COMPOSITIO MATHEMATICA

The André–Oort conjecture for the moduli space of abelian surfaces

Jonathan Pila and Jacob Tsimerman

Compositio Math. **149** (2013), 204–216.

doi:10.1112/S0010437X12000589

FOUNDATION COMPOSITIO MATHEMATICA

LONDON MATHEMATICAL SOCIETY

Compositio Math. 149 (2013) 204–216
doi:10.1112/S0010437X12000589

The André–Oort conjecture for the moduli space of abelian surfaces

Jonathan Pila and Jacob Tsimerman

Abstract

We provide an unconditional proof of the André–Oort conjecture for the coarse moduli space $\mathcal{A}_{2,1}$ of principally polarized abelian surfaces, following the strategy outlined by Pila–Zannier.

1. Introduction and notation

Let $\mathcal{A}_{g,1}$ denote the coarse moduli space of principally polarized abelian varieties of dimension g. Our main theorem is the following, proving the André–Oort conjecture for $\mathcal{A}_{2,1}$.

THEOREM 1.1. *Let $V \subset \mathcal{A}_{2,1}$ be an algebraic subvariety, which is equal to the Zariski closure of its CM points. Then V is a special subvariety.*

Here, a subvariety means a relatively closed irreducible subvariety. Varieties will be identified with their sets of complex-valued points.

We follow the general strategy of Pila–Zannier. Set $\mathbb{H}_g$ to be the Siegel upper half space

$$\mathbb{H}_g = \{Z \in M_{g\times g}(\mathbb{C}) \mid Z = Z^{\mathrm{t}}, \mathrm{Im}(Z) > 0\}.$$

We denote by

$$\pi : \mathbb{H}_g \to \mathcal{A}_{g,1}$$

the natural projection map. The set $\{Z \in M_{g\times g}(\mathbb{C}) | Z = Z^{\mathrm{t}}\}$ is naturally identified with $\mathbb{C}^{g(g+1)/2}$, identifying $\mathbb{H}_g$ with an open domain. We further identify $\mathbb{C}^{g(g+1)/2}$ with $\mathbb{R}^{g(g+1)}$ by means of real and imaginary parts of the complex co-ordinates, and call a set *semialgebraic* if it is a semialgebraic set in $\mathbb{R}^{g(g+1)}$. Note in particular that $\mathbb{H}_g$ is semialgebraic. A semialgebraic set is called *irreducible* if it is not the union of two non-empty relatively closed subsets in the topology induced on it by the Zariski topology of algebraic sets defined over $\mathbb{R}$ [GV95].

One ingredient we need is the following Ax–Lindemann–Weierstrass theorem. Let $V \subset \mathcal{A}_{2,1}$ be an algebraic subvariety. Let $Z = \pi^{-1}(V) \subset \mathbb{H}_2$, and let $Y \subset Z$ be a connected irreducible semialgebraic subset of $\mathbb{H}_2$. We say that Y is *maximal* if every semialgebraic subset Y' with $Y \subset Y' \subset Z$ has Y as a component. The definition of a *weakly special subvariety of* $\mathbb{H}_g$ is given in § 2.

THEOREM 1.2. *Let $V \subset \mathcal{A}_{2,1}$ be an algebraic subvariety, and let $Y \subset \pi^{-1}(V)$ be a maximal connected irreducible semialgebraic subset. Then Y is a weakly special subvariety.*

Let $F_g \subset \mathbb{H}_g$ be the standard fundamental domain [Sie64]. We shall also need the following bound on heights of CM points, where the *height* of $Z \in \mathbb{H}_g \cap M_{g\times g}(\overline{\mathbb{Q}})$, denoted $H(Z)$, is the

Received 8 November 2011, accepted in final form 8 May 2012, published online 11 December 2012.
2010 Mathematics Subject Classification 11G18 (primary), 03C64 (secondary).
Keywords: André–Oort conjecture, moduli of abelian surfaces.

maximum absolute Weil height of its co-ordinates under the natural identification of $M_{g\times g}(\overline{\mathbb{Q}})$ with $\overline{\mathbb{Q}}^{g^2}$.

Theorem 1.3. *There exists an absolute constant* $\delta(g) > 0$ *such that, if* $Z \in F_g$ *is a CM point,*

$$H(Z) \ll_g |\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \cdot \pi(Z)|^{\delta(g)}.$$

The proofs of Theorems 1.1 and 1.2 rely on ideas and results from o-minimality. We refer readers desiring some background on o-minimality to the brief introductory descriptions in [Pil11, PW06] or the sources referred to in [PS11b, PS11a]. By *definable* we mean *definable in some o-minimal structure over the real field* as defined e.g. in [PW06]. All the sets we require are definable in the o-minimal structure $\mathbb{R}_{\mathrm{an,exp}}$ (see [vdDM94]), and except in Lemma 4.3 we take *definable* to mean definable in $\mathbb{R}_{\mathrm{an,exp}}$.

We mention that for the strategy to go through, a basic requirement is the definability of the projection map $\pi : F_g \to \mathbb{A}_{g,1}$ in the case $g = 2$. This was provided (for all g) in the recent work [PS11b] of Peterzil and Starchenko. We also use another result of these authors [PS11a] which says that a definable, globally complex analytic subset of an algebraic variety is algebraic. This can be viewed as an analogue of Chow's theorem, and comes up for us in our proof of Theorem 1.2.

The outline of the paper is as follows: in § 2 we review background concerning Shimura varieties. In § 3 we prove Theorem 1.3. In § 4 we prove Theorem 1.2. Our method of proof of the Ax–Lindemann–Weierstrass result is different to the one in [Pil11] in that we do not use the results of Pila–Wilkie, though we do make use of o-minimality. In § 5 we combine everything to prove our main result Theorem 1.1.

2. Background: Shimura varieties

We recall here some of the basic definitions regarding Shimura varieties; for further details see [Del79, UY, UY11]. Let $\mathbb{S}$ denote the real torus $\mathrm{Res}_{\mathbb{C}/\mathbb{R}}\mathbb{G}_{m\mathbb{C}}$. Let G be a reductive algebraic group over $\mathbb{Q}$, and let X denote a conjugacy class of homomorphisms

$$h : \mathbb{S} \to G_{\mathbb{R}}$$

satisfying the following 3 axioms.

