SOME EFFECTIVITY QUESTIONS FOR PLANE CREMONA TRANSFORMATIONS

N. I. SHEPHERD-BARRON

1 Introduction

The Cremona group $Cr_2(k)$ is the group of k-birational automorphisms of the projective plane $\mathbb{P}^2_k$ over a field k. As such, it is an object of algebraic geometry, but it is also of interest from the viewpoints of dynamics [M1] and group theory, including geometric group theory [CL]. This latter paper introduces a certain infinite-dimensional hyperbolic space $\mathcal{H} = \mathcal{H}(\mathbb{P}^2)$ on which $Cr_2(K)$ acts as a group of isometries and makes it clear that this action is an important tool for studying both $Cr_2(k)$ and its individual elements. However, given an algebraic description of a Cremona transformation g in terms of explicit rational functions, it is not always clear how to calculate anything about g that is relevant to any of these frameworks. For example, there is no known effective procedure for determining whether g is biregular (that is, acts as a biregular automorphism of some projective rational surface) or of calculating the *translation length* $L(g) = L(g_*)$ of the isometry g_* of $\mathcal{H}$ that is associated to g.

Our goal here is to make matters more nearly effective in the very particular case where g is a *special quadratic transformation*. That is, $g = \sigma\alpha$, where σ is the standard quadratic transformation given in terms of homogeneous co-ordinates x, y, z by $\sigma : (x, y, z) \mapsto (yz, xz, xy)$ and α is a linear map with $\alpha^2 = 1$. For this special class of Cremona transformations we give (Theorem 1.1 2 below) a simple and explicit lower bound, in terms of a constant amount of calculation, for $L(g)$. The upper bound $L(h) \leq \log D$ for a Cremona transformation h of degree D is well known [CL]; these bounds together give an arbitrarily fine estimate for $L(g)$. These bounds will also determine the type of g; that is, whether g is elliptic, parabolic or hyperbolic (the word loxodromic is also used for this last class). However, even in this special case, and even over a finite field, we still have no finite means of determining whether g is biregular.

Let P, Q, R denote the base points of σ and that Δ, the fundamental triangle given by $xyz = 0$, is the locus where σ is not biregular. We say that points $x_1, x_2, ..., x_n$ in $\mathbb{P}^2$ are *in general position* if they are all distinct and none of them, except for P, Q, R, lies on Δ. For any natural number n, let w_n denote the reduced word in σ, α that has length n and begins with α when reading from right to left. So, for example, $w_0 = 1, w_1 = \alpha, w_2 = \sigma\alpha$. Set $P_n = w_n(P)$, etc. So $P_0 = P$. We say that α, or g, is *in* (p, q, r)*-general position* if $P_0, ..., P_{p-1}, Q_0, ..., Q_{q-1}, R_0, ..., R_{r-1}$ are in general position. We abbreviate

(r,r,r)-general position to r-general position. Of course, r-general position implies s-general position for any $s \leq r$.

Theorem 1.1 *(1) Suppose that g is in (p,q,r)-general position and that $1/p + 1/q + 1/r < 0$. Then g is hyperbolic.*

(2) For $1 \gg \epsilon > 0$ define $p_0 = p_0(\epsilon) = \log(3/\epsilon)/\log(2-\epsilon)$. Then, if g is in p-general position and $p \geq p_0$, g is hyperbolic and

$$\log 2 - \epsilon < L(g) \leq \log 2.$$

Note [CL] that any transformation g of degree D has $L(g) \leq \log D$, while if g is very general then $L(g) = \log D$ and (this is the main result of [CL]) the normal closure $\langle\langle g^n \rangle\rangle$ in $Cr_2(k)$ of a sufficiently large power g^n of g does not contain g; in particular, it is not all of $Cr_2(k)$, and $Cr_2(k)$ is not simple. By using the ideas and methods of [CL] we can make this more explicit for some special quadratic transformations, as follows.

Theorem 1.2 *Suppose that $P_6 = P_7, Q_6 = Q_7, R_6 = R_7$ and that the 21 points $P_0, ..., P_6, ..., R_6$ are in general position. Assume also that none of P_1, Q_1, R_1 lies on the triangle Δ' defined by the equality of any two of the given homogeneous co-ordinates. Then $g \notin \langle\langle g^n \rangle\rangle$ for $n \gg 0$.*

These hypotheses can be realized over a finite field, since they impose three conditions on the 4-dimensional variety of involutions in PGL_3. In fact, over a finite field more is true.

Theorem 1.3 *Suppose that k is a finite field and that g is a hyperbolic element of $Cr_2(k)$. Then g is tight and $g \notin \langle\langle g^N \rangle\rangle$ for all sufficiently divisible N.*

However, we have no effective bound on N.

We can also give some analogous sufficient conditions for a special quadratic transformation g to be elliptic. Here is an example.

Theorem 1.4 *Suppose that $P_1 = P_2$, $Q_2 = Q_3$, $R_4 = R_5$ and that the ten points $P_0, P_1, ..., R_4$ are in general position. Then g has order 30.*

These results can be summarized by saying that there is a Coxeter–Dynkin diagram associated to the problem and if $P_0, ..., R_{r-1}$ are in general position then the diagram contains the standard tree $T_{p,q,r}$. Since $T_{2,3,5} = E_8$ the number 30 appears as the Coxeter number of E_8. This is a particular instantiation of the very old idea of relating groups of Cremona transformations to Weyl groups, and also of Steinberg's idea [S] of describing Coxeter elements using a description of the Coxeter diagram as a bipartite graph. This viewpoint has also been exploited by Blanc and Cantat [BC], who use an infinite group W_∞ that is something like a Coxeter group of type E_∞ to prove that, for any hyperbolic Cremona transformation g, the *spectral radius* $\lambda(g) = \exp L(g)$ lies in the closure of the set $\mathcal{T}$ of Salem numbers and $\lambda(g) \geq \lambda_{Lehmer}$, the *Lehmer number*, which is the smallest known Salem number. However, although we only consider special quadratic transformations, the Coxeter–Dynkin diagrams and groups that arise here are more general; they include those of all types $T_{p,q,r}$, where $E_{r+3} = T_{2,3,r}$.

2 Other background: dynamical systems and symmetric key cryptography

Hénon introduced certain complex quadratic plane Cremona transformations, now called Hénon maps, as models of (sections of) dynamical systems such as the Lorenz equations. They are of the form

$$f(x, y) = (ay + q(x), x),$$

where q is a quadratic polynomial and a is a non-zero scalar.

Then f might have sensitive dependence on initial conditions in this sense: even if initial points x_0 and y_0 are very close, their images $f^n(x_0)$ and $f^n(y_0)$ can be far apart for large values of n.

In the context where f is a smooth self-map of a compact manifold X this, when stated precisely in terms of Lyapunov exponents, turns out to be equivalent to the topological entropy $h(f)$ being strictly positive. (Recall that the topological entropy $h(g)$ of a self-map g of a compact metric space X is defined by

$$h(g) = \lim_{\epsilon \to 0} \lim_{n \to \infty} \log \left(\frac{1}{n} N(n, \epsilon) \right),$$

where $N(n, \epsilon)$ is the number of orbit segments of length n that are at least a distance ϵ apart. Gromov [G] extended this definition to cover correspondences, which include Cremona transformations, as well.) However, this definition involves two limits, so the questions arise of finding how large n must be taken, and how small ϵ, in order to estimate it in to a given accuracy in a bounded time.

On the other hand, over a finite field, especially one of characteristic 2, Feistel introduced the same kind of Cremona transformations, except that he took the parameter a to be $a = 1$ always and he did not demand that q be quadratic. These maps are also known as *round functions* and they are an essential element of *Feistel ciphers* such as DES, the Data Encryption Standard. The Advanced Encryption Standard, AES, uses different Cremona transformations, but otherwise both DES and AES have a similar structure.

Here is a toy model of DES ("toy" because it omits the key schedule): after Alice and Bob have established a key K (for example, by using some version of public key cryptography based, say, on elliptic curves), the key determines, according to a fixed public procedure that is part of the infrastructure of the algorithm, a round function f_K that is regarded as a non-linear but polynomial automorphism of some affine space $\mathbb{A}^n_k$ over a finite field k.

Then encryption of a message M is this: break M into blocks M_i, each of size n (that is, M_i is a k-point of $\mathbb{A}^n$) and then, for a fixed integer N that is also part of the infrastructure of the algorithm, apply the transformation f_K^N to the plaintext block M_i and transmit $f_K^N(M_i)$. Decryption is: apply $(f_K^{-1})^N$ to each block $f_K^N(M_i)$ that is received.

