THÈSE DE DOCTORAT

présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ PARIS-SUD

Préparée à l'École Doctorale
de Mathématiques de la région Paris-Sud

Spécialité : Mathématiques

par

Martin Orr

La conjecture d'André–Pink : Orbites de Hecke et sous-variétés faiblement spéciales

Soutenue le 25 septembre 2013 devant la Commission d'examen :

M. Jean-Benoît	Bost	Université Paris-Sud	*Examinateur*
M. Antoine	Chambert-Loir	Université Paris-Sud	*Examinateur*
M. Bruno	Klingler	Université Paris-Diderot	*Examinateur*
M. Jonathan	Pila	University of Oxford	*Rapporteur*
M. Emmanuel	Ullmo	Université Paris-Sud, IHES	*Directeur de thèse*
M. Andrei	Yafaev	University College London	*Rapporteur*

Thèse préparée au
Département de Mathématiques d'Orsay
Laboratoire de Mathématiques (UMR 8628), Bât. 425
Université Paris-Sud
91 405 Orsay CEDEX

La conjecture d'André–Pink : Orbites de Hecke et sous-variétés faiblement spéciales

Martin Orr

Résumé

La conjecture d'André–Pink affirme qu'une sous-variété d'une variété de Shimura ayant une intersection dense avec une orbite de Hecke est faiblement spéciale. On démontre cette conjecture dans le cas de courbes dans une variété de Shimura de type abélien, ainsi que dans certains cas de sous-variétés de dimension supérieure. Ceci est un cas spécial de la conjecture de Zilber–Pink. C'est une généralisation de théorèmes d'Edixhoven et Yafaev quand l'orbite de Hecke se compose de points spéciaux, de Pink quand l'orbite de Hecke se compose de points Galois génériques, et de Habegger et Pila quand la variété de Shimura est un produit de courbes modulaires.

Notre démonstration de la conjecture d'André–Pink pour les courbes dans l'espace de modules des variétés abéliennes principalement polarisées est basée sur la méthode de Pila et Zannier, utilisant une variante forte du théorème de comptage de Pila–Wilkie. On obtient les bornes galoisiennes requises grâce au théorème d'isogénie de Masser et Wüstholz. Afin de relier les bornes sur les isogénies aux hauteurs, on démontre également diverses bornes concernant l'arithmétique des formes hermitiennes sur l'anneau d'endomorphismes d'une variété abélienne. Afin d'étendre le résultat sur la conjecture d'André–Pink aux courbes dans les variétés de Shimura de type abélien et à certains cas de sous-variétés de dimension supérieure, on étudie les propriétés fonctorielles de plusieurs variantes des orbites de Hecke.

Un chapitre concerne les rangs des groupes de Mumford–Tate de variétés abéliennes complexes. On y démontre une minoration de ces rangs en fonction de la dimension de la variété abélienne, étant donné que ses sous-variétés abéliennes simples sont deux à deux non isogènes.

The André–Pink conjecture: Hecke orbits and weakly special subvarieties

Martin Orr

Abstract

The André–Pink conjecture predicts that a subvariety of a Shimura variety which has dense intersection with a Hecke orbit is weakly special. We prove this conjecture for curves in a Shimura variety of abelian type, as well as for certain cases for subvarieties of higher dimension. This is a special case of the Zilber–Pink conjecture. It generalises theorems of Edixhoven and Yafaev when the Hecke orbit consists of special points, of Pink when the Hecke orbit consists of Galois generic points, and of Habegger and Pila when the Shimura variety is a product of modular curves.

Our proof of the André–Pink conjecture for curves in the moduli space of principally polarised abelian varieties is based on the Pila–Zannier method, using a strong form of the Pila–Wilkie counting theorem. The necessary Galois bounds are obtained from the Masser–Wüstholz isogeny theorem. In order to relate isogeny bounds to heights, we also prove various bounds concerning the arithmetic of Hermitian forms over the endomorphism ring of an abelian variety. In order to extend the result on the André–Pink conjecture to curves in Shimura varieties of abelian type and to some cases of higher-dimensional subvarieties, we study the functorial properties of Hecke orbits and variations thereof.

One chapter concerns the ranks of Mumford–Tate groups of complex abelian varieties. We prove a lower bound for these ranks in terms of the dimension of the abelian variety, subject to the condition that the simple abelian subvarieties are pairwise non-isogenous.

Remerciements

Je tiens d'abord à remercier Emmanuel Ullmo, mon directeur de thèse. Il a su me proposer des problèmes intéressants et abordables et a perseveré jusqu'à ce que je les croyais potentiellement résolubles.

Je remercie mes rapporteurs, Jonathan Pila et Andrei Yafaev, pour la relecture de ce texte et pour l'intérêt qu'ils ont manifesté envers mes travaux tout au long de la thèse. Je remercie également Jean-Benoît Bost, Antoine Chambert-Loir et Bruno Klingler d'avoir accepté de faire partie du jury de la soutenance.

Je remercie Chris Daw pour les discussions qu'on a eues ainsi que pour notre collaboration récente, et dans la recherche et dans nos aventures aux colloques des logiciens. J'espère que cela se prolongera à Londres. Je remercie Ziyang qui m'a parfois posé des questions intéressantes.

Je remercie Barinder pour ses longues lettres, pour nos échanges quand on a pu se visiter et pour avoir souvent été le premier lecteur de mes textes.

Merci aux logiciens qui m'ont toujours acceuilli chaleureusement, et surtout à Élisabeth Bouscaren pour m'avoir encouragé à partagé mes travaux.

Je remercie les collègues avec lesquels j'ai discuté de la géométrie algébrique et de la théorie des nombres depuis le M2, en particulier Giancarlo, Arne et Arno. Je remercie Clément et Lionel qui ont relu l'introduction en français de cette thèse, ainsi que les autres doctorants qui ont de temps en temps répondu à mes questions de français, surtout Tristan.

Merci à Valérie Blandin-Lavigne pour son aide avec les tâches adminstratives et à David Harari qui m'a d'abord suggéré de parler à Emmanuel.

Contents

1 Introduction (Français)

Le sujet principal de cette thèse est une conjecture d'André et Pink affirmant qu'une sous-variété d'une variété de Shimura est faiblement spéciale si elle a une intersection dense avec une orbite de Hecke. Ceci est une de plusieurs conjectures concernant les sous-variétés de variétés de Shimura mixtes, qui sont des corollaires de la conjecture de Zilber–Pink. Jusqu'à ce jour, les résultats sur ces conjectures dans le cas de variétés de Shimura pures ont été limités, soit aux points spéciaux (la conjecture d'André–Oort), soit aux produits de courbes modulaires. Cette thèse avance donc dans une nouvelle direction en considerant les orbites de Hecke dans une variété de Shimura de type abélien.

Variétés de Shimura. On commence par une esquisse de la définition d'une variété de Shimura connexe. Soit G un groupe réductif connexe défini sur $\mathbb{Q}$. Soit X^+ une composante connexe d'une classe de $G(\mathbb{R})$-conjugaison de **paramètres de Hodge**, c'est à dire des homomorphismes $\mathrm{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m \to G_{\mathbb{R}}$. On appelle **donnée de Shimura connexe** un tel couple (G, X^+) vérifiant les conditions 2.1.1.1–2.1.1.3 de [Del79]. En particulier, ces conditions assurent que X^+ est un domaine hermitien symétrique. Une **variété de Shimura connexe** est le quotient de X^+ par l'action d'un sous-groupe de congruence de $G(\mathbb{Q})$. Une telle variété, à priori définie comme espace analytique complexe, admet en fait une structure naturelle de variété quasi-projective définie sur un corps de nombres.

Les exemples fondamentaux de variétés de Shimura connexes sont les variétés modulaires de Siegel. La variété modulaire de Siegel, notée $\mathcal{A}_g$, est l'espace de modules des variétés abéliennes principalement polarisées de dimension g. Le groupe G associé est le groupe symplectique généralisé GSp_{2g} et la classe de conjugaison X^+ de paramètres de Hodge est isomorphe (comme domaine hermitien symétrique) au demi-espace de Siegel $\mathcal{H}_g$, c'est à dire l'ensemble des matrices complexes symétriques de taille $g \times g$ dont la partie imaginaire est définie positive. On peut obtenir $\mathcal{A}_g(\mathbb{C})$ comme le quotient de $\mathcal{H}_g$ par l'action de $\mathrm{Sp}_{2g}(\mathbb{Z})$.

Dans cette thèse on ne regarde que les variétés de Shimura de type abélien. Une variété de Shimura est dite **de type Hodge** si l'on peut l'interpréter comme un espace de modules de variétés abéliennes munies d'une structure supplémentaire. Une variété de Shimura est dite **de type abélien** si elle admet un revêtement par une variété de Shimura de type Hodge.

La conjecture de Manin–Mumford. La conjecture d'André–Oort et d'autres conjectures semblables sont inspirées par la conjecture de Manin–Mumford sur les variétés abéliennes, dont on rappelle l'énoncé. Cette conjecture a été démontrée par Raynaud. On en connaît aujourd'hui plusieurs démonstrations.

Théorème 1.1 (Conjecture de Manin–Mumford, [Ray83b]). *Soient A une variété abélienne et $Z \subset A$ une sous-variété irréductible. Soit Σ l'ensemble des points de torsion de A.*

Si $\Sigma \cap Z$ est dense dans Z pour la topologie de Zariski, alors Z est le translaté d'une sous-variété abélienne de A par un point de torsion.

Sous-variétés spéciales. Il y a une analogie entre les variétés abéliennes et les variétés de Shimura, due au fait que chacune est une variété quasi-projective qui peut être construite comme quotient d'une variété complexe simplement connexe, munie d'une action transitive d'un groupe de Lie réel, par un réseau dans le groupe de Lie. La définition d'une variété de Shimura va dans ce sens, à l'instar de la définition d'une variété abélienne complexe de dimension g, pouvant être construite comme quotient de l'espace affine $\mathbb{C}^g$ par un réseau dans le groupe vectoriel $\mathbb{R}^{2g}$. Remarquons que cette analogie n'a rien à voir avec le fait que certaines variétés de Shimura s'interprétent comme espaces de modules de variétés abéliennes.

Dans la conclusion de la conjecture de Manin–Mumford, on parle du translaté d'une sous-variété abélienne par un point de torsion. On peut également décrire une telle sous-variété comme une composante connexe d'une sous-variété algébrique de A. Les sous-variétés analogues d'une variété de Shimura s'appellent sous-variétés spéciales. Une sous-variété d'une variété de Shimura S est dite une **sous-variété spéciale** si c'est l'image de X_H^+ dans S pour une sous-donnée de Shimura connexe $(H, X_H^+) \subset (G, X^+)$.

On a ainsi le dictionnaire suivant :

Variété abélienne	Variété de Shimura
$\mathbb{R}^{2g}$ Groupe additif	G Groupe réductif
$\mathbb{C}^{2g}$ Espace affine	X^+ Domaine hermitien symétrique
$\Lambda \subset \mathbb{R}^{2g}$ Réseau	$\Gamma \subset G(\mathbb{Q})^+$ sous-groupe de congruence
Translaté d'une sous-variété abélienne par un point de torsion	Sous-variété spéciale

Il y a deux analogues possibles de la notion de point de torsion dans les variétés de Shimura, donc deux analogues différents de la conjecture de Manin–Mumford : la conjecture d'André–Oort et la conjecture d'André–Pink.

Points spéciaux. La conjecture d'André–Oort. Une première traduction de la notion de point de torsion pour une variété de Shimura est inspirée par le fait que, pour les variétés abéliennes, les points de torsion sont précisément les exemples de dimension zero de translatés de sous-variétés abéliennes par des points de torsion. On définit donc un **point spécial** dans une variété de Shimura comme une sous-variété spéciale de dimension zero. Il se trouve que les points spéciaux sont précisément les sous-variétés spéciales pour lesquelles le sous-groupe H est un tore. Dans le cas de la variété modulaire de Siegel, les points spéciaux sont les points correspondants aux variétés abéliennes avec multiplications complexes.

L'analogie entre points spéciaux et points de torsion est complétée par la constatation suivante : si A est une variété abélienne définie sur un corps de nombres, alors un point $x \in A(\mathbb{C})$ est de torsion si et seulement si x et une préimage quelconque de x dans le revêtement universel $\mathbb{C}^g$ sont tous les deux définis sur $\bar{\mathbb{Q}}$. De même, si S est une variété de Shimura de type abélien, alors un point $x \in S$ est spécial si et seulement si x et une préimage quelconque de x dans X^+ sont tous les deux définis sur $\bar{\mathbb{Q}}$ [Coh96, SW95].

Voici l'énoncé de la conjecture d'André–Oort.

Conjecture 1.2 ([And89] Chapitre X Problème 1, [Oor97] Conjecture 2)**.** *Soient S une variété de Shimura connexe et $Z \subset S$ une sous-variété irréductible. Soit Σ l'ensemble des points spéciaux de S.*

Si $\Sigma \cap Z$ est dense dans Z pour la topologie de Zariski, alors Z est une sous-variété spéciale de S.

La conjecture d'André–Oort a été démontrée par Klingler, Ullmo et Yafaev, en admettant l'hypothèse de Riemann généralisée (GRH) [KY13, UY13a]. Leur démonstration est basée sur des minorations des degrés galoisiens de sous-variétés spéciales, l'équidistribution de certaines suites de sous-variétés spéciales, et un critère géométrique concernant les sous-variétés spéciales contenues dans leur propre image par une correspondance de Hecke. GRH est utilisée deux fois : pour démontrer les bornes galoisiennes et pour assurer qu'il existe un petit nombre premier auquel on peut appliquer le critère géométrique.

Pila [Pil11] a employé une nouvelle méthode, due à Pila et Zannier, pour démontrer la conjecture d'André–Oort pour les produits de courbes modulaires sans GRH. Cette méthode se base sur le théorème de comptage de Pila–Wilkie sur les points rationnels d'ensembles définissables dans des modèles o-minimaux, des minorations galoisiennes pour les points spéciaux et un analogue du théorème d'Ax–Lindemann–Weierstrass pour les variétés de Shimura. Dans des travaux conséquents, Pila, Tsimerman, Ullmo et Yafaev ont développé cette méthode pour obtenir une démonstration inconditionelle de la conjecture d'André–Oort pour les variétés de Shimura qui se plongent dans $\mathcal{A}_6^n$ pour un certain n. On a encore besoin de GRH pour les bornes galoisiennes pour $\mathcal{A}_g$ avec $g > 6$, et il reste quelques détails techniques pour les variétés de Shimura qui ne se plongent pas dans une variété modulaire de Siegel.

Dans cette thèse on applique la méthode de Pila et Zannier à la conjecture d'André–Pink. Le principal nouvel ingrédient consiste en des bornes galoisiennes pour les points d'une orbite de Hecke.

Orbites de Hecke. La conjecture d'André–Pink. Les points d'une orbite de Hecke fournissent un deuxième analogue dans une variété de Shimura des points de torsion. Ici le principe de l'analogue est que les préimages dans $\mathbb{C}^g$ des points de torsion d'une variété abélienne forment une orbite pour le sous-groupe $\mathbb{Q}^{2g}$ de points rationnels dans $\mathbb{R}^{2g}$. Pour que ceci soit vrai, il faut faire attention au choix de l'action de $\mathbb{R}^{2g}$ sur $\mathbb{C}^g$: soit $\{e_1, \ldots, e_{2g}\} \subset \mathbb{C}^g$ une base pour le réseau Λ de périodes de la variété abélienne. Alors l'action correcte de $\mathbb{R}^{2g}$ sur $\mathbb{C}^g$ est donnée par

$$(x_1, \ldots, x_{2g}).v = v + \sum_{i=1}^{2g} x_i e_i \text{ où } (x_1, \ldots, x_{2g}) \in \mathbb{R}^{2g}, v \in \mathbb{C}^g.$$

Appelons une **orbite de torsion** l'image dans A d'une $\mathbb{Q}^{2g}$-orbite dans $\mathbb{C}^g$. Une telle orbite est de la forme $x + A_{\text{tors}}$ pour un point $x \in A$.

Soit S une variété de Shimura connexe associée à la donnée de Shimura connexe (G, X^+). On définit une **orbite de Hecke** comme l'image dans S d'une $G(\mathbb{Q})_+$-orbite dans X^+, où $G(\mathbb{Q})_+$ désigne le groupe d'éléments de $G(\mathbb{Q})$ qui envoient X^+ dans lui-même. Dans la variété modulaire de Siegel $\mathcal{A}_g$, deux points sont dans la même orbite de Hecke si et seulement si les variétés abéliennes principalement polarisées correspondantes (A, λ) et (B, μ) sont liées par une **isogénie polarisée**, c'est à dire une isogénie $f \colon A \to B$ telle que $f^*\mu \in \mathbb{Z}.\lambda$.

Remarquons que si, dans l'énoncé de la conjecture de Manin–Mumford, Σ est une orbite de torsion quelconque au lieu de l'ensemble de points de torsion, alors on peut conclure que Z est un translaté d'une sous-variété abélienne (pas nécessairement un translaté par un point de torsion). Cet énoncé modifié est équivalent à l'énoncé d'origine parce qu'on peut toujours translater Z et Σ de sorte que Σ contienne l'origine.

On appelle sous-variétés faiblement spéciales les sous-variétés d'une variété de Shimura analogues aux translatés arbitraires de sous-variétés abéliennes. On pourrait imaginer qu'une sous-variété faiblement spéciale de S serait une sous-variété qui est l'image de $g.X_H^+$ pour une sous-donnée de Shimura connexe quelconque $(H, X_H^+) \subset (G, X)$ et un élément quelconque $g \in G(\mathbb{R})^+$. Cependant en général, $g.X_H^+$ n'est pas invariant sous $H(\mathbb{R})^+$ et l'image de $g.X_H^+$ n'est pas une sous-variété algébrique de S. On évite ces problèmes en imposant la condition supplémentaire que g normalise H^{der}.

On est donc amené à la définition suivante : une sous-variété de S est dite **faiblement spéciale** si c'est l'image dans S de $g.X_H^+ \subset X^+$ pour une sous-donnée de Shimura $(H, X_H^+) \subset (G, X^+)$ et un élément $g \in G(\mathbb{R})^+$ normalisant H^{der}. On peut également décrire les sous-variétés faiblement spéciales géométriquement comme les images dans S de sous-variétés totalement géodésiques de X^+.

Si l'on prend les sous-variétés faiblement spéciales comme analogues d'orbites de torsion, on obtient l'analogue suivant de la conjecture de Manin–Mumford, qu'on appelle la conjecture d'André–Pink. Comme Pink, on énonce la conjecture pour les orbites de Hecke généralisées, une légère généralisation des orbites de Hecke. La différence entre les orbites de Hecke usuelles et généralisées est discutée en détail dans la thèse, mais ce n'est pas important dans cette introduction.

Conjecture 1.3 ([And89] Chapitre X Problème 3, [Pin05a] Conjecture 1.6)**.** *Soient S une variété de Shimura connexe et $Z \subset S$ une sous-variété irréductible. Soit s un point de S, et soit Σ l'orbite de Hecke généralisée de s.*

Si $\Sigma \cap Z$ est dense dans Z pour la topologie de Zariski, alors Z est une sous-variété faiblement spéciale de S.

Pink a énoncé la conjecture pour les variétés de Shimura mixtes, où l'on se permet des groupes G non-réductifs, X^+ étant un espace de modules de structures de Hodge mixtes au lieu de structures de Hodge pures. La version pour les variétés de Shimura mixtes implique les conjectures de Manin–Mumford et Mordell–Lang. (L'énoncé d'André permet aussi des groupes G non-réductifs mais d'une manière différente, qui implique la conjecture de Manin–Mumford mais pas celle de Mordell–Lang.) Contrairement au cas de variétés de Shimura pures, la différence entre orbites de Hecke usuelles et généralisées est très importante pour la version mixte de la conjecture. Il paraît peu probable que les méthodes de cette thèse puissent démontrer la conjecture de Pink pour les variétés de Shimura mixtes.

La reciproque de la conjecture d'André–Pink est vraie dans le sens suivant : si Z est une sous-variété faiblement spéciale d'une variété de Shimura S, s est un point de Z et Σ est l'orbite de Hecke (usuelle ou généralisée) de s, alors $\Sigma \cap Z$ est dense dans Z pour la topologie de Zariski, et même pour la topologie complexe.

La conjecture de Zilber–Pink. La conjecture de Zilber–Pink est une généralisation de la conjecture d'André–Oort qui implique également la conjecture d'André–Pink. Dans sa version la plus générale, pour les variétés de Shimura mixte, elle implique aussi d'autres conjectures comme des versions relatives de la conjecture de Manin–Mumford et la Conjecture sur les Intersections de Tores de Zilber [Zil02].

Conjecture 1.4 (Conjecture de Zilber–Pink, [Pin05b] Conjecture 1.1)**.** *Soient S une variété de Shimura connexe et $Z \subset S$ une sous-variété fermée. Soit d un entier positif, et soit Ξ la réunion de toutes les sous-variétés spéciales de S de dimension au plus d.*

Si $\Xi \cap Z$ est dense dans Z pour la topologie de Zariski, alors Z est contenu dans une sous-variété spéciale S_Z de S telle que $\dim S_Z \leq \dim Z + d$.

La conjecture d'André–Oort est le cas $d = 0$ de la conjecture de Zilber–Pink. La démonstration que la conjecture de Zilber–Pink implique celle d'André–Pink est plus compliquée. Cette démonstration est due à Pink ([Pin05b] Théorème 3.3).

Les cas connus de la conjecture d'André–Pink. Certains cas restreints de la conjecture d'André–Pink sont déjà connus. Edixhoven et Yafaev [EY03] ont démontré l'intersection des conjectures d'André–Oort et d'André–Pink pour les courbes, c'est à dire la conjecture d'André–Pink sous les hypothèses que s soit un point spécial et Z soit une courbe. Leur démonstration reposait sur des minorations des degrés galoisiens de points spéciaux et un critère géométrique pour démontrer qu'une sous-variété est spécial. C'est la méthode qui a été développée pour la démonstration de la conjecture d'André–Oort sous GRH par Klingler, Ullmo et Yafaev.

Klingler et Yafaev [KY13] ont démontré l'intersection des conjectures d'André–Oort et d'André–Pink pour les sous-variétés de dimension arbitraire : leur démonstration de la conjecture d'André–Oort ne dépend pas de GRH dans le cas où tous les points spéciaux impliqués se trouvent dans la même orbite de Hecke. Edixhoven–Yafaev et Klingler–Yafaev regardaient en fait des classes de points un peu plus générales que les orbites de Hecke, ce qu'on appelle les classes de ρ-isogénie dans la thèse.

Pink ([Pin05a] Théorème 7.6) a démontré la conjecture pour les points s qui sont Galois génériques dans la variété modulaire de Siegel $\mathcal{A}_g$, c'est à dire les points dont l'image de la répresentation galoisienne associée est ouverte dans $\mathrm{GSp}_{2g}(\hat{\mathbb{Z}})$. Cette démonstration employait un résultat de Clozel, Oh et Ullmo sur l'équidistribution des orbites de Hecke.

Habegger et Pila [HP12] ont utilisé la méthode de Pila–Zannier pour démontrer la conjecture pour $\mathcal{A}_1^n$. Ils employaient une borne pour les isogénies de courbes elliptiques due à Pellarin pour obtenir les bornes galoisiennes.

On n'a pas pu étendre les stratégies d'Edixhoven–Yafaev et Pink au cas d'orbites de Hecke dont les points ne sont ni spéciaux ni Galois génériques. Dans la thèse, on emploie la stratégie de Habegger et Pila.

Les résultats principaux de la thèse. On démontre les cas suivants de la conjecture d'André–Pink pour les variétés de Shimura de type abélien.

Théorème 1.5. *La conjecture 1.3 est vérifiée si S est de type abélien et :*

(i) *soit Z est une courbe,*

(ii) *soit la plus petite sous-variété spéciale de S contenant s est égale à la plus petite sous-variété spéciale de S contenant Z.*

On démontre aussi le théorème suivant, qui est la conjecture d'André–Pink pour les variétés de Shimura de type abélien avec la phrase «orbite de Hecke généralisée» remplacée par «P-orbite de Hecke» pour un ensemble fini P de nombres premiers. La définition précise de P-orbite de Hecke se trouve dans le paragraphe 4.5 de la thèse. Pour en donner un exemple, deux points de la variété modulaire de Siegel sont dans la même P-orbite de Hecke si et seulement si les variété abéliennes principalement polarisées correspondantes sont liées par une isogénie polarisée f telle que P contient tous les facteurs premiers de $\deg f$.

Théorème 1.6. *Soient S une variété de Shimura connexe et $Z \subset S$ une sous-variété irréductible. Soit s un point de S et soit Σ_P la P-orbite de Hecke de s, où P est un ensemble fini de nombres premiers.*

Si $\Sigma_P \cap Z$ est Zariski dense dans Z, alors Z est une sous-variété faiblement spéciale de S.

Les orbites de Hecke et les classes d'isogénie. On définit une **classe d'isogénie** dans $\mathcal{A}_g$ comme l'ensemble des points correspondants aux variétés abéliennes isogènes à une variété abélienne fixe, sans tenir compte des polarisations. Puisque les points d'une orbite de Hecke correspondent aux variétés abéliennes liées par des isogénies polarisées, chaque orbite de Hecke de $\mathcal{A}_g$ est contenue dans une classe d'isogénie. Le rapport entre les orbites de Hecke et les classes d'isogénie et leur différentes propriétés de fonctorialité sont très importants dans la démonstration du théorème 1.5.

Si s est un point Hodge générique dans $\mathcal{A}_g$ (c'est à dire un point qui n'est contenu dans aucune sous-variété spéciale propre), alors sa classe d'isogénie et son orbite de Hecke coïncident. Cependant ce n'est pas vrai en général. En fait il existe des classes d'isogénie dans $\mathcal{A}_g$ contenant une infinité d'orbites de Hecke (le lemme 4.3).

En revanche, chaque classe d'isogénie de $\mathcal{A}_g$ est contenue dans une orbite de Hecke de $\mathcal{A}_{4g}$, via le plongement naturel $\mathcal{A}_g \to \mathcal{A}_{4g}$ qui envoie une variété abélienne principalement polarisée sur sa quatrième puissance. Ceci est un résultat non-trivial : c'est la proposition 4.4 de la thèse, dont la démonstration dépend de la classification des formes hermitiennes sur l'algèbre d'endomorphismes d'une variété abélienne. Donc la conjecture suivante serait un corollaire de la conjecture d'André–Pink pour $\mathcal{A}_{4g}$, mais pas de la conjecture d'André–Pink uniquement pour $\mathcal{A}_g$.

Conjecture 1.7. *Soit Z une sous-variété irréductible de $\mathcal{A}_g$. Soient s un point de $\mathcal{A}_g$ et Σ la classe d'isogénie de s.*

Si $\Sigma \cap Z$ est Zariski dense dans Z, alors Z est une sous-variété faiblement spéciale de $\mathcal{A}_g$.

La définition d'une classe d'isogénie qu'on vient d'énoncer ne s'applique qu'à la variété modulaire de Siegel. Cependant on peut définir des notions similaires dans une variété de Shimura quelconque, qui dépendent du choix d'une représentation rationelle du groupe réductif G. Une telle représentation ρ associe une $\mathbb{Q}$-structure de Hodge à chaque point de la variété de Shimura connexe S. On définit une **classe de ρ-isogénie** dans S comme l'ensemble des points dont les $\mathbb{Q}$-structures de Hodge associées sont membres d'une classe d'isomorphisme fixe. Cette notion est surtout intéressante quand la représentation ρ est fidèle. Par exemple, les classes d'isogénie simples de $\mathcal{A}_g$ sont égales aux classes de ρ-isogénie par rapport à la représentation standard de dimension $2g$ de GSp_{2g}.

Pour toute représentation ρ de G, chaque orbite de Hecke de S est contenue dans une classe de ρ-isogénie. Comme le montre le cas de $\mathcal{A}_g$ et la représentation standard, une classe de ρ-isogénie n'est pas forcément contenue dans une orbite de Hecke. L'exemple du plongement $\mathcal{A}_g \to \mathcal{A}_{4g}$ suggère la question ouverte suivante, qui paraît difficile.

Question 1.8. *Soient S une variété de Shimura connexe et ρ une $\mathbb{Q}$-représentation fidèle du groupe réductif sous-jacent G. Soit Σ une classe de ρ-isogénie de S.*

Existe-t-il toujours une variété de Shimura S' et un plongement de Shimura (ou une immersion de Shimura) $f\colon S \to S'$ tels que $f(\Sigma)$ est contenu dans une orbite de Hecke généralisée de S' ?

Expliquons l'obstacle à une démonstration entière de la conjecture d'André–Pink pour les variétés de Shimura de type abélien. Soit $[\iota]\colon S' \to S$ un plongement de Shimura, c'est à dire un morphisme de variétés de Shimura induit par une injection $\iota\colon G' \to G$ des groupes sous-jacents. Les exemples qu'on vient de discuter montrent que si Σ est une orbite de Hecke de S, alors $[\iota]^{-1}(\Sigma)$ n'est pas forcément contenu dans une réunion finie d'orbites de Hecke (même généralisées) de S'. Par conséquent on ne peut pas démontrer la conjecture d'André–Pink en remplaçant la variété de Shimura S par sa plus petite sous-variété spéciale contenant Z ; ce serait la première étape de la démonstration naturelle par récurrence sur la dimension.

Ce problème de la préimage sous un plongement de Shimura disparaît si on travaille avec les classes de ρ-isogénie au lieu des orbites de Hecke. Si ρ est une représentation fidèle de G est Σ est une classe de ρ-isogénie de S, alors $[\iota]^{-1}(\Sigma)$ est évidemment contenu dans une classe de $(\rho \circ \iota)$-isogénie de S'. Donc une réponse positive à la question 1.8, avec les résultats du chapitre 6 de la thèse, impliquerait la conjecture d'André–Pink pour les variétés de Shimura de type abélien.

Les résultats supplémentaire de la thèse : des bornes pour les isogénies. Un ingrédient clé dans la démonstration du théorème 1.5 est le théorème d'isogénie de Masser et Wüstholz [MW93a] : si K est un corps de nombres et A une variété abélienne principalement polarisée définie sur K, alors pour toute autre variété abélienne B définie sur un corps de nombres $L \supset K$ et isogène à A, il existe une isogénie $A \to B$ dont le degré est majoré par un polynôme dans $[L : K]$.

Ce théorème ne dit rien sur la compatibilité des isogénies concernées avec les polarisations. En particulier, même si l'on suppose que A et B sont dans la même classe d'isogénie polarisée, l'isogénie de petit degré dont l'existence est affirmée par le théorème n'est pas forcément une isogénie polarisée. Ceci rend plus compliquée la démonstration de la proposition 6.5.

Le théorème suivant résout ce problème : il affirme que si A et B sont dans la même classe d'isogénie polarisée et il y a une isogénie $A \to B$ de degré donné, alors il existe une isogénie polarisée de degré borné polynomialement. On aurait pu appliquer ce théorème dans la démonstration du théorème 1.5, mais la longueur de la démonstration du théorème 1.9 l'emporte sur la simplification de la démonstration de la proposition 6.5 qui en resultérait. On a néanmoins inclu le théorème dans la thèse au cas où il aurait des applications indépendantes.

Théorème 1.9. *Soit* (A, λ) *une variété abélienne principalement polarisée définie sur un corps de caractéristique* 0*. Il existe des constantes* c*,* k *ne dépendant que de* (A, λ) *telles que, si* (B, λ') *est une variété principalement polarisée et*

1. *il existe une isogénie* $f\colon A \to B$ *compatible avec les polarisations (de degré quelconque), et*

2. *il existe une isogénie* $g\colon A \to B$ *de degré* n *(pas forcément compatible avec les polarisations),*

alors il existe une isogénie $h\colon A \to B$ *compatible avec les polarisations et dont le degré est majoré par* cn^k*.*

En ce qui concerne les isogénies, on a d'ailleurs le théorème 5.2 qui étend le théorème de Masser–Wüstholz des corps de nombres aux corps de type fini de caractéristique 0. Ce résultat est employé dans la démonstration du théorème 1.5 : il nous permet de démontrer le théorème pour l'orbite de Hecke d'un point $s \in S(\mathbb{C})$ et pas seulement $s \in S(\bar{\mathbb{Q}})$.

Les résultats supplémentaires de la thèse : les groupes de Mumford–Tate. Le chapitre 7 est indépendant des autres résultats de la thèse. Il concerne les rangs des groupes de Mumford–Tate de variétés abéliennes.

Soit A une variété abélienne principalement polarisée de dimension g définie sur $\mathbb{C}$. Via la structure de Hodge de $H^1(A, \mathbb{C})$, on peut associer à A un morphisme de groupes algébriques $h_A\colon \mathrm{Res}_{\mathbb{C}/\mathbb{R}}\, \mathbb{G}_m \to \mathrm{GSp}_{2g,\mathbb{R}}$. Ce morphisme est défini à conjugaison par $\mathrm{Sp}_{2g}(\mathbb{Z})$ près, parce qu'on a un choix d'une base symplectique de $H^1(A, \mathbb{Z})$. Ceci identifie l'espace de modules $\mathcal{A}_g$ au quotient d'une classe de conjugaison de paramètres de Hodge de GSp_{2g} par l'action de $\mathrm{Sp}_{2g}(\mathbb{Z})$, comme on a expliqué ci-dessus.

On définit le **groupe de Mumford–Tate** de A comme le plus petit sous-groupe de GSp_{2g} défini sur $\mathbb{Q}$ et contenant l'image de h_A.

Il y a une liaison étroite entre les groupes de Mumford–Tate et les sous-variétés spéciales. À chaque sous-variété spéciale S de $\mathcal{A}_g$, on peut associer un groupe de Mumford–Tate générique H. Si un point de S n'est pas contenu dans une sous-variété spéciale strictement plus petite que S, alors le groupe de Mumford–Tate de la variété abélienne associée est égale à H. Les points des sous-variétés spéciales strictement contenues dans S ont des groupes de Mumford–Tate strictement contenus dans H.

Dans la thèse, on obtient une minoration du rang du groupe de Mumford–Tate d'une variété abélienne, sous l'hypothèse que ses sous-variétés abéliennes simples soient deux à deux non-isogènes. Une telle condition sur les sous-variétés simples est inévitable pour obtenir une minoration du rang qui croît avec la dimension de la variété abélienne, parce que le groupe de Mumford–Tate de A^n est isomorphe à celui de A pour tout entier $n \geq 1$.

Théorème 1.10. *Soit A une variété abélienne de dimension g dont les sous-variétés abéliennes simples sont deux à deux non-isogènes. Soit n le rang du groupe de Mumford–Tate de A.*

Alors

$$n + \alpha(n)\sqrt{n \log_e n} \geq \log_2 g + 2$$

pour une fonction $\alpha\colon \mathbb{N}_{\geq 2} \to \mathbb{R}$ (indépendante de A et g) vérifiant $\alpha(n) < 2$ pour tout n et $\alpha(n) \to 1/\log_e 2 = 1.44\ldots$ quand $n \to \infty$.

Si l'on suppose que l'anneau d'endomorphismes de A est commutatif, alors $n \geq \log_2 g + 2$.

Ce théorème étend une minoration démontrée par Ribet [Rib81] dans le cas de variétés abéliennes avec multiplications complexes.

D'ailleurs, on démontre que la même minoration vaut pour le rang du groupe de monodromie ℓ-adique d'une variété abélienne définie sur un corps de nombres. Ceci entraine une minoration du degré du corps de définition des points de ℓ^n-torsion de la variété abélienne.

Remarques sur la démonstration du théorème 1.5. Pour démontrer le théorème 1.5, on commence par le cas où $S = \mathcal{A}_g$ et Z est une courbe. Dans ce cas, on suit la méthode de Pila et Zannier. Cette méthode consiste en l'étude de la préimage de Z sous une certaine fonction $\alpha\colon U \to \mathcal{A}_g$ définissable dans un ensemble o-minimal (U étant un sous-ensemble définissable de $\mathbb{R}^n$). On choisit la fonction α de sorte que tout point dans l'orbite de Hecke Σ admet une préimage rationelle dans U.

Dans notre cas, on prend U comme un certain sous-ensemble de $\mathrm{GL}_{2g}(\mathbb{R})$. Un des points subtils de la démonstration est le fait qu'on ne peut pas se contenter de prendre U contenu dans $\mathrm{GSp}_{2g}(\mathbb{R})$, grâce à la différence entre les classes d'isogénie et les orbites de Hecke et le fait que le théorème d'isogénie de Masser–Wüstholz ne s'occupe pas des compatibilités des isogénies avec la polarisation. Cependant on définit la fonction $\alpha\colon U \to \mathcal{A}_g$ de manière à ce que sa restriction à $U \to \mathcal{A}_g$ soit pareille à l'action de $\mathrm{GSp}_{2g}(\mathbb{R})^+$ sur un point convenable dans $\mathcal{H}_g$.

Il y a deux autres nouveaux ingrédients de la démonstration : des majorations des hauteurs des représentations rationnelles d'isogénies et l'application du théorème de Masser–Wüstholz pour obtenir une minoration des degrés galoisiens des points de l'orbite de Hecke Σ. Les majorations des hauteurs d'isogénies repose sur des calculs sur les formes hermitiennes sur les algèbres d'endomorphismes de variétés abéliennes, une thème qu'on trouve aussi dans les résultats supplémentaires sur les isogénies polarisées et non-polarisées.

Une fois que les ingrédients ci-dessus sont préparés, on termine la démonstration du théorème 1.5 en appliquant une variante forte du théorème de comptage de Pila–Wilkie ([PW06], [Pil11] Théorème 3.6) concernant les points rationnels d'ensembles définissables ainsi que la caractérisation (due à Ullmo et Yafaev [UY11]) de sous-variétés faiblement spéciales comme sous-variétés algébriques de $\mathcal{A}_g$ dont la préimage dans $\mathcal{H}_g$ a une composante irréductible algébrique.

Pour étendre le résultat de $\mathcal{A}_g$ aux autres variétés de Shimura de type abélien, on démontre que toute variété de Shimura S de type abélien admet un revêtement fini par une autre variété de Shimura S_1 qui se plonge dans $\mathcal{A}_g$ de telle façon que la préimage dans S_1 de chaque orbite de Hecke de S soit contenue dans une réunion finie d'orbites de Hecke de $\mathcal{A}_g$. Ceci est un raffinement d'un résultat de Deligne ([Del79] Proposition 2.3.10) qui fait partie de sa classification des variétés de Shimura de type abélien. Pour obtenir un résultat sur les orbites de Hecke, il faut améliorer le raisonnement de Deligne pour prendre en compte les centres des groupes réductifs concernés.

Enfin la démonstration de théorème 1.5(ii), concernant les sous-variétés de dimension supérieure, passe par la récurrence sur la dimension. Dans cette partie de la démonstation, au lieu de la caractérisation de sous-variétés faiblement spéciales d'Ullmo et Yafaev, il faut utiliser la conjecture d'Ax–Lindemann–Weierstrass hyperbolique pour $\mathcal{A}_g$ dont la démonstration a été annoncée par Pila et Tsimerman [PT12].