- The action of h on the lie algebra of $G_{\mathbb{R}}$ only has Hodge weights $(0, 0)$, $(1, -1)$ and $(-1, 1)$ occurring.
- The adjoint action of $h(i)$ induces a Cartan involution on the adjoint group of $G_{\mathbb{R}}$.
- The adjoint group of $G_{\mathbb{R}}$ has no factors H defined over $\mathbb{Q}$ on which the projection of h becomes trivial.

This guarantees that X acquires a natural structure of a complex analytic space. Moreover, $G(\mathbb{R})$ has a natural action on X given by conjugation, and this turns $G(\mathbb{R})$ into a group of biholomorphic automorphisms of X. We call the pair (G, X) a *Shimura datum.* A Shimura datum (H, X_H) is said to be a *Shimura sub-datum* of (G, X) if $H \subset G$ and $X_H \subset X$.

Fix K to be a compact subgroup of $G(\mathbb{A}_f)$, where $\mathbb{A}_f$ denote the finite adeles. We then define

$$\mathrm{Sh}_K(G, X)(\mathbb{C}) = G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$$

where $G(\mathbb{Q})$ acts diagonally, and K only acts on $G(\mathbb{A}_f)$. It is a theorem of Deligne [Del79] that $\mathrm{Sh}_K(G, X)$ can be given the structure of an algebraic variety over $\overline{\mathbb{Q}}$, and we call $\mathrm{Sh}_K(G, X)$

a *Shimura variety*. Note that $\mathrm{Sh}_K(G,X)$ is in general reducible, and that its components are called *connected Shimura varieties*. Let X^+ be a connected component of X. Then X^+ has the structure of a bounded symmetric domain, and the connected components S of $\mathrm{Sh}_K(G,X)$ arise as quotients

$$\pi : X^+ \to S = \Gamma\backslash X^+$$

where Γ is a suitable subgroup of $G(\mathbb{Q})$. In particular $A_{g,1} = \mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathbb{H}_g$ is a connected Shimura variety; note that $\mathbb{H}_g$ has an alternative realization as a bounded symmetric domain via a biholomorphic algebraic map [UY11, § 3.2].

Given two Shimura varieties $\mathrm{Sh}_{K_i}(G_i, X_i)$, $i = 1, 2$ and a map of algebraic groups $\phi : G_1 \to G_2$ which (induces a map which) takes X_1 to X_2 and K_1 to K_2, we get an induced map $\tilde{\phi} : \mathrm{Sh}_{K_1}(G_1, X_1) \to \mathrm{Sh}_{K_2}(G_2, X_2)$. Given an element $g \in G_2(\mathbb{A}_f)$, right multiplication by g gives a correspondence T_g on $\mathrm{Sh}_{K_2}(G_2, X_2)$. We define a *special subvariety* of $\mathrm{Sh}_{K_2}(G_2, X_2)$ to be an irreducible component of the image of $\mathrm{Sh}_{H_1}(G_1, X_1)$ under $T_g \circ \tilde{\phi}$.

We denote by G^{ad} the adjoint form of a reductive group G. Following Ullmo–Yafaev [UY], we make the following definition.

Definition. An algebraic subvariety Z of $\mathrm{Sh}_K(G,X)$ is *weakly special* if there exists a Shimura sub-datum (H, X_H) of (G,X), a decomposition

$$(H^{\mathrm{ad}}, X_H^{\mathrm{ad}}) = (H_1, X_1) \times (H_2, X_2),$$

and a point $y_2 \in X_2$ such that Z is the image of a connected component of $X_1 \times y_2$.

In this definition, a weakly special subvariety is *special* if and only if it contains a special point and if and only if y_2 is special.

Let $S = X^+\backslash\Gamma$ be a connected Shimura variety. A *weakly special subvariety* of X^+ is a connected component of $\pi^{-1}(Z)$ for some weakly special $Z \subset \mathcal{S}$. Being bounded, X^+ will not contain any positive dimensional algebraic subvarieties of its ambient space $\mathbb{C}^n$. For a connected open domain $\Omega \subset \mathbb{C}^n$, an *algebraic subvariety* of Ω will mean a connected component of $W \cap \Omega$ for some algebraic subvariety $W \subset \mathbb{C}^n$. Then a weakly special subvariety of X is an algebraic subvariety of X; see [UY11].

3. Heights of CM points

The aim of this section is to prove that the moduli point corresponding to a principally polarized CM abelian variety in a fundamental domain for $\mathbb{H}_2$ has polynomial height in terms of the discriminant of its endomorphism algebra, which is essential for us to apply the results of Pila–Wilkie. We only need this result for abelian surfaces, but we give the proof in general since the increased difficulty is only technical, and the theorem is fundamental to the Pila–Zannier strategy. To ease notation, we fix the following convention: given a set X and two functions $f, g : X \to \mathbb{R}$ we say that f is *polynomially bounded* in terms of g if there are positive real constants a, b such that $f(x) \leqslant a g(x)^b$ for all $x \in X$.

Theorem 3.1. *Fix g. For a complex, principally polarized abelian variety A of dimension g with complex multiplication, set $R = Z(\mathrm{End}(A))$ to be the center of its endomorphism algebra. Let $x \in F_g$ be the point representing A. Then, as functions defined on the CM points of F_g, $H(x)$ is polynomially bounded in terms of* $\mathrm{Disc}(R)$ *as functions on the simple CM points of $x \in F_g$.*

Combined with the results of [Tsi12], this proves Theorem 1.3.

Proof. We first handle the case where A is simple.

Case 1. *A is simple.* In this case, there is a CM field K such that A has CM by K. Let $S = \{\phi_1 \ldots, \phi_g\}$ denote the complex embeddings of K that make up the CM type of A, and set F to be the maximal totally real subfield of K, so that K is a quadratic extension of F. In this case R is just the endomorphism ring of A. Now, there is a $\mathbb{Z}$-lattice $I \subset O_K$ such that A is isomorphic to $\mathbb{C}^g/\phi_S(I)$ as a complex torus under the embedding

$$\phi_S : K \hookrightarrow \mathbb{C}^g, \quad \phi(a) = (a^{\phi_1}, a^{\phi_2}, \ldots, a^{\phi_g}).$$

Moreover, the order of I must be R.

