THE GEOMETRIC TORSION CONJECTURE FOR ABELIAN VARIETIES WITH REAL MULTIPLICATION

Benjamin Bakker & Jacob Tsimerman

Abstract

The geometric torsion conjecture asserts that the torsion part of the Mordell–Weil group of a family of abelian varieties over a complex quasi-projective curve is uniformly bounded in terms of the genus of the curve. We prove the conjecture for abelian varieties with real multiplication, uniformly in the field of multiplication. Fixing the field, we furthermore show that the torsion is bounded in terms of the *gonality* of the base curve, which is the closer analog of the arithmetic conjecture. The proof is a hybrid technique employing both the hyperbolic and algebraic geometry of the toroidal compactifications of the Hilbert modular varieties $\overline{X}(1)$ parametrizing such abelian varieties. We show that only finitely many torsion covers $\overline{X}_1(\mathfrak{n})$ contain d-gonal curves outside of the boundary for any fixed d; the same is true for entire curves $\mathbb{C} \to \overline{X}_1(\mathfrak{n})$. We also deduce some results about the birational geometry of Hilbert modular varieties.

Statement of Results

For any elliptic curve $E/\mathbb{Q}$, the group of rational points $E(\mathbb{Q})$ is finitely generated by Mordell's theorem. The free part behaves wildly; it is expected that there are elliptic curves $E/\mathbb{Q}$ with arbitrarily large rank $\operatorname{rk} E(\mathbb{Q})$, and the record to date is an elliptic curve E with $\operatorname{rk} E(\mathbb{Q}) \geq 28$ found by Elkies. On the other hand, by a celebrated theorem of Mazur [**MG78**] the torsion part $E(\mathbb{Q})_{tor}$ is uniformly bounded:

Theorem (Mazur)**.** *For any elliptic curve $E/\mathbb{Q}$, $|E(\mathbb{Q})_{tor}| \leq 16$.*

Mazur's theorem was subsequently generalized to arbitrary number fields $K/\mathbb{Q}$ by Merel [**Mer96**] (building on partial results of [**Kam92**]) who showed a stronger uniformity: there is an integer $N = N(d)$ such that for any degree d number field K and any elliptic curve E/K, every K-rational torsion point has order dividing N, *i.e.* $E(K)_{tor} \subset E(K)[N]$.

Similarly, it is expected that the torsion part of the Mordell–Weil group of an abelian variety A/K is uniformly bounded, though there are few results in this direction. The same question can be asked for

Received June 27, 2017.

the function field $K = k(C)$ of a curve C over any field k: $k = \overline{\mathbb{F}}_p$ is most closely analogous to the number field case, but the $k = \overline{\mathbb{Q}}$ is also interesting:

Conjecture (Geometric torsion conjecture)**.** *Let k be an algebraically closed field of characteristic 0. There is an integer $N = N(g, n)$ such that for any quasi-projective genus g curve C/k and any family of n-dimensional abelian varieties A/C with no isotrivial part, the torsion of the Mordell–Weil group is uniformly bounded:*

$$A(C)_{tor} \subset A(C)[N].$$

Here $A(C)$ is the group of rational sections. By the standard argument, it suffices to consider $k = \mathbb{C}$, and we do so for the remainder. The geometric conjecture is also largely open, though some recent progress has been made by Cadoret and Tamagawa [**CT11**, **CT12**]. Their technique however only applies over a *fixed* base C.

The main goal of this paper is to prove the conjecture for abelian varieties with real multiplication:

Theorem A. *There is an integer $N = N(g, n)$ such that for any quasi-projective genus g complex curve C and any non-isotrival family of n-dimensional abelian varieties A/C with real multiplication, the torsion part of the Mordell–Weil group is uniformly bounded:*

$$A(C)_{tor} \subset A(C)[N].$$

An abelian variety with real multiplication is an n-dimensional abelian variety A together with an injection $\mathcal{O}_F \to \mathrm{End}(A)$ of the ring of integers $\mathcal{O}_F$ in a totally real field $F/\mathbb{Q}$ of degree n. The constant in Theorem A only depends on F through the dimension n and thus has the same uniformity as in the conjecture. There are no reduction hypotheses in Theorem A, so the result is really a statement about abelian varieties over (characteristic 0) function fields:

Corollary B. *For fixed n and g, there are only finitely many (finite) groups occurring as the rational torsion $A(K)_{tor}$ of a non-isotrivial n-dimensional abelian variety A/K with real multiplication over the function field K/k of a genus g curve over k.*

Note that with the real multiplication hypothesis, non-isotriviality is equivalent to having no isotrivial part.

We also prove a version of the conjecture uniformly in the *gonality* of the base curve, at the expense of a dependence on the field of multiplication:

Theorem C. *Fix a totally real field F. There is an integer $N = N(d, F)$ such that for any quasi-projective d-gonal complex curve C and* *any non-isotrivial family of abelian varieties A/C with real multiplication by $\mathcal{O}_F$, the torsion part of the Mordell–Weil group is uniformly bounded:*

$$A(C)_{tor} \subset A(C)[N].$$

Of course, there is a corresponding version of Corollary B. Uniformity in the *gonality* of the base curve C is strictly stronger than uniformity in the genus; gonality is the correct function field analog of the degree of the number field in Merel's theorem. Recall that, for a curve C the gonality $\mathrm{gon}(C)$ is the minimum degree of a (dominant) map to $\mathbb{P}^1$, and since the map $C^{(g+1)} \to \mathrm{Pic}^{g+1}(C)$ has relative dimension 1, we at least[1] have $\mathrm{gon}(C) \leq g(C) + 1$. On the other hand, for any $d > 0$ there are d-gonal curves of genus g for all $g \geq 2d - 3$. Similarly, while there are infinitely many d-gonal curves $C/\mathbb{F}_q$, there are only finitely many for any fixed genus. This should be viewed as the function field analog of the fact that there are infinitely many degree d number fields $K/\mathbb{Q}$ but only finitely many of bounded discriminant.

Our proofs of Theorems A and C are geometric. For a fixed F, there is a Hilbert modular variety $X(1)$ parametrizing abelian varieties with real multiplication by $\mathcal{O}_F$, and a map $C \to X(1)$ is equivalent to a family of such abelian varieties over C. For any ideal $\mathfrak{n} \subset \mathcal{O}_F$, there is a level cover $X_1(\mathfrak{n})$ parametrizing abelian varieties A with real multiplication by $\mathcal{O}_F$ together with a point $x \in A$ whose annihilator is $\mathfrak{n}$. To prove Theorem A, we show that $X_1(\mathfrak{n})$ uniformly contains no genus g curves for $|\,\mathrm{Nm}(\mathfrak{n})|$ large.

Theorem D. *For each $\mathfrak{n} \subset \mathcal{O}_F$ let $X_1(\mathfrak{n})^*$ be the Baily–Borel compactification of the $\mathfrak{n}$-torsion level cover of the Hilbert modular variety $X(1)$. Then for any g, $X_1(\mathfrak{n})^*$ contains no genus g curves for all but finitely many $\mathfrak{n}$, uniformly for all F of a fixed degree. For a fixed F, the same is true of d-gonal curves.*

The proof is conceptually similar to that of [**HT06**], where a version of Theorem D is shown for *full-level* covers $X(\mathfrak{n})$. However, the boundary of $X_1(\mathfrak{n})$ does not totally ramify over $X(1)$ and therefore a new technique is required. The core idea is to prove a bound relating the volume of a curve in a toroidal compactification $\overline{X}_1(\mathfrak{n})$ to its multiplicity along the boundary (see Proposition 2.8) using the metric geometry. For larger arithmetic lattices, intersection with the boundary comes at the price of more volume, which in turn implies larger genus. The bound is proven by constructing a positive singular metric on the log canonical bundle of $\overline{X}_1(\mathfrak{n})$ whose singular support is concentrated on the boundary. For the gonality statement, we must generalize the framework to

[1] In fact, by the Brill–Noether therem $\mathrm{gon}(C) \leq \lfloor \frac{g(C)+1}{2} \rfloor$ with equality for C generic of genus $g(C)$.

higher dimensional locally symmetric varieties which contain nontrivial orbifold loci.

Our method applies more generally to toroidal compactifications $\overline{X}$ of any variety uniformized by $\mathbb{H}^n$, and the above construction has many implications to the birational geometry of these varieties. In particular it implies that the slice of the big cone generated by the canonical bundle $K_{\overline{X}}$ and the boundary grows with the "size" of the lattice. For instance, we have the following:

Theorem E. *Assume $X_1(\mathfrak{n})$ has no elliptic points and let $\overline{X}_1(\mathfrak{n})$ be a smooth toroidal compactification with boundary divisor D. Then for any $\lambda > 0$, we have that $K_{\overline{X}_1(\mathfrak{n})} - (\lambda - 1)D$ is ample modulo the boundary provided $|\operatorname{Nm}(\mathfrak{n})| > (\frac{2\pi\lambda}{n})^{2n}$.*

See Corollary 2.11 for more precise bounds on the ample modulo D cone of any toroidal compactification $\overline{X}$ of a finite-volume quotient of $\mathbb{H}^n$ by an irreducible lattice. Recall that we say that a divisor is ample modulo D if its augmented base locus is contained in D. We note that $X_1(\mathfrak{n})$ uniformly has no elliptic points for large $|\operatorname{Nm}(\mathfrak{n})|$ (see Lemma 3.1).

For some geometric consequences of Theorem E, see Section 4. In particular, we immediately have the following:

Corollary F. *With the hypotheses of Theorem E, $X_1(\mathfrak{n})$ is of general type provided $|\operatorname{Nm}(\mathfrak{n})| > (\frac{2\pi}{n})^{2n}$.*

By essentially taking $\mathfrak{n} = \mathcal{O}_F$, we recover a result of [**Tsu85**]:

Corollary G. *$X(1)$ is of general type provided $n > 6$.*

It is worth noting that the proof of Tsuyumine relies entirely on the theory of modular forms, whereas our proof only involves the metric geometry.

Given Corollary F, the Green–Griffiths conjecture then predicts that all entire curves $\mathbb{C} \to \overline{X}_1(\mathfrak{n})$ have image contained in a strict algebraic subvariety (called the exceptional locus) for those $\mathfrak{n}$. By a theorem of Nadel [**Nad89**], Theorem E indeed implies this is the case:

Corollary H. *Every nontrivial entire curve $\mathbb{C} \to \overline{X}_1(\mathfrak{n})$ has image contained in the boundary provided $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^{2n}$.*

This provides an explicit bound in the genus 0 or 1 case of Theorem A, as well as a similar boundedness statement for entire families of abelian varieties with real multiplication. The Green–Griffiths conjecture for the base Hilbert modular varieties $\overline{X}(1)$ has been addressed recently by [**RT15**] using modular forms and foliation theory, where the conjecture is proven for all but finitely many choices of F. This clearly implies the conjecture for the toroidal compactification $\overline{X}'$ of *any* cover $X' \to X(1)$, although the exceptional locus is not explicit. Corollary H says that the exceptional locus is in fact (contained in) the boundary sufficiently high in the torsion tower, uniformly in F.

$X_1(\mathfrak{n})$ is defined over $\mathbb{Q}$, and the *arithmetic* torsion conjecture for abelian varieties with real multiplication can likewise be phrased as the nonexistence of K-rational points of $X_1(\mathfrak{n})$ for all but finitely many $\mathfrak{n}$, for any number field K. Theorem D and Corollary H are geometric and analytic analogs asserting the nonexistence of rational points valued in the function field of a curve and the field of meromorphic functions on $\mathbb{C}$, respectively. All three are conjecturally related: in particular, assuming the Bombieri–Lang conjecture, Corollary H implies $X_1(\mathfrak{n})(K)$ is *finite* for any number field K.

We finally note that Theorem A (and Corollary B) can be made effective. We also expect Theorem C to be true uniformly in $X(1)$, and that the same idea for the proof should work with some modifications. The methods investigated here apply more generally to rank one lattices (see [**BT15**] for an application to complex ball quotients), and we expect the torsion conjecture in the case of abelian varieties parametrized by a rank one Shimura variety to be proven similarly.