Structure de la thèse. Le chapitre 2 contient un résumé de la théorie des variétés de Shimura et des sous-variétés spéciales et faiblement spéciales, se concentrant sur les aspects traités dans la thèse. Il contient principalement des définitions et des résultats connus. Il contient parfois des démonstrations qui ne se trouvent pas das la litérature.

Dans le chapitre 3 on esquisse la stratégie de Pila et Zannier, rappelant ses ingrédients principaux. On explique la démonstation de la conjecture de Manin–

Mumford pour les tores qui est l'application la plus facile de la stratégie. Il s'agit exclusivement des travaux connus.

Le chapitre 4 concerne les orbites de Hecke et leurs variantes, comme les orbites de Hecke généralisées, les P-orbites de Hecke et les classes d'isogénie. On rappelle les définitions de ces différentes notions et les rapports entre eux, et on démontre quelques résultats sur leurs propriétés de fonctorialité par rapport aux morphismes de Shimura. Ces propriétés de fonctorialité nous permettent de réduire certains cas de la conjecture d'André–Pink à des cas plus simples. La définition et les propriétés élémentaires sont dues à Pink, mais la plupart du chapitre est originale.

Le chapitre 5 comporte plusieurs bornes concernant les isogénies de variétés abéliennes, essentielles à la stratégie de la démonstration de la conjecture d'André–Pink pour les courbes dans $\mathcal{A}_g$. On démontre également une borne pour les isogénies polarisées qui n'est pas utilisée dans cette thèse, mais qui pourrait avoir des applications indépendantes. Ce chapitre contient exclusivement des travaux originaux.

Le chapitre 6 contient les démonstrations des théorèmes principaux sur la conjecture d'André–Oort (les théorèmes 1.5 et 1.6). Ça se fait en combinant les ingrédients des chapitres 3, 4 et 5.

Dans le chapitre 7 on démontre des minorations des rangs des groupes de Mumford–Tate de variétés abéliennes. Ce chapitre est une réproduction d'un article soumis pour publication. Il est indépendant des autres chapitres.

1 Introduction (English)

The primary topic of this thesis is a conjecture of André and Pink asserting that a subvariety of a Shimura variety which has dense intersection with a Hecke orbit is weakly special. This is one of several conjectures on subvarieties of mixed Shimura varieties, all of which are consequences of the Zilber–Pink conjecture. So far, progress on the pure Shimura variety cases of these conjectures has either concerned special points (the André–Oort conjecture) or been restricted to products of modular curves, so this thesis goes in a new direction by considering Hecke orbits in Shimura varieties of abelian type.

Shimura varieties. We must begin by sketching the definition of a connected Shimura variety (for more details, see sections 2.5 and 2.6). Let G be a connected reductive group over $\mathbb{Q}$. Let X^+ be a connected component of a $G(\mathbb{R})$-conjugacy class of **Hodge parameters**, that is, homomorphisms $\operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m \to G_{\mathbb{R}}$. The pair (G, X^+) is called a **connected Shimura datum** if it satisfies the conditions 2.1.1.1–2.1.1.3 of [Del79]. In particular, these conditions ensure that X^+ is a Hermitian symmetric domain. A **connected Shimura variety** is a quotient of X^+ by a congruence subgroup of $G(\mathbb{Q})$ which stabilises X^+. Such a variety, a priori defined as a complex analytic space, in fact has a structure of quasi-projective algebraic variety defined over a number field.

The fundamental examples of connected Shimura varieties are the Siegel modular varieties. The Siegel modular variety $\mathcal{A}_g$ is the moduli space of principally polarised abelian varieties of dimension g. The associated group G is the general symplectic group GSp_{2g} and the conjugacy class of Hodge parameters X^+ is isomorphic (as a Hermitian symmetric domain) to the Siegel upper half space $\mathcal{H}_g$ consisting of symmetric $g \times g$ complex matrices whose imaginary part is positive definite. We can obtain $\mathcal{A}_g(\mathbb{C})$ as the quotient of $\mathcal{H}_g$ by the action of $\mathrm{Sp}_{2g}(\mathbb{Z})$.

In this thesis we shall consider only Shimura varieties of abelian type. A Shimura variety is said to be **of Hodge type** if it can be interpreted as a moduli space of abelian varieties with additional structure. A Shimura variety is **of abelian type** if it has a covering by a Shimura variety of Hodge type. Coverings of Shimura varieties do not change the (weakly) special subvarieties but do change the Hecke orbits.

The Manin–Mumford conjecture. The inspiration for conjectures such as the André–Pink conjecture was the Manin–Mumford conjecture on abelian varieties. This conjecture was first proved by Raynaud. Many different proofs are now known.

Theorem 1.1 (Manin–Mumford Conjecture, [Ray83b]). *Let A be an abelian variety and $Z \subset A$ an irreducible subvariety. Let Σ denote the set of torsion points in A.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a translate of an abelian subvariety of A by a torsion point.

Special subvarieties. There is an analogy between abelian varieties and Shimura varieties because each is a quasi-projective variety which can be constructed as the quotient of a simply connected complex manifold with a transitive action of a real Lie group by a lattice in that Lie group. Shimura varieties are defined in this way, using a reductive group, while a complex abelian variety of dimension g can be constructed as a quotient of the affine space $\mathbb{C}^g$ by a lattice in the vector group $\mathbb{R}^{2g}$. Note that this analogy has nothing to do with the fact that some Shimura varieties can be interpreted as moduli spaces for abelian varieties.

The Manin–Mumford conjecture refers to translates of abelian subvarieties of A by torsion points. Such a subvariety can also be described as a connected component of an algebraic subgroup of A. The analogue in a Shimura variety S is a special subvariety. A subvariety of S is a **special subvariety** if it is the image in S of X_H^+ for some connected Shimura subdatum $(H, X_H^+) \subset (G, X^+)$.

We thus have the following dictionary:

Abelian variety	Shimura variety
Additive group $\mathbb{R}^{2g}$	Reductive group G
Affine space $\mathbb{C}^g$	Hermitian symmetric domain X^+
Lattice $\Lambda \subset \mathbb{R}^{2g}$	Congruence subgroup $\Gamma \subset G(\mathbb{Q})^+$
Translate of an abelian subvariety by a torsion point	Special subvariety

There are two possible analogues of torsion points in Shimura varieties, which lead to two different analogues of the Manin–Mumford conjecture: the André–Oort conjecture and the André–Pink conjecture.

Special points. The André–Oort conjecture. The first analogue of torsion points in Shimura varieties focuses on the fact that torsion points are precisely the zero-dimensional examples of translates of abelian subvarieties by torsion points. We therefore define a **special point** in a Shimura variety to be a zero-dimensional special subvariety. It turns out that these are precisely the special subvarieties for which the group H is a torus. In the case of $\mathcal{A}_g$, special points are those which correspond to abelian varieties with complex multiplication.

A second justification for this analogy between torsion points and special points is that, if A is an abelian variety defined over a number field, then $x \in A(\mathbb{C})$ is torsion if and only if both x and any preimage of x in the universal cover $\mathbb{C}^g$ are defined over $\bar{\mathbb{Q}}$. Similarly if S is a Shimura variety of abelian type, then a point $x \in S$ is special if and only if both x and any of its preimages in X^+ are defined over $\bar{\mathbb{Q}}$ ([Coh96], [SW95]).

Conjecture 1.2 ([And89] Chapter X Problem 1, [Oor97] Conjecture 2)**.** *Let S be a connected Shimura variety and $Z \subset S$ an irreducible subvariety. Let Σ denote the set of special points in S.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a special subvariety of S.

The André–Oort conjecture has been proved by Klingler, Ullmo and Yafaev, assuming the Generalised Riemann Hypothesis [KY13, UY13a]. Their proof relies on lower bounds for Galois degrees of special subvarieties, equidistribution of certain sequences of special subvarieties, and a geometric criterion concerning subvarieties which are contained in their own image under a Hecke correspondence. The Generalised Riemann Hypothesis is needed both to prove the Galois bounds and in order to ensure that there exists a small prime at which we can apply the geometric criterion.

A new method, due to Pila and Zannier, was used by Pila [Pil11] to prove the André–Oort conjecture unconditionally for products of modular curves. This method uses the Pila–Wilkie counting theorem on rational points in sets definable in o-minimal models, Galois lower bounds for special points and an analogue of the Ax–Lindemann–Weierstrass theorem for Shimura varieties. Subsequent work by Pila, Tsimerman, Ullmo and Yafaev has extended this method to give an unconditional proof of the André–Oort conjecture for Shimura varieties which can be embedded in $\mathcal{A}_6^n$ for some n. GRH is still needed to obtain the Galois bounds for $\mathcal{A}_g$ with $g > 6$ and some technical issues remain for Shimura varieties which cannot be embedded in a Siegel modular variety.

It is the method of Pila and Zannier which we will apply to the André–Pink conjecture in this thesis. The primary new ingredient required is Galois bounds for Hecke orbits.

Hecke orbits. The André–Pink conjecture. Hecke orbits in Shimura varieties are defined by analogy with the fact that the preimages in $\mathbb{C}^g$ of torsion points in abelian varieties form an orbit for the subgroup $\mathbb{Q}^{2g}$ of rational points in $\mathbb{R}^{2g}$. In order for this to be true, we need to be careful about the choice of action of $\mathbb{R}^{2g}$ on $\mathbb{C}^g$ when setting up the dictionary between abelian varieties and Shimura varieties: let $\{e_1, \dots, e_{2g}\} \subset \mathbb{C}^g$ be a basis for the period lattice Λ of our abelian variety A. Then the correct action of $\mathbb{R}^{2g}$ on $\mathbb{C}^g$ is defined by

$$(x_1, \dots, x_{2g}).v = v + \sum_{i=1}^{2g} x_i e_i \text{ for } (x_1, \dots, x_{2g}) \in \mathbb{R}^{2g}, v \in \mathbb{C}^g.$$

Let us call the image in A of a $\mathbb{Q}^{2g}$-orbit in $\mathbb{C}^g$ a **torsion orbit**. Such an orbit has the form $x + A_{\mathrm{tors}}$ for some point $x \in A$.

Let S be a connected Shimura variety associated with the connected Shimura datum (G, X^+). We define a **Hecke orbit** to be the image in S of a $G(\mathbb{Q})_+$-orbit in X^+ (where $G(\mathbb{Q})_+$ is the group of points in $G(\mathbb{Q})$ which map X^+ into itself). In the Siegel modular variety $\mathcal{A}_g$, two points are in the same Hecke orbit if and only if the corresponding principally polarised abelian varieties (A, λ), (B, μ) are related by a **polarised isogeny**, that is, an isogeny $f \colon A \to B$ such that $f^*\mu \in \mathbb{Z}.\lambda$.

Note that if, in the statement of the Manin–Mumford conjecture, we allow Σ to be any torsion orbit instead of requiring it to be the set of torsion points, then we may conclude that Z is some translate of an abelian subvariety (not necessarily by a torsion point). This is equivalent to the original statement of the Manin–Mumford conjecture because we can always apply a translation so that our torsion orbit contains the origin.

The analogues for Shimura varieties of arbitrary translates of abelian subvarieties are called weakly special subvarieties. One might think that a weakly special subvariety should be a subvariety of S which is the image of $g.X_H^+$ for any connected Shimura subdatum $(H, X_H^+) \subset (G, X^+)$ and any $g \in G(\mathbb{R})^+$. However usually $H(\mathbb{R})^+$ does not stabilise $g.X_H^+$ and the image of $g.X_H^+$ is not an algebraic subvariety of S. These problems are avoided if g normalises H^{der}.

We therefore make the following definition: a subvariety of S is a **weakly special subvariety** if it is the image in S of $g.X_H^+ \subset X^+$ for some connected Shimura subdatum $(H, X_H^+) \subset (G, X^+)$ and some $g \in G(\mathbb{R})^+$ normalising H^{der}. Geometrically, these can also be described as the images in S of totally geodesic subvarieties of X^+ (see section 2.10 for equivalences between various definitions).

Taking Hecke orbits as analogous to torsion orbits, we get the following analogue for the Manin–Mumford conjecture, which we call the André–Pink conjecture. Following Pink we have stated the conjecture for generalised Hecke orbits, which are a slight generalisation of Hecke orbits. The difference between generalised and usual Hecke orbits is unimportant for this introduction, and indeed when the conjecture is equivalent with or without the word "generalised", when it is applied to all pure connected Shimura varieties or all connected Shimura varieties of abelian type simultaneously.

Conjecture 1.3 ([And89] Chapter X Problem 3, [Pin05a] Conjecture 1.6)**.** *Let S be a connected Shimura variety and $Z \subset S$ an irreducible subvariety. Let s be a point in S and let Σ be the generalised Hecke orbit of s.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of S.

Pink stated this conjecture for mixed Shimura varieties, in which the underlying group G is permitted to be non-reductive and X^+ is a parameter space for mixed Hodge structures instead of pure Hodge structures. The mixed version implies the Manin–Mumford and Mordell–Lang conjectures. (André also allowed non-reductive groups G but in a different way, which implies Manin–Mumford but not Mordell–Lang.) Unlike in the pure case, the difference between generalised and usual Hecke orbits is very important for the mixed version of the conjecture. It seems unlikely that the methods of this thesis could be used to prove Pink's conjecture for mixed Shimura varieties.

The converse to the André–Pink conjecture is true in the following sense: if Z is a weakly special subvariety of the Shimura variety S, s is a point in Z and Σ is the (generalised or usual) Hecke orbit of s, then $\Sigma \cap Z$ is Zariski dense in Z.

The Zilber–Pink conjecture. The Zilber–Pink conjecture is a generalisation of the André–Oort conjecture which also implies the André–Pink conjecture. In its most general version, for mixed Shimura varieties, it has further implications such as relative versions of the Manin–Mumford conjecture and Zilber's Conjecture on Intersections of Tori [Zil02].

Conjecture 1.4 (Zilber–Pink Conjecture, [Pin05b] Conjecture 1.1)**.** *Let S be a connected Shimura variety and $Z \subset S$ a closed subvariety. Let d be any nonnegative integer, and let Ξ denote the union of all special subvarieties of S of dimension at most d.*

If $\Xi \cap Z$ is Zariski dense in Z, then Z is contained in a special subvariety S_Z of S, such that $\dim S_Z \leq \dim Z + d$.

The André–Oort conjecture is the case $d = 0$ of the Zilber–Pink conjecture. The proof that the Zilber–Pink conjecture implies the André–Pink conjecture is more involved and is due to Pink ([Pin05b] Theorem 3.3).

We will outline the proof of this implication. Let $Z \subset S$ be a subvariety and Σ the generalised Hecke orbit of a point $s \in S$. We will apply the Zilber–Pink conjecture to the subvariety $Z \times \{s\} \subset S \times S$. Let S' be the smallest special subvariety of S containing s and let $d = \dim S'$. Then, for each point $t \in \Sigma$, (t, s) is contained in a special subvariety of $S \times S$ of dimension d, specifically an irreducible component of the image of $\mathrm{diag}(S')$ by some Hecke operator on $S \times S$. Thus $\Sigma \times \{s\} \subset \Xi$. So if $Z \cap \Sigma$ is dense in Z, then the Zilber–Pink conjecture implies that the smallest special subvariety $S_Z \subset S \times S$ containing $Z \times \{s\}$ has dimension at most $\dim Z + d$. The image of S_Z under the second projection $\pi_2 \colon S \times S \to S$ is S' and by comparing dimensions, we can deduce that Z is an irreducible component of a fibre of $\pi_{2|S_Z}$ and so is weakly special.

Known cases of the André–Pink conjecture. Several restricted cases of the André–Pink conjecture were previously known. Edixhoven and Yafaev [EY03] proved the intersection of the André–Oort and André–Pink conjectures for curves, that is, the André–Pink conjecture with the hypotheses that s is a special point and Z is a curve. Their proof used lower bounds for the Galois degrees of special points and a geometric criterion for proving that a subvariety is special – this was the method that developed into the proof of André–Oort assuming GRH by Klingler, Ullmo and Yafaev.

The full intersection of the André–Oort and André–Pink conjectures was proved by Klingler and Yafaev [KY13] – their proof of André–Oort does not require GRH in the case that all the special points involved are in a single Hecke orbit. In fact both Edixhoven–Yafaev and Klingler–Yafaev worked with classes of points slightly more general than Hecke orbits, classes which we call ρ-isogeny classes in this thesis.

Pink ([Pin05a] Theorem 7.6) proved the conjecture for Galois generic points s in the Siegel modular variety $\mathcal{A}_g$, that is, points whose associated Galois representation has open image in $\mathrm{GSp}_{2g}(\hat{\mathbb{Z}})$. Pink's proof used a result of Clozel, Oh and Ullmo on equidistribution of Hecke orbits.

Habegger and Pila [HP12] used the Pila–Zannier method to prove the conjecture for $\mathcal{A}_1^n$. They used an isogeny bound for elliptic curves due to Pellarin to obtain the Galois lower bounds.

The strategies of Edixhoven–Yafaev and of Pink do not appear to extend to the Hecke orbits consisting of points which are neither special nor Galois generic. This thesis is based on the strategy of Habegger and Pila.

Primary results of this thesis. We do not prove the full André–Pink conjecture for Shimura varieties of abelian type, but we prove the following cases.

Theorem 1.5. *Conjecture 1.3 holds if S is of abelian type and either:*

(i) Z is a curve, or

(ii) the smallest special subvariety of S containing s is equal to the smallest special subvariety of S containing Z.

We also prove the following theorem, which is the André–Pink conjecture for Shimura varieties of abelian type with "generalised Hecke orbit" replaced by "P-Hecke orbit" for a finite set P of prime numbers. The precise definition of P-Hecke orbits is in section 4.5. In any case, this theorem is weaker than the André–Pink conjecture because P-Hecke orbits are smaller than usual Hecke orbits. As an illustration, two points of the Siegel modular variety $\mathcal{A}_g$ are in the same P-Hecke orbit if and only if the corresponding principally polarised abelian varieties are related by a polarised isogeny whose degree has all of its prime factors in P.

Theorem 1.6. *Let S be a connected Shimura variety and $Z \subset S$ an irreducible subvariety. Let s be a point in S and let Σ_P be the P-Hecke orbit of s, for a finite set P of prime numbers.*

If $\Sigma_P \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of S.

Hecke orbits and isogeny classes. We define an **isogeny class** in $\mathcal{A}_g$ to be the set of points for which the associated abelian varieties are isogenous to some fixed abelian variety, ignoring the polarisations. Since points in a Hecke orbit correspond to abelian varieties which are related by polarised isogenies, each Hecke orbit in $\mathcal{A}_g$ is contained in an isogeny class. The relationship between Hecke orbits and isogeny classes and their different functoriality properties are very important to the proof of Theorem 1.5.

For a Hodge generic point in $\mathcal{A}_g$ (that is, a point which is not contained in any proper special subvariety), its isogeny class and its Hecke orbit coincide. However this is not true in general. Indeed there are isogeny classes in $\mathcal{A}_g$ which contain infinitely many Hecke orbits (Lemma 4.3).

On the other hand, isogeny classes are not too much bigger than Hecke orbits. Every isogeny class in $\mathcal{A}_g$ is contained in a Hecke orbit in $\mathcal{A}_{4g}$, via the natural embedding $\mathcal{A}_g \to \mathcal{A}_{4g}$ which maps a principally polarised abelian variety to its fourth power. (This is a non-trivial result, proved in Proposition 4.4 using the classification of Hermitian forms over the endomorphism algebra of an abelian variety.) Hence the following conjecture would be a corollary of the André–Pink conjecture for $\mathcal{A}_{4g}$ but cannot be deduced directly from the André–Pink conjecture for $\mathcal{A}_g$ alone.

Conjecture 1.7. *Let Z be an irreducible subvariety of $\mathcal{A}_g$. Let s be a point in $\mathcal{A}_g$ and let Σ be the isogeny class of s.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of $\mathcal{A}_g$.

As defined above, the notion of isogeny class applies only to the Siegel modular variety. But we can define related notions in any Shimura variety, depending on the choice of a rational representation of the reductive group G. Such a representation ρ gives rise to a $\mathbb{Q}$-Hodge structure for each point in the connected Shimura variety S. We define a **ρ-isogeny class** in S to be the set of points for which the associated $\mathbb{Q}$-Hodge structures belong to a fixed isomorphism class. This notion is mainly interesting when the representation ρ is faithful. In particular, plain isogeny classes in $\mathcal{A}_g$ are the same as ρ-isogeny classes for the standard $2g$-dimensional representation of GSp_{2g}.

For any representation ρ of G, each Hecke orbit in S is contained in a ρ-isogeny class. As we have seen in the case of $\mathcal{A}_g$ and the standard representation, a ρ-isogeny class need not be contained in a Hecke orbit. The example of the embedding $\mathcal{A}_g \hookrightarrow \mathcal{A}_{4g}$ suggests the following open question, which appears to be difficult.

Question 1.8. *Let S be a connected Shimura variety and ρ a faithful $\mathbb{Q}$-representation of the underlying reductive group G. Let Σ be a ρ-isogeny class in S.*

Is there always a Shimura variety S' and a Shimura embedding (or Shimura immersion) $f \colon S \to S'$ such that $f(\Sigma)$ is contained in a generalised Hecke orbit in S'?

The obstacle to proving the full André–Pink conjecture for Shimura varieties of abelian type is as follows. Let $[\iota] \colon S' \to S$ be a Shimura embedding, that is, a morphism of Shimura varieties induced by an injection $\iota \colon G' \to G$ of the underlying groups. If Σ is a Hecke orbit in S, then $[\iota]^{-1}(\Sigma)$ need not be contained in a finite union of Hecke orbits (or even of generalised Hecke orbits) in S'. Hence, in trying to prove the general André–Pink conjecture, we cannot replace the Shimura variety S by its smallest special subvariety containing Z, which would be the first step in the obvious induction on dimension.

This problem of pulling back by a Shimura embedding goes away if we work with ρ-isogeny classes instead of Hecke orbits. If ρ is a faithful representation of G and Σ is a ρ-isogeny class in S, then $[\iota]^{-1}(\Sigma)$ is trivially contained in a $(\rho \circ \iota)$-isogeny class in S'. This means that a positive answer to Question 1.8, combined with the results of chapter 6, would imply the André–Pink conjecture for Shimura varieties of abelian type.

Additional results of this thesis: isogeny bounds. As well as Theorems 1.5 and 1.6, we prove some other results concerning isogeny bounds and the Mumford–Tate groups of abelian varieties.

A key ingredient in the proof of Theorem 1.5 is the Masser–Wüstholz isogeny theorem [MW93a]: if we fix a number field K and a principally polarised abelian variety A defined over K, then for any other principally polarised abelian variety B defined over a number field $L \supset K$ and isogenous to A, there exists an isogeny $A \to B$ whose degree is polynomially bounded with respect to $[L : K]$.

This theorem says nothing about the compatibility of the isogenies involved with the polarisations. In particular, even if we assume that A and B are in the same polarised isogeny class, the isogeny of small degree whose existence is asserted by the theorem need not be a polarised isogeny. This leads to certain complexities in the proof of Proposition 6.5.

The following theorem rectifies this problem: if A and B are in the same polarised isogeny class and there is an isogeny $A \to B$ of given degree, then there is a polarised isogeny of polynomially bounded degree. This could be used to simplify the proof of Proposition 6.5, but the proof of Theorem 1.9 is so long that this seems like a net loss for the proof of Theorem 1.5. We have nonetheless included the theorem as it might have independent applications.

Theorem 1.9. *Let (A, λ) be a principally polarised abelian variety defined over a field of characteristic* 0*. There exist constants c, k depending only on (A, λ) such that, if (B, λ') is a principally polarised abelian variety for which*

1. *there exists an isogeny $f\colon A \to B$ compatible with the polarisations (of any degree), and*

2. *there exists an isogeny $g\colon A \to B$ of degree n (not necessarily compatible with the polarisations),*

then there exists an isogeny $h\colon A \to B$ compatible with the polarisations and of degree at most cn^k.

On the subject of isogeny bounds, we have also included in Theorem 5.2 an extension of the Masser–Wüstholz theorem from number fields to finitely generated fields of characteristic 0. This is used in the proof of Theorem 1.5, to allow us to prove the theorem for the Hecke orbit of a point $s \in S(\mathbb{C})$ and not just $s \in S(\bar{\mathbb{Q}})$.

Additional results of this thesis: Mumford–Tate groups. Chapter 7 is independent of the other chapters of the thesis. It concerns the ranks of Mumford–Tate groups of abelian varieties.

Let A be a principally polarised abelian variety of dimension g defined over $\mathbb{C}$. Via the Hodge structure on $H^1(A, \mathbb{C})$, we can associate with A a morphism of algebraic groups $h_A\colon \operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m \to \mathrm{GSp}_{2g,\mathbb{R}}$. When doing this, there is a choice of symplectic basis for $H^1(A, \mathbb{Z})$ and so h_A is defined up to conjugation by $\mathrm{GSp}_{2g}(\mathbb{Z})$. This is how the moduli space $\mathcal{A}_g$ is identified with the quotient of a conjugacy class of Hodge parameters of GSp_{2g} by $\mathrm{GSp}_{2g}(\mathbb{Z})$.

The **Mumford–Tate group** of A is defined as the smallest subgroup of GSp_{2g} defined over $\mathbb{Q}$ and containing the image of h_A.

Mumford–Tate groups are closely related to special subvarieties. Each special subvariety S of $\mathcal{A}_g$ has a generic Mumford–Tate group H. This means that every point of S not contained in a smaller special subvariety has Mumford–Tate group equal to H. Those points of S which are contained in smaller special subvarieties have Mumford–Tate groups strictly contained in H.

In this thesis, we prove a lower bound for the rank of the Mumford–Tate group of an abelian variety, subject to the condition that its simple abelian subvarieties are pairwise non-isogenous. Such a condition on the simple subvarieties is required to get a bound which increases with the dimension of the abelian variety, because A^n has the same Mumford–Tate group as A.

Theorem 1.10. *Let A be an abelian variety of dimension g whose simple abelian subvarieties are pairwise non-isogenous. Let n be the rank of the Mumford–Tate group of A.*

Then

$$n + \alpha(n)\sqrt{n \log_e n} \geq \log_2 g + 2$$

for a function $\alpha\colon \mathbb{N}_{\geq 2} \to \mathbb{R}$ satisfying $\alpha(n) < 2$ for all n and $\alpha(n) \to 1/\log_e 2 = 1.44...$ as $n \to \infty$.

If we further assume that the endomorphism ring of A is commutative, then $n \geq \log_2 g + 2$.

This extends a bound proved by Ribet [Rib81] for abelian varieties with complex multiplication.

We prove that the same bound holds for the rank of the ℓ-adic monodromy group of an abelian variety defined over a number field. This implies a lower bound for the degrees of the fields of definition of ℓ^n-torsion points of an abelian variety.

Remarks on the proof of Theorem 1.5. Our proof of Theorem 1.5 begins with the case of $S = \mathcal{A}_g$ and Z a curve. For this case, we follow the method of Pila and Zannier. This consists in studying the preimage of Z under some function $\alpha\colon U \to \mathcal{A}_g$ which is definable in an o-minimal structure (where U is some definable subset of $\mathbb{R}^n$). The function α is chosen so that every point in the Hecke orbit Σ has a rational preimage in U.

In our case, we take U to be a certain subset of $\mathrm{GL}_{2g}(\mathbb{R})$. One of the subtle points in the proof is that we cannot take U to be a subset of $\mathrm{GSp}_{2g}(\mathbb{R})^+$, due to the difference between isogeny classes and Hecke orbits and the fact that the Masser–Wüstholz isogeny theorem says nothing about the compatibility of isogenies with polarisations. The function $\alpha\colon U \to \mathcal{A}_g$ is defined such that its restriction to $U \cap \mathrm{GSp}_{2g}(\mathbb{R})^+$ is the same as the action of $\mathrm{GSp}_{2g}(\mathbb{R})^+$ on a suitably chosen point in $\mathcal{H}_g$.

As well as the above choice of U and α, there are two other new ingredients in our proof: certain bounds on the heights of rational representations of isogenies and the use of the Masser–Wüstholz isogeny theorem to obtain a lower bound for the Galois degrees of points in the Hecke orbit Σ. The height bounds for isogenies form part of a theme of calculations concerning Hermitian forms over endomorphism algebras of abelian varieties, which appear also in the additional results on polarised and unpolarised isogenies.

With the above ingredients in place, the proof of Theorem 1.5 for $S = \mathcal{A}_g$ and Z a curve is completed by applying a strong version of the Pila–Wilkie counting theorem ([PW06], [Pil11] Theorem 3.6) on rational points in definable sets and the

Ullmo–Yafaev characterisation [UY11] of weakly special subvarieties as algebraic subvarieties of $\mathcal{A}_g$ whose preimage in $\mathcal{H}_g$ has an algebraic irreducible component.

To extend the result from $\mathcal{A}_g$ to other Shimura varieties of abelian type, we show that any Shimura variety S of abelian type has a finite covering by some other Shimura variety S_1 which can be embedded in $\mathcal{A}_g$, and such that the preimage in S_1 of each Hecke orbit in S is contained in a finite union of Hecke orbits in $\mathcal{A}_g$. This is a refinement of a result of Deligne ([Del79] Proposition 2.3.10), proved as part of his classification of Shimura varieties of abelian type. In order to get a result on Hecke orbits, we have to extend Deligne's argument to keep track of the centres of the reductive groups involved.

The proof of part (ii) of Theorem 1.5, concerning higher-dimensional subvarieties, is by induction on the dimension. Here in place of Ullmo and Yafaev's characterisation of weakly special subvarieties, we must use the hyperbolic Ax–Lindemann–Weierstrass conjecture for $\mathcal{A}_g$ whose proof has been announced by Pila and Tsimerman [PT12].

Structure of the thesis. Chapter 2 summarises the theory of Shimura varieties and special and weakly special subvarieties as they are used in this thesis. This consists primarily of definitions and of well-known facts, although in some cases I have written proofs where these cannot be found in the literature.

Chapter 3 outlines the Pila–Zannier strategy, recalling the various key ingredients and sketching the proof of the Manin–Mumford conjecture for tori which is the easiest application of this strategy. This is entirely well-known material.

Chapter 4 concerns Hecke orbits and variations thereon, in particular generalised Hecke orbits, P-Hecke orbits and isogeny classes. We recall the definitions of these different types of orbits and the relations between them, and prove some results on their functoriality with respect to Shimura morphisms. These functoriality results are useful for reducing various cases of the André–Pink conjecture to simpler cases. The definitions and basic properties are due to Pink, but much of the chapter is original.

Chapter 5 collects several bounds relating to isogenies of abelian varieties, which are essential to the strategy for the proof of the André–Pink conjecture for curves in $\mathcal{A}_g$. We also prove a bound for polarised isogenies which is not used in this thesis, but could be of independent interest. This consists entirely of original material.

Chapter 6 contains the proofs of our main theorems on the André–Pink conjecture (Theorems 1.5 and 1.6). This is by combining the ingredients of chapters 3, 4 and 5.

Chapter 7 proves lower bounds for the ranks of Mumford–Tate groups of abelian varieties. It is independent of the rest of the thesis.

2 Shimura varieties

In this chapter we recall the definition of connected Shimura varieties and various related definitions and fundamental theorems, mostly without proofs. We begin with the examples of the moduli spaces of principally polarised abelian varieties, which motivate the general definition as well as playing a special role in the proofs of the main theorems of this thesis. We then give the general definition of connected Shimura data and connected Shimura varieties as in [Pin05a]. We prove various properties of morphisms of Shimura varieties which are asserted without proof in [Pin05a]. The chapter ends with discussion of the definitions of special and weakly special subvarieties which appear in the André–Pink and Zilber–Pink conjectures.

We do not discuss canonical models in detail because they are not used in this thesis. Since we do not need canonical models, we also simplify the discussion by only considering connected Shimura varieties and by not using adelic language. (However our definition of connected Shimura data differs from that of Deligne because we do need to allow arbitrary connected reductive groups and not only semisimple ones.)

The primary sources for the theory of Shimura varieties are [Del71] and [Del79]. A more pedagogic source is [Mil05]. The definitions we have chosen (as being the simplest for the purposes of this thesis) come from [Pin05b].

2.1 Period matrices of abelian varieties

The fundamental examples of Shimura varieties are the Siegel modular varieties, the moduli spaces of principally polarised abelian varieties of given dimension. Before discussing these, let us recall briefly the period matrices of abelian varieties.

A **polarisation** of an abelian variety of A is an isogeny $A \to A^\vee$ to the dual variety satisfying a certain positivity condition. Any polarisation induces a symplectic form $H_1(A,\mathbb{Z}) \times H_1(A,\mathbb{Z}) \to \mathbb{Z}$. A polarisation is **principal** if it has degree 1 – that is, if it is an isomorphism $A \to A^\vee$. A **principally polarised abelian variety** means a pair (A, λ) where A is an abelian variety and λ is a principal polarisation of A.

Let (A, λ) be a complex principally polarised abelian variety. There is an exact sequence

$$0 \to H_1(A,\mathbb{Z}) \to T_0A \to A(\mathbb{C}) \to 0$$

which realises $H_1(A,\mathbb{Z})$ (which is a free $\mathbb{Z}$-module of rank $2g$) as a lattice in T_0A (which is a complex vector space of dimension g).

Choose a symplectic basis $\{e_1, \dots, e_g, f_1, \dots, f_g\}$ for $H_1(A,\mathbb{Z})$ with respect to the symplectic form ψ induced by the polarisation. Then $\{e_1, \dots, e_g\}$ form a basis for the complex vector space T_0A. The complex $g \times g$ matrix giving the coordinates of $\{f_1, \dots, f_g\}$ in this basis is called a **period matrix** of A.

According to the Riemann bilinear relations, the period matrix is in the Siegel upper half space

$$\mathcal{H}_g = \{Z \in \mathrm{M}_g(\mathbb{C}) \mid Z \text{ is symmetric and } \operatorname{Im} Z \text{ is positive definite } \}.$$

Conversely, every matrix in $\mathcal{H}_g$ is the period matrix of some principally polarised abelian variety, unique up to isomorphism of polarised abelian varieties.

Different choices of symplectic basis for $H_1(A, \mathbb{Z})$ may give rise to different period matrices. If the change of basis matrix is M, then M is in the symplectic group

$$\mathrm{Sp}_{2g}(\mathbb{Z}) = \{M \in \mathrm{GL}_{2g}(\mathbb{Z}) \mid M^t J M = J\} \text{ where } J = \left(\begin{smallmatrix} 0 & \mathbb{1}_g \\ -\mathbb{1}_g & 0 \end{smallmatrix}\right)$$

because it maps one symplectic basis into another. The period matrices Z, Z' are related by

$$Z' = (AZ + B)(CZ + D)^{-1} \text{ where } M = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right),\ A, B, C, D \in \mathrm{M}_g(\mathbb{Z}).$$

2.2 The Siegel modular varieties

Let $\mathrm{GSp}_{2g}(\mathbb{R})^+$ denote the group of real symplectic similitudes with positive multiplier:

$$\mathrm{GSp}_{2g}(\mathbb{R})^+ = \{M \in \mathrm{GL}_{2g}(\mathbb{R}) \mid M^t J M = \lambda(M) J \text{ for some } \lambda(M) \in \mathbb{R}^+\}.$$

This group acts on the Siegel upper half space $\mathcal{H}_g$ by

$$\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) Z = (AZ + B)(CZ + D)^{-1},\ A, B, C, D \in \mathrm{M}_g(\mathbb{R}).$$

As we saw in the previous section, associating a principally polarised abelian variety with its period matrices gives a bijection between the isomorphism classes of complex principally polarised abelian varieties and the set of orbits $\mathrm{Sp}_{2g}(\mathbb{Z})\backslash\mathcal{H}_g$ for this action. The action of $\mathrm{Sp}_{2g}(\mathbb{Z})$ on $\mathcal{H}_g$ is properly discontinuous, so the quotient is a complex analytic space. Satake [Sat56] and Baily [Bai58] showed that this quotient is isomorphic to a quasi-projective algebraic variety, called the **Siegel modular variety** and denoted $\mathcal{A}_g$.

The Siegel modular variety (equipped with the bijection between its complex points and isomorphism classes of principally polarised abelian varieties) can also be characterised purely in terms of algebraic geometry by the following properties ([Mil08] Theorem IV.7.3):

(a) for each point $s \in \mathcal{A}_g$, there is a Zariski open neighbourhood U of s and a polarised abelian scheme $A \to U$ such that for every $t \in U$, the fibre A_t is isomorphic to the principally polarised abelian variety which corresponds to the point $t \in \mathcal{A}_g$;

(b) for every complex algebraic variety T and every principally polarised abelian scheme $A \to T$ of relative dimension g, if φ denotes the map $T \to \mathcal{A}_g$ which sends $t \in T$ to the point of $\mathcal{A}_g$ corresponding to the principally polarised abelian variety A_t, then φ is a morphism of algebraic varieties.

These properties imply that the variety $\mathcal{A}_g$ is a coarse moduli space for principally polarised abelian varieties. It is not a fine moduli space because the abelian schemes defined over open sets of $\mathcal{A}_g$, as in (a), cannot be glued together into an abelian scheme on all of $\mathcal{A}_g$.

In order to obtain fine moduli spaces, we introduce level structures. Moduli spaces for polarised abelian varieties with appropriate level structures also fix several other deficiencies of $\mathcal{A}_g$, for example that it is not smooth for $g \geq 2$ and that the map $\mathcal{H}_g \to \mathcal{A}_g$ is ramified.

Let N be a positive integer and fix a primitive N-th root of unity ζ_N. We define a **level-N structure** on a principally polarised abelian variety (A, λ) to be a symplectic basis for the N-torsion $A[N]$ with respect to the Weil pairing $e_N \colon A[N] \times A[N] \to \mu_N$. More precisely, a level-N structure is a basis $\{x_1, \dots, x_g, y_1, \dots, y_g\}$ for $A[N]$ as a free $\mathbb{Z}/N\mathbb{Z}$-module such that

$$\psi(x_i, y_i) = \zeta_N \text{ for all } i, \text{ and}$$
$$\psi(x_i, x_j) = \psi(y_i, y_j) = \psi(x_i, y_j) = 1 \text{ for all } i \neq j.$$

Let $\Gamma(N)$ denote the set of matrices in $\mathrm{Sp}_{2g}(\mathbb{Z})$ which are congruent to the identity modulo N. Then the quotient $\Gamma(N) \backslash \mathcal{H}_g$ is also a quasi-projective complex variety and its points are in bijection with the isomorphism classes of principally polarised abelian varieties with level-N structure. We will denote this variety $\mathcal{A}_g(N)$.

