Quantitative reduction theory and unlikely intersections

C. Daw and M. Orr (2021) "Quantitative Reduction Theory and Unlikely Intersections,"
International Mathematics Research Notices, Vol. 00, No. 0, pp. 1–58
https://doi.org/10.1093/imrn/rnab173

Quantitative Reduction Theory and Unlikely Intersections

Christopher Daw[1] and Martin Orr[2]

[1]Department of Mathematics and Statistics, University of Reading, Whiteknights, PO Box 217, Reading, Berkshire RG6 6AH, UK and
[2]Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, UK

*Correspondence to be sent to: e-mail: chris.daw@reading.ac.uk

We prove quantitative versions of Borel and Harish-Chandra's theorems on reduction theory for arithmetic groups. Firstly, we obtain polynomial bounds on the lengths of reduced integral vectors in any rational representation of a reductive group. Secondly, we obtain polynomial bounds in the construction of fundamental sets for arithmetic subgroups of reductive groups, as the latter vary in a real conjugacy class of subgroups of a fixed reductive group. Our results allow us to apply the Pila–Zannier strategy to the Zilber–Pink conjecture for the moduli space of principally polarised abelian surfaces. Building on our previous paper, we prove this conjecture under a Galois orbits hypothesis. Finally, we establish the Galois orbits hypothesis for points corresponding to abelian surfaces with quaternionic multiplication, under certain geometric conditions.

1 Introduction

Reduction theory is concerned with finding small representatives for each orbit in actions of arithmetic groups, for example through constructing fundamental sets. It began with the study of the action of $\mathrm{SL}_n(\mathbb{Z})$ on quadratic forms, which was described by Siegel in terms of a fundamental set for $\mathrm{SL}_n(\mathbb{Z})$ in $\mathrm{SL}_n(\mathbb{R})$. Borel and Harish-Chandra generalised this to arithmetic lattices in arbitrary semisimple Lie groups. This theory

Received June 26, 2020; Revised March 19, 2021; Accepted June 9, 2021
Communicated by Jonathan Pila

has had wide-ranging applications in areas such as the theory of automorphic forms and locally symmetric spaces [1], the arithmetic of algebraic groups [32] and finiteness theorems for abelian varieties [22].

The first goal of this paper is to prove quantitative bounds for the group elements used in Borel and Harish-Chandra's construction of fundamental sets. These bounds are polynomial in terms of suitable input parameters, although they are not fully effective. They generalise the polynomial bounds of Li and Margulis for the reduction theory of quadratic forms [21] and complement the second-named author's polynomial bounds for the Siegel property [29]. It should be noted that while Borel and Harish-Chandra's reduction theory is algorithmic in nature, as made explicit by Grunewald and Segal [17], their arguments give no bounds for the running time or output size of these algorithms.

Our primary theorems on reduction theory are as follows. The first is a quantitative version of [4, Lemma 5.4]. See Section 3 for the relevant definitions and Section 2.3 for discussion of how this theorem is related to [4].

Theorem 1.1. Let $\mathbf{G}$ be a reductive $\mathbb{Q}$-algebraic group and let $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ be a Siegel set. Let $\rho\colon \mathbf{G} \to \mathbf{GL}(V)$ be a representation of $\mathbf{G}$ defined over $\mathbb{Q}$. Let $\Lambda \subset V$ be a $\mathbb{Z}$-lattice. Let $v_0 \in V_{\mathbb{R}}$ be such that:

(i) $\rho(\mathbf{G}(\mathbb{R}))v_0$ is closed in $V_{\mathbb{R}}$;

(ii) the stabiliser $\mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(v_0)$ is self-adjoint.

Then, there exist constants c_1, c_2 such that, for every $v \in \mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})v_0$ and every $w \in \rho(\mathfrak{S})v \cap \Lambda$, we have $|w| \leq c_1|v|^{c_2}$.

We use Theorem 1.1 to prove our second theorem on quantitative reduction theory, which is a quantitative version of Borel and Harish-Chandra's construction of fundamental sets for arithmetic groups [4, Thm. 6.5]. See Section 2.4 for discussion of the details of this theorem, including how it relates to [4].

Theorem 1.2. Let $\mathbf{G}$ be a reductive $\mathbb{Q}$-algebraic group. Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. Let $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ be a Siegel set such that $C\mathfrak{S}$ is a fundamental set for Γ in $\mathbf{G}(\mathbb{R})$, for some finite set $C \subset \mathbf{G}(\mathbb{Q})$.

Let $\rho\colon \mathbf{G} \to \mathbf{GL}(\Lambda_{\mathbb{Q}})$ be a $\mathbb{Q}$-algebraic representation of $\mathbf{G}$, where Λ is a finitely generated free $\mathbb{Z}$-module. Let $\mathbf{H}_0 \subset \mathbf{G}$ be a self-adjoint reductive $\mathbb{Q}$-algebraic subgroup and let $v_0 \in \Lambda$ be a vector such that:

(i) $\mathrm{Stab}_{\mathbf{G},\rho}(v_0) = \mathbf{H}_0$;

(ii) the orbit $\rho(\mathbf{G}(\mathbb{R}))v_0$ is closed in $\Lambda_{\mathbb{R}}$.

Then, there exist positive constants c_3 and c_4 (depending only on $\mathbf{G}$, Γ, $\mathfrak{S}$, C, ρ, $\mathbf{H}_0$ and v_0) with the following property: for every $u \in \mathbf{G}(\mathbb{R})$ and $v_u \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})v_0$ such that $\mathbf{H}_u = u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$ and $\rho(u)v_u \in \Lambda$, there exists a fundamental set for $\Gamma \cap \mathbf{H}_u(\mathbb{R})$ in $\mathbf{H}_u(\mathbb{R})$ of the form

$$B_u C \mathfrak{S} u^{-1} \cap \mathbf{H}_u(\mathbb{R}),$$

where $B_u \subset \Gamma$ is a finite set such that every $b \in B_u$ satisfies

$$|\rho(b^{-1}u)v_u| \leq c_3 |v_u|^{c_4}.$$

We apply Theorem 1.2 to the Zilber–Pink conjecture on unlikely intersections. We prove the Zilber–Pink conjecture for $\mathcal{A}_2$, the (coarse) moduli space of principally polarised abelian surfaces over $\mathbb{C}$, subject to a large Galois orbits conjecture (Conjecture 6.2), which is stated in Section 6.

We recall that $\mathcal{A}_2$ is a Shimura variety of dimension 3 associated with the group $\mathbf{GSp}_4$. The Zilber–Pink conjecture predicts that an irreducible algebraic curve $C \subset \mathcal{A}_2$ that is Hodge generic (i.e. not contained in any proper special subvariety) contains only finitely many points of intersection with the special subvarieties of $\mathcal{A}_2$ having dimension 1 or 0 (see [31, Conjecture 1.3] for the most relevant formulation of the Zilber–Pink conjecture).

Pila and Tsimerman's proof of the André–Oort conjecture for $\mathcal{A}_2$ [33] shows that C contains only finitely many special points. (A special point on $\mathcal{A}_2$ is a point associated with an abelian surface with complex multiplication.) Therefore, in order to prove the Zilber–Pink conjecture for $\mathcal{A}_2$, it suffices to show that C contains only finitely many non-special points belonging also to a special curve.

The special curves in $\mathcal{A}_2$ are of three types:

(1) curves parametrising abelian surfaces with quaternionic multiplication (we refer to these as quaternionic curves);

(2) curves parametrising abelian surfaces isogenous to the square of an elliptic curve ("E^2 curves");

(3) curves parametrising abelian surfaces isogenous to the product of two elliptic curves, at least one of which has complex multiplication ("$E \times \mathrm{CM}$ curves").

In this paper, we study intersections with the quaternionic and E^2 curves. Let Σ_{Quat} (resp. Σ_{E^2}) denote the set of points of $\mathcal{A}_2$ which are Hodge generic in some quaternionic (resp. E^2) curve. Our first main result on unlikely intersections is the following.

Theorem 1.3. Let Σ denote Σ_{Quat} or Σ_{E^2} and let $C \subset \mathcal{A}_2$ denote an irreducible Hodge generic algebraic curve.

If C satisfies Conjecture 6.2 for Σ, then $C \cap \Sigma$ is finite.

Combined with our previous work [12], in which we study intersections with the $E \times \mathrm{CM}$ curves, and Pila and Tsimerman's proof of the André–Oort conjecture for $\mathcal{A}_2$, this completes the proof of the Zilber–Pink conjecture for $\mathcal{A}_2$, subject to Conjecture 6.2 (a large Galois orbits conjecture). As in [12], the general strategy follows the proof of [13, Theorem 14.2], which was an application of the so-called Pila–Zannier method to the Zilber–Pink conjecture for general Shimura varieties. However, we will have to make several modifications, and the end of the proof is closer to [30, Proposition 3.5].

Finally, we show that the large Galois orbits conjecture holds for $\Sigma = \Sigma_{\mathrm{Quat}}$ when the curve under consideration satisfies a multiplicative reduction hypothesis at the boundary of the moduli space. We proved the analogous result for intersections with $E \times \mathrm{CM}$ curves in [12].

Theorem 1.4. Let $C \subset \mathcal{A}_2$ denote an irreducible Hodge generic algebraic curve defined over $\overline{\mathbb{Q}}$ such that the Zariski closure of C in the Baily–Borel compactification of $\mathcal{A}_2$ intersects the 0-dimensional stratum of the boundary.

Then, C satisfies Conjecture 6.2 for Σ_{Quat}, and so, $C \cap \Sigma_{\mathrm{Quat}}$ is finite.

Theorem 1.4 follows from a result of André [2, Ch. X, Thm. 1.3] and the Masser–Wüstholz isogeny theorem. We have not been able to prove an analogue of Theorem 1.4 for $\Sigma = \Sigma_{E^2}$ because the result of André does not apply to abelian surfaces isogenous to the square of an elliptic curve.

We note that the results on quantitative reduction theory in this paper will be an important tool for proving the Zilber–Pink conjecture for other Shimura varieties, which will be the subject of future work by the authors. We expect these results to have further applications, for example a uniform version of the second-named author's bounds for polarisations and isogenies of abelian varieties [28] and bounds for the heights of generators of arithmetic groups by combining them with some of the techniques of homogeneous dynamics from [21].

1.1 Outline of the paper

In Section 2, we give some background on reduction theory to put our Theorems 1.1 and 1.2 into context.

In Section 3, we define notation to be used throughout the paper. We state the various (equivalent) definitions of a Cartan involution that exist in the literature, and we define the notion of a Siegel set.

In Section 4, we prove our main theorems on quantitative reduction theory, namely, Theorems 1.1 and 1.2. The proof follows the stategy of [4] but, in order to obtain a quantitative result, it is necessary to replace "soft" topological ingredients in the former, notably [4, Prop. 5.2], with results from elsewhere.

In order to apply Theorem 1.2 to a specific situation, we must construct a representation ρ of the ambient reductive group and show that the vectors v_u can be chosen suitably bounded. This is the topic of Section 5—we construct a representation of $\mathbf{GSp}_4$ with the properties required to apply Theorem 1.2 to the subgroups associated with quaternionic and E^2 curves.

In Section 6, we prove our theorems on unlikely intersections, namely, Theorems 1.3 and 1.4. The strategy for the former is to parametrise the unlikely intersections by integral vectors of suitably bounded length (or, equivalently, height). This parametrisation is obtained using the results of Sections 4 and 5. Then, as in all versions of the Pila–Zannier method, we consider the parameters for unlikely intersections as a set definable in an o-minimal structure and apply a Pila–Wilkie counting theorem to control the number of such points in terms of their height. Theorem 1.3 follows by comparing this upper bound with the lower bound of the large Galois orbits conjecture.

We emphasise that Sections 2–5 require no knowledge of Shimura varieties and may be read independently of Section 6, while the results of Section 2 are not used elsewhere in the paper.

2 Background on reduction theory

In this section, we outline some of the history of reduction theory, focussing on quantitative results and on the work of Borel and Harish-Chandra, and explain how our Theorems 1.1 and 1.2 fit into this theory. The results of this section are not used later in the paper.

2.1 Reduction theory for quadratic forms

The group $\mathbf{SL}_n(\mathbb{Z})$ acts on the set of integral quadratic forms in n variables via its natural action on the variables. The classical reduction theory of quadratic forms defines a set of *reduced quadratic forms* with the following properties.

Properties 2.1.

(i) Each $\mathbf{SL}_n(\mathbb{Z})$-orbit of non-degenerate integral quadratic forms contains at least one reduced form.

(ii) Each $\mathbf{SL}_n(\mathbb{Z})$-orbit of non-degenerate integral quadratic forms contains only finitely many reduced forms.

A variety of definitions of reduced quadratic forms are used, possessing varying properties in addition to Properties 2.1. The most important definitions are due to Lagrange and Gauss for binary quadratic forms, and to Hermite, Minkowski and Siegel for quadratic forms in any number of variables. In these definitions (except some of Hermite's definitions), the reduced quadratic forms can be defined by finitely many polynomial inequalities in the coefficients of the forms.

Reduction theory behaves better for positive (or negative) definite quadratic forms than for indefinite forms. Definite quadratic forms satisfy a much stronger version of Property 2.1(ii) called the Siegel property. This guarantees that there is a uniform bound on the number of reduced quadratic forms in each $\mathbf{SL}_n(\mathbb{Z})$-orbit (for fixed n). In the nicest case of all, definite binary quadratic forms using Gauss's definition of reduced forms, each $\mathbf{SL}_2(\mathbb{Z})$-orbit contains *exactly one* reduced form.

Quantitative reduction theory for quadratic forms

The discriminant of a quadratic form is invariant under the action of $\mathbf{SL}_n(\mathbb{Z})$. A form is non-degenerate if and only if its discriminant is non-zero. Hence the following lemma implies Property 2.1(i).

Lemma 2.2. For each integer $\Delta \neq 0$, there are only finitely many reduced integral quadratic forms in n variables of discriminant Δ.

The following quantitative version of Lemma 2.2 is classical for anisotropic quadratic forms [8, p. 287, Cor. 1], using Hermite and Minkowski's definitions of reduction (note that anisotropic forms are always definite when $n \geq 5$). Indeed, for binary anisotropic forms, it goes back to Lagrange. For Siegel reduced forms (definite

or indefinite), it can be proved by adapting the proof of Lemma 2.2 found on [8, pp. 322–4] but we are unsure whether this was classically known.

Proposition 2.3. For each positive integer n, there exists a positive real number $c_5(n)$ such that, for every integer $\Delta \neq 0$, all reduced integral n-ary quadratic forms of discriminant Δ have coefficients with absolute values at most $c_5(n)|\Delta|$.

Lemma 2.2 also implies the following lemma, which does not mention reduced forms, and which was historically one of the most important consequences of reduction theory.

Lemma 2.4. For each integer $\Delta \neq 0$, there are only finitely many $\mathbf{SL}_n(\mathbb{Z})$-orbits of integral quadratic forms in n variables of discriminant Δ.

Li and Margulis have proved a quantitative version of Lemma 2.4. The bound is stronger than in Proposition 2.3, but it applies only to at least one form in each $\mathbf{SL}_n(\mathbb{Z})$-orbit, rather than to all reduced forms.

Proposition 2.5. [21, Theorem 3] For each integer $n \geq 3$, there exists a constant $c_6(n)$ such that every $\mathbf{SL}_n(\mathbb{Z})$-orbit of indefinite quadratic forms in n variables of discriminant $\Delta \neq 0$ contains a form whose coefficients have absolute value at most $c_6(n)|\Delta|^{1/n}$.

2.2 Siegel sets

Siegel shifted the emphasis in reduction theory from quadratic forms to arithmetic groups. There is a direct link between the reduction theory of definite quadratic forms and fundamental sets for $\mathbf{SL}_n(\mathbb{Z})$ in $\mathbf{SL}_n(\mathbb{R})$.

Let v_0 denote the standard positive definite quadratic form in n variables

$$v_0(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2.$$

If $\mathcal{F}$ is a fundamental set of *positive definite real* quadratic forms in n variables (that is, a set which satisfies the generalisations of Property 2.1 for positive definite real forms), then

$$\mathfrak{S} = \{g \in \mathbf{SL}_n(\mathbb{R}) : gv_0 \in \mathcal{F}\}$$

is a fundamental set in $\mathbf{SL}_n(\mathbb{R})$ for the action of $\mathbf{SL}_n(\mathbb{Z})$ by multiplication on the left (i.e. every right $\mathbf{SL}_n(\mathbb{Z})$-coset intersects $\mathfrak{S}$ in at least one, and at most finitely many, elements). Conversely, if $\mathfrak{S} \subset \mathbf{SL}_n(\mathbb{R})$ is a fundamental set for $\mathbf{SL}_n(\mathbb{Z})$ which is invariant under right multiplication by $\mathrm{SO}_n(\mathbb{R}) = \mathrm{Stab}_{\mathbf{SL}_n(\mathbb{R})}(v_0)$, then $\mathbb{R}_{>0}\mathfrak{S}v_0$ is a fundamental set of positive definite real quadratic forms.

Siegel defined a family of sets $\mathfrak{S} = \mathfrak{S}_{t,u} \subset \mathbf{SL}_n(\mathbb{R})$, depending on two parameters $t, u \in \mathbb{R}_{>0}$. The set $\mathfrak{S}_{t,u}$ is a fundamental set for $\mathbf{SL}_n(\mathbb{Z})$ whenever $t \leq \sqrt{3}/2$ and $u \geq 1/2$ (according to the conventions used in this paper). We call $\mathfrak{S}_0 = \mathfrak{S}_{\sqrt{3}/2,1/2}$ the **standard Siegel set** in $\mathbf{SL}_n(\mathbb{R})$. Using the construction described in the previous paragraph, we obtain a fundamental set of positive definite real quadratic forms, namely $\mathbb{R}_{>0}\mathfrak{S}_0 v_0$. We say that a positive definite quadratic form is **Siegel reduced** if it lies in $\mathbb{R}_{>0}\mathfrak{S}_0 v_0$.

We also say that an indefinite quadratic form of signature (p, q) (with $p + q = n$) is **Siegel reduced** if it lies in $\mathbb{R}_{>0}\mathfrak{S}_0 v_0^{(p,q)}$, where

$$v_0^{(p,q)}(x_1, \ldots, x_{p+q}) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2. \tag{1}$$

The Siegel reduced indefinite *integral* quadratic forms satisfy Properties 2.1. However, the set of Siegel reduced indefinite real quadratic forms is not a fundamental set because it does not satisfy the generalisation of Property 2.1(ii) to real forms.

Borel and Harish-Chandra generalised the notion of Siegel set from $\mathbf{SL}_n(\mathbb{R})$ to all reductive Lie groups [4, Sec. 4.1] (which they called **Siegel domains**). However, these Siegel domains are not always fundamental sets for arithmetic subgroups — in general, one can only say that there is a fundamental set *contained in* a finite union of translates of Siegel domains [4, Thm. 6.5, Lemma 7.5].

Borel subsequently gave a new definition of Siegel sets for reductive $\mathbb{Q}$-algebraic groups [6,12.3], taking into account the $\mathbb{Q}$-algebraic group structure and not just the Lie group structure. For each reductive $\mathbb{Q}$-algebraic group $\mathbf{G}$ and each arithmetic subgroup $\Gamma \subset \mathbf{G}(\mathbb{Q})$, there is a finite union of $\mathbf{G}(\mathbb{Q})$-translates of a Siegel set which forms a fundamental set for Γ in $\mathbf{G}(\mathbb{R})$ [6, Thm. 13.1] (this is a consequence of Theorem 2.7 below). In this paper, we shall use a minor modification of Borel's definition of Siegel sets, described in Section 3.3.

2.3 Reduction theory for representations of reductive groups

The following result is a key step in Borel and Harish-Chandra's construction of fundamental sets for arithmetic groups.

Theorem 2.6. [4, Lemma 5.4] Let $\mathbf{G}$ be a reductive $\mathbb{Q}$-algebraic group whose $\mathbb{Q}$-rank is equal to its $\mathbb{R}$-rank. Let $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ be a Siegel set. Let $\rho\colon \mathbf{G} \to \mathbf{GL}(V)$ be a representation of $\mathbf{G}$ defined over $\mathbb{Q}$. Let $\Lambda \subset V$ be a $\mathbb{Z}$-lattice. Let $v \in V_{\mathbb{R}}$ be such that:

(i) the orbit $\rho(\mathbf{G}(\mathbb{R}))v$ is closed in $V_{\mathbb{R}}$;

(ii) the stabiliser $\mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(v)$ is self-adjoint.

Then, $\rho(\mathfrak{S})v \cap \Lambda$ is finite.

The restriction on the $\mathbb{Q}$-rank of $\mathbf{G}$ in Theorem 2.6 can be removed with only minor alterations to the proof, provided we use the definition of Siegel sets from [6, 12.3] (or the definition in Section 3.3 of this paper) instead of the definition of Siegel domains from [4, 4.1].

As noted in [4, Example 5.5], Theorem 2.6 implies Lemma 2.2, by applying it to the representation of $\mathbf{SL}_n(\mathbb{R})$ on the vector space of real quadratic forms in n variables, with $v = \lambda v_0^{(p,q)}$, where $\lambda \in \mathbb{R}_{>0}$ and $v_0^{(p,q)}$ is defined by equation (1). Then, the orbit $\mathbf{SL}_n(\mathbb{R})v$ is the set of all quadratic forms of signature (p,q) and discriminant $(-1)^q\lambda^n$, and $\mathfrak{S}v \cap \Lambda$ is the set of Siegel reduced integral quadratic forms of given signature and discriminant.

In general, we may think of $\rho(\mathfrak{S})v$ as a set of "reduced vectors" in the representation $V_{\mathbb{R}}$. However, we should note that this set depends on v.

Quantitative reduction theory for representations

Theorem 1.1 is a quantitative version of Theorem 2.6, bounding the length of "reduced integral vectors," that is, elements of the finite set $\rho(\mathfrak{S})v \cap \Lambda$, in terms of v (for fixed group $\mathbf{G}$ and representation ρ). We are not able to prove such a bound for all $v \in V_{\mathbb{R}}$: we must restrict to a set of v for which $\mathbf{G}(\mathbb{R})$ acts "in a similar way" on all of the permitted vectors v. This is achieved through the condition that v must lie in the $\mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$-orbit of a fixed vector.

