1$ ; in particular,  $d(x_2, \mu x_2) = 2s$ . Hence, under the assumptions of the claim,

$$d(\sigma\mu\sigma\mu x_1, x_1) = d(\mu\sigma x_1, \sigma x_1) = d(\mu x_2, x_2) = 2s.$$

This proves the claim for  $k=2$  and the proof is complete.  $\square$

For reference, we include the following results.

PROPOSITION A.8. *If  $\mathbf{k}$  is algebraically closed, the group  $\mathrm{PGL}_2(\mathbf{k})$  has property  $(\mathrm{FA})_\infty$  but, unless  $\mathbf{k}$  is an algebraic closure of a finite field, does not satisfy  $(\mathrm{FR})_\infty$ .*

*Proof.* If  $\mathbf{k}$  has characteristic zero, the group  $\mathrm{PGL}_2(\mathbf{k})$  has the property that the square of every element is divisible (i.e. has  $n$ th roots for all  $n > 0$ ). This implies that no element can act hyperbolically on a discrete tree: indeed, in the automorphism group of a tree, the translation length of any element is an integer and the translation length of  $x^n$  is  $n$  times the translation length of  $x$ . If  $\mathbf{k}$  has characteristic  $p$ , the same argument holds: for every  $x$  we have that  $x^{2^p}$  is divisible.

On the other hand, let  $I$  be a transcendence basis of  $\mathbf{k}$  and assume that it is non-empty, and let  $x_0 \in I$ . Set  $\mathbf{L} = \mathbf{k}(I - \{x_0\})$ , so that  $\mathbf{k}$  is an algebraic closure of  $\mathbf{L}(x_0)$ . The non-trivial discrete valuation of  $\mathbf{L}((x_0))$  uniquely extends to a non-trivial  $\mathbf{Q}$ -valued valuation on an algebraic closure. It restricts to a non-trivial  $\mathbf{Q}$ -valued valuation on  $\mathbf{k}$ .

The remaining case is the case of an algebraic closure of the rational field  $\mathbf{Q}$ ; pick any prime  $p$  and restrict the  $p$ -valuation from an algebraic closure of  $\mathbf{Q}_p$ .

Now, if  $F$  is any field valued in  $\mathbf{R}$ , then  $\mathrm{PGL}_2(F)$  has a natural action on a real tree, on which an element  $\mathrm{diag}(a, a^{-1})$ , for  $|a| > 1$ , acts hyperbolically.

(If  $\mathbf{k}$  is algebraic over a finite field, then  $\mathrm{PGL}_2(\mathbf{k})$  is locally finite and thus satisfies  $(\mathrm{FA})_\infty$ .)  $\square$

LEMMA A.9. *The three definitions of  $(\mathrm{FR})_\infty$  in the introduction are equivalent.*

*Sketch of proof.* The implications  $(\mathrm{A3}) \Rightarrow (\mathrm{A2}) \Rightarrow (\mathrm{A1})$  are clear.

(A1)  $\Rightarrow$  (A2) This is proved for trees in [51, Chapter I, §6.5], the argument working for real trees.

(A2)  $\Rightarrow$  (A3) Fix a point  $x_0$ . For every finite subset  $F$  of the group, let  $\mathcal{S}_F$  be the segment joining  $x_0$  to the set of  $F$ -fixed points. Then the union of  $\mathcal{S}_F$ , when  $F$  ranges over finite subsets of the group, is a geodesic emanating from 0. If it is bounded, its other extremity (which exists by completeness) is a fixed point. Otherwise, it defines a stably fixed end.  $\square$