For $N \geq 3$, $\mathcal{A}_g(N)$ is a fine moduli space – that is, there is a principally polarised abelian scheme $A \to \mathcal{A}_g(N)$ with $2g$ distinguished N-torsion sections such that each fibre is isomorphic to the correct principally polarised abelian variety with level-N structure. It is no coincidence that the group $\Gamma(N)$ is torsion-free if and only if $N \geq 3$, and that a principally polarised abelian variety with level-N structure has no non-trivial automorphisms if and only if $N \geq 3$.

A coarse moduli space for principally polarised abelian varieties can also be constructed purely algebraically via geometric invariant theory in place of the above analytic construction ([Mum65] Chapter 7). The algebraic construction works over any field and hence gives a canonical model for $\mathcal{A}_g$ as an algebraic variety over $\mathbb{Q}$. Note that, for each field $k \supset \mathbb{Q}$, there is a map from the set of isomorphism classes of principally polarised abelian varieties defined over k to $\mathcal{A}_g(k)$, but in general this map is neither injective nor surjective. Similarly each variety $\mathcal{A}_g(N)$ has a canonical model over $\mathbb{Q}(\mu_N)$, but not over $\mathbb{Q}$ because of the need to choose a primitive N-th root of unity in the definition of level-N structures.

2.3 Hodge structures

Loosely speaking, Shimura varieties are moduli spaces of Hodge structures. Accordingly we recall the definition of a Hodge structure. A **real Hodge structure** is a finite-dimensional real vector space V together with a bigrading of $V \otimes_{\mathbb{R}} \mathbb{C}$:

$$V \otimes_{\mathbb{R}} \mathbb{C} = \bigoplus_{(p,q) \in \mathbb{Z}^2} V^{pq},$$

such that V^{pq} is the complex conjugate of V^{qp} for all p, q.

We say that a real Hodge structure V is

(i) **of type** T if $T = \{(p, q) \in \mathbb{Z}^2 \mid V^{pq} \neq 0\}$;

(ii) **pure of weight** n if $V^{pq} = 0$ for all (p, q) such that $p + q \neq n$;

(iii) **effective** if $V^{pq} = 0$ for all (p, q) such that either $p < 0$ or $q < 0$.

We may equivalently define a real Hodge structure to be a real representation of the **Deligne torus** $\mathbb{S} = \operatorname{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m$. The complex character group of $\mathbb{S}$ is generated by two characters z and $\bar{z}$ which are exchanged by complex conjugation. To translate from a representation $\mathbb{S} \to \mathrm{GL}(V)$ into the primary definition of a Hodge structure, we let V^{pq} be the subspace of $V \otimes_{\mathbb{R}} \mathbb{C}$ on which $\mathbb{S}$ acts by $z^{-p}\bar{z}^{-q}$. (This is the standard sign convention, coming from [Del79] and [Del82]. It is chosen because we want cohomology Hodge structures to have positive weight, but we also want $\mathbb{S}$ to act as z on the holomorphic part $H^{-1,0}(A) \subset H_1(A, \mathbb{C})$ of the first homology of an abelian variety, which can be identified with the tangent space T_0A.)

Let w be the cocharacter $\mathbb{G}_{m,\mathbb{R}} \to \mathbb{S}$ induced by the inclusion $\mathbb{R}^\times \to \mathbb{C}^\times$ (in other words, $z \circ w = \bar{z} \circ w = \mathrm{id}$). Then a Hodge structure (defined as a representation of $\mathbb{S}$) is pure of weight n if and only if $h \circ w \colon \mathbb{G}_{m,\mathbb{R}} \to \mathrm{GL}(V)$ is given by $x \mapsto x^{-n} \mathbb{1}_V$.

If A is any subring of $\mathbb{R}$, an **A-Hodge structure** is defined to be a free A-module V_A of finite rank together with a real Hodge structure on $V_A \otimes_A \mathbb{R}$. In practice, we only use this definition for $A = \mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$. (Note that we do not require the weight grading to be defined over A.)

A **morphism of A-Hodge structures** is defined in the obvious way: it is an A-linear map $f \colon V_A \to W_A$ such that $f(V^{pq}) \subset W^{pq}$ for all p, q. Equivalently, it is an A-linear map $f \colon V_A \to W_A$ such that $f_{\mathbb{R}}$ is a homomorphism of representations of $\mathbb{S}$.

Polarisations. A **polarisation** of an A-Hodge structure is a bilinear form

$$\psi \colon V_A \times V_A \to A$$

which is invariant under $U_1 = \ker(z\bar{z}) \subset \mathbb{S}$ and such that $(x, y) \mapsto \psi(x, h(i)y)$ is a positive-definite symmetric form on $V_{\mathbb{R}}^n$. In practice we only ever consider polarisations on Hodge structures of pure weight. The symmetry of $\psi(x, h(i)y)$ implies that ψ is symmetric if V is purely of even weight and alternating if V is purely of odd weight.

A Hodge structure V is said to be **polarisable** if there exists a polarisation of V.

Example: cohomology. The motivating example for the definition of Hodge structure is the cohomology of a complex projective variety. If X is a smooth projective variety over $\mathbb{C}$, then its cohomology $H^n(X, \mathbb{Z})$ (modulo torsion) carries the structure of an effective $\mathbb{Z}$-Hodge structure of weight n. A choice of ample line bundle on X induces polarisations on the primitive parts of the cohomology.

In this thesis we are interested in the case where X is an abelian variety. We prefer to work with homology instead of cohomology in order to have a covariant functor. For abelian varieties, the homology is dual to the cohomology, so $H_1(X, \mathbb{Z})$ has a Hodge structure of type $\{(-1, 0), (0, -1)\}$. The functor $X \mapsto H_1(X, \mathbb{Z})$ is an equivalence of categories between the category of complex abelian varieties and the category of polarisable $\mathbb{Z}$-Hodge structures of type $\{(-1, 0), (0, -1)\}$.

There are a couple of useful variants on this equivalence: we can add a choice of polarisation on each side, or we can consider the isogeny category of abelian varieties on one side and $\mathbb{Q}$-Hodge structures on the other.

2.4 Mumford–Tate groups and Hodge parameters

Let $(V, h \colon \mathbb{S} \to \mathrm{GL}(V_{\mathbb{R}}))$ be a $\mathbb{Q}$-Hodge structure. The **Mumford–Tate group** of V is the smallest algebraic subgroup $M \subset \mathrm{GL}(V)$, defined over $\mathbb{Q}$, such that h factorises through $M_{\mathbb{R}}$.

If the Hodge structure has a polarisation ψ, then the Mumford–Tate group is reductive and is contained in the group of symplectic similitudes

$$\mathrm{GSp}(V, \psi) = \{g \in \mathrm{GL}(V) \mid \exists \chi(g) \in \mathbb{G}_m. \forall x, y \in V. \, \psi(gx, gy) = \chi(g)\psi(x, y)\}.$$

The bigger the Mumford–Tate group, the more transcendental and generic the Hodge structure is. At one extreme, the Mumford–Tate group can be equal to $\mathrm{GSp}(V, \psi)$. At the other extreme, it may be a torus. A Hodge structure whose Mumford–Tate group is a torus is called a **CM Hodge structure**, because the

Mumford–Tate group of an abelian variety is a torus if and only if the abelian variety has complex multiplication.

In the case of abelian varieties, the bigger the endomorphism ring of the abelian variety, the smaller the Mumford–Tate group. Thus moduli spaces of abelian varieties with given endomorphisms (for example, Hilbert modular surfaces and Shimura curves) can be expressed as moduli spaces of Hodge structures whose Mumford–Tate group is contained in a given group G. Note that for moduli problems, we should always consider Hodge structures whose Mumford–Tate group is contained in G, not just those whose Mumford–Tate group is equal to G, as the first gives closed subvarieties in the moduli space and the second does not.

By definition, if $(V, h\colon \mathbb{S} \to \mathrm{GL}(V_\mathbb{R}))$ is a $\mathbb{Q}$-Hodge structure whose Mumford–Tate group is contained in $G \subset \mathrm{GL}(V)$ then h factors as

$$\mathbb{S} \to G_\mathbb{R} \hookrightarrow \mathrm{GL}(V_\mathbb{R}).$$

In order to study the collection of all such h on a fixed V, we may forget the second half of this factorisation, and focus on $\mathbb{S} \to G_\mathbb{R}$. Accordingly, we make the following definition: if G is a reductive group over $\mathbb{Q}$, then a **Hodge parameter** of G is a morphism $h\colon \mathbb{S} \to G_\mathbb{R}$.

A Hodge parameter should be thought of as a prototype for a $\mathbb{Q}$-Hodge structure in which we have forgotten the vector space V. Indeed given any rational representation $\rho\colon G \to \mathrm{GL}(V)$, each Hodge parameter of G induces a $\mathbb{Q}$-Hodge structure $\rho \circ h$ on V.

We may now define the **Mumford–Tate group** of a Hodge parameter $h\colon \mathbb{S} \to G_\mathbb{R}$ in the same way as the Mumford–Tate group of a Hodge structure: it is the smallest algebraic subgroup $M \subset G$, defined over $\mathbb{Q}$, such that h factorises through $M_\mathbb{R}$.

2.5 Shimura data

Let G be a connected reductive group over $\mathbb{Q}$. We will write:

(i) $G(\mathbb{R})^+$ for the identity component of $G(\mathbb{R})$ in the real topology,

(ii) $G(\mathbb{R})_+$ for the preimage in $G(\mathbb{R})$ of $G^{\mathrm{ad}}(\mathbb{R})^+$, and

(iii) $G(\mathbb{Q})_+$ for $G(\mathbb{R})_+ \cap G(\mathbb{Q})$.

Let X^+ be a connected component of a $G(\mathbb{R})$-conjugacy class of Hodge parameters of G. This is a homogeneous space for the Lie group $G(\mathbb{R})^+$, hence is a real manifold. The reason for begin interested in $G(\mathbb{R})_+$ is that it is the stabiliser of X^+.

Any rational representation $\rho\colon G \to \mathrm{GL}(V)$ induces a smooth family of $\mathbb{Q}$-Hodge structures on V parametrised by X^+. If the representation ρ is faithful, then the set of all Hodge structures on V whose Mumford–Tate group is contained in $\rho(G)$ is a finite union of such conjugacy classes.

A connected Shimura datum is defined to be a pair (G, X^+) as above which satisfies certain conditions ensuring that the associated families of Hodge structures are well-behaved.

Definition. A **connected Shimura datum** is a pair (G, X^+) where G is a connected reductive group over $\mathbb{Q}$ and X^+ is a $G(\mathbb{R})_+$-conjugacy class of Hodge parameters of G, such that for all $h \in X^+$:

(SV1) the Hodge structure $\mathrm{Ad} \circ h$ has type contained in $\{(-1, 1), (0, 0), (1, -1)\}$;

(SV2) $\theta = \mathrm{Ad}\, h(i)$ is a Cartan involution of $G^{\mathrm{ad}}_{\mathbb{R}}$ (this means that the real Lie group $\{g \in G^{\mathrm{ad}}(\mathbb{C}) \mid g = \theta(\bar{g})\}$ is compact, where $\bar{g}$ denotes the complex conjugate of g);

(SV3) G^{ad} has no $\mathbb{Q}$-factor on which the projection of h is trivial.

The notions of Shimura data and connected Shimura data are due to Deligne [Del79]. The definition above of connected Shimura data differs from both that of Deligne and that of Pink ([Pin05a] Definition 2.1) by allowing greater freedom in the choice of the centre of G. Deligne considers only adjoint groups G in his definition of connected Shimura data (although in order to have enough congruence subgroups he also includes a semisimple cover of G in the data), while Pink requires that G should be the generic Mumford–Tate group of X^+. Allowing the centre of G to be larger than necessary does not fundamentally change the possible connected Shimura data because the centre acts trivially on X^+, but it does affect the Hecke orbits. This additional freedom will be important in this thesis.

The following theorem explains the meaning of conditions (SV1) and (SV2) in terms of Hodge structures. (SV3) simply ensures that the semisimple part of G is no larger than necessary.

Theorem 2.1 ([Del79] Proposition 1.1.14, Corollary 1.1.17)**.** *Let G be a connected reductive group over $\mathbb{Q}$ and X^+ a $G(\mathbb{R})_+$-conjugacy class of Hodge parameters of G. Choose a rational representation ρ of G and let $\mathcal{V}_\rho$ denote the induced family of Hodge structures.*

Suppose that the Hodge structure $\mathrm{Ad} \circ h$ *is pure of weight* 0 *(equivalently, the image of $h \circ w$ is contained in the centre of G).*

1. *There is a unique complex structure on X^+ such that $\mathcal{V}_\rho$ varies holomorphically.*

2. *(G, X^+) satisfies (SV1) if and only if $\mathcal{V}_\rho$ satisfies the Griffiths transversality condition (this is a differential equation satisfied by the family of Hodge structures coming from the cohomology of the fibres of a smooth projective morphism of complex varieties).*

3. *(G, X^+) satisfies (SV2) if and only if $\mathcal{V}_\rho$ is polarisable.*

4. *If (G, X^+) satisfies (SV1) and (SV2), then X^+ is a Hermitian symmetric domain.*

Example. Let $G = \mathrm{GSp}_{2g}$, and let ρ be the standard representation $G \to \mathrm{GL}_{2g}$. Let X^+ be the $G(\mathbb{R})_+$-conjugacy classes of Hodge parameters for which $\rho \circ h$ is a $\mathbb{Q}$-Hodge structure of type $\{(-1, 0), (0, -1)\}$ and the standard symplectic form is a polarisation.

Then (G, X^+) is a connected Shimura datum. There is an isomorphism of Hermitian symmetric domains $\phi\colon X^+ \to \mathcal{H}_g$ such that for each $h \in X^+$, $\rho \circ h$ is the polarised Hodge structure $H_1(A, \mathbb{Q})$ associated with the principally polarised abelian variety (A, λ) with period matrix $\phi(h)$. Furthermore ϕ is equivariant with respect to the action of $G(\mathbb{R})_+$ on X^+ by conjugation and its action on $\mathcal{H}_g$ as defined in section 2.2.

2.6 Shimura varieties

Connected Shimura varieties are constructed from connected Shimura data (G, X^+) by taking quotients of X^+ by congruence subgroups of $G(\mathbb{Q})_+$. This generalises the construction of the modular varieties $\mathcal{A}_g(N)$ as quotients of $\mathcal{H}_g$ by congruence subgroups of $\mathrm{GSp}_{2g}(\mathbb{Q})^+$. A connected Shimura variety has an interpretation as a moduli space for polarised Hodge structures whose Mumford–Tate group is contained in G and equipped with certain level structure, but specifying exactly the level structure for a general connected Shimura variety is tricky and is best done using adelic language, which we do not use in this thesis.

Let G be a reductive algebraic group over $\mathbb{Q}$. Choose a faithful $\mathbb{Q}$-representation $\rho\colon G \to \mathrm{GL}_{n,\mathbb{Q}}$ and let

$$\Gamma(N) = \{g \in G(\mathbb{Q}) \mid \rho(g) \in \mathrm{GL}_n(\mathbb{Z}) \text{ and } \rho(g) \equiv \mathbb{1}_n \bmod N\},$$

where N is a positive integer. A **congruence subgroup** of $G(\mathbb{Q})$ is a subgroup which contains some $\Gamma(N)$ as a subgroup of finite index. The groups $\Gamma(N)$ depend on the choice of representation ρ, but the definition of congruence subgroup does not (it is possible to define congruence subgroups independently of ρ by saying that they are subgroups of the form $K \cap G(\mathbb{Q})$ where K is a compact open subgroup of $G(\mathbb{A}_f)$).

We define a **connected Shimura variety** to be a quotient complex variety $\mathrm{Sh}_\Gamma(G, X^+) = \Gamma\backslash X^+$ where (G, X^+) is a connected Shimura datum and $\Gamma \subset G(\mathbb{Q})_+$ is a congruence subgroup. The group Γ acts properly discontinuously on X^+ and so $\Gamma\backslash X^+$ is a normal complex analytic space ([Car57] Theorem 4). Baily and Borel [BB66] used automorphic forms to give $\mathrm{Sh}_\Gamma(G, X^+)$ a structure of quasi-projective variety over $\mathbb{C}$.

The connected Shimura varieties associated with a given connected Shimura datum form a tower, with a finite surjective map $\mathrm{Sh}_\Gamma(G, X^+) \to \mathrm{Sh}_{\Gamma'}(G, X^+)$ whenever $\Gamma' \subset \Gamma$. If Γ is torsion free, then $\mathrm{Sh}_\Gamma(G, X^+)$ is smooth and X^+ is its universal cover.

Deligne [Del79] showed that a large class of connected Shimura varieties (those of abelian type) have models over a number field, so-called **weakly canonical models**, in which the Galois action on special points (special points are defined below in section 2.9) obeys a certain reciprocity law. The field of definition of such a model depends on G, X^+ and Γ. The existence of weakly canonical models for all Shimura varieties was proved by Milne and Shih.

In fact, Deligne worked with non-connected Shimura varieties; over $\mathbb{C}$ these are finite disjoint unions of connected Shimura varieties. The benefit of working with non-connected Shimura varieties is that they have a **canonical model**, whose field of definition depends only on (G, X^+) and not on Γ. This is particularly important if we want to consider the entire tower of Shimura varieties associated with (G, X^+) simultaneously.

For example, the connected Shimura variety $\mathcal{A}_g(N)$ is defined over $\mathbb{Q}(\mu_N)$ while the corresponding non-connected Shimura variety has $\phi(N)$ connected components and is defined over $\mathbb{Q}$.

In this thesis, we only need to know that Shimura varieties are defined over some number field and do not care about what the field of definition is, so it will suffice to consider only connected Shimura varieties.

2.7 Shimura morphisms

A **morphism of Shimura data** $(G_1, X_1^+) \to (G_2, X_2^+)$ is a homomorphism $f\colon G_1 \to G_2$ of algebraic groups over $\mathbb{Q}$ which induces a map $f_*\colon X_1^+ \to X_2^+$. If $\Gamma_1 \subset G_1(\mathbb{Q})$ and $\Gamma_2 \subset G_2(\mathbb{Q})$ are congruence subgroups such that $f(\Gamma_1) \subset \Gamma_2$, then f induces a holomorphic map

$$[f]\colon\ \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+) \to \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+)$$

on the Shimura varieties. Such a map $[f]$ is called a **Shimura morphism**. In fact $[f]$ is a morphism of algebraic varieties; this will be proved below in Theorem 2.4.

Following Pink [Pin05a] for (b), (c) and (d), we say that:

(a) $[f]$ is a **Shimura embedding** if $f: G_1 \to G_2$ is injective;

(b) $[f]$ is a **Shimura immersion** if the identity component of $\ker f$ is a torus (equivalently, if $\ker f$ is contained in the centre of G_1);

(c) $[f]$ is a **Shimura submersion** if $\operatorname{Im} f$ contains G_2^{der};

(d) $[f]$ is a **Shimura covering** if it is both a Shimura immersion and a Shimura submersion.

Note that a Shimura immersion or embedding need not be injective (but as we will prove below, it is finite). As Pink explains, the word "immersion" should be interpreted locally but even then it is somewhat abusive because a Shimura immersion may be ramified if Γ_2 is not neat.

Before discussing the properties of Shimura morphisms, we will construct some important examples. First we need a lemma on the images and preimages of congruence subgroups under group homomorphisms.

Lemma 2.2. *Let $f: G_1 \to G_2$ be a homomorphism of reductive groups over $\mathbb{Q}$.*

(1) If Γ_1 is a congruence subgroup in $G_1(\mathbb{Q})$, then $f(\Gamma_1)$ is contained in a congruence subgroup of $G_2(\mathbb{Q})$.

(2) If Γ_2 is a congruence subgroup in $G_2(\mathbb{Q})$ and Γ_1 is any congruence subgroup of $G_1(\mathbb{Q})$, then $f^{-1}(\Gamma_2) \cap \Gamma_1$ is a congruence subgroup of $G_1(\mathbb{Q})$.

Proof.

(1) Let ρ be a faithful $\mathbb{Q}$-representation $G_2 \to \mathrm{GL}_{n,\mathbb{Q}}$. By [PR94] Proposition 4.2 and the Remark following it, there is a $\rho f(\Gamma_1)$-invariant lattice $L \subset V$. Replacing ρ by a suitable $\mathrm{GL}_n(\mathbb{Q})$-conjugate, we may assume that $L = \mathbb{Z}^n$. Then $f(\Gamma_1)$ is contained in the congruence subgroup $\rho^{-1}(\mathrm{GL}_n(\mathbb{Z}))$.

(2) By [Mar91] Lemma I.3.1.1(ii), $f^{-1}(\Gamma_2)$ contains a congruence subgroup Γ_1'. Then
$$\Gamma_1' \cap \Gamma_1 \subset f^{-1}(\Gamma_2) \cap \Gamma_1 \subset \Gamma_1$$
and both $\Gamma_1' \cap \Gamma_1$ and Γ_1 are congruence subgroups, so $f^{-1}(\Gamma_2) \cap \Gamma_1$ is also a congruence subgroup. ☐

Here are some important constructions of Shimura morphisms.

(1) If $\Gamma_1 \subset \Gamma_2$ are two congruence subgroups of $G(\mathbb{Q})_+$, then the natural projection $\mathrm{Sh}_{\Gamma_1}(G, X^+) \to \mathrm{Sh}_{\Gamma_2}(G, X^+)$ is a Shimura covering.

(2) Let (G_1, X_1^+) be a connected Shimura datum and $\pi\colon G_1 \to G_2$ a surjective morphism of algebraic groups whose kernel is contained in the centre of G_1. In particular, G_2 could be the adjoint group of G_1. Let $X_2^+ = \pi_* X_1^+$; this is a set of Hodge parameters of G_2. By [Mil05] Proposition 5.7, (G_2, X_2^+) is a connected Shimura datum and $\pi_*\colon X_1^+ \to X_2^+$ is bijective.

   (a) Suppose we are given a congruence subgroup $\Gamma_1 \subset G_1(\mathbb{Q})_+$.
   By Lemma 2.2(i), there is a congruence subgroup $\Gamma_2 \subset G_2(\mathbb{Q})_+$ containing $\pi(\Gamma_1)$, and then $[\pi]\colon \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+) \to \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+)$ is a Shimura covering.

   (b) Suppose we are given a congruence subgroup $\Gamma_2 \subset G_2(\mathbb{Q})_+$.
   By Lemma 2.2(ii), there is a congruence subgroup $\Gamma_1 \subset G_1(\mathbb{Q})_+$ contained in $\pi^{-1}(\Gamma_2)$. Then $[\pi]\colon \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+) \to \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+)$ is a Shimura covering.

   Construction (a) shows that, given any connected Shimura variety S with underlying group G, there is always a Shimura variety S^{ad} with underlying group G^{ad} such that there is a Shimura covering $S \to S^{\mathrm{ad}}$.

(3) Let (G, X^+) be a connected Shimura datum and $H \subset G$ a connected reductive subgroup defined over $\mathbb{Q}$. We shall write $X^+[H]$ for the set of Hodge parameters in X^+ whose image is contained in H. Assume that $X^+[H]$ is non-empty and let X_H^+ be a connected component of X_H. Then (H, X_H^+) is a connected Shimura variety.

   If Γ is a congruence subgroup of $G(\mathbb{Q})_+$, then $\Gamma_H = \Gamma \cap H(\mathbb{Q})_+$ is a congruence subgroup of $H(\mathbb{Q})_+$. The inclusion $\iota\colon G \to H$ induces a Shimura embedding $\mathrm{Sh}_{\Gamma_H}(H, X_H^+) \to \mathrm{Sh}_\Gamma(G, X^+)$.

As we mentioned above, Shimura immersions need not be injective. This is true even in the case of Shimura embeddings where $\Gamma_1 = \Gamma_2 \cap G_1(\mathbb{Q})_+$, as in construction (3). For an example of this, let $G = \mathrm{GSp}_4$ and

$$H = \mathbb{G}_m.(\mathrm{SL}_2 \times \mathrm{SL}_2) = \{ \left(\begin{smallmatrix} A & 0 \\ 0 & B \end{smallmatrix}\right) \mid A, B \in \mathrm{GL}_2 \text{ and } \det A = \det B \} \subset G.$$

Here H is the Mumford–Tate group of a generic product of two elliptic curves. If $\Gamma = \mathrm{Sp}_4(\mathbb{Z})$ then $\Gamma_H = \mathrm{SL}_2(\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z})$ and the induced Shimura morphism is

$$[\iota]\colon \mathcal{A}_1 \times \mathcal{A}_1 \to \mathcal{A}_2$$

sending a pair of elliptic curves (E_1, E_2) to their product $E_1 \times E_2$, equipped with the product of the principal polarisations. This is not injective because $E_1 \times E_2 \cong E_2 \times E_1$.

Lemma 2.3. *The composition of two Shimura immersions is a Shimura immersion.*

Proof. Suppose we have homomorphisms of reductive groups $f_1 \colon G_1 \to G_2$ and $f_2 \colon G_2 \to G_3$ each with central kernel. Then $\ker(f_2 \circ f_1)^\circ$ is an extension of one torus by another, so is solvable. It is also a normal subgroup of G_1, so as G_1 is reductive it is a torus. □

Theorem 2.4 (cf. [Pin05a] Facts 2.6)**.** *Let* $[f] \colon \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+) \to \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+)$ *be a Shimura morphism.*

(1) $[f]$ *is a morphism of algebraic varieties.*

(2) If $[f]$ *is a Shimura immersion, then it is finite.*

(3) If $[f]$ *is a Shimura submersion, then it is surjective.*

Proof.

(1) If Γ_2 is torsion free, this follows from the Borel extension theorem [Bor72]. Otherwise, let Γ_2' be a torsion free congruence subgroup which is normal and of finite index in Γ_2. By Lemma 2.2(ii), $f^{-1}(\Gamma_2') \cap \Gamma_1$ is a congruence subgroup of $G_1(\mathbb{Q})$. Hence it contains a congruence subgroup Γ_1' which is normal and of finite index in Γ_1.

We then have a diagram

$$\begin{array}{ccc} \mathrm{Sh}_{\Gamma_1'}(G_1, X_1^+) & \xrightarrow{[f]} & \mathrm{Sh}_{\Gamma_2'}(G_2, X_2^+) \\ \downarrow & & \downarrow \\ \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+) & \xrightarrow{[f]} & \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+) \end{array}$$

in which the top arrow is a morphism of algebraic varieties and the vertical arrows are quotients by finite group actions. It follows that the bottom arrow is a morphism of algebraic varieties.

(2) Using construction 2(i) and Lemma 2.3, we may replace G_2 by its adjoint.

Let $G_3 = f(G_1) \subset G_2$. According to construction (3), we can factor $[f]$ as

$$\mathrm{Sh}_{\Gamma_1}(G_1, X_1^+) \xrightarrow{[f_0]} \mathrm{Sh}_{\Gamma_3}(G_3, X_3^+) \xrightarrow{[\iota]} \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+)$$

where $[f_0]$ is a Shimura covering and $[\iota]$ is a Shimura embedding with $\Gamma_3 = \Gamma_2 \cap G_3(\mathbb{Q})_+$.

We prove first that the Shimura covering $[f_0]$ is finite. By [Mil05] Proposition 5.7, $f_{0*}\colon X_1^+ \to X_3^+$ is a bijection, and the action of $G_1(\mathbb{R})_+$ on X_1^+ factors through $G_3(\mathbb{R})_+$. Hence $[f_0]$ factors as

$$\Gamma_1\backslash X_1^+ \to f_0(\Gamma_1)\backslash X_3^+ \to \Gamma_3\backslash X_3^+$$

where the first map is a homeomorphism and the second is a quotient by a finite group action. Thus $[f_0]$ is finite.

Now we prove that the Shimura embedding $[\iota]$ is finite. This is Proposition 3.8(a) of [Pin89]. Pink's proof appears to be essentially correct despite beginning with an incorrect claim that the centre of G_1 maps into the centre of G_2. In fact the preimage of the centre of G_2 is contained in the centre of G_1; hence we may quotient by the centre of G_2 to assume that G_2 is an adjoint group, but we may not assume that G_1 is adjoint. Here is a corrected version of his proof.

Let $\Gamma_3^\dagger = \Gamma_3 \cap G_3^{\mathrm{der}}(\mathbb{R})_+$. Fix a point $x_3 \in X_3^+$ and let $x_2 = \iota_* x_3 \in X_2^+$. The actions on x_2 and x_3 induce maps $G_2(\mathbb{R})_+ \to X_2^+$ and $G_3^{\mathrm{der}}(\mathbb{R})_+ \to X_3^+$ respectively. These give a commuting square

$$\begin{array}{ccc} \Gamma_3^\dagger\backslash G_3^{\mathrm{der}}(\mathbb{R})_+ & \longrightarrow & \Gamma_2\backslash G_2(\mathbb{R})_+ \\ \downarrow & & \downarrow \\ \Gamma_3\backslash X_3^+ & \xrightarrow{[\iota]} & \Gamma_2\backslash X_2^+ \end{array}$$

The top map is proper by [Rag72] Proposition 10.15. Raghunathan's proposition is only stated for GL_n on the right but it works for all linear algebraic groups G_2. The proposition requires the group on the left to have no characters defined over $\mathbb{Q}$, which is why we use G_3^{der} instead of G_3 itself.

The right vertical map is also proper because it is a quotient by the action of a compact group. The left vertical map is surjective because $G_3^{\mathrm{der}}(\mathbb{R})_+ \to X_3^+$ is the composition of surjections

$$G_3^{\mathrm{der}}(\mathbb{R})_+ \to G_3^{\mathrm{ad}}(\mathbb{R})_+ \to X_3^+.$$

A diagram chase shows that the bottom map is proper.

Now consider the commuting square

$$\begin{array}{ccc} X_3^+ & \longrightarrow & X_2^+ \\ \downarrow & & \downarrow \\ \Gamma_3\backslash X_3^+ & \xrightarrow{[\iota]} & \Gamma_2\backslash X_2^+ \end{array}$$

The top map is an injection, the right vertical map has countable fibres and the left vertical map is a surjection. Hence the bottom map has countable fibres. Since it is a morphism of algebraic varieties, it must have finite fibres.

Properness in the analytic topology and finite fibres are sufficient to imply that a morphism of complex algebraic varieties is finite in the algebraic sense.

(3) It suffices to show that $X_1^+ \to X_2^+$ is surjective. By [Mil05] Proposition 5.7, we may replace G_2 by its adjoint group. Then $f\colon G_1 \to G_2$ is surjective. By [Mil05] Proposition 5.1, $G_1(\mathbb{R})_+ \to G_2(\mathbb{R})_+$ is surjective and so $X_1^+ \to X_2^+$ is also surjective. □

2.8 Shimura varieties of Hodge type and of abelian type

A connected Shimura variety S is said to be **of Hodge type** if there exists a Shimura embedding $S \to \mathcal{A}_g$ for some g. It is **of abelian type** if there exists a Shimura covering $S' \to S$ such that S' is of Hodge type.

Shimura varieties of Hodge type can be interpreted as moduli spaces for abelian varieties with certain additional structure (Hodge tensors and level structure). Shimura varieties of abelian type are a generalisation of Shimura varieties of Hodge type whose main interest is that we have more flexibility in choosing the centre – most usefully, if S is a Shimura variety of Hodge type then the associated adjoint Shimura variety S^{ad} is of abelian type (adjoint Shimura varieties are never of Hodge type themselves).

Deligne classified Shimura varieties of adjoint abelian type in terms of group theory. If the underlying group is a direct product of groups over $\mathbb{Q}$, then the Shimura datum likewise splits as a product, and it is of abelian type if and only if all the factors are. So when classifying Shimura varieties of abelian type, it suffices to consider $\mathbb{Q}$-simple adjoint groups.

Let (G, X^+) be a connected Shimura datum, with G a $\mathbb{Q}$-simple adjoint group. The absolutely simple factors of $G_{\bar{\mathbb{Q}}}$ are permuted transitively by $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ so all have the same type in the Dynkin classification. Thus we can talk about the Dynkin type of G.

We must also consider the irreducible factors of the Hermitian symmetric domain X^+. There are two different irreducible Hermitian symmetric domains associated with a group of type D_n for $n \geq 5$, which are called $D_n^{\mathbb{R}}$ and $D_n^{\mathbb{H}}$. In the case of D_4, there is only one Hermitian symmetric domain, but we can still distinguish two cases $D_4^{\mathbb{R}}$ and $D_4^{\mathbb{H}}$ by taking into account the how $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts on the Dynkin diagram of G. In order for (G, X^+) to be of abelian type when G has type D_n, all irreducible factors of X^+ must be of the same type. There are also multiple irreducible Hermitian symmetric domains associated with groups of type A_n, but mixing these does not prevent a Shimura datum being of abelian type.

Proposition 2.5 ([Del79] Proposition 2.3.10)**.** *Let (G, X^+) be a connected Shimura datum, with G a $\mathbb{Q}$-simple adjoint group. Then (G, X^+) is of abelian type if and only if: G has type A_n, B_n, C_n or D_n; and, if G has type D_n then all irreducible factors of X^+ have the same type (all $D_n^{\mathbb{R}}$ or all $D_n^{\mathbb{H}}$).*

2.9 Special subvarieties

A subvariety of a connected Shimura variety S is called a **special subvariety** if it is the image of some Shimura morphism $S' \to S$.

Suppose we have a special subvariety

$$Z \subset S = \mathrm{Sh}_\Gamma(G, X^+).$$

Let $f\colon (G', X'^+) \to (G, X^+)$ be a morphism of Shimura data such that $[f]$ has image Z. Let

$$X_H^+ = f_*(X'^+) \subset X^+.$$

Observe that X_H^+ is a connected component of $\pi^{-1}(Z)$, where $\pi\colon X^+ \to S$ is the uniformisation map.

Let H be the smallest subgroup of G such that every Hodge parameter in X_H^+ factors through H. Let $\Gamma_H = H(\mathbb{Q})_+ \cap \Gamma$. Then (H, X_H^+) is a connected Shimura datum and $\mathrm{Sh}_{\Gamma_H}(H, X_H^+) \to S$ is a Shimura embedding whose image is Z. Thus every special subvariety is the image of some Shimura embedding $S_H \to S$.

The group H defined above is called the **generic Mumford–Tate group** of Z. It is not uniquely defined, as we could replace f by some other morphism of Shimura data such that $[f]$ has image Z. This would replace X_H^+ by a different connected component of $\pi^{-1}(Z)$, and thus H is defined up to conjugation by Γ.

Thus we get a map

$$\{\text{special subvarieties of } S\} \to \{\Gamma\text{-conjugacy classes of subgroups of } G\}.$$

This map is finite-to-one because for any subgroup $H \subset G$, the set of Hodge parameters in X^+ which factor through H has finitely many connected components. It is not in general a surjective map: some subgroups of G do not contain the image of any Hodge parameters in X^+, while others contain such Hodge parameters but are larger than their generic Mumford–Tate group.

Because the generic Mumford–Tate group of a special subvariety must be defined over $\mathbb{Q}$, there are countably many special subvarieties of S. Hence there are points of S which are not contained in any proper special subvariety. A point or subvariety of S is called **Hodge generic** if it is not contained in any special subvariety of S other than S itself. A point is Hodge generic in S if and only if its Mumford–Tate group is equal to the generic Mumford–Tate group M of S itself. (Note that the generic Mumford–Tate group M need not be equal to G. By condition (SV3), M must contain G^{der}, but it need not contain all of the centre of G.)

Every irreducible component of the intersection of some special subvarieties of S is a special subvariety. Hence every subvariety of S is contained in a unique smallest special subvariety.

Of particular importance are **special points**, which are zero-dimensional special subvarieties. A special subvariety has dimension zero if and only if its generic Mumford–Tate group is a torus. If $S = \mathcal{A}_g$, then special points are the moduli of abelian varieties with complex multiplication.

An example of a positive-dimensional special subvariety is the image of the Shimura embedding $\mathcal{A}_1 \times \mathcal{A}_1 \to \mathcal{A}_2$ considered above. As mentioned above, this special subvariety of $\mathcal{A}_2$ is not isomorphic to $\mathcal{A}_1 \times \mathcal{A}_1$ itself but rather to the symmetric square of $\mathcal{A}_1$. It can be interpreted as the moduli space of principally polarised abelian surfaces which are isomorphic (as polarised abelian varieties) to products of two elliptic curves.

2.10 Weakly special subvarieties

The André–Pink conjecture concerns a slightly more general notion than that of special subvariety, namely that of weakly special subvariety. Weakly special subvarieties were named by Pink [Pin05a], although the concept was essentially considered first by Moonen [Moo98]. In particular, Moonen realised that André's conjecture on subvarieties with a dense intersection with a Hecke orbit would be false if its conclusion was that the subvariety was special instead of weakly special.

A subvariety of a connected Shimura variety S is a **weakly special subvariety** if it is an irreducible component of $[\iota]([\varphi]^{-1}(s_2))$ for some Shimura morphisms

$$\begin{array}{ccc} S' & \xrightarrow{[\iota]} & S \\ {\scriptstyle [\varphi]}\downarrow & & \\ S_2 & & \end{array}$$

and some point $s_2 \in S_2$.

Observe that weakly special subvarieties always come in families, parametrised by the points of a finite cover of S_2. The union of this family of weakly special subvarieties is the special subvariety $[\iota](S')$.

By [Pin05a] Proposition 4.4, we may assume that $[\iota]$ is a Shimura embedding and $[\varphi]$ is a Shimura submersion in the definition of a weakly special subvariety.

We shall now show that the above definition is equivalent to two other definitions: the definition we gave in the introduction and a description used by Moonen [Moo98].

Proposition 2.6. *Let $S = \mathrm{Sh}_\Gamma(G, X^+)$. A subvariety $Z \subset S$ is weakly special if and only if there exist a connected Shimura subdatum $(H, X_H^+) \subset (G, X^+)$ and an element $g \in G(\mathbb{R})^+$ normalising H^{der} such that Z is the image in S of $g.X_H^+ \subset X^+$.*

Proof. Suppose first that Z is weakly special. Let (G', X'^+) and (G_2, X_2^+) be the connected Shimura data associated with S' and S_2 respectively.

Let $Z' \subset S'$ be an irreducible component of $[\varphi]^{-1}(s_2)$ which maps onto Z, and let $W \subset X'^+$ be a connected component of the preimage in X'^+ of Z'.

Choose a special point $\tilde{t_2} \in X_2^+$ in the image of X'^+ and let $H_2 \subset G_2$ be the Mumford–Tate group of $\tilde{t_2}$. Let $H' = \varphi^{-1}(H_2)^\circ$. Since H' is reductive and $\varphi(H')$ is commutative, H'^{der} is equal to the derived group of $(\ker \varphi)^\circ$. In particular H'^{der} is normal in G'.

Let $X_{H'}^+$ be a connected component of the preimage in X'^+ of $\tilde{t_2}$. Observe that $(H', X_{H'}^+)$ is a connected Shimura subdatum of (G', X'^+).

