ELSEVIER Journal of Pure and Applied Algebra 131 (1998) 227–244

JOURNAL OF PURE AND APPLIED ALGEBRA

Correspondences on hyperbolic curves

Shinichi Mochizuki *

Kyoto University, Research Institute for Mathematical Sciences, Kyoto, Japan

Communicated by F. Oort; received 29 November 1996; received in revised form 17 January 1997

Abstract

This paper concerns correspondences on hyperbolic curves, which are analogous to isogenies of abelian varieties. The first main result states that given a fixed hyperbolic curve in characteristic zero and a fixed "type" (g,r) (where $2g-2+r \geq 1$), there are only finitely many hyperbolic curves of type (g,r) that are isogenous to the given curve. The second main result states if $2g-2+r \geq 3$, then the only curves isogenous to a general hyperbolic curve of type (g,r) are the curves that arise as its coverings. Finally, we discuss the meaning of these results relative to the analogy with abelian varieties, especially in light of a certain result of Royden on automorphisms of Teichmüller space. © 1998 Elsevier Science B.V. All rights reserved.

1991 Math. Subj. Class.: Primary 14H35; Secondary 14E10

0. Introduction

The purpose of this paper is to prove several theorems concerning the finiteness and, more generally, the scarcity of correspondences on hyperbolic curves in characteristic zero and to comment on the meaning of these results, especially relative to the analogy with abelian varieties.

We consider hyperbolic curves over an algebraically closed field k of characteristic zero. We call two such curves X, Y *isogenous* if there exists a nonempty scheme C, together with finite étale morphisms $C \to X$, $C \to Y$. (We refer to such a pair $(C \to X, C \to Y)$ as a *correspondence* from X to Y.) It is easy to see that the relation of isogeny is an equivalence relation on the set of isomorphism classes of hyperbolic

* E-mail: motizuki@kurims.kyoto-u.ac.jp.

PII: S0022-4049(97)00078-9

curves over k. Then the first main result of this paper (cf. Lemma 4.1 and Theorem 4.2 in the text) is the following:

Theorem A. *Let k be an algebraically closed field of characteristic zero. Let X be a hyperbolic curve over k. Let (g',r') be a pair of nonnegative integers satisfying $2g'-2+r'>0$. Then (up to isomorphism) there are only finitely many hyperbolic curves over k of type (g',r') that are isogenous to X. Moreover, if K is an algebraically closed field extension of k, then any curve which is isogenous to X over K is defined over k and already isogenous to X over k.*

This result is, technically speaking, a rather trivial consequence of highly nontrivial results of Margulis and Takeuchi [4, 9]. Moreover, it is possible that Theorem A has been known to many experts for some time, but that they simply never bothered to write it down. As for the author, I was dimly aware of Theorem A for some time, without having checked the details of the proof of it, until I was asked explicitly about the finiteness stated in Theorem A by Prof. Frans Oort during my stay at Utrecht University in November 1996. I was then encouraged by Prof. Oort to write down the details; whence the present paper.

In fact, for *general* curves, we can say more: indeed, let $(\mathcal{M}_{g,r})_k$ denote the moduli stack of r-pointed smooth (proper) curves of genus g. Here, the r marked points are *unordered.* (Note that this differs slightly from the usual convention.) The complement of the divisor of marked points of such a curve will be a hyperbolic curve of type (g,r). Thus, we shall also refer (by slight abuse of terminology) to $(\mathcal{M}_{g,r})_k$ as the moduli stack of hyperbolic curves of type (g,r).

Theorem B. *Let k be an algebraically closed field of characteristic zero. Let (g,r) be a pair of nonnegative integers such that $2g-2+r\geq 3$. Let $(\mathcal{M}_{g,r})_k$ be the moduli stack of (hyperbolic) curves of type (g,r). Then there exists a dense open substack $\mathcal{U}\subseteq(\mathcal{M}_{g,r})_k$ with the following property: If X is a hyperbolic curve over some algebraically closed extension field K of k that defines a point of $\mathcal{U}(K)$, then every correspondence $(\alpha: C\to X, \beta: C\to X')$ from X to another hyperbolic curve X' is* trivial *in the sense that there exists a finite étale morphism $\gamma: X'\to X$ such that $\alpha=\gamma\circ\beta$*

In particular, for such an X, every X' isogenous to X can be realized as a finite étale covering of X.

Theorem B follows from Theorem 5.3 in the text. Moreover, in the exceptional cases ruled out in the statement of Theorem B, a general curve always admits nontrivial correspondences (see Theorem 5.7 and the Remark following it).

One aspect of the significance of Theorem A is that it shows that *although "isogeny" of hyperbolic curves is a natural analogue of the notion of isogeny for abelian varieties, the behavior of hyperbolic curves with respect to isogeny is somewhat different from the behavior of abelian varieties with respect to isogeny.* For instance, if one starts with a (principally polarized) abelian variety, and considers all the principally polarized abelian varieties isogenous to it – i.e., a so-called "Hecke orbit"

– such orbits (far from being finite) are dense in the moduli stack of principally polarized abelian varieties in characteristic zero $\mathscr{A}_g$. Indeed, one can see this density in the classical complex topology by using the uniformization of $\mathscr{A}_g$ by $\mathrm{Sp}_{2g}(\mathbf{R})$ (modulo a maximal compact subgroup), and the fact that $\mathrm{Sp}_{2g}(\mathbf{Q})$ is dense in $\mathrm{Sp}_{2g}(\mathbf{R})$.

One way to describe why such "Hecke orbits" tend to be so big is to regard this phenomenon as a consequence of the existence of various natural nontrivial correspondences on $\mathscr{A}_g$, the so-called Hecke correspondences. "Acting on" some initial point with these correspondences gives a natural way of constructing lots of abelian varieties isogenous to the abelian variety corresponding to the initial point. *Given these circumstances, Theorem A then leads one to suspect that unlike $\mathscr{A}_g$, the moduli stack $\mathscr{M}_{g,r}$ of hyperbolic curves of type (g,r) will not have very many correspondences.* In fact, one has the following result (given as Theorem 6.1 in the text):

Theorem C. *Suppose that $2g-2+r \geq 3$. Then $\mathscr{M}_{g,r}$ is generically a scheme, and moreover, does not admit any nontrivial automorphisms or correspondences.*

Technically speaking, this is a trivial consequence of a theorem of Royden, although I have not seen Royden's theorem interpreted in this way – i.e., *as implying a statement about correspondences on $\mathscr{M}_{g,r}$* – elsewhere.

It is intriguing that the exceptional cases ruled out in Theorems B and C (i.e., the cases where $2g-2+r \leq 2$) are precisely the same. That is to say, the existence of nontrivial correspondences on a general curve appears to be related to irregularities in the holomorphic automorphism group of Teichmüller space. Unfortunately, I do not have any theoretical explanation for this phenomenon at the time of writing.

Finally, another interesting aspect of this circle of ideas is the following: in the case of $\mathscr{A}_g$, the *algebraic* Hecke correspondences may be constructed p-adically using the Serre–Tate parameters, or, equivalently, by means of a certain canonical Frobenius lifting over the ordinary locus of the p-adic completion of $\mathscr{A}_g$. (This Frobenius lifting is the Frobenius lifting given by assigning to an abelian variety with ordinary reduction modulo p, the quotient of this abelian variety by the multiplicative portion of the kernel of multiplication by p. For $g=1$, this Frobenius lifting is known as the "Deligne–Tate map".) Put another way, although this canonical Frobenius lifting is essentially p-adic in nature, and cannot be algebraized, by combining it with its transpose, we obtain a correspondence which *can* be algebraized – namely, into a Hecke correspondence. On the other hand, in the case of $\mathscr{M}_{g,r}$, there does exist a direct analogue of the canonical Frobenius lifting on (the ordinary locus of the p-adic completion of) $\mathscr{A}_g$ – namely, the theory of [5–7]. Thus, it is natural to ask whether the canonical modular Frobenius lifting on $\mathscr{M}_{g,r}$ can be algebraized in a similar fashion by forming a correspondence from the union of the Frobenius lifting and its transpose. Theorem C tells us, however, that the answer is no.