CLAIM 3.1. *There is a $\nu \in K$ such that $\nu I \subset O_K$ and $[O_K : \nu I]$ is polynomially bounded in terms of* $\mathrm{Disc}(R)$.

Proof of Claim. If $R = O_K$, then the lemma is a known consequence of Minkowski's bound. For the general case, set $e_R = [O_K : R]$. Note that

$$\mathrm{Disc}(R) = \mathrm{Disc}(O_K)e_R^2$$

and that

$$e_R O_K \subset R \subset O_K.$$

Set $J = O_K \cdot I$, so that J is an O_K-ideal such that

$$e_R J \subset I \subset J.$$

By the above we can find a $\nu \in K$ with $[O_K : \nu J]$ polynomially bounded in terms of $\mathrm{Disc}(O_K)$. Then

$$\nu I \subset \nu J \subset O_K$$

and

$$[O_K : \nu I] \leqslant [O_K : \nu e_R J] \leqslant e_R^{2g}[O_K : \nu J],$$

and the claim follows. □

CLAIM 3.2. *Given a $\mathbb{Z}$-lattice $I \subset O_K$ there is a basis $\alpha_1, \ldots, \alpha_{2g}$ of I, such that the absolute values of all conjugates of the α_i are polynomially bounded in terms of* $\mathrm{Disc}(O_K) \cdot [O_K : I]$.

Proof of Claim. Consider the standard embedding ψ of O_K as a lattice in $\mathbb{C}^g$ given by

$$\psi(\alpha) = (\alpha^{\phi_i})_{1 \leqslant i \leqslant g}.$$

Then the covolume of $\psi(I)$ as a lattice is $\mathrm{Disc}(O_K) \cdot [O_K : I]$, and every vector in I has norm at least 1 (since the norm of every algebraic integer is at least 1).

Now consider the lattice

$$\frac{1}{\mathrm{Vol}(\mathbb{C}^g/I)^{1/2g}} \cdot \psi(I)$$

as an element l in $\mathrm{SL}_{2g}(\mathbb{Z})\backslash\mathrm{SL}_{2g}(\mathbb{R})$. Let N, A, O_{2g} denote the upper triangular subgroup, the diagonal subgroup, and the maximal compact orthogonal group of $\mathrm{SL}_{2g}(\mathbb{R})$. By the theory of Siegel sets, there is a representative nak of l in $N(\mathbb{R})D(\mathbb{R})O_{2g}(\mathbb{R})$ where n has all its elements bounded by $\frac{1}{2}$ in absolute value, and the diagonal matrix d has

$$d_{1,1} \geqslant d_{2,2} \geqslant \cdots \geqslant d_{2g,2g}.$$

Now, since $d_{2g,2g}$ is the norm of an element, we know that

$$d_{2g,2g} \gg \frac{1}{\mathrm{Vol}(\mathbb{C}^g/I)^{1/2g}}.$$

Since the product of the $d_{i,i}$ is 1, we deduce that $d_{1,1}$ is polynomially bounded in terms of $\mathrm{Disc}(O_k) \cdot [O_k : I]$. The basis corresponding to nak thus has all its co-ordinates polynomially bounded in terms of $\mathrm{Disc}(O_K) \cdot [O_K : I]$. $\square$

Now, consider again our abelian variety A. As before, A is isomorphic as a complex torus to $\mathbb{C}^g/\phi_S(I)$ for some $I \subset K$. The principal polarization on A corresponds to a totally imaginary element $\xi \in K$, which induces the Riemann form

$$E_\xi(a, b) = \mathrm{tr}_{K/\mathbb{Q}}(\xi a b^\rho)$$

where ρ denotes complex conjugation. Moreover, since the polarization is principal E_ξ has determinant 1 as a bilinear alternating form on I.

By changing co-ordinates, one can change the pair (I, ξ) to $(I\nu, \xi(\nu\nu^\rho)^{-1})$ where $\nu \in K^\times$. By Claim 3.1 we can change I to be a sublattice of O_K with $e_I = [O_K : I]$ polynomially bounded by $\mathrm{Disc}(R)$, so we assume that I is of this form. Next, the determinant of E_ξ as a binary form on I is

$$N_{K/\mathbb{Q}}(\xi) \cdot \mathrm{Disc}(O_K)[O_K : I]^2,$$

so that $N_{K/Q}(\xi)$ is bounded above by 1. Moreover, $e_I O_K \subset I$, so that

$$\mathrm{tr}(e_I^2 \xi O_K) \in \mathbb{Z},$$

and so $\xi \in e_I^{-2}\,\mathrm{Disc}(K)^{-1} O_K$.

We know that $-\xi^2$ is a totally positive element, so we can consider the lattice

$$L_\xi = \psi(O_F \cdot (-\xi^2)^{\frac{1}{4}}) \subset \mathbb{R}^g.$$

The covolume of L_ξ is polynomially bounded in terms of $\mathrm{Disc}(R)$, and must contain an element inside a sphere with radius polynomially bounded in terms of the covolume. Therefore, there exists an element $\nu \in O_F$ such that $\nu\nu^\rho\xi = \nu^2\xi$ has all its conjugates polynomially bounded in terms of $\mathrm{Disc}(R)$. Since ν must have norm polynomially bounded by $\mathrm{Disc}(R)$, we can and do assume that $I \subset O_K$ with $[O_K : I]$ polynomially bounded by $\mathrm{Disc}(R)$, and that ξ has all its conjugates polynomially bounded by $\mathrm{Disc}(R)$.