There is a parallel toy model of AES. First, some Galois twist σ of the standard non-linear birational involution $(x_0, ..., x_n) \mapsto (x_0^{-1}, ..., x_n^{-1})$ of $\mathbb{P}^n_k$ is given in advance and is public. Then the key K is used to construct a linear transformation of $\mathbb{P}^n_k$ and we set $f_K = \sigma \circ L_K$. Then the algorithm runs as for DES. (If the process ever encounters a base point, meaning that it is trying to invert 0, then it maps 0 to 0. So the iterated Cremona transformation is garbled. However, the security of the scheme does not reside in this garbling.) Since $f_K^{-1} = \sigma \circ L_K^{-1}$, decrypting is then, as with DES, as fast and cheap as encrypting (especially if L_K is an involution). However, in higher dimensions, a general Cremona transformations is of lower degree than its inverse, so slower and more expensive to invert.

Encryption should also mix up the points of projective space thoroughly and quickly; in other words, it is desirable that if x_0 and y_0 are distinct basepoints that are close, then the points $f_K^N(x_0)$ and $f_K^N(y_0)$ should be far apart for some large, but fixed, value of N. In other words, the round function should be sensitive to the parameters that define it, in the sense of (1) above. That is, in terms that might be over-simplified, over the complex numbers certain Cremona transformations serve as simple models of a real world that is known to be chaotic, while over a finite field the same Cremona transformations are used to create the appearance of chaos.

Moreover [M1], positive entropy does not exclude the existence of Siegel discs; Siegel discs are undesirable in cryptography because they are regions consisting of plaintexts that are close and that remain close after encryption. Of course, none of this is as significant as the fact that it is not clear what is the correct definition of entropy for points over finite fields, with distance defined to be the Hamming distance. On the other hand, the theorem of Gromov and Yomdin, that the entropy of an endomorphism g of a smooth projective variety X is the logarithm of the spectral radius of the action of g on the cohomology of X shows that $h(g)$ can be computed by reducing modulo p and counting points.

3 Hyperbolic space

Here we review the construction and basic properties of the infinite hyperbolic space $\mathcal{H} = \mathcal{H}_k = \mathcal{H}(\mathbb{P}^2_k)$ and the action of $Cr_2(k)$ on it. This is taken from [CL]; we repeat it in order to establish notation.

Let V be any smooth projective surface over the field k. Set $\mathcal{Z}(V)_{\mathbb{Z}} = \varinjlim_{Y \to V} \mathrm{NS}(Y)$, where the direct limit is taken over all blow-ups $Y \to V$, and $\mathcal{Z}(V) = \mathcal{Z}(V)_{\mathbb{Z}} \otimes_{\mathbb{Z}} \mathbb{R}$. There is a Lorentzian completion $\widetilde{\mathcal{Z}}(V)$ of $\mathcal{Z}(V)$, given by

$$\widetilde{\mathcal{Z}}(V) = \{\lambda + \sum \deg(P) n_P e_P | \lambda \in \mathrm{NS}(V)_{\mathbb{R}}, n_P \in \mathbb{R}, \sum \deg(P) n_P^2 < \infty\},$$

where e_P is the exceptional curve associated to the closed point P on some blow-up Y and $\deg(P)$ is the degree of the field extension $k(P)/k$. On $\widetilde{\mathcal{Z}}(V)$ there

is a Lorentzian inner product denoted by $(x.y)$. The hyperbolic space $\mathcal{H}(V)$ is one of the two connected components of the locus $\{x \in \widetilde{\mathcal{Z}}(V) | (x.x) = 1\}$. The distance on $\mathcal{H}(V)$ is denoted by d, so that $\cosh d(x,y) = (x.y)$. The isometries of $\mathcal{H}(V)$ are the continuous linear transformations of $\widetilde{\mathcal{Z}}(V)$ that preserve the inner product and the connected component above.

Given any blow–up $Y \to V$, there are natural isomorphisms $\mathcal{Z}(V)_{\mathbb{Z}} \to \mathcal{Z}(Y)_{\mathbb{Z}}$, $\mathcal{H}(V) \to \mathcal{H}(Y)$, etc., so that the group $\mathrm{Bir}(V)$ of birational automorphisms of V acts as a group of isometries of $\mathcal{H}(V)$ via $\gamma \mapsto \gamma_*$.

From now on, we take $V = \mathbb{P}^2_k$. We shall usually drop the subscript k from $\mathcal{H}_k$, but note that the formation of $\mathcal{H}_k$ is functorial in k.

Isometries ϕ of $\mathcal{H}$ are of three types: ϕ is *elliptic* if it has a fixed point in (the interior of) $\mathcal{H}$; ϕ is *parabolic* if it has a unique fixed point on the ideal boundary $\partial\mathcal{H}$ of $\mathcal{H}$ and no fixed point in $\mathcal{H}$; ϕ is *hyperbolic* if the lower bound $L(\phi) = \liminf d(x, \phi(x))$ is strictly positive and is attained in $\mathcal{H}$.

In this last case the set $\{x \in \mathcal{H} | d(x,\phi(x)) = M(\phi)\}$ is a geodesic in $\mathcal{H}$. It is the *axis* of ϕ and is denoted by $\mathrm{Ax}(\phi)$. It is the unique geodesic preserved by ϕ and its endpoints on $\partial\mathcal{H}$ are the unique fixed point on the closure $\overline{\mathcal{H}} = \mathcal{H} \cup \partial\mathcal{H}$. The quantity $L(\phi)$ is the *translation length* of ϕ. On the other hand, a parabolic isometry does not preserve any geodesic.

We record the following lemma.

Lemma 3.1 *If Π is a sub-vector space of $\widetilde{\mathcal{Z}}$ and δ an isometry of $\mathcal{H}$ that preserves V, then δ is hyperbolic on $\widetilde{\mathcal{Z}}$ if and only if it is hyperbolic on the hyperbolic space $\mathcal{H}(\Pi)$ associated to Π. If δ is hyperbolic, then its axis in $\mathcal{H}(\Pi)$ equals its axis in $\mathcal{H}$.*

Suppose that δ is a Cremona transformation, of degree D (in that δ is defined by a net of homogeneous polynomials of degree D). Then [CL] $L(\delta_*)$ equals the dynamical degree of δ, defined as $\lim_{n\to\infty}(\deg(f^n))^{1/n}$. Moreover, the entropy $h(\delta)$ of δ satisfies $h(\delta) \leq L(\delta_*)$ and equals it if δ is biregular on some smooth projective rational surface. Then δ is elliptic, etc., if δ_* is so, and δ is *biregular* if there is some rational surface X on which δ is a biregular automorphism. If δ is elliptic, then δ is biregular, and moreover there is a rational surface X on which δ is biregular and an integer $n > 0$ such that δ^n lies in the connected component of Aut_X.

If δ is parabolic, then there is a rational surface X and a fibration $X \to \mathbb{P}^1$ that is preserved by δ and whose generic fibre has genus at most 1.

Now suppose that δ is hyperbolic. Then the following are equivalent:
(1) δ is biregular;
(2) there is a finite dimensional Lorentzian subspace W of $\mathcal{Z}(\mathbb{P}^2)_{\mathbb{Q}}$, defined over $\mathbb{Q}$, that is preserved by δ_*;
(3) there is a finite dimensional Lorentzian subspace W of $\mathcal{Z}(\mathbb{P}^2)_{\mathbb{Q}}$, defined over $\mathbb{Q}$, such that $\mathrm{Ax}(\delta_*)$ is contained in $W_{\mathbb{R}}$;
(4) the end-points of $\mathrm{Ax}(\delta_*)$ on $\partial\mathcal{H}$ are defined over $\mathbb{Q}$.

Now fix $V = \mathbb{P}^2$. We denote by σ the standard quadratic involution $\sigma : (x, y, z) \mapsto (yz, xz, xy)$ and by α a linear involution. We put $g = \sigma\alpha$ and $\gamma = g_*$.

Lemma 3.2 *(1)* $\mathrm{Fix}(\sigma_*)$ *and* $\mathrm{Fix}(\alpha_*)$ *are hyperbolic subspaces of* $\mathcal{H}$.

(2) $\mathrm{Fix}(\sigma_*)$ *and* $\mathrm{Fix}(\alpha_*)$ *meet in* $\mathcal{H}$ *if and only if* γ *is elliptic if and only if there is a rational surface* X *on which both* α *and* σ *are biregular automorphisms and some power of* g *lies in* Aut^0_X. *That is, both* α *and* σ *act projectively on* X.

(3) $\mathrm{Fix}(\sigma_*)$ *and* $\mathrm{Fix}(\alpha_*)$ *are parallel if and only if they meet in a single ideal boundary point* $[\xi]$ *in* $\partial\mathcal{H}$. *In this case* $[\xi]$ *is defined over* $\mathbb{Q}$ *and is the class of a fibre in a genus* 1 *fibration that is preserved by both* α *and* σ.

(4) If $\mathrm{Fix}(\sigma_*)$ *and* $\mathrm{Fix}(\alpha_*)$ *are ultraparallel then there is a unique geodesic* Γ *perpendicular to both. This geodesic is preserved by both* α *and* σ *and is the axis of* γ.