Thus, although Theorems A–C are technically just concatenations of known results, their significance in the context of the theory of [5–7] appears not to have been noticed by previous authors.

1. Basic definitions

Let k be an *algebraically closed field of characteristic zero*. Let X be a *hyperbolic curve over* k. By this, we mean that X is an open subset of a proper, smooth, connected, one-dimensional k-scheme $\overline{X}$ such that if g is the genus of $\overline{X}$ (i.e., the dimension of $H^1(\overline{X}, \mathcal{O}_{\overline{X}})$ over k), and r is the number of points in $\overline{X} - X$, then we have $2g-2+r>0$. We shall refer to (g,r) as the *type of* X.

Suppose that Y and Z are also hyperbolic curves over k. Then let us give the following definition.

Definition 1.1. We shall refer to as *a correspondence from* X *to* Y any (ordered) pair of finite, étale morphisms $\alpha : C \to X$, $\beta : C \to Y$, where we assume that C is nonempty. Thus, C will necessarily be a finite disjoint union of hyperbolic curves over k. Note that we do not assume that C is connected.

Definition 1.2. We shall refer to a correspondence $(\alpha : C \to X, \beta : C \to Y)$ from X to Y as *trivial* if there exists a finite étale morphism $\gamma : Y \to X$ such that $\alpha = \gamma \circ \beta$.

Definition 1.3. Given a correspondence (α, β) from X to Y, we shall refer to as *the transpose correspondence to* (α, β) the correspondence (from Y to X) given by the pair (β, α).

Definition 1.4. Let $(\alpha_1 : C_1 \to X, \beta_1 : C_1 \to Y)$ (respectively, $(\alpha_2 : C_2 \to Y, \beta_2 : C_2 \to Z)$) be a correspondence from X to Y (respectively, Y to Z). Then we shall refer to as *the composite of these two correspondences* the correspondence given by the following pair of morphisms: the first morphism $C_1 \times_Y C_2 \to X$ is given by composing the projection to C_1 with α_1; the second morphism $C_1 \times_Y C_2 \to Z$ is given by composing the projection to C_2 with β_2. Thus, the composite correspondence is a correspondence from X to Z.

As the terminology "from X to Y" implies, we want to regard correspondences from X to Y as a sort of hyperbolic analogue of isogenies between abelian varieties.

Definition 1.5. We shall call two hyperbolic curves X and Y over k *isogenous* if there exists a correspondence from X to Y.

Note that by taking transposes and composites of correspondences, one sees immediately that the relation of isogeny is an equivalence relation.

2. Review of results of Margulis and Takeuchi

In this section, we assume that k is $\mathbf{C}$, the field of complex numbers. Let X be a hyperbolic curve over k. Let $\mathscr{X}$ be the Riemann surface associated to X. Thus, the underlying set of $\mathscr{X}$ is $X(\mathbf{C})$. Let $\widetilde{\mathscr{X}}$ be the universal covering space of $\mathscr{X}$. Thus,

$\widetilde{\mathscr{X}}$ is a Riemann surface. From elementary complex analysis, one knows that $\widetilde{\mathscr{X}}$ is holomorphically isomorphic to $\mathscr{H} \stackrel{\text{def}}{=} \{z \in \mathbf{C} \mid \text{Im}(z) > 0\}$. Let us choose a holomorphic identification of $\widetilde{\mathscr{X}}$ with $\mathscr{H}$. Recall also from elementary complex analysis that the group of holomorphic automorphisms of $\mathscr{H}$ may be identified with $\text{PSL}_2(\mathbf{R})^0$ (acting via linear fractional transformations). (Here, the superscripted "0" denotes the connected component of the identity.) Let us write Π for the (topological) fundamental group of $\mathscr{X}$ (for some choice of base-point). Then the action of Π on $\widetilde{\mathscr{X}}$ by deck transformations defines an injection $\Pi \hookrightarrow \text{Aut}(\mathscr{H}) = \text{PSL}_2(\mathbf{R})^0$. Let us denote the image of this injection by $\Gamma \subseteq \text{PSL}_2(\mathbf{R})^0$. In the following, we will always think of Γ as a *subgroup* of $\text{PSL}_2(\mathbf{R})^0$.

Next, if Γ_1 and Γ_2 are two subgroups of $\text{PSL}_2(\mathbf{R})^0$, let us write $\Gamma_1 \sim \Gamma_2$ (read "Γ_1 is *commensurable* with Γ_2") if $\Gamma_1 \cap \Gamma_2$ has finite index in both Γ_1 and Γ_2. Also, let us write

$$\text{Comm}(\Gamma) \stackrel{\text{def}}{=} \{\gamma \in \text{PSL}_2(\mathbf{R})^0 \mid (\gamma \cdot \Gamma \cdot \gamma^{-1}) \sim \Gamma\}.$$

Note that $\Gamma \subseteq \text{Comm}(\Gamma)$. Then we give the following definition.

Definition 2.1. We shall say that X, $\mathscr{X}$, or Γ has *infinitely many correspondences* if Γ has infinite index in $\text{Comm}(\Gamma)$.

By a theorem of [4] (see Theorem 2.5 below), X is "arithmetic" if and only if it has infinitely many correspondences. We would like to review this result below, but before we can do this, we need to review what it means for X to be "arithmetic". Unfortunately, for hyperbolic curves, there (at least) two different ways to define arithmeticity. In this paper, we will need to use both definitions, so in the following, we shall review both definitions, and then show that they are equivalent.

We begin with the definition of [4, Chap. 9, Section 1.5]: To do this, first we need to recall some basic terminology. If F is a field of characteristic zero, and G is an algebraic group over F, then we shall say that G is *almost F-simple* if any proper, closed, normal algebraic subgroup of G defined over F is finite. Also, we shall denote by $(\text{PSL}_2)_{\mathbf{R}}$ the algebraic group "PSL_2" over $\mathbf{R}$.

Definition 2.2. We shall call X, $\mathscr{X}$, or Γ *Margulis arithmetic* if there exists a connected non-commutative almost $\mathbf{Q}$-simple algebraic group G over $\mathbf{Q}$, together with a surjection $\tau : G_{\mathbf{R}} \stackrel{\text{def}}{=} G \otimes_{\mathbf{Q}} \mathbf{R} \to (\text{PSL}_2)_{\mathbf{R}}$ of algebraic groups over $\mathbf{R}$ such that the Lie group $(\text{Ker}\,\tau)(\mathbf{R})$ is compact, and the subgroups $\tau(G(\mathbf{Z}))$ and Γ (of $\text{PSL}_2(\mathbf{R})^0$) are commensurable. (Here, by the notation $G(\mathbf{Z})$, we mean the $\mathbf{Z}$-valued points of GL_N that lie inside G for some embedding of $\mathbf{Q}$-algebraic groups $G \hookrightarrow (\text{GL}_N)_{\mathbf{Q}}$. Thus, properly speaking, "$G(\mathbf{Z})$" is an equivalence class of commensurable subgroups of $G(\mathbf{Q})$ (see [4, p. 60, Lemma 3.1.1(iv)] and the following discussion for more details).)