Choose an element $g' \in G'(\mathbb{R})^+$ such that $g'.X_{H'}^+ \cap W$ is non-empty. Since $\phi_*\colon X'^+ \to X_2^+$ is $G'(\mathbb{R})^+$-equivariant and $X_{H'}^+$ and W are each a connected component of the preimage of a point in X_2^+, we must have

$$g'.X_{H'}^+ = W.$$

Furthermore g' normalises H'^{der} because H'^{der} is normal in G'.

Letting $g = \iota(g')$, $H = \iota(H')$ and $X_H^+ = \iota_* X_{H'}^+$ we get a connected Shimura subdatum $(H, X_H^+) \subset (G, X^+)$ and an element $g \in G(\mathbb{R})^+$ normalising H^{der} such that Z is the image in S of $g.X_H^+$.

In the converse direction, suppose that Z is the image in S of $g.X_H^+$. Let $G' \subset G$ be the normaliser of H^{der}; note that $H \subset G'$. Let $G_2 = G'/H^{\mathrm{der}}$, and let $\iota\colon G' \to G$ and $\varphi\colon G' \to G_2$ be the obvious homomorphisms.

By hypothesis, $g \in G'(\mathbb{R})$. By [PR94] Theorem 7.7 there exists $g_0 \in G'(\mathbb{Q})$ such that $gg_0^{-1} \in G'(\mathbb{R})^+$. Replace g by gg_0^{-1}, H by $g_0 H g_0^{-1}$ and X_H^+ by $g_0.X_H^+$. This does not change H^{der}, so does not change G', and hence we may proceed, assuming that $g \in G'(\mathbb{R})^+$.

Let X'^+ be the connected component of $X^+[G']$ containing X_H^+ and let $X_2^+ = \varphi_* X'^+$. Then (G', X'^+) and (G_2, X_2^+) are connected Shimura data and for suitable choices of congruence subgroups, we get Shimura varieties S' and S_2 and Shimura morphisms

$$\begin{array}{ccc} S' & \xrightarrow{[\iota]} & S \\ {\scriptstyle [\varphi]}\big\downarrow & & \\ S_2 & & \end{array}$$

Now $\varphi(H)$ is a torus so the image of X_H^+ in X_2^+ is a point $\tilde{t_2}$ (whose image in S_2 is a special point). Let

$$\tilde{s_2} = \varphi(g).\tilde{t_2} \in X_2^+$$

and let s_2 be the image of $\tilde{s_2}$ in S_2. Then $g.X_H^+$ is a connected component of $\varphi_*^{-1}(\tilde{s_2})$ and so Z is an irreducible component of $[\iota]([\varphi]^{-1}(s_2))$ as required. □

Proposition 2.7. *Let $S = \mathrm{Sh}_\Gamma(G, X^+)$. A subvariety $Z \subset S$ is weakly special if and only if there exist a connected Shimura subdatum $(H, X_H^+) \subset (G, X^+)$ and a decomposition of the adjoint Shimura datum $(H^{\mathrm{ad}}, X_H^{\mathrm{ad}+})$ as a direct product*

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

as well as a point $x_2 \in X_2$ such that Z is the image in S of $X_1^+ \times \{x_2\}$.

Proof. First suppose that Z is weakly special. Assume that $[\iota]$ is a Shimura embedding and $[\varphi]$ is a Shimura submersion. Let (H, X_H^+) and (H_2, X_2^+) be the connected Shimura data associated with S' and S_2 respectively.

We may replace (H_2, X_2^+) by $(H_2^{\mathrm{ad}}, X_2^{\mathrm{ad}+})$; since the Shimura covering $\pi \colon S_2 \to S_2^{\mathrm{ad}}$ is finite, each irreducible component of a fibre of $[\varphi]$ is also an irreducible component of a fibre of $[\pi \circ \varphi]$. Thus we may suppose that H_2 is an adjoint group.

Then we have a surjective homomorphism of adjoint groups $H^{\mathrm{ad}} \to H_2$, so H^{ad} is isomorphic to a direct product $H_1 \times H_2$ for some other group H_1. It is straightforward to finish the proof.

The converse is obvious. □

Moonen ([Moo98] Theorem 4.3) also showed that weakly special subvarieties are precisely the images in S of totally geodesic subvarieties in X^+. Another natural description of weakly special subvarieties, due to Ullmo and Yafaev, is that they are the algebraic subvarieties of S for which a component of the preimage in X^+ is also algebraic (see section 3.4 for a precise definition of algebraic subvarieties of X^+).

All special subvarieties are weakly special. To see this, observe that the pair consisting of the trivial algebraic group $\{1\}$ and the unique Hodge parameter $\mathbb{S} \to \{1\}$ is a connected Shimura datum. Furthermore $\{1\}$ is a congruence subgroup of itself. Taking S_2 to be the resulting one-point Shimura variety shows that every special subvariety is weakly special.

A weakly special subvariety is special if and only if $s_2 \in S_2$ is a special point, and this occurs if and only if the weakly special subvariety contains a special point. The concept of "weakly special point" is uninteresting because all points are weakly special (taking both $[\iota]$ and $[\phi]$ to be the identity).

Examples. The Shimura immersion $[\iota]\colon \mathcal{A}_1 \times \mathcal{A}_1 \to \mathcal{A}_2$ considered previously gives a family of weakly special subvarieties: for each elliptic curve E_0, the subvariety $[\iota](\{E_0\} \times \mathcal{A}_1) \subset \mathcal{A}_2$ is weakly special. In modular terms, this is the variety which parametrises principally polarised abelian surfaces isomorphic to a product $E_0 \times E$ for a varying elliptic curve E. This is a special subvariety if and only if E_0 has complex multiplication.

As we have seen, $\mathcal{A}_1 \times \mathcal{A}_1$ has generic Mumford–Tate group

$$H = \mathbb{G}_m.(\mathrm{SL}_2 \times \mathrm{SL}_2).$$

Thus, in accordance with Proposition 2.7, H^{ad} is a direct product $\mathrm{PSL}_2 \times \mathrm{PSL}_2$.

Not all weakly special subvarieties of $\mathcal{A}_g$ parametrise abelian varieties which split as direct products (even up to isogeny). According to [Sai93] Corollary 8.4 counter-examples cannot occur for $g \leq 7$. A counter-example with $g = 8$ was found (in a different context) by Deligne and described by Faltings [Fal83a] and André [And92]. André attributed this counter-example, seemingly incorrectly, to Borovoi. This counter-example is fully described in the context of totally geodesic subvarieties (equivalent to weakly special subvarieties) in [Moo98].

Here is an outline of the construction. Let F be a real quadratic field with archimedean places σ_1, σ_2. Let D_1 and D_2 be quaternion algebras over F such that D_1 is split at σ_1 and unsplit at σ_2, while D_2 is split at σ_2 and unsplit at σ_1. Then there is a Shimura variety of Hodge type, embedded in $\mathcal{A}_8$, such that the adjoint of the underlying group is $\mathrm{PSL}_1(D_1) \times \mathrm{PSL}_1(D_2)$, giving rise to a family of weakly special subvarieties of $\mathcal{A}_8$ on which the generic abelian variety is simple.

3 The Pila–Zannier strategy

This chapter summarises the ingredients from model theory and transcendence theory which go into the Pila–Zannier strategy for proving conjectures on unlikely intersections.

We begin with a rapid overview of o-minimal structures. We will not deal with such structures directly in this thesis: all we need to know is the definability of the uniformisation map $\mathcal{H}_g \to \mathcal{A}_g$ in $\mathbb{R}_{\mathrm{an,exp}}$ and a suitably strong version of the Pila–Wilkie theorem. Therefore we do not go into the properties of definable sets in o-minimal models in detail but pass directly to the statement of the Pila–Wilkie theorem.

We then discuss the Ax–Lindemann–Weierstrass theorem on the transcendence of the exponential function and its analogue for Shimura varieties, the hyperbolic Ax–Lindemann–Weierstrass conjecture. To show how the ingredients fit together, we end the chapter with an outline of the simplest application of the Pila–Zannier strategy, namely the torus analogue of the Manin–Mumford conjecture.

3.1 O-minimality

In this section we recall the definition and some important examples of o-minimal structures. A general introduction to o-minimality can be found in [vdD98]. For an account which focusses on the Pila–Wilkie theorem and Diophantine applications, see [Sca12].

A subset of $\mathbb{R}^n$ is **semialgebraic** if it can be defined by a finite boolean combination of polynomial inequalities.

A **structure $\mathcal{S}$ (over $\mathbb{R}$)** is a sequence $\mathcal{S}_n$ of collections of subsets of $\mathbb{R}^n$ (for each natural number n) such that

1. $\mathcal{S}_n$ is closed under finite unions, intersections and complements;
2. $\mathcal{S}_n$ contains all semialgebraic subsets of $\mathbb{R}^n$;
3. if $A \in \mathcal{S}_m$ and $B \in \mathcal{S}_n$ then $A \times B \in \mathcal{S}_{m+n}$;
4. if $m \geq n$ and $A \in \mathcal{S}_m$ then $\pi(A) \in \mathcal{S}_n$, where $\pi \colon \mathbb{R}^m \to \mathbb{R}^n$ is projection onto the first n coordinates.

The sets in $\mathcal{S}_n$ are called the **definable sets** of the structure $\mathcal{S}$. A function $f \colon A \to B$ for definable sets $A \subset \mathbb{R}^m$, $B \subset \mathbb{R}^n$ is said to be **definable** if its graph is a definable subset of $\mathbb{R}^{m+n}$.

A structure is **o-minimal** if every set in $\mathcal{S}_1$ (that is, every definable subset of $\mathbb{R}$ itself) is a finite union of points and intervals. The basic example of an o-minimal structure is the structure of semialgebraic sets. The only non-obvious condition for semialgebraic sets is closure under projections, which is the Tarski–Seidenberg theorem.

All definable sets in an o-minimal structure are topologically well-behaved: for example they have finitely many connected components and a finite cell decomposition and can be stratified and triangulated (this is proved in [vdD98]).

The only o-minimal structure we shall use in this thesis is the structure $\mathbb{R}_{\text{an,exp}}$ generated by the graphs of restricted analytic functions and the real exponential function. A **restricted analytic function** is a function $[0,1]^n \to \mathbb{R}$ which extends to a real analytic function on some open neighbourhood of $[0,1]^n$. This structure was shown to be o-minimal by van den Dries and Miller [vdDM94].

Throughout this thesis, when the word **definable** is used without specifying a structure, it will mean definable in $\mathbb{R}_{\text{an,exp}}$.

Let $\pi \colon \mathcal{H}_g \to \mathcal{A}_g \hookrightarrow \mathbb{P}^N(\mathbb{C})$ be the map constructed by Baily and Borel [BB66]. Let $\mathcal{F}_g$ denote the Siegel fundamental set in $\mathcal{H}_g$; this is a semialgebraic fundamental set for the action of $\mathrm{Sp}_{2g}(\mathbb{Z})$. Crucially for us, Peterzil and Starchenko have shown that the restriction of π to $\mathcal{F}_g$ is definable in $\mathbb{R}_{\text{an,exp}}$ [PS13]. This implies that, if $Z \subset \mathcal{A}_g$ is an algebraic subvariety, then $\pi^{-1}(Z) \cap \mathcal{F}_g$ is definable.

It seems that an analogous result should hold for other Shimura varieties, i.e. the restriction of the Baily–Borel map $X^+ \to \Gamma\backslash X^+ \to \mathbb{P}^N(\mathbb{C})$ to a suitable fundamental set should be definable in $\mathbb{R}_{\text{an,exp}}$. Klingler, Ullmo and Yafaev have recently announced a proof of this.

3.2 The Pila–Wilkie theorem

We never use o-minimality directly in this thesis. We use it only for the Pila–Wilkie theorem, which we shall now state. We require a strong version of this theorem using definable blocks, but to illustrate the core of the theorem we begin with a simple version.

The general idea of the theorem is that a transcendental definable set should only contain few rational points. However a transcendental set may contain semialgebraic subsets of positive dimension, and these may contain many rational points, so we need to exclude such subsets. We therefore make the following definition: if X is a subset of $\mathbb{R}^n$, then the **semialgebraic part** X^{alg} of X is the union of all connected semialgebraic subsets of X of positive dimension.

The idea that there are few rational points is made precise by counting points up to a given height. It does not matter much how we define the height, but for concreteness we shall adopt the following naïve definition: if

$$x = \left(\frac{a_1}{b_1}, \frac{a_2}{b_2}, \dots, \frac{a_n}{b_n}\right) \in \mathbb{Q}^n,$$

where each of the fractions is written in lowest terms, then the **height** of x is

$$\max(|a_1|, |b_1|, |a_2|, |b_2|, \dots, |a_n|, |b_n|).$$

For any set $X \subset \mathbb{R}^n$, we write $X(\mathbb{Q}, T)$ for the set of rational points in X of height at most T.

Theorem 3.1 ([PW06] Theorem 1.8)**.** *Let $X \subset \mathbb{R}^n$ be a definable set and let $\epsilon > 0$. There is a constant $c = c(X, \epsilon)$ such that for all $T \geq 1$,*

$$\#(X - X^{\mathrm{alg}})(\mathbb{Q}, T) \leq cT^{\epsilon}.$$

The set X^{alg} need not itself be semialgebraic or even definable, but stronger versions of the Pila–Wilkie theorem assert that most of the rational points of X are contained in nicely behaved subsets of X^{alg}. This can be made this precise using definable blocks.

A definable block is a definable set which is connected and almost semialgebraic. We use the definition from [Pil11], which is called a "basic block" in [Pil09a]: a **(definable) block** of dimension w in $\mathbb{R}^n$ is a connected definable subset $W \subset \mathbb{R}^n$ of dimension w, regular at every point, such that there is a semialgebraic set $A \subset \mathbb{R}^n$ of dimension w, regular at every point, with $W \subset A$.

A **definable block family** is a definable subset W of $\mathbb{R}^n \times \mathbb{R}^m$ such that for each $\eta \in \mathbb{R}^m$, the fibre $W_\eta = \{x \in \mathbb{R}^n \mid (x, \eta) \in W\}$ is a definable block.

We have simplified the statement of the following theorem with respect to that given by Pila by considering only a single set X instead of a definable family, as this is all that is required in this thesis.

Theorem 3.2 ([Pil09a] Theorem 3.5)**.** *Let $X \subset \mathbb{R}^n$ be a definable set and $\epsilon > 0$. There are a finite number $J = J(X, \epsilon)$ of definable block families*

$$\mathcal{W}^{(j)} \subset \mathbb{R}^n \times \mathbb{R}^m, \ j = 1, \dots, J$$

and a constant $c = c(X, \epsilon)$ such that:

(1) for all $\eta \in \mathbb{R}^m$,

$$\mathcal{W}^{(j)}_\eta \subset X;$$

(2) for all $T \geq 1$, $X(\mathbb{Q}, T)$ is contained in the union of at most cT^ϵ definable blocks of the form $\mathcal{W}^{(j)}_\eta$ (for some $j \in \{1, \dots, J\}$ and $\eta \in \mathbb{R}^m$).

3.3 The Ax–Lindemann–Weierstrass theorem

We wish to apply the Pila–Wilkie theorem to the preimage $\widetilde{Z}$ in $\mathcal{H}_g$ of an algebraic subvariety of $\mathcal{A}_g$ (or more precisely to the intersection of this with a fundamental set). In order to do this, we must understand the semialgebraic part $\widetilde{Z}^{\mathrm{alg}}$ of such a set. This is the subject of the hyperbolic Ax–Lindemann–Weierstrass conjecture. In order to motivate this conjecture, we shall begin with the classical Lindemann–Weierstrass theorem and with the Ax–Lindemann–Weierstrass theorem for the exponential function.

Theorem 3.3 (Lindemann–Weierstrass)**.** *Let $z_1, \dots, z_n$ be algebraic numbers. If $z_1, \dots, z_n$ are linearly independent over $\mathbb{Q}$, then $\exp(z_1), \dots, \exp(z_n)$ are algebraically independent over $\mathbb{Q}$.*

The following analogue of this result for functions on a complex algebraic variety was given the name of Ax–Lindemann–Weierstrass by Pila. Since it does not appear in this form in [Ax71], we show how to deduce it from the so-called Ax–Schanuel theorem. In this theorem, we say that $f_1, \dots, f_n$ are **$\mathbb{Q}$-linearly independent modulo constants** if there do not exist $a_1, \dots, a_n \in \mathbb{Q}$, not all zero, such that $a_1 f_1 + \dots + a_n f_n \in \mathbb{C}$.

Theorem 3.4 (Ax–Lindemann–Weierstrass)**.** *Let W be an irreducible complex algebraic variety and $f_1, \dots, f_n \in \mathbb{C}[W]$ regular functions on W. If $f_1, \dots, f_n$ are $\mathbb{Q}$-linearly independent modulo constants, then the functions*

$$\exp f_1, \dots, \exp f_n \colon W \to \mathbb{C}$$

are algebraically independent over $\mathbb{C}$.

Proof. Let $f \colon W \to \mathbb{C}^n$ denote the function $(f_1, \dots, f_n)$, and let $V \subset \mathbb{C}^n$ be the Zariski closure of the image of f. Let $m = \dim W$ and $r = \dim V$.

By [Har77] Corollary II.8.16 and Proposition III.10.6 there is a Zariski open subset $U \subset W$ such that U is non-singular and for all $w \in U$, the morphism of tangent spaces $df_{|w} \colon T_{W,w} \to T_{V,f(w)}$ has rank at least r.

Since U is non-singular, we may choose an isomorphism of complex manifolds $\psi \colon \Delta^m \to U_0$ from the polydisc to an open subset of U. Let $y_i = f_i \circ \psi \colon \Delta^m \to \mathbb{C}$.

The condition that the f_i are $\mathbb{Q}$-linearly independent modulo constants implies that the functions $y_i - y_i(0)$ are $\mathbb{Q}$-linearly independent. Hence we can apply Corollary 2 of Theorem 3 in [Ax71] to deduce that

$$\mathrm{trdeg}_{\mathbb{C}}(y_1, \dots, y_n, \exp y_1, \dots, \exp y_n) \geq n + \mathrm{rk} \left(\frac{\partial y_i}{\partial t_j} \right)_{1 \leq i \leq n, 1 \leq j \leq m}$$

Our condition on the rank of $df_{|w}$ for $w \in U$ implies that $\mathrm{rk}(\partial y_i/\partial t_j) \geq r$. Furthermore, $\mathrm{trdeg}_{\mathbb{C}}(y_1, \dots, y_n) = \dim V = r$. Hence the above inequality implies

$$r + \mathrm{trdeg}_{\mathbb{C}}(\exp y_1, \dots, \exp y_n) \geq n + r$$

or in other words $\exp y_1, \dots, \exp y_n$ are algebraically independent, as required. □

We can translate this theorem into geometric terms as follows.

Theorem 3.5. *Let $\pi = (\exp, \exp, \dots, \exp) \colon \mathbb{C}^n \to (\mathbb{C}^\times)^n$. Let W be an irreducible algebraic subvariety of $\mathbb{C}^n$.*

If W is not contained in a translate of a proper $\mathbb{Q}$-linear subspace of $\mathbb{C}^n$, then $\pi(W)$ is Zariski dense in $\mathbb{G}_{m,\mathbb{C}}^n$.

The form in which the Ax–Lindemann–Weierstrass theorem is used in the Pila–Zannier method is as follows.

Corollary 3.6. *Let $\pi = (\exp, \exp, \dots, \exp) \colon \mathbb{C}^n \to (\mathbb{C}^\times)^n$. Let Z be an algebraic subvariety of $\mathbb{G}_{m,\mathbb{C}}^n$.*

If W is a maximal irreducible algebraic subvariety of $\mathbb{C}^n$ contained in $\pi^{-1}(Z)$, then W is a translate of a $\mathbb{Q}$-linear subspace of $\mathbb{C}^n$. Equivalently, $\pi(W)$ is a translate of an algebraic subgroup of $\mathbb{G}_{m,\mathbb{C}}^n$.

Proof. By translating, we may assume that $0 \in W$. Let $V \subset \mathbb{C}^n$ be the smallest $\mathbb{Q}$-linear subspace containing W and let $G = \pi(V)$. Then $V \cong \mathbb{C}^r$ and $G \cong \mathbb{G}_{m,\mathbb{C}}^r$ for some r so we may apply Theorem 3.5 to $\pi \colon V \to G$ and W to deduce that $\pi(W)$ is Zariski dense in G.

It follows that $G \subset Z$ so $V \subset \pi^{-1}(Z)$. Since V is an irreducible algebraic subvariety of $\mathbb{C}^n$, the maximality of W implies that $W = V$. □

3.4 Hyperbolic Ax–Lindemann–Weierstrass conjecture

The hyperbolic Ax–Lindemann–Weierstrass conjecture is the analogue of the Ax–Lindemann–Weierstrass theorem in which the exponential map $\mathbb{C}^n \to \mathbb{G}_{m,\mathbb{C}}^n$ is replaced by the uniformisation map $X^+ \to S$ of a Shimura variety. In order to state the conjecture, we must first define what we mean by algebraic subvarieties of a Hermitian symmetric domain X^+; this definition is not entirely clear because X^+ is a complex manifold but is not quasi-projective.

Fix a connected semialgebraic open subset $U \subset \mathbb{C}^n$ (in our application, U will be a suitable realisation of X^+). We say that $W \subset U$ is an **irreducible algebraic subset** of U if it is a connected component of $V \cap U$ for some complex algebraic variety $V \subset \mathbb{C}^n$.

For any subset $Y \subset U$, we define the **complex algebraic part** of Y, denoted Y^{ca}, to be the union of all positive-dimensional irreducible algebraic subsets of U contained in Y. Clearly Y^{ca} is contained in the semialgebraic part Y^{alg}. In fact these are equal if Y is complex analytic.

Lemma 3.7 ([Pil09b] Lemma 2.1)**.** *Let $U \subset \mathbb{C}^n$ be a connected semialgebraic open set. If $Y \subset U$ is complex analytic then*

$$Y^{\mathrm{ca}} = Y^{\mathrm{alg}}.$$

A **realisation** of a Hermitian symmetric domain X^+ (associated with the reductive group G) is an open analytic subset $\mathcal{X}^+$ of a complex algebraic variety, equipped with an isomorphism of complex manifolds $X^+ \to \mathcal{X}^+$ and such that the induced action of $G(\mathbb{R})_+$ on $\mathcal{X}^+$ is semialgebraic. Every Hermitian symmetric domain has a Harish-Chandra realisation as a bounded domain in the tangent space of a base point $x_0 \in X^+$. In the case of the Siegel upper half-space, $\mathcal{H}_g \subset \mathrm{M}_g(\mathbb{C})$ is a realisation which is isomorphic to the Harish-Chandra realisation.

A choice of realisation allows us to talk about irreducible algebraic subsets of a Hermitian symmetric domain X^+, using the above definition. According to [Ull12] Lemma 2.1, realisations are semialgebraic sets and isomorphisms of realisations are semialgebraic maps. Hence if $Y \subset X^+$ is complex analytic, it follows from Lemma 3.7 that Y^{ca} is independent of the choice of realisation (provided that the realisation is isomorphic to the Harish-Chandra realisation).

Conjecture 3.8 (Hyperbolic Ax–Lindemann–Weierstrass)**.** *Let (G, X^+) be a connected Shimura datum and $S = \mathrm{Sh}_\Gamma(G, X^+)$ a connected Shimura variety. Let $\pi \colon X^+ \to S$ be the uniformisation map. Choose a realisation of X^+ isomorphic to the Harish-Chandra realisation.*

Let Z be an algebraic subvariety of S.

If W is a maximal irreducible algebraic subset of X^+ contained in $\pi^{-1}(Z)$, then $\pi(W)$ is a weakly special subvariety of S.

Ullmo and Yafaev [UY11] have proved the important case of this conjecture in which W consists of an entire irreducible component of $\pi^{-1}(Z)$: in other words if $Z \subset S$ is algebraic and some irreducible component of $\pi^{-1}(Z)$ is algebraic, then Z is weakly special. In particular this proves the full hyperbolic Ax–Lindemann–Weierstrass conjecture when Z is a curve.

Ullmo and Yafaev have also proved the hyperbolic Ax–Lindemann–Weierstrass conjecture for compact Shimura varieties S [UY13b]. Pila and Tsimerman have announced a proof of the conjecture for $S = \mathcal{A}_g$ [PT12], and Klingler, Ullmo and Yafaev have recently announced a proof of the full conjecture.

Theorem 3.9 ([PT12])**.** *The hyperbolic Ax–Lindemann–Weierstrass conjecture holds for $\mathcal{A}_g$.*

Let Z^{ws} denote **weakly special part** of Z, i.e. the union of the weakly special subvarieties of S contained in Z. The theorem of Pila and Tsimerman has the following corollary.

Corollary 3.10. *Let $\pi \colon \mathcal{H}_g \to \mathcal{A}_g$ be the uniformisation map. Let Z be an algebraic subvariety of $\mathcal{A}_g$, and let $\widetilde{Z} = \pi^{-1}(W)$.*

Then $Z^{\mathrm{ws}} = \pi(\widetilde{Z}^{\mathrm{ca}})$.

3.5 Proof of the Manin–Mumford conjecture for tori

We now sketch a simple application of the method of Pila and Zannier [PZ08]. This shows the overall structure we will use for the proof of the André–Pink conjecture for curves in $\mathcal{A}_g$. In order to make the sketch as simple as possible, we consider the analogue of the Manin–Mumford conjecture for the torus $\mathbb{G}_{m,\mathbb{C}}^n$.

Theorem 3.11. *Let $G = \mathbb{G}_{m,\mathbb{C}}^n$ and let $Z \subset G$ be an algebraic subvariety. Let Σ denote the set of torsion points of G.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a translate of an algebraic subgroup of G by a torsion point.

Proof.

1. Let $\pi = (\exp(2\pi i z_1), \dots, \exp(2\pi i z_n)) \colon \mathbb{C}^n \to \mathbb{G}_{m,\mathbb{C}}^n$. Let $\mathcal{F}$ denote the fundamental domain

$$\mathcal{F} = \{(z_1, \dots, z_n) \in \mathbb{C}^n \mid 0 \leq \operatorname{Re} z_i < 1 \ \forall i\}.$$

   Then $\pi_{|\mathcal{F}}$ is definable in $\mathbb{R}_{\mathrm{an,exp}}$. Hence

$$\widetilde{Z} = \pi^{-1}(Z) \cap \mathcal{F}$$

   is a definable set.

2. Let $x \in \mathbb{G}_{m,\mathbb{C}}^n$ be a torsion point of order N. Then there is a unique $\tilde{x} \in \mathcal{F}$ lifting x. The coordinates of $\tilde{x}$ are rational numbers in $[0, 1)$ with denominators dividing N, and so $H(\tilde{x}) \leq N$.

3. We can now apply the Pila–Wilkie theorem in its simplest form (Theorem 3.1) to deduce that there exists c such that for all $T \geq 1$, $\pi(\widetilde{Z} - \widetilde{Z}^{\mathrm{alg}})$ contains fewer than $cT^{1/2}$ torsion points of order at most T.

4. If $x \in \mathbb{G}_{m,\mathbb{C}}^n$ is a torsion point of order N, then the coordinates of x generate the cyclotomic field $\mathbb{Q}(\mu_N)$ which has degree $\phi(N)$ over $\mathbb{Q}$.

Since Z contains a dense set of points defined over number fields, Z itself is defined over a number field K. It is easy to show that

$$\phi(N)/[K : \mathbb{Q}] > cN^{1/2} \tag{*}$$

for large enough N.

5. Since $\Sigma \cap Z$ is infinite and the number of torsion points in $\mathbb{G}_{m,\mathbb{C}}^n$ of given order is finite, we can find some N satisfying (*) and such that $\Sigma \cap Z$ contains a point x of order N.

   All $\mathrm{Gal}(\bar{\mathbb{Q}}/K)$-conjugates of x are also torsion points of order N and are also in Z. Hence by (*), Z contains more than $cN^{1/2}$ torsion points of order N and therefore by step 3, $\widetilde{Z}^{\mathrm{alg}}$ is non-empty.

6. By Lemma 3.7, $\widetilde{Z}^{\mathrm{alg}} = \widetilde{Z}^{\mathrm{ca}}$ so $\widetilde{Z}^{\mathrm{ca}}$ is non-empty. Assume now that Z is a curve. It follows that some irreducible component of $\widetilde{Z}$ is algebraic.

7. By Corollary 3.6, Z is a translate of an algebraic subgroup.

8. Subvarieties Z of dimension greater than 1 can be dealt with by an induction on the dimension. □

The same method can be used to prove the Manin–Mumford conjecture for an abelian variety A. In step 1, instead of the exponential map, we use the uniformisation $\pi \colon \mathbb{C}^n \to A$. In order for step 2 to work, we must choose real coordinates on $\mathbb{C}^n$ which map the lattice $\ker \pi$ onto $\mathbb{Z}^{2n}$. Step 4 is replaced by a lower bound for the Galois degrees of torsion points on A, proved by Masser [Mas84a]. In step 7 we use the analogue of the Ax–Lindemann–Weierstrass theorem for abelian varieties ([Ax72] Theorem 3).

In our proof of André–Pink for curves in $\mathcal{A}_g$, we will apply the same outline to the uniformisation $\mathcal{H}_g \to \mathcal{A}_g$. However step 2 cannot be carried out directly as points in the Hecke orbit Σ do not have rational preimages in $\mathcal{H}_g$. They do have rational preimages in $\mathrm{GSp}_{2g}(\mathbb{R})^+$ and so we will apply the Pila–Wilkie theorem there, but this requires us to use the blocks version of Pila–Wilkie. Some calculations are also required to relate the heights of these preimages to the degrees of isogenies. In place of step 4 we deduce a Galois lower bound from the Masser–Wüstholz isogeny theorem. Finally the induction to extend the result from curves to higher-dimensional subvarieties turns out to be more difficult and we have not been able to do this in all cases.

4 Hecke orbits

In this chapter we discuss Hecke orbits in connected Shimura varieties and certain related concepts, namely generalised Hecke orbits, P-Hecke orbits and isogeny classes. If (G, X^+) is a connected Shimura variety, then a Hecke orbit in an associated connected Shimura variety is the image of a $G(\mathbb{Q})_+$-orbit in X^+.

We begin by recalling the definitions of Hecke orbits and generalised Hecke orbits, the relations between them and their basic functoriality properties; this is largely taken from [Pin05a]. In the case of the Siegel modular variety $\mathcal{A}_g$, a Hecke orbit is the same as a polarised isogeny class. The difference between polarised and unpolarised isogeny classes of abelian varieties is very important in this thesis, so we discuss this in more detail. In particular we prove that an unpolarised isogeny class of principally polarised abelian varieties may contain infinitely many polarised isogeny classes, but that the natural map $\mathcal{A}_g \to \mathcal{A}_{4g}$ maps unpolarised isogeny classes into polarised isogeny classes. The latter is useful because it allows us to reduce questions about unpolarised isogeny classes to questions about Hecke orbits by replacing $\mathcal{A}_g$ with $\mathcal{A}_{4g}$.

We then introduce P-Hecke orbits, where P is a finite set of primes. A P-Hecke orbit is the image of an orbit in X^+ for the subgroup of $G(\mathbb{Q})_+$ consisting of elements which are "trivial at primes outside P." The reason for introducing these is that the preimage of a P-Hecke orbit by a Shimura immersion is contained in a finite union of P-Hecke orbits, while the analogous statement for usual Hecke orbits is false. This allows us to prove the André–Pink conjecture for arbitrary subvarieties of Shimura varieties of abelian type, provided that we replace Hecke orbits by P-Hecke orbits in the statement of the conjecture.

4.1 Definitions of Hecke orbits

Let $S = \mathrm{Sh}_\Gamma(G, X^+)$ be a connected Shimura variety. Let φ be an automorphism of the Shimura datum (G, X^+) (that is, an automorphism of the $\mathbb{Q}$-algebraic group G which maps X^+ into itself). Let $\Gamma_\varphi = \Gamma \cap \varphi^{-1}(\Gamma)$ (this is a congruence subgroup). We have a diagram of Shimura coverings

$$\begin{array}{ccc} & \mathrm{Sh}_{\Gamma_\varphi}(G, X^+) & \\ {}^{[\mathrm{id}]}\swarrow & & \searrow^{[\varphi]} \\ \mathrm{Sh}_\Gamma(G, X^+) & & \mathrm{Sh}_\Gamma(G, X^+) \end{array}$$

This induces a finite correspondence $S \to S$ denoted T_φ. Correspondences of this form are called **generalised Hecke operators**.

The correspondence is called a **(usual) Hecke operator** if φ is an inner automorphism $h \mapsto ghg^{-1}$ for some $g \in G(\mathbb{Q})_+$. In this case we may also write T_g instead of T_φ.

Observe that $G(\mathbb{Q})_+$ stabilises X^+, so every $g \in G(\mathbb{Q})_+$ gives rise to a usual Hecke operator. Furthermore if $\gamma \in \Gamma$ then conjugation by γ induces the identity morphism $S \to S$, so the Hecke operator T_g depends only on the double coset $\Gamma g \Gamma$.

Let $s \in S$. The **(usual) Hecke orbit** of s is the union of $T_g.s$ for all usual Hecke operators T_g. Similarly the **generalised Hecke orbit** of s is the union of $T_\varphi.s$ for all generalised Hecke operators T_φ.

We may also describe Hecke orbits via their preimages under $\pi \colon X^+ \to S$. Let $\tilde{s}$ be any point of X^+ in the preimage of s. Then the usual Hecke orbit of s is $\pi(G(\mathbb{Q})_+.\tilde{s})$, and the generalised Hecke orbit is $\pi(\mathrm{Aut}(G, X^+).\tilde{s})$.

4.1.1 Comparison of definitions

Generalised Hecke operators are more general than usual Hecke operators in two ways:

1. The automorphism φ may become an inner automorphism of G after extension of scalars to $\bar{\mathbb{Q}}$, even though it is not an inner automorphism over $\mathbb{Q}$. (We will call such a generalised Hecke operator an **adjoint Hecke operator**.)

2. The automorphism φ may be an outer automorphism of G, even after extending scalars to $\bar{\mathbb{Q}}$.

The first generalisation can make generalised Hecke orbits significantly larger than usual Hecke orbits, although it has no effect if G is an adjoint group. The second generalisation is unimportant, insofar as a generalised Hecke orbit is a finite union of orbits with respect to adjoint Hecke operators. (This is only true because we are only considering pure Shimura varieties; outer automorphisms make a big difference to generalised Hecke orbits for mixed Shimura varieties.)

With regard to the first generalisation, suppose that $\varphi \colon G \to G$ is conjugation by an element $g \in G(\bar{\mathbb{Q}})$. The condition that φ should be defined over $\mathbb{Q}$ is equivalent to the image of g in G^{ad} being defined over $\mathbb{Q}$, and the condition that φ should preserve X^+ is equivalent to the image of g in G^{ad} being in $G^{\mathrm{ad}}(\mathbb{R})^+$. Hence each adjoint Hecke operator is induced by an element of $G^{\mathrm{ad}}(\mathbb{Q})_+$, explaining the name.

Clearly an adjoint Hecke orbit is the same as a usual Hecke orbit if $G = G^{\mathrm{ad}}$. More generally, each adjoint Hecke orbit in S is the preimage of a usual Hecke orbit in S^{ad} (an associated adjoint Shimura variety). An adjoint Hecke orbit contains infinitely many usual Hecke orbits whenever the image of $G(\mathbb{Q})_+$ has infinite index in $G^{\mathrm{ad}}(\mathbb{Q})_+$. The standard example of a group for which the image of $G(\mathbb{Q})$ has infinite index in $G^{\mathrm{ad}}(\mathbb{Q})$ is SL_2, but there are no Shimura varieties associated with

SL_2. To obtain an example of a Shimura variety where adjoint Hecke orbits contain infinitely many usual Hecke orbits, let F be a real quadratic field and take

$$G = \mathbb{G}_{m,\mathbb{Q}}.\operatorname{Res}_{F/\mathbb{Q}} \mathrm{SL}_2 .$$

This is the Mumford–Tate group of an abelian surface whose endomorphism algebra is the real quadratic field F. The cosets of $G(\mathbb{Q})_+$ in $G^{\mathrm{ad}}(\mathbb{Q})_+$ are in bijection with $F^{\times,+}/\mathbb{Q}^{\times,+}F^{\times 2}$ which is infinite. (Here $F^{\times,+}$ means the totally positive elements in $F^\times$.)

For the second generalisation, consider first the case of an adjoint group. Such a group is semisimple so has finite outer automorphism group. Hence each generalised Hecke orbit in a connected Shimura variety of adjoint type is a finite union of adjoint (or equivalently usual) Hecke orbits. We can deduce that each generalised Hecke orbit in any connected Shimura variety S is a finite union of adjoint Hecke orbits, because its image in S^{ad} is contained in a generalised Hecke orbit for S^{ad} (because every automorphism of G induces an automorphism of G^{ad}).

Let us also compare our definition of usual Hecke orbits with that of Deligne. Deligne (working with non-connected Shimura varieties) associates a Hecke operator with each $g \in G(\mathbb{A}_f)$. This Hecke operator depends only on the double coset KgK, where K is the closure of Γ in $G(\mathbb{A}_f)$. Hence every Hecke operator induced by an element of $G(\mathbb{Q})_+K$ is the same as one induced by an element of $G(\mathbb{Q})_+$. The additional Hecke operators induced by elements of $G(\mathbb{A}_f)$ not in $G(\mathbb{Q})_+K$ permute the connected components of the Shimura variety non-trivially, and so do not make sense in the world of connected Shimura varieties.

4.1.2 Isogeny classes

Another related notion is that of ρ-isogeny classes,where ρ is a faithful representation of the group G. These have the advantage over usual and generalised Hecke orbits that they have better functorial properties (for suitably chosen representations ρ), but the disadvantages that they depend on the choice of a representation ρ and that it seems difficult to exploit the information that points are in the same isogeny class, except for some special cases. Edixhoven and Yafaev [EY03] used ρ-isogeny classes when proving the André–Pink conjecture for special points because of the functorial properties.

Let ρ be a faithful representation of G defined over $\mathbb{Q}$. We say that two points $s, t \in S$ are **ρ-isogenous** if the associated $\mathbb{Q}$-Hodge structures $\rho \circ \tilde{s}$ and $\rho \circ \tilde{t}$ are isomorphic, for points $\tilde{s}, \tilde{t} \in X^+$ in the preimages of s and t respectively. This is independent of the choice of lifts $\tilde{s}$, $\tilde{t}$.

For every representation ρ, each generalised Hecke orbit is contained in a finite union of ρ-isogeny classes ([Pin05a] Proposition 3.6). In the converse direction, a ρ-isogeny class may contain infinitely many generalised Hecke orbits (Lemma 4.3).

4.1.3 Example

The most important examples of ρ-isogeny classes are when $G = \mathrm{GSp}_{2g}$ and ρ is the standard $2g$-dimensional representation of G. Then two points of $\mathcal{A}_g$ are ρ-isogenous if and only if the associated abelian varieties are isogenous (forgetting the polarisations).