Outline. The proof follows the general strategy developed in [**BT16**] to prove the geometric analog of another arithmetic uniformity conjecture, the Frey–Mazur conjecture. In Section 1 we detail the local structure of cusps of cofinite-volume quotients of $\mathbb{H}^n$ and their toroidal compactifications. We then prove a volume bound on the boundary multiplicity of curves in the toroidal compactification in Section 2. These bounds are similar to those proven by [**HT02**] for *interior* points. In Section 3 we show that the torsion covers of Hilbert modular varieties hyperbolically "expand," from which it follows that the multiplicity bound of Section 1 improves in the torsion tower. We then deduce some geometric results in Section 4 including Theorem E and Corollaries F and H. In Section 5 we assemble the previous results to prove the first part of Theorem D and conclude Theorem A (and Corollary B). Finally, in Section 6 we prove a volume bound on the diagonal multiplicity of curves in products of Hilbert modular varieties and use it to prove the second part of Theorem D, and thus Theorem C. The presence of degenerations complicates the analysis of the diagonal, and the eventual bound we obtain in Theorem C is not uniform in the field of multiplication F.

Acknowledgements. The first named author would like to thank G. van der Geer and A. J. de Jong for enlightening discussions.

1. Lattices in $\mathrm{SL}_2(\mathbb{R})^n$

Let $\mathbb{H} = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$ be the upper half-plane. $G = \mathrm{SL}_2(\mathbb{R})^n$ is the group of holomorphic automorphisms of $\mathbb{H}^n$ acting by Möbius transformations component-wise. A discrete cofinite-volume subgroup

$\Gamma \subset G$ is *nondegenerate* if it is not commensurable to a product, and in this section we will only consider such Γ. In this case, Γ has only isolated parabolic fixed points, called cusps, and there are only finitely many up to the action by Γ. The Baily–Borel compactification X^* of the quotient $X = \Gamma\backslash\mathbb{H}^n$ is obtained by adding a point to compactify each equivalence class of cusps; it is a normal projective variety [**BB66**]. Note that for $n > 1$, Γ is commensurable to an arithmetic lattice (up to conjugation) by a theorem of Margulis [**Mar84**].

Local models of the cusp. The Siegel model $\mathbb{H}^n$ has a preferred cusp at $\infty := (\infty, \dots, \infty) \in \mathbb{H}^n$, whose parabolic stabilizer U_∞ is the upper triangular matrices. For $z \in \mathbb{H}^n$, define $N(z) := \prod_i (\operatorname{Im} z_i)$. Natural neighborhoods of the cusp ∞ are given by the sets

$$U(s) = \{z \in \mathbb{H}^n | N(z) > 1/s\}$$

and we refer to $U(s)$ as the horoball around ∞ of depth s.

Given a cusp $*$ of Γ, we may move it to ∞ by conjugating Γ. This is only unique up to conjugation by an upper-triangular element

$$\gamma = \begin{pmatrix} a_1 & b_1 \\ 0 & a_1^{-1} \end{pmatrix} \times \cdots \times \begin{pmatrix} a_n & b_n \\ 0 & a_n^{-1} \end{pmatrix}$$

Note that this sends the horoball $U(s)$ to $\gamma \cdot U(s) = U(s')$ with $s' = s/\prod_i a_i^2$, as $N(\gamma \cdot z) = N(z)\prod_i a_i^2$.

The stabilizer $\Gamma_\infty = \Gamma \cap U_\infty$ of ∞ has unipotent radical Λ_∞ a lattice of real translations. N yields a norm form Nm on Λ_∞, and we can identify $\mathbb{H}^n$ with

$$\mathbb{H}_\infty := \{\zeta \in \Lambda_\infty \otimes \mathbb{C} \mid (\operatorname{Im}\zeta)_i > 0 \text{ for all } i\} \subset \Lambda_\infty \otimes \mathbb{C}$$

with Λ_∞ acting by translations by $\Lambda_\infty \otimes 1$. Here $(\operatorname{Im}\zeta)_i$ is the ith coordinate of $\operatorname{Im}\zeta$, which is not defined up to conjugation by γ, as it scales by a_i^2, but its positivity is. The norm on Λ_∞ clearly scales by $\prod_i a_i^2$.

Thus, we can scale so that the shortest vector of Λ_∞ has length 1, and we thereby associate to any cusp $*$ of Γ with stabilizer Γ_* and unipotent radical Λ_* a canonically determined norm form Nm_* on Λ_* normalized by the condition that the length of the shortest vector is 1. The coordinates $\sigma_*^i : \Lambda_* \to \mathbb{R}$ for which

$$\mathrm{Nm}_*(\lambda) = \prod_i \sigma_*^i(\lambda)$$

are not well-defined as they scale individually, but as above their positivity is. We also have a canonically determined function $N_*(\zeta) = \prod_i \sigma_*^i(\operatorname{Im}\zeta)$ defined on the canonical horoball $U_*(s)$, which has the form

$$U_*(s) := \{\zeta \in \Lambda_* \otimes \mathbb{C} \mid N_*(\zeta) > 1/s\}.$$

Definition 1.1. For a cusp $*$ of Γ, let $W_*(s) := \Gamma_* \backslash U_*(s)$. We say the horoball $U_*(s)$ is *precisely invariant* if $W_*(s)$ injects into $\Gamma\backslash\mathbb{H}^n$. The *canonical depth* s_* of $*$ is the largest s such that $U_*(s)$ is precisely invariant. When there's no chance of confusion we just refer to s_* as the depth.

The following lemma gives us a bound on the canonical depth.

Lemma 1.2. *For $\Gamma \subset G$ nondegenerate, suppose s is the minimal nonzero value of $\prod_i |c_i|$ over all group elements*

$$\gamma = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \times \cdots \times \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \in \Gamma$$

Then $U(s)$ is precisely invariant under Γ.

Proof. We need to show that if $\gamma \cdot z = w$ for some $\gamma \in \Gamma$ and $z, w \in U(s)$ then γ is upper triangular. We have

$$N(w) = \prod_i \frac{y_i}{|c_i z_i + d_i|^2}.$$

Note that since Γ is nondegenerate, one of the c_i is zero only if all of them are zero. If they are not all zero, then as $|c_i z_i + d_i| \geq |c_i||z_i| \geq |y_i|$ we get

$$N(w) \leq \frac{1}{\prod_i |c_i|^2 N(z)} \leq \frac{1}{\prod_i |c_i|}$$

which is a contradiction.

q.e.d.

Toroidal compactifications. We briefly describe the geometry of the toroidal compactifications of X constructed by [**AMRT10**].

For a cusp $*$ of Γ, the partial quotient of $\mathbb{H}^n$ by Λ_* naturally sits in the torus $T_* = \Lambda_* \backslash \Lambda_* \otimes \mathbb{C}$, and there is a log map valued in $\Lambda_* \otimes \mathbb{R}$ defined by taking the imaginary part:

$$\begin{array}{ccccccc} \mathbb{H}_* & \hookrightarrow & \Lambda_* \otimes \mathbb{C} & \longrightarrow & T_* & \xrightarrow{\log} & \Lambda_* \otimes \mathbb{R} \\ & & & & & & \\ & & \lambda \otimes z & & \longmapsto & & \lambda \otimes \operatorname{Im} z \end{array}$$

The coordinate ring of T_* is canonically $\mathbb{C}[\Lambda_*^\vee]$, via the identification between $\Lambda_*^\vee$ and the character group $\operatorname{Hom}(T_*, \mathbb{G}_m)$. For an element $\chi \in \Lambda_*^\vee$ we denote the corresponding character by the symbol q^χ, concretely given by

$$q^\chi(\lambda \otimes z) = e(\chi(\lambda) z)$$

and for any $t \in T_*$, we have

$$\chi(\log(t)) = -\frac{1}{2\pi} \log |q^\chi(t)|$$

The function N_* descends (and extends) to the torus T_*, and in fact further descends to $\Lambda_* \otimes \mathbb{R}$; it is given by

$$N_*(t) = \left(\frac{-1}{2\pi}\right)^n \prod_i \log |q^{\sigma_*^i}(t)| = \mathrm{Nm}_*(\log t)$$

where on the right hand side we mean Nm_* extended to $\Lambda_* \otimes \mathbb{R}$ in the obvious way. The horoballs $U_*(s)$ are likewise stable under the action of Λ_* and we let $V_*(s)$ be the image in T_*. Note that $\log(V_*(s))$ lies inside the positive cone $C(\Lambda_* \otimes \mathbb{R})$ defined by

$$C(\Lambda_* \otimes \mathbb{R}) = \{x \in \Lambda_* \otimes \mathbb{R} \mid \sigma_*^i(x) > 0\}$$

where we extend σ_*^i $\mathbb{R}$-linearly. Importantly, $C(\Lambda_* \otimes \mathbb{R})$ is *not* an integral cone.

The group $\Delta_* := \Gamma_*/\Lambda_*$ acts on $C(\Lambda_* \otimes \mathbb{R})$. A toroidal compactification at $*$ is specified by a subdivision of $C(\Lambda_* \otimes \mathbb{R})$ into a fan of integral polyhedral cones $\Sigma_* = \{\tau\}$ that is stable under the action of Δ_*. The compactification is smooth if and only if each full-dimensional τ is generated by an integral basis of Λ_*, and any fan can by sufficiently subdivided to yield a smooth compactification. For each full-dimensional cone τ, this provides us with a coordinate chart of the compactification which looks like a neighborhood of $0 \in \mathbb{A}^n$ with coordinates $q^{\chi_1}, \ldots, q^{\chi_n}$, where the χ_i are the basis dual to the basis of Λ_* given by the primitive generators of the rays of τ, and in this chart the boundary is the union of the coordinate axes $q^{\chi_i} = 0$.

Taking s smaller than the depth of $*$, the quotient $W_*(s) = \Delta_* \backslash V_*(s)$ injects into the toroidal compactification, and N_* descends to $W_*(s)$. We therefore obtain a function $N_* : W_*(s) \to [-\infty, \infty)$ by declaring N_* to be $-\infty$ along the boundary. To compute the Lelong number of $N_*^{1/n}$, we use the $\mathbb{A}^n$ neighborhoods of the boundary in the toroidal compactification. Let $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, and for $z \in \mathbb{C}^n$ denote $|z|^{\mathbf{a}} = \prod_i |z_i|^{a_i}$. Then we have:

Lemma 1.3.

$$\liminf_{z \to x} \frac{\prod_i \log |z|^{\mathbf{a}^{(i)}}}{\log^n |z - x|} = \prod_j \sum_{\substack{i \\ x_i = 0}} a_i^{(j)}.$$

In the surface case this means for $x_2 \neq 0$

$$\liminf_{(z_1,z_2)\to(0,x_2)} \frac{\sqrt{(a_1 \log |z_1| + a_2 \log |z_2|)(b_1 \log |z_1| + b_2 \log |z_2|)}}{\log \sqrt{|z_1|^2 + |z_2 - x_2|^2}} = \sqrt{a_1 b_1}$$

and

$$\liminf_{(z_1,z_2)\to(0,0)} \frac{\sqrt{(a_1\log|z_1|+a_2\log|z_2|)(b_1\log|z_1|+b_2\log|z_2|)}}{\log\sqrt{|z_1|^2+|z_2|^2}} = \sqrt{(a_1+a_2)(b_1+b_2)}$$

For any cone $\tau \in \Sigma_*$, let $Z(\tau)$ be the corresponding T_*-orbit (*not* the orbit closure).

Lemma 1.4. *The Lelong number $\nu(-N_*^{1/n}, x)$ at a point x in the boundary component compactifying $*$ is constant along $Z(\tau)$ and we have*

$$\nu(-N_*^{1/n}, x) = \frac{1}{2\pi}\,\mathrm{Nm}_*(\lambda(\tau))^{1/n}, \text{ for all } x \in Z(\tau)$$

where $\lambda_i \in \Lambda_*$ are the primitive generators of the rays of τ and $\lambda(\tau) = \sum \lambda_i$.

In the future we denote $\nu(-N_*^{1/n}, \tau) := \nu(-N_*^{1/n}, x)$ for $x \in Z(\tau)$.