For an example application of Theorem 1.1, let ρ be the representation of $\mathbf{G} = \mathbf{SL}_n$ on the quadratic forms in n variables, and let $v_0 = v_0^{(p,q)}$. Noting that every scalar $\lambda \in \mathbb{R}^\times$ is in $\mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$, we deduce that there are constants c_7 and c_8 (depending on n) such that

$$|w| < c_7|\lambda|^{c_8} \text{ for all } \lambda \in \mathbb{R}^\times \text{ and } w \in \mathfrak{S}\lambda v_0 \cap L.$$

Thus, Theorem 1.1 implies a weakened version of Proposition 2.3—Proposition 2.3 is stronger because it gives a bound which is linear in the discriminant, while the constants in Theorem 1.1, even the exponent, are ineffective (see Remark 4.3).

In this example, the closed orbits in $V_{\mathbb{R}}$ are those which consist of non-degenerate quadratic forms. These orbits are parameterised by their signature and discriminant, so every closed orbit intersects $\mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})v_0^{(p,q)}$ for some signature (p,q). Thus, in this case, Theorem 1.1 is sufficient to give a polynomial bound for integral elements of a reduced set in every closed orbit. In general, however, there is no finite subset of $V_{\mathbb{R}}$ whose $\mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$-orbit intersects every closed $\mathbf{G}(\mathbb{R})$-orbit, and then, Theorem 1.1 does not allow us to compare all closed orbits.

In this example, the representation ρ is absolutely irreducible so its only endomorphisms are scalars. In general, there may be more endomorphisms of ρ, and it will be important for our applications that we allow v to be any element of $\mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})v_0$, not just a scalar multiple of v_0.

2.4 Fundamental sets for arithmetic groups

The central result of Borel and Harish-Chandra's reduction theory was the construction of fundamental sets for $\Gamma_{\mathbf{H}}\backslash\mathbf{H}(\mathbb{R})$, where $\mathbf{H}$ is a reductive $\mathbb{Q}$-algebraic group and $\Gamma_{\mathbf{H}} \subset \mathbf{H}(\mathbb{Q})$ is an arithmetic subgroup. These fundamental sets are constructed by embedding $\mathbf{H}$ into some $\mathbf{GL}_n$, where we already know how to obtain fundamental sets using standard Siegel sets. (Note that [4, Thm. 6.5] used the notation $\mathbf{G}$ where we write $\mathbf{H}$ in this theorem.)

Theorem 2.7. [4, Thm. 6.5] Let $\mathbf{H}$ be a reductive $\mathbb{Q}$-algebraic subgroup of $\mathbf{GL}_{n,\mathbb{Q}}$ and let $\Gamma_{\mathbf{H}} = \mathbf{GL}_n(\mathbb{Z}) \cap \mathbf{H}(\mathbb{R})$. Let $\mathfrak{S}_0$ be the standard Siegel set in $\mathbf{GL}_n(\mathbb{R})$. Let $u \in \mathbf{GL}_n(\mathbb{R})$ be such that $u^{-1}\mathbf{H}(\mathbb{R})u$ is self-adjoint.

Then, there exists a finite set $B \subset \mathbf{GL}_n(\mathbb{Z})$ such that

$$B\mathfrak{S}_0 u^{-1} \cap \mathbf{H}(\mathbb{R})$$

is a fundamental set for $\Gamma_{\mathbf{H}}$ in $\mathbf{H}(\mathbb{R})$.

The ambient group $\mathbf{GL}_n$ in Theorem 2.7 can be replaced by an arbitrary reductive $\mathbb{Q}$-algebraic group $\mathbf{G}$ containg $\mathbf{H}$ with only minor alterations to the proof, where $\mathfrak{S}_0$ is replaced by a sufficiently large Siegel set in $\mathbf{G}(\mathbb{R})$.

Quantitative fundamental sets for arithmetic groups

Theorem 1.2 is a quantitative version of Theorem 2.7, where $\mathbf{H}$ varies over the $\mathbb{Q}$-algebraic members of a $\mathbf{G}(\mathbb{R})$-conjugacy class of subgroups of some fixed reductive group $\mathbf{G}$. This theorem is "quantitative" in the sense that it controls a measure of the size of the elements of the finite set B. Ideally, we would like to bound the height of elements of B but we have not yet achieved this (it may be possible by combining the methods of this paper with tools of homogeneous dynamics as in [21]). Instead, we measure the size of elements of B in terms of how they act on a vector v_u (whose stabiliser is $\mathbf{H}$) in a suitable representation of $\mathbf{G}$. This turns out to be sufficient for our applications to unlikely intersections.

The theorem will only apply to subgroups $\mathbf{H} \subset \mathbf{G}$ which are defined over $\mathbb{Q}$ because these are the subgroups for which $\Gamma \cap \mathbf{H}(\mathbb{R})$ is a lattice in $\mathbf{H}(\mathbb{R})$. However, it is very important that $\mathbf{H}$ varies over a $\mathbf{G}(\mathbb{R})$-conjugacy class, not just a $\mathbf{G}(\mathbb{Q})$-conjugacy class, because this allows the $\mathbb{Q}$-algebraic subgroups in the conjugacy class to belong to more than one isomorphism class over $\mathbb{Q}$. A striking consequence of allowing this is that the conjugacy class may contain both $\mathbb{Q}$-anisotropic and $\mathbb{Q}$-isotropic groups, so the fundamental set in $\mathbf{H}(\mathbb{R})$ is sometimes compact and sometimes not compact, yet the same bounds apply to fundamental sets for all $\mathbf{H}$ in the conjugacy class. For example, $\mathbf{SL}_{2,\mathbb{Q}}$ and unit groups of quaternion algebras can be found in the same $\mathbf{SL}_4(\mathbb{R})$-conjugacy class of subgroups of $\mathbf{SL}_4$.

Note also that the semisimple subgroups of $\mathbf{G}$ belong to only finitely many $\mathbf{G}(\mathbb{R})$-conjugacy classes [3, Cor. 0.2]. This is not true for reductive subgroups, as may be seen by considering the torus $\mathbb{G}_m^2$, which contains infinitely many non-conjugate subgroups isomorphic to $\mathbb{G}_m$ – see [3, Remark 1.2].

For an example application of Theorem 1.2, consider the case where $\mathbf{G} = \mathbf{GL}_n$ and $\mathbf{H}_0 \subset \mathbf{GL}_n$ is the orthogonal group of the quadratic form $v_0^{p,q}$. The representation $\rho\colon \mathbf{G} \to \mathbf{GL}(\Lambda_{\mathbb{Q}})$, where Λ is the $\mathbb{Z}$-module of integral quadratic forms of signature (p,q), satisfies the conditions of Theorem 1.2. In particular, (iii) holds because if $\mathbf{H}_u$ is defined over $\mathbb{Q}$ then it is the orthogonal group of the integral quadratic form $v = \rho(u)\lambda v_0^{(p,q)}$ for some $\lambda \in \mathbb{R}^\times$.

As noted in [4, 6.7], the space of Hermite majorants of v is

$$\Sigma_u = \mathbf{H}_u(\mathbb{R})/(u\mathbf{O}_n(\mathbb{R})u^{-1} \cap \mathbf{H}_u(\mathbb{R})).$$

The image of $B_u C \mathfrak{S} u^{-1} \cap \mathbf{H}_u(\mathbb{R})$ in Σ_u is a fundamental set for $\Gamma \cap \mathbf{H}_u(\mathbb{R})$ in Σ_u. Theorem 1.2 allows us to control the set B_u used to construct this fundamental set in

that for each $b \in B_u$, the coefficients of the quadratic form $\rho(b^{-1})v$ are polynomially bounded in terms of $\mathrm{disc}(v)$. A related result can be found in [21, Section 9.5], which bounds the entries of the matrices $b \in B_u$ (stronger than bounding $\rho(b^{-1})v$), although its bound involves the coefficients of v as well as the discriminant.

As with Theorem 1.1, the constants in Theorem 1.2 are ineffective.

3 Preliminaries

3.1 Notation

If Λ is a $\mathbb{Z}$-module, we write $\Lambda_{\mathbb{Q}}$ for $\Lambda \otimes_{\mathbb{Z}} \mathbb{Q}$ and $\Lambda_{\mathbb{R}}$ for $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$. If Λ is free and finitely generated and $\phi\colon \Lambda_{\mathbb{Q}} \times \Lambda_{\mathbb{Q}} \to \mathbb{Q}$ is a $\mathbb{Q}$-bilinear form, we denote by $\mathrm{disc}(\Lambda, \phi)$ the determinant of the matrix $(\phi(e_i, e_j))_{i,j}$, where $\{e_1, \dots, e_n\}$ is a $\mathbb{Z}$-basis for Λ (the determinant is independent of the choice of basis).

If R is an order in a semisimple $\mathbb{Q}$-algebra D, then we write $\mathrm{disc}(R)$ for the discriminant of the $\mathbb{Z}$-module R with respect to the bilinear form $\phi(x, y) = \mathrm{Tr}_{D/\mathbb{Q}}(xy)$ where $\mathrm{Tr}_{D/\mathbb{Q}}$ is the (nonreduced) trace of the regular representation of D. See Section 5.4 for more details.

If $V = \Lambda_{\mathbb{Q}}$ (or $V = \Lambda_{\mathbb{R}}$) and $\mathbf{G}$ is an algebraic group over $\mathbb{Q}$ (or $\mathbb{R}$, respectively), then, for any representation $\rho\colon \mathbf{G} \to \mathbf{GL}(V)$ and $v \in V$, we write $\mathrm{Stab}_{\mathbf{G},\rho}(v)$ for the stabiliser of v in G with respect to ρ, that is

$$\mathrm{Stab}_{\mathbf{G},\rho}(v) = \{g \in \mathbf{G} : \rho(g)v = v\}.$$

If W is a subspace of V, we write $\mathrm{Stab}_{\mathbf{G},\rho}(W)$ for the subgroup preserving W. Similarly, we write $\mathrm{End}_{\mathbf{G},\rho}(V)$ and $\mathrm{Aut}_{\mathbf{G},\rho}(V)$ for the endomorphisms and automorphisms, respectively, of V commuting with $\rho(G)$, and we also write $\mathrm{End}_{\mathbf{G},\rho}(\Lambda)$ for the endomorphisms of Λ commuting with $\rho(G)$. If ρ is an inclusion $\mathbf{G} \hookrightarrow \mathbf{GL}(V)$, then we omit it from the subscripts. If D is a ring acting on V, we denote by $\mathrm{End}_D(V)$ the endomorphisms of V commuting with the action of D.

We denote by $\mathbf{G}^{\mathrm{der}}$ the derived subgroup of $\mathbf{G}$, by $Z(\mathbf{G})$ the centre of $\mathbf{G}$ and by $\mathbf{G}(\mathbb{R})^+$ the connected component of $\mathbf{G}(\mathbb{R})$ (in the archimedean topology) containing the identity. If $\mathbf{S}$ is a split $\mathbb{Q}$-subtorus of $\mathbf{G}$, we denote by $Z_{\mathbf{G}}(\mathbf{S})$ the centraliser of $\mathbf{S}$ in $\mathbf{G}$ and by $X^*(\mathbf{S})$ the character group of $\mathbf{S}$.

If $V = \Lambda_{\mathbb{R}}$, we write $|\cdot|$ for a norm on V. Unless otherwise specified, it does not matter which norm we choose, except that the values of constants will depend on the norm. Whenever the statement of a theorem involves a norm $|\cdot|$, we implicitly assume

that such a norm has been chosen, and the constants in the theorem implicitly depend on this choice. We write $\|\cdot\|$ for the associated operator norm on $\mathrm{End}_{\mathbb{R}}(V)$. In other words, for $f \in \mathrm{End}_{\mathbb{R}}(V)$,

$$\|f\| = \sup\{|f(v)| : v \in V, |v| = 1\}.$$

3.2 Cartan involutions

The theory of Cartan involutions is well-known for connected semisimple groups. However, for reductive real algebraic groups, several definitions of Cartan involutions are used in the literature. The following seems to us to be the most elegant definition.

Definition. Let $\mathbf{G}$ be a reductive $\mathbb{R}$-algebraic group. A **Cartan involution** of $\mathbf{G}$ is an involution $\theta\colon \mathbf{G} \to \mathbf{G}$ in the category of $\mathbb{R}$-algebraic groups such that the set of fixed points of θ in $\mathbf{G}(\mathbb{R})$ is a maximal compact subgroup of $\mathbf{G}(\mathbb{R})$.

The fundamental example is the standard Cartan involution $x \mapsto (x^t)^{-1}$ on $\mathbf{GL}_n$, whose real fixed point set is $\mathbf{O}_n(\mathbb{R})$.

The commonly used definitions of Cartan involutions for reductive real algebraic groups are all equivalent to this one, but the equivalences are not obvious and it is difficult to find proofs for all of the equivalences. For convenience, we provide a list of equivalent definitions, and we will post a self-contained proof of this lemma on arXiv. In the following lemma: (ii) is the definition of Cartan involution used in [6, 11.17]; (iii) is the definition used in [4] and [34]; while (iv) is the definition from [10, p. 255], commonly used in the study of Shimura varieties.

Lemma 3.1. Let $\mathbf{G}$ be a reductive $\mathbb{R}$-algebraic group and let $\theta\colon \mathbf{G}(\mathbb{R}) \to \mathbf{G}(\mathbb{R})$ be an involution in the category of real Lie groups. Let $\mathbf{Z}_d$ denote the maximal $\mathbb{R}$-split torus in the centre of $\mathbf{G}$. The following are equivalent:

(i) θ is a Cartan involution as defined above;

(ii) the set of fixed points of θ in $\mathbf{G}(\mathbb{R})$ is a maximal compact subgroup of $\mathbf{G}(\mathbb{R})$ and $\theta(z) = z^{-1}$ for all $z \in \mathbf{Z}_d(\mathbb{R})$;

(iii) there exists a faithful representation $\rho\colon \mathbf{G} \to \mathbf{GL}_{n,\mathbb{R}}$ in the category of $\mathbb{R}$-algebraic groups such that

$$\rho(\theta(g)) = (\rho(g)^t)^{-1}$$

for all $g \in \mathbf{G}(\mathbb{R})$;

(iv) θ is a morphism in the category of $\mathbb{R}$-algebraic groups and the real form $G^\theta = \{g \in \mathbf{G}(\mathbb{C}) : \theta(\bar{g}) = g\}$ is compact and intersects every connected component of $\mathbf{G}(\mathbb{C})$, where $\bar{\cdot}$ denotes complex conjugation.

Furthermore, for each maximal compact subgroup $K \subset \mathbf{G}(\mathbb{R})$, there is a unique Cartan involution of $\mathbf{G}$ whose set of real fixed points is K.

Given a reductive $\mathbb{R}$-algebraic group $\mathbf{G}$ and a Cartan involution θ of $\mathbf{G}$, we say that an algebraic subgroup $\mathbf{H} \subset \mathbf{G}$ is **self-adjoint** (with respect to θ) if $\theta(\mathbf{H}) = \mathbf{H}$.

In several of our theorem statements (including Theorems 1.1 and 1.2), we are given a reductive $\mathbb{Q}$-algebraic group $\mathbf{G}$ and a Siegel set $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$. It will be seen in Section 3.3 that the definition of a Siegel set involves the choice of a maximal compact subgroup $K \subset \mathbf{G}(\mathbb{R})$. In such a situation, we say that a subgroup of $\mathbf{G}$ is self-adjoint if it is self-adjoint with respect to the Cartan involution whose fixed point set is the K used in the construction of the Siegel set.

3.3 Siegel sets

We use the definition of Siegel sets from [29, sec. 2.2], which is a minor modification of definitions used in [6, Def. 12.3] and [1, Ch. II, sec. 4.1]. For a comparison between these definitions, see [29, sec. 2.3].

Let $\mathbf{G}$ be a reductive $\mathbb{Q}$-algebraic group. In order to define a Siegel set in $\mathbf{G}(\mathbb{R})$, we begin by making choices of the following subgroups of $\mathbf{G}$:

(1) $\mathbf{P}$ a minimal parabolic $\mathbb{Q}$-subgroup of $\mathbf{G}$;
(2) K a maximal compact subgroup of $\mathbf{G}(\mathbb{R})$.

As a consequence of [1, Ch. II, Lemma 3.12], there is a unique $\mathbb{R}$-torus $\mathbf{S} \subset \mathbf{P}$ satisfying the following conditions:

(i) $\mathbf{S}$ is $\mathbf{P}(\mathbb{R})$-conjugate to a maximal $\mathbb{Q}$-split torus in $\mathbf{P}$.
(ii) $\mathbf{S}$ is self-adjoint with respect to the Cartan involution associated with K.

These conditions could equivalently be stated as:

(i) $\mathbf{S}$ is a lift of the unique maximal $\mathbb{Q}$-split torus in $\mathbf{P}/\mathbf{U}$, where $\mathbf{U}$ denotes the unipotent radical of $\mathbf{P}$.
(ii) $\mathrm{Lie}\,\mathbf{S}(\mathbb{R})$ is orthogonal to $\mathrm{Lie}\,K$ with respect to the Killing form of $\mathbf{G}$.

Define the following further pieces of notation:

(1) $\mathbf{M}$ is the preimage in $Z_{\mathbf{G}}(\mathbf{S})$ of the maximal $\mathbb{Q}$-anisotropic subgroup of $\mathbf{P}/\mathbf{U}$. (Note that by [7, Corollaire 4.16], $Z_{\mathbf{G}}(\mathbf{S})$ is a Levi subgroup of $\mathbf{P}$ and hence maps isomorphically onto $\mathbf{P}/\mathbf{U}$.)
(2) Δ is the set of simple roots of $\mathbf{G}$ with respect to $\mathbf{S}$, using the ordering induced by $\mathbf{P}$. (The roots of $\mathbf{G}$ with respect to $\mathbf{S}$ form a root system because $\mathbf{S}$ is conjugate to a maximal $\mathbb{Q}$-split torus in $\mathbf{G}$.)
(3) $A_t = \{\alpha \in \mathbf{S}(\mathbb{R})^+ : \chi(\alpha) \geq t \text{ for all } \chi \in \Delta\}$ for any real number $t > 0$.

A **Siegel set** in $\mathbf{G}(\mathbb{R})$ (with respect to $(\mathbf{P}, \mathbf{S}, K)$) is a set of the form

$$\mathfrak{S} = \Omega A_t K$$

where

(1) Ω is a compact subset of $\mathbf{U}(\mathbb{R})\mathbf{M}(\mathbb{R})^+$;
(2) t is a positive real number.

We say that a set $\Omega \subset \mathbf{G}(\mathbb{R})$ is a **fundamental set** for Γ if the following conditions are satisfied:

(F1) $\Gamma\Omega = \mathbf{G}(\mathbb{R})$;
(F2) for every $g \in \mathbf{G}(\mathbb{Q})$, the set $\{\gamma \in \Gamma : \gamma\Omega \cap g\Omega \neq \emptyset\}$ is finite (the **Siegel property**).

The following two theorems show that if we make suitable choices of Siegel set $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ and finite set $C \subset \mathbf{G}(\mathbb{Q})$, then $C\mathfrak{S}$ is a fundamental set for Γ in $\mathbf{G}(\mathbb{R})$.

Theorem 3.2. [6, Théorème 13.1] Let Γ be an arithmetic subgroup of $\mathbf{G}(\mathbb{Q})$. For any minimal parabolic $\mathbb{Q}$-subgroup $\mathbf{P} \subset \mathbf{G}$ and maximal compact subgroup $K \subset \mathbf{G}(\mathbb{R})$, there exist a Siegel set $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ with respect to $(\mathbf{P}, \mathbf{S}, K)$ and a finite set $C \subset \mathbf{G}(\mathbb{Q})$ such that

$$\mathbf{G}(\mathbb{R}) = \Gamma C \mathfrak{S}.$$

Theorem 3.3. [6, Théorème 15.4] Let Γ be an arithmetic subgroup of $\mathbf{G}(\mathbb{Q})$. Let $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ be a Siegel set. For any finite set $C \subset \mathbf{G}(\mathbb{Q})$ and any element $g \in \mathbf{G}(\mathbb{Q})$, the set

$$\{\gamma \in \Gamma : \gamma C\mathfrak{S} \cap gC\mathfrak{S} \neq \emptyset\}$$

is finite.

A quantitative version of Theorem 3.3 can be found at [29, Thm. 1.1].

4 Quantitative reduction theory

In this section, we will prove Theorems 1.1 and 1.2. The proof follows the same strategy as that of [4, Lemma 5.3 and Thm. 6.5]. We replace the purely topological proof of [4, Prop. 5.2] by an orbit growth bound of Eberlein [14] using Riemannian geometry in $\mathbf{GL}_n(\mathbb{R})^+$. We also prove a lemma bounding the norm of $\tau \in \mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$ in terms of the length of $v = \tau(v_0)$—this is a calculation in a semisimple $\mathbb{R}$-algebra. For the rest, the proof closely follows the method of Borel and Harish-Chandra, keeping track of quantitative information and the action of $\mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$ throughout and with some small adaptations to generalise to reductive groups whose $\mathbb{R}$-rank is greater than their $\mathbb{Q}$-rank.

4.1 Bound for orbits of real reductive groups

We begin by proving the following bound for orbits in representations of real reductive groups, not yet considering any arithmetic subgroup.