PROOF: This is standard elementary hyperbolic geometry. □

Say that P, Q, R are the base points of σ and $Y = \mathrm{Bl}_{P,Q,R} V$, with exceptional curves e_P, e_Q, e_R. Note that σ is biregular on Y and α is biregular on V. Let ℓ denote the class of a line in V. Since $\mathcal{Z}(Y)_{\mathbb{Z}} \cong \mathcal{Z}(V)_{\mathbb{Z}}$, the lattice $\mathcal{Z}(V)_{\mathbb{Z}}$ has a $\mathbb{Z}$-basis $\{\ell\} \cup \{e_P, e_Q, e_R\} \cup E$, where E is the set of all exceptional curves e_x as x runs over all points of all blow-ups of Y.

From now on we shall not always be careful to distinguish between α and α_*, nor between σ and σ_*.

Lemma 3.3 α *permutes the set* $\{e_P, e_Q, e_R\} \cup E$ *and* σ *permutes* E.

PROOF: Immediate, from the facts that α is biregular on V and that σ is biregular on Y. □

Lemma 3.4 α *preserves* ℓ *and* σ *acts on the lattice* $\mathbb{Z}.\{\ell, e_P, e_Q, e_R\}$ *as the reflection* s_{v_0} *in the root* $v_0 = \ell - e_P - e_Q - e_R$.

PROOF: Calculation. □

4 Graphs and lattices

As above, σ and α are fixed. Our aim is to construct a bipartite Coxeter–Dynkin diagram H that depends upon the choice of α. Then $g = \sigma\alpha$ will act as something close to a Coxeter element in the corresponding Coxeter group.

We begin by constructing a bipartite graph $\widetilde{\Gamma} = \widetilde{\Gamma}_\alpha \cup \widetilde{\Gamma}_\sigma$, as follows.

The vertices of $\widetilde{\Gamma}_\sigma$ are v_0 and one representative $e_x - \sigma(e_x)$ of each non-zero pair $\pm(e_x - \sigma(e_x))$ as x runs over E. The vertices of $\widetilde{\Gamma}_\alpha$ are one representative $e_y - \alpha(e_y)$ of each non-zero pair $\pm(e_y - \alpha(e_y))$ as y runs over $E \cup \{e_P, e_Q, e_R\}$. Finally, if v lies in $\widetilde{\Gamma}_\alpha$ and $\pm v$ lies in $\widetilde{\Gamma}_\sigma$, then normalize through the process of replacing $\pm v$ by v.

We join two vertices in $\widetilde{\Gamma}$ by an edge of multiplicity equal to their intersection number, if that number is non-zero. If the intersection number is zero, then the corresponding vertices are left disjoint. So every edge has multiplicity ± 1 or 2. Define the valency of a vertex v to be the sum of the absolute values of the edges meeting v.

Lemma 4.1 *If v lies in the intersection $\widetilde{\Gamma}_\alpha \cap \widetilde{\Gamma}_\sigma$ then v is disjoint from all other vertices in $\widetilde{\Gamma}$.*

Now define $G_\alpha = \widetilde{\Gamma}_\alpha - (\widetilde{\Gamma}_\alpha \cap \widetilde{\Gamma}_\sigma)$ and define G_σ similarly. Put $G = G_\alpha \coprod G_\sigma$. Note that v_0 lies in G_σ.

Lemma 4.2 *G is a bipartite graph; there are no edges within either of G_α or G_σ; the vertex v_0 is of valency at most 3; every other vertex of G is of valency at most 2.*

PROOF: It is enough to notice that in $\widetilde{\Gamma}$ the vertex v_0 has valency at most 3 and that every other vertex has valency at most 2, so that deleting $\widetilde{\Gamma}_\alpha \cap \widetilde{\Gamma}_\sigma$ amounts to deleting those vertices that are joined to themselves and to no other vertex. Equivalently, deleting $\widetilde{\Gamma}_\alpha \cap \widetilde{\Gamma}_\sigma$ amounts to deleting all double bonds and the corresponding vertices. □

Now define $\widetilde{H}$ to be the connected component of G that contains v_0.

Lemma 4.3 *$\widetilde{H}$ is bipartite and is* either *a tree $T_{p,q,r}$ consisting of v_0 and three arms of lengths $p, q, r \leq \infty$ attached to v_0* or *the union $\Delta_{m,r}$ of a cycle of finite length m together with an arm of length $r \leq \infty$ attached to the cycle at v_0.*

PROOF: Immediate. □

Lemma 4.4 *m is even.*

PROOF: Suppose that $m = 2q + 1$ is odd. Then start at v_0 and travel around Δ in the two different directions through a distance q. We arrive at vertices $e_{sw(P)} - e_{w(P)}$ and $e_{sw(Q)} - e_{w(Q)}$ for some word w in the group W generated by σ, α such that $\ell(w) = q - 1$ and $\ell(w\alpha) = l(w) - 1$, s equals either σ or α and $(e_{sw(P)} - e_{w(P)}).(e_{sw(Q)} - e_{w(Q)}) = \pm 1$. Since $w(P) \neq w(Q)$ we have $w(P) = sw(Q)$. But then $e_{sw(P)} - e_{w(P)} = \pm(e_{sw(Q)} - e_{w(Q)})$, a contradiction. □

Write $m = 2n$. From $\widetilde{H}$, construct a graph H with the same vertices as $\widetilde{H}$, but where every edge except at most one has multiplicity $+1$, by starting at v_0 and proceeding either outwards along one arm at a time (in the case of $T_{p,q,r}$) or around the cycle and then along the arm (in the case of $\Delta_{2n,r}$) as follows: at each step, change an edge of multiplicity -1 into an edge of multiplicity $+1$ by replacing a vector v by its negative, $-v$. If H is of type $T_{p,q,r}$ then all its edges have multiplicity $+1$, and we write $H = T_{p,q,r}$; in the other case all edges, except possibly one edge that meets v_0 and lies in the cycle, are of multiplicity $+1$, and we write $H = \Delta^{\pm}_{2n,r}$ accordingly.

Lemma 4.5 *(1) H is either of type $T_{p,q,r}$ with $p,q,r \leq \infty$ or of type $\Delta^-_{2n,r}$. That is, $\Delta^+_{2n,r}$ cannot occur.*

(2) If p,q,r are finite and the points $P_0,...,P_{p-1},Q_0,...,Q_{q-1},R_0,...,R_{r-1}$ are in general position, then H contains the diagram $T_{p,q,r}$.

PROOF: We use the notation of the preceding proof, and in addition let s' denote the element of $\{\alpha,\sigma\}$ distinct from s. Then there are consecutive nodes in the cycle that are of the form $e_{sw(P)} - e_{w(P)}$, $e_{s'sw(P)} - e_{sw(P)}$ and $e_{sw(Q)} - e_{w(Q)}$, and also $e_{s'sw(P)} - e_{sw(P)} = e_{s'sw(Q)} - e_{sw(Q)}$. However, this is absurd.

The second part is an immediate observation. □

Denote by $L(H)$ the lattice on the vertices of H, with pairing determined by the edges of H and their multiplicities. Put $H_\alpha = H \cap G_\alpha$ and $H_\sigma = H \cap G_\sigma$, so that $v_0 \in H_\sigma$.

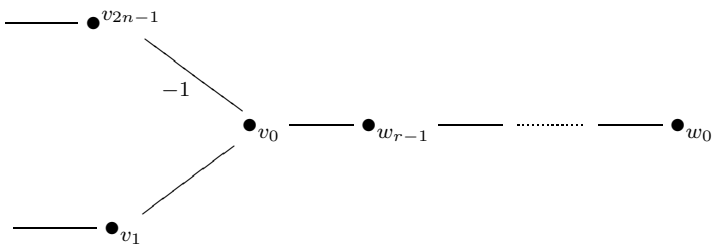
Lemma 4.6 *σ acts on $L(H)$ via the product $S = \prod_{v\in H_\sigma} s_v$ of the reflections in the vectors v in H_σ and α acts on $L(H)$ via the product $S = \prod_{w\in H_\alpha} s_w$ of the reflections in the vectors w in H_α.*

PROOF: Immediate observation. Notice that because all the reflections in each product commute with each other, the order in which they are taken is immaterial. The fact that each product contains infinitely many factors is also immaterial, since there are only finitely many terms in either product that act non-trivially on any given element of $L(H)$. □

Lemma 4.7 *$L(H)$ is degenerate only when H is of finite type $T_{p,q,r}$ with $1/p + 1/q + 1/r = 1$.*

PROOF: The $T_{p,q,r}$ case is well known. The degenerate possibilities correspond to various types of singular Kodaira–Néron fibres.

Suppose that $H = \Delta_{2n,r}$ and has diagram



(the vertices $v_0,\dots,v_{2n-1}$ are arranged in a cycle). Suppose also that $r \geq 1$ and that $\eta = \sum_0^{2n-1} a_i v_i + \sum_0^{r-1} b_j w_j$ is in the radical, so that $\eta.v_i = \eta.w_j = 0$ for all i,j. Put $w_r = v_0$.