Now consider the representative (I, ξ) of A. To pick a symplectic basis of $(\nu^{-1}I)$, we simply take a basis α_i of $I \cap O_F$ as in Claim 3.2. Next we consider the lattice $\mathrm{Im}(\psi_s(I))$ in $(i\mathbb{R})^g$. Pick a basis as in Lemma 3.2, and refine it to the dual symplectic basis β_i' to α_i. Since all the conjugates of ξ are polynomially bounded in terms of $\mathrm{Disc}(R)$, the basis β_i' has all its components polynomially bounded in terms of $\mathrm{Disc}(R)$ as well. Lift β_i' to elements $\beta_i = \beta_i' + c$, where c is an element in F, such that $\beta_i \in I$. Note that c is an element of $e_I^{-1}O_F/O_F$, and can thus be chosen to have all its components polynomially bounded by $\mathrm{Disc}(R)$. Finally, since the values $E_\xi(\beta_i, \beta_j)$ might not be 0, we replace β_i by $\beta_i - \sum_{i \leqslant j} E_\xi(\beta_i, \beta_j)\alpha_j$.

Now consider the matrix $Z \in \mathbb{H}_g$ which represents the elements β_i in terms of the α_i. Then, Z is the matrix representing A. Moreover, the above construction gives $Z = X + iY$, where $X \in M_g(\mathbb{Q})$ with all its denominators polynomially bounded by $\mathrm{Disc}(R)$, and $Y \in M_g(K)$ such that all the entries of Y have all their complex conjugates polynomially bounded by $\mathrm{Disc}(R)$, and likewise for the denominators of the entries of Y (the denominator of an algebraic number α is the smallest integer n with $n\alpha$ an algebraic integer).

https://doi.org/10.1112/S0010437X12000589 Published online by Cambridge University Press

It is evident that Z has height polynomially bounded by $\mathrm{Disc}(R)$, and that the degree of all the entries in Z is at most $4g$. All that is left to see is that putting Z in its fundamental domain does not increase the height by too much, and so the following lemma completes the proof. The determinant of a matrix Y is denoted $|Y|$ or $\det Y$.

Lemma 3.2. *Let $g \geqslant 1$ be a natural number, and $Z = X + iY$ be an element in Siegel upper half space $\mathbb{H}_g(\mathbb{C})$. Set $h(Z) = \mathrm{Max}(|z_{ij}|, 1/|Y|)$. Then for $\gamma \in \mathrm{Sp}_{2g}(\mathbb{Z})$ such that $\gamma \cdot Z \in F_g$, all the co-ordinates of γ are polynomially bounded in terms of $h(Z)$.*

Proof. Set $\gamma = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$. The proof involves going through and quantifying Siegel's proof in [Sie64] of the fact that F_g is a fundamental domain. As is well known, C, D are such that $|\det(CZ + D)|$ is minimal, and by perturbing Z slightly we can ensure that this determines C and D up to the left action of $\mathrm{GL}_g(\mathbb{Z})$.

As in Siegel, pick an element $U \in \mathrm{GL}_g(\mathbb{Z})$ such that $UY^{-1}U^t$ is Minkowski reduced, and set

$$\gamma_0 = \begin{pmatrix} A_0 & B_0 \\ C_0 & D_0 \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & U^{-t} \end{pmatrix} \gamma$$

and

$$\gamma_0 \cdot Z = X_0 + iY_0.$$

Set $y_1, y_2, \ldots, y_n$ to be the diagonal elements of Y_0^{-1}, and c_l, d_l to be the rows of C_0, D_0. By [Sie64, p. 40, Equations (87), (88)], we have

$$y_l = Y^{-1}[Xc_l + d_l] + Y[c_l] \tag{1}$$

and

$$\prod_{i=1}^{n} y_i \ll |Y|^{-1}. \tag{2}$$

Now, the eigenvalues of Y are polynomially bounded in terms of the co-ordinates of Y, and their product is equal to the determinant of Y, hence the inverses of the eigenvalues of Y are polynomially bounded in terms of $h(Z)$. Thus, since integer vectors have euclidean norm at least 1, (1) implies that y_l is polynomially bounded below by $h(Z)$, which is to say that y_l^{-1} is polynomially bounded in terms of $h(Z)$. But now (2) implies that y_l is polynomially bounded in terms of $h(Z)$, and thus so are the norms of c_l and d_l. Hence all the co-ordinates of C_0 and D_0 are polynomially bounded in terms of $h(Z)$.

We can find A_1, B_1 with polynomially bounded entries such that the matrix

$$\gamma_1 := \begin{pmatrix} A_1 & B_1 \\ C_0 & D_0 \end{pmatrix}$$

is in $\mathrm{Sp}_{2g}(\mathbb{Z})$. Set $Z_1 = \gamma_1 \cdot Z$. There is then an upper triangular matrix $\gamma_2 = \gamma\gamma_1^{-1}$ which takes Z_1 into F_g. The lemma now follows from Lemma 3.3, which is well known but we include for completeness. $\square$

Lemma 3.3. *Let $U \in \mathrm{GL}_g(\mathbb{Z})$ be such that UYU^t is Minkowski reduced. Then U is polynomially bounded in terms of $h(Y) = \mathrm{Max}(|y_{ij}|, 1/|Y|)$.*

https://doi.org/10.1112/S0010437X12000589 Published online by Cambridge University Press

Proof. Let $Y' = UYU^t$, and set $y_1, \ldots, y_n$ to be the diagonal elements of Y'. Letting u_l be the rows of U, we have $Y[u_l] = y_l$, and as before

$$\prod_{i=1}^{n} y_i \ll |Y'| = |Y|. \tag{3}$$

Now, since the u_l have euclidean norm at least 1, we have that the y_l^{-1} are polynomially bounded in terms of $h(Y)$, and hence by (3) the y_l are polynomially bounded in terms of $h(Y)$. Thus we can conclude that the euclidean norms of the u_l are polynomially bounded in terms of $h(Y)$, which implies the lemma. □