Proposition 4.1. Let $\mathbf{G}$ be a reductive $\mathbb{R}$-algebraic group and let $\rho\colon \mathbf{G} \to \mathbf{GL}(V_{\mathbb{R}})$ be an $\mathbb{R}$-algebraic representation. Let $v_0 \in V_{\mathbb{R}}$ be a non-zero vector whose orbit $\rho(\mathbf{G}(\mathbb{R}))v_0$ is closed. Then, there exist constants c_9 and c_{10} (depending on $\mathbf{G}$, ρ and v_0) such that, for every $w \in \rho(\mathbf{G}(\mathbb{R}))v_0$, there exists $g \in \mathbf{G}(\mathbb{R})$ satisfying $w = \rho(g)v_0$ and

$$\max(\|\rho(g)\|, \|\rho(g^{-1})\|) \leq c_9|w|^{c_{10}}.$$

Proposition 4.1 provides a quantitative version of [4, Prop. 5.2], which asserts that if $w \in \rho(\mathbf{G}(\mathbb{R}))v_0 \cap Q$ for some compact subset $Q \subset V_{\mathbb{R}}$, then in fact $w \in \rho(\Omega)v_0$ for some compact subset $\Omega \subset \mathbf{G}(\mathbb{R})$ (independent of w). Here, we show that the operator norm of elements of $\rho(\Omega)$ is polynomially bounded with respect to the length of vectors in Q.

We define a Riemannian metric on $\mathbf{GL}_n(\mathbb{R})^+$ as follows. The positive definite bilinear form $(A, B) \mapsto \mathrm{tr}(AB^t)$ on $\mathrm{M}_n(\mathbb{R})$, which is the Lie algebra of $\mathbf{GL}_n(\mathbb{R})$, induces a right-invariant Riemannian metric on the Lie group $\mathbf{GL}_n(\mathbb{R})^+$. Let d_{R} denote the distance function on $\mathbf{GL}_n(\mathbb{R})^+$ induced by this Riemannian metric.

Eberlein's theorem relates $|\rho(g)v_0|$ to the Riemannian distance between g and the stabiliser of v_0. We will combine this with the following lemma bounding $\|\rho(g)\|$ in terms of the Riemannian distance.

Lemma 4.2. Let I denote the identity matrix in $\mathbf{GL}_n(\mathbb{R})$. There exists a constant $c_{11}(n)$ such that every $g \in \mathbf{GL}_n(\mathbb{R})^+$ satisfies

$$\|g\| \leq c_{11}(n)\, \exp(d_{\mathrm{R}}(g, I)).$$

Proof. Let $|g|_{\mathrm{F}}$ denote the Frobenius norm of g, that is

$$|g|_{\mathrm{F}} = \sqrt{\mathrm{tr}(gg^t)}.$$

Using the Cartan decomposition, we can write $g = k \exp(X)$ for some $k \in \mathbf{SO}_n(\mathbb{R})$ and some symmetric matrix $X \in \mathrm{M}_n(\mathbb{R})$. Let $\lambda_{\max}$ denote the largest eigenvalue of X (note that X is diagonalisable and all its eigenvalues are real because it is symmetric).

By [14, Prop. 4.8], we have

$$n^{-1/2} \exp(-c_{12}(n)) \exp(|X|_{\mathrm{F}} - \lambda_{\max}) \leq \frac{\exp(d_{\mathrm{R}}(g, I))}{|g|_{\mathrm{F}}} \tag{2}$$

for some constant $c_{12}(n)$ which depends only on n. Since X is symmetric, we have

$$|X|_{\mathrm{F}} = \sqrt{\mathrm{tr}(XX^t)} = \sqrt{\mathrm{tr}(X^2)} = \sqrt{\sum_{i=1}^{n} \lambda_i^2} \geq \lambda_{\max},$$

where $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of X. Hence, $\exp(|X|_{\mathrm{F}} - \lambda_{\max}) \geq 1$, so (2) implies that

$$|g|_{\mathrm{F}} \leq c_{13}(n)\, \exp(d_{\mathrm{R}}(g, I)),$$

where $c_{13}(n) = n^{-1/2} \exp(-c_{12}(n))$.

Since $|\cdot|_{\mathrm{F}}$ and $\|\cdot\|$ are norms on the finite-dimensional vector space $\mathrm{M}_n(\mathbb{R})$, they are equivalent, so this proves the lemma. ∎

Proof of Proposition 4.1. Let $G = \mathbf{G}(\mathbb{R})^+$. Fix a finite list of representatives $a_1, \ldots, a_r$ for the connected components of $\mathbf{G}(\mathbb{R})$. Then, given $w \in \rho(\mathbf{G}(\mathbb{R}))v_0$, we can write $w = a_i w'$ for some $i \leq r$ and some $w' \in \rho(G)v_0$. Hence, it suffices to prove the proposition for $w \in \rho(G)v_0$.

By [24], we can choose an inner product on $V_{\mathbb{R}}$ with respect to which $\rho(G)$ is self-adjoint. Since all norms on a finite-dimensional vector space are equivalent, it suffices to prove the proposition under the assumption that the norm on $V_{\mathbb{R}}$ is induced by such

an inner product. In particular, the stabiliser of the norm in G is a maximal compact subgroup K and if $\mathfrak{p}$ denotes the -1 eigenspace of the associated Cartan involution θ on $\mathrm{Lie}(G)$, then for every $X \in \mathfrak{p}$, $\mathrm{d}\rho(X)$ is self-adjoint. Thus, the conditions of [34, sec. 3] are satisfied.

Since $\rho(G)v_0$ is closed, it contains a minimal vector, that is, a vector whose length is minimal among all elements of the orbit. (Indeed, a theorem of Richardson and Slodowy [34, Thm. 4.4] states that the two properties are, in fact, equivalent). Replacing v_0 by another vector in its orbit changes the element g such that $w = \rho(g)v_0$ by a fixed element of G, so we may assume that v_0 itself is a minimal vector.

Let $H_0 = \mathrm{Stab}_{\rho(G)}(v_0) \subset \mathbf{GL}_n(\mathbb{R})$. Note that H_0 is self-adjoint with respect to our chosen inner product on $V_{\mathbb{R}}$ (see [34, Thm. 4.3], e.g.).

If $\rho(G)v_0$ is bounded, then it is compact, so by [4, Prop. 5.2], there exists a compact set $\Omega \subset G$ such that $\rho(G)v_0 = \rho(\Omega)v_0$. The elements $g \in \Omega$ satisfy a uniform bound for $\max(\|\rho(g)\|, \|\rho(g^{-1})\|)$, proving the proposition in this case since $|w| \geq |v_0| > 0$.

From now on, assume that the orbit $\rho(G)v_0$ is unbounded. Then, [14, sec. 2.5] defines an associated value $\lambda^-(v_0) \in \mathbb{R}$. By [14, Thm. 3.1 (1)], since v_0 is minimal, $\lambda^-(v_0) > 0$. By [14, Thm. 3.1 (2)], we also have

$$\lambda^-(v_0) \leq \liminf_{d_{\mathrm{R}}(\rho(g),H_0)\to\infty} \frac{\log|\rho(g)v_0|}{d_{\mathrm{R}}(\rho(g),H_0)}.$$

Hence, there exists a constant $c_{14} > 0$ (depending only on G, ρ and v_0) such that

$$\frac{\log|\rho(g)v_0|}{d_{\mathrm{R}}(\rho(g),H_0)} > \tfrac{1}{2}\lambda^-(v_0)$$

for all $g \in G$ satisfying $d_{\mathrm{R}}(\rho(g),H_0) > c_{14}$.

On the other hand, if $d_{\mathrm{R}}(\rho(g),H_0) \leq c_{14}$, then because v is minimal, we have

$$\frac{\log|\rho(g)v_0|}{d_{\mathrm{R}}(\rho(g),H_0)} \geq \frac{\log|v_0|}{d_{\mathrm{R}}(\rho(g),H_0)} \geq \frac{\log|v_0|}{c_{14}},$$

which is a positive constant.

Combining the above two inequalities, we deduce that there is a positive constant c_{15} such that the following inequality holds for all $g \in G$:

$$\frac{\log|\rho(g)v_0|}{d_{\mathrm{R}}(\rho(g),H_0)} \geq c_{15}$$

or in other words,

$$|\rho(g)v_0|^{1/c_{15}} \geq \exp(d_{\mathrm{R}}(\rho(g), H_0)). \tag{3}$$

Given $w \in \rho(G)v_0$, write $w = \rho(g')v_0$, where $g' \in G$. Since H_0 is closed, we can choose $h \in H_0$ such that $d_{\mathrm{R}}(\rho(g'), H_0) = d_{\mathrm{R}}(\rho(g'), h)$.

Since $H_0 \subset \rho(G)$, we can choose $g \in G$ such that $\rho(g) = \rho(g')h^{-1}$. Since $h \in H_0$, we have $\rho(g)v_0 = w$. Since d_{R} is right invariant, we have

$$d_{\mathrm{R}}(\rho(g'), h) = d_{\mathrm{R}}(\rho(g), I) = d_{\mathrm{R}}(I, \rho(g^{-1})).$$

Thus, (3) (applied to g') becomes

$$|w|^{1/c_{15}} \geq \exp(d_{\mathrm{R}}(\rho(g), I)) = \exp(d_{\mathrm{R}}(\rho(g^{-1}), I)).$$

Applying Lemma 4.2 to both $\rho(g)$ and $\rho(g^{-1})$ completes the proof of the proposition. ■

Remark 4.3. The proof of Proposition 4.1 is ineffective for two reasons.

(1) It depends on the value of $\lambda^-(v_0)$. We do not know a general method for calculating this value, although it seems to be feasible to calculate it in particular cases.

(2) The value c_{14} depends on the speed of convergence of the limit in [14, Thm. 3.1], which is ineffective.

4.2 Quantitative reduction theory for representations

We now prove Theorem 1.1. The proof follows that of [4, Lemma 5.3], keeping track of quantitative information and some minor generalisations. For the sake of clarity, we have broken it down into a series of lemmas, each proved by a short calculation.

We use the notation for Siegel sets from Section 3.3. Then, $\mathbf{S}$ is $\mathbf{P}(\mathbb{R})$-conjugate to a maximal $\mathbb{Q}$-split torus $\mathbf{T}$ in $\mathbf{P}$. Since $\mathbf{P}(\mathbb{R}) = Z_{\mathbf{G}}(\mathbf{S})(\mathbb{R}).\mathbf{U}(\mathbb{R})$, we can choose n in $\mathbf{U}(\mathbb{R})$ such that $\mathbf{S} = n\mathbf{T}n^{-1}$.

We also adopt some notation from the proof of [4, Lemma 5.3] (bearing in mind that we have reversed the order of multiplication in our Iwasawa decomposition relative to [4]). By the Iwasawa and Langlands decompositions, the multiplication map

$$\mathbf{U}(\mathbb{R}) \times \mathbf{S}(\mathbb{R})^+ \times \mathbf{M}(\mathbb{R})^+ K \to \mathbf{G}(\mathbb{R})$$

is bijective. Given $x \in \mathbf{G}(\mathbb{R})$, we write it as $x = n_x a_x k_x$ according to this decomposition. Let

$$y_x = a_x^{-1} n x, \quad z_x = a_x^{-2} n x.$$

For each character $\chi \in X^*(\mathbf{S})$, let V_χ denote the corresponding eigenspace in $V_\mathbb{R}$. We have $V_\mathbb{R} = \bigoplus_\chi V_\chi$ and we let $\pi_\chi : V \to V_\chi$ denote the projection maps. Since all norms on the finite-dimensional vector space $V_\mathbb{R}$ are equivalent, we may assume without loss of generality that the norm is chosen so that the spaces V_χ are orthogonal to each other.

In the lemmas which follow, τ denotes an element of $\mathrm{Aut}_{\rho(\mathbf{G})}(V_\mathbb{R})$. Constants labelled c_n depend only on $\mathbf{G}$, $\mathfrak{S}$, ρ, Λ and v_0, and not on τ, x, v or w.

Lemma 4.4. There exists a constant $c_{16} > 0$ such that for all $v' \in V_\mathbb{R}$ and all $\chi \in X^*(\mathbf{S})$, if $\tau v' \in \Lambda$, then either $\pi_\chi(\rho(n)v') = 0$ or $|\pi_\chi(\rho(n)v')| \geq c_{16}/\|\tau\|$.

Proof. Since $\mathbf{T}$ is $\mathbb{Q}$-split, its eigenspaces V_ψ are defined over $\mathbb{Q}$ and V decomposes as $\oplus_{\psi \in X^*(\mathbf{T})} V_\psi$. For $\psi \in X^*(\mathbf{T})$, let π_ψ denote the projection $V \to V_\psi$ in this direct sum. Because the V_ψ are defined over $\mathbb{Q}$, the image $\pi_\psi(\Lambda)$ is a lattice in V_ψ. Hence, there is a constant $c_{17} > 0$ such that, if $\tau v' \in \Lambda$, then

$$\pi_\psi(\tau v') = 0 \text{ or } |\pi_\psi(\tau v')| \geq c_{17}.$$

Since $V_\psi \neq 0$ for only finitely many characters $\psi \in X^*(\mathbf{T})$, it is possible to choose a single constant $c_{17} > 0$ which works for every ψ.

For each $\chi \in X^*(\mathbf{S})$, the eigenspace V_χ of $\mathbf{S}$ is equal to $\rho(n)V_\psi$ for some $\psi \in X^*(\mathbf{T})$. It follows that $\pi_\chi = \rho(n) \circ \pi_\psi \circ \rho(n)^{-1}$. Therefore, there is a constant $c_{16} > 0$ (namely, $c_{17}\|\rho(n)^{-1}\|^{-1}$) such that, if $\tau v' \in \Lambda$, then

$$\pi_\chi(\tau\rho(n)v') = 0 \text{ or } |\pi_\chi(\tau\rho(n)v')| \geq c_{16}.$$

Since τ commutes with $\rho(\mathbf{G}(\mathbb{R}))$, it preserves the eigenspaces V_χ and hence commutes with π_χ. Therefore, either

$$\tau(\pi_\chi(\rho(n)v')) = \pi_\chi(\tau\rho(n)v') = 0,$$

which implies $\pi_\chi(\rho(n)v')) = 0$, or

$$\|\tau\||\pi_\chi(\rho(n)v')| \geq |\tau(\pi_\chi(\rho(n)v'))| = |\pi_\chi(\tau\rho(n)v')| \geq c_{16}.$$

■

Lemma 4.5. There exists a constant c_{18} such that, for all $x \in \mathfrak{S}$, we have

$$|\rho(y_x)v_0| \leq c_{18}.$$

Proof. From the definition of a Siegel set, $\{n_x : x \in \mathfrak{S}\}$ is relatively compact. Hence, $\{nn_x : x \in \mathfrak{S}\}$ is a relatively compact subset of $\mathbf{U}(\mathbb{R})$. Therefore, by [6, Lemme 12.2], $\{a_x^{-1}nn_xa_x : x \in \mathfrak{S}\}$ is relatively compact. Furthermore, $\{k_x : x \in \mathfrak{S}\}$ is also relatively compact. Since

$$y_x = a_x^{-1}nx = a_x^{-1}nn_xa_xk_x,$$

we conclude that $\{y_x : x \in \mathfrak{S}\}$ is relatively compact. ■

Lemma 4.6. There exists a constant c_{19} such that, for all $x \in \mathfrak{S}$, if $\tau\rho(x)v_0 \in \Lambda$, then

$$|\rho(z_x)v_0| \leq c_{19}\|\tau\|.$$

Proof. Let $\chi \in X^*(\mathbf{S})$. From the definitions of y_x and z_x, we can calculate

$$\pi_\chi(\rho(y_x)v_0) = \chi(a_x)^{-1}\pi_\chi(\rho(nx)v_0), \quad \pi_\chi(\rho(z_x)v_0) = \chi(a_x)^{-2}\pi_\chi(\rho(nx)v_0).$$

Therefore, either $\pi_\chi(\rho(nx)v_0) = 0$, in which case $\pi_\chi(\rho(z_x)v_0) = 0$, or else, by Lemma 4.4 (applied to $v' = \rho(x)v_0$) and Lemma 4.5, we have

$$|\pi_\chi(\rho(z_x)v_0)| = \frac{|\pi_\chi(\rho(y_x)v_0)|^2}{|\pi_\chi(\rho(nx)v_0)|} \leq \frac{c_{18}^2}{c_{16}/\|\tau\|} = c_{20}\|\tau\|.$$

Since $V_\mathbb{R}$ is the orthogonal direct sum of the V_χ, the lemma follows by squaring and summing over χ. ■

Lemma 4.7. There exist constants c_{21} and c_{22} such that, for every $x \in \mathfrak{S}$, if $\tau(\rho(x)v_0) \in \Lambda$, then there exists $g \in \mathbf{G}(\mathbb{R})$ satisfying

$$\rho(g)v_0 = \rho(a_x^{-1}k_x)v_0 \text{ and } \max(\|\rho(g)\|, \|\rho(g)^{-1}\|) \leq c_{21}\|\tau\|^{c_{22}}.$$

Proof. By Proposition 4.1 and Lemma 4.6, there exists $g' \in \mathbf{G}(\mathbb{R})$ such that $\rho(z_x)v_0 = \rho(g')v_0$ and

$$\max(\|\rho(g')\|, \|\rho(g'^{-1})\|) \leq c_9|\rho(z_x)v_0|^{c_{10}} \leq c_9 c_{19}^{c_{10}}\|\tau\|^{c_{10}}. \tag{4}$$

Let

$$g = a_x^{-2}n_x^{-1}n^{-1}a_x^2 g'.$$

Then,

$$\begin{aligned}\rho(g)v_0 &= \rho(a_x^{-2}n_x^{-1}n^{-1}a_x^2)\rho(g')v_0 = \rho(a_x^{-2}n_x^{-1}n^{-1}a_x^2)\rho(z_x)v_0\\ &= \rho(a_x^{-2}n_x^{-1}x)v_0 = \rho(a_x^{-1}k_x)v_0.\end{aligned}$$

Meanwhile, by [6, Lemma 12.2], $\{a_x^{-2}n_x^{-1}na_x^2 : x \in \mathfrak{S}\}$ is relatively compact so (4) implies the required bound on $\max(\|\rho(g)\|, \|\rho(g)^{-1}\|)$. ∎

Let θ denote the Cartan involution of $\mathbf{G}$ whose set of real fixed points is K.

Lemma 4.8. There exists a compact set $\Phi \subset \mathbf{G}(\mathbb{R})$ such that, for all $x \in \mathfrak{S}$, we have $\theta(a_x^{-1}k_x) \in \Phi x$.

Proof. By definition $k_x \in \mathbf{M}(\mathbb{R})^+K$, so we can write $k_x = m_x\ell_x$ where $m_x \in \mathbf{M}(\mathbb{R})^+$ and $\ell_x \in K$. This is not a unique decomposition, but the definition of Siegel set guarantees that we can choose m_x in a fixed compact subset of $\mathbf{M}(\mathbb{R})^+$. (Recall that, by definition, $\mathbf{M}$ commutes with $\mathbf{S}$.)

By definition, θ acts trivially on K and stabilises $\mathbf{S}(\mathbb{R})$. Since $\mathbf{S}$ is an $\mathbb{R}$-split torus, the latter implies that $\theta(a) = a^{-1}$ for all $a \in \mathbf{S}(\mathbb{R})$. Hence,

$$\theta(a_x^{-1}k_x) = \theta(m_x a_x^{-1}\ell_x) = \theta(m_x)a_x\ell_x = \theta(m_x)m_x^{-1}n_x^{-1}x.$$

Since m_x and n_x lie in compact sets independent of x, this proves the lemma. ∎

We are now ready to prove a version of Theorem 1.1 in which the bound is expressed in terms of the operator norm of $\tau \in \mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$, instead of the length of $v = \tau(v_0)$.

Proposition 4.9. Let $\mathbf{G}$ be a reductive $\mathbb{Q}$-algebraic group and let $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})$ be a Siegel set. Let $\rho\colon \mathbf{G} \to \mathbf{GL}(V)$ be a representation of $\mathbf{G}$ defined over $\mathbb{Q}$. Let $\Lambda \subset V$ be a $\mathbb{Z}$-lattice. Let $v_0 \in V_{\mathbb{R}}$ be such that:

(i) $\rho(\mathbf{G}(\mathbb{R}))v_0$ is closed in $V_{\mathbb{R}}$;

(ii) the stabiliser $\mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(v_0)$ is self-adjoint.

Then, there exist constants c_{23}, c_{24} such that, for every $\tau \in \mathrm{Aut}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$ and every $w \in \rho(\mathfrak{S})\tau(v_0) \cap \Lambda$, we have $|w| \leq c_{23}\|\tau\|^{c_{24}}$.

Proof. Write $w = \rho(x)\tau(v_0) = \tau(\rho(x)v_0)$, with $x \in \mathfrak{S}$. Then, we get g as in Lemma 4.7.

By [4, Prop. 13.5], there is a Cartan involution θ' of $\mathbf{GL}(V_{\mathbb{R}})$ such that $\theta' \circ \rho = \rho \circ \theta$. With respect to a suitable basis of $V_{\mathbb{R}}$, θ' is given by $g \mapsto (g^{-1})^t$. The norms $\|X\|$ and $\|X^t\|$ on the finite dimensional vector space $\mathrm{End}(V_{\mathbb{R}})$ are equivalent, so there exists a constant c_{25} such that $\|\rho(\theta(g))\| \leq c_{25}\|\rho(g^{-1})\|$.

We have $g = a_x^{-1}k_x h$, where $h \in H_0 = \mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(v_0)$. Hence by Lemma 4.8, we get

$$\theta(g) = \theta(a_x^{-1}k_x)\theta(h) \in \Phi x \theta(h),$$

where Φ is a fixed compact set. Hence, we get

$$\|\rho(x\theta(h))\| \leq c_{26}\|\rho(\theta(g))\| \leq c_{27}\|\rho(g^{-1})\| \leq c_{28}\|\tau\|^{c_{29}}.$$

By hypothesis, H_0 is self-adjoint so $\theta(h) \in H_0$. Hence, $\rho(x\theta(h))v_0 = \rho(x)v_0$ so $w = \tau(\rho(x\theta(h)v_0)$ and

$$|w| \leq \|\tau\|\|\rho(x\theta(h))\||v_0|,$$

which is polynomially bounded with respect to $\|\tau\|$, as required. ∎

To conclude, we show that it is possible to choose τ such that $\tau(v_0) = v$ and $\|\tau\|$ is bounded in terms of $|v|$. Theorem 1.1 follows by combining Proposition 4.9 with Lemma 4.7, applied to $E = \mathrm{End}_{\rho(\mathbf{G})}(V_{\mathbb{R}})$.