Proof. Let τ_0 be a full-dimensional cone with associated basis $\lambda_1, \dots, \lambda_n$ of Λ_*, so if $\chi_1, \dots, \chi_n$ is the dual basis, then the corresponding coordinates are $x_i = q^{\chi_i}$. Faces τ of τ_0 are indexed by subsets $S \subset \{1, \dots, n\}$, where the corresponding face is generated by the rays $\{\lambda_s | s \in S\}$, and the associated orbit is locally the coordinate plane given by the intersection of $x_s = 0$ for all $s \in S$. We have

$$\sigma_*^i = \sum_j \sigma_*^i(\lambda_j)\chi_j$$

and therefore in these coordinates

$$\log|q^{\sigma_*^i}(t)| = \log\prod_j |x_j|^{\sigma_*^i(\lambda_j)}$$

Thus, since

$$\prod_i \sum_{s\in S} \sigma_*^i(\lambda_s) = \prod_i \sigma_*^i(\lambda) = \mathrm{Nm}_*(\lambda(\tau))$$

for $\lambda(\tau) = \sum_{s\in S} \lambda_s$, the result follows from the previous lemma.

q.e.d.

2. Curves in quotients of $\mathbb{H}^n$

Metric geometry. Endow $\mathbb{H}$ with its hyperbolic hermitian metric $ds^2_{\mathbb{H}}$ of constant sectional curvature -1; explicitly, the associated Kähler form is

$$\omega_{\mathbb{H}} = \frac{1}{2}\,\mathrm{Im}\, ds^2_H = \frac{i dz \wedge d\overline{z}}{2y^2} = i\partial\overline{\partial}(-2\log y)$$

Likewise endow $\mathbb{H}^n$ with the invariant metric

$$ds^2_{\mathbb{H}^n} = \sum_i \pi_i^* ds^2_{\mathbb{H}}$$

where $\pi_i : \mathbb{H}^n \to \mathbb{H}$ is the ith projection, and we again denote by $\omega_{\mathbb{H}^n}$ the associated Kähler form. Note that $-2\log N$ (and therefore also $-2\log N_*$) is a global potential for $\omega_{\mathbb{H}^n}$. The distance function we use on $\mathbb{H}^n$ is the Kobayashi distance—namely, the distance between $z, w \in \mathbb{H}^n$ is the maximum of the distances $d_{\mathbb{H}}(z_i, w_i)$ of the coordinate projections, where $d_{\mathbb{H}}$ is the usual hyperbolic distance on $\mathbb{H}$:

$$\left|\frac{z-w}{z-\overline{w}}\right|^2 = \tanh^2(d_{\mathbb{H}}(z,w)/2)$$

For $\Gamma \subset G$ a discrete nondegenerate cofinite-volume subgroup and $X = \Gamma\backslash\mathbb{H}^n$, let ds^2_X be the induced metric with Kähler form ω_X. For $\overline{X}$ a smooth toroidal compactification of X, a theorem of Mumford [**Mum77**] tells us that the current $[\omega_X] \in H^{1,1}(\overline{X}, \mathbb{R})$ defined by integration against ω_X on the open part X is represented by a multiple of the log-canonical bundle:

$$\mathrm{c}_1(K_{\overline{X}} + D) = \frac{1}{2\pi}[\omega_X] \in H^{1,1}(\overline{X}, \mathbb{R}) \tag{1}$$

Take $f : C \to \overline{X}$ a map from a smooth proper curve of genus $g(C)$ whose image is not contained in the boundary, and let $U \subset C$ be the open subset mapping to the interior X.

Lemma 2.1. *In the above situation,*

$$\frac{1}{n}(K_{\overline{X}} - (n-1)D).C \leq 2g(C) - 2$$

Proof. U is necessarily uniformized by $\mathbb{H}$, so let ds^2_U be it's uniformized metric of constant curvature -1. By Schwarz's Lemma,

$$\frac{1}{n} f^* ds^2_X \leq ds^2_U$$

Integrating the Kähler forms, we get by Gauss–Bonnet

$$\frac{1}{2\pi n}\,\mathrm{vol}_X(U) \leq -\chi(U) = -\chi(C) + |C \smallsetminus U| \tag{2}$$

The left-hand side is $\frac{1}{n}(K_{\overline{X}} + D).C$, while $|C \smallsetminus U| \leq D.C$, and the result follows. q.e.d.

Remark 2.2. Lemma 2.1 still holds in the form of (2) with the same proof if Γ has elliptic elements, as long as we treat X and C as orbifolds and $\chi(C)$ as the orbifold Euler characteristic.

Multiplicity bounds. For the rest of this section, we additionally assume Γ is torsion-free. We begin by recalling a theorem of Hwang–To which says that the volume of a curve in a neighborhood of a point in the interior of a quotient of $\mathbb{H}^n$ (or in fact any bounded symmetric domain) is bounded by its multiplicity at the point.

Theorem 2.3 (See [**HT02**])**.** *Let $x \in X$ and take $B(x,r)$ the Kobayashi ball around x of radius*[2] *$r < \rho_x$. Then for any irreducible k-dimensional subvariety $Y \subset X$, we have*

$$\mathrm{vol}_X(Y \cap B(x,r)) \geq \frac{(4\pi)^k}{k!} \sinh^{2k}(r/2) \cdot \mathrm{mult}_x Y$$

By (1) this theorem allows us to estimate the degree of a curve $C \subset \overline{X}$ not meeting the boundary in terms of its multiplicity at a point $x \in X$:

$$(K_{\overline{X}} + D).C = K_{\overline{X}}.C = \frac{1}{2\pi n} \mathrm{vol}_X(C \cap X) \geq \frac{2}{n} \sinh^2(\rho_x) \cdot \mathrm{mult}_x C \tag{3}$$

The main goal of this section is to provide a similar volume bound on the multiplicity along the boundary, as well as a relative version.

For any cusp $*$ of Γ, denote by $\widetilde{W}_*(s)$ the interior closure of $W_*(s)$ in $\overline{X}$ for s smaller than the canonical depth s_* of $*$.

Proposition 2.4. *Let $*$ be a cusp of Γ. Then for any irreducible k-dimensional analytic variety Y in $\widetilde{W}_*(s_*)$ not contained in the boundary,*

$$s^{-k/n} \mathrm{vol}_X(Y \cap W_*(s))$$

is an increasing function of s for $s < s_*$.

Before the proof we recall a lemma of Demailly [**Dem12**]. Let X be a complex manifold and $\varphi : X \to [-\infty, \infty)$ a continuous plurisubharmonic function. Define

$$B_\varphi(r) = \{x \in X \mid \varphi(x) < r\}$$

We say φ is semi-exhaustive if the balls $B_\varphi(r)$ have compact closure in X. Further, for T a closed positive current of type (k,k), we say φ is semi-exhaustive on $\mathrm{Supp}\, T$ if $\varphi \not\equiv -\infty$ on $\mathrm{Supp}\, T$ and each $B_\varphi(r) \cap \mathrm{Supp}\, T$ has compact closure. In this case, the integral

$$\int_{B_\varphi(r)} T \wedge (i\partial\overline{\partial}\varphi)^k := \int_{B_\varphi(r)} T \wedge (i\partial\overline{\partial} \max(\varphi, s))^k$$

is well-defined and independent of $s < r$ [**Dem12**, §III.5]. We then have the following:

Lemma 2.5 (Formula III.5.5 of [**Dem12**])**.** *For any convex increasing function $f : \mathbb{R} \to \mathbb{R}$,*

$$\int_{B(r)} T \wedge (i\partial\overline{\partial} f \circ \varphi)^k = f'(r-0)^k \int_{B(r)} T \wedge (i\partial\overline{\partial}\varphi)^k$$

where $f'(r-0)$ is the derivative of f from the left at r.

We also need:

Lemma 2.6. *$-N^{1/n}$ is plurisubharmonic on $\mathbb{H}^n$.*

[2] ρ_x is the injectivity radius, see Section 3.

Proof. Let I_n be the identity $n\times n$ matrix and Z_n be the matrix which has 0's along the diagonal, and 1's elsewhere. Now note that I_n+tZ_n is positive semi-definite exactly when $-\frac{1}{n-1}\leq t\leq 1$—this can be seen by computing its determinant.

For any function $g:\mathbb{R}\to\mathbb{R}$, we compute by the chain rule

$$\begin{aligned} i\partial\overline{\partial}g(N) &= g''(N)\cdot i\partial N\wedge\overline{\partial}N+g'(N)\cdot i\partial\overline{\partial}N \\ &= \frac{1}{4}N^2g''(N)(I_n+Z_n)+\frac{1}{4}Ng'(N)Z_n \end{aligned}$$

in the renormalized basis $(\operatorname{Im} z_i)\frac{\partial}{\partial z}$. Setting $g(x)=-x^t$, we get

$$\begin{aligned} i\partial\overline{\partial}(-N^t) &= -\frac{t(t-1)N^t}{4}\left((I_n+Z_n)+\frac{1}{t-1}Z_n\right) \\ &= \frac{t(1-t)N^t}{4}\left(I_n-\frac{t}{1-t}Z_n\right) \end{aligned}$$

which is positive semi-definite for $t=1/n$. q.e.d.

Proof of Proposition 2.4.

$$\begin{aligned} \operatorname{vol}_X(Y\cap W_*(s)) &= \frac{1}{k!}\int_{Y\cap W_*(s)}\omega_X^k \\ &= \frac{1}{k!}\int_{Y\cap W_*(s)}(i\partial\overline{\partial}(-2\log N_*))^k \\ &= \frac{1}{k!}\int_{W_*(s)}[Y]\wedge(i\partial\overline{\partial}(-2\log N_*))^k \\ &= \frac{(2n)^ks^{k/n}}{k!}\int_{W_*(s)}[Y]\wedge(i\partial\overline{\partial}(-N_*^{1/n}))^k \\ &= \frac{(2n)^ks^{k/n}}{k!}\int_{Y\cap W_*(s)}(i\partial\overline{\partial}(-N_*^{1/n}))^k \end{aligned}$$

As $-N_*^{1/n}$ is plurisubharmonic,

$$s^{-k/n}\operatorname{vol}_X(Y\cap W_*(s))=\frac{(2n)^k}{k!}\int_{Y\cap W_*(s)}(i\partial\overline{\partial}(-N_*^{1/n}))^k$$

is an increasing function of s. q.e.d.

Remark 2.7. Proposition 2.4 is also valid in the orbifold context just by pulling up to a finite neat cover. It is optimal in the sense that for Y a the projection of a linear geodesic $\mathbb{H}\subset\mathbb{H}^n$ (no coordinate of which is constant), $s^{-1/n}\operatorname{vol}(Y\cap W_*(s))$ will be constant. Indeed, in this case $N_*^{1/n}$ restricts to a multiple of y, which is a potential for (a multiple of) the current of integration at the cusp.

Taking the limit of Proposition 2.4 as $s \to 0$ and using the results from Section 1, we obtain a multiplicity bound in the style of Theorem 2.3.

Proposition 2.8. *Let $*$ be a cusp of Γ with canonical depth s_*. Then for any irreducible analytic curve C in $\widetilde{W}_*(s_*)$ not contained in the boundary and any $s < s_*$,*

$$\frac{1}{n}\operatorname{vol}_X(C \cap W_*(s)) \geq \sum_{\tau \in \Sigma_*} (s \cdot \mathrm{Nm}_*(\lambda(\tau)))^{1/n} \operatorname{mult}_{Z(\tau)} C$$

Proof. From the proof of the previous proposition, we have

$$\frac{1}{n}\operatorname{vol}_X(C \cap W_*(s)) \geq 2s^{1/n} \cdot \lim_{s \to 0} \int_{C \cap W_*(s)} i\partial\overline{\partial}(-N_*^{1/n})$$

For s sufficiently small, $C \cap \widetilde{W}_*(s)$ is a union of pure 1-dimensional analytic sets, each component of which is normalized by a disk $f : \Delta \to C \cap \widetilde{W}_*(s)$. We may assume $f(0) \in Z(\tau)$ and that $f|_{\Delta^*}$ is an isomorphism onto an open set of C. If $\chi_1, \ldots, \chi_s$ are the dual basis to the generators $\lambda_1, \ldots, \lambda_s$ of τ, then write $m_j = \operatorname{ord} f^* q^{\chi_j}$. Note that $m = \min(m_j)$ is the contribution of the branch f to $\operatorname{mult}_{Z(\tau)} C$.