### References

- [1] AKHIEZER, D. N., *Lie Group Actions in Complex Analysis*. Aspects of Mathematics, E27. Vieweg, Braunschweig, 1995.
- [2] BLANC, J., Sous-groupes algébriques du groupe de Cremona. *Transform. Groups*, 14 (2009), 249–285.
- [3] —, Groupes de Cremona, connexité et simplicité. *Ann. Sci. Éc. Norm. Supér.*, 43 (2010), 357–364.
- [4] BOUCKSOM, S., FAVRE, C. & JONSSON, M., Degree growth of meromorphic surface maps. *Duke Math. J.*, 141 (2008), 519–538.
- [5] BRIDSON, M. R. & HAEFLIGER, A., *Metric Spaces of Non-Positive Curvature*. Grundlehren der Mathematischen Wissenschaften, 319. Springer, Berlin–Heidelberg, 1999.
- [6] CANTAT, S., Sur les groupes de transformations birationnelles des surfaces. *Ann. of Math.*, 174 (2011), 299–340.
- [7] CANTAT, S. & DOLGACHEV, I., Rational surfaces with a large group of automorphisms. *J. Amer. Math. Soc.*, 25 (2012), 863–905.
- [8] CANTAT, S. & LAMY, S., Normal subgroups in the Cremona group (long version). Preprint, 2010. [arXiv:1007.0895 \[math.AG\]](https://arxiv.org/abs/1007.0895).
- [9] CERVEAU, D. & DÉSERTE, J., Transformations birationnelles de petit degré. Unpublished notes, 2009.
- [10] CHAYNIKOV, V., On the generators of the kernels of hyperbolic group presentations. *Algebra Discrete Math.*, 11 (2011), 18–50.
- [11] CHISWELL, I., *Introduction to  $\Lambda$ -Trees*. World Scientific, River Edge, NJ, 2001.
- [12] COBLE, A. B., *Algebraic Geometry and Theta Functions*. American Mathematical Society Colloquium Publications, 10. Amer. Math. Soc., Providence, RI, 1982.
- [13] COORNAERT, M., DELZANT, T. & PAPADOPOULOS, A., *Géométrie et théorie des groupes*. Lecture Notes in Mathematics, 1441. Springer, Berlin–Heidelberg, 1990.
- [14] COSSEC, F. R. & DOLGACHEV, I. V., *Enriques Surfaces*. I. Progress in Mathematics, 76. Birkhäuser, Boston, MA, 1989.
- [15] DAHMANI, F., GUIARDEL, V. & OSIN, D., Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces. Preprint, 2011. [arXiv:1111.7048 \[math.GR\]](https://arxiv.org/abs/1111.7048).
- [16] DANILOV, V. I., Non-simplicity of the group of unimodular automorphisms of an affine plane. *Mat. Zametki*, 15 (1974), 289–293 (Russian); English translation in *Math. Notes*, 15 (1974), 165–167.
- [17] DELZANT, T., Sous-groupes distingués et quotients des groupes hyperboliques. *Duke Math. J.*, 83 (1996), 661–682.

- [18] DÉSERTI, J., Groupe de Cremona et dynamique complexe: une approche de la conjecture de Zimmer. *Int. Math. Res. Not.*, 2006 (2006), Art. ID 71701, 27 pp.
- [19] — Sur les automorphismes du groupe de Cremona. *Compos. Math.*, 142 (2006), 1459–1478.
- [20] — Le groupe de Cremona est hopfien. *C. R. Math. Acad. Sci. Paris*, 344 (2007), 153–156.
- [21] DILLER, J. & FAVRE, C., Dynamics of bimeromorphic maps of surfaces. *Amer. J. Math.*, 123 (2001), 1135–1169.
- [22] DOLGACHEV, I. V., On automorphisms of Enriques surfaces. *Invent. Math.*, 76 (1984), 163–177.
- [23] — Infinite Coxeter groups and automorphisms of algebraic surfaces, in *The Lefschetz Centennial Conference*, Part I (Mexico City, 1984), *Contemp. Math.*, 58, pp. 91–106. Amer. Math. Soc., Providence, RI, 1986.
- [24] — Reflection groups in algebraic geometry. *Bull. Amer. Math. Soc.*, 45 (2008), 1–60.
- [25] DOLGACHEV, I. V. & ORTLAND, D., Point sets in projective spaces and theta functions. *Astérisque*, 165 (1988).
- [26] DOLGACHEV, I. V. & ZHANG, D.-Q., Coble rational surfaces. *Amer. J. Math.*, 123 (2001), 79–114.
- [27] ENRIQUES, F., *Conferenze di Geometria*. Pongetti, Bologna, 1895.  
<http://enriques.mat.uniroma2.it/opere/950061f1.html>.
- [28] ERSHOV, M. & JAIKIN-ZAPIRAIN, A., Property (T) for noncommutative universal lattices. *Invent. Math.*, 179 (2010), 303–347.
- [29] FAVRE, C., Le groupe de Cremona et ses sous-groupes de type fini, in *Séminaire Bourbaki*, Vol. 2008/2009, Exp. No. 998, pp. 11–43. *Astérisque*, 332. Soc. Math. France, Paris, 2010.
- [30] FRÖHLICH, A. & TAYLOR, M. J., *Algebraic Number Theory*. Cambridge Studies in Advanced Mathematics, 27. Cambridge University Press, Cambridge, 1993.
- [31] FURTER, J.-P. & LAMY, S., Normal subgroup generated by a plane polynomial automorphism. *Transform. Groups*, 15 (2010), 577–610.
- [32] GHYS, É. & DE LA HARPE, P., *Sur les groupes hyperboliques d'après Mikhael Gromov* (Bern, 1988). *Progr. Math.*, 83. Birkhäuser, Boston, MA, 1990.
- [33] GIZATULLIN, M. H., Rational  $G$ -surfaces. *Izv. Akad. Nauk SSSR Ser. Mat.*, 44 (1980), 110–144, 239 (Russian); English translation in *Math. USSR-Izv.*, 16 (1981), 103–134.
- [34] — The decomposition, inertia and ramification groups in birational geometry, in *Algebraic Geometry and its Applications* (Yaroslavl', 1992), *Aspects Math.*, E25, pp. 39–45. Vieweg, Braunschweig, 1994.
- [35] GUEDJ, V., Propriétés ergodiques des applications rationnelles, in *Quelques aspects des systèmes dynamiques polynomiaux*, *Panor. Synthèses*, 30, pp. 97–202. Soc. Math. France, Paris, 2010.
- [36] HALPHÉN, G., Sur les courbes planes du sixième degré à neuf points doubles. *Bull. Soc. Math. France*, 10 (1882), 162–172.
- [37] HARTSHORNE, R., *Algebraic Geometry*. Graduate Texts in Mathematics, 52. Springer, New York, 1977.
- [38] HIGMAN, G., NEUMANN, B. H. & NEUMANN, H., Embedding theorems for groups. *J. London Math. Soc.*, 24 (1949), 247–254.
- [39] HUBBARD, J. H., *Teichmüller Theory and Applications to Geometry, Topology, and Dynamics*. Vol. 1. Matrix Editions, Ithaca, NY, 2006.
- [40] JUNG, H. W. E., Über ganze birationale Transformationen der Ebene. *J. Reine Angew. Math.*, 184 (1942), 161–174.
- [41] KOLLÁR, J., SMITH, K. E. & CORTI, A., *Rational and Nearly Rational Varieties*. Cambridge Studies in Advanced Mathematics, 92. Cambridge Univ. Press, Cambridge, 2004.