Lemma 4.10. Let $V_{\mathbb{R}}$ be a real vector space and let E be a semisimple $\mathbb{R}$-subalgebra of $\mathrm{End}(V_{\mathbb{R}})$. Let $v_0 \in V_{\mathbb{R}}$.

Then, there exists a constant c_{30} such that, for every $v \in E^{\times} v_0$, there exists $e \in E^{\times}$ satisfying $v = ev_0$ and $\|e\| \leq c_{30}|v|$.

Note that $E^{\times}$ is the group of $\mathbb{R}$-points of a reductive $\mathbb{R}$-algebraic group. However, this lemma does not follow from Proposition 4.1 because the orbit $E^{\times} v_0$ is not closed.

Proof of Lemma 4.10. Write E as a product of simple $\mathbb{R}$-algebras $\prod_{i=1}^{m} E_i$. There is a corresponding decomposition $V_{\mathbb{R}} = \bigoplus_{i=1}^{m} V_i$, where the action of E_i on $V_{\mathbb{R}}$ factors through V_i. If $v_0 = \sum_{i=1}^{m} v_i \in V_{\mathbb{R}}$ and $e = (e_1, \ldots, e_m) \in E$, then $ev_0 = \sum_{i=1}^{m} e_i v_i$.

Because all norms on a finite-dimensional real vector space are equivalent, we may assume without loss of generality that the norm of each element of $V_{\mathbb{R}}$ is the maximum of the norms of its projections to the V_i. Then, the operator norm satisfies $\|e\| = \max\{\|e_i\| : 1 \leq i \leq m\}$. Hence, it suffices to prove the lemma for each pair (E_i, V_i). In other words, we may assume that E is a simple $\mathbb{R}$-algebra.

Then, $E = \mathrm{M}_n(D)$ for some positive integer n, where D is a division algebra isomorphic to $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. There is a unique simple E-module, namely D^n, such that we can identify $V_{\mathbb{R}}$ (as a left E-module) with $(D^n)^r$ for some positive integer r.

Again, since all norms on a finite-dimensional real vector space are equivalent, we may assume that the norm on $V_{\mathbb{R}} \cong D^{nr}$ is induced by a norm on D by letting the norm of an element of D^{nr} be the maximum of the norms of its coordinates.

Via this identification, write

$$v_0 = (x_1, \ldots, x_r), \quad v = (y_1, \ldots, y_r),$$

where $x_1, \ldots, x_r, y_1, \ldots, y_r \in D^n$.

Reordering $x_1, \ldots, x_r$ (and the corresponding $y_1, \ldots, y_r$), we may assume that $\{x_1, \ldots, x_s\}$ forms a maximal right D-linearly independent subset of $\{x_1, \ldots, x_r\}$ for a suitable positive integer $s \leq r$. Note that $s \leq n$. Then, there exist $a_{ij} \in D$ $(1 \leq i \leq s < j \leq r)$ such that

$$x_j = \sum_{i=1}^{s} x_i a_{ij} \text{ for } s+1 \leq j \leq r. \tag{5}$$

By hypothesis, there exists $e' \in E^\times$ such that $v = e'v_0$ or in other words $y_i = e'x_i$ for all i. Consequently,

$$y_j = e'x_j = \sum_{i=1}^{s} e'x_i a_{ij} = \sum_{i=1}^{s} y_i a_{ij} \text{ for } s+1 \le j \le r. \tag{6}$$

Since the set $\{x_1, \dots, x_s\}$ is right D-linearly independent, it can be extended to form a right D-basis of D^n. Let $h \in \mathrm{M}_n(D)$ denote the matrix formed using such a basis as its columns. Then, h is invertible and $hb_i = x_i$ for $1 \le i \le s$, where $\{b_1, \dots, b_n\}$ denotes the standard D-basis of D^n. Note that the choices made in constructing h can be made depending only on v_0, not on v.

Note that, since $e' \in E^\times = \mathrm{GL}_n(D)$, the set $\{y_1, \dots, y_s\}$ is also right D-linearly independent. Hence, it can be extended to a right D-basis of D^n and we can assume, by scaling if necessary, that the norms of the additional vectors are at most $|y_1| + \cdots + |y_s|$.

Let $f \in \mathrm{M}_n(D)$ be the invertible matrix that has this basis as its columns, the first s equal to the $y_1, \dots, y_s$. Then,

$$fh^{-1}x_i = fb_i = y_i \text{ for } 1 \le i \le s.$$

Using (5) and (6), we deduce that also

$$fh^{-1}x_j = y_j \text{ for } s+1 \le j \le r.$$

In other words, $e = fh^{-1} \in E^\times$ satisfies $ev_0 = v$.

By construction,

$$\|f\| \le |y_1| + \cdots + |y_s| + (n-s)(|y_1| + \cdots + |y_s|) \le n|v|.$$

Since h is independent of v, the proof is complete. ■

4.3 Quantitative fundamental sets for arithmetic groups

The proof of Theorem 1.2 follows the proof of [4, Thm. 6.5]. All we have to do is use the quantitative information from Theorem 1.1 in place of the finiteness statement [4, Lemma 5.4]. There are also some minor additional technical steps due to the need to keep track of the finite set $C \subset \mathbf{G}(\mathbb{Q})$ such that $C\mathfrak{S}$ is a fundamental set in the ambient group $\mathbf{G}(\mathbb{R})$—this was not needed in [4, Thm. 6.5] because there $\mathbf{G} = \mathbf{GL}_n$ and so $C = \{1\}$.

Proof of Theorem 1.2. Suppose we are given $u \in \mathbf{G}(\mathbb{R})$ and $v_u \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})v_0$ such that $\mathbf{H}_u = u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$ and $\rho(u)v_u \in \Lambda$. Let $v_u = \tau(v_0)$ where $\tau \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})$, and let $v = \rho(u)v_u$.

Thanks to [4, Cor. 6.3], we may enlarge the lattice $\Lambda \subset \Lambda_{\mathbb{Q}}$ so that it is $\rho(\Gamma)$-stable. For each $c \in C$, $c^{-1}\Lambda$ is a lattice in $\Lambda_{\mathbb{Q}}$. Hence, we can choose a lattice $\Lambda' \subset \Lambda_{\mathbb{Q}}$ such that $c^{-1}\Lambda \subset \Lambda'$ for all $c \in C$.

By Theorem 1.1, every $w \in \rho(\mathfrak{S})v_u \cap \Lambda'$ has length polynomially bounded with respect to $|v_u|$. In particular, for each $c \in C$, the set

$$\rho(\mathfrak{S})v_u \cap \rho(c^{-1}\Gamma)v \subset \rho(\mathfrak{S})v_u \cap \Lambda'$$

is finite, so we can choose a finite set $\{b_{c,1}, \dots, b_{c,m_c}\} \subset \Gamma$ such that

$$\rho(\mathfrak{S})v_u \cap \rho(c^{-1}\Gamma)v = \{\rho(c^{-1}b_{c,1}^{-1})v, \dots, \rho(c^{-1}b_{c,m_c}^{-1})v\}.$$

Let $B_u = \bigcup_{c\in C}\{b_{c,1}, \dots, b_{c,m_c}\}$, which is a finite subset of Γ.

By Theorem 1.1, we have

$$|\rho(c^{-1}b_{c,i}^{-1})v| \leq c_{31}|v_u|^{c_{32}}$$

for all $c \in C$ and $i \leq m_c$. Since c comes from a fixed finite set, we deduce that

$$|\rho(b_{c,i}^{-1})v| \leq c_{33}|v_u|^{c_{34}}.$$

This is the length bound on $\rho(b^{-1}u)v_u$ for $b \in B_u$ which is required by the statement of the theorem.

Let $\Gamma_u = \Gamma \cap \mathbf{H}_u(\mathbb{R})$ and $\mathcal{F}_{\mathbf{H}_u} = B_u C\mathfrak{S}u^{-1} \cap \mathbf{H}_u(\mathbb{R})$. It remains to show that $\mathcal{F}_{\mathbf{H}_u}$ is a fundamental set for Γ_u in $\mathbf{H}_u(\mathbb{R})$.

Let $h \in \mathbf{H}_u(\mathbb{R}) \subset \mathbf{G}(\mathbb{R})$. By hypothesis, $C\mathfrak{S}$ is a fundamental set for Γ in $\mathbf{G}(\mathbb{R})$, so we can write

$$hu = \gamma cs,$$

where $\gamma \in \Gamma$, $c \in C$ and $s \in \mathfrak{S}$. Since $h \in \mathbf{H}_u(\mathbb{R}) = \mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(\rho(u)v_0)$, we obtain

$$\rho(\gamma cs)v_0 = \rho(hu)v_0 = \rho(u)v_0.$$

Applying τ, we get

$$\rho(\gamma cs)v_u = v$$

or in other words

$$\rho(s)v_u = \rho(c^{-1}\gamma^{-1})v \in \rho(\mathfrak{S})v_u \cap \rho(c^{-1}\Gamma)v.$$

Hence, there exists $b_{c,i} \in B_u$ such that

$$\rho(c^{-1}b_{c,i}^{-1})v = \rho(s)v_u = \rho(c^{-1}\gamma^{-1})v.$$

In particular, $\gamma b_{c,i}^{-1} \in \mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(v)$, and we also have $\gamma b_{c,i}^{-1} \in \Gamma$. Since $\tau \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})$, we have $\mathrm{Stab}_{\mathbf{G}(\mathbb{R}),\rho}(v) = \mathbf{H}_u(\mathbb{R})$. Thus, $\gamma b_{c,i}^{-1} \in \Gamma_u$ and

$$h = \gamma b_{c,i}^{-1}.b_{c,i}csu^{-1} \in \Gamma_u \mathcal{F}_{\mathbf{H}_u}.$$

Thus, the Γ_u-translates of $\mathcal{F}_{\mathbf{H}_u}$ cover $\mathbf{H}_u(\mathbb{R})$. The fact that there are only finitely many $\gamma \in \Gamma_u$ for which $\gamma\mathcal{F}_{\mathbf{H}_u} \cap \mathcal{F}_{\mathbf{H}_u} \neq \emptyset$ follows from the Siegel property for $\mathfrak{S}$ (and indeed this implies that $\mathcal{F}_{\mathbf{H}_u}$ also satisfies the Siegel property). Thus, $\mathcal{F}_{\mathbf{H}_u}$ is a fundamental set for Γ_u in $\mathbf{H}_u(\mathbb{R})$. ■

5 Quantitative reduction theory for quaternion algebras

In order to apply Theorem 1.2, it is necessary to choose a representation ρ and a vector v_0 having the properties described in the theorem. In this section, we will explain how to construct a suitable representation for our application to unlikely intersections with E^2 and quaternionic curves. This illustrates a method for constructing representations which will be useful for applying Theorem 1.2 to other problems of unlikely intersections in the future while avoiding many technical complications which occur in more general situations.

Borel and Harish-Chandra's reduction theory considered only a fixed reductive subgroup $\mathbf{H}_0 \subset \mathbf{G}$ (and not its conjugates $u\mathbf{H}_0u^{-1}$). As such, [4, Thm. 3.8] constructs a representation satisfying the properties (i) and (ii) of Theorem 1.2 but does not construct the vectors v_u. Another construction of representations satisfying (i) is given by [11, Prop. 3.1] (based on [9, Exp. 10, Prop. 5]), and Deligne's construction can easily be modified to yield the vectors v_u.

However, it is not enough to know just that the vectors v_u exist. Theorem 1.2 gives bounds in terms of $|v_u|$ so, in order to apply these, we need to control the length $|v_u|$ in terms of some more intrinsic quantity attached to the subgroup $u\mathbf{H}_0u^{-1}$. For example, in our application, the subgroups $u\mathbf{H}_0u^{-1}$ will be associated with quaternion algebras and we will bound $|v_u|$ in terms of the discriminants of (orders in) these algebras.

5.1 The set-up: quaternionic subgroups of $\mathbf{GSp}_4$

Let $\mathbf{G} = \mathbf{GSp}_4$, the algebraic group whose $\mathbb{Q}$-points are the invertible linear transformations of $\mathbb{Q}^4$ which multiply the standard symplectic form by a scalar. For the standard symplectic form, we use $\psi : \mathbb{Q}^4 \times \mathbb{Q}^4 \to \mathbb{Q}$ represented by the matrix

$$\begin{pmatrix} J & 0 \\ 0 & J \end{pmatrix}, \text{ where } J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The subgroup $\mathbf{H}_0$ is equal to $\mathbf{GL}_2$, embedded block diagonally in $\mathbf{GSp}_4$:

$$\mathbf{H}_0 = \left\{ \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \in \mathbf{GSp}_4 : A \in \mathbf{GL}_2 \right\}. \tag{7}$$

If a $\mathbf{G}(\mathbb{R})$-conjugate $u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$, then its $\mathbb{Q}$-points form the multiplicative group of a (perhaps split) indefinite quaternion algebra over $\mathbb{Q}$. We shall prove the following proposition.

Proposition 5.1. Let $\mathbf{G} = \mathbf{GSp}_{4,\mathbb{Q}}$ and let $\mathbf{H}_0 = \mathbf{GL}_{2,\mathbb{Q}}$, embedded in $\mathbf{G}$ as in (7). Let $\Gamma = \mathbf{Sp}_4(\mathbb{Z})$. Let $L = \mathbb{Z}^4$ and let $\mathbf{G}$ act on $L_{\mathbb{Q}}$ in the natural way.

There exist a $\mathbb{Q}$-algebraic representation $\rho\colon \mathbf{G} \to \mathbf{GL}(\Lambda_{\mathbb{Q}})$, where Λ is a finitely generated free $\mathbb{Z}$-module stabilised by Γ, a vector $v_0 \in \Lambda$ and constants c_{35}, c_{36}, c_{37}, c_{38} such that:

(i) $\mathrm{Stab}_{\mathbf{G},\rho}(v_0) = \mathbf{H}_0$;

(ii) the orbit $\rho(\mathbf{G}(\mathbb{R}))v_0$ is closed in $\Lambda_{\mathbb{R}}$;

(iii) for each $u \in \mathbf{G}(\mathbb{R})$, if the group $\mathbf{H}_u = u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$, then

(a) there exists $v_u \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})v_0$ such that $\rho(u)v_u \in \Lambda$ and

$$|v_u| \leq c_{35}|\mathrm{disc}(R_u)|^{c_{36}};$$

(b) there exists $\gamma \in \Gamma$ and $h \in \mathbf{H}_0(\mathbb{R})$ such that

$$\|\gamma u h\| \leq c_{37} |\mathrm{disc}(R_u)|^{c_{38}},$$

where R_u denotes the order $\mathrm{End}_{\mathbf{H}_u}(L)$ of the quaternion algebra $\mathrm{End}_{\mathbf{H}_u}(L_{\mathbb{Q}})$.

The condition $v_u \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}}) v_0$ in Proposition 5.1(iii)(a) ensures that the element $\rho(u)v_u \in \Lambda$ satisfies $\mathrm{Stab}_{\mathbf{G},\rho}(\rho(u)v_u) = \mathbf{H}_u$. Proposition 5.1(iii)(a) is the bound we need to apply Theorem 1.2. Proposition 5.1(iii)(b) is not required for our application to unlikely intersections but may be useful in its own right—we can replace u by $\gamma u h$ if we replace $\mathbf{H}_u$ by $\gamma \mathbf{H}_u \gamma^{-1}$, a subgroup of $\mathbf{G}$ which gives rise to the same special subvariety of $\mathcal{A}_2$ as $\mathbf{H}_u$.

The proof of Proposition 5.1 will proceed in three steps: first, we construct ρ and v_0 satisfying property (i), then we show that the representation we have constructed possesses property (ii) and then (iii). The proofs of properties (ii) and (iii) are independent of each other, while (iii)(b) is a by-product of the proof of (iii)(a).

5.2 Construction of representation of $\mathbf{GSp}_4$

We construct the representation ρ of $\mathbf{GSp}_4$ and the vector v_0, and define notation which we shall use throughout the rest of the section.

Let $W = \mathrm{M}_4(\mathbb{Q})$, considered as a $\mathbb{Q}$-vector space. Define two representations $\sigma_L, \sigma_R \colon \mathbf{G} = \mathbf{GSp}_4 \to \mathbf{GL}(W)$ by multiplication on the left and on the right:

$$\sigma_L(g)w = gw, \quad \sigma_R(g)w = wg^{-1}.$$

(The inverse in the formula for σ_R is so that σ_R is a left representation of $\mathbf{G}$.)

Let $E_0 = \mathrm{M}_2(\mathbb{Q})$ and define $\iota_0 \colon E_0 \to \mathrm{M}_4(\mathbb{Q})$ by

$$\iota_0(A) = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}.$$

Thus, $\mathbf{H}_0 = \iota_0(\mathbf{GL}_2)$. Let $Z = \iota_0(E_0)$, a four-dimensional $\mathbb{Q}$-linear subspace of W. Observe that

$$\mathrm{Stab}_{\mathbf{G},\sigma_L}(Z) = \iota_0(E_0) \cap \mathbf{G} = \mathbf{H}_0.$$

Similarly, $\mathrm{Stab}_{\mathbf{G},\sigma_R}(Z) = \mathbf{H}_0$ but we shall not need this latter fact.

Let $V = \bigwedge^4 W$ and let $\rho_L, \rho_R \colon \mathbf{G} \to \mathbf{GL}(V)$ be the representations

$$\rho_L = \bigwedge^4 \sigma_L \otimes \det^{-1}, \quad \rho_R = \bigwedge^4 \sigma_R \otimes \det.$$

Then, $\bigwedge^4 Z$ is a one-dimensional $\mathbb{Q}$-linear subspace of V, with

$$\mathrm{Stab}_{\mathbf{G},\rho_L}(\bigwedge^4 Z) = \mathrm{Stab}_{\mathbf{G},\sigma_L}(Z) = \mathbf{H}_0.$$

The action of $\mathbf{GL}_2(\mathbb{Q})$ on Z via $\sigma_L \circ \iota_0$ is the restriction of the left regular representation of $\mathrm{M}_2(\mathbb{Q})$. Hence, the action of $\mathbf{GL}_2(\mathbb{Q})$ on $\bigwedge^4 Z$ via $\bigwedge^4 \sigma_L \circ \iota_0$ is multiplication by $(\det_{\mathbf{GL}_2})^2 = \det_{\mathbf{GL}_4} \circ \iota_0$. Therefore, the action of $\mathbf{H}_0$ on $\bigwedge^4 Z$ via ρ_L is trivial, so each non-zero vector in $\bigwedge^4 Z$ has stabiliser equal to $\mathbf{H}_0$.

Let $\Lambda = \bigwedge^4 \mathrm{M}_4(\mathbb{Z}) \subset V$. For later use, we choose a specific element $v_0 \in (\bigwedge^4 Z) \cap \Lambda$. Let e_1, e_2, e_3 and e_4 denote the following $\mathbb{Z}$-basis for $\mathrm{M}_2(\mathbb{Z})$:

$$e_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \; e_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \; e_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8}$$

Then, $\iota_0(e_1), \iota_0(e_2), \iota_0(e_3)$ and $\iota_0(e_4)$ form a $\mathbb{Z}$-basis for $Z \cap \mathrm{M}_4(\mathbb{Z})$, so

$$v_0 = \iota_0(e_1) \wedge \iota_0(e_2) \wedge \iota_0(e_3) \wedge \iota_0(e_4)$$

is a generator of the rank-1 $\mathbb{Z}$-module $(\bigwedge^4 Z) \cap \Lambda$. Then, ρ_L and v_0 satisfy Proposition 5.1(i).

Given $u \in \mathbf{G}(\mathbb{R})$, we can easily find a vector $v_u \in \mathrm{Aut}_{\rho_L(\mathbf{G})}(\Lambda_{\mathbb{R}}) v_0$ such that $\rho_L(u) v_u \in \Lambda$ (i.e. the first part of Proposition 5.1(iii)(a)). The vector

$$\rho_L(u)\rho_R(u)v_0 = u\iota_0(e_1)u^{-1} \wedge u\iota_0(e_2)u^{-1} \wedge u\iota_0(e_3)u^{-1} \wedge u\iota_0(e_4)u^{-1} \in V_{\mathbb{R}}$$

generates the line $\bigwedge^4 u Z_{\mathbb{R}} u^{-1} \subset V_{\mathbb{R}}$. If the subgroup $u\mathbf{H}_{0,\mathbb{R}}u^{-1} \subset \mathbf{G}_{\mathbb{R}}$ is defined over $\mathbb{Q}$, then so is the linear subspace $uZ_{\mathbb{R}}u^{-1} \subset W_{\mathbb{R}}$. (This follows from the fact that Z is the $\mathbb{Q}$-linear span of $\mathbf{H}_0(\mathbb{Q})$.) Consequently $(\bigwedge^4 uZ_{\mathbb{R}}u^{-1}) \cap \Lambda$ is non-empty, so there exists $d_u \in \mathbb{R}^\times$ such that

$$d_u \rho_L(u)\rho_R(u)v_0 \in \Lambda.$$

Now $d_u \rho_R(u) \in \mathrm{Aut}_{\rho_L}(V_{\mathbb{R}})$, so $d_u \rho_R(u) v_0$ has the required property. This algebraic construction does not control the size of d_u and hence does not control $|v_u|$.

Later, in Section 5.5, we will choose a slightly different v_u (making use of γ as in Proposition 5.1(iii)(b)) allowing us to bound $|v_u|$.