Being in the same Hecke orbit in $\mathcal{A}_g$ is a stricter condition. Two points of $\mathcal{A}_g$ are in the same Hecke orbit if and only if there is a **polarised isogeny** between the associated principally polarised abelian varieties (A, λ) and (B, μ): that is, an isogeny $f \colon A \to B$ such that $f^*\mu = n\lambda$ for some $n \in \mathbb{Z}$. There is no difference between Hecke orbits and generalised Hecke orbits for GSp_{2g} because $\mathrm{GSp}_{2g}(\mathbb{Q}) \to \mathrm{PGSp}_{2g}(\mathbb{Q})$ is surjective and because GSp_{2g} has no outer automorphisms which map X^+ into X^+.

We will prove in Lemma 4.3 that an isogeny class in $\mathcal{A}_g$ may contain infinitely many polarised isogeny classes – this is essentially the same example as the generalised and usual Hecke orbits for $\mathbb{G}_m.\operatorname{Res}_{F/\mathbb{Q}} \mathrm{SL}_2$ considered above. The difference between isogeny classes and polarised isogeny classes will be important throughout this thesis. However it is possible to embed isogeny classes in $\mathcal{A}_g$ in Hecke orbits in a larger Shimura variety, using the natural Shimura embedding $\mathcal{A}_g \to \mathcal{A}_{4g}$ which sends a principally polarised abelian variety to its fourth power (see Proposition 4.4).

4.2 Functoriality of Hecke orbits

The functorial properties of Hecke orbits are complicated, varying between the different kinds of orbits. We begin with a list of the basic properties.

Lemma 4.1. *(1) The image of a usual Hecke orbit by any Shimura morphism is contained in a Hecke orbit.*

(2) The preimage of a usual Hecke orbit may contain infinitely many Hecke orbits, even for a Shimura embedding or Shimura covering.

(3) The image of a generalised Hecke orbit by a Shimura embedding may intersect infinitely many generalised Hecke orbits.

(4) The image of a generalised Hecke orbit by a Shimura submersion is contained in a finite union of generalised Hecke orbits.

(5) The preimage of a generalised Hecke orbit by a Shimura embedding may intersect infinitely many generalised Hecke orbits.

(6) The preimage of a generalised Hecke orbit by a Shimura covering is contained in a finite union of generalised Hecke orbits.

(7) *For every faithful representation ρ of G, there exists a faithful representation ρ^{ad} of G^{ad} such that the image of each ρ-isogeny class by the Shimura covering associated with $G \to G^{\mathrm{ad}}$ is contained in a ρ^{ad}-isogeny class.*

(8) *For every faithful representation ρ of G, the preimage of each ρ-isogeny class by a Shimura embedding $[\iota]$ is contained in a $\rho\iota$-isogeny class.*

Proof. Property (1) is obvious. Property (2) for a Shimura embedding follows from Lemma 4.3 and Proposition 4.4. (2) for a Shimura covering and (3) may be proved by modifications of the $\mathbb{G}_m.\operatorname{Res}_{F/\mathbb{Q}}\mathrm{SL}_2$ example above. (4) can be proved by reducing it to a question for usual Hecke orbits in Shimura varieties of adjoint type and using (1). (5) for a Shimura embedding follows from Lemma 4.3 and Proposition 4.4. (6) can be proved by replacing the codomain of the Shimura covering by an associated adjoint Shimura variety; it is then clear that the preimage of a generalised Hecke orbit is a finite union of adjoint Hecke orbits. (7) is proved by Construction 2.3 in [EY03]. (8) is obvious. □

As in (5) above, the preimage of a generalised Hecke orbit by a Shimura embedding $S' \to S$ may intersect infinitely many generalised Hecke orbits. This prevents us proving the full André–Pink conjecture for Shimura varieties of abelian type. However this problem does not occur if the generalised Hecke orbit in question is that of a Hodge generic point in S'.

Lemma 4.2. *Let $[\iota]\colon S' \to S$ be a Shimura immersion of connected Shimura varieties. Let s be a point in S' and Σ the generalised Hecke orbit of $[\iota](s)$ in S.*

If s is Hodge generic in S', then $[\iota]^{-1}(\Sigma)$ is contained in a finite union of generalised Hecke orbits in S'.

Proof. Let (G, X^+) and (G', X'^+) be the connected Shimura data describing S and S' respectively. As in Theorem 6.6, we may assume that G is an adjoint group and that Σ is a usual Hecke orbit in S. Similarly we may assume that G' is the generic Mumford–Tate group of X'^+ (the inclusion of the generic Mumford–Tate group in G' is a Shimura covering). We may also replace G' by its image $\iota(G')$ and hence assume that $[\iota]$ is a Shimura embedding and G' is a subgroup of G.

Choose $\tilde{s} \in X'^+$ above s. Let t be any point in $[\iota]^{-1}(\Sigma)$, and let $\tilde{t}$ be a point in X'^+ above t. Since $[\iota](t)$ and $[\iota](s)$ are in the same Hecke orbit in S, there is some $g \in G(\mathbb{Q})_+$ such that $g.\iota_*\tilde{s} = \iota_*\tilde{t}$.

Since s is Hodge generic in S' and G' is the generic Mumford–Tate group for X'^+, the Mumford–Tate group of $\tilde{s}$ is G'. Hence the Mumford–Tate group of $\iota_*\tilde{t}$ is $gG'g^{-1}$. But since $\tilde{t} \in X'^+$, its Mumford–Tate group is contained in G'. By comparing dimensions, we get that

$$gG'g^{-1} = G'.$$

Hence conjugation by g induces an automorphism of G' defined over $\mathbb{Q}$. Hence s and t are in the same generalised Hecke orbit in S'. $\square$

4.3 Polarised and unpolarised isogeny classes

As mentioned above, a Hecke orbit or generalised Hecke orbit in $\mathcal{A}_g$ is the same as a polarised isogeny class of principally polarised abelian varieties. We use the arithmetic of endomorphism rings of abelian varieties to prove two results on the relationship between polarised and unpolarised isogeny classes. First we show that the difference between the two is important, because an isogeny class may contain infinitely many polarised isogeny classes. Then we show that it is possible to work round this difference by embedding $\mathcal{A}_g$ in $\mathcal{A}_{4g}$.

Lemma 4.3. *An isogeny class of principally polarised abelian varieties may contain infinitely many polarised isogeny classes.*

Proof. Consider a principally polarised abelian surface (A, λ) whose endomorphism ring is the ring of integers of a real quadratic field F.

For each totally positive $d \in \mathfrak{o}_F$, there is a principally polarised abelian surface (A_d, λ_d) and an isogeny $f_d \colon A \to A_d$ such that $f_d^* \lambda_d = \lambda \circ d$ (apply [Mil86] Corollary 16.10 to $(A, \lambda \circ d)$).

By definition, (A_d, λ_d) and (A_e, λ_e) are in the same polarised isogeny class if and only if there is some $g \colon A_d \to A_e$ such that $g^* \lambda_e = n \lambda_d$ for some $n \in \mathbb{Z}$. Thus, if $u = f_e^{-1} g f_d \in \operatorname{End} A \otimes \mathbb{Q}$, we must have $nd = u^\dagger e u$. So (A_d, λ_d) and (A_e, λ_e) are in the same polarised isogeny class if and only if there exist $n \in \mathbb{Z}$ and $u \in \mathfrak{o}_F - \{0\}$ such that $nd = u^\dagger e u = u^2 e$.

Hence the polarised isogeny classes contained in the isogeny class of (A, λ) are parameterised by $F^{+,\times} / \mathbb{Q}^\times F^{\times 2}$, which is infinite: there are infinitely many rational primes which split as a product of two principal prime ideals $(p) = (a_p)(a'_p)$ in $\mathfrak{o}_F$; we can always choose a_p totally positive, and the a_p for different p are in different classes in $F^{+,\times} / \mathbb{Q}^\times F^{\times 2}$. $\square$

Let $[\iota]$ denote the Shimura embedding $\mathcal{A}_g \to \mathcal{A}_{4g}$ induced by the block diagonal embedding

$$M \mapsto \begin{pmatrix} M & & & \\ & M & & \\ & & M & \\ & & & M \end{pmatrix} : \ \mathrm{GSp}_{2g} \to \mathrm{GSp}_{8g}.$$

In terms of moduli of abelian varieties, this sends a principally polarised abelian variety to its fourth power. The following proposition shows that if Σ is an isogeny class in $\mathcal{A}_g$, then $[\iota](\Sigma)$ is contained in a polarised isogeny class in $\mathcal{A}_{4g}$.

Proposition 4.4. *Let (A, λ) and (B, μ) be principally polarised abelian varieties. If A and B are isogenous, then there is a polarised isogeny $(A, \lambda)^4 \to (B, \mu)^4$.*

Let $f \colon A \to B$ be an isogeny. There is some positive symmetric endomorphism $d \in \operatorname{End} A$ such that $f^*\mu = \lambda d$. We shall show that there are $u \in \mathrm{M}_4(\operatorname{End} A)$ and $q \in \mathbb{Z}$ such that

$$u^\dagger \operatorname{diag}_4(d) u = q. \tag{*}$$

Then $(\operatorname{diag}_4(f)u)^*(\operatorname{diag}_4(\mu)) = \operatorname{diag}_4(\lambda)q$ so $\operatorname{diag}_4(f)u$ is the desired polarised isogeny $A^4 \to B^4$.

Let $E = \operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{Q}$. It will suffice to find $u \in \mathrm{M}_4(E)$ and $q \in \mathbb{Q}$ satisfying (*) because we can multiply up by a rational integer to clear the denominators. Furthermore E is a direct product of simple $\mathbb{Q}$-algebras (each of which is preserved by $\dagger$) and we can solve (*) independently in each component, so we may assume that E is simple. Then $E = \mathrm{M}_n(D)$ for some division algebra D with positive involution $\dagger$.

Let $\psi \colon D^n \times D^n \to D$ be the $(D, \dagger)$-Hermitian form $\psi(a, b) = adb^\dagger$. Then (*) is the claim that $\psi^{\oplus 4}$ is equivalent to a Hermitian form represented by a matrix $\operatorname{diag}_{4n}(q)$ for some $q \in \mathbb{Q}$. Thus Proposition 4.4 is implied by the lemma below.

Lemma 4.5. *Let $(D, \dagger)$ be a division algebra over $\mathbb{Q}$ with positive involution. Let $\psi \colon D^n \times D^n \to D$ be a $(D, \dagger)$-Hermitian form, positive definite in each real place. Then $\psi^{\oplus 4}$ is equivalent to the Hermitian form represented by the matrix* $\operatorname{diag}_{4n}(1)$.

Proof. We shall prove this using the classification of $(D, \dagger)$-Hermitian forms, as in chapter 10 of [Sch85].

By construction, $\psi^{\oplus 4}$ and $\operatorname{diag}_{4n}(1)$ have the same dimension. Since ψ is positive at each real place they also have the same signature. The other invariants in the classification of Hermitian forms differ depending on the type of $(D, \dagger)$ in the Albert classification, we must split into cases following the Albert classification.

Type I. D is a totally real number field and the Rosati involution is trivial, so $(D, \dagger)$-Hermitian forms are just quadratic forms over D. Quadratic forms over D are classified by their dimension, determinant in $D^\times / D^{\times 2}$, Hasse invariant in $\operatorname{Br} D$ and signature at real places ([Sch85] Corollary 6.6.6). The determinant of $\psi^{\oplus 4}$ is the fourth power of $\det \psi$, so is a square.

It remains to check that Hasse invariant $s(\psi^{\oplus 4})$ is 1. To compute the Hasse invariant, choose a diagonal form $\operatorname{diag}(a_1, \dots, a_n)$ equivalent to ψ. When we take unordered pairs from $\{a_1, \dots, a_n\}$ repeated four times, each pair $\{a_i, a_j\}$ with $i \neq j$ occurs 16 times and each pair $\{a_i, a_i\}$ occurs six times. Hence

$$s(\psi^{\oplus 4}) = \prod_{i<j} s(a_i, a_j)^{16} \prod_i s(a_i, a_i)^6.$$

Since each $s(a_i, a_j)$ and $s(a_i, a_i)$ has order 1 or 2 in $\operatorname{Br} D$ it follows that $s(\psi^{\oplus 4}) = 1$.

Type II. D is a quaternion algebra over a number field F, indefinite at all real places, and $\dagger$ is an orthogonal involution.

This case is harder than the others because the isometry class of a $(D, \dagger)$-Hermitian form not determined by its localisations. There is a localisation map on the Witt group of $(D, \dagger)$-Hermitian forms

$$r \colon W(D, \dagger) \to \prod_{\mathfrak{p}} W(D_{\mathfrak{p}}, \dagger)$$

where the product is over all places of F, but it is not injective. We shall show first that $\psi^{\oplus 2}$ is equivalent to $\mathrm{diag}_{2n}(1)$ modulo $\ker r$, then that this kernel has exponent 2.

Let $*$ denote the canonical involution on D. Then the classification of $\dagger$-Hermitian forms is equivalent to the classification of $*$-skew-Hermitian forms. Hence for each non-archimedean place $\mathfrak{p}$, regular $(D_{\mathfrak{p}}, \dagger)$-Hermitian forms are classified by their dimension and determinant in $F^\times / F^{\times 2}$ ([Sch85] Theorem 10.3.6). The determinant of $\psi^{\oplus 2}$ is a square, so $\psi^{\oplus 4}$ is equivalent to $\mathrm{diag}_{2n}(1)$ locally in each non-archimedean place.

At archimedean places $\mathfrak{p}$, $(D_{\mathfrak{p}}, \dagger) \cong (\mathrm{M}_2(\mathbb{R}), \text{transpose})$ so $(D_{\mathfrak{p}}, \dagger)$-Hermitian forms are just real quadratic forms and are classified by their dimension and signature. By assumption, $\psi^{\oplus 2}$ and $\mathrm{diag}_{2n}(1)$ have the same signature at each archimedean place.

Hence $\psi^{\oplus 2}$ and $\mathrm{diag}_{2n}(1)$ are equivalent in every localisation; in other words, they are equivalent modulo the kernel of r.

According to [Lew82] Proposition 3, $\ker r \cong (\mathbb{Z}/2\mathbb{Z})^{s-2}$ where s is the number of places at which D does not split. (Lewis only states that the order of $\ker r$ is equal to 2^{s-2} and that $\ker r \cong q^{-1}(\operatorname{Im} \delta) / \operatorname{Im} \alpha$ for certain maps q, δ, α. But the proof shows that q is injective and $\operatorname{Im} \delta$ has exponent 2, so $\ker r$ has exponent 2.)

Hence $2([\psi^{\oplus 2}] - [\mathrm{diag}_{2n}(1)]) = 0$ in $W(D, \dagger)$. Hence $\psi^{\oplus 4}$ and $\mathrm{diag}_{4n}(1)$ are equivalent as $(D, \dagger)$-Hermitian forms.

Type III. D is a quaternion algebra over a number field F, definite at all real places, and $\dagger$ is the canonical involution.

By [Sch85] 10.1.8, $(D, \dagger)$-Hermitian forms are classified by their dimension and signature at all real places of F. We have already observed that these are equal for $\psi^{\oplus 4}$ and $\mathrm{diag}_{4n}(1)$.

Type IV. D is a simple algebra with centre F and $\dagger$ is a unitary involution. Let F_0 be the fixed field of $\dagger$ in F.

By [Sch85] Corollary 10.6.6, $(D, \dagger)$-Hermitian forms are classified by their dimension, determinant in $F_0^\times / N_{F/F_0}(F^\times)$ and signature for all real places of F_0

which do not decompose in F. We know that the dimension and signature of $\psi^{\oplus 4}$ and $\mathrm{diag}_{4n}(1)$ agree.

For the determinant, $\det(\psi^{\oplus 4}) = (\det \psi)^4$ is the square of an element of $F_0^\times$ and hence is in $N_{F/F_0}(F^\times)$ as required. ☐

4.4 Embedding Shimura varieties of abelian type in $\mathcal{A}_g$

By definition, every connected Shimura variety S of abelian type has a covering $[p]\colon S_1 \to S$ by a Shimura variety for which there is an embedding $[\iota]\colon S_1 \to \mathcal{A}_g$ (for some g). However, if Σ is a generalised Hecke orbit in S then $[\iota][p]^{-1}(\Sigma)$ may intersect infinitely many Hecke orbits in $\mathcal{A}_g$ (thanks to Lemma 4.1 (2) and (3)).

We shall show that this can be worked around by suitably choosing the centre of the reductive group associated with S_1. This is necessary to reduce the André–Pink conjecture for Shimura varieties of abelian type to the case of $\mathcal{A}_g$.

Theorem 4.6. *Let S be a connected Shimura variety of adjoint abelian type. Then there exist a Shimura variety S_1, a Shimura covering $[p]\colon S_1 \to S$ and a Shimura immersion $[\iota]\colon S_1 \to \mathcal{A}_g$ for some $g \in \mathbb{N}$ such that, for every Hecke orbit Σ in S, $[\iota][p]^{-1}(\Sigma)$ is contained in a Hecke orbit in $\mathcal{A}_g$.*

We will prove this using the following proposition.

Proposition 4.7. *Let (G, X^+) be a connected Shimura datum of adjoint abelian type. Then there exist reductive $\mathbb{Q}$-algebraic groups G_1 and G_3 fitting into the following commutative diagram*

$$\begin{array}{ccccc} & & G_3 & \hookrightarrow & \mathrm{GL}_{2g} \\ & \swarrow & \uparrow & & \uparrow \\ G & \xleftarrow{\;p\;} & G_1 & \xhookrightarrow{\;\iota\;} & \mathrm{GSp}_{2g} \end{array}$$

such that

(1) the four arrows of the right-hand square are injections;

(2) there is a connected Shimura datum (G_1, X_1^+) such that $G_1 \to G$ and $G_1 \to \mathrm{GSp}_{2g}$ induce morphisms of Shimura data;

(3) $G_1^{\mathrm{ad}} = G_3^{\mathrm{ad}} = G$; and

(4) $G_3(\mathbb{Q})_+ \to G(\mathbb{Q})_+$ is surjective.

Proof of Theorem 4.6. Apply Proposition 4.7 to get G_1 and G_3. Choose a congruence subgroup $\Gamma_1 \subset G_1(\mathbb{Q})_+$ such that $p(\Gamma_1) \subset \Gamma$, and let $S_1 = \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+)$.

Suppose that $s, t \in \Sigma$ and choose $s_1, t_1 \in S_1$ such that $[p](s_1) = s$ and $[p](t_1) = t$. We claim that $[\iota](s_1)$ and $[\iota](s_2)$ are isogenous.

Choose $\tilde{s}, \tilde{t} \in X_1^+$ lifting s_1, t_1.

By [Mil05] Proposition 5.7, $p\colon G_1 \to G$ induces an isomorphism $p_*\colon X_1^+ \to X^+$. Because s and t are in the same Hecke orbit, we have $p_*\tilde{s} \in G(\mathbb{Q})_+.p_*(\tilde{t})$. Since $G_3(\mathbb{Q})_+ \to G(\mathbb{Q})_+$ is surjective, it follows that $\tilde{s} \in G_3(\mathbb{Q})_+.\tilde{t}$.

Thus $\iota \circ \tilde{s}$ and $\iota \circ \tilde{t}$ are $\mathrm{GL}_{2g}(\mathbb{Q})$-conjugate. In other words, they define isomorphic $\mathbb{Q}$-Hodge structures. Hence the abelian varieties corresponding to $[\iota](s_1)$ and $[\iota](s_2)$ are isogenous (forgetting the polarisations).

We have proved that $[\iota][p]^{-1}(\Sigma)$ is contained in a single isogeny class in $\mathcal{A}_g$. By Proposition 4.4, composing $[\iota]$ with the natural embedding $\mathcal{A}_g \to \mathcal{A}_{4g}$ maps this into a single Hecke orbit in $\mathcal{A}_{4g}$. □

Proof of Proposition 4.7. Suppose that G is $\mathbb{Q}$-simple; if not, we may deal with each $\mathbb{Q}$-simple factor of G separately. Thus $G = \mathrm{Res}_{F/\mathbb{Q}} H$ for some totally real number field F and absolutely simple F-group H.

Let D be the Dynkin diagram of $G_{\mathbb{C}}$. It has $[F : \mathbb{Q}]$ irreducible components D_σ corresponding to the embeddings $\sigma\colon F \hookrightarrow \mathbb{R}$.

Choose some $\sigma\colon F \hookrightarrow \mathbb{R}$ such that G_σ is non-compact. Then the conjugacy class of Hodge parameters of G_σ obtained by projecting X^+ onto G_σ determines a so-called special node in D_σ: that is, a circled node in [Del79] Table 1.3.9.

Choose a corresponding symplectic node s_0 in D_σ: that is, one of the underlined nodes in Table 1.3.9. Let S be the $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-orbit of s_0 in the Dynkin diagram. Then the set of singletons $\{\{s\} \mid s \in S\}$ satisfies Deligne's conditions 2.3.7(b) and (c).

Let K be the number field such that $S = \mathrm{Hom}(K, \bar{\mathbb{Q}})$ as $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$-sets. Then K is either equal to F or is a CM quadratic extension of F. (In fact, K is a CM quadratic extension of F if and only if the opposite involution of D does not fix s_0; in other words, if and only if (G, X^+) has type A_n for $n \geq 2$ or $D_n^{\mathbb{R}}$ for odd n.)

Let $\tilde{G}$ be the simply connected cover of G. Let (ρ, V) be the irreducible $\mathbb{Q}$-representation of $\tilde{G}$ which has the weight corresponding to s_0 as a highest weight. Since $S = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}).s$, we have

$$V_{\bar{\mathbb{Q}}} = \bigoplus_{s \in S} V_s$$

where V_s is a power of the irreducible representation of $\tilde{G}_{\bar{\mathbb{Q}}}$ whose highest weight corresponds to s. The endomorphism algebra of V is a division algebra with centre isomorphic to K. We thus get an injection $i\colon \mathrm{Res}_{K/\mathbb{Q}} \mathbb{G}_m \to \mathrm{GL}(V)$.

Let G_2 be the subgroup of $\mathrm{GL}(V)$ generated by $\rho(\tilde{G})$ and $i(\mathrm{Res}_{K/\mathbb{Q}}\,\mathbb{G}_m)$. Because $\ker\rho$ is contained in the centre of $\tilde{G}$, the adjoint group of G_2 is isomorphic to G. According to the following lemma, whose proof is obvious, the centre of G_2 is $i(\mathrm{Res}_{K/\mathbb{Q}}\,\mathbb{G}_m)$.

Lemma 4.8. *Let $\rho\colon G \to \mathrm{GL}(V)$ be a finite-dimensional representation of a group, let Z be the centre of $\mathrm{End}\,\rho$ and $i\colon Z \to \mathrm{GL}(V)$ its action on V. Let G_2 be the subgroup of $\mathrm{GL}(V)$ generated by $\rho(G)$ and $i(Z)$. Then the centre of G_2 is $i(Z)$.*

Choose a Hodge parameter $h \in X^+$. As in [Del79] 2.3.10, we can find $h_2\colon \mathbb{S} \to G_{2\mathbb{R}}$ lifting h such that induced Hodge structure on V_s has type $(0,0)$ when s is in a component of the Dynkin diagram corresponding to a compact factor of $G_{\mathbb{R}}$ and type $\{(-1,0),(0,-1)\}$ when s is in a component of the Dynkin diagram corresponding to a non-compact factor.

Let $E = K$ if K is a CM field, and let E be any CM quadratic extension of F otherwise. Choose a Hodge parameter h_E of $\mathrm{Res}_{E/\mathbb{Q}}\,\mathbb{G}_m$ which induces a Hodge structure of type $\{(-1,0),(0,-1)\}$ on E_σ for those $\sigma\colon F \to \mathbb{R}$ for which G_σ is compact, and of type $(0,0)$ for those σ for which G_σ is non-compact.

If K is a CM field, then let $V' = V$, $G_3 = G_2$ and $h_3 = h_2.h_E$.

If $K = F$, then let $V' = V \otimes_F E$, let G_3 be the subgroup of $\mathrm{GL}(V')$ generated by G_2 and $\mathbb{G}_{m,E}$ and let $h_3 = h_2 \otimes h_E\colon \mathbb{S} \to G_{3\mathbb{R}}$.

In each case we get a reductive group G_3, a Hodge parameter $h_3\colon \mathbb{S} \to G_{3\mathbb{R}}$ and a faithful representation V' of G_3 such that the induced Hodge structure has type $\{(-1,0),(0,-1)\}$. The adjoint group of G_3 is G and the centre is $\mathrm{Res}_{E/\mathbb{Q}}\,\mathbb{G}_m$.

The centre of G_3 splits over the CM field E so by [Del79] Corollary 2.3.3, there exist a subgroup $G_1 \subset G_3$ and a symplectic form ψ on V' such that

(1) $G_1^{\mathrm{ad}} = G_3^{\mathrm{ad}} = G$,

(2) h_3 factors through G_1, and

(3) $G_1 \subset \mathrm{GSp}(V', \psi)$.

It remains to show that $G_3(\mathbb{Q})_+ \to G(\mathbb{Q})_+$ is surjective. Because the centre Z of G_3 is a quasi-split torus, $H^1(\mathbb{Q}, Z) = 0$ and so $G_3(\mathbb{Q}) \to G(\mathbb{Q})$ is surjective. By [Mil05] Proposition 5.1, $G_3(\mathbb{R})_+ \to G(\mathbb{R})_+$ is also surjective. Hence any $g \in G(\mathbb{Q})_+$ has preimages $x \in G_3(\mathbb{R})_+$ and $y \in G_3(\mathbb{Q})$. Then $xy^{-1} \in Z(\mathbb{R})$. By [PR94] Theorem 7.7, $Z(\mathbb{Q})$ is dense in $Z(\mathbb{R})$. In particular we may choose $z \in Z(\mathbb{Q})$ in the same connected component of $Z(\mathbb{R})$ as xy^{-1}. Then zy is in $G_3(\mathbb{Q})_+$ and has image g in $G(\mathbb{Q})$. □

4.5 P-Hecke orbits

Let $S = \mathrm{Sh}_\Gamma(G, X^+)$ be a connected Shimura variety and P a finite set of prime numbers. The P-Hecke orbit of a point $s \in S$ is a subset of the usual Hecke orbit of s, consisting of those points related to s by Hecke operators which are in a sense "trivial at primes outside P".

For example, consider the case of the modular curve $\mathcal{A}_1 = \mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}_g$. The Hecke operators are T_N for $N \in \mathbb{N}$, represented by the matrices

$$\left(\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Q})_+.$$

A P-Hecke orbit is an orbit for the Hecke operators T_N which have all prime factors of N in P.

To define P-Hecke orbits, let $\mathbb{A}_f$ denote the ring of finite adèles and $\mathbb{A}_f^P$ the projection of $\mathbb{A}_f$ into $\prod_{l \notin P} \mathbb{Q}_\ell$. Let K^P denote the closure of Γ in $G(\mathbb{A}_f^P)$ (via the diagonal embedding $G(\mathbb{Q}) \to G(\mathbb{A}_f^P)$). We say that a Hecke operator T_g is a **P-Hecke operator** if $g \in G(\mathbb{Q})_+ \cap K^P$. The **$P$-Hecke orbit** of a point $s \in S$ is the union of its images under P-Hecke operators.

We can give an equivalent definition of P-Hecke operators which does not mention adèles. Choose a faithful $\mathbb{Q}$-representation $\rho\colon G \to \mathrm{GL}_n$. We suppose that $\rho(\Gamma) \subset \mathrm{GL}_n(\mathbb{Z})$ – this is always possible for some choice of integral structure on ρ. Recall that Γ contains the basic congruence subgroup $\Gamma(N)$ for some integer N as a subgroup of finite index. Thus there is a finite subgroup $\bar{\Gamma} \subset \mathrm{GL}_n(\mathbb{Z}/N\mathbb{Z})$ such that

$$\Gamma = \{g \in G(\mathbb{Q}) \mid \rho(g) \in \mathrm{GL}_n(\mathbb{Z}) \text{ and } (\rho(g) \bmod N) \in \bar{\Gamma}\}.$$

Let N' be the largest factor of N which has no prime factors in P. Let $\bar{\Gamma}'$ be the image of $\bar{\Gamma}$ in $\mathrm{GL}_n(\mathbb{Z}/N'\mathbb{Z})$. Then

$$K^P = \{\mathbf{g} \in G(\mathbb{A}_f^P) \mid \rho(\mathbf{g}) \in \mathrm{GL}_n(\prod_{\ell \notin P} \mathbb{Z}_\ell) \text{ and } (\rho(\mathbf{g}) \bmod N') \in \bar{\Gamma}'\}$$

so a Hecke operator T_g (for $g \in G(\mathbb{Q})_+$) is a P-Hecke operator if and only if

$$\rho(g) \in \mathrm{GL}_n(\mathbb{Z}[1/p \mid p \in P]) \text{ and } (\rho(g) \bmod N') \in \bar{\Gamma}'.$$

If we use Deligne's definition of Hecke operators in which they are induced by elements of $G(\mathbb{A}_f)$, a Hecke operator is a P-Hecke operator if and only if it can be induced by an element of $\prod_{p \in P} G(\mathbb{Q}_p)$.

The image of a P-Hecke orbit by any Shimura morphism $[f]\colon S_1 \to S_2$ is contained in a P-Hecke orbit, because f maps Γ_1 into Γ_2 and hence K_1^P into K_2^P. Unlike for usual Hecke orbits, the preimage of a P-Hecke orbit by a Shimura immersion is contained in a finite union of P-Hecke orbits. This will be proved in Theorem 4.9.

Example. Two points of $\mathcal{A}_g$ are in the same P-Hecke orbit if and only if the corresponding principally polarised abelian varieties are related by a polarised P-isogeny i.e. a polarised isogeny whose degree has all of its prime factors in P.

Combining Proposition 4.4 and Theorem 4.9 shows that each P-isogeny class in $\mathcal{A}_g$ is a finite union of polarised P-isogeny classes.

4.6 P-Hecke orbits and Shimura immersions

Theorem 4.9. *Let $[f]\colon S_1 \to S_2$ be a Shimura immersion. Let P be a finite set of prime numbers and Σ_P a P-Hecke orbit in S_2.*

Then $[f]^{-1}(\Sigma_P)$ is contained in a finite union of P-Hecke orbits in S_1.

The idea of the proof of this theorem is that the number of P-Hecke orbits in S' in the preimage of Σ_P can be bounded above by certain Galois cohomology groups (specifically the cohomology groups of $\ker f$ and of the centraliser of the Mumford–Tate group of a point in Σ_P). These are cohomology groups over the base field $\mathbb{Q}_p$ for each $p \in P$, so are finite. On the other hand, the analogous argument for usual Hecke orbits would involve Galois cohomology groups over $\mathbb{Q}$, which may be infinite.

4.6.1 Finiteness lemmas

We begin with two general finiteness lemmas for algebraic groups.

Lemma 4.10. *Let k be a field of characteristic 0 such that $H^1(k, G)$ is finite for all linear algebraic groups G (in particular, k could be $\mathbb{Q}_p$). Let G and H be connected reductive groups over k such that $H \subset G$ and let M be any linear algebraic group over k. Let f be a homomorphism $M \to H$.*

Then $\mathrm{Hom}_k(M, H) \cap (\mathrm{Inn}\, G(k)).f$ is a finite union of $H(k)$-conjugacy classes.

Proof. The groups (G, H) form a reductive pair in the sense of [Ric67]. Hence the result is true over the algebraic closure $\bar{k}$ by [Ric67] Theorem 7.1. So it suffices to show that

$$\mathrm{Hom}_k(M, H) \cap (\mathrm{Inn}\, H(\bar{k})).f$$

is a finite union of $H(k)$-conjugacy classes.

Let Z be the centraliser in H of $f(M)$, a linear algebraic subgroup of H. Then $(\mathrm{Inn}\, H(\bar{k}).f$ can be identified with the $\bar{k}$-points of the homogeneous space H/Z, and $\mathrm{Hom}_k(M, H) \cap (\mathrm{Inn}\, H(\bar{k})).f$ with the k-points of H/Z.

We have an exact sequence

$$H(k) \to (H/Z)(k) \to H^1(k, Z)$$

and we are done because $H^1(k, Z)$ is finite. □

Interpreting the following lemma requires some care because $G(\mathbb{Q})$ is usually regarded as a subgroup of $G(\mathbb{A}_f)$ via the diagonal embedding, but this is not the composition of the natural inclusions $G(\mathbb{Q}) \to G(\mathbb{Q}_P) \to G(\mathbb{A}_f)$. In the statement of the lemma, when we write $G(\mathbb{Q}) \cap K^P$ as a subgroup of $G(\mathbb{Q}_P)$ this means that we take the preimage of K^P by the inclusion $G(\mathbb{Q}) \to G(\mathbb{A}_f^P)$ then apply the inclusion $G(\mathbb{Q}) \to G(\mathbb{Q}_P)$ to this. In the proof, when we write elements of $G(\mathbb{A}_f)$, we shall write them as two components using the direct product

$$G(\mathbb{A}_f) = G(\mathbb{A}_f^P) \times G(\mathbb{Q}_P).$$

Lemma 4.11. *Let G be a reductive group over $\mathbb{Q}$, and let K_P, K^P be compact open subgroups of $G(\mathbb{Q}_P)$ and $G(\mathbb{A}_f^P)$ respectively.*

Then

$$(G(\mathbb{Q}) \cap K^P)\backslash G(\mathbb{Q}_P)/K_P$$

is finite.

Proof. Let $C \subset G(\mathbb{Q}_P)$ be a set of representatives for the fibres of the map

$$G(\mathbb{Q}_P) \to G(\mathbb{Q})\backslash G(\mathbb{A}_f)/K^P K_P$$

induced by the inclusion $G(\mathbb{Q}_P) \to G(\mathbb{A}_f)$ (using the diagonal embedding $G(\mathbb{Q}) \to G(\mathbb{A}_f)$ on the right). The set on the right is finite by [PR94] Theorem 5.1, so C is finite.

Then for any $g \in G(\mathbb{Q}_P)$, we have a decomposition

$$(1, g) = (q, q)(1, c)(k^P, k_P) \text{ in } G(\mathbb{A}_f) = G(\mathbb{A}_f^P) \times G(\mathbb{Q}_P)$$

with $q \in G(\mathbb{Q})$, $c \in C$, $k^P \in K^P$, $k_P \in K_P$.

Looking at the $G(\mathbb{A}_f^P)$ component gives $q = (k^P)^{-1}$ so $q \in G(\mathbb{Q}) \cap K^P$.

Looking at the $G(\mathbb{Q}_P)$ component gives $g = qck_P$ and so

$$g \in (G(\mathbb{Q}) \cap K^P)\, C\, K_P.$$

Since C is finite this proves the lemma. □

4.6.2 Proof of Theorem 4.9

We have a Shimura immersion $[f]\colon S_1 \to S_2$ and a P-Hecke orbit $\Sigma_P \subset S_2$. We will need various further pieces of notation.

(i) Let $\mathbb{Q}_P = \prod_{p \in P} \mathbb{Q}_p$.

(ii) Let $S_1 = \mathrm{Sh}_{\Gamma_1}(G_1, X_1^+)$ and $S_2 = \mathrm{Sh}_{\Gamma_2}(G_2, X_2^+)$.

(iii) Let K_1^P and $K_{1,P}$ be the closures of Γ_1 in $G_1(\mathbb{A}_f^P)$ and in $G_1(\mathbb{Q}_P)$ respectively.

(iv) Let $\Gamma_1^P = G_1(\mathbb{Q})_+ \cap K_1^P$.

(v) Define K_2^P, $K_{2,P}$ and Γ_2^P similarly.

(vi) Choose points $s \in [f]^{-1}(\Sigma_P)$ and $\tilde{s} \in X_1^+$ lifting s.

(vii) Let M be the Mumford–Tate group of $f_*\tilde{s}$ in G_2.

(viii) Let Z be the centraliser of M in G_2.

(ix) Define K_Z^P, $K_{Z,P}$ and Γ_Z^P as in (iii) and (iv), with G_1 replaced by Z and Γ_1 replaced by $Z(\mathbb{Q}) \cap \Gamma_2$.

Now consider any point $t \in [f]^{-1}(\Sigma_P)$. Let $\tilde{t}$ be a point in X_1^+ lifting t. By the definition of Σ_P, there is some $g \in \Gamma_2^P$ such that

$$f_*\tilde{t} = g.(f_*\tilde{s}) \text{ in } X_2^+.$$

Lemma 4.12. *There is a finite set $C_3 \subset G_2(\mathbb{Q}_P)$, depending on f and $\tilde{s}$ but not on t, such that*

$$g = f(g_1)c_3z \text{ in } G_2(\mathbb{Q}_P)$$

for some $g_1 \in G_1(\mathbb{Q}_P)$, $c_3 \in C_3$ and $z \in Z(\mathbb{Q}_P)$.

Proof. It suffices to prove the result with $\mathbb{Q}_p$ in place of $\mathbb{Q}_P$ for each $p \in P$, as we can simply take the Cartesian product of the resulting sets C_3.

Let $H = f(G_1)$, an algebraic subgroup of G_2. Observe that $M \subset H$, and let ι be the inclusion map $M \to H$.

By Lemma 4.10, the set of homomorphisms $M_{\mathbb{Q}_p} \to H_{\mathbb{Q}_p}$ which are $G_2(\mathbb{Q}_p)$-conjugate to ι is a finite union of $H(\mathbb{Q}_p)$-conjugacy classes. Let C_1 be a finite subset of $G_2(\mathbb{Q}_p)$ such that $\{\operatorname{Inn} c \circ \iota \mid c \in C_1\}$ is a set of representatives for these $H(\mathbb{Q}_p)$-conjugacy classes.

Now $(\operatorname{Inn} g)(M) \subset G_2$ is the Mumford–Tate group of $f_*\tilde{s}$, and hence is contained in H. Thus $\operatorname{Inn} g \circ \iota$ is a homomorphism $M_{\mathbb{Q}_p} \to H_{\mathbb{Q}_p}$ which is $G_2(\mathbb{Q}_p)$-conjugate to ι. Therefore

$$\operatorname{Inn} g \circ \iota = \operatorname{Inn}(hc_1) \circ \iota$$

for some $h \in H(\mathbb{Q}_p)$ and $c_1 \in C_1$.

Now $z = (hc_1)^{-1}g$ centralises ι and so is in $Z(\mathbb{Q}_p)$.

It remains to replace $h \in H(\mathbb{Q}_p)$ by $f(g_1)$ for some $g_1 \in G_1(\mathbb{Q}_p)$. We have an exact sequence

$$G_1(\mathbb{Q}_p) \to H(\mathbb{Q}_p) \to H^1(\mathbb{Q}_p, \ker f)$$

in which the last set is finite. So we may choose a finite set C_2 of representatives for the image of $G_1(\mathbb{Q}_p)$ in $H(\mathbb{Q}_p)$. We then have

$$h = f(g_1)c_2$$

for some $g_1 \in G_1(\mathbb{Q}_p)$ and $c_2 \in C_2$.