Now for sufficiently small s, we have

$$\int_{C \cap \widetilde{W}_*(s)} i\partial\overline{\partial}(-N_*^{1/n}) = \sum_f \int_\Delta f^* i\partial\overline{\partial}(-N_*^{1/n})$$

but of course

$$\begin{aligned}\int_\Delta f^* i\partial\overline{\partial}(-N_*^{1/n}) &\geq \pi\nu(f^*(-N_*^{1/n}), 0)\\ &\geq \pi m\nu(-N_*^{1/n}, \tau)\\ &= \frac{m}{2}\mathrm{Nm}_*(\lambda(\tau))^{1/n}\end{aligned}$$

by Lemma 1.4, and the claim follows.

q.e.d.

Remark 2.9. In fact we can obtain a more precise estimate if we remember more of the local behavior of C. In the notation of the proof, for the branch $f : \Delta \to C \cap \widetilde{W}_*(s)$ let t be a uniformizer in the local ring of the disk at 0. We have

$$\begin{aligned}2\pi\nu(f^*(-N_*^{1/n}), 0) &= \lim_{t \to 0} \frac{\prod_i \log|f^* q^{\sigma_*^i}|}{\log^n |t|}\\ &= \lim_{t \to 0} \frac{\prod_i \sum_j \sigma_*^i(m_j\lambda_j)\log|t|}{\log^n |t|}\\ &= \mathrm{Nm}_*\left(\sum m_j\lambda_j\right)\end{aligned}$$

which is a version of the multiplicity weighted by the local geometry of $Z(\tau)$.

Remark 2.10. Proposition 2.8 is also optimal in the sense that the geodesic in Remark 2.7 will realize the equality with the coefficient from Remark 2.9.

The Lelong number jumps up along smaller T_*-orbits in the boundary. Indeed, for any totally positive $\lambda_i \in \Lambda_*$

$$\mathrm{Nm}_*\left(\sum \lambda_i\right) \geq \sum_i \mathrm{Nm}_*(\lambda_i).$$

For the main theorems in Sections 5 and 6, Proposition 2.8 is sufficient, but with a slightly more detailed analysis we can package the above result into the positivity of a large slope divisor. The main point is that the proof of Corollary 2.13 can alternatively be seen as explicitly constructing a singular hermitian metric on $K_{\overline{X}} - tD$ as a twist of the canonical singular metric coming from the uniformized geometry. Recall that a line bundle M on a smooth variety Y is ample modulo a subvariety $Z \subset Y$ (see for example Section 5 of [**DCDC15**]) if either: (1) a sufficiently large multiple of M defines a map to $\mathbb{P}^N$ which is an embedding on $Y \setminus Z$; (2) the augmented base locus of M is contained in Z.

Proposition 2.11. *For a fixed cusp $*$, let ρ run through the (nonequivalent) one-dimensional cones of Σ_* and let D_ρ be the corresponding irreducible boundary divisor. Then*

$$(K_{\overline{X}} + D) - \frac{n}{2\pi} \sum_{\rho \in \Sigma_*} \left(s_* \mathrm{Nm}_*(\rho)\right)^{1/n} \cdot D_\rho$$

is ample modulo the boundary.

Proof. Let $L = K_{\overline{X}} + D$. Since L is big and the big cone is the interior of the pseudo-effective cone, it suffices to show that $M = L - E$ is pseudo-effective where $E = \sum a_\rho D_\rho$ for appropriately chosen $a_\rho \in \mathbb{Q}$. By [**Dem92**] its enough for (a multiple of) $L - E$ to possess a singular hermitian metric h with chern form $\mathrm{c}_1(M, h) \geq 0$ (as a current). Taking the canonical metric on D_ρ given by a section (for which $\mathrm{c}_1(D_\rho) = [D_\rho]$), this is achieved if we construct such an h on L with $\mathrm{c}_1(L, h) \geq \sum a_\rho [D_\rho]$.

To finish, we need the following lemma, which is easily proven along the lines of [**HT02**]. Recall that a potential for $\omega_{\mathbb{H}^n}$ is given by $-2\log N$.

Lemma 2.12. *Fix $S_0 > 0$. For any $\epsilon > 0$, there exists a smooth function $\varphi : \mathbb{R}_{>0} \to \mathbb{R}$ such that:*

a) $\varphi(N) = -2\log N$ *outside of* $[1/S_0 - \epsilon, \infty)$*;*
b) $i\partial\overline{\partial}\varphi(N) \geq 0$*;*
c) $\varphi(N) \sim -S_0^{1/n} N^{1/n}$ *as* $N \to \infty$*.*

By [Mum77], the hyperbolic metric on K_X extends to a singular metric h_0 on L whose curvature is $-i\partial\overline{\partial}\log h_0 = [\omega_X]$. For any particular cusp of $\overline{X}$, the function N_* is defined in the horoball neighborhood $W(s_*)$, we define $h = e^{-\varphi(N_*)+2\log N_*}h_0$ on $W(s_*)$, and $h = h_0$ on the complement of all the cuspidal neighborhoods. By the Lemma and the above Lelong number computations, h satisfies the desired properties. q.e.d.

Provided we have for each cusp $*$ a depth $t_* < s_*$ such that the horoball neighborhoods $W_*(t_*)$ are disjoint, then we can apply the same argument to each cusp simultaneously:

Corollary 2.13. *For t_* chosen as above,*

$$(K_{\overline{X}} + D) - \frac{n}{2\pi}\sum_{*}\sum_{\rho\in\Sigma_*}(t_*\,\mathrm{Nm}_*(\rho))^{1/n}\cdot D_\rho \tag{4}$$

is ample modulo the boundary. In particular,

$$(K_{\overline{X}} + D) - \frac{n}{2\pi}\sum_{*} t_*^{1/n}\cdot D_*$$

is ample modulo the boundary.

Proof. By definition, Nm_* is at least 1 on nonzero integral lattice points. q.e.d.

As mentioned above, for the main result all we will need is that the above divisors are nef modulo the boundary—that is, nonnegative on curves not contained in the boundary.

Remark 2.14. We note here that for quotients with orbifold points in dimension $n > 1$, the orbifold points are isolated and therefore the above conclusions are still valid.

3. Hilbert modular varieties

Fix F a totally real field of degree n with ring of integers $\mathcal{O}_F \subset F$ and denote the real embeddings by $\sigma_i : F \hookrightarrow \mathbb{R}$. We will typically suppress $\mathcal{O}_F, F$ from the notation as much as possible. For more a more detailed account of Hilbert modular groups, we refer the reader to [vdG88, Gor02].

Hilbert modular groups. To any projective rank 2 module[3] M over $\mathcal{O}_F$ we can associate a Hilbert modular group $\mathrm{SL}(M) \subset G = \mathrm{SL}_2(\mathbb{R})^n$ after choosing an isomorphism $M \otimes_{\sigma_i} \mathbb{R} \cong \mathbb{R}^2$ for each embedding σ_i. Any such M is isomorphic to $M \cong \mathcal{O}_F \oplus \mathfrak{a}$ for an ideal $\mathfrak{a} \subset \mathcal{O}_F$, and therefore up to conjugation we need only consider $\Gamma(1) := \mathrm{SL}(\mathcal{O}_F \oplus$

[3] Strictly speaking M should be endowed with a symplectic pairing as well.

$\mathfrak{a}$) where the embedding in G is obtained on the ith factor from the embedding $\mathcal{O}_F \oplus \mathfrak{a} \to \mathbb{R}^2$ via σ_i.

Our eventual goal is to prove Theorem A uniformly in the choice of F and M for fixed $[F:\mathbb{Q}]$, and we will use the phrase "uniformly in $\Gamma(1)$" to mean uniformity in this sense. Note that Hilbert modular groups exist for non-maximal orders as well, though we do not pursue uniformity in this level of generality here.

For any ideal $\mathfrak{n} \subset \mathcal{O}_F$, we define the congruence subgroups

$$\Gamma_0(\mathfrak{n}) := \left\{ A \in \Gamma(1) \mid A \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \bmod \mathfrak{n} \right\}$$

$$\Gamma_1(\mathfrak{n}) := \left\{ A \in \Gamma(1) \mid A \equiv \begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix} \bmod \mathfrak{n} \right\}$$

of elements $A \in \Gamma(1)$ which fix a primitive line or vector mod $\mathfrak{n}$, respectively. Concretely, $\Gamma_0(\mathfrak{n})$ consists of determinant 1 matrices $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ for which $\alpha, \delta \in \mathcal{O}_F$, $\beta \in \mathfrak{a}^{-1}$, and $\gamma \in \mathfrak{a}\mathfrak{n}$.

The Hilbert modular stack associated to $\Gamma(1)$ is the quotient $X(1) := \Gamma(1)\backslash\mathbb{H}^n$, and we also have level covers

$$X_0(\mathfrak{n}) := \Gamma_0(\mathfrak{n})\backslash\mathbb{H}^n \qquad\qquad X_1(\mathfrak{n}) := \Gamma_1(\mathfrak{n})\backslash\mathbb{H}^n$$

$X(1), X_0(\mathfrak{n})$, and $X_1(\mathfrak{n})$ are *a priori* analytic Deligne–Mumford stacks.

Let Γ be any of the Hilbert modular groups described above. The Baily–Borel compactification $(\Gamma\backslash\mathbb{H}^n)^*$ of the coarse space of $\Gamma\backslash\mathbb{H}^n$ is obtained by adding a finite set of points corresponding to the equivalences classes of the rational boundary components of $\mathbb{H}^n$ under the action of Γ. The Baily–Borel compactifications are normal projective varieties, so in fact $X(1), X_0(\mathfrak{n}), X_1(\mathfrak{n})$ are smooth algebraic Deligne–Mumford stacks. For our purposes the distinction between $X_1(\mathfrak{n})$ and its coarse space is unnecessary since for $|\operatorname{Nm}\mathfrak{n}|$ sufficiently large (uniformly in $\Gamma(1)$), $X_1(\mathfrak{n})$ will have no stabilizers:

Lemma 3.1. *For $\mathfrak{n} \subset \mathcal{O}_F$ with $|\operatorname{Nm}(\mathfrak{n})| > 4^n$, $X_1(\mathfrak{n})$ has no elliptic points.*

Proof. Suppose $\gamma \in \Gamma_1(\mathfrak{n})$ fixes a point of $\mathbb{H}^n$. Then the eigenvalues of γ must be roots of unity. However, the characteristic polynomial of γ is $x^2 - \alpha x + 1$ where $\alpha \in \mathcal{O}_F$ with $\alpha - 2 \in \mathfrak{n}$. It thus follows that the absolute value of the norm of $(\alpha - 2)$, if non-zero, is at least as large as the norm of $\mathfrak{n}$. However, since α is the sum of two roots of unity it is of absolute value at most 2 in any archimedean embedding, and thus has norm at most 4^n. By our assumption on $\mathfrak{n}$ the norm of $(\alpha - 2)$ must be 0 and thus $\alpha = 2$. It follows that γ is unipotent, and for γ to have fixed points we conclude that γ is the identity element, as desired.

q.e.d.

Moduli of abelian varieties with real multiplication. We briefly describe the moduli interpretation of the Hilbert modular stacks introduced in the previous subsection.

Definition 3.2. Let S be a scheme. An *abelian scheme with real multiplication* by $\mathcal{O}_F$ over S is an abelian scheme A/S of (relative) dimension n together with an injection $\mathcal{O}_F \to \mathrm{End}_S(A)$. We simply say A/S has real multiplication if it has real multiplication by $\mathcal{O}_F$ for some totally real field F.

Again, we restrict ourselves to the case of multiplication by the maximal order here.