To place this representation in a more general context, we compare it with [9, Exp. 10, Prop. 5]. Let $\mathbf{G}$ be an arbitrary affine $\mathbb{Q}$-algebraic group and $\mathbf{H}_0 \subset \mathbf{G}$ an algebraic subgroup. Chevalley considers the ring of regular functions $\mathbb{Q}[\mathbf{G}]$, on which $\mathbf{G}$ acts by right translations. The stabiliser of the ideal $I(\mathbf{H}_0)$ is equal to $\mathbf{H}_0$. Choose a finite-dimensional subrepresentation $W \subset \mathbb{Q}[\mathbf{G}]$, which contains a generating set for $I(\mathbf{H}_0)$. Then, $\mathbf{H}_0$ is also the stabiliser of $Z = W \cap I(\mathbf{H}_0)$. Let $d = \dim_{\mathbb{Q}}(Z)$. Then, $\bigwedge^d W$ is a representation of $\mathbf{G}$ in which the line $\bigwedge^d Z$ is defined over $\mathbb{Q}$ and has stabiliser equal to $\mathbf{H}_0$. If $\mathbf{H}_0$ is semisimple, it has no non-trivial characters so each non-zero vector in $\bigwedge^d Z$ also has stabiliser equal to $\mathbf{H}_0$. [11, Prop. 3.1] describes how this construction can be modified to obtain a vector v_0 (not just a line) with stabiliser equal to $\mathbf{H}_0$ whenever $\mathbf{H}_0$ is reductive.

If we choose W to be stable under left as well as right translations (denoting the representations by ρ_L and ρ_R, respectively), then the same argument as in the special case above shows that the line $\mathbb{R}^\times \rho_L(u)\rho_R(u)v_0$ is defined over $\mathbb{Q}$ whenever $u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$, and so this line contains non-zero rational vectors.

Comparing this general construction with our special case of $\mathbf{G} = \mathbf{GSp}_4$, $\mathbf{H}_0 = \iota_0(\mathbf{GL}_2)$, we note that in the special case, $I(\mathbf{H}_0)$ is generated by linear functions on M_4. Thus following Chevalley's method, we could choose W to be the linear dual of $\mathrm{M}_4(\mathbb{Q})$. In fact, we chose W to be $\mathrm{M}_4(\mathbb{Q})$ itself, and Z to be the linear subspace of $\mathrm{M}_4(\mathbb{Q})$ which is annihilated by $I(\mathbf{H}_0) \cap \mathrm{M}_4(\mathbb{Q})^\vee$. The choice of $\mathrm{M}_4(\mathbb{Q})$ instead of its dual is a matter of convenience.

The representations constructed by Chevalley's method do not necessarily contain a closed orbit $\rho_L(\mathbf{G}(\mathbb{R}))v_0$, although this can often be achieved by carefully choosing $W \subset \mathbb{Q}[\mathbf{G}]$ and perhaps making some minor modifications using linear algebra constructions. On the other hand, finding a suitable v_u with bounded length requires much more detailed arithmetic information about the groups $\mathbf{H}_u$.

5.3 Closed orbit

We now show that Proposition 5.1(ii) holds, that is, the orbit $\rho_L(\mathbf{G}(\mathbb{R}))v_0$ is closed in $V_{\mathbb{R}}$. By [4, Prop. 2.3], it suffices to prove that $\rho_L(\mathbf{G}(\mathbb{C}))v_0$ is closed in $V_{\mathbb{C}}$.

We use the following definitions. If $V_{\mathbb{C}}$ is a vector space over $\mathbb{C}$, we say that a subset of $V_{\mathbb{C}}$ is **homogeneous** if it is non-empty and stable under multiplication by scalars. In other words, a subset of $V_{\mathbb{C}}$ is homogeneous if and only if it is the cone over

some subset of $\mathbb{P}(V_{\mathbb{C}})$. For a non-negative integer d, a set-theoretic function between vector spaces $f\colon V'_{\mathbb{C}} \to V_{\mathbb{C}}$ is **homogeneous of degree** d if

$$f(\lambda v) = \lambda^d f(v) \text{ for all } \lambda \in \mathbb{C},\ v \in V'_{\mathbb{C}}.$$

Homogeneous sets and homogeneous maps are useful because of the following lemma, which is equivalent to the fact that a morphism of projective algebraic varieties maps Zariski closed sets to Zariski closed sets.

Lemma 5.2. Let $V_{\mathbb{C}}$, $V'_{\mathbb{C}}$ be vector spaces over $\mathbb{C}$ (or any algebraically closed field), let $X \subset V'_{\mathbb{C}}$ be a homogeneous Zariski closed subset and let $f\colon X \to V_{\mathbb{C}}$ be a morphism of algebraic varieties which is homogeneous. If $f(x) \neq 0$ for all $x \in X \setminus \{0\}$, then $f(X)$ is a homogeneous Zariski closed subset of $V_{\mathbb{C}}$.

Let $U = \mathbb{C}^4$. We define a sequence $(u_1, u_2, u_3, u_4) \in U^4$ to be **quasi-symplectic** if it satisfies the conditions

$$\psi(u_1, u_3) = \psi(u_1, u_4) = \psi(u_2, u_3) = \psi(u_2, u_4) = 0,$$
$$\psi(u_1, u_2) = \psi(u_3, u_4).$$

If (u_1, u_2, u_3, u_4) is a quasi-symplectic sequence, then either:

(1) $\psi(u_1, u_2) = \psi(u_3, u_4) \neq 0$, in which case (u_1, u_2, u_3, u_4) is a non-zero scalar multiple of a symplectic basis for (U, ψ); or
(2) $\psi(u_1, u_2) = \psi(u_3, u_4) = 0$, in which case u_1, u_2, u_3, u_4 are contained in an isotropic subspace of U for ψ; in particular, they are linearly dependent.

Let

$$\begin{aligned}\mathbf{\Omega} &= \{g \in \mathrm{M}_4(\mathbb{C}) : \text{the columns of } g \text{ form a quasi-symplectic sequence}\}\\ &= \{g \in \mathrm{M}_4(\mathbb{C}) : \exists \nu(g) \in \mathbb{C} \text{ such that } \psi(gx, gy) = \nu(g)\psi(x, y)\}.\end{aligned}$$

The set $\mathbf{\Omega}$ is closed under multiplication, but not all of its elements are invertible so it is not a group. We have $\mathbf{\Omega} \cap \mathbf{GL}_4(\mathbb{C}) = \mathbf{GSp}_4(\mathbb{C})$.

Let σ_L denote the action of $\mathrm{M}_4(\mathbb{C})$ on $W_{\mathbb{C}} = \mathrm{End}_{\mathbb{C}}(U)$ by left multiplication (this extends our earlier definition of σ_L as a representation of $\mathbf{G} = \mathbf{GSp}_4$). Let ρ'_L denote the induced action $\bigwedge^4 \sigma_L$ of $\mathrm{M}_4(\mathbb{C})$ on $V_{\mathbb{C}} = \bigwedge^4 W_{\mathbb{C}}$ (this is a representation of $\mathrm{M}_4(\mathbb{C})$ as a

multiplicative monoid but not as a $\mathbb{C}$-algebra). Note that $\rho_L = \rho'_L \otimes \det^{-1}$, but $\rho_L(g)$ is only defined for $g \in \mathbf{GL}_4(\mathbb{C})$, while $\rho'_L(g)$ is defined for all $g \in \mathrm{M}_4(\mathbb{C})$. In particular, ρ'_L is defined on $\mathbf{Q}$.

In order to prove that $\rho_L(\mathbf{G}(\mathbb{C}))v_0$ is closed, we find a homogeneous Zariski closed set $X \subset \bigwedge^2 U^2$ such that $\rho'_L(\mathbf{Q})v_0$ is the image of X under a homogeneous morphism of varieties ζ. Hence, $\rho'_L(\mathbf{Q})v_0$ is Zariski closed. We conclude by showing that $\rho_L(\mathbf{G}(\mathbb{C}))v_0$ is the intersection of $\rho'_L(\mathbf{Q})v_0$ with a hyperplane in $V_{\mathbb{C}}$.

Lemma 5.3. The following homogeneous subset of $\bigwedge^2 U^2$ is Zariski closed:

$$X = \{(u_1, u_3) \wedge (u_2, u_4) : (u_1, u_2, u_3, u_4) \text{ is quasi-symplectic}\}.$$

Proof. Let $\bigwedge^2_{\mathrm{dec}} U^2$ denote the set of decomposable vectors in $\bigwedge^2 U^2$:

$$\bigwedge\nolimits^2_{\mathrm{dec}} U^2 = \{x \wedge y : x, y \in U^2\}.$$

This is the cone over the Grassmannian $\mathrm{Gr}(2, U^2)$ (embedded in $\mathbb{P}(\bigwedge^2 U^2)$ via the Plücker embedding), so it is a homogeneous Zariski closed subset of $\bigwedge^2 U^2$.

Define a quadratic form $q\colon U^2 \to \mathbb{C}$ by $q((u, v)) = \psi(u, v)$. Let

$$X' = \{x \wedge y \in \bigwedge\nolimits^2 U^2 : q_{|\langle x,y\rangle} = 0\},$$

where $\langle x, y\rangle$ denotes the linear subspace of U^2 spanned by x and y. Then, X' is the cone over the orthogonal Grassmannian $\mathrm{OGr}(2, U^2, q) \subset \mathbb{P}(\bigwedge^2 U^2)$, so it is a homogeneous Zariski closed subset of $\bigwedge^2_{\mathrm{dec}} U^2$.

For an element $(u_1, u_3) \wedge (u_2, u_4) \in \bigwedge^2_{\mathrm{dec}} U^2$, we have:

$$\begin{aligned}
&(u_1, u_3) \wedge (u_2, u_4) \in X' \\
\Leftrightarrow\ & q(\lambda(u_1, u_3) + \mu(u_2, u_4)) = 0 \text{ for all } \lambda, \mu \in \mathbb{C} \\
\Leftrightarrow\ & \psi(\lambda u_1 + \mu u_2, \lambda u_3 + \mu u_4) = 0 \text{ for all } \lambda, \mu \in \mathbb{C} \\
\Leftrightarrow\ & \lambda^2 \psi(u_1, u_3) + \lambda\mu\big(\psi(u_2, u_3) + \psi(u_1, u_4)\big) + \mu^2 \psi(u_2, u_4) = 0 \text{ for all } \lambda, \mu \in \mathbb{C} \\
\Leftrightarrow\ & \psi(u_1, u_3) = \psi(u_2, u_3) + \psi(u_1, u_4) = \psi(u_2, u_4) = 0. \qquad (9)
\end{aligned}$$

It follows that we can define linear maps $\Psi_{ij}\colon X' \to \mathbb{C}$ for $i = 1, 3$ and $j = 2, 4$ by

$$\Psi_{ij}((u_1, u_3) \wedge (u_2, u_4)) = \psi(u_i, u_j).$$

(When $i = 1$ and $j = 2$ or $i = 3$ and $j = 4$, the map Ψ_{ij} is defined on the whole of $\bigwedge^2 U^2$. When $i = 1$ and $j = 4$ or $i = 3$ and $j = 2$, we require (9).) Using (9), we conclude that the set X from the statement of the lemma is equal to

$$X' \cap \ker(\Psi_{14}) \cap \ker(\Psi_{23}) \cap \ker(\Psi_{12} - \Psi_{34}).$$

Thus, X is homogeneous and Zariski closed. ■

Lemma 5.4. $\rho'_L(\mathbf{\Omega})v_0$ is a Zariski closed subset of $V_{\mathbb{C}}$.

Proof. Define two linear maps $\beta_1, \beta_2\colon U^2 \to W_{\mathbb{C}} = \mathrm{M}_4(\mathbb{C})$ by

$$\beta_1(u_1, u_2) = \begin{pmatrix} u_1 & 0 & u_2 & 0 \end{pmatrix},$$
$$\beta_2(u_1, u_2) = \begin{pmatrix} 0 & u_1 & 0 & u_2 \end{pmatrix}.$$

This notation means that $\beta(u_1, u_2)$ is the 4×4 matrix with columns $u_1, 0, u_2, 0$, and similarly for β_2.

If z_1, z_2, z_3, z_4 denote the standard basis of U, then

$$\iota_0(e_1) = \beta_1(z_1, z_3), \quad \iota_0(e_2) = \beta_2(z_1, z_3), \quad \iota_0(e_3) = \beta_1(z_2, z_4), \quad \iota_0(e_4) = \beta_2(z_2, z_4). \tag{10}$$

The maps β_1 and β_2 commute with the action of $\mathbf{G}$ by left multiplication in the sense that

$$g\beta_i(u_1, u_2) = \beta_i(gu_1, gu_2)$$

for all $u_1, u_2 \in U$ and $g \in \mathbf{G}(\mathbb{C})$. Consequently,

$$\begin{aligned}\rho'_L(\mathbf{\Omega})v_0 &= \{g\iota_0(e_1) \wedge g\iota_0(e_2) \wedge g\iota_0(e_3) \wedge g\iota_0(e_4) : g \in \mathbf{\Omega}\} \\ &= \{\beta_1(gz_1, gz_3) \wedge \beta_2(gz_1, gz_3) \wedge \beta_1(gz_2, gz_4) \wedge \beta_2(gz_2, gz_4) : g \in \mathbf{\Omega}\} \\ &= \{\beta_1(u_1, u_3) \wedge \beta_2(u_1, u_3) \wedge \beta_1(u_2, u_4) \wedge \beta_2(u_2, u_4) : \\ &\quad (u_1, u_2, u_3, u_4) \text{ is quasi-symplectic}\}. \end{aligned} \tag{11}$$

Define $f\colon \bigwedge^2_{\mathrm{dec}} U^2 \to V_{\mathbb{C}} = \bigwedge^4 W_{\mathbb{C}}$ by

$$f(x \wedge y) = \beta_1(x) \wedge \beta_2(x) \wedge \beta_1(y) \wedge \beta_2(y).$$

This is well-defined, homogeneous of degree 2 and a morphism of varieties. Thanks to (11), we have $\rho'_L(\mathbf{\Omega})v_0 = f(X)$.

If $x, y \in U^2$ are linearly independent, then it is easy to check that $\beta_1(x), \beta_1(y)$ are linearly independent and that $\beta_2(x), \beta_2(y)$ are linearly independent. Furthermore, $\mathrm{im}(\beta_1) \cap \mathrm{im}(\beta_2) = \{0\}$. Hence if $x, y \in U^2$ are linearly independent, then $\beta_1(x), \beta_1(y), \beta_2(x), \beta_2(y) \in W_{\mathbb{C}}$ are linearly independent. In other words, if $x \wedge y \in \left(\bigwedge^2_{\mathrm{dec}} U^2\right) \setminus \{0\}$, then $f(x \wedge y) \neq 0$.

Hence by Lemmas 5.2 and 5.3, $f(X)$ is a Zariski closed subset of $V_{\mathbb{C}}$. ■

Lemma 5.5. There exists a linear map $s\colon V_{\mathbb{C}} \to \mathbb{C}$ such that

$$\rho_L(\mathbf{G}(\mathbb{C}))v_0 = \rho'_L(\mathbf{\Omega})v_0 \cap s^{-1}(1).$$

Proof. We continue to use the functions β_1 and β_2 from the proof of Lemma 5.4.

Let $\delta\colon W_{\mathbb{C}} \to U$ be the linear map which sends the matrix with columns $\begin{pmatrix} C_1 & C_2 & C_3 & C_4 \end{pmatrix}$ to the sum $C_1 + C_4$. This map is equivariant with respect to multiplication by $\mathrm{M}_4(\mathbb{C})$ on the left and the compositions $\delta \circ \beta_1, \delta \circ \beta_2\colon U^2 \to U$ are the projections on to the two copies of U.

Taking the fourth exterior power, δ induces a linear map

$$s\colon V_{\mathbb{C}} = \bigwedge^4 W_{\mathbb{C}} \to \bigwedge^4 U \cong \mathbb{C}.$$

By (10) and the descriptions of $\delta \circ \beta_1$, $\delta \circ \beta_2$, we have

$$s(v_0) = \delta\beta_1(z_1, z_3) \wedge \delta\beta_2(z_1, z_3) \wedge \delta\beta_1(z_2, z_4) \wedge \delta\beta_2(z_2, z_4) = -z_1 \wedge z_2 \wedge z_3 \wedge z_4.$$

Since z_1, z_2, z_3, z_4 form a basis for U, $s(v_0) \neq 0$. Hence, we can choose the isomorphism $\bigwedge^4 U \cong \mathbb{C}$ so that $s(v_0) = 1$.

The linear map δ is $\mathrm{M}_4(\mathbb{C})$-equivariant with respect to left multiplication. Consequently, s is $\mathrm{M}_4(\mathbb{C})$-equivariant with respect to ρ_L' on $V_{\mathbb{C}}$ and multiplication by the determinant on $\bigwedge^4 U$. Twisting by $\det^{-1}$, we deduce that s is $\mathbf{GL}_4(\mathbb{C})$-equivariant with respect to ρ_L on $V_{\mathbb{C}}$ and the trivial action on $\bigwedge^4 U$. Thus,

$$s(\rho_L(g)v_0) = s(v_0) = 1 \text{ for all } g \in \mathbf{G}(\mathbb{C}).$$

Furthermore, if $g \in \mathbf{G}(\mathbb{C})$, then $g = \lambda g'$ for some $g' \in \mathbf{G}(\mathbb{C}) \cap \mathbf{SL}_4(\mathbb{C})$ and $\lambda \in \mathbb{C}^\times$. Then, $\rho_L(g) = \rho_L'(g')$. Since $\mathbf{G}(\mathbb{C}) \subset \mathbf{\Omega}$, we conclude that $\rho_L(g)v_0 \in \rho_L'(\mathbf{\Omega})v_0 \cap s^{-1}(1)$.

Conversely, if $v \in \rho_L'(\mathbf{\Omega})v_0 \cap s^{-1}(1)$, then we can write $v = \rho_L'(g)v_0$ for some $g \in \mathbf{\Omega}$. Then, $s(v) = \det(g)s(v_0) = \det(g)$. Therefore, $s(v) = 1$ forces $\det(g) = 1$. Thus, $g \in \mathbf{\Omega} \cap \mathbf{SL}_4(\mathbb{C}) \subset \mathbf{G}(\mathbb{C})$ and $\rho_L(g)v_0 = \rho_L'(g)v_0 = v$. ■

5.4 Discriminants and orders

Before the proof of Proposition 5.1(iii), we prove some results on discriminants, involutions and orders in semisimple algebras.

Let D be a semisimple $\mathbb{Q}$-algebra. We define a symmetric $\mathbb{Q}$-bilinear form $\phi\colon D \times D \to \mathbb{Q}$ by

$$\phi(x, y) = \mathrm{Tr}_{D/\mathbb{Q}}(xy),$$

where $\mathrm{Tr}_{D/\mathbb{Q}}$ is the (non-reduced) trace of the regular representation of D. This form is non-degenerate by [19, Ch. I, Prop. (1.8)] (note that [19] refers to the reduced trace, which is a non-zero multiple of $\mathrm{Tr}_{D/\mathbb{Q}}$ on each simple factor). For any order R in D, we define the **discriminant** of R, denoted $\mathrm{disc}(R)$, to be $\mathrm{disc}(R, \phi)$.

For any involution $\dagger$ of D, we define another $\mathbb{Q}$-bilinear form $\phi^\dagger\colon D \times D \to \mathbb{Q}$ by

$$\phi^\dagger(x, y) = \mathrm{Tr}_{D/\mathbb{Q}}(xy^\dagger). \tag{12}$$

This form is symmetric by [19, Ch. I, Cor. (2.2) and Cor. (2.16)].

Lemma 5.6. Let R be an order in a semisimple $\mathbb{Q}$-algebra D and let $\dagger$ be any involution of D. Then, $\mathrm{disc}(R, \phi^\dagger) = \pm\mathrm{disc}(R)$.

Proof. This is based on the proof of [16, Prop. 2.9].

Let $\{d_1, \dots, d_n\}$ be a $\mathbb{Z}$-basis for R and let $A \in \mathbf{GL}_n(\mathbb{Q})$ be the matrix such that $d_j^\dagger = \sum_{i=1}^n A_{ij} d_i$. Since $\dagger$ is $\mathbb{Q}$-linear and an involution, we have $A^2 = I$, hence, $\det(A) = \pm 1$. Now $\phi^\dagger(d_i, d_j) = \sum_{k=1}^n A_{kj} \phi(d_i, d_k)$, so

$$\operatorname{disc}(R, \phi^\dagger) = \det(A) \operatorname{disc}(R) = \pm \operatorname{disc}(R).$$

■

The following lemma is restricted to quaternion algebras because its proof makes use of the fact that the reduced norm is a quadratic form on a quaternion algebra.

Lemma 5.7. There exists an absolute constant c_{39} with the following property.

Let R be an order in a quaternion algebra D over $\mathbb{Q}$. Let L be a left R-module such that $L_{\mathbb{Q}}$ is isomorphic to the left regular representation of D.

Then, there exists a left R-ideal $I \subset R$ such that I is isomorphic to L as a left R-module and

$$[R : I] \leq c_{39} |\operatorname{disc}(R)|^{3/2}.$$

Proof. This is a generalisation to quaternion algebras of Minkowski's bound for ideal classes in a number field, and the proof is similar.

Choose an isomorphism of D-modules $\eta_1 : D \to L_{\mathbb{Q}}$ and let $I_1 = \eta_1^{-1}(L)$.

Since D is a quaternion algebra, it possesses a canonical involution $*$ defined by $d^* = \operatorname{Trd}_{D/\mathbb{Q}}(d) - d$, where we write $\operatorname{Trd}_{D/\mathbb{Q}}$ for the reduced trace. The canonical involution has the property that

$$\phi^*(d, d) = \operatorname{Tr}_{D/\mathbb{Q}}(dd^*) = 4\operatorname{Nrd}_{D/\mathbb{Q}}(d) = \pm 4 \operatorname{Nm}_{D/\mathbb{Q}}(d)^{1/2}. \tag{13}$$

By [5, Lemma 1], there is an element $s \in I_1$ satisfying

$$0 < |\phi^*(s, s)| \leq c_{40} |\operatorname{disc}(I_1, \phi^*)|^{1/4}$$

(the exponent is $1/\operatorname{rk}_{\mathbb{Z}}(I_1)$). Hence by (13), there is a constant c_{41} such that

$$0 < |\operatorname{Nm}_{D/\mathbb{Q}}(s)| \leq c_{41} |\operatorname{disc}(I_1, \phi^*)|^{1/2}. \tag{14}$$

Since $\mathrm{Nm}_{D/\mathbb{Q}}(s) \neq 0$, s is invertible in D. Let $I_2 = I_1 s^{-1} \subset D$. Then, I_2 is a left R-module isomorphic to L. Since $1 = ss^{-1} \in I_2$, we have $R \subset I_2$.