By letting i run from 0 to $2n-1$, we see that a_i is a linear function of i; say $a_i = \lambda i + a_0$ for $i = 0,...,2n-1$. By letting j run from 0 to r, we see that $b_j = \mu j + b_0$ for $j = 0,...,r$.

Since $w_r = v_0$, we get $a_0 = \mu r + rb + 0$. Also, $b_1 = 2b_0$, so that

$$\mu = b_0,\ a_0 = \mu(r+1).$$

From $\eta.v_{2n-1} = 0$ we get $a_{2n-2} - 2a_{2n-1} - a_0 = 0$, so that $2n\lambda + a_0 = 0$.
From $\eta.v_0 = 0$ we get $a_1 - a_{2n-1} - 2a + 0 + b_{r-1} = 0$, so that

$$0 = -2(n-1)\lambda - 2a_0 + \mu r.$$

Then $\mu = -2\lambda(n+1)/2r$, so that $\mu = -2n\lambda/(r+1)$. Hence $(n+1)(r+1) = nr$, which is absurd.

Finally, the negative definiteness of $\Delta^-_{2n,0}$ follows from a comparison of it with the affine lattice $\Delta^+_{2n,0}$. □

There is an obvious natural homomorphism $\beta : L(H) \to \mathcal{Z}(V)_{\mathbb{Z}}$ of lattices.

Lemma 4.8 *β is injective unless H is of type $T_{p,q,r}$ with $1/p + 1/q + 1/r = 1$.*

PROOF: A lattice homomorphism whose domain is non-degenerate is injective. □

Let $T^{(\lambda)}_{p,q,r}$ and $\Delta^{-(\lambda)}_{2n,r}$ denote the lattices corresponding to the diagrams $T_{p,q,r}$ and $\Delta^-_{2n,r}$, but where each vertex v has $v^2 = -\lambda$ and the other intersection numbers are unchanged.

Lemma 4.9 *$T^{(2)}_{p,q,r}$ is hyperbolic if $1/p + 1/q + 1/r < 1$ and $\Delta^{-(2)}_{2n,r}$ is hyperbolic if $2/m + 1/r < 1$.*

PROOF: For $T^{(2)}_{p,q,r}$ this is well known. For $\Delta^{-(2)}_{2n,r}$, note that deleting v_0 leaves a negative definite lattice, while deleting the vertex opposite v_0 in the cycle leaves a $T_{n,n,r}$ diagram. □

Say that a lattice L is *affine* if L is degenerate, its radical $R(L)$ is of rank 1 and $L/R(L)$ is negative definite.

When $1/p + 1/q + 1/r \leq 1$, define $\mu(p,q,r)$ to be the minimum value of λ such that $T^{(\mu)}_{p,q,r}$ is affine. So, for example, $\mu(2,3,6) = \mu(2,4,4) = \mu(3,3,3) = 2$.

Lemma 4.10 *$\mu(p,q,r)$ is a strictly increasing function of each of p, q, r.*

PROOF: Say $s = \mu(p,q,r-1)$, so that $T^{(s)}_{p,q,r-1}$ is affine. Then there is a non-zero vector $\sum n_i v_i$ in the radical of $T^{(s)}_{p,q,r-1}$, and it is easy to see that we can take every n_i to be positive.

Say that w is the vertex adjoined in passing from $T_{p,q,r-1}$ to $T_{p,q,r}$. Then

$$\left(\sum n_i v_i + \epsilon w\right)^2 = 2n_2\epsilon - \epsilon^2 s > 0$$

for $0 < \epsilon \ll 1$, so that $T^{(s)}_{p,q,r}$ is Lorentzian. □

Write $\mu(p,p,p) = \mu_p$.

Lemma 4.11 *$\limsup \mu(p,q,r) = \lim_{p,q,r\to\infty} \mu(p,q,r) = 3/\sqrt{2}$ and $\mu_p \to (3/\sqrt{2})^-$ faster than $3/\sqrt{2} - 2y_0^p$ for any $y_0 \in (1/2, 1)$.*

PROOF: Say that $f_1, f_2, ...; g_1, g_2, ...; h_1, h_2, ...$ are the vertices of $T_{p-1,q-1,r-1}$, reading outwards along the arms from the central vertex v_0. We can suppose

that $p = q = r$ and then, by symmetry, that

$$\xi = b_p(0)v_0 + \sum_{i=1}^{\infty} b_p(i)(f_i + g_i + h_i)$$

is in the radical of $T^{(\mu_p)}_{p-1,q-1,r-1}$.

Set $b_p(0) = 3$ and $b_p(p) = 0$. Then

$$0 = \xi.v_0 = -3\mu_p + 3b_p(1),\ 0 = \xi.f_n = b_p(n-1) + b_p(n+1) - \mu_p.b_p(n)\ \forall\ n \geq 1.$$

Therefore

$$b_p(n) = -C_p e^{\alpha_p n} + D_p e^{-\alpha_p n}$$

for constants C_p, D_p, α_p, with $\alpha_p > 0$, and $\mu_p = e^{\alpha_p} + e^{-\alpha_p}$. Since $\mu_p > 2$, b_p is a decreasing function of n; moreover, $b_p(n) \geq 0$ for $n \leq p$. So as $p \to \infty$, we get $b_p(n) \to 3e^{-cn}$ for some c, and $\mu_p \to e^c + e^{-c}$. Since $b_p(1) = \mu_p$, we get $\mu_p \to (3/\sqrt{2})^-$.

For the speed of convergence, note that the boundary conditions give

$$m_p^p = \frac{2m_p - 1}{2 - m_p}$$

where $m_p = e^{2\alpha_p} > 1$. The same equation is satisfied by m_p^{-1}.

Lemma 4.12 *$m_p \to 2^-$, and does so faster than $2 - 4y_0^p$ for any $y_0 \in (1/2, 1)$. Similarly $m_p^{-1} \to (1/2)^+$ and does so faster than $1/2 + y_0^p$ for any $y_0 \in (1/2, 1)$.*

PROOF: Write $p = x$ and $m_p^{-1} = y = y(x)$. So $y^x = (2y-1)/(2-y)$. Then $y^x = y + \frac{y^2-1}{2-y}$, so that

$$1/2 < y = y^x + \frac{1 - y^2}{2 - y} < y^x + 1/2,$$

since $1/2 < y < 1$. That is, $1/2 < m_p^{-1} < (m_p^{-1})^p + 1/2$. We know that μ_p is an increasing function of p; since $\mu_p^2 = m_p + m_p^{-1} + 2$ and $m_p > 1$, it follows that m_p is also an increasing function of p and that m_p^{-1} is a decreasing function of p. So

$$1/2 < m_p^{-1} < (m_{p_0}^{-1})^p + 1/2$$

for any fixed $p_0 \leq p$. It follows that

$$2 > m_p > 2 - 4(m_p^{-1})^p > 2 - 4(m_{p_0}^{-1})^p$$

for all $p \geq p_0$. □

Taking square roots shows that $e^{-\alpha_p} \to (1/\sqrt{2})^+$ faster than $y_0^{p/2} + 1/\sqrt{2}$ for any $y_0 \in (1/2, 1)$. Write $z = 1/y = \nu_p^{-1}$; then

$$\sqrt{2} > z > \sqrt{2} - 2y^{p/2}$$

and the result follows from $\mu_p = y + z$. □

Corollary 4.13 *If* $1/p + 1/q + 1/r \leq 1$, *then* $2 \leq \mu(p,q,r) \leq 3/\sqrt{2} \approx 2.1213$.

Calculation reveals that $\mu(4,4,4) \approx 2.0743$, $\mu(5,5,5) = \sqrt{3+\sqrt{2}} \approx 2.101$, $\mu(6,6,6) \approx 2.112$. Of course, $\mu(3,3,3) = \mu(2,4,4) = \mu(2,3,6) = 2$.

5 Estimating the length of a special quadratic transformation

Suppose that $H = T_{p,q,r}$. (The case of $\Delta^-_{2n,r}$ will be treated later; it turns out to be very similar.) The adjacency matrix M of H is, according to the bipartite decomposition $H = H_\sigma \coprod H_\alpha$, of the form

$$M = \begin{bmatrix} 0 & {}^tC \\ C & 0 \end{bmatrix},$$

where each row and column of C contains at most three non-zero entries, and each non-zero entry is ± 1. Put $\mu = \mu(p,q,r)$, so that $2 \leq \mu \leq 3/\sqrt{2}$, and $B = -\mu + M$ is the Gram matrix of the lattice $L(H^{(\mu)})$. This lattice is affine, so that an eigenvector of B corresponding to the eigenvalue 0, again of B, generates the radical $R(L(H^{(\mu)})$. So the other eigenvalues of the symmetric matrix B are real and strictly negative. In particular, μ is the maximum eigenvalue of M, and is of multiplicity 1.