Thus

$$g = f(g_1)c_2c_1z$$

and we are done, taking $C_3 = C_2C_1$. □

Lemma 4.13. *There is a finite set $C_6 \subset G_2(\mathbb{Q}_p)$, depending on f and $\tilde{s}$ but not on t, such that*

$$g = f(g_1^P)\gamma_2 c_6 z^P$$

for some $g_1^P \in \Gamma_1^P$, $\gamma_2 \in \Gamma_2$, $c_6 \in C_6$ and $z^P \in \Gamma_Z^P$.

Proof. Apply Lemma 4.11 to both G_1 and Z: let C_4 and C_5 be finite sets of representatives for

$$\Gamma_1^P \backslash G_1(\mathbb{Q}_P)/K_{1,P} \text{ and } K_{Z,P}\backslash Z(\mathbb{Q}_P)/\Gamma_Z^P$$

respectively (note that we have reversed the order of the decomposition for Z).

Write $g = f(g_1)c_3z$ as in Lemma 4.12 and then apply the above decompositions to g_1 and z:

$$g_1 = g_1^P c_4 k_1 \text{ for some } g_1^P \in \Gamma_1^P, c_4 \in C_4 \text{ and } k_1 \in K_{1,P}$$

and

$$z = k_2 c_5 z^P \text{ for some } k_2 \in K_{Z,P}, c_5 \in C_5 \text{ and } z^P \in \Gamma_Z^P.$$

Therefore

$$g = f(g_1^P)f(c_4)f(k_1)c_3k_2c_5z^P.$$

Observe that $f(k_1)$ and k_2 are both in $K_{2,P}$. Each $K_{2,P}$-$K_{2,P}$-double coset in $G_2(\mathbb{Q}_P)$ is a finite union of finitely many right $K_{2,P}$ cosets (because $K_{2,P}$ is compact and open in $G_2(\mathbb{Q}_P)$) so we can choose a finite set $C_6 \subset G_2(\mathbb{Q}_P)$ such that

$$f(C_4)\, K_{2,P}\, C_3\, K_{2,P}\, C_5 \subset K_{2,P}\, C_6.$$

For every right $K_{2,P}$-coset which intersects Γ_2^P, we choose an element of Γ_2^P for its representative in C_6.

We can therefore write

$$g = f(g_1^P)k_3c_6z^P \tag{*}$$

with $g_1^P \in \Gamma_1^P$, $k_3 \in K_{2,P}$, $c_6 \in C_6$ and $z^P \in \Gamma_Z^P$.

Now g, $f(g_1^P)$ and z^P are all in Γ_2^P so $k_3 c_6 \in \Gamma_2^P$. Hence by the condition on the choice of C_6, we have $c_6 \in \Gamma_2^P$. Therefore

$$k_3 \in \Gamma_2^P \cap K_{2,P} = G(\mathbb{Q})_+ \cap (K_2^P \times K_{2,P}).$$

Now $K_2^P \times K_{2,P}$ is the closure in $G_2(\mathbb{A}_f)$ of the congruence subgroup Γ_2, so in fact

$$G(\mathbb{Q})_+ \cap (K_2^P \times K_{2,P}) = \Gamma_2.$$

Thus $k_3 \in \Gamma_2$, and the lemma is proved by the decomposition (*). $\square$

Proof of Theorem 4.9. The condition that $z^P \in Z(\mathbb{R})$ implies that $z^P.(f_*\tilde{s}) = f_*\tilde{s}$, and so the decomposition of Lemma 4.13 gives

$$f_*\tilde{t} = g_1^P \gamma_2 c_6.(f_*\tilde{s}).$$

Let π_1, π_2 denote the uniformisation maps $X_1^+ \to S_1$ and $X_2^+ \to S_2$.
Let $\tilde{u} = (g_1^P)^{-1}.\tilde{t} \in X_2^+$, and $u = \pi_1(\tilde{u})$. We then have

$$f_*\tilde{u} = \gamma_2 c_6.(f_*\tilde{s}).$$

Since π_2 is left Γ_2-invariant, this gives that

$$[f](u) = \pi_2(c_6.(f_*\tilde{s})).$$

Now $\pi_2(c_6.(f_*\tilde{s}))$ lies in a finite subset of S_2 depending on f and $\tilde{s}$ but not on t, and $[f]$ is a finite morphism. Hence there are finitely many possibilities for u.

Since $g_1^P \in \Gamma_1^P$, t is in the P-Hecke orbit of u in S_1. This completes the proof of Theorem 4.9. $\square$

5 Isogeny bounds

In this chapter we prove several bounds concerning isogenies between abelian varieties. The starting point is the Masser–Wüstholz isogeny theorem.

Theorem 5.1 ([MW93a]). *Let K be a number field and A a principally polarised abelian variety defined over K. There exist constants $c(A, K)$ and κ (with κ depending only on* $\dim A$*) such that:*

If B is any principally polarised abelian variety defined over a finite extension L of K and isogenous over $\bar{K}$ to A, then there exists an isogeny $A \to B$ defined over $\bar{K}$ of degree at most

$$c(A, K)[L : K]^{\kappa}.$$

We begin by extending this theorem from abelian varieties defined over number fields to abelian varieties defined over finitely generated fields of characteristic 0. This is needed in order to prove the André–Pink conjecture for points $s \in \mathcal{A}_g(\mathbb{C})$ instead of just points in $\mathcal{A}_g(\bar{\mathbb{Q}})$.

In the remainder of the chapter, we discuss bounds for heights and degrees of isogenies which will be useful when applying the Masser–Wüstholz theorem to the André–Pink conjecture. For such an application, we must translate a bound on the degree of an isogeny $A \to B$ into a bound for the height of a matrix mapping a point of $\mathcal{H}_g$ corresponding to A into a point of $\mathcal{H}_g$ corresponding to B. Such applications are made more complicated by the fact that, even if we assume that there is a polarised isogeny $A \to B$, the isogeny of bounded degree which we obtain from the conclusion of the Masser–Wüstholz theorem might not be a polarised isogeny. We give two approaches to dealing with this problem.

First, in Proposition 5.3, we convert a bound for the degree of an isogeny into a bound for the height of a rational representation of an isogeny. In doing this, we need to take care to work with symplectic bases for the period lattices of the abelian varieties but we do not restrict our attention to polarised isogenies. Instead we will deal with the issue of unpolarised isogenies in the proof of Proposition 6.1.

For an alternative approach, we prove Theorem 5.6. This asserts that, if there exists a polarised isogeny $A \to B$ of arbitrary degree and a not-necessarily-polarised isogeny of given degree, then there also exists a polarised isogeny of bounded degree. We do not include the full details of how to prove Proposition 6.5 using Theorem 5.6; essentially, when working with polarised isogenies it is much easier to obtain a height bound from a degree bound and the proof of Proposition 6.1 can also be simplified considerably. However the proof of Theorem 5.6 itself is so long and complicated that this seems to lead to a longer proof of Proposition 6.5 overall.

5.1 Isogeny theorem over finitely generated fields

The Masser–Wüstholz isogeny theorem [MW93a] gives a bound for the minimum degree of an isogeny between two abelian varieties over number fields, as a function of one of the varieties and the degree of their joint field of definition. In order to prove Proposition 6.5 for points $s \in \mathcal{A}_g$ defined over $\mathbb{C}$ and not merely over $\bar{\mathbb{Q}}$, we need to extend the isogeny theorem to abelian varieties defined over finitely generated fields of characteristic 0. We will do this by a specialisation argument, using the fact that any abelian scheme has a closed fibre in which the specialisation map of endomorphism rings is surjective. The proof is based on Raynaud's proof [Ray83a] that the Manin–Mumford conjecture over $\bar{\mathbb{Q}}$ implies the conjecture over $\mathbb{C}$.

A key feature of the theorem of Masser and Wüstholz is the explicit dependence of the bound on the abelian variety A, via the Faltings height. Our theorem does not make this explicit, and it is not apparent that there is any analogy of the Faltings height over a finitely generated field which would enable it to be made explicit. Instead what matters to us is the dependence on the field of definition of B.

Theorem 5.2. *Let K be a finitely generated field of characteristic 0 and A an abelian variety defined over K. There exist constants $c(A, K)$ and κ (with κ depending only on $\dim A$) such that:*

If B is any abelian variety defined over a finite extension L of K and isogenous over $\bar{K}$ to A, then there exists an isogeny $A \to B$ defined over $\bar{K}$ of degree at most

$$c(A, K)[L : K]^{\kappa}.$$

In Masser and Wüstholz's theorem, the constant c depended also on the degrees of polarisations of A and B. This dependence has been eliminated by Gaudron and Rémond [GR12] who also showed that we can take $\kappa = 2^{10}(\dim A)^3 + \epsilon$ for the exponent.

Proof. Let R be a finitely generated normal $\mathbb{Q}$-algebra whose field of fractions is K, and let $S = \operatorname{Spec} R$. There is an abelian scheme $\mathcal{A}$ over some open subset $U \subset S$ whose generic fibre is isomorphic to A. (Note that R is a finitely generated $\mathbb{Q}$-algebra, not a finitely generated $\mathbb{Z}$-algebra, because we have no need to reduce modulo p while Noot's specialisation result requires the base to be a variety over $\mathbb{Q}$.)

By replacing L by a larger extension of bounded degree (the bound depending only on $\dim A$), we may assume that all homomorphisms $A \to B$ are defined over L ([MW93b] Lemma 2.1). Let R' be the integral closure of R in L and $S' = \operatorname{Spec} R'$. Let $\pi \colon S' \to S$ be the obvious finite morphism and let $U' = \pi^{-1}(U)$.

Because A and B are isogenous, there is an abelian scheme $\mathcal{B}$ over U' with generic fibre isomorphic to B, and such that $\mathcal{B}$ is isogenous to $\mathcal{A}$. We can construct this as follows: let N be the kernel of an isogeny $A \to B$. We can extend N to a finite flat subgroup scheme $\mathcal{N} \subset \mathcal{A}$. Then let $\mathcal{B}$ be the quotient $\mathcal{A}/\mathcal{N}$.

For any closed points $s' \in U'$ and $s = \pi(s') \in U$, the fibres $\mathcal{A}_s$ and $\mathcal{B}_{s'}$ are abelian varieties over the number fields k_s and $k_{s'}$, isogenous over $k_{s'}$. We can apply the Masser–Wüstholz theorem to deduce that there are constants $c(\mathcal{A}_s, k_s)$ and $\kappa(\dim A)$ and an isogeny $\mathcal{A}_s \to \mathcal{B}_{s'}$ of degree at most

$$c(\mathcal{A}_s, k_s)[k_{s'} : k_s]^\kappa.$$

Observe that $[k_{s'} : k_s] \leq [L : K]$.

In order to prove the theorem, all we have to do is show that this isogeny $\mathcal{A}_s \to \mathcal{B}_{s'}$ lifts to an isogeny $A \to B$ (which will have the same degree). Hence it will suffice to show that there is some closed point s such that the specialisation map

$$\operatorname{Hom}_{\bar{K}}(A, B) \to \operatorname{Hom}_{\bar{k_s}}(\mathcal{A}_s, \mathcal{B}_{s'}) \tag{*}$$

is surjective. Because we want a bound which depends only on A and not on B, we have to show that there is a single point $s \in U$ which will work for all B.

We choose a closed point $s \in U$ such that $\operatorname{End}_{\bar{K}} A \to \operatorname{End}_{\bar{k_s}} \mathcal{A}_s$ is surjective. Such an s exists by [Noo95] Corollary 1.5 (this is proved using the Hilbert irreducibility theorem).

Let f_s be a $\bar{k}_s$-homomorphism $\mathcal{A}_s \to \mathcal{B}_{s'}$. To prove that (*) is surjective, we have to show that f_s lifts to $\bar{K}$-homomorphism $A \to B$.

We are assuming that A and B are isogenous. Choose any isogeny $g_\eta : A \to B$ and let g_s be its specialisation at s. Let

$$\alpha_s = g_s^{-1} \circ f_s \in \operatorname{End}_{\bar{k_s}} \mathcal{A}_s \otimes_{\mathbb{Z}} \mathbb{Q}.$$

By our choice of s, this lifts to some $\alpha_\eta \in \operatorname{End}_{\bar{K}} A \otimes_{\mathbb{Z}} \mathbb{Q}$. Then $f_\eta = g_\eta \circ \alpha_\eta$ is a quasi-isogeny $A \to B$ specialising to f_s.

All we have to do is check that f_η is an isogeny and not just a quasi-isogeny. Choose an integer m such that mf_η is an isogeny. The kernel of mf_s contains $\mathcal{A}_s[m]$ so lifting to the generic fibre, the kernel of mf_η contains $A[m]$. Hence mf_η factorises as $f'_\eta \circ [m]$ for an isogeny $f'_\eta : A \to B$, and we must have $f'_\eta = f_\eta$. □

5.2 Heights of rational representations of isogenies

Let (A, λ) and (A', λ') be principally polarised abelian varieties over $\mathbb{C}$ related by an isogeny of degree n (not necessarily a polarised isogeny). In this section we show that we can choose an isogeny $f \colon A \to A'$ and bases for the period lattices

$H_1(A, \mathbb{Z})$ and $H_1(A', \mathbb{Z})$ such that the height of the rational representation of f is polynomially bounded in n. In particular we prove the following proposition, which will form part of the proof of the André–Pink conjecture for curves in $\mathcal{A}_g$.

The notation $H(f, \mathcal{B}', \mathcal{B})$ in the proposition refers to the height of the rational representation of the isogeny f with respect to bases $\mathcal{B}'$, $\mathcal{B}$ of the period lattices. This is defined below in section 5.2.1.

Proposition 5.3. *Let (A, λ) be a principally polarised abelian variety over $\mathbb{C}$ and fix a symplectic basis $\mathcal{B}$ for $H_1(A, \mathbb{Z})$. There exist constants c, k depending only on (A, λ) such that:*

If (A', λ') is any principally polarised abelian variety for which there exists an isogeny $A \to A'$ of degree n, then there exist an isogeny $f\colon A' \to A$ and a symplectic basis $\mathcal{B}'$ for $H_1(A', \mathbb{Z})$ such that

$$H(f, \mathcal{B}', \mathcal{B}) \leq cn^k.$$

In this proposition, the isogeny whose existence is assumed and the isogeny whose existence is asserted in the conclusion go in opposite directions. This is the most convenient formulation for our proof and application, but it is not important since any isogeny $A \to A'$ of degree n gives rise to an isogeny in the opposite direction of degree n^{2g-1}.

5.2.1 Rational representations and heights

We define the **rational representation** of an isogeny $f\colon A' \to A$ (with respect to bases $\mathcal{B}, \mathcal{B}'$ for $H_1(A, \mathbb{Z})$ and $H_1(A', \mathbb{Z})$) to be the matrix of the induced morphism

$$f_*\colon H_1(A', \mathbb{Z}) \to H_1(A, \mathbb{Z})$$

in terms of the chosen bases. This gives a $2g \times 2g$ integer matrix. We write

$$H(f, \mathcal{B}', \mathcal{B})$$

for the **height** of the rational representation of f, meaning simply the maximum of the absolute values of the entries of the matrix.

Suppose that the bases $\mathcal{B}$, $\mathcal{B}'$ are symplectic with respect to the polarisations λ, λ'. In this case, if $\tilde{s}, \tilde{t} \in \mathcal{H}_g$ are the period matrices of (A, λ) and (A', λ') with respect to the chosen bases and γ is the rational representation of an isogeny $A' \to A$, then

$$\tilde{t} = (A\tilde{s} + B)(C\tilde{s} + D)^{-1} \text{ where } \gamma^t = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right).$$

If the bases are symplectic, then an isogeny is polarised if and only if its rational representation is in $\mathrm{GSp}_{2g}(\mathbb{Q})$.

5.2.2 Polarisations

Let A be an abelian variety. A polarisation $\lambda\colon A \to A^\vee$ induces an involution, called the **Rosati involution**, of $\operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{Q}$ defined by

$$a^\dagger = \lambda^{-1} \circ a^\vee \circ \lambda$$

where $a^\vee$ means the morphism dual to a. This involution reverses the order of multiplication in $\operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{Q}$. It gives an involution of $\operatorname{End} A$ itself if λ is principal.

Having fixed a principal polarisation λ of A, every other polarisation has the form $\lambda \circ q$ for some $q \in \operatorname{End} A$ which is **symmetric**, i.e. $q^\dagger = q$, and **positive definite**, i.e. each component of q in

$$\operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{R} \cong \prod \mathrm{M}_{l_i}(\mathbb{R}) \times \prod \mathrm{M}_{m_i}(\mathbb{C}) \times \prod \mathrm{M}_{n_i}(\mathbb{H})$$

has eigenvalues which are positive real numbers.

5.2.3 Outline of proof of Proposition 5.3

In the situation of Proposition 5.3, let $h\colon A \to A'$ be an isogeny of degree n. Then $h^*\lambda'$ is a polarisation of A, so there is a symmetric positive definite endomorphism $q \in \operatorname{End} A$ such that

$$h^*\lambda' = \lambda \circ q.$$

We can identify $H_1(A', \mathbb{Z})$ with a submodule of $H_1(A, \mathbb{Z})$ of index n^{2g-1}, and so find a basis for $H_1(A', \mathbb{Z})$ whose height is at most n^{2g-1}. However this need not be a symplectic basis. We apply the standard algorithm for finding a symplectic basis: the height of this new basis is controlled by $h^*\lambda'$, in other words by q.

So we would like to bound the height of the rational representation of q in terms of $\deg h$. However this is not possible: let A be an abelian variety whose endomorphism ring is the ring of integers $\mathfrak{o}$ of a real quadratic field. In particular $\mathfrak{o}$ has infinitely many units. Let h be a unit in $\mathfrak{o}$ – in other words, an isomorphism $A \to A$. If we take the same polarisation on each copy of A, then $q = h^2$ and the rational representation of this can have arbitrarily large height.

We can avoid this by replacing h by $h \circ u$ for some automorphism u of A – recall that all we have supposed about h is that it is an isogeny $A \to A'$ of degree n. This replaces q by $u^\dagger q u$. We will show that we can choose u so that the height of the rational representation of $u^\dagger q u$ is bounded by a multiple of $\deg q = n^2$.

5.2.4 Heights in the endomorphism ring

The following proposition is motivated by the theorem [Mil86] that the symmetric elements of $\operatorname{End} A$ of a given norm fall into finitely many orbits under the action of

$(\operatorname{End} A)^\times$ given by $(u, q) \mapsto uqu^\dagger$. In geometric terms, Milne's theorem says that if we fix A and $\deg \mu$ then there are finitely many isomorphism classes of polarised abelian varieties (A, μ). Our proposition strengthens this by saying that each orbit contains an element whose height is bounded by a multiple of the norm. Milne's theorem is proved using the reduction theory of arithmetic groups. We also use reduction theory, but in order to get height bounds we have to go deeper into the structure of $\operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{R}$.

The representation ρ appears in the proposition solely to give us a convenient definition of heights and norms of elements of R. Specifically, $H(x)$ means the height of the matrix $\rho(x)$ and $\mathrm{N}(x)$ means $\det \rho(x)$ for $x \in R$. In our application, we take ρ to be the rational representation of $\operatorname{End} A$ on $H_1(A, \mathbb{Z})$; then $\mathrm{N}(f) = \deg f$ whenever $f \colon A \to A$ is an isogeny.

Proposition 5.4. *Let $(E, \dagger)$ be a semisimple $\mathbb{Q}$-algebra with a positive involution, let R be a $\dagger$-stable order in E and let $\rho \colon R \to \mathrm{M}_N(\mathbb{Z})$ be a faithful representation of R.*

There is a constant c depending only on $(R, \dagger, \rho)$ such that for any symmetric positive definite $q \in R$, there is some $u \in R^\times$ such that

$$H(u^\dagger qu) \leq c\,\mathrm{N}(q).$$

Proof. We begin by checking that it suffices to prove the proposition for simple algebras E. In general, $E = \prod E_i$ for some simple $\mathbb{Q}$-algebras E_i. Let $R_i = R \cap E_i$. Then $R' = \prod R_i$ is an order of E contained in R. Let $m = [R : R']$. Given $q \in R$, we look at $mq \in R'$. Suppose that the proposition holds for each R_i; then clearly it holds for R', so there is $u \in R'^\times$ (a fortiori $u \in R^\times$) such that

$$H(umqu^\dagger) \leq c\,\mathrm{N}(mu).$$

Hence the proposition holds for R with constant $c\,\mathrm{N}(m)/m$.

So we suppose that E is simple. Then $E = \mathrm{M}_n(D)$ for some division algebra D, and the involution $\dagger$ is matrix transposition composed with some involution of D. We may also suppose that R is contained in the maximal order $\mathrm{M}_n(\mathfrak{o})$, where $\mathfrak{o}$ is a maximal order in D.

By the Albert classification of division algebras with positive involution, $E \otimes_{\mathbb{Q}} \mathbb{R}$ is isomorphic to one of $\mathrm{M}_{nd}(\mathbb{R})^r$, $\mathrm{M}_{nd}(\mathbb{C})^r$ or $\mathrm{M}_{nd}(\mathbb{H})^r$. Because q is symmetric, its projection onto each simple factor of $E \otimes_{\mathbb{Q}} \mathbb{R}$ is a Hermitian matrix. By the theory of Hermitian forms over $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$, there exist $x, d \in E \otimes_{\mathbb{Q}} \mathbb{R}$ such that d is diagonal with real entries in each factor and

$$q = x^\dagger dx.$$

Since q is positive definite, all the diagonal entries of d are positive so we can multiply each row of x by the square root of the corresponding entry of d to suppose that $d = 1$. We then have $q = x^\dagger x$.

Let G be the $\mathbb{Z}$-group scheme representing the functor

$$G(A) = (R \otimes_{\mathbb{Z}} A)^\times.$$

Over $\mathbb{Q}$ this is the reductive group $\mathrm{Res}_{D/\mathbb{Q}} \mathrm{GL}_n$. We will use the following notations for subgroups of G:

(i) S is the maximal $\mathbb{Q}$-split torus of G whose $\mathbb{Q}$-points are the diagonal matrices of $\mathrm{GL}_n(D)$ with entries in $\mathbb{Q}$;

(ii) P is the minimal parabolic $\mathbb{Q}$-subgroup of G consisting of upper triangular matrices;

(iii) $U = R_u(P)$ is the group of upper triangular matrices with ones on the diagonal;

(iv) Z is the centraliser of S in G; that is, $Z(\mathbb{Q})$ consists of the diagonal matrices in $\mathrm{GL}_n(D)$;

(v) M is the maximal $\mathbb{Q}$-anisotropic subgroup of Z; that is, $M(\mathbb{Q})$ consists of the diagonal matrices in $\mathrm{GL}_n(D)$ whose diagonal entries have reduced norm ± 1;

(vi) $K = \{g \in G(\mathbb{R}) \mid g^\dagger g = 1\}$ is a maximal compact subgroup of $G(\mathbb{R})$.

By Proposition 13.1 of [Bor69], there exist a positive real number t, a finite set $C \subset G(\mathbb{Q})$ and a compact neighbourhood ω of 1 in $M^\circ(\mathbb{R})U(\mathbb{R})$ such that

$$G(\mathbb{R}) = K A_t \omega C G(\mathbb{Z})$$

where

$$A_t = \{a \in S(\mathbb{R}) \mid a_i > 0, a_i/a_{i+1} \leq t \text{ for all } i\}.$$

We note that $M^\circ(\mathbb{R})U(\mathbb{R})$ is the group of upper triangular matrices in $\mathrm{M}_n(D \otimes_{\mathbb{Q}} \mathbb{R})$ whose diagonal entries have reduced norm 1.

Hence we can write

$$x = kaz\nu\gamma$$

where $k \in K$, $a \in A_t$, $z \in \omega$, $\nu \in C$ and $\gamma \in G(\mathbb{Z}) = R^\times$.

Let $u = \gamma^{-1}$ and

$$q' = u^\dagger q u.$$

In order to prove the proposition, it will suffice to show that $H(q') \leq c\,\mathrm{N}(q)$.

Since $k^\dagger k = 1$, and using the decomposition of x, we get that

$$q' = \nu^\dagger z^\dagger a^\dagger a z \nu.$$

Fix some $\mathbb{Z}$-basis of R. We will show below that the (real) coordinates of $a^\dagger a$ with respect to this basis are bounded above by a constant multiple of $\mathrm{N}(q)$. The coordinates of z and ν are uniformly bounded because z is in the compact set ω and ν is in the finite set C. Hence the coordinates of q' in this basis are bounded by a multiple of $\mathrm{N}(q)$, so $H(q')$ is likewise linearly bounded.

Let $a^\dagger a = \operatorname{diag}(a_1, \dots, a_n)$ with $a_i \in \mathbb{R}$. In order to show that the coordinates of $a^\dagger a$ in the chosen basis are bounded, it will suffice to show that the a_i are bounded by a multiple of $\mathrm{N}(q)$. We shall show that the a_i are bounded *below* by a constant, and that their product $\prod a_i$ is bounded above by a multiple of $\mathrm{N}(q)$. These two facts together imply that the a_o are bounded above by a multiple of $\mathrm{N}(q)$.

Choose an integer m such that $m\nu^{-1} \in R$ for all $\nu \in C$. Then

$$m^2 z^\dagger a^\dagger a z = (m\nu^{\dagger -1}) q' (m\nu^{-1}) \in R$$

so every entry of $m^2 z^\dagger a^\dagger a z$, viewed as a matrix in $\mathrm{M}_n(D)$, is in $\mathfrak{o}$.

Let z_{11} denote the upper left entry of $z \in \mathrm{M}_n(D \otimes_\mathbb{Q} \mathbb{R})$. Because z is upper triangular, the upper left entry of $m^2 z^\dagger a^\dagger a z$ is $m^2 z_{11}^\dagger a_1 z_{11}$. So $m^2 z_{11}^\dagger a_1 z_{11} \in \mathfrak{o}$ and

$$\left| \mathrm{Nrd}_{D/\mathbb{Q}}(m^2 z_{11}^\dagger a_1 z_{11}) \right| \geq 1.$$

But $\mathrm{Nrd}(z_{11}) = 1$ because $z \in \omega$, so

$$\left| \mathrm{Nrd}_{D \otimes_\mathbb{Q} \mathbb{R}/\mathbb{R}}(m^2 a_1) \right| \geq 1.$$

Since $m^2 a_1$ is a positive real number, $\mathrm{Nrd}_{D \otimes_\mathbb{Q} \mathbb{R}/\mathbb{R}}(m^2 a_1)$ is just some fixed positive power of $m^2 a_1$ so we conclude that

$$m^2 a_1 \geq 1.$$

From the definition of A_t, it follows that $a_i \geq m^{-2} t^{2-2i}$ for all i and we have established that the a_i are uniformly bounded below.

Hence there is a constant c_1 such that for every j,

$$a_j \leq c_1 \prod a_i.$$

Since ρ is faithful $\dim \rho \geq n$. Together with the fact that $\prod a_i$ is bounded below this implies that

$$\prod a_i \leq c_2 \left(\prod a_i \right)^{\dim \rho / n} = c_2 \, \mathrm{N}(a^\dagger a).$$

Now $\mathrm{N}(z) = \mathrm{N}(u) = 1$ and $\mathrm{N}(\nu)$ is bounded because ν comes from a finite set, so $\mathrm{N}(a^\dagger a)$ is bounded above by a constant multiple of $\mathrm{N}(q)$. Combining all this we have proved that each a_i is bounded above by a constant multiple of $\mathrm{N}(q)$, and as remarked above this suffices to establish the proposition. $\square$

5.2.5 Height of a symplectic basis

We will need the following bound for the height of a symplectic basis for a symplectic free $\mathbb{Z}$-module in terms of the values of the symplectic pairing on the standard basis. The proof is simply to apply the standard recursive algorithm for finding a symplectic basis, verifying that the new vectors introduced always have polynomially bounded heights.

Lemma 5.5. *Let $L = \mathbb{Z}^{2g}$ and let $\{e_1, \dots, e_{2g}\}$ be a basis for L. There exist constants c, k depending only on g such that:*

For any perfect symplectic pairing $\psi\colon L \times L \to \mathbb{Z}$ with

$$N = \max_{i,j} |\psi(e_i, e_j)|,$$

there exists a symplectic basis for (L, ψ) whose coordinates with respect to the basis $\{e_1, \dots, e_{2g}\}$ are at most cN^k.

Proof. For any $x \in L$, we write $H(x)$ for the maximum of the absolute values of the coordinates of x with respect to the basis $\{e_1, \dots, e_{2g}\}$.

First let $e'_1 = e_1$ and choose e'_2 such that $\psi(e'_1, e'_2) = 1$ and $H(e'_2) \leq N$. We can do this because ψ is perfect, so that $\gcd_{i=2}^{n}(\psi(e_1, e_i)) = 1$. Hence there are integers a_i such that $|a_i| \leq N$ and

$$\sum a_i \psi(e_1, e_i) = 1$$

We let $e'_2 = \sum a_i e_i$.

Then find $e'_3, \dots, e'_{2g}$ orthogonal to e'_1 and to e'_2 such that $\{e'_1, \dots, e'_{2g}\}$ is a basis for L and $H(e'_i) \leq 2gN^2$. We can do this by setting

$$e'_i = e_i + \psi(e'_2, e_i)e'_1 + \psi(e'_1, e_i)e'_2.$$

Here we have $|\psi(e'_2, e_i)| \leq \sum_{j=2}^{n} |a_j \psi(e_j, e_i)| \leq (2g-1)N^2$ and $\psi(e'_1, e_i)e'_2$ has height at most N^2 so $H(e'_i) \leq 2gN^2$.

Finally apply the algorithm recursively to $L' = \mathbb{Z}\langle e'_3, \dots, e'_{2g}\rangle$. We have

$$\left|\psi(e'_i, e'_j)\right| \leq gNH(e'_i)H(e'_j) \leq 4g^3N^5.$$

Hence by induction L' has a symplectic basis whose coordinates with respect to $\{e'_3, \dots, e'_{2g}\}$ are bounded by a constant multiple of $N^{5k'}$, where k' is the exponent in the lemma for $\mathbb{Z}^{2(g-1)}$. Converting these into coordinates with respect to $\{e_1, \dots, e_{2g}\}$, we get that the elements of this symplectic basis for L' have height bounded by a constant multiple of $N^{2+5k'}$. This proves the lemma.

We remark that the recurrence $k(g) = 2 + 5k(g-1)$, $k(0) = 0$ is satisfied by $k(g) = (5^g - 1)/2$, so this provides a suitable choice of exponent for the lemma. □

5.2.6 Proof of Proposition 5.3

Let $h\colon A \to A'$ be an isogeny of degree n. There is $q \in \operatorname{End} A$ such that

$$h^*\lambda' = \lambda \circ q.$$

Apply Proposition 5.4 to get $u \in (\operatorname{End} A)^\times$ such that

$$H(u^\dagger q u) \leq c\,\mathrm{N}(q).$$

Then hu is an isogeny $A \to A'$ of degree n, so there is also an isogeny $f\colon A' \to A$ of degree n^{2g-1} such that

$$hu \circ f = [n]_A.$$

The image of $f_*\colon H_1(A', \mathbb{Z}) \to H_1(A, \mathbb{Z})$ is a submodule of index n^{2g-1}. By the structure theory of finitely generated $\mathbb{Z}$-modules there is a basis $\{e'_1, \dots, e'_{2g}\}$ for $H_1(A', \mathbb{Z})$ with respect to which the rational representation of f is upper triangular and has height at most n^{2g-1}. But this need not be a symplectic basis.

Let ψ, ψ' be the symplectic forms on $H_1(A, \mathbb{Z})$ and $H_1(A', \mathbb{Z})$ induced by λ, λ' respectively. Let $q' = u^\dagger q u$. Then

$$n^2\lambda' = [n]_{A'}^*\lambda' = f^*u^*h^*\lambda' = f^*(\lambda \circ q').$$

In terms of symplectic forms this says that

$$n^2\psi'(x, y) = \psi(f_*x, q'_*f_*y).$$

In particular, since the entries of the matrix q'_* and the coordinates with respect to $\mathcal{B}$ of $\{f_*e'_1, \dots, f_*e'_{2g}\}$ are bounded by a polynomial in n, the values $\psi'(e'_i, e'_j)$ are linear combinations of $\psi(e_i, e_j)$ (for $e_i, e_j \in \mathcal{B}$) with polynomially bounded coefficients. Since $\mathcal{B}$ is a symplectic basis for ψ, we conclude that the values

$$\left|\psi'(e'_i, e'_j)\right|$$

are polynomially bounded.

Hence by Lemma 5.5 there is a symplectic basis $\mathcal{B}'$ for $H_1(A', \mathbb{Z})$ whose coordinates with respect to $\{e'_1, \dots, e'_{2g}\}$ are polynomially bounded. Using again that the coordinates with respect to $\mathcal{B}$ of $\{f_*e'_1, \dots, f_*e'_{2g}\}$ are polynomially bounded, we deduce that $H(f, \mathcal{B}', \mathcal{B})$ is also polynomially bounded. $\square$

5.3 Degree bound for polarised isogenies

Let (A, λ) be a principally polarised abelian variety defined over a number field. If B is another abelian variety isogenous to A, then the Masser–Wüstholz isogeny theorem gives an upper bound for the minimum degree of an isogeny $A \to B$. But if we make the stronger assumption that (B, λ') is in the same polarised isogeny class as A, the isogeny of small degree given by Masser–Wüstholz will not necessarily be a polarised isogeny. In this section, we prove that in such a situation, there does exist a polarised isogeny of small degree (polynomial in the Masser–Wüstholz bound).

This offers an alternative approach to the proof of Proposition 5.3: section 5.2 can be replaced by a much simpler bound for the height of a rational representation of a polarised isogeny, and in the proof of Proposition 6.1 we can work directly with the action of $\mathrm{GSp}_{2g}(\mathbb{R})^+$ on $\mathcal{H}_g$ instead of working with a partially defined action of $\mathrm{GL}_{2g}(\mathbb{R})$. Overall this seems to lead to a longer proof of Proposition 6.1, but we have included the theorem in case it is of independent interest.

Theorem 5.6. *Let (A, λ) be a principally polarised abelian variety defined over a field of characteristic 0. There exist constants c, k depending only on (A, λ) such that, if (B, λ') is a principally polarised abelian variety for which*

1. *there exists an isogeny $f : A \to B$ compatible with the polarisations (of any degree), and*
2. *there exists an isogeny $g : A \to B$ of degree n (not necessarily compatible with the polarisations),*

then there exists an isogeny $h : A \to B$ compatible with the polarisations and of degree at most cn^k.

We begin by reducing the theorem to an algebraic problem in $\operatorname{End} A$. Under the conditions of the theorem, let $a = g^{-1} \circ f \in \operatorname{End} A \otimes \mathbb{Q}$. Let q be in $\operatorname{End} A$ such that $g^* \lambda' = \lambda \circ q$. Then condition (1) says that $a^\dagger q a \in \mathbb{Z}$. To prove the theorem, it will suffice to find $b \in \operatorname{End} A$ such that $b^\dagger q b$ is also in $\mathbb{Z}$ and $\deg b$ is bounded by a polynomial in n: then $h = g \circ b$ satisfies the required conditions. Using Proposition 5.7 gives an exponent $k = 4 \dim A$ in Theorem 5.6.

As for Proposition 5.4, the fact that there exists some function $C(A, \lambda, n)$ such that there is a $b \in R$ satisfying $b^\dagger q b \in \mathbb{Z} - \{0\}$ and $\mathrm{N}_E(b) \leq C(A, \lambda, n)$ is an immediate consequence of the theorem [Mil86] that there are finitely many polarisations of A of given degree (up to isomorphisms of polarised abelian varieties). The content of the theorem is that the bound is polynomial in n.

Proposition 5.7. *Let $(E, \dagger)$ be a semisimple $\mathbb{Q}$-algebra with involution, $R \subset E$ a $\dagger$-stable order, and ρ a faithful $\mathbb{Z}$-representation of R of rank d.*

There exists a constant c depending only on $(R, \dagger, \rho)$ such that: For every $q \in R$, if there exists $a \in E$ such that $a^\dagger q a \in \mathbb{Q}^\times$, then there exists $b \in R$ such that

$$b^\dagger q b \in \mathbb{Z} - \{0\} \text{ and } \mathrm{N}_E(b) \leq c\, \mathrm{N}_E(q)^{d-1/2}.$$

As in Proposition 5.4, the representation ρ in the proposition is purely a technical device which we use to define a norm on E. Specifically, if E is a semisimple $\mathbb{Q}$-algebra with faithful $\mathbb{Q}$-representation ρ, then we define N_E to be the function $E \to \mathbb{Q}$ given by

$$\mathrm{N}_E(x) = |\det \rho(x)| \,.$$

The norms associated with different representations are polynomially bounded with respect to each other. In our application, we take ρ to be the rational representation of $\operatorname{End} A$ on $H_1(A, \mathbb{Z})$; then $\mathrm{N}_E(f) = \deg f$ whenever $f \colon A \to A$ is an isogeny.

We also define norms for semisimple $\mathbb{Q}_p$-algebras: if E_p is a semisimple $\mathbb{Q}_p$-algebra with faithful $\mathbb{Q}_p$-representation ρ_p, then N_{E_p} denotes the function $E_p \to \mathbb{Q}$ given by

$$\mathrm{N}_{E_p}(x) = |\det \rho_p(x)|_p^{-1} \,.$$

Thus the norms are always rational and satisfy

$$\mathrm{N}_E(x) = \prod_p \mathrm{N}_{E \otimes \mathbb{Q}_p}(x) \text{ for all } x \in E.$$

If R is a maximal order in E and ρ is an integer representation of R, then an element $x \in R$ is invertible in R if and only if $\mathrm{N}_E(x) = 1$.

Before looking at the proof of Proposition 5.7 in general, let us first consider the case in which E is a number field. By looking at the prime factorisation of the ideal (a), we can show that there is an ideal $\mathfrak{b} \subset R$ of norm polynomially bounded by $\mathrm{N}_E(q)$ such that $\mathfrak{b}^\dagger q \mathfrak{b} = (n)$ for some $n \in \mathbb{Z}$. Using finiteness of the class group, we may suppose that the ideal $\mathfrak{b}$ is principal with the cost of a constant factor in the norm bound; thus there are $b \in \mathfrak{o}_E$, $u \in \mathfrak{o}_E^\times$ and $n \in \mathbb{Z}$ such that $b^\dagger q b = un$. Using Dirichlet's unit theorem we can remove the unit u at the cost of another constant factor.

Such an argument using ideals is difficult to generalise to noncommutative E. Instead, we will first prove the result in each localisation $R \otimes_{\mathbb{Z}} \mathbb{Z}_p$, then use the adelic formulation of the finiteness of the class group to pass from local to global results.