Using (14), we can calculate

$$|\mathrm{disc}(I_2, \phi^*)| = |\mathrm{Nm}_{D/\mathbb{Q}}(s^{-1})|^2 \, |\mathrm{disc}(I_1, \phi^*)| \geq c_{41}^{-2}.$$

Consequently, using Lemma 5.6,

$$[I_2 : R]^2 = \frac{|\mathrm{disc}(R, \phi^*)|}{|\mathrm{disc}(I_2, \phi^*)|} \leq c_{41}^2 |\mathrm{disc}(R)|.$$

Finally let $I = [I_2 : R] I_2$. This is contained in R and is a left R-submodule of D, so it is a left R-ideal. It satisfies

$$[R : I] = \frac{[I_2 : I]}{[I_2 : R]} = \frac{[I_2 : R]^4}{[I_2 : R]} \leq c_{41}^3 |\mathrm{disc}(R)|^{3/2}.$$

■

In the following lemma, we note that $S = \mathrm{End}_R(L)$ is an order in $\mathrm{End}_D(L_{\mathbb{Q}}) \cong \mathrm{End}_D(D) \cong D^{\mathrm{op}}$. Thus, S is an order in a quaternion algebra, so $\mathrm{disc}(S)$ is defined.

Lemma 5.8. There exists an absolute constant c_{42} with the following property.

Let R be an order in a quaternion algebra D over $\mathbb{Q}$. Let L be a left R-module such that $L_{\mathbb{Q}}$ is isomorphic to the left regular representation of D. Let $S = \mathrm{End}_R(L)$. Then,

$$|\mathrm{disc}(S)| \leq c_{42} |\mathrm{disc}(R)|^4.$$

Proof. By Lemma 5.7, L is isomorphic to a left R-ideal $I \subset R$ which satisfies

$$[R : I] \leq c_{39} |\mathrm{disc}(R)|^{3/2}.$$

Then, $S = \mathrm{End}_R(L) = \mathrm{End}_R(I) \subset \mathrm{End}_D(D)$, where the latter is the ring of endomorphisms of D as a left D-module.

We can define a multiplication-reversing function $\mu : D \to \mathrm{End}_D(D)$ by

$$\mu(d)x = xd \text{ for all } d, x \in D.$$

This is a $\mathbb{Q}$-algebra isomorphism $D^{\mathrm{op}} \to \mathrm{End}_D(D)$ [35, Ch. 8, Lemma 1.10].

If $r \in R$ and $x \in I$, then we have $\mu(r)x = xr \in R$ since $I \subset R$ and R is closed under multiplication. Hence,

$$[R : I]\mu(r)x \in I.$$

Thus, $[R : I]\mu(r) \in \mathrm{End}_R(I)$ for all $r \in R$.

Let ϕ_S denote the trace form on $\mathrm{End}_D(D) = S_{\mathbb{Q}}$. Since μ is an algebra isomorphism, it pulls back ϕ_S to the trace form on D^{op}, which is equal to the trace form on D. Hence, $\mathrm{disc}(\mu(R), \phi_S) = \mathrm{disc}(R)$.

Since $[R : I]\mu(R) \subset S$, we conclude that

$$|\mathrm{disc}(S)| \leq |\mathrm{disc}([R : I]\mu(R), \phi_S)| = [R : I]^2 \, |\mathrm{disc}(R)| \leq c_{39}^2 |\mathrm{disc}(R)|^4.$$

■

5.5 Choice of v_u, γ and h

Throughout this section, c_n will denote absolute constants (in particular, independent of u).

We will now prove Proposition 5.1(iii). Thus, we are given $u \in \mathbf{G}(\mathbb{R}) = \mathbf{GSp}_4(\mathbb{R})$ such that the algebraic group $\mathbf{H}_u = u\mathbf{H}_{0,\mathbb{R}}u^{-1} \subset \mathbf{G}_{\mathbb{R}}$ is defined over $\mathbb{Q}$. Multiplying u by a scalar does not change $\mathbf{H}_u$, so we may assume that u multiplies the symplectic form ψ by ± 1; consequently $\det(u) = 1$.

Since $\mathbf{H}_u$ is defined over $\mathbb{Q}$, the $\mathbb{R}$-vector space $u\iota_0(E_{0,\mathbb{R}})u^{-1}$ is also defined over $\mathbb{Q}$. Hence, the $\mathbb{Q}$-algebra

$$E = \mathrm{M}_4(\mathbb{Q}) \cap u\iota_0(E_{0,\mathbb{R}})u^{-1}$$

satisfies $E_{\mathbb{R}} = u\iota_0(E_{0,\mathbb{R}})u^{-1}$.

Let D_0 and D denote the centralisers in $\mathrm{M}_4(\mathbb{Q})$ of $\iota_0(E_0)$ and E, respectively. Then, $D_{\mathbb{R}} = uD_{0,\mathbb{R}}u^{-1}$, so there is an isomorphism of $\mathbb{R}$-algebras $\alpha : D_{\mathbb{R}} \to D_{0,\mathbb{R}}$ defined by

$$\alpha(d) = u^{-1}du.$$

Note that

$$D_0 = \left\{ \begin{pmatrix} aI & bI \\ cI & dI \end{pmatrix} \in \mathrm{M}_4(\mathbb{Q}) : a, b, c, d \in \mathbb{Q} \right\}.$$

Hence, the "transpose" involution of $\mathrm{M}_4(\mathbb{Q})$ restricts to an involution of D_0, which we denote by t. This involution is positive, which is to say that the corresponding trace form (see (12)) is positive definite. Let

$$\dagger = \alpha^{-1} \circ t \circ \alpha \colon D_{\mathbb{R}} \to D_{\mathbb{R}}.$$

Let $L = \mathbb{Z}^4$. Let $R = \mathrm{End}_E(L)$, which is the order in D consisting of those elements preserving L.

Lemma 5.9. The quadratic form $\phi^\dagger$ takes integer values on R.

Proof. Thanks to our choice of symplectic form ψ on $\mathbb{Q}^4$, we have $\psi(d_0 x, y) = \psi(x, d_0^t y)$ for all $x, y \in \mathbb{R}^4$ and $d_0 \in D_{0,\mathbb{R}}$. Using this, the fact that u multiplies ψ by ± 1, and the definition of $\dagger$, we can calculate, for $d \in D_{\mathbb{R}}$,

$$\begin{aligned}\psi(dx, y) &= \psi(u\alpha(d)u^{-1}x,\, y) = \pm\psi(\alpha(d)u^{-1}x,\, u^{-1}y)\\ &= \pm\psi(u^{-1}x,\, \alpha(d)^t u^{-1}y) = \pm\psi(u^{-1}x,\, \alpha(d^\dagger)u^{-1}y)\\ &= \pm\psi(x,\, u\alpha(d^\dagger)u^{-1}y) = \pm\psi(x, d^\dagger y).\end{aligned}$$

Since ψ is a perfect pairing on L, this implies that $\dagger$ maps R into R. It follows that $\phi^\dagger(x, y) = \mathrm{Tr}_{D/\mathbb{Q}}(xy^\dagger) \in \mathbb{Z}$ for all $x, y \in R$. ∎

Observe that $L_{\mathbb{Q}}$ is isomorphic to the left regular representation of D_0. Hence, α induces an isomorphism between $L_{\mathbb{R}}$ and the left regular representation of $D_{\mathbb{R}}$ and it follows easily that this isomorphism can be scaled to produce an isomorphism between $L_{\mathbb{Q}}$ and the left regular representation of D. Therefore, by Lemma 5.7, there is a left R-ideal $I \subset R$ which is isomorphic to L as a left R-module, such that

$$[R : I] \leq c_{39} |\mathrm{disc}(R)|^{3/2}.$$

Choose a left R-module isomorphism $\eta \colon I \to L$.

Fix an isomorphism $\eta_0 \colon D_0 \to \mathbb{Q}^4$ of left D_0-modules (independent of u).

Lemma 5.10. There exists $h \in \mathbf{H}_0(\mathbb{R})$ such that $\eta\alpha^{-1}\eta_0^{-1} = uh$ in $\mathrm{Aut}(\mathbb{R}^4)$.

Proof. We can view $D_{\mathbb{R}}$ as a left $D_{0,\mathbb{R}}$-module with the action given by

$$d \cdot x = \alpha^{-1}(d)x. \tag{15}$$

Now $\alpha^{-1}\eta_0^{-1}\colon \mathbb{R}^4 \to D_{\mathbb{R}}$ is an isomorphism of left $D_{0,\mathbb{R}}$-modules with respect to the natural action on $\mathbb{R}^4$ and the action (15) on $D_{\mathbb{R}}$.

Since $\eta\colon D_{\mathbb{R}} \to \mathbb{R}^4$ is an isomorphism of $D_{\mathbb{R}}$-modules, it is also an isomorphism of $D_{0,\mathbb{R}}$-modules with respect to the action (15) on $D_{\mathbb{R}}$ and the natural action conjugated by u on $\mathbb{R}^4$. (We use here the fact that α^{-1} is conjugation by u.)

Finally $u^{-1}\colon \mathbb{R}^4 \to \mathbb{R}^4$ is an isomorphism of $D_{0,\mathbb{R}}$-modules with respect to the natural action conjugated by u on the domain and the natural action on the target.

Composing these, we deduce that $u^{-1}\eta\alpha^{-1}\eta_0^{-1}$ is an automorphism of $\mathbb{R}^4$ with its natural action of $D_{0,\mathbb{R}}$. In other words, $u^{-1}\eta\alpha^{-1}\eta_0^{-1}$ lies in the centraliser of $D_{0,\mathbb{R}}$ in $\mathrm{M}_4(\mathbb{R})$. By the double centraliser theorem, this centraliser is equal to $\iota_0(E_{0,\mathbb{R}})$ and so its group of invertible elements is equal to $\iota_0(\mathbf{GL}_2(\mathbb{R})) = \mathbf{H}_0(\mathbb{R})$. ■

Lemma 5.11. The absolute value of $\det(h)$ is uniformly bounded.

Proof. Let ϕ_0^t denote the bilinear form $\phi_0^t(x,y) = \mathrm{Tr}_{D_0/\mathbb{Q}}(xy^t)$ on D_0, and let ϕ_L denote the bilinear form on $\mathbb{Q}^4$ given by $\phi_L(x,y) = \phi_0^t(\eta_0^{-1}(x), \eta_0^{-1}(y))$.

Since α is an isomorphism of $\mathbb{R}$-algebras, it preserves traces, so

$$\phi^\dagger(x,y) = \phi_0^t(\alpha(x),\alpha(y)) = \phi_L(\eta_0\alpha(x), \eta_0\alpha(y)). \tag{16}$$

Consequently,

$$\mathrm{disc}(\phi^\dagger, I) = \mathrm{disc}(\phi^\dagger, \eta^{-1}(L)) = \mathrm{disc}(\phi_L, \eta_0\alpha\eta^{-1}(L)) = \det(\eta_0\alpha\eta^{-1})^2\mathrm{disc}(\phi_L, L).$$

Thanks to Lemma 5.10 and noting that $\det(u) = 1$, this can be rewritten as

$$\det(h)^2 = \det(uh)^2 = \mathrm{disc}(\phi_L, L)/\mathrm{disc}(\phi^\dagger, I).$$

By Lemma 5.9, $\phi^\dagger$ takes integer values on R and hence on I. Therefore, $\mathrm{disc}(\phi^\dagger, I)$ is a positive integer. Therefore, $\det(h)^2 \leq \mathrm{disc}(\phi_L, L)$, which is a constant. ■

Lemma 5.12. There exists a $\mathbb{Z}$-basis $\{d'_1, d'_2, d'_3, d'_4\}$ for I such that the coordinates of the vectors $\eta_0\alpha(d'_1), \eta_0\alpha(d'_2), \eta_0\alpha(d'_3), \eta_0\alpha(d'_4) \in \mathbb{R}^4$ are polynomially bounded in terms of $|\mathrm{disc}(R)|$.

Proof. By Lemma 5.9, $\phi^\dagger$ takes integer values on I. Furthermore, $\dagger$ is a positive involution (because t is a positive involution on D_0), so $\phi^\dagger$ is a positive definite quadratic form. Hence by [37, Thm. 5], there exists a $\mathbb{Z}$-basis $\{d'_1, d'_2, d'_3, d'_4\}$ for I satisfying

$$\phi^\dagger(d'_i, d'_i) \leq c_{43}\mathrm{disc}(I, \phi^\dagger)$$

for $i = 1, 2, 3, 4$. Hence by (16), the values $\phi_L(\eta_0\alpha(d'_i), \eta_0\alpha(d'_i))$ are bounded by a constant multiple of $\mathrm{disc}(I, \phi^\dagger)$. Since ϕ_L is a fixed positive definite quadratic form on $\mathbb{R}^4$, this implies that the coordinates of the vectors $\eta_0\alpha(d'_i)$ are polynomially bounded in terms of $\mathrm{disc}(I, \phi^\dagger)$.

Finally, by Lemmas 5.6 and 5.7, we have

$$\mathrm{disc}(I, \phi^\dagger) = [R : I]^2 \, |\mathrm{disc}(R)| \leq c_{44}|\mathrm{disc}(R)|^4.$$

■

Let $\{\ell_1, \ell_2, \ell_3, \ell_4\}$ denote the standard basis for L. Since $\{\eta(d'_1), \eta(d'_2), \eta(d'_3), \eta(d'_4)\}$ is a $\mathbb{Z}$-basis for L, there is a matrix $\gamma' \in \mathbf{GL}_4(\mathbb{Z})$ such that $\ell_i = \gamma'\eta(d'_i)$ for each i.

Lemma 5.13. The entries of the matrices $\gamma' uh, (\gamma' uh)^{-1} \in \mathbf{GL}_4(\mathbb{R})$ are polynomially bounded in terms of $|\mathrm{disc}(R)|$.

Proof. Let $A = \gamma' uh = \gamma'\eta\alpha^{-1}\eta_0^{-1} \in \mathbf{GL}_4(\mathbb{R})$. We have

$$\ell_i = \gamma'\eta(d'_i) = A\eta_0\alpha(d'_i).$$

By Lemma 5.12, the coordinates of the vectors $A^{-1}\ell_i = \eta_0\alpha(d'_i)$ are polynomially bounded, or in other words, the entries of the matrix A^{-1} are polynomially bounded in terms of $|\mathrm{disc}(R)|$.

Meanwhile, $|\det(\gamma')| = \det(u) = 1$ so $|\det(A)| = |\det(h)|$. By Lemma 5.11, we deduce that $|\det(A^{-1})|$ is bounded below by a positive constant. Hence by Cramer's rule, the entries of the matrix A are also polynomially bounded in terms of $|\mathrm{disc}(R)|$. ■

The following lemma establishes Proposition 5.1(iii)(b). Note that it is not required for the proof of Proposition 5.1(iii)(a): the subsequent arguments proving Proposition 5.1(iii)(a) do not use the fact that $\gamma \in \mathbf{Sp}_4(\mathbb{Z})$, so they would still work with γ' instead of γ.

Lemma 5.14. There exists $\gamma \in \Gamma = \mathbf{Sp}_4(\mathbb{Z})$ such that the entries of γuh and γuh^{-1} are polynomially bounded in terms of $|\mathrm{disc}(R)|$.

Proof. We have $uh \in \mathbf{GSp}_4(\mathbb{R})$. Consequently,

$$\psi(\gamma'^{-1}\ell_i, \gamma'^{-1}\ell_j) = \pm \det(uh)^{1/2}\, \psi((uh)^{-1}\gamma'^{-1}\ell_i,\ (uh)^{-1}\gamma'^{-1}\ell_j).$$

Using Lemmas 5.11 and 5.13, we conclude that the values $\psi(\gamma'^{-1}\ell_i, \gamma'^{-1}\ell_j)$ are polynomially bounded in terms of $|\mathrm{disc}(R)|$.

Hence by [27, Lemma 4.3], there exists a symplectic $\mathbb{Z}$-basis $\{s_1, s_2, s_3, s_4\}$ for (L, ψ) whose coordinates with respect to $\{\gamma'^{-1}\ell_1, \gamma'^{-1}\ell_2, \gamma'^{-1}\ell_3, \gamma'^{-1}\ell_4\}$ are polynomially bounded in terms of $|\mathrm{disc}(R)|$. Applying γ', we deduce that the coordinates of $\gamma's_1, \gamma's_2, \gamma's_3, \gamma's_4$ with respect to the standard basis are polynomially bounded.

Let $\gamma \in \mathbf{GL}_4(\mathbb{Z})$ be the matrix such that $\ell_i = \gamma s_i$ for each i. Since $\{s_1, s_2, s_3, s_4\}$ is a symplectic basis, we have $\gamma \in \Gamma$. We have just shown that the coordinates of $\gamma' s_i = \gamma'\gamma^{-1}\ell_i$ are polynomially bounded. In other words, the entries of the matrix $\gamma'\gamma^{-1}$ are polynomially bounded in terms of $|\mathrm{disc}(R)|$.

Multiplying $(\gamma' uh)^{-1}$ by $\gamma'\gamma^{-1}$ and applying Lemma 5.13, we deduce that the entries of $(\gamma uh)^{-1}$ are polynomially bounded in terms of $|\mathrm{disc}(R)|$. Thanks to Lemma 5.11, $|\det((\gamma uh)^{-1})|$ is bounded below by a positive constant, so it follows that the entries of (γuh) are also polynomially bounded in terms of $|\mathrm{disc}(R)|$. ∎

Let $S = E \cap \mathrm{M}_4(\mathbb{Z}) = \mathrm{End}_R(L)$ and $S_0 = \iota_0(E_0) \cap \mathrm{M}_4(\mathbb{Z})$. Set

$$d_u = (\mathrm{disc}(S)/\mathrm{disc}(S_0))^{1/2} \text{ and } v_u = d_u \rho_R(\gamma u) v_0 \in V_{\mathbb{R}}.$$

We shall use this v_u to prove Proposition 5.1(iii)(a). Note first that $d_u \rho_R(\gamma u) \in \mathrm{Aut}_{\rho_L}(\Lambda_{\mathbb{R}})$.

Lemma 5.15. $\rho_L(u) v_u \in \Lambda$.

Proof. Since $\rho_R(\gamma)$ stabilises Λ, in order to show that $\rho_L(u)v_u \in \Lambda$, it suffices to show that

$$d_u \rho_R(u)\rho_L(u)v_0 \in \Lambda.$$

Since v_0 is a generator for the rank-1 $\mathbb{Z}$-module $(\bigwedge^4 Z) \cap \Lambda = \bigwedge^4 S_0$, it follows that $\rho_R(u)\rho_L(u)v_0$ is a generator for $\bigwedge^4 uS_0u^{-1} \subset \bigwedge^4 E_{\mathbb{R}}$.

Consider a matrix $B \in \mathbf{GL}(E_{\mathbb{R}})$ such that $B(uS_0u^{-1}) = S$. Conjugation by u maps the trace form ϕ_{E_0} on $\iota_0(E_{0,\mathbb{R}})$ to the trace form ϕ_E on $E_{\mathbb{R}}$, so we have

$$\operatorname{disc}(uS_0u^{-1}, \phi_E) = \operatorname{disc}(S_0, \phi_{E_0}) = \operatorname{disc}(S_0).$$

Consequently, $\det(B)^2 = \operatorname{disc}(S)/\operatorname{disc}(S_0)$. In other words, $\det(B) = \pm d_u$.

It follows that

$$\bigwedge^4 S = \det(B) \bigwedge^4 uS_0u^{-1} = \pm d_u \bigwedge^4 uS_0u^{-1}$$

and so $d_u\rho_R(u)\rho_L(u)v_0$ is a generator for $\bigwedge^4 S \subset \Lambda$. ∎

Lemma 5.16. $|v_u| \leq c_{45} * |\operatorname{disc}(R)|^{c_{46}}$.

Proof. The action of $\mathbf{H}_0$ on the line $\bigwedge^4 Z$ via ρ_R is trivial, for the same reasons as the action via ρ_L is trivial. Therefore,

$$\begin{aligned} v_u &= d_u \rho_R(\gamma u)v_0 \\ &= d_u \rho_R(\gamma uh)v_0 \\ &= d_u \iota_0(e_1)\,(\gamma uh)^{-1} \wedge \cdots \wedge \iota_0(e_4)\,(\gamma uh)^{-1}. \end{aligned}$$

By Lemma 5.8, d_u is polynomially bounded in terms of $|\operatorname{disc}(R)|$. By Lemma 5.14, the entries of $(\gamma uh)^{-1}$ are polynomially bounded in terms of $|\operatorname{disc}(R)|$. We conclude that the coordinates, and hence the length, of v_u are polynomially bounded in terms of $|\operatorname{disc}(R)|$. ∎

Lemmas 5.15 and 5.16 complete the proof of Proposition 5.1(iii)(a).

6 Unlikely intersections in $\mathcal{A}_2$

In this section, we prove Theorems 1.3 and 1.4. We first need some preliminary material.

6.1 Realising $\mathcal{A}_2$ as a Shimura variety

Recall that $\mathcal{A}_2$ denotes the (coarse) moduli space of principally polarised abelian surfaces. To realise $\mathcal{A}_2$ as a Shimura variety, we let $(\mathbf{G}, X)$ denote the Shimura datum for which $\mathbf{G} = \mathbf{GSp}_4$ and X is isomorphic to $\mathcal{H}_2 \cup \mathcal{H}_2^-$, where $\mathcal{H}_2$ and $\mathcal{H}_2^-$ are, respectively, the Siegel upper and lower half-spaces of genus 2. (Recall that X is a $\mathbf{G}(\mathbb{R})$-conjugacy class of morphisms $\mathbb{S} \to \mathbf{G}_{\mathbb{R}}$, where $\mathbb{S} = \mathrm{Res}_{\mathbb{C}/\mathbb{R}}\mathbb{G}_{m,\mathbb{C}}$. We will henceforth identify X with $\mathcal{H}_2 \cup \mathcal{H}_2^-$.) We let $K = \mathbf{G}(\hat{\mathbb{Z}})$, where $\hat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$, the product ranging over all finite primes p. Then, K is a compact open subgroup of $\mathbf{G}(\mathbb{A}_f)$, where $\mathbb{A}_f$ denotes the finite rational adeles, and $\mathcal{A}_2$ is equal to the Shimura variety whose complex points are given by

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

As is easily seen, this is isomorphic to the quotient $\mathbf{Sp}_4(\mathbb{Z})\backslash\mathcal{H}_2$.