Suppose that $v = \begin{bmatrix} u \\ z \end{bmatrix}$ is an non-zero eigenvalue of M that belongs to μ. Then ${}^tCz = \mu u$ and $Cu = \mu z$.

Recall that $g = \sigma\alpha$ and notice that, from the bipartite description of H, α acts as the matrix $A = \begin{bmatrix} 1 & 0 \\ C & -1 \end{bmatrix}$, while σ acts as the matrix $S = \begin{bmatrix} -1 & {}^tC \\ 0 & 1 \end{bmatrix}$. Then

$$S\begin{bmatrix} u \\ 0 \end{bmatrix} = \begin{bmatrix} -u \\ 0 \end{bmatrix},\ S\begin{bmatrix} 0 \\ z \end{bmatrix} = \begin{bmatrix} \mu u \\ z \end{bmatrix},$$

$$A\begin{bmatrix} u \\ 0 \end{bmatrix} = \begin{bmatrix} u \\ \mu z \end{bmatrix},\ A\begin{bmatrix} 0 \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ -z \end{bmatrix}.$$

So S, A preserve the real 2-plane Π based by $\left(\begin{bmatrix} u \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ z \end{bmatrix}\right)$ and act on Π with respect to this basis as the matrices $\begin{bmatrix} -1 & \mu \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ \mu & -1 \end{bmatrix}$, respectively. The matrix of $\gamma = g_*$ is then $SA = \begin{bmatrix} \mu^2 - 1 & -\mu \\ \mu & -1 \end{bmatrix} = \delta^2$, where $\delta = \begin{bmatrix} \mu & -1 \\ 1 & 0 \end{bmatrix}$. Since $\operatorname{Tr}\delta = \mu > 2$, it follows that δ, and so γ, is hyperbolic.

Now suppose that m is the greater of the two eigenvalues of γ (both are real). Then $\mu^2 = m + m^{-1} + 2$, so that

$$\log m = 2\cosh^{-1}(\mu/2).$$

Moreover, m is, by construction, the maximum eigenvalue of a Coxeter element of the Coxeter group $W(T_{p,q,r})$ and so [M2] is a Salem number.

Corollary 5.1 *(1) If p, q, r are finite and $1/p + 1/q + 1/r < 1$, then g is a hyperbolic element of $Cr_2(k)$ and*

$$0 < L(g) = \log m < 2\cosh^{-1}(3/2\sqrt{2}) = \log 2.$$

(2) If $p = q = r \geq 3$ then $L(g) \geq \log m_p$, where $1 < m_p \leq 2$ and

$$m_p^p = \frac{2m_p - 1}{2 - m_p}.$$

In particular, $L(g) \to (\log 2)^-$ as $p \to \infty$ and the convergence is fast, in the sense that $\exp L(g) \to 2^-$ faster than $2 - 4y_0^p$ for any $y_0 \in (1/2, 1)$.

PROOF: Note that δ preserves the Lorentzian quadratic form Q on Π defined by the matrix $Q = \begin{bmatrix} 1 & -\mu/2 \\ -\mu/2 & 1 \end{bmatrix}$. The vector $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ satisfies $(x.x) = 1$ and $(x.\delta(x)) = \mu/2$, so that $d(x, \delta(x)) = \cosh^{-1}(\mu/2)$. That is, $L(\delta) = \cosh^{-1}(\mu/2)$, so that, by Lemma 3.1, γ is hyperbolic and $L(\gamma) = 2\cosh^{-1}(\mu/2) = \log m$.

For the second part, note that, in the notation of Lemma 4.12 and the formula immediately preceding it, $L(g) = \log m_p = 2\alpha_p$, since $\mu_p^2 = m_p + m_p^{-1} + 2$. That lemma states that m_p converges fast to 2 from below, and we are done. □

Proposition 5.2 *If any or all of p, q, r are infinite, then again g is hyperbolic and $L(g) = \log 2 \approx 0.6931$.*

PROOF: Even if p, q, r are infinite, the adjacency matrix M still represents a bounded linear operator, since its rows and columns have at most 3 non-zero entries, all ± 1, and its spectrum is $3/\sqrt{2}$. □

Of course, this is not an effective result, because we have no effective means of deciding whether any of p, q, r is infinite.

Now suppose that $H = \Delta_{2n,r}^-$, and assume that $2/n + 1/r < 1$.

Proposition 5.3 *g is hyperbolic and $\lim_{n,r\to\infty} L(g) = \log 2$. If $n, r \geq 5$, then*

$$0.6329 \approx 2\cosh^{-1}(\sqrt{3 + \sqrt{2}}/2) \leq L(g).$$

PROOF: Deleting the vertex opposite v_0 in the cycle leaves a diagram of type $T_{n,n,r}$. Now the previous arguments apply. □

Finally we give the proof of Theorem 1.1. Recall the statement:

Theorem 5.4 *(= Theorem 1.1)*

(1) Suppose that the twelve points $P_0, P_1, P_2, P_3, Q_0, ..., R_3$ are in general position. Then g is hyperbolic and

$$0.5435 \approx \log m \leq L(g) \leq \log 2 \approx 0.6932,$$

where $m = m_4$ is the maximal real zero of the quartic polynomial $m^4 - m^3 - m^2 - m + 1$.

(2) For $1 \gg \epsilon > 0$ define $p_0 = p_0(\epsilon) = \log(3/\epsilon)\log(2-\epsilon)$. Then, if the $3p$ points $P_0, ..., P_{p-1}, ..., R_{p-1}$ are in general position, the Cremona transformation $g = \sigma\alpha$ is hyperbolic and

$$\log 2 - \epsilon_0 < L(g) \leq \log 2.$$

PROOF: For any p, consider the diagram $T_{p,p,p}$. Its bipartite nature means that we can find a Coxeter element $\gamma = \gamma_p$ of $W(T_{p,p,p})$ that is exhibited as a product $\gamma = SA$ of two involutions, each represented as a 2×2 matrix that is computed from the adjacency matrix of the diagram, exactly as above. Then compute μ_p as above, so that

$$\mu_p^2 = m_p + m_p^{-1} + 2.$$

We define $p(\epsilon_0)$ to be the least value of p such that

$$\log 2 - \epsilon_0 < \log m_p.$$

Now Theorem 1.1 follows from Corollary 5.1. □

6 Constructing Cremona transformations with prescribed properties

Here we prove Theorems 1.4 and 1.2.

Fix (finite) integers p, q, r. Then we can force H to be equal to the finite tree $T_{p,q,r}$ by putting stringent conditions on the linear involution α, as follows. Assume that $P_0, ..., P_{p-1}, ..., R_{r-1}$ are in general position and that $P_p = P_{p-1}, Q_q = Q_{q-1}, R_r = R_{r-1}$. Note that, since σ has only finitely many fixed points while α has a line of fixed points, these are two conditions on α for each of p, q, r that is odd, and one condition on α for each that is even. As before, ℓ is the class of a line in $\mathbb{P}^2$.

Lemma 6.1 *H is a finite diagram of type $T_{p,q,r}$ and its vertices are $v_0 = \ell - e_{P_0} - e_{Q_0} - e_{R_0}$; $e_{P_0} - e_{P_1}, ..., e_{P_{p-2}} - e_{P_{p-1}}$; $e_{Q_0} - e_{Q_1}, ..., e_{Q_{q-2}} - e_{Q_{q-1}}$; $e_{R_0} - e_{R_1}, ..., e_{R_{r-2}} - e_{R_{r-1}}$.*

PROOF: Immediate. □

Lemma 6.2 *Both α and σ are biregular on the blow-up Y of $\mathbb{P}^2$ at the $p+q+r$ points $P_0, ..., R_{r-1}$.*

PROOF: Also immediate. □

Of course, g is then also biregular on Y.

Lemma 6.3 *The orthogonal complement $L(H)^\perp$ of $L(H)$ in $\mathrm{NS}(Y)$ has a $\mathbb{Z}$-basis (x_P, x_Q, x_R) given by $x_P = \ell - \sum e_{P_i}$, etc. Each of α, σ acts trivially on $L(H)^\perp$.*

Theorem 6.4 *(= 1.4) g acts on $\mathrm{NS}(Y)$ as a Coxeter element in the Weyl group of the lattice $T^{(2)}_{p,q,r}$. In particular, if $(p,q,r) = (2,3,5)$ then g has order 30.*

PROOF: This follows from the discussion in the previous sections. □

Remark: Moreover, if $(p,q,r) = (2,3,7)$ then g is hyperbolic and its length $L(g)$ equals $\log \lambda_{Lehmer}$ where λ_{Lehmer}, the *Lehmer number*, is the smallest known Salem number. Moreover, this value of $L(g)$ is minimal over all hyperbolic elements of $Cr_2(k)$ [BC].

7 Rigidity and tightness over finite fields

We refer to [CL] for the crucial notions of *rigidity* and *tightness* for elements of $Cr_2(k)$. Instead of repeating the definitions here we explain what needs to be proved.