Over $\mathbb{C}$, the Hilbert modular stack $X(1)$ associated to $\Gamma(1) = \mathrm{SL}(M)$ is naturally identified with the moduli stack of abelian varieties A with real multiplication by $\mathcal{O}_F$ such that $H_1(A,\mathbb{Z}) \cong M$ as $\mathcal{O}_F$-modules[4] . Analytically, the isomorphism is described as follows; for simplicity take $M = \mathcal{O}_F \oplus \mathfrak{a}$. To any $\tau = (\tau_1, \dots, \tau_n) \in \mathbb{H}^n$ we can associate a lattice

$$L := \mathcal{O}_F \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \mathfrak{a} \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_n \end{pmatrix} \subset \mathbb{C}^n$$

where $\mathcal{O}_F$ (and $\mathfrak{a}$) acts on the ith factor by multiplication via the embedding σ_i. The complex torus $A = \mathbb{C}^n/L$ evidently has an action $\mathcal{O}_F \to \mathrm{End}(A)$ with $H_1(A,\mathbb{Z}) \cong \mathcal{O}_F \oplus \mathfrak{a}$, and it can be shown that any n-dimensional complex torus with a (faithful) action by $\mathcal{O}_F$ admits a polarization.

Let $\mathfrak{n} \subset \mathcal{O}_F$ be an ideal. For A an abelian variety with real multiplication by $\mathcal{O}_F$, we define the $\mathfrak{n}$-torsion $A[\mathfrak{n}]$ to be the sub-group scheme annihilated by $\mathfrak{n}$. For two such abelian varieties A, A' a cyclic $\mathfrak{n}$-isogeny is an isogeny $f : A \to A'$ such that the induced map $\mathrm{End}(A)_{\mathbb{Q}} \to \mathrm{End}(A')_{\mathbb{Q}}$ is a map of $\mathcal{O}_F$-modules and the sub-group scheme $\ker f$ is cyclic and contained in $A[\mathfrak{n}]$. Over a characteristic 0 algebraically closed field, a $\mathfrak{n}$-isogeny is equivalent to specifying a cyclic $\mathcal{O}_F/\mathfrak{n}$ submodule of $A[\mathfrak{n}]$. As above, over $\mathbb{C}$ we can identify $X_0(\mathfrak{n})$ (resp. $X_1(\mathfrak{n})$) with the stack of abelian varieties with real multiplication by $\mathcal{O}_F$ and a cyclic $\mathfrak{n}$-isogeny from A (resp. a point of $A[\mathfrak{n}]$ with annihilator $\mathfrak{n}$).

Canonical depths. The rational boundary components of $\Gamma(1)$ (and its subgroups) are naturally identified with $\mathbb{P}(M_F) \cong \mathbb{P}^1(F)$, on which $\mathrm{SL}_2(F)$ acts via the standard action. For an ideal $\mathfrak{n} \subset \mathcal{O}_F$, we would like to show that the canonical depths of the cusps of $\Gamma_0(\mathfrak{n})$ grow uniformly in $|\mathrm{Nm}(\mathfrak{n})|$. For any fractional ideal $\mathfrak{b} \subset F$, denote by $|\mathfrak{b}|$ the smallest nonzero value of $|\mathrm{Nm}(x)|$ for $x \in \mathfrak{b}$. Clearly we have $|\mathfrak{b}| \geq |\mathrm{Nm}(\mathfrak{b})|$.

[4]Again, strictly speaking M should be endowed with a symplectic pairing, in which case the isomorphism is one of polarized $\mathcal{O}_F$-modules.

Given two distinct (possibly equivalent) cusps $\xi_1 = (\alpha : \beta)$, $\xi_2 = (\gamma : \delta) \in \mathbb{P}^1(F)$ of $\Gamma_0(\mathfrak{n})$, we may assume by clearing denominators that both have integral coordinates. Consider the matrix

$$T = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$$

and set $\Delta = \alpha\delta - \beta\gamma$. By conjugating

$$T^{-1}\Gamma_0(\mathfrak{n})T$$

we have moved ξ_1 to ∞ and ξ_2 to 0. Let $\mathfrak{b}_1 = \alpha\mathfrak{a} + (\beta)$ and $\mathfrak{b}_2 = \gamma\mathfrak{a} + (\delta)$.

Lemma 3.3. *The unipotent stabilizer of* ∞ *in* $T^{-1}\Gamma_0(\mathfrak{n})T$ *is*

$$\begin{pmatrix} 1 & \Delta\mathfrak{a}(\mathfrak{b}_1^{-2} \cap \beta^{-2}\mathfrak{n}) \\ & 1 \end{pmatrix}$$

and that of 0 *is*

$$\begin{pmatrix} 1 & \\ \Delta\mathfrak{a}(\mathfrak{b}_2^{-2} \cap \delta^{-2}\mathfrak{n}) & 1 \end{pmatrix}$$

Proof. We need to know what matrices $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \in G$ conjugate back to elements of $\Gamma_0(\mathfrak{n})$, and we find

$$T\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}T^{-1} = \begin{pmatrix} 1 - \frac{\alpha\beta}{\Delta}t & \frac{\alpha^2}{\Delta}t \\ -\frac{\beta^2}{\Delta}t & 1 + \frac{\alpha\beta}{\Delta}t \end{pmatrix} \tag{5}$$

which is in $\Gamma(1)$ if and only if $t \in \Delta\mathfrak{a}\mathfrak{b}_1^{-2}$, and additionally in $\Gamma_0(\mathfrak{n})$ if we have $t \in \Delta\mathfrak{a}\beta^{-2}\mathfrak{n}$ as well. Similarly for 0 and elements $\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} \in G$.

q.e.d.

Now the horoball

$$U(s) = \{z \in \mathbb{H}^n \mid N(z) > 1/s\}$$

around ∞ is mapped under the inversion $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ to the horoball

$$U^{-1}(s) := \left\{ z \in \mathbb{H}^n \mid \prod_i |z_i|^2 < s \cdot N(z) \right\}$$

and furthermore that $U(s) \cap U^{-1}(s') = \varnothing$ if and only if $ss' \leq 1$.

The *canonical* horoballs at ξ_1 and ξ_2 are then given in the above coordinates by

$$U_{\xi_1}(s) = U\left(\frac{s}{|\Delta\mathfrak{a}(\mathfrak{b}_1^{-2} \cap \beta^{-2}\mathfrak{n})|}\right)$$

and

$$U_{\xi_2}(s) = U^{-1}\left(\frac{s}{|\Delta\mathfrak{a}(\mathfrak{b}_2^{-2} \cap \delta^{-2}\mathfrak{n})|}\right).$$

Lemma 3.4. *If ξ_1 and ξ_2 are distinct (possibly equivalent) cusps of $\Gamma_0(\mathfrak{n})$, the canonical horoballs $U_{\xi_1}(|\operatorname{Nm}(\mathfrak{n})|^{1/2})$ and $U_{\xi_2}(|\operatorname{Nm}(\mathfrak{n})|^{1/2})$ are disjoint. In particular, every cusp of $\Gamma_0(\mathfrak{n})$ has canonical depth at least $|\operatorname{Nm}(\mathfrak{n})|^{1/2}$.*

Proof. We know the horoballs $U_{\xi_1}(s_1)$ and $U_{\xi_2}(s_2)$ are disjoint if

$$s_1 s_2 \leq |\Delta\mathfrak{a}(\mathfrak{b}_1^{-2} \cap \beta^{-2}\mathfrak{n})||\Delta\mathfrak{a}(\mathfrak{b}_2^{-2} \cap \delta^{-2}\mathfrak{n})|$$

so it is sufficient to show that the right hand side is at least of size $|\operatorname{Nm}(\mathfrak{n})|$. Now, recall that for two non-zero fractional ideals I, J, we have $(I \cap J)^{-1} = I^{-1} + J^{-1}$, and also that by factorization we have that $(I + J)^2 = I^2 + J^2$. Thus

$$(\mathfrak{b}_1^2 + \beta^2\mathfrak{n}^{-1}) \cdot (\mathfrak{b}_2^2 + \delta^2\mathfrak{n}^{-1}) = (\alpha^2\mathfrak{a}^2 + \beta^2\mathfrak{n}^{-1}) \cdot (\gamma^2\mathfrak{a}^2 + \delta^2\mathfrak{n}^{-1})$$

since we have equality of the two first factors (and the two second factors). The left hand side contains $\Delta^2\mathfrak{a}^2\mathfrak{n}^{-1}$, and

$$\begin{aligned}
|\Delta\mathfrak{a}(\mathfrak{b}_1^{-2} \cap \beta^{-2}\mathfrak{n})||\Delta\mathfrak{a}(\mathfrak{b}_2^{-2} \cap \delta^{-2}\mathfrak{n})| &= |\Delta\mathfrak{a}(\mathfrak{b}_1^2 + \beta^2\mathfrak{n}^{-1})^{-1}||\Delta\mathfrak{a}(\mathfrak{b}_2^2 + \delta^2\mathfrak{n}^{-1})^{-1}| \\
&\geq \frac{|\operatorname{Nm}(\Delta\mathfrak{a})|^2}{|\operatorname{Nm}(\mathfrak{b}_1^2 + \beta^2\mathfrak{n}^{-1})||\operatorname{Nm}(\mathfrak{b}_2^2 + \delta^2\mathfrak{n}^{-1})|} \\
&\geq |\operatorname{Nm}(\mathfrak{n})|
\end{aligned}$$

as desired. q.e.d.

Now consider the étale cover $X_1(\mathfrak{n}) \to X_0(\mathfrak{n})$. Disjoint precisely invariant horoballs pull back to disjoint precisely invariant horoballs, and the canonical depth can only increase in covers. Thus we have the following:

Corollary 3.5. *Each cusp of $X_1(\mathfrak{n})$ has canonical depth at least $|\operatorname{Nm}(\mathfrak{n})|^{1/2}$. Moreover, the horoball neighborhoods $W_*(|\operatorname{Nm}(\mathfrak{n})|^{1/2})$ are pairwise disjoint.*

Remark 3.6. Lemma 3.4 is far from optimal. For example, it is not difficult to show by a similar analysis that for a prime $\mathfrak{p}$ the cusps of $X_0(\mathfrak{p})$ that ramify over $X(1)$ have canonical depth $|\operatorname{Nm}(\mathfrak{p})|$, and therefore the same is true for all cusps (since an involution interchanges the ramifying and non-ramifying cusps of $X_0(\mathfrak{p})$, though possibly corresponding to a different choice of M). These horoball neighborhoods will however intersect, albeit at most with multiplicity two. Furthermore, Corollary 3.5 is even less optimal because it is pulled back from $X_0(\mathfrak{n})$. For our purposes we only need the uniform growth.

Finally, using Corollary 2.13 we conclude that a divisor of growing slope on $X_1(\mathfrak{n})$ is nef modulo the boundary, uniformly in $\Gamma(1)$:

Proposition 3.7. *For $\mathfrak{n} \subset \mathcal{O}_F$ such that $X_1(\mathfrak{n})$ has no elliptic points,*

a) *Each cusp of* $X_1(\mathfrak{n})$ *has canonical depth at least* $|\operatorname{Nm}(\mathfrak{n})|^{1/2}$*;*

b)
$$K_{\overline{X}_1(\mathfrak{n})} + \left(1 - \frac{n}{2\pi}|\operatorname{Nm}(\mathfrak{n})|^{1/2n}\right) D$$
is ample modulo the boundary.

Proof. The second statement follows from Corollary 2.13. q.e.d.

Remark 3.8. By a minor modification of the proof of Lemma 3.4, we see that Proposition 3.7 is true for the principal cover $X(\mathfrak{n})$ with $|\operatorname{Nm}(\mathfrak{n})|^{1/2}$ replaced by $|\operatorname{Nm}(\mathfrak{n})|$.