Case 2. *General A.* In general, there are simple abelian varieties A_i of dimension g_i with complex multiplication by K_i of CM type S_i such that A is isogenous to

$$\prod_i A_i^{n_i}$$

so that $\sum_i n_i g_i = g$. We assume that the types (K_i, S_i) are inequivalent (which does *not* mean that the fields K_i are all distinct!) so that

$$\mathrm{End}(A) \subset \prod_i M_{g_i}(K_i)$$

and

$$R \subset \bigoplus_i O_{K_i}.$$

For simplicity of notation we set $O_A = \bigoplus_i O_{K_i}$. As before, define e_R to be the index of R in O_A so that

$$\mathrm{Disc}(R) = e_R^2 \cdot \prod_i \mathrm{Disc}(K_i).$$

There is an embedding

$$\psi_S : \bigoplus_i K_i^{n_i} \to \mathbb{C}^g$$

given by $\psi_S = \bigoplus_i \psi_{S_i}^{n_i}$, and a lattice $I \subset \bigoplus_i K_i^{n_i}$ such that

$$A(\mathbb{C}) \cong \mathbb{C}^g / \psi_S(I).$$

Moreover, I is invariant under multiplication by R. Consider $J = O_A \cdot I$. Then J is an O_A module with

$$e_R \cdot J \subset I \subset J.$$

We thus have a direct sum decomposition

$$J = \bigoplus_i J_i$$

with J_i an O_{K_i} ideal, and so in fact we can decompose further

$$J_i = \bigoplus_{j=1}^{n_i} P_{ij}$$

with each P_{ij} an O_{K_i} module. By scaling with elements of $K_i^{\times}$, we can guarantee that $P_{ij} \subset O_{k_i}$ of index at most $\mathrm{Disc}(K_i)^{1/2}$. Therefore, there is an integer N polynomially bounded in terms

https://doi.org/10.1112/S0010437X12000589 Published online by Cambridge University Press

of $\mathrm{Disc}(O_A)$ such that the ring

$$\bigoplus_i (O_{K_i} + N \cdot M_{n_i}(O_{K_i}))$$

preserves J, and thus the ring

$$\mathbb{Z} + Ne_R \bigoplus_i \cdot M_{n_i}(O_{K_i})$$

preserves I.

The following lemma allows us to reduce to the case where $A \cong A_1^{n_1}$.

Lemma 3.4. *There are principally polarized abelian varieties B_i isogenous to $A_i^{n_i}$, and an isogeny $\lambda : A \to \bigoplus_i B_i$ compatible with polarizations such that λ has degree polynomially bounded in terms of* $\mathrm{Disc}(O_A)$.

Proof. Consider $I_0 = O_A \cdot I$, and let A^0 be the abelian variety whose complex points are

$$\mathbb{C}^g / \psi_S(I_0).$$

Thus A has an isogeny λ^0 to A^0 of degree polynomially bounded in terms of $\mathrm{Disc}(O_A)$. Since A^0 has an action by O_A, it splits as $A^0 \cong \bigoplus_i A_i^0$, where A_i^0 has an action of O_{K_i}, and has CM type (K_i, S_i). The polarization on A induces a polarization η on A^0 via λ^0 of degree polynomially bounded in terms of $\mathrm{Disc}(O_A)$, and as the A_i^0 have no non-trivial homomorphisms between them, η splits as $\eta \cong \bigoplus_i \eta_i$. There is then an isogeny from A_i^0 to B_i of degree $\deg(\eta)^{\frac{1}{2}}$ such that B_i is principally polarized. Composing with λ^0 completes the proof. □

If Z_1, Z_2 are two points in the fundamental domain F_g corresponding to principally polarized abelian varieties with an isogeny of degree C between them, then $H(Z_1)/H(Z_2)$ is polynomially bounded in terms of C. Thus, by Lemma 3.4 we can and do restrict to the case where $A \cong A_1^{n_1}$.

Now, as before we can and do assume that $I \subset O_{K_1}^{n_1}$, I is invariant by $\mathbb{Z} + Ne_R M_{n_1 \times n_1}(O_{K_1})$, where N is an integer polynomially bounded in terms of $\mathrm{Disc}(K_1)$, and $[O_{K_1}^{n_1} : I]$ is polynomially bounded in terms of $\mathrm{Disc}(R)$. Then the polarization on A is given by a matrix $E \in M_{g \times g}(K_1)$ satisfying $E = -E^*$, where E^* denotes the transpose-conjugate matrix. As the symplectic form defined by E takes integer values on I, we deduce that there is an integer polynomially bounded by $\mathrm{Disc}(R)$, which we may take to be N, such that NE has entries which are algebraic integers. Moreover, as E defines a principal polarization of A, the determinant of E is polynomially bounded in terms of $\mathrm{Disc}(R)$.

Moreover, as before let $\zeta \in O_{K_1}$ be a totally imaginary element, with entries polynomially bounded by $\mathrm{Disc}(K_1)$ and $-i\phi(\zeta) > 0$ for all $\phi \in S_1$. Then the quadratic form Q on K_1^n defined by

$$Q(v_1, v_2) = \mathrm{tr}_{K_1/\mathbb{Q}}(\zeta \cdot v_1 E v_2^*)$$

is positive definite.

Lemma 3.5. *There exists an invertible matrix $G \in M_{g \times g}(O_{K_1})$ such that GEG^* has entries all of whose conjugates are polynomially bounded in terms of* $\mathrm{Disc}(R)$.

Proof. Consider the quadratic form

$$Q(v_1, v_2) = \mathrm{tr}_{K_1/\mathbb{Q}}(\zeta \cdot v_1 E v_2^*)$$

as a positive-definite quadratic form on ${O_K}^{n_1}$ thought of as $\mathbb{Z}^{n_1 \cdot [K:\mathbb{Q}]}$. Since NE has entries which are algebraic integers, the smallest non-zero value Q can take is $1/N$. By repeating the proof of Lemma 3.2, we can find a basis of $n_1 \cdot [K:\mathbb{Q}]$ elements v_i in $O_K^{n_1}$ such that $Q(v_i, v_i)$ is polynomially bounded by the determinant of Q, which is in turn polynomially bounded in terms of $\mathrm{Disc}(R)$. Pick a subset

$$\{w_j, 1 \leqslant j \leqslant n_1\}$$

of the v_i which are linearly independent over K, and make them the rows of G. For $\phi \in S_1$, consider the positive-definite matrix $E_\phi := \phi(\zeta \cdot GEG^*)$. By construction, E_ϕ is Hermitian, positive definite, and has diagonal elements polynomially bounded in terms of $\mathrm{Disc}(R)$. Thus all the entries are automatically polynomially bounded in terms of $\mathrm{Disc}(R)$. As the conjugates of ζ^{-1} are also polynomially bounded in terms of $\mathrm{Disc}(R)$, this completes the proof. □