Next, we review the definition of arithmeticity given in [9]:

Definition 2.3. We shall call X, $\mathscr{X}$, or Γ *Shimura arithmetic* if the following data exist:

(1) a totally real algebraic number field F;

(2) a quaternion algebra A over F which is trivial at one of the infinite places of F and nontrivial at all the other infinite places;

(3) a trivialization of A at the infinite place of F at which A is trivial; this trivialization will be used to regard A as a subalgebra of $M_2(\mathbf{R})$;

(4) an order $\mathcal{O}_A \subseteq A$ such that the intersection of $\mathcal{O}_A \subseteq A \subseteq M_2(\mathbf{R})$ with $SL_2(\mathbf{R}) \subseteq M_2(\mathbf{R})$ has image in $PSL_2(\mathbf{R})^0$ commensurable with Γ.

(The reason for this terminology is that the situation described in this definition (used in [9]) was studied extensively by Shimura in, for instance, [8, Chap. 8].)

The following result is well-known, but I do not know of an adequate reference:

Proposition 2.4. *The Riemann surface $\mathscr{X}$ is Margulis arithmetic if and only if it is Shimura arithmetic.*

Proof. That Shimura arithmeticity implies Margulis arithmeticity is clear. Thus, let us assume that Γ is Margulis arithmetic, and prove that it is also Shimura arithmetic. Let us suppose that we have a G and a $\tau : G_{\mathbf{R}} \to (PSL_2)_{\mathbf{R}}$ as in Definition 2.2. First, let us observe that the fact that G is almost $\mathbf{Q}$-simple implies that $G_{\overline{\mathbf{Q}}}$ is the almost direct product of its almost simple factors $H_i \subseteq G_{\overline{\mathbf{Q}}}$ (where $i = 1, \ldots, n$) – see, e.g., [4, p. 21]. Moreover, since the almost simple factors are canonical, it follows that the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $G_{\overline{\mathbf{Q}}}$ (given by the fact that $G_{\overline{\mathbf{Q}}}$ is defined over G) permutes these almost simple factors. Since G is almost $\mathbf{Q}$-simple, it even follows that $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts *transitively* on the almost simple factors of $G_{\overline{\mathbf{Q}}}$. Thus, the stabilizer of, for instance, H_1 in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is $\mathrm{Gal}(\overline{\mathbf{Q}}/F)$, for some finite extension F of $\mathbf{Q}$. Moreover, the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/F)$ on H_1 (which is an algebraic group over $\overline{\mathbf{Q}}$) defines an F-rational structure on H_1, i.e., there is some F-algebraic group H_F such that $H_1 = (H_F) \otimes_F \overline{\mathbf{Q}}$. In fact, it follows from the definitions that the other H_i's are just the Galois conjugates of H_1, hence that the inclusion $(H_F)_{\overline{\mathbf{Q}}} = H_1 \hookrightarrow G_{\overline{\mathbf{Q}}}$ induces an isogeny of $G' \stackrel{\mathrm{def}}{=} \mathrm{Rest}_{F/\mathbf{Q}}(H_F)$ (where "$\mathrm{Rest}_{F/\mathbf{Q}}$" denotes "Weil restriction of scalars from F to $\mathbf{Q}$") onto G.

Next, we would like to observe that $\mathrm{Lie}(H_F)_{\mathbf{C}}$ is isomorphic to $sl_2(\mathbf{C})$. To see this, we argue as follows. First, note that $\tau \otimes_{\mathbf{R}} \mathbf{C}$ induces a surjection of Lie algebras from $\mathrm{Lie}(G_{\mathbf{C}})$ onto $sl_2(\mathbf{C})$. Since $sl_2(\mathbf{C})$ is a simple Lie algebra, it thus follows that at least one of the $\mathrm{Lie}(H_i)_{\mathbf{C}}$'s is isomorphic to $sl_2(\mathbf{C})$. But this implies that $\mathrm{Lie}(H_F)_{\mathbf{C}} \cong sl_2(\mathbf{C})$, as desired.

Now let H_F^* be the quotient of H_F by its centre. Then it follows from the elementary theory of algebraic groups, plus what we did in the preceding paragraph, that H_F^* is some twisted form of $(PGL_2)_{\overline{\mathbf{Q}}}$ over F. In other words, H_F^* defines a class in the nonabelian Galois cohomology set $H^1(F, PGL_2)$, hence an element of the Brauer group of F of order two. Put another way, there exists a quaternion algebra A over F such that H_F^* may be identified with (the F-algebraic group corresponding to) $A^\times/F^\times$.

Next, we would like to show that F is *totally real*, and that A is the sort of quaternion algebra that appears in Definition 2.3. To do this, we consider $G(\mathbf{R})$. The above analysis of G' and H_F^* shows that for each complex infinite place of F, there appears in $G_{\mathbf{R}}$ an almost $\mathbf{R}$-simple factor which is isogenous to $\mathrm{Rest}_{\mathbf{C}/\mathbf{R}}(\mathrm{PGL}_2)_{\mathbf{C}}$. If τ were trivial on this factor, then $(\mathrm{Ker}\,\tau)(\mathbf{R})$ would contain $\mathrm{SL}_2(\mathbf{C})$ or $\mathrm{PSL}_2(\mathbf{C})$, hence would be noncompact. Thus, we obtain that τ is nontrivial on such a factor. On the other hand, this implies that there exists a nontrivial morphism $\mathrm{PGL}_2(\mathbf{C}) \to \mathrm{PGL}_2(\mathbf{R})^0$ of real Lie groups. Moreover, since $\mathrm{PGL}_2(\mathbf{R})^0$ has an $\mathbf{R}$-simple Lie algebra, it follows that such a morphism is surjective. But since the kernel of this morphism is compact and of real dimension three, this implies that the maximal compact subgroup of $\mathrm{PGL}_2(\mathbf{C})$ is normal, which is absurd. This contradiction implies that F has no complex places.

Similarly, if the quaternion algebra A were trivial at two real places of F, then we would get a surjection $\mathrm{PGL}_2(\mathbf{R})^0 \times \mathrm{PGL}_2(\mathbf{R})^0 \to \mathrm{PGL}_2(\mathbf{R})^0$ (of real Lie groups) with compact kernel. But since the kernel of such a surjection is necessarily isomorphic to $\mathrm{PGL}_2(\mathbf{R})^0$, this is absurd. Thus, we see that A is as in Definition 2.3. Now one sees easily that τ defines a trivialization (datum (3) of Definition 2.3), and that there exist representatives of the equivalence class "$G(\mathbf{Z})$" that arise in the fashion described in (4) of Definition 2.3. This shows that Γ is Shimura arithmetic, thus completing the proof of the proposition. □

In the future, we shall refer to X, $\mathscr{X}$, or Γ as arithmetic if it is either Margulis arithmetic or Shimura arithmetic (since we now know that these two notions of arithmeticity are equivalent).

Now we are ready to state the main results that we wanted to review in this section:

Theorem 2.5 (Margulis [4, p. 337, Theorem 27; p. 60, Lemma 3.1.1(v)]). *The hyperbolic Riemann surface $\mathscr{X}$ is arithmetic if and only if it has infinitely many correspondences (in the sense of Definition* 2.1).

Theorem 2.6 (Takeuchi [9, Theorem 2.1]). *There are only finitely many arithmetic X over $\mathbf{C}$ of a given type (g,r).*

The first main result of this paper will essentially be a consequence of the above two results, plus various elementary manipulations, to be discussed in the following section.