Specifically, we will prove the following local result. Part (1) is simply the local version of Proposition 5.7. Part (2) is needed because the global constant c in Proposition 5.7 depends on the product of the c_p, so in order to prove Proposition 5.7 we must have that $c_p = 1$ for almost all p. Observe that if $R_p = R \otimes_{\mathbb{Z}} \mathbb{Z}_p$ then the conditions of part (2) are indeed satisfied for all but finitely many p.

Proposition 5.8. *Let $(E_p, \dagger)$ be a semisimple $\mathbb{Q}_p$-algebra with involution, $R_p \subset E_p$ a $\dagger$-stable order and ρ_p a faithful $\mathbb{Z}_p$-representation of R_p of rank d.*

(1) There exists a constant $c_p \in \mathbb{N}$ depending only on $(R_p, \dagger, \rho_p)$ such that: For every $q \in R_p$, if there exists $a \in E_p$ such that $a^\dagger q a \in \mathbb{Q}_p^\times$, then there exists $b_p \in R_p$ such that

$$b_p^\dagger q b_p \in \mathbb{Z}_p - \{0\} \text{ and } \mathrm{N}_{E_p}(b_p) \leq c_p \, \mathrm{N}_{E_p}(q)^{d-1/2}.$$

(2) If E_p is split (that is, it is isomorphic to a direct product of matrix algebras over fields), its centre is a product of unramified extensions of $\mathbb{Q}_p$, R_p is a maximal order in E_p and $p \neq 2$, then (1) holds with $c_p = 1$.

We will deduce Proposition 5.8 from the following simpler lemma. One case will also require Lemma 5.10.

Lemma 5.9. *Let $(E_p, \dagger)$ be a semisimple $\mathbb{Q}_p$-algebra with involution and $R_p \subset E_p$ a $\dagger$-stable order.*

(1) There exists a constant $c_{p,1} \in \mathbb{N}$ depending only on $(R_p, \dagger)$ such that: For every $x \in R_p$, if $x = yy^\dagger$ for some $y \in E_p$, then $c_{p,1} x = zz^\dagger$ for some $z \in R_p$.

(2) If E_p is split, its centre is a product of unramified extensions of $\mathbb{Q}_p$, R_p is a maximal order in E_p and $p \neq 2$, then (1) holds with $c_{p,1} = 1$.

Lemma 5.10. *Let $(E_p, \dagger)$ be a semisimple $\mathbb{Q}_p$-algebra with involution and $R_p \subset E_p$ a $\dagger$-stable order. We suppose that E_p is split, its centre is a product of unramified extensions of $\mathbb{Q}_p$, R_p is a maximal order in E_p and $p \neq 2$.*

For every $q \in R_p$, if $\mathrm{N}_{E_p}(q) = 1$ (in other words $q \in R_p^\times$) and there exists $a \in R_p$ such that $a^\dagger q a \in \mathbb{Q}_p^\times$, then there exist $y \in R_p$ and $u \in \mathbb{Z}_p^\times$ such that $uq = yy^\dagger$.

We begin by proving that Proposition 5.8 follows from Lemmas 5.9 and 5.10. The proofs of parts (1) and (2) of Lemma 5.9 (which are independent of each other) follow in the subsequent sections, as well as Lemma 5.10.

Proof of Proposition 5.8. Let

$$c_{p,2} = [\rho_p(E_p) \cap \mathrm{M}_d(\mathbb{Z}_p) : \rho_p(R_p).]$$

In part (2), R_p is a maximal order in E_p and so $c_{p,2} = 1$.

Let $n = \mathrm{N}_{E_p}(q)$ and $m = a^\dagger q a$. Both n and m are in $\mathbb{Q}_p^\times$ so there is some $e \in \mathbb{Z}_p$ such that $c_{p,2} n m^{-1} e$ is a square in $\mathbb{Q}_p^\times$ and $v_p(e) = 0$ or 1. Choose $s \in \mathbb{Q}_p^\times$ such that

$$s^2 = c_{p,2} n m^{-1} e.$$

Now

$$q = a^{\dagger -1} m a^{-1} = a^{\dagger -1} a^{-1} m$$

since m is in the centre of E_p, and so

$$c_{p,2} n e q^{-1} = s^2 m q^{-1} = s^2 a a^\dagger = (sa)(sa)^\dagger.$$

Furthermore $n\rho_p(q)^{-1} \in \mathrm{M}_d(\mathbb{Z}_p)$, hence the same is true of $n\rho_p(eq^{-1})$. Therefore $c_{p,2} n \rho_p(eq^{-1}) \in \rho_p(R_p)$, and so

$$c_{p,2} n e q^{-1} \in R_p.$$

Hence by Lemma 5.9, there is $b_p \in R_p$ such that

$$c_{p,1} c_{p,2} n e q^{-1} = b_p b_p^\dagger.$$

Reversing the above calculations, we get that

$$b_p^\dagger q b_p = c_{p,1} c_{p,2} n e \in \mathbb{Z}_p - \{0\}.$$

It remains to bound $\mathrm{N}_{E_p}(b_p)$. We have

$$\begin{aligned} \mathrm{N}_{E_p}(b_p)^2 &= \mathrm{N}_{E_p}(c_{p,1} c_{p,2} n e q^{-1}) \\ &= c_{p,1}^d c_{p,2}^d n^d \, \mathrm{N}_{E_p}(e) n^{-1} \\ &\leq c_{p,1}^d c_{p,2}^d p^d n^{d-1} \qquad (*) \end{aligned}$$

because $\mathrm{N}_{E_p}(e) = |e|_p^{-d} \leq p^d$.

Thus part (1) of the proposition is proved, with $c_p = (c_{p,1} c_{p,2} p)^{d/2}$. We get an exponent for n of $(d-1)/2$, which is better than the claimed $d - 1/2$. The weaker exponent is needed for part (2).

For part (2), there are two cases. First if q is not invertible in R_p, then we can reuse the above argument. Recall that $c_{p,1} = c_{p,2} = 1$ in part 2. Since q is not invertible, $n = \mathrm{N}_{E_p}(q) \geq p$ and so the bound (*) becomes

$$\mathrm{N}_{E_p}(b_p)^2 \leq n^{2d-1}$$

as required.

Finally if q is invertible in R_p, it seems not to be possible to control $\mathrm{N}_{E_p}(e)$ in the argument above. So instead we apply Lemma 5.10 to $q^{-1} \in R_p^\times$ to get $u \in \mathbb{Z}_p^\times$ and $b_p \in R_p$ such that

$$b_p b_p^\dagger = u q^{-1}.$$

Thus

$$b_p^\dagger q b_p = u \in \mathbb{Z}_p - \{0\}$$

and

$$\mathrm{N}_{E_p}(b_p)^2 = \mathrm{N}_{E_p}(u q^{-1}) = 1. \qquad \square$$

5.3.1 Local calculations – non-split case

We prove part (1) of Lemma 5.9. First we need a lemma.

Lemma 5.11. *Let E be a semisimple $\mathbb{Q}_p$-algebra and $R \subset E$ an order. Let G be a subgroup of $E^\times$ isomorphic to $\mathbb{Q}_p^{\times r}$ for some r.*

There is a positive integer n such that: for all $a \in G$, if $a^2 \in R$ then $na \in R$.

Proof. Choose a faithful representation $\rho\colon R \to \mathrm{GL}_d(\mathbb{Q}_p)$. Then $\rho(G)$ is a subgroup of $\mathrm{GL}_m(\mathbb{Q}_p)$ isomorphic to $\mathbb{Q}_p^{\times r}$ and so is $\mathrm{GL}_d(\mathbb{Q}_p)$-conjugate to a subgroup of the diagonal matrices of $\mathrm{GL}_d(\mathbb{Q}_p)$. Replacing ρ by this conjugate, we may assume that $\rho(G)$ is contained in the diagonal matrices.

Since R is an order in E, $\rho(R)$ and $\rho(E) \cap \mathrm{M}_m(\mathbb{Z}_p)$ are commensurable. Let

$$n_1 = [\rho(R) : \rho(R) \cap \mathrm{M}_d(\mathbb{Z}_p)] \text{ and } n_2 = [\rho(E) \cap \mathrm{M}_d(\mathbb{Z}_p) : \rho(R) \cap \mathrm{M}_d(\mathbb{Z}_p)].$$

Since $a^2 \in R$, we get $n_1 \rho(a)^2 \in \rho(R) \cap \mathrm{M}_d(\mathbb{Z}_p)$. A fortiori $n_1^2 \rho(a)^2 \in \mathrm{M}_d(\mathbb{Z}_p)$. Furthermore $\rho(a)$ is a diagonal matrix. Thus $n_1\rho(a)$ is a diagonal matrix whose square has entries in $\mathbb{Z}_p$. It follows that $n_1\rho(a)$ itself has entries in $\mathbb{Z}_p$.

So $n_1\rho(a) \in \rho(E) \cap \mathrm{M}_d(\mathbb{Z}_p)$ and hence $n_2 n_1 \rho(a) \in \rho(R) \cap \mathrm{M}_d(\mathbb{Z}_p)$. Thus

$$n_2 n_1 a \in R$$

and the lemma is proved with $n = n_2 n_1$. □

Lemma 5.9 (1). *Let $(E_p, \dagger)$ be a semisimple $\mathbb{Q}_p$-algebra with involution and $R_p \subset E_p$ a $\dagger$-stable order.*

There exists a constant $c_{p,1} \in \mathbb{N}$ depending only on $(R_p, \dagger)$ such that: For every $x \in R_p$, if $x = yy^\dagger$ for some $y \in E_p$, then $c_{p,1} x = zz^\dagger$ for some $z \in R_p$.

Proof. Let G be the $\mathbb{Z}_p$-group scheme of invertible elements of R_p

$$G(A) = (R_p \otimes_{\mathbb{Z}_p} A)^\times$$

and H the unitary subgroup scheme

$$H(A) = \{h \in G(A) \mid hh^\dagger = 1\}.$$

We say that a torus $S \subset G$ is $(\mathbb{Q}_p, \dagger)$**-split** if it is split over $\mathbb{Q}_p$ and $s^\dagger = s$ for all $s \in S$. By a theorem of Helminck and Wang [HW93] there are finitely many $H(\mathbb{Q}_p)$-conjugacy classes of maximal $(\mathbb{Q}_p, \dagger)$-split tori in G. Choose representatives A_i for these $H(\mathbb{Q}_p)$-conjugacy classes. Note that each $(\mathbb{Q}_p, \dagger)$-split torus (in particular each A_i) contains the scalars $\mathbb{G}_{m,\mathbb{Q}_p}$.

By the p-adic polar decomposition of Benoist and Oh [BO07], there is a compact subset $K \subset G(\mathbb{Q}_p)$ such that

$$G(\mathbb{Q}_p) = K \left(\bigcup_i A_i(\mathbb{Q}_p) \right) H(\mathbb{Q}_p).$$

Since K is compact, it has bounded denominators; so replacing it by a rational multiple, we may assume that $K \subset R_p$. Since K^{-1} also has bounded denominators, we can choose $c_{p,3} \in \mathbb{N}$ such that $c_{p,3}K^{-1} \subset R_p$.

Using the decomposition, write

$$y = kah$$

with $k \in K$, $a \in \bigcup_i A_i(\mathbb{Q}_p)$ and $h \in H(\mathbb{Q}_p)$.

Using that $hh^\dagger = 1$ and $a^\dagger = a$, we therefore have

$$x = ka^2k^\dagger$$

so

$$c_{p,3}^2 a^2 = (c_{p,3}k^{-1})x(c_{p,3}k^{\dagger-1}) \in R_p.$$

By Lemma 5.11 we can choose $c_{p,4} \in \mathbb{N}$ for all $x \in \bigcup_i A_i(\mathbb{Q}_p)$, if $x^2 \in R_p$, then $c_{p,4}x \in R_p$. In particular $c_{p,3}a \in \bigcup_i A_i(\mathbb{Q}_p)$ and its square is in R_p, so

$$c_{p,4}c_{p,3}a \in R_p.$$

Let

$$z = kc_{p,4}c_{p,3}a \in R_p.$$

Then

$$zz^\dagger = c_{p,4}^2 c_{p,3}^2 x$$

and the result is proved with $c_{p,1} = c_{p,4}^2 c_{p,3}^2$. □

5.3.2 Local calculations – split case

Now we prove part (2) of Lemma 5.9 and Lemma 5.10.

Lemma 5.9 (2). *Let $(E_p, \dagger)$ be a semisimple $\mathbb{Q}_p$-algebra with involution, $R_p \subset E_p$ a $\dagger$-stable order. Suppose further that E_p is split, its centre is a product of unramified extensions of $\mathbb{Q}_p$, R_p is a maximal order in E_p and $p \neq 2$.*

For every $x \in R_p$, if $x = yy^\dagger$ for some $y \in E_p$, then $x = zz^\dagger$ for some $z \in R_p$.

Proof. Write the algebra E_p as a product

$$E_p = \prod_i E_i$$

such that each E_i is either simple and preserved by † or is a product of two simple algebras which are exchanged by †. Since R_p is a maximal order in E_p, it is a direct product of maximal orders $R_i \subset E_i$. Hence it suffices to prove the lemma for a single E_i.

Let F be the centre of E_i and F_0 the subfield of F fixed by †. Since E_i is split, it is isomorphic to $\mathrm{M}_n(F)$ for some n. Let $V = F^n$. According to [KMRT98] Propositions 2.14 and 4.2, there is an F_0-bilinear form $h\colon V \times V \to F$ such that † is the adjoint involution with respect to h. This bilinear form may be of the following types (labelled according to the types of algebras in the Albert classification for which E_i can arise as a localisation).

Type I/II. $F = F_0$ and h is symmetric;

Type III. $F = F_0$ and h is skew-symmetric;

Type IVa. F is a quadratic extension of F_0 and h is Hermitian with respect to the non-trivial element of $\mathrm{Gal}(F/F_0)$;

Type IVb. $F = F_0 \times F_0$ and h is Hermitian with respect to the automorphism of F which exchanges the two factors.

Let Λ be a lattice in V whose stabiliser is R_i, and let $\mathfrak{a}$ be the norm ideal of Λ, that is, the ideal in F_0 generated by $\{h(v,v) \mid v \in \Lambda\}$. Because $p \neq 2$ and F/F_0 is unramified, the ideal in F generated by $\{h(v,w) \mid v, w \in \Lambda\}$ is equal to $\mathfrak{a}\mathfrak{o}_F$.

We claim that Λ is a maximal lattice with respect to h – that is, there is no lattice properly containing Λ and also having norm ideal $\mathfrak{a}$. Let

$$\Lambda^{\mathfrak{a}} = \{v \in V \mid h(v,w) \in \mathfrak{a}\mathfrak{o}_F \text{ for all } w \in \Lambda\}$$

and observe that any lattice containing Λ with norm ideal $\mathfrak{a}$ must be contained in $\Lambda^{\mathfrak{a}}$. Hence in order to prove that Λ is maximal, it suffices to show that $\Lambda^{\mathfrak{a}} = \Lambda$.

Since R_i is †-stable, we have

$$h(R_i v, w) = h(v, R_i w) \subset \mathfrak{a} \text{ for all } v \in \Lambda^{\mathfrak{a}} \text{ and } w \in \Lambda.$$

Hence R_i stabilises $\Lambda^{\mathfrak{a}}$. Thus Λ and $\Lambda^{\mathfrak{a}}$ are lattices in V with the same stabiliser. It follows that $\Lambda^{\mathfrak{a}} = u\Lambda$ for some scalar $u \in F^\times$. But then the ideal generated by $h(\Lambda^{\mathfrak{a}}, \Lambda)$ is $u\mathfrak{a}\mathfrak{o}_F$, so $u \in \mathfrak{o}_F^\times$ and $\Lambda^{\mathfrak{a}} = \Lambda$.

Since $yy^\dagger = x \in R_i$, we have

$$h(y^\dagger u, y^\dagger v) = h(xu, v) \in \mathfrak{a} \text{ for all } u, v \in \Lambda.$$

Hence $y^\dagger\Lambda$ is a lattice on which h takes values in $\mathfrak{a}$.

So $y^\dagger\Lambda$ is contained in an $\mathfrak{a}$-maximal lattice Λ' ([Shi97] Lemma 4.8 for types I/II, IVa and IVb; text of [Shi63] for type III).

Any two $\mathfrak{a}$-maximal lattices in V are isometric (by [Shi97] Lemma 5.9 for types I/II and IVa, [Shi97] Lemma 4.12 for type IVb and [Shi63] Proposition 1.4 for type III). In particular Λ is isometric to Λ'.

Hence Λ contains a lattice Λ_x which is isometric to $y^\dagger\Lambda$. Thus there is $z \in E_i^\times$ such that $z\Lambda = \Lambda_x$ and $x = zz^\dagger$. Since $z\Lambda \subset \Lambda$, we have $z \in R_i$. □

Lemma 5.12. *Let $(E_p, \dagger)$ be a semisimple $\mathbb{Q}_p$-algebra with involution and $R_p \subset E_p$ a $\dagger$-stable order. We suppose that E_p is split, its centre is a product of unramified extensions of $\mathbb{Q}_p$, R_p is a maximal order in E_p and $p \neq 2$.*

For every $q \in R_p$, if $\mathrm{N}_{E_p}(q) = 1$ (in other words $q \in R_p^\times$) and there exists $a \in R_p$ such that $a^\dagger qa \in \mathbb{Q}_p^\times$, then there exist $y \in R_p$ and $u \in \mathbb{Z}_p^\times$ such that $uq = yy^\dagger$.

Proof. We use the same notation E_i, R_i, F, F_0, V, h as in the proof of Lemma 5.9 (2). Let Λ be the lattice $\mathfrak{o}_F^n$ in $V = F^n$. The claim that there exist $y \in R_i$ and $u \in \mathbb{Z}_p^\times$ such that $uq = yy^\dagger$ says that, for some $u \in \mathbb{Z}_p^\times$, the restriction to Λ of the bilinear form

$$h_{uq} \colon (v, w) \mapsto h(v, uqw)$$

is integrally equivalent to h. Observe that the condition $a^\dagger qa \in \mathbb{Q}_p^\times$ implies that $(uq)^\dagger = uq$ so h_{uq} is of the same type (symmetric, skew-symmetric or Hermitian) as h.

By Lemma 5.9 (2), it suffices to show that $uq = yy^\dagger$ for some $y \in E_i$. Thus we only need to show that h and h_{uq} are rationally equivalent. In order to prove rational equivalence, the condition $q \in R_i$ is not required save in the case of type I/II.

We proceed in cases according to the type of h as in Lemma 5.9 (2).

Type I/II. The condition that $q \in R_p^\times$ implies that Λ is a unimodular lattice for h_q. According to [O'M63] 92:1, isometry classes of unimodular lattices are classified by the determinant of the quadratic form modulo squares. Thus it will suffice to show that there is some $u \in \mathbb{Z}_p^\times$ such that $\det(uq)$ is a square in $F^\times$.

Let $m = a^\dagger qa \in \mathbb{Q}_p^\times$. Then

$$m^d = \det m = (\det a)^2 \det q.$$

If d is even then we see immediately that $\det q$ is a square.

If d is odd: $v_p(\det q) = 0$ since $q \in R_p^\times$ and $v_p(\det a)$ is an integer because $F/\mathbb{Q}_p$ is unramified. So $v_p(\det m)$ is even. Hence there is $u \in \mathbb{Z}_p^\times$ such that $u \det m$ is a square in $\mathbb{Q}_p^\times$. Because d is odd, $\det(um) = u^d \det m$ is also a square in $\mathbb{Q}_p^\times$ and so $\det(uq)$ is a square in $F^\times$.

Type III. All symplectic forms of given dimension over a field are equivalent, in particular h and h_q.

Type IVa. Over local fields F, F_0, equivalence classes of Hermitian forms of dimension n are classified by their determinant in $F_0^\times / \mathrm{N}_{F/F_0}(F^\times)$ ([Sch85] 10.1.6(ii)).

Since the extension F/F_0 is unramified, $\mathrm{N}_{F/F_0}(F^\times)$ contains the unit group $\mathfrak{o}_{F_0}^\times$ so $\det(q) \in \mathrm{N}_{F/F_0}(F^\times)$ and h_q is equivalent to h.

Type IVb. By [KMRT98] Proposition 2.14, there is an isomorphism $E_i \cong \mathrm{M}_d(F_0) \times \mathrm{M}_d(F_0)$ under which the involution $\dagger$ becomes

$$(A, B)^\dagger = (B^t, A^t).$$

Since $q = q^\dagger$, q must have the form (A, A^t) for some $A \in \mathrm{M}_d(F_0)$. Then taking $y = (A, 1)$ gives $q = yy^\dagger$. $\square$

5.3.3 Global arguments

We deduce Proposition 5.7 from Proposition 5.8 using reduction theory in the adelic points of the $\mathbb{Z}$-group scheme U of $\dagger$-quasi-unitary elements of R. That is,

$$U(A) = \{u \in (R \otimes_{\mathbb{Z}} A)^\times \mid uu^\dagger \in A^\times\}.$$

Proof of Proposition 5.7. By [PR94] Theorem 5.1, the double coset space

$$U(\mathbb{Q}) \backslash U(\mathbb{A}^f) / \prod U(\mathbb{Z}_p)$$

is finite. Choose representatives $\mathbf{g}_1, \dots, \mathbf{g}_r$ for these double cosets. By multiplying them by suitable elements of $\mathbb{Q}^\times$, we clear denominators so that $\mathbf{g}_i^{-1} \in \prod(R \otimes_{\mathbb{Z}} \mathbb{Z}_p)$ for each i.

By the hypothesis of the proposition, we have $q \in R$ and $a \in E$ such that

$$a^\dagger q a \in \mathbb{Q}^\times.$$

Hence we can apply Proposition 5.8 (1) to q in each localisation $R_p = R \otimes_{\mathbb{Z}} \mathbb{Z}_p$. So there is $b_p \in R_p$ such that

$$b_p^\dagger q b_p \in \mathbb{Z}_p - \{0\} \text{ and } \mathrm{N}_{E_p}(b_p) \leq c_p \, \mathrm{N}_{E_p}(q)^{d-1/2}.$$

For almost all p, Proposition 5.8 (2) says that $c_p = 1$. Also for almost all p, $\mathrm{N}_{E_p}(q) = 1$ and R_p is a maximal order in E_p. It follows that $b_p \in R_p^\times$ for almost all p.

Let

$$u_p = a^{-1} b_p \in E_p^\times .$$

We claim that $u_p \in U(\mathbb{Q}_p)$. To prove this, let $m = a^\dagger q a$. Then $a^{\dagger -1} m a^{-1} = q$. Since $m \in \mathbb{Q}^\times$ is in the centre of R, we deduce that

$$a^{\dagger -1} a^{-1} = m^{-1} q.$$

It follows that

$$u_p^\dagger u_p = b_p^\dagger a^{\dagger -1} a^{-1} b_p = b_p^\dagger m^{-1} q b_p = m^{-1} (b_p^\dagger q b_p) \in \mathbb{Q}_p^\times .$$

We have $a \in R_p^\times$ and $b_p \in R_p^\times$ for almost all p, and so $u_p \in U(\mathbb{Z}_p)$ for almost all p. Hence $\mathbf{u} \in U(\mathbb{A}^f)$ and we can decompose it as

$$\mathbf{u} = x \mathbf{g}_i \mathbf{y}$$

for some $x \in U(\mathbb{Q})$, $\mathbf{g}_i$ among our chosen set of double coset representatives and $\mathbf{y} \in \prod U(\mathbb{Z}_p)$.

We claim that $b' = ax \in E$ satisfies the conditions for b in the proposition.

For each prime p, we have that

$$b' = b_p y_p^{-1} g_{i,p}^{-1}$$

which is a product of elements of R_p, so is in R_p. Hence $b' \in \bigcap R_p = R$.

Now

$$b'^\dagger q b' = x^\dagger a^\dagger q a x = (a^\dagger q a) x^\dagger x.$$

The second equality holds because $a^\dagger q a \in \mathbb{Q}^\times$, which is in the centre of E. But we know also that $x^\dagger x \in \mathbb{Q}^\times$ because $x \in U(\mathbb{Q})$, so

$$b'^\dagger q b' \in \mathbb{Q}^\times .$$

In fact $b'^\dagger q b' \in \mathbb{Z} - \{0\}$ because $b', q \in R$ and $R \cap \mathbb{Q}^\times = \mathbb{Z} - \{0\}$.

Finally

$$\mathrm{N}_E(b') = \prod_p \left(\mathrm{N}_{E_p}(b_p) \, \mathrm{N}_{E_p}(y_p)^{-1} \, \mathrm{N}_{E_p}(g_{i,p})^{-1} \right).$$

We have that $\mathrm{N}_{E_p}(y_p) = 1$ for all p because $y_p \in R_p^\times$.

For each i, $\mathrm{N}_{E_p}(g_{i,p}) = 1$ for almost all p, and so the following constant is well-defined:

$$c_\infty = \max_{1 \le i \le r} \prod_p \mathrm{N}_{E_p}(g_{i,p})^{-1}.$$

Recall that we can choose $\mathbf{g}_i$ depending only on $(R, \dagger)$ so c_∞ depends only on $(R, \dagger, \rho)$.

So using the bound on $\mathrm{N}_{E_p}(b_p)$ from Proposition 5.8, we get that

$$\mathrm{N}_E(b') \leq c_\infty \prod_p \left(c_p \, \mathrm{N}_{E_p}(q)^{d-1/2}\right) = c_\infty \left(\prod_p c_p\right) \mathrm{N}_E(q)^{d-1/2}$$

and the proposition holds with $c = c_\infty \prod c_p$. ☐

6 Some cases of the André–Pink conjecture

In this chapter we will prove our main theorems, namely the André–Pink conjecture for Shimura varieties of abelian type in the following cases:

(i) the subvariety Z is a curve;

(ii) the point s is Hodge generic in the smallest special subvariety of S containing Z;

(iii) the generalised Hecke orbit Σ is replaced by a P-Hecke orbit.

We begin by restricting our attention to $\mathcal{A}_g$, because our Galois bounds (from the Masser–Wüstholz isogeny theorem) rely heavily on the theory of abelian varieties; furthermore the hyperbolic Ax–Lindemann–Weierstrass conjecture and the definability of the uniformisation $X^+ \to S$ are only known for $\mathcal{A}_g$.

Let $\mathcal{F}_g$ denote the Siegel fundamental set in $\mathcal{H}_g$, and let $\widetilde{Z}$ be the preimage of Z in $\mathcal{F}_g$. We shall use a strong form of the Pila–Wilkie counting theorem and our isogeny bounds to prove that, if a Hecke orbit intersects Z densely, then the complex algebraic part of $\widetilde{Z}$ has Zariski dense image in Z. (In fact, our argument is valid without change for isogeny classes and not merely Hecke orbits in $\mathcal{A}_g$. Because we can always map an isogeny class into a Hecke orbit via the embedding $\mathcal{A}_g \to \mathcal{A}_{4g}$, we do not gain anything from this generalisation.)

We then apply the hyperbolic Ax–Lindemann–Weierstrass conjecture for $\mathcal{A}_g$, to deduce that the weakly special part of Z is Zariski dense in Z. Using Theorem 4.6 we can extend this result from $\mathcal{A}_g$ to all Shimura varieties of abelian type.

We can deduce directly that Z is weakly special if it is a curve. In other cases we get a diagram as in the definition of a weakly special subvariety:

$$\begin{array}{ccc} S' & \xrightarrow{[\iota]} & S \\ {\scriptstyle [\varphi]}\downarrow & & \\ S_2 & & \end{array}$$

with $[\iota]$ a Shimura immersion and φ a Shimura submersion, but we can only conclude that Z has the form $[\iota]([\varphi]^{-1}(Z_2))$ where Z_2 is a subvariety of S_2, not necessarily a single point in S_2 as for a weakly special subvariety.

In cases (ii) and (iii) above, our results on functoriality of Hecke orbits show there is a generalised Hecke orbit in S' which intersects $[\varphi]^{-1}(Z_2)$ densely. We can then deduce that there is a generalised Hecke orbit in S_2 which intersects Z_2 densely, and prove by induction that Z is weakly special. In other cases the preimage by $[\iota]$ of a generalised Hecke orbit in S may intersect infinitely many generalised Hecke orbits in S' and we are unable to complete the proof of the André–Pink conjecture.

6.1 Application of the Pila–Wilkie theorem

In this section we will prove the following proposition, using a strong form of the Pila–Wilkie counting theorem and the isogeny bounds of Theorem 5.2 and Proposition 5.3. In the next section, we will deduce cases of the André–Pink conjecture from the proposition using the hyperbolic Ax–Lindemann–Weierstrass conjecture.

Proposition 6.1. *Let $\pi\colon \mathcal{H}_g \to \mathcal{A}_g$ be the uniformisation map, and let $\mathcal{F}_g$ be the Siegel fundamental set in $\mathcal{H}_g$. Let Z be a closed subvariety of $\mathcal{A}_g$ and let $\tilde{Z} = \pi^{-1}(Z) \cap \mathcal{F}_g$. Let $\tilde{Z}^{\mathrm{ca}}$ denote the complex algebraic part of $\tilde{Z}$.*

Let Σ be the Hecke orbit of a point $s \in \mathcal{A}_g$. If $Z \cap \Sigma$ is Zariski dense in Z, then $\pi(\tilde{Z}^{\mathrm{ca}})$ is also Zariski dense in Z.

6.1.1 Outline of proof of Proposition 6.1

We begin with some definitions and notation. Fix a point $s \in \mathcal{A}_g(\mathbb{C})$ and let Σ be its isogeny class. Let $Z \subset \mathcal{A}_g$ be an irreducible closed algebraic subvariety such that $Z \cap \Sigma$ is Zariski dense in Z.

Let $\pi\colon \mathcal{H}_g \to \mathcal{A}_g$ denote the uniformisation map and $\mathcal{F}_g$ the Siegel fundamental set in $\mathcal{H}_g$ for the action of $\mathrm{Sp}_{2g}(\mathbb{Z})$. Let

$$\widetilde{Z} = \pi^{-1}(Z) \cap \mathcal{F}_g \quad \text{and} \quad \widetilde{\Sigma} = \pi^{-1}(\Sigma) \cap \mathcal{F}_g.$$

Fix a point $\tilde{s} \in \mathcal{H}_g$ such that $\pi(\tilde{s}) = s$.

We define the **complexity** of a point $t \in \Sigma$ to be the minimum degree of an isogeny $A_s \to A_t$ between the abelian varieties corresponding to the points s and t of $\mathcal{A}_g$ (note that this isogeny need not be polarised). We may also talk about the complexity of a point in $\widetilde{\Sigma}$, meaning the complexity of its image in Σ.

For a matrix $\gamma \in \mathrm{M}_{n\times n}(\mathbb{Q})$, the **height** $H(\gamma)$ will mean the maximum of the standard multiplicative heights of the entries of γ. A straightforward calculation shows that if $\gamma_1, \gamma_2 \in \mathrm{M}_{n\times n}(\mathbb{Q})$ then

$$H(\gamma_1\gamma_2) \leq nH(\gamma_1)H(\gamma_2).$$

The key step in the proof of Proposition 6.1 is Proposition 6.2: the points of $\widetilde{Z} \cap \widetilde{\Sigma}$ of a given complexity are contained in subpolynomially many definable blocks, these blocks themselves contained in $\widetilde{Z}$. This is proved using the Pila–Wilkie theorem and the matrix height bounds of section 5.2.

Proposition 6.2. *Let Z be a subvariety of $\mathcal{A}_g$ and $\tilde{s}$ a point in $\mathcal{H}_g$. Let $\epsilon > 0$.*

There is a constant $c = c(Z, \tilde{s}, \epsilon)$ such that for every $n \geq 1$, there is a collection of at most cn^ϵ definable blocks $W_i \subset \widetilde{Z}$ such that the union $\bigcup W_i$ contains all points of $\widetilde{Z} \cap \widetilde{\Sigma}$ of complexity n.

On the other hand, the Masser–Wüstholz isogeny theorem gives a polynomial lower bound for the Galois degree of points in Σ in terms of their complexity. Combining these two bounds, once the complexity gets large enough there are more points in $\widetilde{Z} \cap \widetilde{\Sigma}$ than there are blocks to contain them. Hence most points of $\widetilde{Z} \cap \widetilde{\Sigma}$ are contained in blocks of positive dimension. In particular the union of the positive-dimensional blocks contained in $\widetilde{Z}$ has Zariski dense image in Z, and by a lemma of Pila the same is true for the complex algebraic part of $\widetilde{Z}$.

Let us outline the proof of this Proposition 6.2. We cannot apply the counting theorem to $\widetilde{\Sigma} \subset \widetilde{Z}$ directly, because the points of $\widetilde{\Sigma}$ are transcendental. Instead we construct a definable subset Y of $\mathrm{GL}_{2g}(\mathbb{R})$ and a semialgebraic map $\sigma\colon Y \to \widetilde{Z}$ such that points of $\widetilde{Z} \cap \widetilde{\Sigma}$ have rational preimages in Y, with heights polynomially bounded in terms of their complexity. This idea is due to Habegger and Pila [HP12].

Consider first the case $\operatorname{End} A_s = \mathbb{Z}$. This case is easier because all isogenies between A_s and any abelian variety are polarised. In this case we let

$$Y = \{\gamma \in \mathrm{GSp}_{2g}(\mathbb{R})^+ \mid \gamma.\tilde{s} \in \widetilde{Z}\},$$

and let $\sigma\colon Y \to \widetilde{Z}$ be the map $\sigma(\gamma) = \gamma.\tilde{s}$.

Let $\tilde{t} \in \widetilde{Z} \cap \widetilde{\Sigma}$ and $t = \pi(\tilde{t})$. Then there is an isogeny $f\colon A_t \to A_s$ whose degree is equal to the complexity of t. By the hypothesis $\operatorname{End} A_s = \mathbb{Z}$ this isogeny is polarised. Hence the rational representation of f (explained in section 5.2) gives a matrix $\gamma \in \mathrm{GSp}_{2g}(\mathbb{Q})_+$ such that $\pi(\gamma.\tilde{s}) = t$ and whose height is polynomially bounded with respect to the complexity. We can also find $\gamma_1 \in \mathrm{Sp}_{2g}(\mathbb{Z})$ of polynomially bounded height such that $\gamma_1\gamma.\tilde{s} = \tilde{t}$. Hence every point in $\widetilde{Z} \cap \widetilde{\Sigma}$ has a rational preimage in Y of polynomially bounded height. This is precisely what we need to apply the Pila–Wilkie theorem to Y.

If we drop the assumption $\operatorname{End} A_s = \mathbb{Z}$ then this no longer works, because the rational representation of a non-polarised isogeny is not in $\mathrm{GSp}_{2g}(\mathbb{Q})_+$. Recall that even though t is in the Hecke orbit of s, so that there is some polarised isogeny $A_s \to A_t$, the isogeny of minimum degree need not be polarised. Thus we do not get an element of $\mathrm{GSp}_{2g}(\mathbb{Q})_+$ whose height is polynomially bounded in terms of the complexity.

To avoid this problem we will take Y to be a subset of $\mathrm{GL}_{2g}(\mathbb{R})$ instead of $\mathrm{GSp}_{2g}(\mathbb{R})$. This will allow us to carry out the same proof using the rational representation of a not-necessarily-polarised isogeny. Of course $\mathrm{GL}_{2g}(\mathbb{R})$ does not act on $\mathcal{H}_g$ but this does not matter: the map

$$\sigma\begin{pmatrix} A & B \\ C & D \end{pmatrix} = (A\tilde{s} + B)(C\tilde{s} + D)^{-1} \quad \text{for } A, B, C, D \in \mathrm{M}_{g \times g}(\mathbb{R})$$

is defined on a Zariski open subset of $\mathrm{GL}_{2g}(\mathbb{R})$, and we will only consider matrices in $\mathrm{GL}_{2g}(\mathbb{R})$ where σ is defined and has image in $\mathcal{H}_g$. In particular let

$$Y = \sigma^{-1}(\widetilde{Z}).$$

6.1.2 Proof of Proposition 6.2

Before proving Proposition 6.2, we need to check that every element of $\widetilde{Z} \cap \widetilde{\Sigma}$ has a rational preimage in Y whose height is polynomially bounded with respect to the complexity. Proposition 5.3 says that there is some preimage in $\mathrm{GL}_{2g}(\mathbb{R})$ with this property, and all we need to do is move it into the preimage of the fundamental set $\mathcal{F}_g$.

Lemma 6.3. *There exist constants c, k depending only on g and $\tilde{s}$ such that:*

For any $\tilde{t} \in \widetilde{Z} \cap \widetilde{\Sigma}$ of complexity n, there is a rational matrix $\gamma \in Y$ such that $\sigma(\gamma) = \tilde{t}$ and $H(\gamma) \leq cn^k$.

Proof. Let $t = \pi(\tilde{t})$. Let $\mathcal{B}$ be a symplectic basis for $H_1(A_s, \mathbb{Z})$ with period matrix $\tilde{s}$.

By Proposition 5.3 there is an isogeny $f\colon A_t \to A_s$ and a symplectic basis $\mathcal{B}'$ for $H_1(A_t, \mathbb{Z})$ such that the rational representation γ_1 of f has polynomially bounded height. As remarked in section 5.2.1, $\sigma({\gamma_1}^t)$ is the period matrix for (A_t, λ_t) with respect to the basis $\mathcal{B}'$. In particular,

$$\pi\sigma({\gamma_1}^t) = t.$$

Hence there is $\gamma_2 \in \mathrm{Sp}_{2g}(\mathbb{Z})$ such that $\gamma_2.\sigma({\gamma_1}^t) = \tilde{t}$. By [PT13] Lemma 3.2, $H(\gamma_2)$ is polynomially bounded. Then $\gamma = \gamma_2{\gamma_1}^t$ satisfies the required conditions. □

Now we are ready to prove Proposition 6.2. We simply apply the Theorem 3.2 to Y, using Lemma 6.3 to relate heights of rational points in Y to complexities of points in $\widetilde{Z} \cap \widetilde{\Sigma}$. We then use the fact that σ is semialgebraic, and that the blocks in Y can be chosen uniformly from finitely many definable families, to go from Y to $\widetilde{Z}$.

Proof of Proposition 6.2. The set

$$Y = \sigma^{-1}(\pi^{-1}(Z) \cap \mathcal{F}_g)$$

is definable because σ is semialgebraic and $\pi_{|\mathcal{F}_g}$ is definable.

Hence we can apply Theorem 3.2 to Y: for every $\epsilon > 0$, there are finitely many definable block families $\mathcal{W}^{(j)}(\epsilon) \subset Y \times \mathbb{R}^m$ and a constant $c_1(Y, \epsilon)$ such that for every $T \geq 1$, the rational points of Y of height at most T are contained in the union of at most $c_1 T^\epsilon$ definable blocks $W_i(T, \epsilon)$, taken from the families $\mathcal{W}^{(j)}(\epsilon)$.