Take G as in Lemma 3.5, and note that G must have determinant polynomially bounded in terms of $\mathrm{Disc}(R)$. We can thus replace (I, E) by $(G^{-1}(I), GEG^*)$. Now we can pick a basis for $G^{-1}(I)$ of vectors whose entries have all their conjugates and denominators polynomially bounded in terms of $\mathrm{Disc}(R)$ as in Claim 3.2. The rest of the proof follows as in case (i). □

4. Ax–Lindemann–Weierstrass

The aim of this section is to prove Theorem 1.2. We begin by showing that we can restrict our attention to complex algebraic subvarieties (as defined in § 2). The following lemma strengthens a result obtained in [PZ08, Lemma 2.1] for curves, showing directly that semialgebraic subsets are contained in complex subvarieties, without fibring by curves. This is more convenient for us and may admit other applications.

Lemma 4.1. *Let $\Omega \subset \mathbb{C}^n$ be a connected open domain and let Z be a complex analytic subvariety of Ω. Let $W \subset Z$ be a maximal connected irreducible semialgebraic set. Then W is a complex algebraic subvariety of Ω.*

Proof. We may assume that $W \not\subset \mathrm{sing}(Z)$ (the singular locus; otherwise replace Z by $\mathrm{sing}(Z)$). Take U to be the Zariski closure of W, and let $O \in W$ be a smooth point of X. Let $m = \dim U = \dim W$. Let $z_1, z_2, \ldots, z_n$ be the usual co-ordinates on $\mathbb{C}^n$, with x_j, y_j being real co-ordinates such that $z_j = x_j + iy_j$ as usual. We first want to 'complexify' U into a complex variety inside $\mathbb{C}^n$. Since U is a real algebraic variety over $\mathbb{R}$, we consider the set of its complex points $U(\mathbb{C})$ as an abstract complex algebraic variety. Moreover, the inclusion map $i: U \to \mathbb{R}^{2n}$ is given by n pairs of polynomial maps (f_i, g_i) from U to $\mathbb{R}$, so that $i(u) = (f_1(u), g_1(u), \ldots, f_n(u), g_n(u))$. Thus we can consider the complexified map $i_{\mathbb{C}}: U(\mathbb{C}) \to \mathbb{C}^{2n}$ via

$$i_{\mathbb{C}}(u) = (f_1(u) + ig_1(u), \ldots, f_n(u) + ig_n(u)).$$

The map $i_{\mathbb{C}}$ is the identity map on the real points $U(\mathbb{R})$, and its image on the whole of $U(\mathbb{C})$ is a complex algebraic variety.[1]

Now, pick local *real* co-ordinates $u_1, \ldots, u_m$ for U around O, so that the u_i become *complex* co-ordinates for $U(\mathbb{C})$ around O. Define Y to be the pullback of Z along $i_{\mathbb{C}}$, so that

$$Y := i_{\mathbb{C}}^{-1}(Z \cap i_{\mathbb{C}} U(\mathbb{C})).$$

Then $O \in Y$, and locally around O in the co-ordinates u_i, Y is a complex manifold which contains $\mathbb{R}^m$. Since Y is a complex manifold its tangent space at O is a complex subspace, and

[1] For those familiar with Weil restriction, this simply reflects that Weil restriction is the right adjoint to the base change functor.

since it contains $\mathbb{R}^m$ it must be all of $\mathbb{C}^m$. Thus Y contains an open neighborhood of $U(\mathbb{C})$, and thus Z contains an open neighborhood of $i_{\mathbb{C}}(U(\mathbb{C}))$. Since W was assumed to be maximal, W must be of the same dimension as $i_{\mathbb{C}}(U(\mathbb{C}))$, and this completes the proof. □

We shall make use of the following 2 lemmas.

Lemma 4.2. *Suppose $W \subset \mathcal{A}_{2,1}$ is an algebraic variety such that $\pi^{-1}(W)$ has an algebraic component. Then W is weakly special.*

Proof. This is the main theorem of [UY11]. □

In view of Lemma 4.1, the condition that a component of $\pi^{-1}(W)$ is algebraic is equivalent to it being semialgebraic. In the following lemma, 'definable' means definable in any o-minimal expansion of the real field.

Lemma 4.3. *Let W be a complex algebraic variety, and $D \subset W$ be definable, complex analytic and closed in W. Then D is algebraic.*

Proof. By taking an affine open set in W, it suffices to consider the case where W is an affine subset of projective space. Now, one can express W as $M \backslash E$ where M is a projective variety and E is an algebraic subvariety of M. Theorem 5.3 in [PS11a] then implies that the closure of D in M is a definable, complex analytic subset of M, and thus D must be algebraic by Chow's theorem. □

Proof of Theorem 1.2. First, by Lemma 4.1 we can assume Y is a complex algebraic subvariety of Z. Note also that if $\dim_{\mathbb{C}} Z = \dim_{\mathbb{C}} Y$, then Z must equal Y and be semialgebraic itself, so that we are done by Lemma 4.2. We can thus assume that $\dim_{\mathbb{C}} Z = 2$ and $\dim_{\mathbb{C}} Y = 1$.

Define Z^0 to be the connected component of Z containing Y. Now consider a fundamental domain F_2 which intersects Z^0 and define $Z_0 = Z \cap F_2$. We know that Z_0 is definable by the main result of [PS11b].

We define

$$X = \{g \in \mathrm{Sp}_{2g}(\mathbb{R}) \mid \dim_{\mathbb{C}}(g \cdot Y \cap Z_0) = 1\},$$

so X is a definable subset of $\mathrm{Sp}_{2g}(\mathbb{R})$. Moreover, set $\Gamma_0 \subset \Gamma$ to be the monodromy group of V, so that Γ_0 preserves a connected component of Z. Then for all elements of $g \in \Gamma_0$ such that $Y \cap gF$ is not empty, we must have $g \in X$.