3. The non-arithmetic case

We maintain the notation of Section 2. Moreover, in this section, we assume that X is *not arithmetic*. Thus, we have $\Gamma \subseteq \mathrm{Comm}(\Gamma) \subseteq \mathrm{PSL}_2(\mathbf{R})^0$, and Γ is *of finite index* in $\mathrm{Comm}(\Gamma)$. Now we would like to form the quotient of $\mathscr{H}$ by $\mathrm{Comm}(\Gamma)$ *in the sense of stacks*. (We refer to Chap. 1, Section 4 of [2] for generalities on stacks.) Let us denote this quotient by $\mathscr{Y}$. Note that since Γ has finite index in $\mathrm{Comm}(\Gamma)$, it follows that we get a finite étale morphism $\mathscr{X} \to \mathscr{Y}$. Moreover, this finite étale morphism gives the analytic stack $\mathscr{Y}$ an algebraic structure, so we obtain an algebraic stack Y together with a finite étale morphism $X \to Y$ corresponding to $\mathscr{X} \to \mathscr{Y}$.

Definition 3.1. Suppose that X is *not arithmetic*. Then we shall refer to Y (respectively $\mathscr{Y}$) as the *hyperbolic core of* X (*respectively*, $\mathscr{X}$).

Next, we would like to suppose that we have been given a correspondence ($\alpha : C \to X$, $\beta : C \to Z$) from X to some other hyperbolic curve Z; we assume here, for simplicity, that C is connected. This gives rise to corresponding analytic morphisms $\mathscr{C} \to \mathscr{X}$, $\mathscr{C} \to \mathscr{Z}$. Moreover, these two morphisms induce isomorphisms between the respective universal covering spaces. Also, we get various groups of deck transformations $\Gamma_{\mathscr{X}}, \Gamma_{\mathscr{C}}, \Gamma_{\mathscr{Z}} \subseteq \mathrm{Aut}(\mathscr{H}) = \mathrm{PSL}_2(\mathbf{R})^0$, together with various inclusion relations: $\Gamma_{\mathscr{C}} \subseteq \Gamma_{\mathscr{X}}$; $\Gamma_{\mathscr{C}} \subseteq \Gamma_{\mathscr{Z}}$. (Note that the object that we have been referring to up till now by the notation "Γ" will now be referred to as "$\Gamma_{\mathscr{X}}$".) Now we have the following result:

Proposition 3.2. *We have* $\Gamma_{\mathscr{Z}} \subseteq \mathrm{Comm}(\Gamma_{\mathscr{X}})$.

Proof. First observe that for the purpose of proving this proposition, we may assume that C is *Galois* over Z. Thus, $\Gamma_{\mathscr{C}}$ is normal (and of finite index) in $\Gamma_{\mathscr{Z}}$. Now let $\gamma \in \Gamma_{\mathscr{Z}}$. Then $\Gamma_{\mathscr{X}} \cap (\gamma \cdot \Gamma_{\mathscr{X}} \cdot \gamma^{-1})$ contains $\Gamma_{\mathscr{C}} \cap (\gamma \cdot \Gamma_{\mathscr{C}} \cdot \gamma^{-1}) = \Gamma_{\mathscr{C}}$. In particular, it follows that $\Gamma_{\mathscr{X}} \cap (\gamma \cdot \Gamma_{\mathscr{X}} \cdot \gamma^{-1})$ is of finite index in $\Gamma_{\mathscr{X}}$ (and hence also – by replacing γ by γ^{-1} – of finite index in $\gamma \cdot \Gamma_{\mathscr{X}} \cdot \gamma^{-1}$). This completes the proof of the proposition. □

Interpreting this proposition in terms of Riemann surfaces, we see that there exists a unique finite étale morphism $\mathscr{Z} \to \mathscr{Y}$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathscr{C} & \longrightarrow & \mathscr{Z} \\ \downarrow & & \downarrow \\ \mathscr{X} & \longrightarrow & \mathscr{Y} \end{array}$$

(Here, the upper horizontal and left-hand vertical morphisms are the analytic morphisms associated to β and α, respectively, and the lower horizontal morphism is the morphism that appeared in the construction of the hyperbolic core of $\mathscr{X}$.) Moreover, this diagram can be algebraized. Thus, in particular, we obtain an (algebraic) finite étale morphism $Z \to Y$.

Write (g_Z, r_Z) for the type of Z. Observe from the Riemann–Hurwitz formula that there exists a positive rational number $e_Y \in \mathbf{Q}$ such that if T is any hyperbolic curve, of type (g_T, r_T), and $f : T \to Y$ is finite étale of degree d, then $2g_T - 2 + r_T = e_Y \cdot d$. Now we are ready to prove the following result:

Theorem 3.3. *Suppose that X is not arithmetic. Fix a pair of nonnegative integers (g', r') such that $2g' - 2 + r' > 0$. Then there exist (up to isomorphism) only finitely many hyperbolic curves Z of type (g', r') that are isogenous to Z.*

Proof. Indeed, by the above discussion, we see that we get (for such Z) a finite étale morphism $Z \to Y$. Moreover, the degree of this morphism (and hence also of the corresponding analytic morphism $\mathscr{X} \to \mathscr{Y}$) is bounded by a number that depends only on g', r', and e_Y (i.e., X). On the other hand, note that since $\mathrm{Comm}(\Gamma_{\mathscr{X}})$ has a finite index subgroup which is finitely generated – namely, $\Gamma_{\mathscr{X}}$ – it follows that $\mathrm{Comm}(\Gamma_{\mathscr{X}})$ is itself finitely generated. Moreover, since one may think of $\mathrm{Comm}(\Gamma_{\mathscr{X}})$ as the fundamental group of $\mathscr{Y}$, the fact that this group is finitely generated implies that there are (up to isomorphism) only finitely many finite étale coverings of $\mathscr{Y}$ of degree less than some fixed number. This observation completes the proof of the Theorem. □

Remark. Let X be *nonarithmetic*. Then let us observe that *if the automorphism group* $G \stackrel{\text{def}}{=} \mathrm{Aut}(X)$ *of* X *is nontrivial, then* X *is not equal to its hyperbolic core.* Indeed, since it is clear that the morphism $X \to Y$ that defines Y as the hyperbolic core of X is *natural*, it follows that this morphism is equivariant with respect to the given action of G on X and the trivial action of G on Y. Thus, $X \to Y$ necessarily factors through the quotient (*in the sense of stacks*) $X \to X/G$, which implies that $X \to Y$ has degree > 1, as claimed.

On the other hand, *the converse to this statement, i.e., that "if the degree of* $X \to Y$ *is* > 1, *then* X *admits nontrivial automorphisms", is false in general.* Indeed, one can construct such an X as follows: Let X' be a nonarithmetic affine hyperbolic curve which is equal to its hyperbolic core (such X' exist by Theorem 5.3 below). Then the fundamental group of $\mathscr{X}'$ will be a nonabelian finitely generated free group, so it is easy to see that it admits a finite étale covering $X \to X'$ (where X is connected, and $X \to X'$ has degree > 1) such that there are no intermediate *Galois* coverings $X \to X''$ (except $X = X$). (For instance, take $X \to X'$ to be non-Galois of prime degree.) Then I *claim* that $X \to X'$ exhibits X' as the hyperbolic core of X. Indeed, if $X \to Y$ is the morphism defining Y as the hyperbolic core of X, then $X \to Y$ must factor through $X \to X'$; but this gives us a finite étale morphism $X' \to Y$ which must be an isomorphism (cf. the discussion preceding Theorem 3.3) since X' is equal to its own hyperbolic core. This proves the claim. Thus, X has no automorphisms (for if it did, then by the argument of the preceding paragraph $X \to X'$ would admit a nontrivial intermediate Galois covering $X \to X''$), but is not equal to its hyperbolic core.

4. The main theorem

Now we return to the situation where k is any algebraically closed field of characteristic zero. Let X be a hyperbolic curve over k. Let K be an algebraically closed field of characteristic zero that contains k. Write X_K for $X \otimes_k K$.