- [42] LAMY, S., Une preuve géométrique du théorème de Jung. *Enseign. Math.*, 48 (2002), 291–315.
- [43] — *Groupes de transformations birationnelles de surfaces*. Mémoire d’habilitation, Université Lyon 1, Lyon, 2010.
- [44] LAZARSFELD, R., *Positivity in Algebraic Geometry*. I. Classical setting: line bundles and linear series. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 48. Springer, Berlin–Heidelberg, 2004.
- [45] LYNDON, R. C. & SCHUPP, P. E., *Combinatorial Group Theory*. *Classics in Mathematics*. Springer, Berlin–Heidelberg, 2001.
- [46] MANIN, Y. I., *Cubic Forms*. North-Holland Mathematical Library, 4. North-Holland, Amsterdam, 1986.
- [47] MUMFORD, D., Hilbert’s fourteenth problem—the finite generation of subrings such as rings of invariants, in *Mathematical Developments Arising from Hilbert Problems* (De Kalb, IL, 1974), *Proc. Sympos. Pure Math.*, 28, pp. 431–444. Amer. Math. Soc., Providence, RI, 1976.
- [48] NGUYEN, D. D., *Composantes irréductibles des transformations de Cremona de degré d*. Thèse de Mathématique, Univ. Nice, Nice, 2009.
- [49] OL’SHANSKII, A. Y., SQ-universality of hyperbolic groups. *Mat. Sb.*, 186 (1995), 119–132 (Russian); English translation in *Sb. Math.*, 186 (1995), 1199–1211.
- [50] OSIN, D., Small cancellations over relatively hyperbolic groups and embedding theorems. *Ann. of Math.*, 172 (2010), 1–39.
- [51] SERRE, J.-P., Arbres, amalgames,  $SL_2$ . *Astérisque*, 46 (1977).
- [52] — Le groupe de Cremona et ses sous-groupes finis, in *Séminaire Bourbaki*, Vol. 2008/2009. Exp. No. 1000, pp. 75–100. *Astérisque*, 332. Soc. Math. France, Paris, 2010.

SERGE CANTAT  
 Université de Rennes I  
 Campus de Beaulieu  
 Bâtiment 22-23  
 FR-35042 Rennes Cedex  
 France

[serge.cantat@univ-rennes1.fr](mailto:serge.cantat@univ-rennes1.fr)

*Current address:*

Département de Mathématiques et Appl.  
 École Normale Supérieure  
 45 rue d’Ulm  
 FR-75230 Paris Cedex 5  
 France

STÉPHANE LAMY  
 Mathematics Institute  
 University of Warwick  
 Coventry CV4 7AL  
 U.K.

*Current address:*

Institut de Mathématiques de Toulouse  
 Université Paul Sabatier  
 118 route de Narbonne  
 FR-31062 Toulouse Cedex 9  
 France  
[slamy@math.univ-toulouse.fr](mailto:slamy@math.univ-toulouse.fr)

YVES DE CORNULIER  
 CNRS & Université Paris-Sud 11  
 Laboratoire de Mathématiques d’Orsay  
 Bâtiment 425  
 FR-91405 Orsay  
 France  
[yves.cornulier@math.u-psud.fr](mailto:yves.cornulier@math.u-psud.fr)

*Received July 7, 2010*

*Received in revised form May 20, 2011*