If V is not Hodge-generic in $\mathcal{A}_{2,1}$, it must be contained in a 2-dimensional special subvariety, which would mean that Z is special, contradicting the maximality of Y. Therefore V is Hodge-generic, and so Γ_0 is Zariski dense in $\mathrm{Sp}_4(\mathbb{R})$.

Now, since X is definable (in $\mathbb{R}_{\mathrm{an\ exp}}$) it admits an analytic cell decomposition [vdDM94]. Thus X is a union of finitely many definable components

$$X = \bigcup_{i=1}^{m} X_i,$$

such that each X_i is real-analytically homeomorphic to an open ball of some dimension. Note that some of the X_i may be points. By analytic continuation, we have $X_i \cdot Y \subset Z$ for all $1 \leqslant i \leqslant m$.

4.1 Case 1: $\forall 1 \leqslant i \leqslant m, \dim_{\mathbb{R}} X_i \cdot Y = 2$

Since everything is locally real analytic, we must have $\forall 1 \leqslant i \leqslant m$, $X_i \cdot Y = x_i \cdot Y$, where $x_i \in X_i$ is an arbitrary point.

https://doi.org/10.1112/S0010437X12000589 Published online by Cambridge University Press

Claim 4.1. *Under the assumptions above, $\pi(Y)$ is an algebraic subvariety of V.*

Proof of claim. We have that

$$\pi(Y)=\bigcup_{g\in\Gamma}\pi(Y\cap g\cdot F_2)=\bigcup_{g\in\Gamma}\pi(gY\cap F_2).$$

Now, if $g\in\Gamma$ and $gY\cap F_2\neq 0$, then in fact

$$Y\cap F_2\subset F_2\cap g\cdot Z=Z_0$$

and so $g\in X$. Thus, there exists an i with $g\in X_i$. We thus have that

$$\pi(Y)=\bigcup_{1\leqslant i\leqslant m}\pi(x_i\cdot Y\cap F_2),$$

and thus $\pi(Y)$ is a finite union of closed sets, and therefore closed. It is also the closed image of the definable complex analytic set $\bigcup_{1\leqslant i\leqslant m}(x_i\cdot Y\cap F_2)$ under the definable map π, and is thus definable by [PS11a, Theorem 5.6.2]. Thus $\pi(Y)$ is algebraic by Lemma 4.3. □

We now have that Y is a semialgebraic subvariety such that $\pi(Y)$ is also algebraic. By Lemma 4.2, Y must be special. This completes the proof in this case.

4.2 Case 2: for some $1\leqslant i\leqslant m$, $\dim_{\mathbb{R}} X_i\cdot Y>2$

Without loss of generality, we assume $\dim_{\mathbb{R}} X_1\cdot Y>2$. Take a small real analytic curve $I\subset X_1$, and consider a local complexification $I_{\mathbb{C}}\subset \mathrm{Sp}_4(\mathbb{C})$. Define Y^0 to be a connected component of $Y_2=I_{\mathbb{C}}\cdot Y\cap\mathbb{H}_2$. By analyticity, Y^0 is contained in Z. Moreover, the complex dimension of Y^0 must be at least 2, and so Y_2 is an open component of Z. Since $I_{\mathbb{C}}$ is definable, Y_2 is also definable. Now define

$$X_2:=\{g\in\mathrm{Sp}_4(\mathbb{R})\mid \dim_{\mathbb{C}} g\cdot Y_2\cap Z_0=2\}.$$

Note that for any point $g\in X_2$, we must have $g\cdot Z^0=Z^0$. We now prove that $X_2\cap\Gamma$ is infinite. Assume not. Since $X_2\cap\Gamma$ is finite, then $I\cdot Y$ intersects only finitely many fundamental domains. Pick $p\in I$, so that $p\cdot Y$ intersects finitely many fundamental domains and hence, by Lemma 4.2, we have that $p\cdot Y$ is a weakly special variety. But weakly special subvarieties are invariant by infinitely many elements of Γ and hence intersect infinitely many fundamental domains. This contradiction proves that $X_2\cap\Gamma$ is infinite.

Since X_2 is also definable, it must contain a real analytic curve $U\subset X_2$. Consider now the group

$$G_Z=\{g\in\mathrm{Sp}_4(\mathbb{R})\mid g\cdot Z^0=Z^0\}.$$

Thus G_Z contains a 1-parameter subgroup, and so the lie algebra $\mathrm{lie}(G_Z)$ is a positive dimensional vector space. Moreover, since $\Gamma_0\subset G_Z$, we must have $\mathrm{lie}(G_Z)$ is invariant under conjugation by Γ_0, and therefore also by the Zariski closure of Γ_0. Thus $\mathrm{lie}(G_Z)$ is invariant under conjugation by $\mathrm{Sp}_4(\mathbb{R})$. Since $\mathrm{Sp}_4(\mathbb{R})$ is simple, this means that $\mathrm{lie}(G_Z)=\mathrm{lie}(\mathrm{Sp}_4(\mathbb{R}))$, and so $G_Z=\mathrm{Sp}_4(\mathbb{R})$, which is a contradiction. □

5. Proof of Theorem 1.1

Proof of Theorem 1.1. Let $V\subset\mathcal{A}_{2,1}$ be a subvariety which is the Zariski-closure of the CM points inside it. Then V is defined over some number field K. Consider $Z=\pi^{-1}(V)$ and let $Z_0=F_2\cap Z$.

Then Z_0 is definable. Now, let $x \in V$ be a CM point and set $\{x_i\}_{1 \leqslant i \leqslant m}$ to be the Galois orbit of x over K, listed without repetitions. Set $w_i \in Z_0$ to be the pre-image of x_i in F_2, so that $\pi(w_i) = x_i$. By Theorem 1.3, there is a $\delta(2) > 0$ such that the heights of the x_i are at most

$$H(w_i) \ll m^{1/\delta(2)}.$$

Thus, we can conclude by Pila–Wilkie [PW06, Theorem 1.8] that at least one (in fact, most, but all we need is one) w_i is contained in a positive dimensional algebraic variety. By Theorem 1.2 this must be a weakly special subvariety. Thus all but finitely many CM points in V must have a Galois conjugate which is contained in a positive dimensional weakly special subvariety of V.