Lemma 4.1. *Suppose that* X_K *is isogenous to some hyperbolic curve* Z_K *over* K. *Then* Z_K *is the result of base-extending some hyperbolic curve* Z *over* k *from* k *to* K, *and,* *moreover, any correspondence from X_K to Z_K descends to a correspondence from X to Z.*

Proof. Indeed, since it only takes "finitely many equations" to define a curve or a correspondence, it follows that any $\alpha_K : C_K \to X_K$, $\beta_K : C_K \to Z_K$ descends to a pair of finite étale morphisms $\alpha_R : C_R \to X_R$, $\beta_R : C_R \to Z_R$ of curves over R, where R is a finitely generated k-subalgebra of K. Now observe that since α_R is finite étale, and the étale site of a scheme is rigid with respect to deformations, it follows that α_R descends to a finite étale morphism $\alpha : C \to X$. Moreover, if we restrict β_R to a closed point s of Spec(R), the tangent space to the space of deformations of $\beta_s : C_s = C \to Z_s$ is given by the kernel of the pull-back map $H^1(Z_s, \tau_{Z_s}) \to H^1(C, \tau_C = \tau_{Z_s}|_C)$ (where "τ" denotes "tangent bundle"). But I claim that this pull-back map is injective: Indeed, this follows from the existence of the trace map which gives a one-sided inverse of the pull-back map. (Note that here we use the fact that the degree of β_s is invertible in k – a consequence of the assumption that k is of characteristic zero.) Thus, there are no nontrivial deformations of β_s, so again we conclude that $\beta_R = \beta_s \otimes_k R$. This completes the proof of the lemma. □

Now we are ready to prove the first main result of the paper:

Theorem 4.2. *Let k be an algebraically closed field of characteristic zero. Let X be a hyperbolic curve over k. Let (g', r') be a pair of nonnegative integers satisfying $2g' - 2 + r' > 0$. Then (up to isomorphism) there are only finitely many hyperbolic curves over k of type (g', r') that are isogenous to X.*

Proof. First, observe that given any finite set of curves isogenous to X, there exists a subfield k' of k which is finitely generated over $\mathbf{Q}$ over which all the curves of that finite set, together with X itself, are defined. Thus, it suffices to show that the number of curves of type (g', r') that are isogenous to X over $\overline{k'}$ (i.e., the algebraic closure of k') is bounded by a number independent of the choice of subfield k'. On the other hand, since there always exists an embedding $\overline{k'} \subseteq \mathbf{C}$, the uniform boundedness statement of the preceding sentence will be proven if we can prove the finiteness statement of the theorem in the case $k = \mathbf{C}$. Thus, we may assume $k = \mathbf{C}$. Then either X is arithmetic or it is not arithmetic. If X is arithmetic, it follows easily from the definitions that any curve isogenous to X will also be arithmetic. Thus, in this case, the theorem follows from Theorem 2.6. If X is not arithmetic, then the theorem is simply Theorem 3.3. □

5. Isogenies of general curves

In this section, we show that (if one rules out certain exceptional cases, then) the only curves isogenous to a general hyperbolic curve are the finite étale coverings of the curve. This essentially amounts to a straightforward elementary calculation involving the Riemann–Hurwitz formula, but nevertheless we give full details below. We remark that although the statement proven below (Theorem 5.3) that a general curve (of all but a few exceptional types) is equal to its hyperbolic core is strictly stronger than the statement that such a curve has no nontrivial automorphisms (cf. the remark at the end of Section 3), this calculation involving the Riemann–Hurwitz formula is *exactly the same* as in the proof that such a curve has no nontrivial automorphisms. Thus, in principle, this calculation is "well-known." Nevertheless, I have chosen to give full details below partly for the convenience of the reader, and partly because of the following set of circumstances:

> In the case $r=0$, the calculation is *much simpler* and is contained, for instance, in [1]. Moreover, the result on automorphisms of a general curve for $r=0$ *immediately implies* the result on automorphisms of a general curve for $r>0$. Thus, if one is only interested in automorphisms, there is no need to carry out this calculation in the more difficult case $r>0$. On the other hand, the result that a general curve is equal to its hyperbolic core when $r=0$ *does not formally imply* the corresponding result when $r>0$. Thus, to obtain the result on the hyperbolic core, one must carry out this calculation in complete generality (i.e., allowing that r might be nonzero). Since I do not know of a reference that gives this calculation in this generality, I decided to give full details here.

Lemma 5.1. *Let k be algebraically closed of characteristic zero. Suppose that k is a subfield of $\mathbf{C}$. Let, X be a hyperbolic curve over k. Suppose that $X_{\mathbf{C}} \stackrel{\text{def}}{=} X \otimes_k \mathbf{C}$ is* not arithmetic. *Then the morphism $X_{\mathbf{C}} \to Y_{\mathbf{C}}$ appearing in the discussion of the hyperbolic core of $X_{\mathbf{C}}$ (see Definition* 3.1) *descends to some morphism $X \to Y$ over k. Moreover, $X \to Y$ has the universal property that any correspondence $(C \to X, C \to Z)$ over k, fits uniquely into a commutative diagram*:

$$\begin{array}{ccc} C & \longrightarrow & Z \\ \downarrow & & \downarrow \\ X & \longrightarrow & Y \end{array}$$

Finally, the morphism $X \to Y$ is independent (up to canonical isomorphism) of the embedding of k into $\mathbf{C}$.

Proof. Observe that (from the definition of $X_{\mathbf{C}} \to Y_{\mathbf{C}}$) there exists a finite étale Galois covering $X'_{\mathbf{C}} \to X_{\mathbf{C}}$ such that $X'_{\mathbf{C}} \to Y_{\mathbf{C}}$ is Galois. Since étale coverings are rigid, $X'_{\mathbf{C}} \to X_{\mathbf{C}}$ descends to some $X' \to X$ over k. Moreover, since automorphisms of hyperbolic curves are rigid, $\text{Aut}_{\mathbf{C}}(X'_{\mathbf{C}}) = \text{Aut}_k(X')$. Thus, $G \stackrel{\text{def}}{=} \text{Gal}(X'_{\mathbf{C}}/Y_{\mathbf{C}})$ acts on X', so that we may form the quotient (*in the sense of stacks*) $X' \to Y \stackrel{\text{def}}{=} X''/G$. Moreover, this quotient clearly factors through X, so we obtain a morphism $X \to Y$ that descends $X_{\mathbf{C}} \to Y_{\mathbf{C}}$, as desired. The universal property follows immediately by descending (cf. the argument of Lemma 4.1) from $\mathbf{C}$ to k the corresponding analytic universal property discussed in the paragraph following Proposition 3.2. The fact that $X \to Y$ does not depend on the embedding of k into $\mathbf{C}$ follows from the existence/uniqueness assertion inherent in the statement of the universal property. □

Definition 5.2. Suppose that we are in the situation of Lemma 5.1. Then we shall refer to the stack Y constructed in Lemma 5.1 as the *hyperbolic core* Y of X.

Notation. Let Y be a smooth, one-dimensional algebraic stack over a field k. Suppose further that Y is generically a scheme. Then we shall use the following notation for objects related to Y: let us write Y^c for the "course moduli space" associated to Y (see, e.g., [2, Chap. 1, Section 4.10], for a discussion of the course moduli space associated to a stack). Thus, Y^c is a smooth, connected, one-dimensional scheme over k, and we have a natural morphism $Y \to Y^c$. Let us write g_Y for the genus of the compactification of Y^c, and r_Y for the number of points that need to be added to Y^c to compactify it. Let us write Σ_Y for the set of points of Y^c over which $Y \to Y^c$ is not étale. For $\sigma \in \Sigma_Y$, let i_σ be the ramification index of $Y \to Y^c$ at σ. Thus, i_σ will always be an integer ≥ 2. Let $j_\sigma \stackrel{\text{def}}{=} (i_\sigma - 1)/i_\sigma$. Thus, j_σ is a rational number $\geq \frac{1}{2}$ and <1. We shall refer to the data $(g_Y; r_Y; \{i_\sigma\}_{\sigma\in\Sigma_Y})$ as the *type of the stack* Y. Finally, we define

$$e_Y \stackrel{\text{def}}{=} 2g_Y - 2 + r_Y + \sum_{\sigma\in\Sigma_Y} j_\sigma.$$

Thus, one may think of e_Y as the *Euler characteristic* of Y.