Injectivity radii. The final ingredient of] the proof of the main theorem is some uniform estimates on the injectivity radii of Hilbert modular groups. Recall that for any discrete $\Gamma \subset G = \mathrm{SL}_2(\mathbb{R})^n$, the injectivity radius ρ at a point $x \in X = \Gamma\backslash\mathbb{H}^n$ is defined as half the length of the smallest loop through x, where the length is taken with respect to the Kobayashi distance on X. We can write

$$\rho_x = \frac{1}{2} \inf_{1\neq\gamma\in\Gamma} d_{\mathbb{H}^n}(\widetilde{x}, \gamma\cdot\widetilde{x})$$

for $\widetilde{x} \in \mathbb{H}^n$ any lift of x. Equivalently, ρ_x is the largest radius r such that the Kobayashi ball $B(x, r) \subset \mathbb{H}^n$ of radius r around x injects in the quotient X. The global injectivity radius ρ_X is the global infimum:

$$\rho_X := \inf_{x\in X} \rho_x$$

For a group Γ with cusps, $\rho_X = 0$ because unipotent elements of Γ correspond to homotopy classes of geodesics wrapping the cusps, and the length of these clearly go to 0. We therefore define the semi-simple injectivity radii by only considering semi-simple elements $\Gamma^{ss} \subset \Gamma$: for $1 \neq \gamma \in \Gamma^{ss}$, define

$$\rho_\gamma := \frac{1}{2} \inf_{z\in\mathbb{H}^n} d_{\mathbb{H}^n}(z, \gamma\cdot z)$$

and then

$$\rho_X^{ss} := \inf_{1\neq\gamma\in\Gamma^{ss}} \rho_\gamma.$$

Lemma 3.9. *For* $\gamma \in G$ *a semi-simple element with largest eigenvalue* λ*,*

$$\rho_\gamma \geq \log|\lambda|$$

Proof. A semisimple $\gamma \in \mathrm{SL}_2(\mathbb{R})$ can be conjugated to a scaling by the square $a = |\lambda|^2$ of its largest eigenvalue, in which case

$$\begin{aligned}\inf_z d_{\mathbb{H}}(z, \gamma z) &= \inf_z d_{\mathbb{H}}(z, az)\\ &= \inf_z \operatorname{arcosh}\left(1 + \frac{(a-1)^2|z|^2}{2a(\operatorname{Im} z)^2}\right)\\ &\geq \log a\end{aligned}$$

Since the Kobayashi distance on $\mathbb{H}^n$ is the maximum of the coordinate-wise distances, we are done. q.e.d.

Corollary 3.10. $\rho^{ss}_{X_1(\mathfrak{n})} \geq \frac{1}{n}\log|\operatorname{Nm}(\mathfrak{n})| - 1$.

Proof. An element $\gamma \in \Gamma_1(\mathfrak{n})$ reduces to $\begin{pmatrix}1 & * \\ 0 & 1\end{pmatrix} \bmod \mathfrak{n}$; if γ is not unipotent, neither eigenvalue is 1. If $\sigma_i(\alpha)$ is the largest eigenvalue of $\sigma_i(\gamma)$ over all i, then since $\alpha - 1 \in \mathfrak{n}$ we have

$$\begin{aligned}|\sigma_i(\alpha)| &\geq |\operatorname{Nm}(\alpha - 1)|^{1/n} - 1 \\ &\geq |\operatorname{Nm}(\mathfrak{n})|^{1/n} - 1\end{aligned}$$

q.e.d.

For any cusp $*$ with unipotent stabilizer Λ_*, we also define at each point $x \in X$

$$\rho^*_x := \frac{1}{2}\inf_{\gamma\in\Lambda_*} d_{\mathbb{H}^n}(\widetilde{x}, \gamma\cdot\widetilde{x})$$

We'd like to continue thinking of $1/N_*$ as the "distance" to $*$, but N_* isn't defined on all of X. If s_* is the canonical depth of $*$, we can simply define N_* to be $1/s_*$ outside of $\widetilde{W}_*(s_*)$.

Lemma 3.11. $\rho^*_x \geq N_*(x)^{-1/n}\cdot(1 + O(N_*(x)^{-1/n}))$.

Proof. For a real translation by a in $\mathbb{H}$, we have

$$d_{\mathbb{H}}(z, z+a) = 2\tanh^{-1}\left(\frac{a}{\sqrt{a^2 + 4(\operatorname{Im} z)^2}}\right)$$

and so

$$d_{\mathbb{H}}(z, z+a) = \frac{a}{\operatorname{Im} z}\cdot(1 + O(a/\operatorname{Im} z))$$

in a region where $\operatorname{Im} z$ is bounded away from 0. Thus for $\lambda \in \Lambda_*$,

$$d_{\mathbb{H}^n}(z, z+\lambda) \geq \frac{\operatorname{Nm}_*(\lambda)^{1/n}}{N_*(z)^{1/n}}\cdot(1 + O(\operatorname{Nm}_*(\lambda)^{1/n}/N_*(z)^{1/n}))$$

By definition the smallest nonzero value of $\operatorname{Nm}_*(\lambda)$ is 1, and the result follows. q.e.d.

4. Geometric consequences

The results of the previous two sections allow us to conclude some interesting geometric results, including Theorem E and Corollaries F, G, and H. These results are not necessary for the proof of the main theorem in the following section.

We choose once and for all smooth toroidal compactifications $\overline{X}_1(\mathfrak{n})$ for all $X(1)$ and $\mathfrak{n} \subset \mathcal{O}_F$ (for which $X_1(\mathfrak{n})$ has no elliptic points). Each of the following results has a strengthening for *principal* covers $\overline{X}(\mathfrak{n})$ using Remark 3.8 which for the most part we leave implicit.

The following is an immediate consequence of Proposition 3.7:

Proposition 4.1. *Assume $X_1(\mathfrak{n})$ has no elliptic points. For any $\lambda > 0$, $K_{\overline{X}_1(\mathfrak{n})} - (\lambda - 1)D$ is ample modulo the boundary provided $|\operatorname{Nm}(\mathfrak{n})| > (\frac{2\pi\lambda}{n})^{2n}$.*

Note that by Lemma 3.1 we may uniformly assume $X_1(\mathfrak{n})$ has no elliptic points. By a general result of Tai [**AMRT10**], an arithmetic quotient $\Gamma\backslash\Omega$ of a bounded symmetric domain has an étale cover which is of general type. Proposition 4.1 implies almost all of the torsion covers of Hilbert modular varieties are of general type:

Corollary 4.2. *Assume $X_1(\mathfrak{n})$ has no elliptic points. Then it is of general type provided $|\operatorname{Nm}(\mathfrak{n})| > \left(\frac{2\pi}{n}\right)^{2n}$.*

In particular, this will be true of *every* torsion cover without elliptic points for $n \geq 6$ (as in this case $2 > (\frac{2\pi}{n})^{2n}$). More generally, we need only assume the singularities are canonical, which they are by [**Tsu85**] provided $n > 6$. By Remark 3.6 the canonical depths of $X(1)$ itself are at least 1, and therefore we conclude:

Corollary 4.3. *$X(1)$ is of general type provided $n > 6$.*

There is also an interesting consequence pertaining to the hyperbolicity of the torsion covers $\overline{X}_1(\mathfrak{n})$. Recall that the Green–Griffiths conjecture asserts that for any general type variety X, there is a strict subvariety $Z \subset X$ such that every entire map $\mathbb{C} \to X$ has image contained in Z.

Corollary 4.4. *Every entire map $\mathbb{C} \to \overline{X}_1(\mathfrak{n})$ has image contained in the boundary provided $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^{2n}$.*

Proof. By a theorem of Nadel [**Nad89**, Theorem 2.1], an entire map $\mathbb{C} \to \overline{X}$ into a toroidal compactification of a quotient of a bounded symmetric domain must be contained in the boundary as soon as $K_{\overline{X}} + (1 - 1/\gamma)D$ is big, where the sectional curvature is bounded from above by $-\gamma$ (with the normalization $\operatorname{Ric}(h) = -h$). For us $\gamma = 1/n$ is sufficient, and by Proposition 3.7 we see that $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^{2n}$ is enough. By Lemma 3.1, this rules out elliptic points as well. q.e.d.

We can deduce from Corollary 4.4 the genus 0 and 1 case of the geometric torsion conjecture. Our proof in the next section for this special case will be independent of Corollary 4.4.

Corollary 4.5. *Every rational or elliptic curve in $\overline{X}_1(\mathfrak{n})$ is contained in the boundary provided $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^{2n}$. The same is true of $\overline{X}(\mathfrak{n})$ provided $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^n$.*

This improves a result of Freitag [**Fre80**] that the statement is true for sufficiently large principal congruence groups with constant depending on $X(1)$. In the $n = 2$ case, a Hilbert modular surface X possesses a canonical smooth model Y that has no -1 curves in the cuspidal or elliptic resolutions, but Y is often not minimal. Hirzebruch and Zagier [**HZ77**] conjectured that as long as the surface is irrational, the minimal model is obtained by blowing down "known" curves—that is, modular curves or curves arising from the desingularization. Van der Geer proves that in fact the principal congruence covers $Y(\mathfrak{n})$ are minimal [**vdG79**] for $|\operatorname{Nm}(\mathfrak{n})|$ larger than a computable constant depending on $X(1)$, and we obtain a uniform improvement:

Corollary 4.6. *For $n = 2$, $Y_1(\mathfrak{n})$ is minimal provided $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^4$. The same is true of $Y(\mathfrak{n})$ provided $|\operatorname{Nm}(\mathfrak{n})| > (2\pi)^2$.*

5. Geometric torsion: uniformity in genus

Let C be a quasi-projective curve over $k = \mathbb{C}$. The generic fiber of an abelian scheme A/C yields an abelian variety over the function field $A/k(C)$, and conversely any abelian variety $A/k(C)$ yields an abelian scheme A/U over an open set $U \subset C$. We define the Mordell–Weil group to be the group of sections, or equivalently the groups of rational sections $A(C) = A(k(C))$, and denote by $A(C)_{tor}$ the torsion subgroup. An abelian scheme A/C is isotrivial if the fibers A_c are generically isomorphic, or equivalently if $A/k(C)$ is the base-change of an abelian variety over $\mathbb{C}$, and A/C is said to have no isotrivial part if it has no isotrivial isogeny factor. Note that an abelian scheme A/C with real multiplication has an isotrivial part only if it is isotrivial. By a folklore result, abelian schemes A/C with no isotrivial part have finitely generated Mordell–Weil group $A(C)$.

Given our current setup, to prove Theorems A and D it will be enough to show the following:

Theorem 5.1. *Let $\overline{X}_1(\mathfrak{n})$ be a smooth toroidal compactification of $X_1(\mathfrak{n})$ for any n-dimensional $X(1)$. For any g, every curve $C \to \overline{X}_1(\mathfrak{n})$ with (geometric) genus $g(C) < g$ is contained in the boundary for all but finitely many $\mathfrak{n}$, uniformly in $X(1)$.*

In the following we'll use the phrase "uniformly in $X(1)$" if the statement holds for all $X(1)$ with constant only depending on the dimension n (and not on the choices of the field F and the module M). For any $\lambda > 0$, define

$$\ell(K_{\overline{X}} - \lambda D) = \inf_{C \not\subset D} (K_{\overline{X}} - \lambda D).C$$

where the infimum is taken over all integral curves not contained in the boundary. We show:

Proposition 5.2. *For any $\lambda, B > 0$, we have*

$$\ell(K_{\overline{X}_1(\mathfrak{n})} - \lambda D) > B$$

for all but finitely many $\mathfrak{n}$*, uniformly in* $X(1)$*.*

Theorem 5.1 then follows from Proposition 5.2 and Lemma 2.1 by taking $\lambda = n - 1$.

Proof of Proposition 5.2. For curves $C \to \overline{X}_1(\mathfrak{n})$ not contained in the boundary we define

$$d(C) := \sup_{x \in C} \sup_{*} \frac{1}{N_*(x)}$$

We divide such curves into three categories depending on how close to the boundary they are:

i) Those meeting the boundary: $d(C) = 0$.
ii) Those not meeting the boundary with a point far from the boundary: $d(C) \geq E$, for some $E > 0$ to be determined.
iii) Those entirely lying close to but not touching the boundary: $0 < d(C) < E$.

We just need to uniformly show that

$$(*) \qquad (K_{\overline{X}_1(\mathfrak{n})} - \lambda D).C > B \quad \text{for} \quad |\operatorname{Nm} \mathfrak{n}| \gg 0$$

for all $C \not\subset D$.

Case i): In this case $D.C > 0$, and therefore Proposition 3.7 immediately implies $(*)$.