Since Galois conjugates of weakly special subvarieties are weakly special, we conclude that all but finitely many CM points lie on positive dimensional weakly special subvarieties S_i of V, which are then special by virtue of containing CM points. If V is 1 dimensional, than any special subvariety that V contains must in fact equal V. So we assume from now on that the dimension of V is 2, and each of the S_i has dimension 1. Assume for the sake of contradiction that V is not special.

Now, say S_i is a weakly special subvariety. Then there exist a semisimple subgroup $H_i \subset \mathrm{Sp}_4(\mathbb{R})$ and an element $z \in F_2$ such that $S_i = \pi(H_i \cdot z)$.

LEMMA 5.1. *The set of groups H_i is finite.*

Proof. There are finitely many semisimple lie algebras which embed into $\mathrm{lie}(\mathrm{Sp}_4(\mathbb{R}))$, and by [EMV09, Lemma A.1.1] these come in finitely many sets of conjugacy classes, so we can assume without loss of generality that there is a fixed semisimple lie group $H \subset \mathrm{Sp}_4(\mathbb{R})$ and elements $t_i \in \mathrm{Sp}_4(\mathbb{R})$ with $H_i = t_i H t_i^{-1}$. Now, as S_i is a special subvariety, the group $\Gamma_i = H_i \cap \mathrm{Sp}_4(\mathbb{Z})$ is Zariski-dense in H_i. Since Γ_i is also finitely generated, the set of such groups is countable and hence the set of possible H_i is countable.

Now, consider the set

$$B = \{(t, z) \in \mathrm{Sp}_4(\mathbb{R}) \times F_2 \mid tHt^{-1} \cdot z \subset Z^0\},$$

which is definable. If $(t, z) \in B$, then either $tHt^{-1}z$ is special, or by Theorem 1.2 it must be contained in a special variety. But by dimension considerations, that special variety must then have dimension at least 2, and so it must be V. Since we are assuming that V is not special, we conclude that the special subvarieties S_i are precisely the images of $tHt^{-1} \cdot z$ for $(t, z) \in B$.

Since a countable definable set (in $\mathbb{R}$) is finite, this proves the lemma. □

By Lemma 5.1 there are finitely many groups $H_1, \ldots, H_m$ such that every weakly special subvariety contained in Z which intersects the upper half plane is an H_i orbit. Define U to be the pre-image of all weakly special subvarieties in V restricted to the fundamental domain F_2 so that by the above

$$U = \{w \in Z_0 \mid \exists i \in \{1, 2, \ldots, m\}, H_i \cdot w \subset Z\}.$$

We therefore have that U is definable. Moreover, since U is contained in Z^0 its dimension is everywhere locally at most 2. Moreover, since U cannot be a finite union of weakly special subvarieties of dimension 1, its dimension must somewhere be 2. Now, let $W_i, i \in \mathbb{N}$ denote the countably many special subvarieties of $\mathcal{A}_{2,1}$ which have dimension at most 2. Every weakly special subvariety of $\mathcal{A}_{2,1}$ is contained in one of the W_i, so we know that

$$U = \bigcup_{i \in \mathbb{N}} U \cap \pi^{-1}(W_i).$$

https://doi.org/10.1112/S0010437X12000589 Published online by Cambridge University Press

But now, if V is not special then $\bigcup_{i\in\mathbb{N}} V \cap W_i$ is a countable union of algebraic varieties of dimension at most 1. Since U must somewhere have dimension 2, this is a contradiction. □

References

Del79 P. Deligne, *Variétés de Shimura: interpretation modulaire, et techniques de construction de modèles canoniques*, Proc. Symp. Pure Math. **33**(part 2) (1979), 247–290.

EMV09 M. Einsiedler, G. A. Margulis and A. Venkatesh, *Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces*, Invent. Math. **177** (2009), 137–212.

GV95 A. Galligo and N. Vorobjov, *Complexity of finding irreducible components of a semialgebraic set*, J. Complexity **11** (1995), 174–193.

PS11a Y. Peterzil and S. Starchenko, *Tame complex analysis and o-minimality*, in *Proceedings of The International Congress of Mathematicians 2010 (ICM 2010): Vol. 2, Hyderabad, India* (World Scientific, 2011).

PS11b Y. Peterzil and S. Starchenko, *Definability of restricted theta functions and families of abelian varieties*, Duke Math. J., to appear, http://arxiv.org/abs/1103.3110.

Pil11 J. Pila, *O-minimality and the André–Oort conjecture for $\mathbb{C}^n$*, Annals Math. **173** (2011), 1779–1840.

PW06 J. Pila and A. Wilkie, *The rational points of a definable set*, Duke Math. J. **133** (2006), 591–616.

PZ08 J. Pila and U. Zannier, *Rational points in periodic analytic sets and the Manin–Mumford conjecture*, Rend. Lincei Mat. Appl. **19** (2008), 149–162.

Sie64 C-L. Siegel, *Symplectic geometry* (Academic Press, New York–London, 1964).

Tsi12 J. Tsimerman, *Brauer–Siegel for arithmetic tori and lower bounds for Galois orbits of special points*, J. Amer Math. Soc. **25** (2012), 1091–1117.

UY E. Ullmo and A. Yafaev, *Galois orbits and equidistribution of special subvarieties: towards the André–Oort conjecture*, http://www.math.u-psud.fr/∼ullmo/.

UY11 E. Ullmo and A. Yafaev, *A characterization of special subvarieties*, Mathematika **57** (2011), 263–273.

vdDM94 L. van den Dries and C. Miller, *On the real exponential field with restricted analytic functions*, Israel J. Math. **85** (1994), 19–56.

Jonathan Pila Jonathan.Pila@maths.ox.ac.uk
Mathematical Institute, University of Oxford, 24-29 St Giles, Oxford OX1 3LB, UK

Jacob Tsimerman jacobt@math.harvard.edu
Mathematics Department, Harvard University, One Oxford Street, Cambridge, MA 02138, USA

https://doi.org/10.1112/S0010437X12000589 Published online by Cambridge University Press