Theorem 5.3. *Let k be algebraically closed of characteristic zero. Suppose that k is a subfield of $\mathbf{C}$. Fix nonnegative integers g and r such that $2g - 2 + r \geq 3$. Then there exists an open dense substack $\mathscr{U} \subseteq (\mathscr{M}_{g,r})_k$ (where $(\mathscr{M}_{g,r})_k$ is the moduli stack of (hyperbolic) curves of type (g,r) over k) with the following property: If X is a hyperbolic curve over some extension algebraically closed field K of k corresponding to a point $\in \mathscr{U}(K)$, then the hyperbolic core of X is equal to X. Thus, in particular, (if K is algebraically closed, then) for such an X, every hyperbolic curve isogenous to X can be realized as a finite étale covering of X.*

Remark. The exceptional cases ruled out by the assumption that $2g - 2 + r \geq 3$ are precisely the cases where (g,r) is equal to $(0,3)$, $(0,4)$, $(1,1)$, $(1,2)$, or $(2,0)$.

Proof of Theorem 5.3. It suffices to find some $\mathscr{U}$ as in the statement of the theorem with the property that for X corresponding to a K-valued point of $\mathscr{U}$ (where K is an algebraically closed extension field of k), the hyperbolic core (Definition 3.1) of X is equal to X itself. To do this, let us consider the case of an X which is nonarithmetic and whose natural morphism $X \to Y$ to its hyperbolic core Y has degree $d > 1$. From the Riemann–Hurwitz formula, we have

$$2g - 2 + r = d\left(2g_Y - 2 + r_Y + \sum_{\sigma\in\Sigma_Y} j_\sigma\right).$$

Since $2g-2+r>0$, it follows that the expression in parentheses, which is simply e_Y, is also >0. Now we have the following (well-known):

Lemma 5.4. *The expression in parentheses e_Y is bounded below by an absolute positive constant, independent of X, g, and r.*

Proof. This is a simple combinatorial exercise. If $2g_Y-2+r_Y \geq 1$, then $e_Y \geq 1$. If $2g-2+r_Y=0$, then $e_Y \geq \frac{1}{2}$. If $2g_Y-2+r_Y=-1$, then either Σ_Y has at least three elements, in which case $e_Y \geq \frac{1}{2}$, or Σ_Y has precisely two elements, in which case $e_Y \geq \frac{1}{6}$. If $2g_Y-2+r_Y=-2$ (so $g_Y=r_Y=0$), then we have the following possibilities: If Σ_Y has at least five elements, then $e_Y \geq \frac{1}{2}$. If Σ_Y has precisely four elements, then $e_Y \geq \frac{1}{6}$. Otherwise, Σ_Y has precisely three elements. In this last case, observe that it is never the case that two i_σ's are $=2$. This observation implies that if the largest i_σ is greater than or equal to 7, then $e_Y \geq \frac{1}{6}-\frac{1}{7}>0$. But there are only finitely many possibilities for Σ_Y for which the largest i_σ is less than or equal to 6. This completes the proof. □

Lemma 5.5. *If g and r are fixed, then there is only a finite number of possibilities for d, g_Y, r_Y, and Σ_Y.*

Proof. Since $2g-2+r=d\cdot e_Y$, and (by Lemma 5.4) e_Y is bounded below by positive constant, it follows that d is bounded above. Since d is a positive integer, it thus follows that there is only a finite number of possibilities for d. Thus, there is only a finite number of possibilities for e_Y. Since $2g_Y-2+r_Y+\frac{1}{2}|\Sigma_Y| \leq e_Y$ (where $|\Sigma_Y|$ is the cardinality of Σ_Y), it thus follows that there is only a finite number of possibilities for g_Y, r_Y, and $|\Sigma_Y|$. But since each $i_\sigma \leq d$, it thus follows that there is only a finite number of possibilities for Σ_Y. This completes the proof. □

Lemma 5.6. *The locus (inside $(\mathcal{M}_{g,r})_k$) of nonarithmetic curves that are not equal to their own hyperbolic cores is constructible (in $(\mathcal{M}_{g,r})_k$).*

Proof. Indeed, for each possible d, g_Y, r_Y, Σ_Y, one considers the moduli stack $\mathcal{N}$ of smooth, one-dimensional hyperbolic stacks Y with invariants g_Y, r_Y, Σ_Y. (Note – for later use – that the dimension of this moduli stack is equal to $3g_Y-3+r_Y+|\Sigma_Y|$.) Then the moduli stack $\mathcal{N}'$ of pairs consisting of such Y together with a finite étale covering $X \to Y$ of degree d (where X is of type (g,r)) forms a finite étale covering $\mathcal{N}' \to \mathcal{N}$ over $\mathcal{N}$. Moreover, the morphism that assigns to such a covering $X \to Y$ the curve X defines a morphism $\mathcal{N}' \to (\mathcal{M}_{g,r})_k$. Thus, the locus in question is the image of a finite (by Lemma 5.5) number of such $\mathcal{N}' \to (\mathcal{M}_{g,r})_k$, hence is constructible. □

Thus, it follows from the proof of Lemma 5.6 that in order to prove Theorem 5.3, it suffices to prove that for all possible g_Y, r_Y, and Σ_Y, we have $3g-3+r>3g_Y-3+r_Y+|\Sigma_Y|$ (at least under the hypotheses placed on (g,r) in the statement of the theorem). We proceed to do this in the paragraphs that follow:

First, let us consider the case $g_Y \geq 1$. In this case, if we multiply the formula $2g-2+r=d\cdot e_Y$ by $\frac{3}{2}$ and then subtract $\frac{1}{2}r$, we obtain

$$\begin{aligned}
3g-3+r &= d\left\{3(g_Y-1)+\frac{3}{2}r_Y+\frac{3}{2}\sum_{\sigma\in\Sigma_Y} j_\sigma\right\}-\frac{1}{2}r \\
&= 3d(g_Y-1)+\frac{3d}{2}\left(\sum_{\sigma\in\Sigma_Y} j_\sigma\right)+dr_Y+\left(\frac{1}{2}dr_Y-\frac{1}{2}r\right) \\
&\geq 3(g_Y-1)+\frac{3d}{2}\left(\sum_{\sigma\in\Sigma_Y} j_\sigma\right)+dr_Y \\
&\geq 3g_Y-3+\frac{3\times 2}{2}\frac{1}{2}|\Sigma_Y|+dr_Y \\
&\geq 3g_Y-3+r_Y+|\Sigma_Y|.
\end{aligned}$$

(Here we use that $dr_Y \geq r$, $d \geq 2$, $j_\sigma \geq \frac{1}{2}$.) Moreover, if $g_Y \geq 2$, then the first "$\geq$" may be replaced with a "$>$", while if $g_Y = 1$ (so that $r_Y+|\Sigma_Y| \geq 1$), then the third "$\geq$" may be replaced with a "$>$". Thus, either way, we obtain that as long as $g_Y \geq 1$, we have $3g-3+r > 3g_Y-3+r_Y+|\Sigma_Y|$, as desired.