6.2 Quaternionic curves and E^2 curves

Quaternionic curves and E^2 curves are the images in $\mathcal{A}_2$ of Shimura varieties of PEL type, via maps induced by morphisms of Shimura data. We recall the construction of Shimura varieties of PEL-type attached to quaternion algebras over $\mathbb{Q}$, following [23, sec. 8].

Let B denote a quaternion algebra over $\mathbb{Q}$ such that $B \otimes_{\mathbb{Q}} \mathbb{R}$ is isomorphic to $\mathbf{M}_2(\mathbb{R})$ and let $\dagger$ be a positive involution of B. (Positive involutions exist for any such B, as explained in [25, pp. 195–6].) As explained in [25, p. 196], we can choose the isomorphism $B \otimes_{\mathbb{Q}} \mathbb{R} \to \mathbf{M}_2(\mathbb{R})$ in such a way that $\dagger$ corresponds to transpose of matrices, so $(B \otimes_{\mathbb{Q}} \mathbb{C}, \dagger)$ has type C in the sense of [23, Prop. 8.3].

Choose $\alpha \in B$ such that $\alpha = -\alpha^\dagger$. Define a symplectic form on B by the formula

$$\psi_\alpha(x, y) = \mathrm{Tr}_{B/\mathbb{Q}}(x\alpha y^\dagger).$$

If $\alpha \in B^\times = B \setminus \{0\}$, then ψ_α is non-degenerate and (B, ψ_α) (with B acting via the left regular representation) is a symplectic $(B, \dagger)$-module as in [23, sec. 8].

Let $\mathbf{H}_B$ denote the centraliser of $B^\times$ (acting on B via the left regular representation) in $\mathbf{GL}(B)$, seen as a $\mathbb{Q}$-algebraic group. Since every symplectic form $\psi : B \times B \to \mathbb{Q}$

which satisfies $\psi(bx,y) = \psi(x,b^\dagger y)$ has the form ψ_α for some α such that $\alpha = -\alpha^\dagger$, and (for a given positive involution $\dagger$) the set of such αs forms a one-dimensional $\mathbb{Q}$-linear subspace of B, every element of $\mathbf{H}_B$ preserves ψ_α up to multiplication by a scalar. Therefore, $\mathbf{H}_B$ is the group of $\mathbb{Q}$-linear automorphisms of B commuting with the B-action and preserving the symplectic form ψ_α up to similitudes. In other words, $\mathbf{H}_B$ is equal to the group denoted G in [23, sec. 8]. (Note: if B^{op} denotes the opposite algebra of B, we have $\mathbf{H}_B(\mathbb{Q}) \cong (B^{\mathrm{op}})^\times \cong B^\times$, where the second isomorphism uses the fact that B is a quaternion algebra.)

By [23, Prop. 8.14], there is a unique Shimura datum $(\mathbf{H}_B, X_B)$ such that each $h \in X_B$ defines a complex structure on $B \otimes_{\mathbb{Q}} \mathbb{R}$ for which the symmetric form $\psi(x,h(i)y)$ is positive or negative definite. As a Hermitian symmetric domain, X_B is isomorphic to the union of the upper and lower half-planes in $\mathbb{C}$.

Choosing a symplectic $\mathbb{Q}$-basis for (B,ψ_α), the tautological action of $\mathbf{H}_B$ on B gives rise to an injective group homomorphism $\mathbf{H}_B \to \mathbf{G}$. Thanks to the properties of X_B given to us by [23, Prop. 8.14], this induces an embedding of Shimura data $(\mathbf{H}_B, X_B) \to (\mathbf{G}, X)$. Letting $K_B = \mathbf{H}_B(\mathbb{A}_f) \cap K$, we obtain a morphism

$$\mathrm{Sh}_{K_B}(\mathbf{H}_B, X_B) \to \mathrm{Sh}_K(\mathbf{G}, X)$$

of algebraic varieties. The irreducible components of the images of such morphisms are, by definition, special curves in $\mathcal{A}_2$. If B is isomorphic (over $\mathbb{Q}$) to $\mathbf{M}_2(\mathbb{Q})$, we obtain E^2 curves, and otherwise, we obtain quaternionic curves. Any such curve parametrises abelian surfaces with multiplication by an order in B.

The Shimura data $(\mathbf{G}, X)$ and $(\mathbf{H}_B, X_B)$ all have reflex field $\mathbb{Q}$. Therefore, $\mathrm{Sh}_{K_B}(\mathbf{H}_B, X_B)$, $\mathrm{Sh}_K(\mathbf{G}, X)$ and $\mathrm{Sh}_{K_B}(\mathbf{H}_B, X_B) \to \mathrm{Sh}_K(\mathbf{G}, X)$ are all defined over $\mathbb{Q}$, but $\mathrm{Sh}_{K_B}(\mathbf{H}_B, X_B)$ often has geometrically irreducible components which are not defined over $\mathbb{Q}$. Hence, the action of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ on $\mathcal{A}_2(\mathbb{C})$ preserves the image of $\mathrm{Sh}_{K_B}(\mathbf{H}_B, X_B) \to \mathrm{Sh}_K(\mathbf{G}, X)$ but permutes its irreducible components and so acts on the set of quaternionic curves and on the set of E^2 curves in $\mathcal{A}_2$. From the theory of complex multiplication of abelian varieties, we know that $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ acts on the set of special points in $\mathcal{A}_2$. Consequently, $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ acts on

$$\Sigma_{\mathrm{Quat}} = \bigcup_{Z \in \mathcal{S}} Z \setminus Z^{sp},$$

where $\mathcal{S}$ denotes the set of quaternionic curves in $\mathcal{A}_2$ and Z^{sp} denotes the set of the special points contained in Z. Similarly, $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ acts on Σ_{E^2}.

Another way to obtain these families of special subvarieties is as follows. Let $B_0 = \mathrm{M}_2(\mathbb{Q})$. Let B_0 act on $\mathbb{Q}^4$ via the left regular representation, with respect to the basis given by (8) (which is a symplectic basis with respect to the form ψ_α where $\alpha = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$). Let $\mathbf{H}_0 \subset \mathbf{G}$ be the centraliser of this action of B_0 in $\mathbf{G}$. Then, $\mathbf{H}_0$ is equal to the image of $\mathbf{GL}_2$ embedded block diagonally, as in (7). Let

$$X_0 = \left\{ \begin{pmatrix} \tau & 0 \\ 0 & \tau \end{pmatrix} \in \mathcal{H}_2 : \mathrm{Im}(\tau) > 0 \right\},$$

$$X_0^{\pm} = \left\{ \begin{pmatrix} \tau & 0 \\ 0 & \tau \end{pmatrix} \in \mathcal{H}_2 : \mathrm{Im}(\tau) \neq 0 \right\}.$$

Then, $(\mathbf{H}_0, X_0^{\pm})$ is the unique Shimura subdatum of $(\mathbf{G}, X)$ with underlying group $\mathbf{H}_0$, and X_0 is the only connected component of $X_0^{\pm}$ contained in $\mathcal{H}_2$. We obtain a morphism of Shimura varieties $\mathcal{A}_1 \to \mathcal{A}_2$ (where $\mathcal{A}_1$ denotes the moduli space of elliptic curves), which, in terms of moduli, sends an elliptic curve E with its principal polarisation λ to the principally polarised abelian surface $(E \times E, \lambda \times \lambda)$.

For any point $x_0 \in X_0$, we have $X_0 = \mathbf{H}_0^{\mathrm{der}}(\mathbb{R})x_0$ (recall that $\mathbf{H}_0^{\mathrm{der}}(\mathbb{R}) = \mathbf{SL}_2(\mathbb{R})$ is connected) and its image in $\mathcal{A}_2$ is an E^2 curve. For any $g \in \mathbf{G}(\mathbb{R})$ such that $\mathbf{H} = g\mathbf{H}_{0,\mathbb{R}}g^{-1}$ is defined over $\mathbb{Q}$, the image of gX_0 in $\mathcal{A}_2$ is a special curve, and $\mathbf{H}$ is isomorphic (as a $\mathbb{Q}$-group) to $\mathbf{H}_B$ for some quaternion algebra B as above. If $\mathbf{H}$ is isomorphic to $\mathbf{GL}_{2,\mathbb{Q}}$, then we obtain an E^2 curve, and, otherwise, we obtain a quaternionic curve.

Lemma 6.1. Every quaternionic or E^2 curve in $\mathcal{A}_2$ is the image of gX_0 for some $g \in \mathbf{G}(\mathbb{R})$ such that $g\mathbf{H}_{0,\mathbb{R}}g^{-1}$ is defined over $\mathbb{Q}$.

Proof. Let Z be a quaternionic or E^2 curve in $\mathcal{A}_2$. Let B be the generic endomorphism algebra of the abelian surfaces parametrised by Z, and let $\dagger$ be the Rosati involution of B. Choose an analytic irreducible component Y of the preimage of Z in X.

The inclusion $\mathbf{G} \to \mathbf{GL}_4$ induces a variation $\mathcal{V}$ of $\mathbb{Q}$-Hodge structures on X with trivial underlying local system $X \times \mathbb{Q}^4$. The restriction $\mathcal{V}_{|Y}$ has endomorphism algebra B and its generic Mumford–Tate group $\mathbf{H} \subset \mathbf{G}$ is the centraliser of B in $\mathbf{G}$. Thus, $\mathbf{H}$ is the image of one of the homomorphisms $\mathbf{H}_B \to \mathbf{G}$ defined above.

The choice of basis (8) induces an isomorphism of $\mathbb{Q}$-vector spaces $\mathbb{Q}^4 \to B_0$. Choose an isomorphism of $\mathbb{R}$-algebras with involutions $(B_0 \otimes_{\mathbb{Q}} \mathbb{R}, t) \to (B \otimes_{\mathbb{Q}} \mathbb{R}, \dagger)$. The action of B on $\mathcal{V}_{|Y}$ gives rise to a B-module structure on $\mathbb{Q}^4$, and this is isomorphic

to the left regular representation of B on itself. Thus, we get an isomorphism of B-modules $B \to \mathbb{Q}^4$. Composing these isomorphisms (after extending scalars to $\mathbb{R}$), we get an isomorphism of $\mathbb{R}$-vector spaces

$$\mathbb{R}^4 \to B_0 \otimes_{\mathbb{Q}} \mathbb{R} \to B \otimes_{\mathbb{Q}} \mathbb{R} \to \mathbb{R}^4$$

or in other words an element $g \in \mathbf{GL}_4(\mathbb{R})$. Via the isomorphism $\mathbb{R}^4 \to B_0 \otimes_{\mathbb{Q}} \mathbb{R}$, the standard symplectic form on $\mathbb{R}^4$ satisfies $\psi(bx, y) = \psi(x, b^t y)$ for all $b \in B_0$, and the isomorphism $B \otimes_{\mathbb{Q}} \mathbb{R} \to \mathbb{R}^4$ behaves similarly with respect to $(B, \dagger)$. Since the spaces of symplectic forms satisfying these conditions are one-dimensional, the composed isomorphism maps the standard symplectic form to a multiple of itself; in other words, $g \in \mathbf{G}(\mathbb{R})$.

Comparing the actions of $B \otimes_{\mathbb{Q}} \mathbb{R}$ and $B_0 \otimes_{\mathbb{Q}} \mathbb{R}$, we see that $\mathbf{H}_{\mathbb{R}} = g\mathbf{H}_{0,\mathbb{R}}g^{-1}$. It follows that $g^{-1}Y$ is a connected component of a Shimura subdatum of $(\mathbf{G}, X)$ with underlying group $\mathbf{H}_0$. The only such Shimura subdatum is $(\mathbf{H}_0, X_0^{\pm})$, so $g^{-1}Y$ is a connected component of $X_0^{\pm}$.

If $g^{-1}Y \neq X_0$, replace g by $g\mathrm{diag}(1,-1,1,-1)$. Since $\mathrm{diag}(1,-1,1,-1) \in \mathbf{H}_0(\mathbb{R})$ and it swaps the two connected components of $X_0^{\pm}$, after this replacement, we will still have $\mathbf{H}_{\mathbb{R}} = g\mathbf{H}_{0,\mathbb{R}}g^{-1}$ but now $g^{-1}Y = X_0$.

Thus, we get $Y = gX_0$ and Z is the image of Y in $\mathcal{A}_2$. ■

6.3 Complexity

As in [13], we will need to define a notion of complexity. That is, to each E^2 or quaternionic curve Z in $\mathcal{A}_2$, we attach a natural number $\Delta(Z)$, which we refer to as the complexity of Z. The complexity is defined in terms of the generic endomorphism algebra of abelian surfaces parametrised by Z.

Let $g \in \mathbf{G}(\mathbb{R})$ be such that $\mathbf{H} = g\mathbf{H}_{0,\mathbb{R}}g^{-1}$ is defined over $\mathbb{Q}$. Then, the image Z of gX_0 in $\mathcal{A}_2$ is an E^2 or quaternionic curve, and by Lemma 6.1, every E^2 or quaternionic curve is obtained this way. We define the complexity $\Delta(Z)$ of Z to be $|\mathrm{disc}(R)|$, where R denotes the ring $\mathrm{End}_{\mathbf{H}}(\mathbb{Z}^4)$ of $\mathbb{Z}$-linear endomorphisms of $\mathbb{Z}^4 \subset \mathbb{Q}^4$ commuting with $\mathbf{H}(\mathbb{Q}) \subset \mathbf{G}(\mathbb{Q}) \subset \mathrm{M}_4(\mathbb{Q})$. Note that this ring R is the generic endomorphism ring of the abelian surfaces parameterised by Z. Indeed, for every non-special point of Z, the associated abelian surface (over $\mathbb{C}$) has endomorphism ring isomorphic to R.

We are now in a position to state the Galois orbits conjecture which appears in Theorem 1.3.

Conjecture 6.2. Let Σ denote Σ_{Quat} or Σ_{E^2} and let $C \subset \mathcal{A}_2$ denote an irreducible Hodge generic algebraic curve. Let L be a finitely generated subfield of $\mathbb{C}$ over which C is defined.

There exist positive constants c_{47} and c_{48} such that, for any point $s \in C \cap \Sigma$, if we let Z denote the (unique) special curve containing s, then

$$\#\mathrm{Aut}(\mathbb{C}/L) \cdot s \geq c_{47}\Delta(Z)^{c_{48}}.$$

6.4 The fixed data

We write Γ for the subgroup $\mathbf{Sp}_4(\mathbb{Z}) \subset \mathbf{G}(\mathbb{Q})$. We let $\pi : \mathcal{H}_2 \to \mathcal{A}_2$ denote the (transcendental) uniformisation map. We choose a Siegel set $\mathfrak{S} \subset \mathbf{G}(\mathbb{R})^+$ (associated with the standard Siegel triple) such that, for some finite set $C \subset \mathbf{G}(\mathbb{Q})$, $\mathcal{F}_{\mathbf{G}} = C\mathfrak{S}$ is a fundamental set for Γ in $\mathbf{G}(\mathbb{R})^+$. We write $\mathcal{F}$ for $\mathcal{F}_{\mathbf{G}}x_0$, where $x_0 \in \mathcal{H}_2$ is the point whose stabiliser in $\mathbf{G}(\mathbb{R})^+$ is the maximal compact subgroup appearing in the definition of $\mathcal{F}_{\mathbf{G}}$.

By Proposition 5.1, we can fix a finitely generated, free $\mathbb{Z}$-module Λ, a representation $\rho : \mathbf{G} \to \mathbf{GL}(\Lambda_{\mathbb{Q}})$ such that Λ is stabilised by $\rho(\Gamma)$, an element $v_0 \in \Lambda$ and positive constants c_{49} and c_{50} such that

(i) $\mathrm{Stab}_{\mathbf{G},\rho}(v_0) = \mathbf{H}_0$;

(ii) the orbit $\rho(\mathbf{G}(\mathbb{R}))v_0$ is closed in $\Lambda_{\mathbb{R}}$;

(iii) for each $u \in \mathbf{G}(\mathbb{R})$, if the group $\mathbf{H}_u = u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$, then there exists $v_u \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})v_0$ such that $\rho(u)v_u \in \Lambda$ and

$$|v_u| \leq c_{49}|\mathrm{disc}(R_u)|^{c_{50}},$$

where R_u denotes the order $\mathrm{End}_{\mathbf{H}_u}(\mathbb{Z}^4)$ of the quaternion algebra $\mathrm{End}_{\mathbf{H}_u}(\mathbb{Q}^4)$.

By Theorem 1.2, we can then fix positive constants c_{51} and c_{52} with the following property: for every $u \in \mathbf{G}(\mathbb{R})$, if $\mathbf{H}_u = u\mathbf{H}_{0,\mathbb{R}}u^{-1}$ is defined over $\mathbb{Q}$, then there exists a fundamental set for $\Gamma \cap \mathbf{H}_u(\mathbb{R})$ in $\mathbf{H}_u(\mathbb{R})$ of the form

$$B_u C\mathfrak{S}u^{-1} \cap \mathbf{H}_u(\mathbb{R}),$$

where $B_u \subset \Gamma$ is a finite set such that

$$|\rho(b^{-1}u)v_u| \leq c_{51}|v_u|^{c_{52}}$$

for every $b \in B_u$.

Choosing a basis, we obtain $\Lambda = \mathbb{Z}^d$ and we may refer to the height $\mathrm{H}(v)$ of any $v \in \Lambda$ (defined as the maximum of the absolute values of the coordinates). For any $v \in \Lambda_{\mathbb{R}}$, we write $\mathbf{G}(v) = \mathrm{Stab}_{\mathbf{G}_{\mathbb{R}},\rho}(v)$.

Proposition 6.3. Let $P \in \Sigma_{\mathrm{Quat}} \cup \Sigma_{E^2}$. Then, there exists $z \in \pi^{-1}(P) \cap \mathcal{F}$ and $v \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})\rho(\mathbf{G}(\mathbb{R}))v_0 \cap \Lambda$ such that $z(\mathbb{S}) \subset \mathbf{G}(v)$ and

$$\mathrm{H}(v) \leq c_{51} c_{49}^{c_{52}} |\mathrm{disc}(R)|^{c_{50}c_{52}},$$

where R denotes the ring $\mathrm{End}_{\mathbf{G}(v)}(L) \subset \mathrm{M}_4(\mathbb{Z})$.

Proof. Let $z \in \pi^{-1}(P) \cap \mathcal{F}$ and let Y denote the smallest pre-special subvariety of $\mathcal{H}_2$ containing z. Then, $\pi(Y)$ is an E^2 or quaternionic curve and so

$$Y = g\mathbf{H}_0^{\mathrm{der}}(\mathbb{R})x_0 = g\mathbf{H}_0^{\mathrm{der}}(\mathbb{R})g^{-1} \cdot gx_0 = \mathbf{H}^{\mathrm{der}}(\mathbb{R}) \cdot gx_0,$$

where $g \in \mathbf{G}(\mathbb{R})$, $\mathbf{H}$ is a $\mathbb{Q}$-subgroup of $\mathbf{G}$ isomorphic to $\mathbf{H}_B$ for some quaternion algebra B, as above, and $\mathbf{H}_{\mathbb{R}} = g\mathbf{H}_{0,\mathbb{R}}g^{-1}$. Since $\mathbf{H}$ is the Mumford–Tate group of Y (i.e. the smallest $\mathbb{Q}$-subgroup of $\mathbf{G}$ containing $x(\mathbb{S})$ for all $x \in Y$), we have $z(\mathbb{S}) \subset \mathbf{H}(\mathbb{R})$. By Proposition 5.1, we obtain $v \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})v_0$ such that $\rho(g)v \in \Lambda$ and

$$|v| \leq c_{49} |\mathrm{disc}(R)|^{c_{50}},$$

where $R = \mathrm{End}_{\mathbf{H}}(L)$. Note that

$$\mathbf{G}(\rho(g)v) = g\mathbf{G}(v)g^{-1} = g\mathbf{G}(v_0)g^{-1} = \mathbf{H}_{\mathbb{R}}.$$

By Theorem 1.2 (with $u = g$), we obtain a finite set $B_g \subset \Gamma$ such that

$$\mathcal{F}_{\mathbf{H}} = B_g C\mathfrak{S}g^{-1} \cap \mathbf{H}(\mathbb{R})$$

is a fundamental set for $\Gamma_{\mathbf{H}} = \Gamma \cap \mathbf{H}(\mathbb{R})$ in $\mathbf{H}(\mathbb{R})$, and

$$|\rho(b^{-1}g)v| \leq c_{51} |v|^{c_{52}}$$

for all $b \in B_g$. In particular, since $z \in \mathbf{H}(\mathbb{R})gx_0$, we can write

$$z = \gamma bsx_0$$

for some $\gamma \in \Gamma_{\mathbf{H}}$, $b \in B_g$ and $s \in C\mathfrak{S}$. Hence,

$$z' := b^{-1}\gamma^{-1}z \in \mathcal{F} \cap \pi^{-1}(P).$$

Furthermore, we have

$$z'(\mathbb{S}) \subset \mathbf{G}(\rho(b^{-1}\gamma^{-1}g)v) = \mathbf{G}(\rho(b^{-1}g)v) = b^{-1}\mathbf{H}_{\mathbb{R}}b,$$

where we use the fact that $\gamma \in \mathbf{H}(\mathbb{R}) = \mathbf{G}(\rho(g)v)(\mathbb{R})$. Finally, from the above, we have

$$|\rho(b^{-1}g)v| \leq c_{51}|v|^{c_{52}} \leq c_{51}(c_{49}|\mathrm{disc}(R)|^{c_{50}})^{c_{52}} = c_{51}c_{49}^{c_{52}}|\mathrm{disc}(R)|^{c_{50}c_{52}}$$

and

$$R = \mathrm{End}_{\mathbf{H}}(L) = \mathrm{End}_{b^{-1}\mathbf{H}b}(b^{-1}L) = \mathrm{End}_{b^{-1}\mathbf{H}b}(L) = \mathrm{End}_{\mathbf{G}(\rho(b^{-1}g)v)}(L).$$

Therefore, since $\rho(b^{-1}g)v \in \Lambda$, we conclude that z' and $\rho(b^{-1}g)v$ satisfy the conditions of the proposition. ■

Corollary 6.4. Let $b \in \mathbb{R}$. The set of E^2 or quaternionic curves Z satisfying $\Delta(Z) \leq b$ is finite.