Case ii): If $d(C) \geq E$, then there is some point $x \in C$ with $N_*(x) \leq E$ for all cusps $*$. It follows from Corollary 3.10 and Lemma 3.11 that ρ_x is large in this region uniformly in $|\operatorname{Nm} \mathfrak{n}|$ if we take E sufficiently large only depending on B. By (3),

$$K_{\overline{X}}.C \geq \frac{2}{n} \sinh^2(\rho_x/2) \operatorname{mult}_x C \geq \frac{2}{n} \sinh^2(\rho_x/2)$$

and this is sufficient for $(*)$ since $D.C = 0$.

Case iii): For $|\operatorname{Nm}(\mathfrak{n})|$ large, $E < s_*$ for each cusp $*$ and the above chosen E. Then any curve C in Case iii) is entirely contained in the horoball neighborhood $\widetilde{W}_*(s)$ for any $E < s < s_*$, which is impossible. Indeed, by Proposition 2.4

$$\begin{aligned} \operatorname{vol}(C) = \operatorname{vol}(C \cap W_*(s_*)) &\geq \left(\frac{s_*}{s}\right)^{1/n} \cdot \operatorname{vol}_X(C \cap W_*(s)) \\ &\geq \left(\frac{s_*}{s}\right)^{1/n} \cdot \operatorname{vol}(C) \end{aligned}$$

q.e.d.

The following easy corollary will be useful for proving uniformity in gonality in the next section. It essentially says that the error term in Schwarz's lemma (2) coming from intersections with the boundary becomes negligible sufficiently high in the torsion level tower:

Corollary 5.3. *For any $\epsilon > 0$ and any curve $C \to \overline{X}_1(\mathfrak{n})$ not contained in the boundary we have*

$$(1-\epsilon) \cdot \frac{1}{2\pi n} \operatorname{vol}(C) \leq 2g(C) - 2$$

for all but finitely many $\mathfrak{n}$, uniformly in $X(1)$.

Proof. Take $\lambda = \frac{n}{\epsilon} - 1$. Then by the proposition and Lemma 2.1,

$$\begin{aligned}
(1-\epsilon) \cdot \frac{1}{2\pi n} \operatorname{vol}(C) &= \frac{1-\epsilon}{n} \left(K_{\overline{X}} + D\right).C \\
&\leq \frac{1-\epsilon}{n} \left(K_{\overline{X}} + D\right).C + \frac{\epsilon}{n}(K_{\overline{X}} - \lambda D).C \\
&= \frac{1}{n} \left(K_{\overline{X}} - (n-1)D\right).C \\
&\leq 2g(C) - 2
\end{aligned}$$

for $|\operatorname{Nm}(\mathfrak{n})| \gg 0$. q.e.d.

6. Geometric torsion: uniformity in gonality

Preparations. The idea of the proof of Theorem C is the same as in Section 5: we show that the torsion level covers $\overline{X}_1(\mathfrak{n})$ don't admit maps from d-gonal curves C for large $|\operatorname{Nm}\mathfrak{n}|$. Note that given such a map $C \to X_1(\mathfrak{n})^*$, the degree d map $C \to \mathbb{P}^1$ gives us a map into the d-fold symmetric product:

$$\mathbb{P}^1 \to \operatorname{Sym}^d X_1(\mathfrak{n})^*$$

This can be thought of as Weil restriction of the associated (rational) family of abelian varieties over C down to $\mathbb{P}^1$. The main theorem then obviously follows from the following

Proposition 6.1. *Fix $X(1)$. Then every rational curve in $\operatorname{Sym}^d X_1(\mathfrak{n})^*$ is contained in a diagonal for all but finitely many $\mathfrak{n}$.*

We will again prove the proposition by showing that the only rational curves in $\operatorname{Sym}^d \overline{X}_1(\mathfrak{n})$ are contained in the boundary or a diagonal. By the boundary of $\operatorname{Sym}^d \overline{X}_1(\mathfrak{n})$, we mean the image under the quotient $q : \overline{X}_1(\mathfrak{n})^d \to \operatorname{Sym}^d \overline{X}_1(\mathfrak{n})$ of the locus of points that project to the boundary in some projection.

Let $\overline{X}$ be the toroidal compactification of a quotient $X = \Gamma\backslash\mathbb{H}^n$ of $\mathbb{H}^n$ by a rank 1 lattice Γ. The first step is to relate the genus of a curve $C \to \operatorname{Sym}^d \overline{X}$ to its volume in the same spirit as Lemma 2.1. Of course, as mentioned in Remark 2.2, the *orbifold* symmetric product $\mathcal{S}\mathrm{ym}^d X$ is a perfectly valid quotient of $\mathbb{H}^{dn}$, and if we consider a proper map $U \to \mathcal{S}\mathrm{ym}^d X$ from a punctured orbifold curve, then Lemma 2.1

still applies. As every curve C in the coarse space $\mathrm{Sym}^d X$ is the coarse space of a minimal orbifold curve mapping to $\mathcal{S}\mathrm{ym}^d X$, we obtain a lower bound to the multiplicity of C along the diagonals and the boundary in terms of the volume. For clarity we give a second argument only involving the coarse spaces.

Consider a map $\alpha : \mathbb{P}^1 \to \mathrm{Sym}^d \overline{X}$ whose image is not contained in any diagonal, and let C be the normalization of the fiber product

$$\begin{tikzcd} C \arrow[r, "\alpha'"] \arrow[d, "\pi"'] & \overline{X}^d \arrow[d, "q"] \\ \mathbb{P}^1 \arrow[r, "\alpha"'] & \mathrm{Sym}^d \overline{X} \end{tikzcd}$$

We may assume C is irreducible. Define the total ramification

$$\mathrm{Ram}(\pi) := \sum_{p \in C} (r_p - 1)$$

where r_p is the order of ramification of π at p, so that Riemann–Hurwitz says

$$2g(C) - 2 = d!(-2) + \mathrm{Ram}(\pi)$$

Now π can only ramify at points p on a diagonal, meaning $\pi(p)$ projects to the diagonal $\Delta_{\overline{X}}$ via one of the projections to $\overline{X} \times \overline{X}$, and the ramification order is less than $d!$. Let $C_{ij} = C \to \overline{X} \times \overline{X}$ be the map obtained by composing α' with the ijth projection. Thus, we have

$$\frac{1}{d!}\,(2g(C) - 2) \leq \max_{ij} \mathrm{mult}_{\Delta_{\overline{X}}} C_{ij} \tag{6}$$

To prove the proposition it is enough to show the following diagonal multiplicity bound:

Proposition 6.2. *Fix $X(1)$. For any $M > 0$ and any curve $C \to \overline{X}_1(\mathfrak{n}) \times \overline{X}_1(\mathfrak{n})$ not contained in the boundary or the diagonal $\Delta_{\overline{X}_1(\mathfrak{n})}$ we have*

$$\mathrm{vol}_{X_1(\mathfrak{n}) \times X_1(\mathfrak{n})} C \geq M \cdot \mathrm{mult}_{\Delta_{\overline{X}}} C$$

for all but finitely many $\mathfrak{n}$.

Proof of Proposition 6.1 given Proposition 6.2. Continuing with the above notation, denote by $C_i = C \to \overline{X}$ the composition of α' with the projection to the ith factor. Certainly we have

$$\mathrm{vol}_{X \times X} C_{ij} = \mathrm{vol}_X C_i + \mathrm{vol}_X C_j$$

Proposition 6.2 and (6) then clearly contradict the asymptotic Schwarz lemma in Corollary 5.3. q.e.d.

Diagonal multiplicity estimate. To prove Proposition 6.2, we first summarize the setup of [**HT12**] where a diagonal multiplicity estimate for compact curves is proven. Recall that $d_{\mathbb{H}}(z,w)$ is the hyperbolic distance function on $\mathbb{H}\times\mathbb{H}$. For any $R>0$, denote by

$$T_{\mathbb{H}}(R)=\{(z,w)\in\mathbb{H}\times\mathbb{H}\mid d_{\mathbb{H}}(z,w)<R\}$$

the radius R tubular neighborhood of the diagonal $\Delta_{\mathbb{H}}\subset\mathbb{H}\times\mathbb{H}$. Hwang and To construct a continuous increasing function $\mu:\mathbb{R}\to[-\infty,\infty)$ such that, setting $f=\mu\circ d_{\mathbb{H}}$,

i) μ is supported on $[0,R]$ with $\mu(0)=-\infty$ and f is smooth and finite valued on the complement of $\Delta_{\mathbb{H}}$;
ii) $i\partial\overline{\partial}[f]\geq-\omega_{\mathbb{H}\times\mathbb{H}}$ on the complement of $\Delta_{\mathbb{H}}$, and for any point $\xi\in\Delta_{\mathbb{H}}$,
$$\nu(\omega_{\mathbb{H}\times\mathbb{H}}+i\partial\overline{\partial}[f],\xi)=L_R$$
iii) $\lim_{R\to\infty}L_R=\infty$.

By $[f]$ we mean the distribution associated to f, which is sensible since a) and b) imply that $f+\varphi$ is plurisubharmonic, where φ is a potential for $\omega_{\mathbb{H}\times\mathbb{H}}$. In general, for any complex manifold X with a (closed) positive real (1,1) current ω, we say that a continuous locally integrable function $g:X\to[-\infty,\infty)$ is ω-plurisubharmonic (ω-psh) if $i\partial\overline{\partial}[g]\geq-\omega$. Thus, f is $\omega_{\mathbb{H}\times\mathbb{H}}$-psh on $\mathbb{H}\times\mathbb{H}$.

f allows us to prove a diagonal multiplicity estimate for a *compact* quotient $X=\Gamma\backslash\mathbb{H}$. Indeed, for $R<\rho_X$, then the quotient of $T_{\mathbb{H}}(R)$ by the diagonal action of Γ embeds as $T_X(R)$ in $X\times X$, and since f is diagonally invariant it descends to a $\omega_{X\times X}$-psh function f on $X\times X$ supported on $T_X(R)$. For any curve $C\subset X\times X$ we then have (*cf.* [**HT12**])

$$\mathrm{vol}_{X\times X}(C\cap T_X(R))=\int_{C\cap T(R)}\omega_{X\times X}+i\partial\overline{\partial}[f]\geq L_R\cdot\mathrm{mult}_{\Delta_X}C \tag{7}$$

In fact, Hwang and To construct f so that L_R is optimal, but we only need property iii) above. Since the maximum of plurisubharmonic functions is plurisubharmonic, and the Kobayashi distance $d_{\mathbb{H}^n}$ on $\mathbb{H}^n\times\mathbb{H}^n$ is the maximum of the coordinate-wise distances, properties i)–iii) continue to hold for $f=\mu\circ d_{\mathbb{H}^n}$ on $\mathbb{H}^n\times\mathbb{H}^n$.

For noncompact quotients $X=\Gamma\backslash\mathbb{H}^n$, the above approach fails because $\rho_X=0$. The idea is to uniformly introduce a new metric on $\overline{X}_1(\mathfrak{n})$ so that $\rho_{\overline{X}_1(\mathfrak{n})}$ is nonzero and growing. The key point is that for a fixed F the unipotent part of the parabolic stabilizer of a cusp of $X_1(\mathfrak{n})$ is one of only finitely many lattices up to scale, corresponding to ideal classes of F. A new metric in a cuspidal neighborhood can therefore be glued in uniformly in $\mathfrak{n}$ (for fixed F), and Proposition 2.4 will show us that the difference between volumes of curves with respect to the old and the new metric is negligible sufficiently high in the tower.