Now, we consider the case $g_Y=0$. First of all, just as above, we obtain that $3g-3+r \geq d(-3+r_Y+\frac{3}{2}\sum_\sigma j_\sigma)$. Since we wish to show that $3g-3+r > -3+r_Y+|\Sigma_Y|$, it suffices to show that the quantity

$$Q_Y \stackrel{\text{def}}{=} (d-1)(r_Y-3)+\left(\frac{3d}{4}-1\right)|\Sigma_Y|$$

is *positive*. (Here we use that $j_\sigma \geq \frac{1}{2}$.) Next, let us observe that if $|\Sigma_Y| \geq 7$, then $Q_Y \geq -3d+3+\frac{21}{4}d-7=\frac{9}{4}d-4 \geq \frac{1}{2} > 0$. Thus, it suffices to consider the case $|\Sigma_Y| \leq 6$. *Note that at this point, we still have not used the assumption that* $2g-2+r \geq 3$.

Now note that if $r=0$ and the desired inequality is false, then $r_Y=0$, so $3g-3 \leq -3+r_Y+|\Sigma_Y| \leq 3$, so $(g,r)=(2,0)$, but this case was ruled out in the hypothesis of the theorem. This completes the proof of the theorem when $r=0$.

Thus, for the rest of the proof, we assume that $r \neq 0$. Then $r_Y \neq 0$. Now if $|\Sigma_Y| \geq 5$, then $Q_Y \geq -2d+2+\frac{15}{4}d-5 \geq \frac{1}{2} > 0$. Thus, we obtain that $|\Sigma_Y| \leq 4$. Now if $r_Y \geq 3$, and the desired inequality is false, then we obtain (*under the assumption that* $(g,r) \neq (0,3)$) that $0 < 3g-3+r \leq (r_Y-3)+|\Sigma_Y|$, so it follows immediately that $Q_Y > 0$; thus, $r_Y \leq 2$. Thus, in summary, we have that $|\Sigma_Y| \leq 4$, $r_Y \leq 2$. Moreover, if $|\Sigma_Y| \in \{3,4\}$ and $r_Y=2$, then $Q_Y \geq 1-d+\frac{9}{4}d-3=\frac{5}{4}d-2 \geq \frac{1}{2} > 0$. Thus, in summary we see that $|\Sigma_Y|+r_Y \leq 5$, $(|\Sigma_Y|,r_Y) \neq (3,2)$. *Note that at this point (in our treatment of the case* $r \neq 0$*), the only assumption that we have used concerning* (g,r) *is that it is not equal to* $(0,3)$.

Now we invoke the assumption that $2g-2+r \geq 3$. If the desired inequality is false, then $3g-3+r \leq -3+r_Y+|\Sigma_Y| \leq 2$, so the only (g,r) that is still possible (and which is not ruled out in the hypothesis of the theorem) is $(g,r)=(0,5)$. Thus, for the rest of the proof, we assume that $(g,r)=(0,5)$.

It remains only to examine the case $|\Sigma_Y|=4$, $r_Y=1$. In this case, $Q_Y=d-2$, so $Q_Y \leq 0$ implies $d=2$. Thus, $5=r \leq dr_Y=2$, which is absurd. This completes the proof of Theorem 5.3. □

Remark. It is not difficult to check that in the exceptional cases (i.e., the cases where $2g-2+r \leq 2$) ruled out in Theorem 5.3, the hyperbolic core of a general curve is not equal to the core itself. Indeed, we have the following:

Theorem 5.7. *For a "general" (in the same sense as in the statement of Theorem 5.3) hyperbolic curve X of type (g,r), the canonical morphism $X \to Y$ to the hyperbolic core of X may be described as follows: $g_Y=0$ and*

(1) If $(g,r)=(0,4)$, then $X \to Y$ has degree 4, $r_Y=1$, $|\Sigma_Y|=3$, all the i_σ are 2, and the ramification index at the point at infinity of Y is 1.

(2) If $(g,r)=(1,1)$, then $X \to Y$ has degree 2, $r_Y=1$, $|\Sigma_Y|=3$, all the i_σ are 2, and the ramification index at the point at infinity of Y is 2.

(3) If $(g,r)=(1,2)$, then $X \to Y$ has degree 2, $r_Y=1$, $|\Sigma_Y|=4$, all the i_σ are 2, and the ramification index at the point at infinity of Y is 1.

(4) If $(g,r)=(2,0)$, then $X \to Y$ has degree 2, $r_Y=0$, $|\Sigma_Y|=6$, and all the i_σ are 2. Finally, if $(g,r)=(0,3)$, then X is arithmetic, so the hyperbolic core is not defined.

Proof. We continue computing with the notation at the end of the proof of Theorem 5.3. Thus, first of all, we have that $3g-3+r=r_Y+|\Sigma_Y|-3$. We begin with the case $(g,r)=(2,0)$. In this case, $r_Y=0$ and $|\Sigma_Y|=6$. Thus, $Q_Y=\frac{3}{2}d-3$, so $Q_Y \leq 0$ implies $d=2$. Since a general proper curve of genus 2 is well-known to be hyperelliptic, this completes the case of $(g,r)=(2,0)$.

Thus, it remains to consider those (g,r) for which $r \neq 0$. Let us also assume (until the second to last paragraph of the proof) that $(g,r) \neq (0,3)$. Then it follows from the proof of Theorem 5.3 that $g_Y=0$ and r_Y is either 1 or 2. Moreover, if $r_Y=2$, then $|\Sigma_Y|=2$, while if $r_Y=1$, then $|\Sigma_Y|$ is 3 or 4.

If $r_Y=2$ and $|\Sigma_Y|=2$, then $Q_Y=\frac{1}{2}d-1$, so $Q_Y \leq 0$ implies $d=2$. Thus, $2g-2+r=de_Y=2$, while $3g-3+r=r_Y+|\Sigma_Y|-3=1$, i.e., $(g,r)=(0,4)$. We shall see later that in fact, this case cannot arise under the assumption that $X \to Y$ is the canonical map defining Y as the hyperbolic core of X.

From now on, we assume that $r_Y=1$. If $|\Sigma_Y|=4$, then $Q_Y=d-2$, so $Q_Y \leq 0$ implies $d=2$. Thus, $2g-2+r=de_Y=2$, while $3g-3+r=r_Y+|\Sigma_Y|-3=2$, i.e., $(g,r)=(1,2)$.

If $|\Sigma_Y|=3$, then $Q_Y=\frac{1}{4}d-1$, so $Q_Y \leq 0$ implies $d \in \{2,3,4\}$. Since $3g-3+r=r_Y+|\Sigma_Y|-3=1$, it follows that $2g-2+r$ is 1 or 2. I claim that $d \neq 3$. Indeed, if d were 3, then all the i_σ would be $=3$, so we would get $2g-2+r=de_Y=3$, which is absurd. This proves the claim. Now observe that all the i_σ are equal to 2. Indeed, since the only possibilities for each i_σ are 2 and 4, if there were even one $i_\sigma \neq 2$, then we would have $d=4$, $e_Y \geq \frac{3}{4}$, so $3 \leq de_Y = 2g-2+r \in \{1,2\}$, which is absurd. Thus, all the $i_\sigma = 2$, as claimed. Moreover, $e_Y = \frac{1}{2}$, and $d = 2(2g-2+r)$. In other words, either $(g,r)=(0,4)$, in which case $d=4$, or $(g,r)=(1,1)$, in which case $d=2$.