Proof. Let Z be an E^2 or quaternionic curve satisfying $\Delta(Z) \leq b$ and let $P \in Z$ be a Hodge generic point on Z. Therefore, $P \in \Sigma_{\mathrm{Quat}} \cup \Sigma_{E^2}$ and, applying Proposition 6.3 and the first paragraph of its proof, we obtain $v \in \Lambda$ satisfying

$$\mathrm{H}(v) \leq c_{51}c_{49}^{c_{50}}b^{c_{50}c_{52}}$$

such that Z is the image in $\mathcal{A}_2$ of an orbit of $\mathbf{G}(v)^{\mathrm{der}}(\mathbb{R})$. As in the proof of Lemma 6.1, there is only one Shimura subdatum of $(\mathbf{G}, X)$ associated with $\mathbf{G}(v)$ and so the result follows. ■

6.5 Proof of Theorem 1.3 for quaternionic curves

Let $\mathcal{C} = \pi^{-1}(C) \cap \mathcal{F}$—a set definable in the o-minimal structure $\mathbb{R}_{\mathrm{an,exp}}$ (see [20] for more details). Let L be a finitely generated field of definition for C. Let $P \in C \cap \Sigma_{\mathrm{Quat}}$. Varying over $\sigma \in \mathrm{Aut}(\mathbb{C}/L)$, we obtain points $\sigma(P) \in C \cap \Sigma_{\mathrm{Quat}}$ and, for each σ, we let $z_\sigma \in \mathcal{F} \cap \pi^{-1}(\sigma(P))$ and we let $v_\sigma \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})\rho(\mathbf{G}(\mathbb{R}))v_0 \cap \Lambda$ be the elements afforded to us by Proposition 6.3. That is, $z_\sigma(\mathbb{S}) \subset \mathbf{G}(v_\sigma)$ and

$$\mathrm{H}(v_\sigma) \leq c_{51} c_{49}^{c_{50}} |\mathrm{disc}(R_\sigma)|^{c_{50}c_{52}},$$

where R_σ denotes the ring $\mathrm{End}_{\mathbf{G}(v_\sigma)}(L) \subset \mathrm{M}_4(\mathbb{Z})$. As above, $|\mathrm{disc}(R_\sigma)| = \Delta(\sigma(Z))$. Note that we also have $z_\sigma \in \mathcal{C}$.

We obtain a set Θ of tuples $(v_\sigma, z_\sigma) \in \Lambda \times \mathcal{C}$ belonging to the definable set

$$D = \{(v, z) \in \Lambda_{\mathbb{R}} \times \mathcal{C} : v \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})\rho(\mathbf{G}(\mathbb{R}))v_0,\ z(\mathbb{S}) \subset \mathbf{G}(v)\}.$$

Let $\pi_1 : D \to \Lambda_{\mathbb{R}}$ and let $\pi_2 : D \to \mathcal{C}$ denote the projection maps. By Conjecture 6.2, we have

$$\begin{aligned} A := \#\pi_2(\Theta) = \#\mathrm{Aut}(\mathbb{C}/L) \cdot P = \#\mathrm{Aut}(\mathbb{C}/L) \cdot \sigma(P) \\ \geq c_{47} |\mathrm{disc}(R_\sigma)|^{c_{48}} \geq c_{53} H(v_\sigma)^{c_{48}/c_{50}c_{52}}. \end{aligned}$$

Applying [13, Theorem 9.1] (a variant of [18, Corollary 7.2]), in the case $l = 0$, $k = 1$, $T = (1/c_{54}A)^{c_{50}c_{52}/c_{48}}$ and $\varepsilon < c_{48}/c_{50}c_{52}$, we conclude that either

(1) $A = \#\pi_2(\Theta)$ is bounded, hence $\Delta(Z)$ is bounded and the theorem holds, or
(2) there exists a continuous definable function

$$\beta : [0, 1] \to D$$

such that $\beta_1 = \pi_1 \circ \beta$ is semi-algebraic, $\beta_2 = \pi_2 \circ \beta$ is non-constant, $\beta(0) \in \Theta$, and $\beta_{|(0,1)}$ is real analytic.

Therefore, it suffices to rule out the latter possibility. To that end, suppose that we have such a function. By definable choice, there exists a semi-algebraic function

$$\tilde{\beta}_1 : [0, 1] \to \mathrm{Aut}_{\rho}(\Lambda_{\mathbb{R}})\rho(\mathbf{G}(\mathbb{R}))$$

such that $\tilde{\beta}_1(t) \cdot v_0 = \beta_1(t)$ for all $t \in [0,1]$. We let $v_t = \beta_1(t)$, $g_t = \tilde{\beta}_1(t)$ and $z_t = \beta_2(t)$. Since $z_t(\mathbb{S}) \subset \mathbf{G}(v_t)$ and $g_t \in \mathrm{Aut}_{\rho(\mathbf{G})}(\Lambda_{\mathbb{R}})\rho(\mathbf{G}(\mathbb{R}))$, we have

$$(g_t^{-1}z_t)(\mathbb{S}) \subset g_t^{-1}\mathbf{G}(v_t)g_t = \mathbf{G}(g_t^{-1}v_t) = \mathbf{G}(v_0) = \mathbf{H}_0(\mathbb{R}).$$

We conclude that $g_t^{-1}z_t$ lies on the unique pre-special subvariety of $\mathcal{H}_2$ associated with $\mathbf{H}_0^{\mathrm{der}}(\mathbb{R})$, namely, X_0.

On the one hand, there exists $0 < t_1 \leq 1$ such that $\beta_2([0,t_1])$ is contained in a single irreducible analytic component $\tilde{C}$ of $\pi^{-1}(C)$. By [36, Theorem 1.3] (the inverse Ax–Lindemann conjecture), $\mathcal{H}_2$ is the smallest algebraic subset of $\mathcal{H}_2$ containing $\tilde{C}$.

Let $B \subset \Lambda_{\mathbb{C}}$ denote the Zariski closure of $\beta_1([0,t_1]) \subset \Lambda_{\mathbb{R}}$. By definable choice, there exists a complex algebraic set $\tilde{B} \subset \mathrm{Aut}_{\rho}(\Lambda_{\mathbb{C}})\rho(\mathbf{G}(\mathbb{C}))$ of dimension at most 1 whose image under the algebraic map $g \mapsto g \cdot v_0$ is B. Using the superscript $^\vee$ to denote the compact dual of a hermitian symmetric domain, we obtain a complex algebraic set $\tilde{B} \times X_0^\vee$ of dimension at most 2. Note that $\tilde{B} \cdot \mathcal{H}_2^\vee = \mathcal{H}_2^\vee$. Hence, $\tilde{B} \cdot X_0^\vee \subset \mathcal{H}_2^\vee$ is algebraic of dimension at most 2.

On the other hand, $\tilde{C}$ is an irreducible complex analytic curve having an uncountable intersection with $\tilde{B} \cdot X_0^\vee$ (in particular, it includes $\beta_2([0,t_1])$ because $g_t^{-1}z_t \in X_0$ and $z_t \in \tilde{C}$ for all $t \in [0,t_1]$). Therefore, $\tilde{C}$ is contained in $\tilde{B} \cdot X_0^\vee$, hence, so is $\mathcal{H}_2^\vee$. However, $\dim \mathcal{H}_2^\vee = 3$, and we arrive at a contradiction.

6.6 Proof of Theorem 1.3 for E^2 curves

The proof is the same as in the previous section, working with Σ_{E^2} instead of Σ_{Quat}.

6.7 Proof of Theorem 1.4

If C is an algebraic curve over a number field and $\mathfrak{A} \to C$ is an abelian scheme of relative dimension 2, we say that $s \in C(\overline{\mathbb{Q}})$ is a quaternionic point if the endomorphism algebra of the fiber $\mathfrak{A}_s$ is a quaternion algebra over $\mathbb{Q}$ not isomorphic to $\mathrm{M}_2(\mathbb{Q})$.

We claim that it suffices to prove the following theorem.

Theorem 6.5. Let C be an irreducible algebraic curve and let $\mathfrak{A} \to C$ be a principally polarised non-isotrivial abelian scheme of relative dimension 2 such that $\mathrm{End}(\mathfrak{A}_{\overline{\eta}}) = \mathbb{Z}$, where $\overline{\eta}$ denotes a geometric generic point of C.

Suppose that C and $\mathfrak{A}$ are defined over a number field L and that there exist a curve C', a semiabelian scheme $\mathfrak{A}' \to C'$ and an open immersion $\iota\colon C \to C'$, all defined

over $\overline{\mathbb{Q}}$, such that $\mathfrak{A} \cong \iota^*\mathfrak{A}'$ and there is a point $s_0 \in C'(\overline{\mathbb{Q}}) \setminus C(\overline{\mathbb{Q}})$ for which the fibre $\mathfrak{A}'_{s_0}$ is a torus.

Then, there exist positive constants c_{55} and c_{56} such that, for any quaternionic point $s \in C$,

$$\#\mathrm{Aut}(\mathbb{C}/L) \cdot s \geq c_{55}|\mathrm{disc}(\mathrm{End}(\mathfrak{A}_s))|^{c_{56}}.$$

To see that Theorem 6.5 implies Theorem 1.4, consider C as in Theorem 1.4. Then, C is defined over a number field L and, furthermore, we can construct a curve $\tilde{C}'$, a finite surjective morphism $q : \tilde{C} \to C$, $s_0 \in \tilde{C}'$ and a semiabelian scheme $\mathfrak{A}' \to \tilde{C}'$ as in [12, Proposition 9.4]. We can find a finite extension $\tilde{L}/L$ such that $\tilde{C}'$, $q : \tilde{C} \to C$, s_0 and $\mathfrak{A}'$ are all defined over $\tilde{L}$. The abelian scheme $\mathfrak{A}'_{|\tilde{C}} \to \tilde{C}$ and the point $s_0 \in \tilde{C}'(\overline{\mathbb{Q}})$ satisfy the conditions of Theorem 6.5 and so, for any quaternionic point $\tilde{s} \in \tilde{C}$,

$$\#\mathrm{Aut}(\mathbb{C}/\tilde{L}) \cdot \tilde{s} \geq c_{55}|\mathrm{disc}(\mathrm{End}(\mathfrak{A}'_{\tilde{s}}))|^{c_{56}}.$$

If $s \in C \cap \Sigma_{\mathrm{Quat}}$, then we can find a quaternionic point $\tilde{s} \in \tilde{C}$ such that $q(\tilde{s}) = s$, and since q is finite,

$$\#\mathrm{Aut}(\mathbb{C}/L) \cdot s \geq c_{57}\#\mathrm{Aut}(\mathbb{C}/\tilde{L}) \cdot \tilde{s}.$$

Let Z denote the unique special curve in $\mathcal{A}_2$ containing s. Since s is a Hodge generic point of Z, the endomorphism ring of the associated abelian surface A_s is isomorphic to the generic endomorphism ring of Z and so $\Delta(Z) = |\mathrm{disc}(\mathrm{End}(A_s))|$. Since also $A_s = \mathfrak{A}'_{\tilde{s}}$, we can combine the above inequalities to obtain

$$\#\mathrm{Aut}(\mathbb{C}/L) \cdot s \geq c_{55}c_{57}\Delta(Z)^{c_{56}},$$

that is, Conjecture 6.2.

Therefore, it remains to prove Theorem 6.5.

Proof of Theorem 6.5. After a finite extension, we may assume that C', $\mathfrak{A}' \to C'$, $\iota\colon C \to C'$ and s_0 are all defined over L. Since $\mathrm{End}(\mathfrak{A}_{\overline{\eta}}) = \mathbb{Z}$ and $\dim(\mathfrak{A}_{\overline{\eta}}) = 2$, the Mumford–Tate group of $\mathfrak{A}_{\overline{\eta}}$ is $\mathbf{GSp}_{4,\mathbb{Q}}$ (see [12, Section 2.F]). Thus, $\mathfrak{A} \to C$ satisfies the conditions of [12, Theorem 8.1], as modified in [12, Remark 8.6].

Let $s \in C$ be a quaternionic point. The image of s under the map $C \to \mathcal{A}_2$ induced by $\mathfrak{A} \to C$ is in the intersection between the image of C and a quaternionic curve. We deduce that $s \in C(\overline{\mathbb{Q}})$.

Now $\mathrm{End}(\mathfrak{A}_s) \otimes \mathbb{Q}$ is a non-split quaternion algebra, so cannot inject into $\mathrm{M}_2(\mathbb{Q})$. Hence, A_s is exceptional in the sense of [12, Section 8]. Therefore, by [12, Theorem 8.1], $h(s)$ is polynomially bounded in terms of $[L(s) : L]$, where h denotes a Weil height on C'. Let h_F denote the stable Faltings height. As proved in [15, p. 356],

$$|h_F(\mathfrak{A}_s) - h(s)| = \mathrm{O}(\log h(s)).$$

We conclude that $h_F(\mathfrak{A}_s)$ is polynomially bounded in terms of $[L(s) : L]$.

In order to deduce a bound for $\mathrm{disc}(\mathrm{End}(A_s))$, we use the following theorem of Masser and Wüstholz.

Theorem 6.6. [26, p. 641] Given positive integers n, d and δ, there are constants $c_{58} = c_{58}(n, d, \delta)$ and $c_{59} = c_{59}(n)$, with the following property. Let A be an abelian variety of dimension n defined over a number field k of degree d, equipped with a polarisation of degree δ. Let $\dagger$ be the Rosati involution of $\mathrm{End}(A)$ associated with this polarisation and let $\phi^\dagger$ be the bilinear form on $\mathrm{End}(A)$ defined by (12). Then, $\mathrm{disc}(\mathrm{End}_k(A), \phi^\dagger)$ is at most $c_{58} \max(1, h_F(A))^{c_{59}}$.

As remarked in [26] immediately following the statement of this theorem, one can replace $\mathrm{End}_k(A)$ by $\mathrm{End}_{\mathbb{C}}(A)$ (the endomorphism ring which appears in the statement of Theorem 6.5) because one can find a finite extension K/k of degree bounded only in terms of n such that $\mathrm{End}_K(A) = \mathrm{End}_{\mathbb{C}}(A)$. Furthermore, as stated near the bottom of [26, p. 650], the constant $c_{58}(n, d, \delta)$ is polynomial in d and δ, for a polynomial which depends only on n. Using also Lemma 5.6 to see that $|\mathrm{disc}(\mathrm{End}(A), \phi^\dagger)| = |\mathrm{disc}(\mathrm{End}(A))|$, we conclude that there are constants c_{60}, c_{61} depending only on n such that, for all A as in the theorem, we have

$$|\mathrm{disc}(\mathrm{End}(A))| \leq c_{60} \max(\delta, d, h_F(A))^{c_{61}}. \tag{17}$$

In our case, we have always $n = 2$ and $\delta = 1$. So applying (17) together with the fact that $h_F(A_s)$ is polynomially bounded in terms of $[L(s) : L]$ completes the proof of Theorem 6.5. ∎

Acknowledgements

Both authors are grateful to the anonymous referees for their suggestions which have improved the paper. They would also like to thank the Universities of Reading, Oxford and Warwick.

Funding

This work was supported by the Engineering and Physical Sciences Research Council [EP/S029613/1 to C.D., EP/T010134/1 to M.O.].

References

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

[2] André, Y. *G-functions and geometry.* Aspects of Mathematics, E13, Friedr. Braunschweig: Vieweg & Sohn, 1989.

[3] Borovoi, M., C. Daw, and J. Ren. *Conjugation of semisimple subgroups over real number fields of bounded degree* to appear in Proc. Amer. Math. Soc., available at arXiv:1802.05894.

[4] Borel, A. and Harish-Chandra. "Arithmetic subgroups of algebraic groups." *Annals of Mathematics. Second Series* 75 (1962): 485–535.

[5] Blaney, H. "Indefinite quadratic forms in n variables." *J. London Math. Soc.* 23 (1948): 153–60.

[6] Borel, A. *Introduction aux groupes arithmétiques*. Publications de l'Institut de Mathématique de l'Université de Strasbourg, XV. Actualités Scientifiques et Industrielles, No. 1341, Hermann, Paris, 1969.

[7] Borel, A. and J. Tits. "Groupes réductifs." *Inst. Hautes études Sci. Publ. Math.* 27 (1965): 55–150.

[8] Cassels, J. W. S. "Rational quadratic forms." *London Mathematical Society Monographs*, Vol. 13. London-New York: Academic Press, Inc., 1978.

[9] Chevalley, C. *Classification des groupes de Lie algébriques*, Vol. 1. 1956–8. Paris: Secrétariat mathématique, 1958 Séminaire C. Chevalley.

[10] Deligne, P. "Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques." *Automorphic forms, representations and L-functions (Part 2), Proc. Sympos. Pure Math.* XXXIII. 247–89. Providence, R.I: AMS, 1979.

[11] Deligne, P. *Hodge cycles on abelian varieties*, *Hodge cycles, motives, and Shimura varieties.* Lecture Notes in Mathematics, vol. 900. Springer-Verlag, 1982, Notes by J. S. Milne, pp. 9–100.

[12] Daw, C. and M. Orr. *Unlikely intersections with $E\times XM$ curves in A_2*. to appear in *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, available at arXiv:1902.10483.

[13] Daw, C. and J. Ren. "Applications of the hyperbolic Ax–Schanuel conjecture." *Compositio Math.* 154 (2018): 1843–88.

[14] Eberlein, P. "Growth estimates for orbits of self-adjoint groups." *Geom. Dedicata* 170 (2014): 87–117.

[15] Faltings, G. "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern." *Invent. Math.* 73, no. 3 (1983): 349–66.

[16] Gaudron, É. and G. Rémond. "Polarisations et isogénies." *Duke Math. J.* 163, no. 11 (2014): 2057–108.

[17] Grunewald, F. and D. Segal. "Some general algorithms. I. Arithmetic groups." *Annals of Mathematics. Second Series* 112, no. 3 (1980): 531–83.

[18] Habegger, P. and J. Pila. "O-minimality and certain atypical intersections." *Annales Scientifiques de l'École Normale Supérieure. Quatrième Série* 49, no. 4 (2016): 813–58.

[19] Knus, M.-A., A. Merkurjev, M. Rost, and J.-P. Tignol. "The book of involutions." *American Mathematical Society Colloquium Publications*, Vol. 44. Providence, RI: American Mathematical Society, 1998.

[20] Klingler, B., E. Ullmo, and A. Yafaev. "The hyperbolic Ax-Lindemann-Weierstrass conjecture." *Publications mathématiques de l'IHÉS* 123, no. 1 (2016): 333–60.

[21] Li, H. and G. A. Margulis. "Effective estimates on integral quadratic forms: Masser's conjecture, generators of orthogonal groups, and bounds in reduction theory." *Geom. Funct. Anal.* 26, no. 3 (2016): 874–908.

[22] Milne, J. S. "Abelian varieties." *Arithmetic geometry (Storrs, Conn., 1984)*. 103–50. New York: Springer, 1986.

[23] Milne, J. S. "Introduction to Shimura varieties." *Harmonic analysis, the trace formula, and Shimura varieties*, Clay Math. Proc, Vol. 4. 265–378. Providence, RI: Amer. Math. Soc, 2005.

[24] Mostow, G. D. "Self-adjoint groups." *Annals of Mathematics. Second Series* 62 (1955): 44–55.

[25] Mumford, D. *Abelian varieties*, second ed. Oxford University Press, 1974.

[26] Masser, D. W. and G. Wüstholz. "Endomorphism estimates for abelian varieties." *Math. Z.* 215, no. 4 (1994): 641–53 MR 1269495.

[27] Orr, M. "Families of abelian varieties with many isogenous fibres." *J. Reine Angew. Math.*, no. 705 (2015): 211–31.

[28] Orr, M. "On compatibility between isogenies and polarizations of abelian varieties." *Int. J. Number Theory* 13, no. 3 (2017): 673–704.

[29] Orr, M. "Height bounds and the Siegel property." *Algebra & Number Theory* 12, no. 2 (2018): 455–78.

[30] Orr, M. "Unlikely intersections with Hecke translates of a special subvariety." *J. Eur. Math. Soc.* 23, no. 1 (2021): 1–28.

[31] Pink, R. *A common generalization of the conjectures of André–Oort, Manin–Mumford, and Mordell–Lang*. 2005. preprint, available at http://www.math.ethz.ch/~pink/ftp/AOMMML.pdf.

[32] Platonov, V. and A. Rapinchuk. Algebraic groups and number theory, *Pure and Applied Mathematics*, vol. 139. Boston, MA: Academic Press, Inc., 1994, Translated from the 1991 Russian original by R. Rowen.

[33] Pila, J. and J. Tsimerman. "The André–Oort conjecture for the moduli space of abelian surfaces." *Compositio Math.* 149, no. 2 (2013): 204–16.

[34] Richardson, R. W. and P. J. Slodowy. "Minimum vectors for real reductive algebraic groups." *J. London Math. Soc. (2)* 42, no. 3 (1990): 409–29.

[35] Scharlau, W. "Quadratic and Hermitian forms." *Grundlehren der Mathematischen Wissenschaften*, Vol. 270. Berlin: Springer-Verlag, 1985.

[36] Ullmo, E. and A. Yafaev. "Algebraic flows on Shimura varieties." *Manuscripta Math.* 155, no. 3–4 (2018): 355–67 MR 3763410.

[37] Weyl, H. "Theory of reduction for arithmetical equivalence." *Trans. Amer. Math. Soc.* 48 (1940): 126–64.