Now suppose that the ground field k is finite.

Proposition 7.1 *For every point $z \in \mathcal{H}$ and every $r > 0$, the set of $f \in Cr_2(k)$ such that $d(z, f(z)) < r$ is finite. In particular, $Cr_2(k)$ acts properly discontinuously on $\mathcal{H}$.*

PROOF: Fix z, r and suppose that $d(z, f(z)) < r$. Then

$$d(\ell, f(\ell)) \leq d(\ell, z) + d(z, f(z)) + d(f(z), f(\ell)),$$

so that $d(\ell, f(\ell)) < 2d(\ell, z) + r$. Since $\cosh d(\ell, f(\ell)) = (\ell.f(\ell)) = \deg(f)$, it follows that $\deg(f)$ is bounded. Because k is finite, we are done. □

Now fix a hyperbolic element g of $Cr_2(k)$. We know that such g exist; for example, the special quadratic transformations.

Theorem 7.2 *The axis $\mathrm{Ax}(g)$ is rigid.*

PROOF: Suppose that $\mathrm{Ax}(g)$ is not rigid. Then, by Proposition 3.3 of [CL], for all bounded segments Σ of $\mathrm{Ax}(g)$ and all $\epsilon > 0$ there exists $h \in Cr_2(k)$ such that $d(y, h(y)) \leq \epsilon$ for all $y \in \Sigma$ and h does not preserve $\mathrm{Ax}(g)$.

Take t to be the point on $\mathrm{Ax}(g)$ that is closest to the class ℓ; say that $d(\ell, t) = a$. Take any segment Σ containing t, and let $\epsilon > 0$. Take $h \in Cr_2(k)$ as given above. Then, by the triangle inequality,

$$d(\ell, h(\ell)) \leq d(\ell, t) + d(t, h(t)) + d(h(t), h(\ell)),$$

so that $d(\ell, h(\ell)) < 2a + \epsilon$. Then

$$(\ell.h(\ell)) = \cosh d(\ell, h(\ell)) < \cosh(2a + \epsilon).$$

Since $(\ell.h(\ell))$ is just $\deg(h)$, it follows that $\deg(h)$ is bounded. Since k is finite, the set of all such h is finite, so that (pigeon-hole principle) there exists $h \in Cr_2(k)$ that works for arbitrarily long Σ and arbitrarily small ϵ. Then $h(s) = s$ for all $s \in \mathrm{Ax}(g)$; however, this contradicts the fact that h does not preserve $\mathrm{Ax}(g)$. □

Let $Fix(\mathrm{Ax}(g))$ denote the set of elements f of $Cr_2(k)$ that fix every point on $\mathrm{Ax}(g)$.

Lemma 7.3 *$Fix(\mathrm{Ax}(g))$ is finite.*

PROOF: The preceding argument with the triangle inequality shows that $(\ell.f(\ell))$ is bounded independently of $f \in Fix(\mathrm{Ax}(g))$, and now finiteness follows again. □

Let M denote the size of $Fix(\mathrm{Ax}(g))$ and $N = M!$.

Theorem 7.4 *g^N is tight.*

PROOF: We know that $\mathrm{Ax}(g)$ is rigid. Suppose that $h \in Cr_2(k)$ preserves $\mathrm{Ax}(g)$; we must show that $hg^Nh^{-1} = g^{\pm N}$.

Assume first that h preserves the orientation of $\mathrm{Ax}(g)$. Then $hg^rh^{-1}g^{-r}$ lies in $Fix(\mathrm{Ax}(g))$ for every r, so that there are distinct integers r, s between 1 and M such that $hg^rh^{-1}g^{-r} = hg^sh^{-1}g^{-s}$. Then $hg^{r-s}h^{-1} = g^{r-s}$, so that $hg^Nh^{-1} = g^N$.

If h reverses the orientation of $\mathrm{Ax}(g)$ then $hg^rh^{-1}g^r \in Fix(\mathrm{Ax}(g))$ for every r, and the same argument shows that $hg^Nh^{-1} = g^{-N}$. □

Theorem 7.5 *(= Theorem 1.3) If k is a finite field and g is a hyperbolic element of $Cr_2(k)$ then g does not lie in the normal closure $\langle\langle g^N\rangle\rangle$ whenever N is sufficiently divisible. In particular, $\langle\langle g^N\rangle\rangle$ is a non-trivial normal subgroup of $Cr_2(k)$.*

PROOF: This follows from tightness, exactly as in Theorems C and 2.10 of [CL]. (Note the typo in Theorem 2.10 of *loc. cit.*: the phrase "either h is a conjugate of g, or..." should read "either h is a conjugate of $g^{\pm 1}$, or...".) □

In particular, every hyperbolic element of $Cr_2(k)$ lies outside some non-trivial normal subgroup of $Cr_2(k)$, and we have exhibited hyperbolic special quadratic transformations over sufficiently large finite fields.

8 Rigidity, tightness and non-normal subgroups over an arbitrary field

Recall that ℓ is the class of a line in $\mathbb{P}^2_k$.

We begin by recalling some elementary hyperbolic geometry. A *skew right trapezium* is a quadrilateral $xyzt$ in hyperbolic space with a *base* $[z,t]$ of length b, a *summit* $[x,y]$ of length s and *sides* $[y,z]$, $[x,t]$, of respective lengths a and a', both of which are orthogonal to the base (at the points z, t). When $a = a'$ this is a *skew Saccheri quadrilateral*.

Lemma 8.1 *(1) In a skew right trapezium we have*

$$\cosh a \cosh a'(\cosh b - 1) \leq \cosh s - \cosh(a - a') \leq \cosh s - 1$$

(so, in particular, $b \leq s$).

(2) Equality holds if and only if the trapezium is planar and $a = a'$.

(3) In a skew Saccheri quadrilateral we have

$$\frac{\cosh s + 1}{\cosh b + 1} \leq \cosh^2 a \leq \frac{\cosh s - 1}{\cosh b - 1}.$$

PROOF: Draw the diagonal $d = [x, z]$. Denote the angles $\angle xzy$ and $\angle xzt$ by θ, ϕ, respectively. Note that $\theta + \phi \geq \pi/2$, by the triangle inequality on the sphere S^2 of curvature $+1$, and $\phi \leq \pi/2$ and $\theta \leq \pi$. So $-\pi/2 \leq \pi - \theta \leq \phi \leq \pi/2$ and therefore $\sin(\pi/2 - \theta) \leq \sin\phi$. The trigonometry of the hyperbolic triangle $\triangle xyz$ gives

$$\sinh a \sinh d \cos\theta = \cosh a \cosh d - \cosh s.$$

The trigonometry of $\triangle xtz$ gives $\sin\phi = \sinh a'/\sinh d$, so that $\sinh d \cos\theta \leq \sinh a'$. The same triangle also gives $\cosh d = \cosh a' \cosh b$. So

$$\cosh a \cosh d - \cosh s = \sinh a \sinh d \cos\theta \leq \sinh a \sinh a'$$

and then substituting $\cosh d = \cosh a' \cosh b$ and using the identity $\sinh a \sinh a' - \cosh a \cosh a' = -\cosh(a - a')$ gives

$$\cosh a \cosh a'(\cosh b - 1) \leq \cosh s - \cosh(a - a').$$

Since $\cosh(a - a') \geq 1$ we have the inequality of the lemma. If equality holds, then $\cosh(a - a') = 1$ and $\cos\theta = \sin\phi$, and we are done, except for the first inequality of the last part.

For this, let w be the midpoint of the base $[z, t]$ and draw the segments $[y, w]$, $[x, w]$. By symmetry, they have the same length, say f. By the triangle inequality $s \leq 2f$, so that $\cosh s \leq \cosh 2f$ and $\cosh s + 1 \leq 2\cosh^2 f$. The trigonometry of $\triangle yzw$ shows that $\cosh f = \cosh a \cosh b/2$, so that

$$\cosh s + 1 \leq 2\cosh^2 a \left(\frac{\cosh b + 1}{2}\right) = \cosh^2 a(\cosh b + 1).$$

□

We fix a hyperbolic element g of $Cr_2(k)$ of degree D. Say that its translation length $L(g) = L$, that t is the point on $\mathrm{Ax}(g)$ closest to ℓ and that $a = d(\ell, t) = d(\ell, \mathrm{Ax}(g))$. We know that $L \leq \log D$; write $L = \log D - \eta_1$. Assume that for some given $n \geq 1$ we know that $\deg(g^n) = D^n$.