Next, we pause to remark that it is not difficult to show that a general curve of type $(0,4)$ can actually be obtained as a degree four covering of a stack Y with $g_Y=0$, $r_Y=1$, $|\Sigma_Y|=3$, and all the $i_\sigma=2$. Indeed, consider the covering of $\mathbf{P}^1$ minus four points defined by the permutations (12)(34); (13)(24); (14)(23); id. (Note that the product of these permutations is the identity.) Here we think of the points corresponding to the first three permutations as the points at which Y is not a scheme, and the point corresponding to the last permutation as the point at infinity of Y. Thus, the existence of such a covering shows that the case $r_Y=2$, $|\Sigma_Y|=2$ (where the degree is necessarily 2, which is <4) could not arise under the assumption that the map $X \to Y$ is the canonical map defining Y as the hyperbolic core of X.

The only remaining case to consider is the case $(g,r)=(0,3)$. In this case, it is well known that X is arithmetic. (In fact, it appears as a finite étale covering of the moduli stack of elliptic curves.) Thus, the hyperbolic core is not defined.

Finally, we observe that it is easy to see that morphisms $X \to Y$ as stated in the theorem always exist. Thus, the above case analysis shows that such morphisms are necessarily the hyperbolic cores in each of the respective cases. This completes the proof of the theorem. □

Remark. It is not difficult to see that all the exceptional cases listed in Theorem 5.7 have the following property: a general curve X admits a correspondence ($\alpha : Z \to X$, $\beta : Z \to X'$) which is *nontrivial* in the sense that there does not exist a finite étale $\gamma : X' \to X$ such that $\alpha = \gamma \circ \beta$ (cf. Definition 1.2). Indeed, in the cases $(g,r)=(0,4)$, $(1,1)$, since the hyperbolic cores are of the same type, $(0,4)$-curves and $(1,1)$-curves provide "X"'s/"X'"'s for each other. Next, we consider the case $(g,r)=(1,2)$. If $X \to Y$ is the hyperbolic core of a general curve X of type $(1,2)$, then let $X' \to Y$ be the covering of Y of degree 4 defined by the permutations: (12)(34), (13)(24), (14)(23), (12)(34), (12)(34). (Here, one thinks of the first four permutations as describing the ramification over the four points of Y^c at which Y is not a scheme, and the last permutation as describing the ramification over the point at infinity of Y.) Then consideration of the inertia groups at the various points of X' shows that $X \to Y$ and $X' \to Y$ are linearly disjoint, so ($Z \stackrel{\text{def}}{=} X \times_Y X' \to X, Z \to X'$) gives the desired nontrivial correspondence. Finally, we consider the case $(g,r)=(2,0)$. In this case, we take for $X' \to Y$ the covering of degree 4 defined by the permutations: (12)(34), (13)(24), (14)(23), (12)(34), (13)(24), (14)(23). (Here, one thinks of these permutations as describing the ramification over the six points of Y^c at which Y is not a scheme.) Then consideration of the inertia groups at the various points of X' shows that $X \to Y$ and $X' \to Y$ are linearly disjoint, so ($Z \stackrel{\text{def}}{=} X \times_Y X' \to X, Z \to X'$) gives the desired nontrivial correspondence.

6. Interpretation of a theorem of Royden

Let (g,r) be a pair of nonnegative integers such that $2g-2+r>0$. Let $\mathcal{M}_{g,r}$ denote the moduli stack of r-pointed smooth (proper) curves of genus g. Here, the r marked points are *unordered*. (Note that this differs slightly from the usual convention.) The complement of the divisor of marked points of such a curve will be a hyperbolic curve of type (g,r). Thus, we shall also refer (by slight abuse of terminology) to $(\mathcal{M}_{g,r})_k$ as the moduli stack of (hyperbolic) curves of type (g,r).

Let us refer to as a *correspondence on* $\mathcal{M}_{g,r}$ an (ordered) pair of finite étale morphisms $\alpha : E \to \mathcal{M}_{g,r}$, $\beta : E \to \mathcal{M}_{g,r}$, where E is nonempty. We shall call a correspondence (α,β) on $\mathcal{M}_{g,r}$ *trivial* if $\alpha=\beta$. Note that this definition of what it means for a "correspondence on a (single) object" to be trivial is a bit different from the definition (Definition 1.2) that we gave earlier for what it means for a "correspondence from one object to another object" to be trivial.

Then we have the following result (essentially a consequence of a theorem of Royden):

Theorem 6.1. *Suppose that* $2g-2+r\geq 3$. *Then* $\mathcal{M}_{g,r}$ *is generically a scheme, and moreover, does not admit any nontrivial automorphisms or correspondences.*

Proof. Write $\mathcal{T}$ for the universal covering space of the analytic stack associated to the algebraic stack $\mathcal{M}_{g,r}$. Thus, $\mathcal{T}$ is what is usually referred to as "Teichmüller space". Let us write $\mathrm{Aut}(\mathcal{T})$ for the group of holomorphic automorphisms of $\mathcal{T}$, and Γ for the fundamental group of the analytic stack associated to $\mathcal{M}_{g,r}$. Thus, we have a natural morphism $\Gamma \to \mathrm{Aut}(\mathcal{T})$. According to a theorem of Royden [3, Section 9.2, p. 169, Theorem 2], this morphism is, in fact, an isomorphism (under the given hypotheses on (g,r)). The injectivity of this morphism implies that $\mathcal{M}_{g,r}$ is generically a scheme; the surjectivity of this morphism implies that $\mathcal{M}_{g,r}$ has no nontrivial automorphisms. Moreover, it is a matter of well-known general nonsense (see, e.g., the discussion of [4, p. 337]) that the existence of a nontrivial correspondence on $\mathcal{M}_{g,r}$ would imply the existence of an element of $\mathrm{Aut}(\mathcal{T})-\Gamma$ such that $\Gamma\cap(\gamma\Gamma\gamma^{-1})$ has finite index in Γ and in $\gamma\Gamma\gamma^{-1}$. Thus, we see that there are no nontrivial correspondences on $\mathcal{M}_{g,r}$. This completes the proof of the result. □

Remark. Note that the conclusion of the theorem is false in the exceptional cases ruled out in the hypothesis of the theorem.

Acknowledgements

I would like to thank Prof. Frans Oort for proposing the finiteness question answered in Theorem A to me during my stay at Utrecht University in November 1996, and for making various useful comments concerning both the substance and the expository style of this paper. Also, I would like to thank the University for its generosity and hospitality during my stay, and the Nederlandse Organizatie voor Wetenschappelijk Onderzoek for financial support. Finally, I would like to thank Prof. Y. Ihara for informing me of the paper [9] in the Spring of 1995.

References

[1] W.L. Baily Jr., On the automorphism group of a generic curve of genus > 2, J. Math. Kyoto Univ. 1, (1961/2) 101–108 (correction p. 325).
[2] G. Faltings, C.-L. Chai, Degenerations of Abelian Varieties, Springer, Berlin, 1990.
[3] F. Gardiner, Teichmüller Theory and Quadratic Differentials, Wiley, New York, 1987.
[4] G.A. Margulis, Discrete Subgroups of Semisimple Lie Groups, Ergebnisse der Mathematik under ihrer Grenzgebiete, vol. 17, Springer, Berlin, 1990.
[5] S. Mochizuki, The generalized ordinary moduli of p-adic hyperbolic curves, RIMS Preprint 1051 (1995).
[6] S. Mochizuki, Combinatorialization of p-adic Teichmüller theory, RIMS Preprint 1076 (1996).
[7] S. Mochizuki, A theory of ordinary p-adic curves, Publ. RIMS, Kyoto Univ. 32 (6) (1996).
[8] G. Shimura, Introduction to the Arithmetic Theory of Automorphic Forms, Publ. Math. Soc. of Japan, vol. 11, Iwanami Shoten and Princeton University Press, 1971.
[9] K. Takeuchi, Arithmetic Fuchsian groups with signature $(1;e)$, J. Math. Soc. Japan 35 (3) (1983) 381–407.