Metrics at the cusp. We have seen that each cusp of $\Gamma_1(\mathfrak{n})$ can be conjugated to infinity so that its stabilizer has the form

$$\begin{pmatrix} u & \lambda \\ 0 & u^{-1} \end{pmatrix} \text{ for } u \in H, \lambda \in \Lambda$$

where Λ is some fractional ideal and H is a subgroup of $\mathcal{O}_F^*$. By scaling we can assume $N_* = N$. Denote by $W_{\Lambda,H}(s)$ the quotient of $U(s) = \{z \in \mathbb{H}^n \mid N(z) > 1/s\}$ by this group. If we take $\{\Lambda_i\}$ to be a set of fractional ideals representing the ideal classes of $\mathrm{Cl}(F)$, then it follows that for each cusp $*$ of $X_1(\mathfrak{n})$ the horoball neighborhood $W_*(s)$ is isomorphic to some $W_{\Lambda_i,H}(s)$. We fix once and for all a smooth toroidal compactification of each $W_{\Lambda_i,\mathcal{O}_F^*}(s)$. Note that the same fan yields a smooth toroidal compactification of $W_{\Lambda_i,H}(s)$ for any finite index $H \subset \mathcal{O}_F^*$ and the resulting map $q : \widetilde{W}_{\Lambda_i,H}(s) \to \widetilde{W}_{\Lambda_i,\mathcal{O}_F^*}(s)$ is étale. Using these fixed cuspidal resolutions, it will therefore be the case that for each cusp $*$ of $X_1(\mathfrak{n})$ we have $\widetilde{W}_*(s) \cong \widetilde{W}_{\Lambda_i,H}(s)$ for some i and H.

Fixing Λ, for simplicity we drop the subscript $\Lambda, \mathcal{O}_F^*$ from the notation for the moment. Given $s_0 > 0$, there is a continuous function $\psi : \widetilde{W}(\infty) \times \widetilde{W}(\infty) \to [-\infty, \infty)$ which is smooth and finite valued on $W(\infty) \times W(\infty)$, $-\infty$ on the boundary, and additionally satisifes the following properties: i) ψ is supported on $\widetilde{W}(s_0)$; ii) $\omega' = \omega_{W(\infty)\times W(\infty)} + i\partial\overline{\partial}[\psi] \geq \omega_0$ for some smooth (complete) Kähler form ω_0 on $\widetilde{W}(\infty) \times \widetilde{W}(\infty)$. This follows from the fact that the boundary can be contracted to a point.

Given $R > 0$, choose s_R so that the injectivity radius of every point in $\widetilde{W}(\infty) \smallsetminus \widetilde{W}(s_R)$ is greater than R. Now, there is a continuous function $f' : \widetilde{W}(\infty) \times \widetilde{W}(\infty) \to [-\infty, \infty)$ which is smooth and finite valued on $\widetilde{W}(\infty) \times \widetilde{W}(\infty) \smallsetminus \Delta_{\widetilde{W}(\infty)}$, and satisfies: i) $f' = f$ outside of $\widetilde{W}(s_R) \times \widetilde{W}(s_R)$; and ii) $f' - L_R \log d_{\widetilde{W}(\infty)}$ is smooth in some neighborhood of $\Delta_{\widetilde{W}(\infty)}$, where $d_{\widetilde{W}(\infty)}$ is the distance function with respect to ω_0. By compactness and the properties of the function f, there is some constant $B > 0$ such that f' is $B\omega'$-psh—and therefore $\psi + f'$ is $B\omega_{W(\infty)\times W(\infty)}$-psh—in $\widetilde{W}(s_R) \times \widetilde{W}(s_R)$.

We are now in a position to prove the following:

Lemma 6.3. *Fix $X(1)$. For each $R > 0$, there are constants $s_R, B_R > 0$ such that for $|\mathrm{Nm}(\mathfrak{n})| \gg 0$, there is a continuous function $g : \overline{X}_1(\mathfrak{n}) \times \overline{X}_1(\mathfrak{n}) \to [-\infty, \infty)$ satisfying:*

a) *g is supported on*

$$T_{X_1(\mathfrak{n})}(R) \cup \bigcup_* \widetilde{W}_*(s_R) \times \widetilde{W}_*(s_R)$$

and smooth and finite valued on the complement of the diagonal and the boundary;

b) *g is $B_R\omega_{X_1(\mathfrak{n})\times X_1(\mathfrak{n})}$-psh and for any point $\xi \in \Delta_{\overline{X}_1(\mathfrak{n})}$,*

$$\nu(i\partial\overline{\partial}[g] + B_R\omega_{X_1(\mathfrak{n})\times X_1(\mathfrak{n})}, \xi) = L_R \tag{8}$$

c) *g is $\omega_{X_1(\mathfrak{n})\times X_1(\mathfrak{n})}$-psh outside of $\bigcup_* \widetilde{W}_*(s_R) \times \widetilde{W}_*(s_R)$, and for any point $\xi \in \Delta_{\overline{X}_1(\mathfrak{n})}$ outside of $\bigcup_* \widetilde{W}_*(s_R) \times \widetilde{W}_*(s_R)$,*

$$\nu(i\partial\overline{\partial}[g] + \omega_{X_1(\mathfrak{n})\times X_1(\mathfrak{n})}, \xi) = L_R.$$

d) *$\lim_{R\to\infty} L_R = \infty$.*

Proof. For each cusp $*$, the pullback g of the function $\psi + f'$ constructed above to $\widetilde{W}_*(\infty)$ satisfies all four properties, and we may glue each of these functions into the cusps of $\overline{X}_1(\mathfrak{n})$. q.e.d.

Conclusion of the proof. We are now ready to prove the diagonal multiplicity inequality:

Proof of Proposition 6.2. Given $M > 0$, choose R such that $L(R) > 2M$. For simplicity write $X = X_1(\mathfrak{n})$ and let

$$\widetilde{W}_R = \bigcup_* \widetilde{W}_*(s_R) \times \widetilde{W}_*(s_R) \subset \overline{X} \times \overline{X}$$

For any curve $C \to \overline{X}\times\overline{X}$ not contained in the diagonal or the boundary, we have for $|\operatorname{Nm}(\mathfrak{n})| \gg 0$

$$\begin{aligned} B_R \cdot \operatorname{vol}_{X\times X} C &= \int_C B_R \cdot \omega_{X\times X} \\ &= \int_C i\partial\overline{\partial}[g] + B_R \cdot \omega_{X\times X} \\ &\geq \int_{C\cap X\smallsetminus \widetilde{W}_R} (B_R - 1)\omega_{X\times X} + L_R \cdot \operatorname{mult}_{\Delta_{\overline{X}}} C \\ &= (B_R - 1)\cdot \operatorname{vol}_{X\times X}(C \smallsetminus \widetilde{W}_R) + L_R \cdot \operatorname{mult}_{\Delta_{\overline{X}}} C \end{aligned}$$

and therefore

$$\operatorname{vol}_{X\times X}(C \smallsetminus \widetilde{W}_R) + B_R \cdot \operatorname{vol}_{X\times X}(C \cap \widetilde{W}_R) \geq 2M \cdot \operatorname{mult}_{\Delta_{\overline{X}}} C \tag{9}$$

Once again letting $C_i = C \to \overline{X}$ be the projection to the ith factor, we have by Proposition 2.4

$$\begin{aligned}\mathrm{vol}_{X\times X}(C\cap \widetilde{W}_R) &\leq \sum_{i=1,2}\sum_{*}\mathrm{vol}_X(C_i\cap W_*(s_R))\\ &\leq \sum_{1,2}\sum_{*}\left(\frac{s_R}{s_*}\right)^{1/n}\mathrm{vol}_X(C_i\cap W_*(s_*))\\ &\leq \frac{1}{B_R}\cdot \mathrm{vol}_{X\times X} C\end{aligned}$$

where in the last line we've taken $|\,\mathrm{Nm}(\mathfrak{n})|$ large enough so that $s_* > s_R\cdot B_R^n$, using Proposition 3.7. Combining this with (9), we obtain

$$\mathrm{vol}_{X\times X} C \geq M\cdot \mathrm{mult}_{\Delta_{\overline{X}}} C$$

for $|\,\mathrm{Nm}(\mathfrak{n})| \gg 0$, as desired. q.e.d.

References

[AMRT10] A. Ash, D. Mumford, M. Rapoport, and Y.-S. Tai. *Smooth compactifications of locally symmetric varieties.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2010. With the collaboration of Peter Scholze.

[BB66] W. L. Baily, Jr. and A. Borel. Compactification of arithmetic quotients of bounded symmetric domains. *Ann. of Math. (2)*, 84:442–528, 1966.

[BT15] B. Bakker and J. Tsimerman. The Kodaira dimension of complex hyperbolic manifolds with cusps. arXiv:1503.05654, 2015.

[BT16] B. Bakker and J. Tsimerman. p-torsion monodromy representations of elliptic curves over geometric function fields. *Ann. of Math. (2)*, 184(3):709–744, 2016.

[CT11] A. Cadoret and A. Tamagawa. On a weak variant of the geometric torsion conjecture. *J. Algebra*, 346:227–247, 2011.

[CT12] A. Cadoret and A. Tamagawa. Uniform boundedness of p-primary torsion of abelian schemes. *Invent. Math.*, 188(1):83–125, 2012.

[DCDC15] G. Di Cerbo and L. F. Di Cerbo. Positivity questions in Kähler-Einstein theory. *Math. Proc. Cambridge Philos. Soc.*, 159(2):321–338, 2015.

[Dem92] J.-P. Demailly. Singular Hermitian metrics on positive line bundles. In *Complex algebraic varieties (Bayreuth, 1990)*, volume 1507 of *Lecture Notes in Math.*, pages 87–104. Springer, Berlin, 1992.

[Dem12] J.-P. Demailly. *Analytic methods in algebraic geometry*, volume 1 of *Surveys of Modern Mathematics.* International Press, Somerville, 2012.

[Fre80] E. Freitag. Eine Bemerkung zur Theorie der Hilbertschen Modulmannigfaltigkeiten hoher Stufe. *Math. Z.*, 171(1):27–35, 1980.

[Gor02] Eyal Z Goren. *Lectures on Hilbert modular varieties and modular forms.* American Mathematical Soc., 2002.

[HT02] J. Hwang and W. To. Volumes of complex analytic subvarieties of Hermitian symmetric spaces. *American Journal of Mathematics*, 124(6):1221–1246, 2002.

[HT06] J.-M. Hwang and W.-K. To. Uniform boundedness of level structures on abelian varieties over complex function fields. *Mathematische Annalen*, 335(2):363–377, 2006.

[HT12] J. Hwang and W. To. Injectivity radius and gonality of a compact Riemann surface. *American Journal of Mathematics*, 134(1):259–283, 2012.

[HZ77] F. Hirzebruch and D. Zagier. Classification of Hilbert modular surfaces. In *Complex analysis and algebraic geometry*, pages 43–77. Iwanami Shoten, Tokyo, 1977.

[Kam92] S. Kamienny. Torsion points on elliptic curves over fields of higher degree. *Internat. Math. Res. Notices*, (6):129–133, 1992.

[Mar84] G. A. Margulis. Arithmeticity of the irreducible lattices in the semi-simple groups of rank greater than 1. *Inventiones mathematicae*, 76(1):93–120, 1984.

[Mer96] L. Merel. Bornes pour la torsion des courbes elliptiques sur les corps de nombres. *Invent. Math.*, 124(1-3):437–449, 1996.

[MG78] B. Mazur and Appendix by D. Goldfeld. Rational isogenies of prime degree. *Inventiones mathematicae*, 44(2):129–162, 1978.

[Mum77] D. Mumford. Hirzebruch's proportionality theorem in the noncompact case. *Invent. Math.*, 42:239–272, 1977.

[Nad89] A. M. Nadel. The nonexistence of certain level structures on abelian varieties over complex function fields. *Ann. of Math. (2)*, 129(1):161–178, 1989.

[RT15] E. Rousseau and F. Touzet. Curves in Hilbert modular varieties. arXiv:1501.03261, 2015.

[Tsu85] Shigeaki Tsuyumine. On the Kodaira dimensions of Hilbert modular varieties. *Invent. Math.*, 80(2):269–281, 1985.

[vdG79] G. van der Geer. Minimal models for Hilbert modular surfaces of principal congruence subgroups. *Topology*, 18(1):29–39, 1979.

[vdG88] G. van der Geer. *Hilbert modular surfaces*, volume 16. Springer-Verlag Berlin-Heidelberg-New York, 1988.

B. Bakker: Department of Mathematics, University of Georgia.

E-mail address: bakker@math.uga.edu

J. Tsimerman: Mathematics Department, University of Toronto.

E-mail address: jacobt@math.toronto.edu