Now consider the skew Saccheri quadrilateral with vertices $x = \ell$, $y = g^n(\ell)$, t the point just described and $z = g^n(t)$. Say $b = nL$ and $s = \cosh^{-1}(\ell.g^n(\ell))$,

so that the notation matches that of Lemma 8.1 above. Put $\psi = \exp(\eta)$.Then Lemma 8.1 gives

$$2\frac{\psi^n(1+D^{-n})}{(1+\psi^n D^{-n})^2} \leq \cosh^2 a \leq 2\frac{\psi^n(1-D^{-n})}{(1-\psi^n D^{-n})^2},$$

which we abbreviate to

$$2A \leq \cosh^2 a \leq 2B.$$

Since $\cosh 2a = 2\cosh^2 a - 1$, this can also be written as

$$4A - 1 \leq \cosh 2a \leq 4B - 1.$$

Now suppose that $g = \sigma\alpha$ as before, so that $D = 2$, and that g is in n-general position for some $n \geq 4$. Then g is hyperbolic, $\deg(g^n) = 2^n$ and

$$\log 2 \geq L = L(g) \geq \log m_n = 2\cosh^{-1}(\mu_n/2),$$

with m_n and μ_n as before.

Define functions A, B, Y, Z of two variables by

$$A(n,\psi) = \frac{\psi^n(1+2^{-n})}{(1+\psi^n 2^{-n})^2},$$

$$B(n,\psi) = \frac{\psi^n(1-2^{-n})}{(1-\psi^n 2^{-n})^2},$$

$$Y(n,\psi) = 4\left(A(n,\psi)(2\psi^{-2} + \psi^2/8) - 1\right),$$

$$Z(n,\psi) = 4\left(B(n,\psi)(2\psi^{-2} + \psi^2/8) - 1\right).$$

Lemma 8.2 *(1) If $n \geq 4$ then A, B, Y, Z are increasing functions of ψ for $\psi \in [1, 2)$.*

(2) For $1 \leq \psi \ll 2$, B and Z are decreasing functions of n.

PROOF: Calculate logarithmic partial derivatives. □

So, since $\psi \leq \psi_n$, where as before ψ_n satisfies $\psi_n^n = 2^{1+n\frac{(\psi_n-1)}{(4-\psi_n)}}$, we have

$$2.\frac{1}{1+2^{-n}} \leq \cosh^2 a \leq 2B(n,\psi_n).$$

Also, $Y \geq 4\left(\frac{1}{1+2^{-n}}.2\frac{1}{8} - 1\right)$, so that $Y \geq 4.2$ for $n \geq 5$. Moreover, calculation reveals that for $n = 8$ we have $Z(n,\psi_n) \approx 4.869$, so that $Z < 4.9$ for all $n \geq 8$.

Proposition 8.3 *If $n = 8$ then $\mathrm{Ax}(g)$ is rigid.*

PROOF: The basic structure of the argument is taken from the proof of Prop. 5.7 of [CL].

Suppose that $\mathrm{Ax}(g)$ is not rigid; then (Proposition 3.3 of [CL]) for all $\epsilon > 0$ (however small) and for all $\Theta > 0$ (however large) there exists $h \in Cr_2(k)$ such that the ϵ-neighbourhood of $\mathrm{Ax}(g)$ and the ϵ-neighbourhood of $h(\mathrm{Ax}(g))$ intersect in something of diameter at least Θ, while at the same time $h(\mathrm{Ax}(g))$ is distinct from $\mathrm{Ax}(g)$. In particular, for all $\epsilon > 0$ there exists $h \in Cr_2(k)$ such that $d(h(x), x) < \epsilon$ for $x = t, g^{\pm}(t)$. However, we make no use of the assumption that $h(\mathrm{Ax}(g)) \neq \mathrm{Ax}(g)$.

First, calculation reveals that for $n = 8$ we have $\psi_8 \approx 1.00614$, $B \approx 1.05477$, $Z \approx 4.86889$

Lemma 8.4 $\deg(h) \leq 3$.

PROOF: By the triangle inequality

$$d(h(\ell), \ell) \leq d(h(\ell), h(t)) + d(h(t), t) + d(t, \ell) < 2d(\ell, t) + \epsilon = 2a + \epsilon.$$

Define ϵ' by $\cosh(2a + \epsilon) = \cosh 2a + \epsilon'$, so that

$$(h(\ell).\ell) < \cosh 2a + \epsilon' = 2\cosh^2 a - 1 + \epsilon'.$$

Since

$$\cosh^2 a \leq 2B = 2\psi_8^8(1 - 2^{-8})/(1 - \psi_8^8 2^{-8}) \approx 2.1095$$

it follows that $(h(\ell).\ell) <\leq 3.2189 + \epsilon' < 4$. □

Similarly $\deg(ghg^{-1}) \leq 3$ and $\deg(g^{-1}hg) \leq 3$.

Assume that $\deg(h) = 3$. We shall derive a contradiction to this when ϵ is sufficiently small.

Extend the geodesic segment $[\ell, t]$ to a point s so that t is the midpoint of the segment $[\ell, s]$. Say that $\delta = d(h(\ell), s)$.

Lemma 8.5 $\cosh \delta \leq 2\cosh a \cosh(a + \epsilon) - 3$.

PROOF: By the triangle inequality, if we write $d(t, h(\ell)) = a + \epsilon''$, then $\epsilon'' \leq \epsilon$. Say that α is the angle of the triangle formed by $t, \ell, h(\ell)$ at the vertex ℓ; then

$$\cosh(a + \epsilon'') = 3\cosh a - \sqrt{8}\sinh a \cos\alpha$$

and

$$\cosh \delta = 3\cosh 2a - \sqrt{8}\sinh 2a \cos\alpha.$$

This reduces to

$$\cosh \delta = 2\cosh a \cosh(a + \epsilon'') - 3 \leq 2\cosh a \cosh(a + \epsilon) - 3.$$

□

Consider the configuration of points

$$\ell, t, s, h(\ell), g^2(\ell), g^2(t).$$

Let y denote the length of the segment $[t, g^2(\ell)]$ and ν the angle of the triangle $\triangle(g^2(\ell), \ell, t)$ at the vertex ℓ; note that ν is also the angle of the triangle $\triangle(g^2(\ell), \ell, s)$ at the vertex ℓ. The trigonometry of the triangles $\triangle(g^2(\ell), g^2(t), t)$ and $\triangle(g^2(\ell), \ell, t)$ gives

$$\cosh y = \cosh a \cosh(2(\log 2 - \eta))$$

and

$$\cosh y = 4 \cosh a - \sinh a \sinh(\cosh^{-1}(4)) \cos \nu,$$

respectively, so that

$$\cos \nu = \cosh a(4 - \cosh(2(\log 2 - \eta))/ \sinh a\sqrt{2^4 - 1}.$$

The trigonometry of $\triangle(g^2(\ell), \ell, s)$ then yields

$$(s.g^2(\ell)) = 2 \cosh^2 a \cosh(2(\log 2 - \eta)) - 4.$$

We shall derive a contradiction by showing that $(s.g^2(\ell))$ is far from any integer, so that when ϵ is sufficiently small $(h(\ell).g^2(\ell))$ is not an integer, which is certainly absurd.

We have

$$4A^2 \cosh(\log 4 - 2\eta) - 4 < (s.g^2(\ell)) < 4B^2 \cosh(\log 4 - 2\eta) - 4.$$

That is, $Y < (s.g^2(\ell)) < Z$, so that, for $n = 8$,

$$4.2 < Y < (s.g^2(\ell)) < Z < 4.9.$$

At the same time $(h(\ell).g^2(\ell))$ is an integer and is close, in terms of ϵ, to $(s.g^2(\ell))$. This is impossible and so $\deg(h) \leq 2$. The same argument shows that $\deg(ghg^{-1})$ and $\deg(g^{-1}hg)$ are both at most 2.

Now assume that $\deg(h) = 2$**.** We shall derive a contradiction in this situation too.

As before, there is a point s such that $d(s, t) = a$ and s is arbitrarily close, in terms of ϵ, to $h(\ell)$. Take w to be the midpoint of the segment $[t, g^m(t)]$ for $m = \pm 1$ or ± 2. Then the right-angled triangles $\triangle(t, s, w)$ and $\triangle(t, g^m(t), g^m(\ell)))$ are congruent; let z denote the common length of their hypotenuses. Then

$$(s.g^m(\ell)) \leq \cosh 2z = 2 \cosh^2 a \cosh^2 \left(\frac{1}{2}|m|(\log 2 - \eta)\right) - 1$$

which leads to $(s.g^m(\ell)) \leq B(2^m + 2^{-m} + 2) - 1$, so that $(s.g^{\pm 1}(\ell)) \leq 4\frac{1}{2}B - 1 < 4$ and $(s.g^{\pm 2}(\ell)) \leq 6\frac{1}{4}B - 1 < 6$. Hence $h(\ell).g^{\pm 1}(\ell)) \leq 3$ and $h(\ell).g^{\pm 2}(\ell)) \leq 5$.

We can write $h(\ell) = 2\ell - E_1 - E_2 - E_3$. We also know that $g(\ell) = 2\ell - F_1 - F_2 - F_3$ and $g^{-1}(\ell) = 2\ell - F_1' - F_2' - F_3'$, so that

$$g^2(\ell) = 4\ell - 2\sum F_i - \sum g(F_i).$$

Moreover, by our general position assumption on g, the classes $F_1, ..., g(F_3)$ are distinct exceptional curves. Since $(h(\ell).g(\ell)) \leq 3$ it follows that $\{E_1, E_2, E_3\} \cap \{F_1, F_2, F_3\}$ has at least one element. Similarly the inequality $(h(\ell).g^2(\ell)) \leq 5$ shows that $\{E_1, E_2, E_3\} \cap S$ has at least two elements, where

$$S = \{F_1, F_2, F_3, g(F_1), g(F_2), g(F_3)\}.$$

Again similarly, $(h(\ell).g^{-2}(\ell)) \leq 5$ shows that $\{E_1, E_2, E_3\} \cap S'$ has at least two elements, where

$$S' = \{F_1', F_2', F_3', g^{-1}(F_1'), g^{-1}(F_2'), g^{-1}(F_3')\}.$$

It follows that $S \cap S'$ is non-empty, which contradicts what we have assumed about g. So $\deg h = 2$ is impossible, so that $\deg h = 1$. Similarly $\deg(ghg^{-1}) = \deg(g^{-1}hg) = 1$. Now the argument of [CL], p. 48, applies to show that $h = 1$. (They denote by P what is denoted here by t.) This completes the proof of Proposition 8.3. □

Proposition 8.6 *If $n = 8$ then g is tight.*

PROOF: We must show that if $f \in Cr_2(k)$ preserves $\mathrm{Ax}(g)$, then $fgf^{-1} = g^{\pm}$.

Suppose first that f preserves the orientation of $\mathrm{Ax}(g)$. Then put $h = fgf^{-1}g^{-1}$, so that h fixes every point on $\mathrm{Ax}(g)$. The proof of rigidity given above shows that then $h = 1$, as required.

If f reverses the the orientation of $\mathrm{Ax}(g)$ take $h = fgf^{-1}g$ instead and then use the same argument. □

Corollary 8.7 *Suppose that $g = \sigma\alpha$ is in 8-general position. Then for some integer m the normal subgroup $\langle\langle g^{rm}\rangle\rangle$ of $Cr_2(k)$ generated by g^m is a non-trivial normal subgroup of $Cr_2(k)$.*

PROOF: This is an immediate application of Theorem C of [CL] to our situation. □

9 A lower bound for parabolic elements

In [BC] Blanc and Cantat prove the lower bound $L(g) \geq \log \lambda_{Lehmer}$ for hyperbolic elements g of $Cr_2(k)$, for any field k. Here $\lambda_{Lehmer} = \lambda$ is *Lehmer's number*, the smallest known Salem number. In particular, $d(x, g(x)) \geq \log \lambda$ for hyperbolic g

and any $x \in \mathcal{H}(\mathbb{P}^2)$. In this section we prove a result for parabolic elements that can be regarded as an analogue of this.

Recall that if v_0 is a primitive isotropic vector in the lattice $\mathbb{L}$ and z is a positive real number, then there is a *horosphere* in $\mathcal{H}(\mathbb{P}^2)$ defined by $(x.v_0) = z$ and a *horoball* $B(v_0, z)$ defined by $(x.v_0) < z$. These horospheres and horoballs are preserved by those isometries that preserve v_0.

Theorem 9.1 *Suppose that f is a parabolic element of $Cr_2(k)$, that $x \in \mathcal{H}(\mathbb{P}^2)$ and that $d(x, f(x)) < \log \lambda$. Then there is a unique primitive isotropic vector $v_0 \in \Lambda$ such that x lies in the horoball $B(v_0, \lambda^{1/2} - \lambda^{-1/2})$. Moreover, v_0 is preserved by f.*

PROOF: The existence of v_0 is, by now, a well known result about Cremona transformations: a parabolic Cremona transformation is biregular and preserves a fibration by curves of genus at most 1. Take v_0 to be the class of a fibre in such a fibration. So there is a primitive isotropic element v_0 of the infinite Lorentzian lattice $\mathbb{L}$ such that f preserves v_0. With respect to the filtration

$$0 \subset \mathbb{Z}v_0 \subset v_0^{\perp} \subset \Lambda$$

we can write f as a matrix

$$f = \begin{bmatrix} 1 & -{}^t\zeta Q' \widetilde{f} & -\frac{1}{2}||\zeta||^2 \\ 0 & \widetilde{f} & \zeta \\ 0 & 0 & 1 \end{bmatrix}$$

where $\widetilde{f}$ is in the orthogonal group $O(v_0^{\perp}/\mathbb{Z}v_0)$, Q' is a Gram matrix of the intersection pairing on $v_0^{\perp}/\mathbb{Z}v_0$ and $\zeta \in v_0^{\perp}/\mathbb{Z}v_0$.

Suppose that $x = \begin{bmatrix} a \\ y \\ z \end{bmatrix}$ lies in $\mathcal{H}(\mathbb{P}^2)$. Here $a, z \in \mathbb{R}$ and $y \in (v_0^{\perp}/\mathbb{Z}v_0)_{\mathbb{R}}$, so that $2az + ||y||^2 = 1$ and $a, z > 0$. Then

$$\cosh d(f(x), x) = (f(x).x) = 2az - (z\zeta.\widetilde{f}y) - \frac{1}{2}(z\zeta.z\zeta) + (y.\widetilde{f}y) + (y.z\zeta).$$

Since $1 = (x.x) = 2az + (y.y)$ and $(y.y) = (\widetilde{f}y.\widetilde{f}y)$, this gives

$$(f(x).x) = 1 - \frac{1}{2}\left|\left|z\zeta - y + \widetilde{f}y\right|\right|^2.$$

Since f is parabolic, the affine transformation

$$f_{aff} : (v_0^{\perp}/\mathbb{Z}v_0)_{\mathbb{R}} \to (v_0^{\perp}/\mathbb{Z}v_0)_{\mathbb{R}} : \; q \mapsto \widetilde{f}(q) + \zeta$$

has no fixed points. Moreover, f_{aff} preserves the lattice $v_0^{\perp}/\mathbb{Z}v_0$, and so shifts everything in $(v_0^{\perp}/\mathbb{Z}v_0)_{\mathbb{R}}$ through a distance of at least 1. Hence

$$\left|\left|\widetilde{f}\left(\frac{y}{z}\right) + \zeta - \left(\frac{y}{z}\right)\right|\right|^2 \leq -1,$$

which gives $(f(x).x) \geq 1 + z^2/2$. Since $d(x, f(x)) < \log \lambda$, it follows that $z < \lambda^{1/2} - \lambda^{-1/2}$, as required.

For the uniqueness, suppose that x also lies in the horoball $B(w_0, \lambda^{1/2} - \lambda^{-1/2})$. Then $(x.x) = 1$ and $((v_0 + w_0).(v_0 + w_0)) \geq 2$, while $(x.(v_0 + w_0)) < 2(\lambda^{1/2} - \lambda^{-1/2})$. However, this is impossible. □

Recall [CL, Prop. 3.3] that if $g \in G = Cr_2(k)$ is hyperbolic and its axis $\Gamma = \mathrm{Ax}(g)$ is not rigid, then for every bounded segment Σ of Γ and every $\epsilon > 0$, there exists $f \in G$ such that $d(x, f(x)) < \epsilon$ for all $x \in \Sigma$.

Corollary 9.2 *If g is a hyperbolic element of $Cr_2(k)$, Σ is a bounded segment of $\mathrm{Ax}(g)$ and $f \in G$ such that $d(x, f(x)) < \log \lambda$ for all $x \in \Sigma$, then f is elliptic.*

PROOF: The horoballs of the preceding theorem are disjoint. □

I am grateful to Anthony Manning, Geoff Robinson, Caroline Series and Colin Sparrow for several valuable conversations and emails.

References

[BC] J. Blanc and S. Cantat, *Dynamical degrees of birational transformations of projective surfaces*, on Cantat's web page:

http://perso.univ-rennes1.fr/serge.cantat/Articles

[CL] S. Cantat and S. Lamy, *Normal subgroups of the Cremona group*, Acta Math. **210** (2013), 31–94.

[DF] J. Diller and C. Favre, *Dynamics of meromorphic maps of surfaces*, American J. Math. **123** (2001), 1135–1169.

[G] M. Gromov, *On the entropy of holomorphic maps*, L'Enseignement Math. **49** (2003), 217-235.

[M1] C. McMullen, *Dynamics on blow-ups of the projective plane*, Publ. Math. IHES **105** (2007), 49–89.

[M2] ———, *Coxeter groups, Salem numbers and the Hilbert metric*, Publ. Math. IHES **95** (2002), 151–183.

[S] R. Steinberg, *Finite reflection groups*, Trans. AMS **91** (1959), 493–504.

Mathematics Department, King's College London, Strand, London WC2R 2LS, U.K.

E-mail address: nisb@dpmms.cam.ac.uk