Since σ is semialgebraic, the image under σ of a definable block in Y is a finite union of definable blocks in $\widetilde{Z}$. Furthermore the number of blocks in the image is uniformly bounded in each definable block family $\mathcal{W}^{(j)}(\epsilon)$. Hence $\sigma(\bigcup W_i(T, \epsilon))$ is the union of at most $c_2 T^\epsilon$ blocks in $\widetilde{Z}$, for some new constant $c_2(Z, \tilde{s}, \epsilon)$.

But by Lemma 6.3, for suitable constants c, k, every point of $\widetilde{Z} \cap \widetilde{\Sigma}$ of complexity n is in $\sigma(\bigcup W_i(cn^k, \epsilon))$. □

6.1.3 End of proof of Proposition 6.1

Proposition 6.2 tells us that the points of $\widetilde{Z} \cap \widetilde{\Sigma}$ of complexity n are contained in fewer than $c(\epsilon)n^\epsilon$ blocks for every $\epsilon > 0$. On the other hand, the Masser–Wüstholz isogeny theorem implies that the number of such points grows at least as fast as $n^{1/k}$ for some constant k. Hence most points of $\widetilde{Z} \cap \widetilde{\Sigma}$ are contained in a block of positive dimension. We check that this is sufficient, with the hypothesis that $Z \cap \Sigma$ is Zariski dense in Z, to deduce that the union of positive-dimensional blocks in $\widetilde{Z}$ has Zariski dense image in Z.

Proposition 6.4. *If $Z \cap \Sigma$ is Zariski dense in Z, then the union of positive-dimensional blocks in $\widetilde{Z}$ has Zariski dense image in Z.*

Proof. Let Σ_1 be the set of points $t \in Z \cap \Sigma$ such that there is a positive-dimensional block $W \subset \widetilde{Z}$ such that $t \in \pi(W)$. (We do not yet know that Σ_1 is non-empty.) Let Z_1 denote the Zariski closure of Σ_1.

Let (A_s, λ_s) be a polarised abelian variety corresponding to the point $s \in \mathcal{A}_g(\mathbb{C})$, defined over a finitely generated field K. We choose K large enough that the varieties Z and Z_1 are also defined over K.

Let t be a point in $Z \cap \Sigma$ of complexity n. The polarised abelian variety corresponding to t might not have a model over the field of moduli $K(t)$, but it has a model (A_t, λ_t) over an extension L of $K(t)$ of uniformly bounded degree. This follows from the fact that a polarised abelian variety with full level-3 structure has no non-trivial automorphisms, so is defined over its field of moduli; and the field of moduli of a full level-3 structure on the polarised abelian variety corresponding to t is an extension of $K(t)$ of degree at most $\# \mathrm{Sp}_{2g}(\mathbb{Z}/3)$.

By Theorem 5.2, the complexity n is bounded above by a polynomial $c[L : K]^k$ in $[L : K]$, with c and k depending only on A_s and K. Hence for a different constant c_1, we have

$$[K(t) : K] \geq c_1 n^{1/k}.$$

But all $\mathrm{Gal}(\bar{K}/K)$-conjugates of t are contained in $Z \cap \Sigma$ and have complexity n. By Proposition 6.2, the preimages in $\mathcal{F}_g$ of these points are contained in the union of $c_2(Z, \tilde{s}, 1/2k)n^{1/2k}$ definable blocks, each of these blocks being contained in $\widetilde{Z}$.

For large enough n, we have

$$c_1 n^{1/k} > c_2 n^{1/2k}.$$

For such n, by the pigeonhole principle there is a definable block $W \subset \widetilde{Z}$ such that $\pi(W)$ contains at least two Galois conjugates of t. Since blocks are connected by definition, $\dim W > 0$. So those conjugates of t in $\pi(W)$ are in Σ_1. Since Z_1 is defined over K, it follows that t itself is also in Z_1.

In other words all points of $Z \cap \Sigma$ of large enough complexity are in Σ_1. But this excludes only finitely many points of $Z \cap \Sigma$. So as $Z \cap \Sigma$ is Zariski dense in Z, we conclude that $Z_1 = Z$. □

Proposition 6.4 says that $\pi(\widetilde{Z}^{\mathrm{alg}})$ is Zariski dense in Z. By Lemma 3.7, $\widetilde{Z}^{\mathrm{ca}} = \widetilde{Z}^{\mathrm{alg}}$ so $\pi(\widetilde{Z}^{\mathrm{ca}})$ is dense in Z and Proposition 6.1 is proved.

6.2 The André–Pink conjecture for curves

We are now ready to prove the André–Pink conjecture for curves in Shimura varieties of abelian type. First for curves in $\mathcal{A}_g$, where it is an immediate corollary of Proposition 6.1 and the characterisation of weakly special subvarieties in [UY11]. Then we use Theorem 4.6 to deduce the result for other Shimura varieties of abelian type.

Proposition 6.5. *Let Z be a closed algebraic curve in $\mathcal{A}_g$. Let s be a point in $\mathcal{A}_g$ and Σ the Hecke orbit of s.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of $\mathcal{A}_g$.

Proof. According to Proposition 6.1, $\pi(\widetilde{Z}^{\mathrm{ca}})$ is Zariski dense in Z. In particular $\widetilde{Z}^{\mathrm{ca}}$ has positive dimension. Since $\dim \widetilde{Z} = 1$, it follows that some connected component $\widetilde{Z}^\circ$ of $\widetilde{Z}$ is an irreducible algebraic subset of $\mathcal{F}_g$ (as defined in section 3.4).

By analytic continuation, the connected component of $\pi^{-1}(Z)$ containing $\widetilde{Z}^\circ$ is an irreducible algebraic subset of $\mathcal{H}_g$. In other words, Z itself is algebraic and a connected component of $\pi^{-1}(Z)$ is also algebraic. Hence by [UY11] Theorem 1.2, Z is weakly special. □

Theorem 6.6. *Let S be a connected Shimura variety of abelian type and $Z \subset S$ a closed algebraic curve. Let s be a point in S and let Σ be the generalised Hecke orbit of s.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of S.

Proof. Suppose that $S = \mathrm{Sh}_\Gamma(G, X^+)$. We first replace G by its adjoint group: let $S^{\mathrm{ad}} = \mathrm{Sh}_{\Gamma^{\mathrm{ad}}}(G^{\mathrm{ad}}, X^{\mathrm{ad}+})$ for some congruence subgroup $\Gamma^{\mathrm{ad}} \subset G^{\mathrm{ad}}(\mathbb{Q})^+$ containing the image of Γ. Then the image of a generalised Hecke orbit in S is contained in a finite union of usual Hecke orbits in S^{ad} and a subvariety of S is weakly special if and only if its image in S^{ad} is weakly special. Hence it will suffice to assume that G is an adjoint group and that Σ is a usual Hecke orbit.

Then we can apply Theorem 4.6 to get a Shimura datum (G_1, X_1^+) and morphisms

$$(G, X^+) \xleftarrow{p} (G_1, X_1^+) \xrightarrow{\iota} (\mathrm{GSp}_{2g}, \mathcal{H}_g)$$

such that $[\iota][p]^{-1}(\Sigma)$ is contained in a Hecke orbit in $\mathcal{A}_g$.

Meanwhile $[\iota][p]^{-1}(Z)$ is a finite union of curves in $\mathcal{A}_g$. Thus there is an irreducible component $Z' \subset [\iota][p]^{-1}(Z)$ which has infinite intersection with some Hecke orbit in $\mathcal{A}_g$. By Proposition 6.5, Z' is weakly special in $\mathcal{A}_g$. It follows that Z is weakly special in S. □

6.3 Higher-dimensional subvarieties

Let S be a connected Shimura variety of abelian type and Z a subvariety of S, now of any dimension. Let Z^{ws} denote the union of the positive-dimensional weakly special subvarieties of S contained in Z.

In order to draw conclusions from Proposition 6.1 when $\dim Z > 1$, we must use the hyperbolic Ax–Lindemann–Weierstrass conjecture (Conjecture 3.8) for $\mathcal{A}_g$, whose proof has been announced by Pila and Tsimerman [PT12]. This tells us that if $S = \mathcal{A}_g$ then $Z^{\mathrm{ws}} = \pi(\widetilde{Z}^{\mathrm{ca}})$ (Corollary 3.10).

Combining this with Proposition 6.1, we deduce that if $S = \mathcal{A}_g$ there is a Hecke orbit Σ in S which intersects Z densely, then Z^{ws} is Zariski dense in Z. By the same argument as in the proof of Theorem 6.6 we can extend this from $\mathcal{A}_g$ to any Shimura variety of abelian type. In other words we have proved the following.

Proposition 6.7. *Let S be a connected Shimura variety of abelian type and $Z \subset S$ a closed algebraic curve. Let s be a point in S and let Σ be the generalised Hecke orbit of s.*

If $\Sigma \cap Z$ is Zariski dense in Z, then Z^{ws} is also Zariski dense in Z.

Now add the further assumption that Z is Hodge generic in S – that is, Z is not contained in any proper special subvariety of S. The following result of Ullmo tells us what Z looks like, given that Z^{ws} is Zariski dense in Z. We have translated Ullmo's statement to match our definition of weakly special subvarieties in a similar way to the translation between definitions of weakly special subvarieties in Proposition 2.7.

Theorem 6.8 ([Ull12] Theorem 1.3)**.** *Let S be a Shimura variety and $Z \subset S$ a Hodge generic subvariety. Suppose that the hyperbolic Ax–Lindemann–Weierstrass conjecture holds for S.*

If Z^{ws} is Zariski dense in Z, then there exist a Shimura variety S_2, a Shimura submersion $[\varphi]\colon S \to S_2$ and a subvariety $Z' \subset S_2$ such that $Z = [\varphi]^{-1}(Z')$ and $\dim S_2 < \dim S$.

If there is a generalised Hecke orbit $\Sigma \subset S$ such that $Z \cap \Sigma$ is Zariski dense in Z, then $[\varphi](\Sigma)$ is a finite union of generalised Hecke orbits in S_2 (Lemma 4.1(4)) and $[\varphi](\Sigma) \cap Z'$ is Zariski dense in Z'. Thus we can prove the following by induction on $\dim S$.

Theorem 6.9. *Let S be a connected Shimura variety of abelian type and $Z \subset S$ an algebraic subvariety. Let Σ be a generalised Hecke orbit in S.*

If Z is Hodge generic in S and $\Sigma \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of S.

Now return to a subvariety Z which is not necessarily Hodge generic. Then we might try to prove the André–Pink conjecture by replacing S by S_Z, a connected Shimura variety with a Shimura immersion $[\iota]\colon S_Z \to S$ whose image is equal to the smallest special subvariety of S containing Z. The problem with this is that the preimage in S_Z of Σ, a generalised Hecke orbit in S, might contain infinitely many generalised Hecke orbits in S_Z.

By Lemma 4.2, this problem does not arise if one (and hence any) point in $[\iota]^{-1}(\Sigma)$ is Hodge generic in S_Z. Combining Lemma 4.2 and Theorem 6.9 yields the following corollary.

Corollary 6.10. *Let S be a connected Shimura variety of abelian type and let Z be a closed algebraic subvariety of S. Let s be a point in S and Σ the generalised Hecke orbit of s.*

If $\Sigma \cap Z$ is Zariski dense in Z and s is Hodge generic in the smallest special subvariety of S containing Z, then Z is a weakly special subvariety of S.

In the same way we may obtain the following corollary by combining Theorems 4.9 and 6.9.

Corollary 6.11. *Let S be a connected Shimura variety of abelian type and let Z be a closed algebraic subvariety of S. Let s be a point in S and Σ_P the P-Hecke orbit of s.*

If $\Sigma_P \cap Z$ is Zariski dense in Z, then Z is a weakly special subvariety of S.

7 Ranks of Mumford–Tate groups

This chapter reproduces an article submitted for publication.

Let A be a complex abelian variety and G its Mumford–Tate group. Supposing that the simple abelian subvarieties of A are pairwise non-isogenous, we find a lower bound for the rank $\operatorname{rk} G$ of G, which is a little less than $\log_2 \dim A$. If we suppose that $\operatorname{End} A$ is commutative, then we show that $\operatorname{rk} G \geq \log_2 \dim A + 2$, and this latter bound is sharp. We also obtain the same results for the rank of the ℓ-adic monodromy group of an abelian variety defined over a number field.

7.1 Introduction

Let A be a complex abelian variety of dimension g, whose simple abelian subvarieties are pairwise non-isogenous. In this paper we will establish a lower bound for the rank of the Mumford–Tate group of A. The Mumford–Tate group is an algebraic group over $\mathbb{Q}$ defined via the Hodge theory of A (see section 7.2 below for the definition). The same argument will also establish a lower bound for the rank of the ℓ-adic monodromy groups G_ℓ, in the case where A is defined over a number field. The ℓ-adic monodromy group is the Zariski closure of the image of the Galois representation on the ℓ-adic Tate module of A. Our main theorems are the following:

Theorem 7.1. *Let A be an abelian variety of dimension g such that* $\operatorname{End} A$ *is commutative. Let G be the Mumford–Tate group or the ℓ-adic monodromy group of A. Then* $\operatorname{rk} G \geq \log_2 g + 2$.

Theorem 7.2. *Let A be an abelian variety of dimension g whose simple abelian subvarieties are pairwise non-isogenous. Let G be the Mumford–Tate group or the ℓ-adic monodromy group of A. If $n = \operatorname{rk} G$, then*

$$n + \alpha(n)\sqrt{n \log_e n} \geq \log_2 g + 2$$

for a function $\alpha : \mathbb{N}_{\geq 2} \to \mathbb{R}$ satisfying $\alpha(n) < 2$ for all n and $\alpha(n) \to 1/\log_e 2 = 1.44...$ as $n \to \infty$.

Each of these theorems is an instance of a more general bound for weak Mumford–Tate triples, which are defined in section 7.2. These more general bounds are Theorems 7.6 and 7.9 respectively. These would apply also for example to the analogue of the Mumford–Tate group for a Hodge–Tate module of weights 0 and 1.

Theorem 7.1 was proved by Ribet in the case of an abelian variety with complex multiplication [Rib81]. Our proof is a generalisation of his, relying on the fact that the defining representation of the Mumford–Tate group or ℓ-adic monodromy group has minuscule weights.

The condition on simple subvarieties in Theorem 7.2 is necessary: taking products of copies of the same simple abelian variety increases the dimension without changing the rank of the Mumford–Tate group. Indeed, if A is isogenous to $\prod_i A_i^{m_i}$ where the A_i are simple and pairwise non-isogenous, then according to [HR10] Lemme 2.2,

$$\mathrm{MT}(A) \cong \mathrm{MT}(\prod_i A_i).$$

Hence Theorem 7.2 implies that for a general abelian variety A, if n denotes the rank of either the Mumford–Tate group or the ℓ-adic monodromy group of A, then

$$n + \alpha(n)\sqrt{n \log_e n} \geq \log_2 \left(\sum_i \dim A_i \right) + 2$$

where the A_i are one representative of each isogeny class of simple abelian subvarieties of A.

The condition of having pairwise non-isogenous simple abelian subvarieties can be interpreted via the endomorphism algebra like the condition in Theorem 7.1: it is equivalent to $\operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{Q}$ being a product of division algebras. Note also that $\operatorname{End} A$ being commutative implies the condition of Theorem 7.2. (Throughout this paper, $\operatorname{End} A$ means the endomorphisms of A after extension of scalars to an algebraically closed field.)

Let G be either the Mumford–Tate group or the ℓ-adic monodromy group of A. It is well known that the rank of G is at most $g+1$, and that this upper bound is achieved for a generic abelian variety. Indeed, if g is odd and $\operatorname{End} A = \mathbb{Z}$, then $\operatorname{rk} G$ is always $g + 1$ [Ser85]. So in this case the bound in Theorem 7.1 is far from sharp.

On the other hand if g is a power of 2, then there are abelian varieties for which the bound in Theorem 7.1 is achieved (even with $\operatorname{End} A = \mathbb{Z}$). We construct such examples in section 7.5. The exact bound for a given g is very sensitive to the prime factors of g. Equality can happen only when g is a power of 2 (for the trivial reason that otherwise $\log_2 g \notin \mathbb{Z}$) but even near-equality can only occur when g has many small prime factors. This was made precise by Dodson in the complex multiplication case [Dod87], and it is possible that something similar could be proved in general.

Theorem 7.2 is not sharp. The function $\alpha(n)$ is specified exactly in section 7.4, but it is likely that this could be improved on, perhaps to something which goes to 0 as $n \to \infty$. In section 7.5, we construct a family of examples showing that Theorem 7.2 cannot be improved to $n + k \geq \log_2 g$ for any constant k.

We can deduce a lower bound for the growth of the degrees of the division fields $K(A[\ell^n])$ (for ℓ a fixed prime number) as a straightforward consequence of Theorem 7.1.

Corollary 7.3. *Let A be an abelian variety of dimension g over a number field K, and ℓ a prime number. If* $\operatorname{End} A$ *is commutative, then there is a constant $C(A, K, \ell)$ such that*

$$[K(A[\ell^n]) : K] \geq C(A, K, \ell)\, \ell^{n(\log_2 g+2)}.$$

Theorem 7.2 implies a similar bound for the degree of $K(A[\ell^n])$ whenever A is an abelian variety whose simple abelian subvarieties are pairwise non-isogenous. One would like to extend these results to lower bounds on the degrees of $K(A[N])$ for N not a prime power, but this cannot be done without knowing how $C(A, K, \ell)$ varies with ℓ. The primary obstacle here is the index of the image of $\operatorname{Gal}(\bar{K}/K)$ in $G_\ell(\mathbb{Z}_\ell)$, which is conjectured to be bounded by a constant $C_1(A, K)$ independent of ℓ.

In section 7.2 we recall the definitions of Mumford–Tate groups, ℓ-adic monodromy groups and weak Mumford–Tate triples; the latter are an axiomatisation of the properties of the groups and representations we will consider. In section 7.3 we bound the number of distinct characters of a maximal torus which can appear in such a representation. In section 7.4 we bound the multiplicity of absolutely irreducible components of this representation. This is straightforward for the Mumford–Tate group but more difficult for the ℓ-adic monodromy group. Combining these two bounds gives Theorems 7.1 and Theorem 7.2. Finally in section 7.5 we give some examples to show that Theorem 7.1 is sharp and to place a limit on the possible improvements of Theorem 7.2.

7.2 Mumford–Tate triples: Definitions

We recall the definition of a weak Mumford–Tate triple, which abstracts the key properties of a Mumford–Tate group which we will use. We recall also the definitions of the two examples of Mumford–Tate triple we will consider, namely the Mumford–Tate group and the ℓ-adic monodromy group of an abelian variety.

The following definition is a slight modification of those used by Serre [Ser79] and Wintenberger [Win86].

Definition. Let F be a field of characteristic zero and E an algebraically closed field containing F.

A **weak Mumford–Tate triple** is a triple (G, ρ, Ψ) where G is an algebraic group over F, ρ is a rational representation of G and Ψ is a set of cocharacters of $G \times_F E$ satisfying the following conditions:

(i) G is a connected reductive group;

(ii) ρ is faithful;

(iii) the images of all $G(E)$-conjugates of elements of Ψ generate G_E.

The **weights** of a Mumford–Tate triple (G, ρ, Ψ) are the integers which appear as weights of $\rho \circ \nu$ (a representation of $\mathbb{G}_m$) for some $\nu \in \Psi$.

A weak Mumford–Tate triple (G, ρ, Ψ) is called **pure** if $\rho(G)$ contains the torus $\mathbb{G}_m.\mathrm{id}$ of homotheties.

The Mumford–Tate group. Let A be an abelian variety over $\mathbb{C}$, of dimension g. The singular cohomology group $H^1(A(\mathbb{C}), \mathbb{Q})$ is a vector space of dimension $2g$ over $\mathbb{Q}$. Hodge theory gives a decomposition of $\mathbb{C}$-vector spaces

$$H^1(A(\mathbb{C}), \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} = H^{1,0}(A) \oplus H^{0,1}(A)$$

with $H^{1,0}(A)$ and $H^{0,1}(A)$ being mapped onto each other by complex conjugation (so each has dimension g).

We define a cocharacter $\mu : \mathbb{G}_{m,\mathbb{C}} \to \mathrm{GL}_{2g,\mathbb{C}}$ by:

$$\begin{aligned} \mu(z) \text{ acts as multiplication by } z \text{ on } H^{1,0}(A) \\ \text{and as the identity on } H^{0,1}(A). \end{aligned}$$

The **Mumford–Tate group** of A is defined to be the smallest algebraic subgroup M of GL_{2g} defined over $\mathbb{Q}$ and such that $M_{\mathbb{C}}$ contains the image of μ.

The triple consisting of the Mumford–Tate group, its defining representation $\rho : M \to \mathrm{GL}_{2g}$, and the set of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$-conjugates of the cocharacter μ form a pure weak Mumford–Tate triple of weights $\{0, 1\}$. This is immediate from the definitions.

The functor $A \mapsto H^1(A(\mathbb{C}), \mathbb{Z})$ is an equivalence of categories between complex abelian varieties and polarisable $\mathbb{Z}$-Hodge structures of type $\{(-1, 0), (0, -1)\}$. Furthermore the endomorphism ring of ρ as a representation of the Mumford–Tate group is equal to the endomorphism ring of $H^1(A(\mathbb{C}), \mathbb{Q})$ as a $\mathbb{Q}$-Hodge structure, so

$$\mathrm{End}\, \rho = \mathrm{End}\, A \otimes_{\mathbb{Z}} \mathbb{Q}.$$

The ℓ-adic algebraic monodromy group. Now suppose that the abelian variety A is defined over a number field K. Its first ℓ-adic cohomology group is a $\mathbb{Q}_\ell$-vector space of dimension $2g$, isomorphic to the dual of the ℓ-adic Tate module:

$$H^1(A_{\bar{K}}, \mathbb{Q}_\ell) \cong (T_\ell A \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell)^\vee .$$

The Galois group $\mathrm{Gal}(\bar{K}/K)$ acts on the torsion points of $A(\bar{K})$, and this induces an action on $H^1(A_{\bar{K}}, \mathbb{Q}_\ell)$, or in other words a continuous representation

$$\rho_\ell : \mathrm{Gal}(\bar{K}/K) \to \mathrm{GL}_{2g}(\mathbb{Q}_\ell).$$

The **ℓ-adic algebraic monodromy group** of A is the smallest algebraic subgroup G_ℓ of $\mathrm{GL}_{2g,\mathbb{Q}_\ell}$ whose $\mathbb{Q}_\ell$-points contain the image of ρ_ℓ. By working with the ℓ-adic monodromy group instead of the image of ρ_ℓ directly, we gain the ability to use the structure theory of algebraic groups. On the other hand, we do not lose very much because $\mathrm{Im}\,\rho_\ell$ is known [Bog81] to be an open (and hence finite-index) subgroup of $G_\ell(\mathbb{Q}_\ell) \cap \mathrm{GL}_{2g}(\mathbb{Z}_\ell)$.

Pink [Pin98] has proved that the identity component G_ℓ° together with the representation ρ_ℓ and a certain set Ψ of cocharacters form a pure weak Mumford–Tate triple of weights $\{0, 1\}$.

By Faltings' Theorem [Fal83b],

$$\mathrm{End}\,\rho_\ell = \mathrm{End}\,A \otimes_{\mathbb{Z}} \mathbb{Q}_\ell.$$

7.3 Bound for the number of characters

Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights $\{0, 1\}$, and let T be a maximal torus of G. In this section we will give an upper bound for the number of distinct characters in $\rho_{|T}$ as a function of $\mathrm{rk}\,G$.

If A has complex multiplication (in other words if G is a torus) then this bound was obtained by Ribet [Rib81]. Our method of proving the bound is inspired by applying Ribet's method to a maximal torus of G, but it is convenient to arrange it differently.

Proposition 7.4. *Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights $\{0, 1\}$. The number of distinct characters in $\rho_{|T}$ is at most $2^{\mathrm{rk}\,G - 1}$.*

Proof. Let $Y = \mathrm{Hom}(\mathbb{G}_{m,E}, T_E) \otimes_{\mathbb{Z}} \mathbb{Q}$ be the quasi-cocharacter space of T, where E is an algebraically closed field of definition for (G, ρ, Ψ).

Let Ψ' be the set of all cocharacters of T_E which are $G(E)$-conjugate to an element of Ψ. Every cocharacter of G has a $G(E)$-conjugate whose image is contained in T_E, so Ψ' still satisfies condition (iii) in the definition of a weak Mumford–Tate triple. Replacing Ψ by Ψ' does not change the weights of our Mumford–Tate triple.

Furthermore Ψ' is closed under the action of the Weyl group of G_E on Y. So condition (iii) implies that Ψ' spans Y as a $\mathbb{Q}$-vector space.

Let Θ be a basis of Y contained in Ψ'. The character space of T is dual to Y, so any character ω is determined by its inner products $\langle \omega, \mu \rangle$ for $\mu \in \Theta$.

Because our Mumford–Tate triple has weights $\{0, 1\}$, if μ is a character in $\rho_{|T}$ then these inner products can only have the values 0 or 1. So there are at most $2^{|\Theta|}$ distinct characters in $\rho_{|T}$, and $|\Theta| = \operatorname{rk} G$.

We can use the fact that our Mumford–Tate triple is pure to improve the exponent to $\operatorname{rk} G - 1$. We know that $\rho(G)$ contains the homotheties. Since ρ is faithful, there is a unique cocharacter $\mu_0 : \mathbb{G}_m \to G$ such that $\rho \circ \mu_0(z) = z.\mathrm{id}$. We take Θ' to be a subset of Ψ such that $\Theta' \cup \{\mu_0\}$ is a basis of Y. Now $\langle \omega, \mu_0 \rangle = 1$ for all characters ω in $\rho_{|T}$, so ω is determined by the values $\langle \omega, \mu \rangle$ for $\mu \in \Theta'$. We may repeat the previous argument with Θ replaced by Θ'. □

Corollary 7.5. *Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights $\{0, 1\}$. Let M be the maximum of the multiplicities of the irreducible components of ρ (working over an algebraically closed base field). Then*

$$\dim \rho \leq M \cdot 2^{\operatorname{rk} G - 1}.$$

Proof. Serre [Ser79] showed that each irreducible component σ in a weak Mumford–Tate triple of weights $\{0, 1\}$ is **minuscule**, that is, the characters in $\sigma_{|T}$ form a single orbit under the action of the Weyl group. Serre only treated strong Mumford–Tate triples, i.e. weak Mumford–Tate triples satisfying the additional condition that all the cocharacters in Ψ are contained in a single $\mathrm{Aut}(E/F)$-orbit. However this extra condition is not used in his argument (see also [Pin98] Section 4 and [Zar84]).

The characters of T in a minuscule representation have multiplicity 1, and non-isomorphic minuscule representations contain disjoint characters. So the multiplicity of any character in $\rho_{|T}$ is equal to the multiplicity of the unique irreducible component which contains that character, and so

$$\dim \rho \leq M \cdot \Big(\text{the number of distinct characters in } \rho_{|T}\Big).$$

The corollary now follows from Proposition 7.4. □

7.4 Bound for the multiplicities

Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights $\{0, 1\}$. In this section we will bound the multiplicities of the absolutely irreducible components of $\rho \otimes_F \bar{F}$. If $\operatorname{End} \rho$ is commutative, then it is immediate that all absolutely irreducible components of $\rho \otimes_F \bar{F}$ have multiplicity 1.

Most of the section concerns the case in which the irreducible components of ρ are pairwise non-isomorphic. Because we use a result on division algebras coming from class field theory, we must assume that the field of definition of ρ is a local field or a number field. If $n = \operatorname{rk} G$, then each absolutely irreducible component has multiplicity at most $\alpha(n)\sqrt{n \log_e n}$ for a function $\alpha(n)$ satisfying the conditions of Theorem 7.2.

To establish this bound, we introduce an invariant $u(G)$ for a reductive group G such that for any F-irreducible representation of G, the multiplicity of its irreducible components over $\bar{F}$ is at most $u(G)$. Then we use Landau's function (the maximum LCM of a set of positive integers with given sum) to obtain a bound for $u(G)$.

The above bounds together with Corollary 7.5 suffice to prove Theorem 7.1 for both the Mumford–Tate group and ℓ-adic monodromy groups, and Theorem 7.2 for the Mumford–Tate group. Proving Theorem 7.2 for the ℓ-adic monodromy group requires additional work because even when an abelian variety satisfies the condition of Theorem 7.2, its associated ℓ-adic representations might not satisfy the corresponding condition of Theorem 7.9.

7.4.1 The commutative endomorphism case

Theorem 7.6. *Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights $\{0, 1\}$. If $\operatorname{End} \rho$ is commutative, then $\operatorname{rk} G \geq \log_2 \dim \rho + 1$.*

Proof. Let F be the field of definition of ρ. Since $\operatorname{End} \rho$ is commutative, each irreducible component of $\rho \otimes_F \bar{F}$ has multiplicity 1. So the theorem follows immediately from Corollary 7.5. ☐

Let A be an abelian variety, G its Mumford–Tate group or ℓ-adic monodromy group, and ρ the associated representation. We have observed that $\operatorname{End} \rho = \operatorname{End} A \otimes_{\mathbb{Z}} F$ where $F = \mathbb{Q}$ or $\mathbb{Q}_\ell$ as appropriate, so that if $\operatorname{End} A$ is commutative the same is true of $\operatorname{End} \rho$. Hence Theorem 7.1 follows from Theorem 7.6. The $\log_2 \dim \rho + 1$ becomes $\log_2 \dim A + 2$ because $\dim \rho = 2 \dim A$.

7.4.2 Multiplicity of irreducible representations and $u(G)$

Definition. Let G be a reductive group defined over the field F. Let T be a maximal torus of G and $\Lambda = \operatorname{Hom}(T_{\bar{F}}, \mathbb{G}_m)$ the character group of T. Let Λ_0 be the subgroup of Λ generated by the roots of G and characters which vanish on $T \cap G^{\mathrm{der}}$. The roots of G span the quasi-character space of $T_{\bar{F}} \cap G^{\mathrm{der}}_{\bar{F}}$ as a $\mathbb{Q}$-vector space so Λ_0 spans $\Lambda \otimes_{\mathbb{Z}} \mathbb{Q}$. It follows that Λ/Λ_0 is finite. (In fact Λ/Λ_0 is canonically isomorphic to the dual of the centre of $G^{\mathrm{der}}(\bar{F})$, which is a finite abelian group.)

Hence we can define $u(G)$ to be the exponent of Λ/Λ_0.

Lemma 7.7. *Let G be a reductive group over a field F and ρ an F-irreducible representation of G. Let D be the endomorphism ring of ρ and E the centre of D. Then the order of $[D]$ in $\operatorname{Br} E$ divides $u(G)$.*

Proof. Fix a base Δ for the root system of G with respect to T. When we refer to the action of $\mathrm{Gal}(\bar{F}/F)$ on the character group Λ below, this is the natural action twisted by the Weyl group so that it preserves the set Δ (this is the same action used in [Tit71]).

Let σ be an absolutely irreducible component of $\rho \otimes_F \bar{F}$, and $\lambda_\sigma \in \Lambda$ the highest weight of σ. Let Γ be the subgroup of $\mathrm{Gal}(\bar{F}/F)$ fixing λ. Then E is isomorphic to the subfield of $\bar{F}$ fixed by Γ.

Tits defined a map

$$\alpha_{G,E} : \Lambda^\Gamma \to \mathrm{Br}\, E$$

as follows: if $\lambda \in \Lambda^\Gamma$ is dominant then there is a unique isomorphism class of E-irreducible representations of G with highest weight λ. The endomorphism ring of such a representation is a division algebra with centre E. We define $\alpha_{G,E}(\lambda)$ to be the inverse of the class of this division algebra in $\mathrm{Br}\, E$. Tits showed that this map on dominant weights is additive so it extends to a homomorphism $\Lambda^\Gamma \to \mathrm{Br}\, E$. He also showed that $\alpha_{G,E}$ is trivial on Λ_0^Γ ([Tit71] Corollary 3.5).

In our case we have $[D]^{-1} = \alpha_{G,E}(\lambda_\sigma)$. Since $[D]$ is in the image of $\alpha_{G,E}$, it follows that the order of $[D]$ in $\mathrm{Br}\, E$ divides the exponent of $\Lambda^\Gamma/\Lambda_0^\Gamma$. But the latter is a subgroup of Λ/Λ_0, so its exponent divides $u(G)$. □

Corollary 7.8. *Let G be a reductive group defined over a number field or a local field F. Let ρ be an F-irreducible representation of G. Then the multiplicity of each absolutely irreducible component of $\rho \otimes_F \bar{F}$ divides $u(G)$.*

Proof. Let $D = \mathrm{End}\, \rho$ and let E be the centre of D. Then the multiplicity of any absolutely irreducible component of $\rho \otimes_F \bar{F}$ is $\sqrt{\dim_E D}$.

Since F is a number field or a local field, it follows from class field theory that $\sqrt{\dim_E D}$ is equal to the order of $[D]$ in $\mathrm{Br}\, E$ (see e.g. [Pie82] Theorem 18.6).

Now apply Lemma 7.7. □

The following theorem is obtained by combining Corollaries 7.5 and 7.8.

Theorem 7.9. *Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights $\{0, 1\}$ defined over a number field or a local field F. If the F-irreducible components of ρ are pairwise non-isomorphic, then*

$$\mathrm{rk}\, G + \log_2 u(G) \geq \log_2 \dim \rho + 1.$$

If A is a complex abelian variety, G its Mumford–Tate group and ρ the associated representation, then $\mathrm{End}\, \rho = \mathrm{End}\, A \otimes_{\mathbb{Z}} \mathbb{Q}$ so the hypothesis that the simple abelian subvarieties of A are pairwise non-isogenous implies that the irreducible components of ρ are pairwise non-isomorphic. Hence Theorem 7.9, together with the bounds for $u(G)$ in section 7.4.4, implies Theorem 7.2 for the Mumford–Tate group.

7.4.3 Multiplicities in ℓ-adic representations

Let A be an abelian variety over a number field whose simple abelian subvarieties are pairwise non-isogenous. Let G_ℓ be the ℓ-adic monodromy group of A and ρ_ℓ the associated ℓ-adic representation. We shall show that the multiplicities of irreducible components of $\rho_\ell \otimes_{\mathbb{Q}_\ell} \bar{\mathbb{Q}}_\ell$ are bounded above by $u(G_\ell^\circ)$ and hence prove Theorem 7.2.

By Faltings' Theorem, if B and B' are non-isogenous simple abelian varieties, then the associated ℓ-adic representations have no common subrepresentations. Hence it will suffice to suppose that A is simple.

By Faltings' Theorem, $\operatorname{End} \rho_\ell = \operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{Q}_\ell$. This implies that the multiplicities of absolutely irreducible components of $\rho_\ell \otimes_{\mathbb{Q}_\ell} \bar{\mathbb{Q}}_\ell$ are independent of ℓ. We will use results of Serre and Pink to show that $u(G_\ell^\circ)$ is also independent of ℓ, and then we can consider all ℓ at once to show that the multiplicities are bounded above by $u(G_\ell^\circ)$.

Lemma 7.10. *$u(G_\ell^\circ)$ is independent of ℓ.*

Proof. Let ℓ, ℓ' be any two rational primes. Via ρ_ℓ, we view G_ℓ° as a subgroup of $\mathrm{GL}_{2g,\mathbb{Q}_\ell}$.

For a finite place v of K, let T_v be the Frobenius torus of A in the sense of Serre [Ser81]. Serre showed that we can choose v such that $T_{v,\mathbb{Q}_\ell}$ is $\mathrm{GL}_{2g,\mathbb{Q}_\ell}$-conjugate to a maximal torus of G_ℓ°, and such that the analogous property holds for ℓ'.

Hence we get maximal tori $T_{v,\ell}$ of G_ℓ and $T_{v,\ell'}$ of G'_ℓ together with an isomorphism $\Lambda(T_{v,\ell}) \cong \Lambda(T_v) \cong \Lambda(T_{v,\ell'})$. Furthermore, under this isomorphism, the formal character of ρ_ℓ corresponds to the formal character of $\rho_{\ell'}$.

As observed by Larsen-Pink [LP90], the formal character of a faithful irreducible representation of a reductive group determines the root lattice Λ_0. Hence $\Lambda/\Lambda_0(G_\ell^\circ) \cong \Lambda/\Lambda_0(G_{\ell'}^\circ)$ so $u(G_\ell^\circ) = u(G_{\ell'}^\circ)$. $\square$

We will also need the following lemma on pure weak Mumford–Tate triples. Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple. Because it is pure, there is a cocharacter μ_0 of G such that $\rho \circ \mu_0(z) = z.\mathrm{id}$. Let H be the identity component of $\ker \det \rho \subset G$. Then the quasi-cocharacter space of a maximal torus T splits as

$$(\operatorname{Hom}(\mathbb{G}_m, T \cap H) \otimes_{\mathbb{Z}} \mathbb{Q}) \oplus \mathbb{Q}.\mu_0. \tag{*}$$

Lemma 7.11. *Let (G, ρ, Ψ) be a pure weak Mumford–Tate triple of weights 0 and 1, with multiplicities g_0 and g_1 respectively. Choose $\mu \in \Psi$ and let T be a maximal torus of G containing the image of μ. Suppose that μ splits as $\mu_H + r\mu_0$ in the decomposition (*). Then for all characters ω in $\rho_{|T}$,*

$$\langle \omega, \mu_H \rangle = \frac{g_0}{g_0 + g_1} \text{ or } \frac{-g_1}{g_0 + g_1}.$$

Proof. By the definition of μ_0, $\langle \omega, \mu_0 \rangle = 1$ for every character ω in $\rho_{|T}$. Hence

$$\langle \det \rho, \mu_0 \rangle = \dim \rho = g_0 + g_1.$$

Because $\det \rho$ is trivial on H, $\langle \det \rho, \mu_H \rangle = 0$. Therefore

$$\langle \det \rho, \mu \rangle = \langle \det \rho, r\mu_0 \rangle = r(g_0 + g_1).$$

On the other hand,

$$\langle \det \rho, \mu \rangle = g_0.0 + g_1.1 = g_1$$

so $r = g_1/(g_0 + g_1)$. Combining with $\langle \omega, \mu \rangle = 0$ or 1 gives the result. □

Proposition 7.12. *Let A be a simple abelian variety defined over a number field, and G_ℓ its ℓ-adic monodromy group. The multiplicity of every absolutely irreducible component of $\rho_\ell \otimes_{\mathbb{Q}_\ell} \bar{\mathbb{Q}}_\ell$ divides $u(G_\ell^\circ)$.*

Proof. Let $D = \operatorname{End} A \otimes_{\mathbb{Z}} \mathbb{Q}$ be the endomorphism algebra of A, and let E be the centre of D. Let $m^2 = \dim_E D$.

By Faltings' Theorem, $\operatorname{End} \rho_\ell = D \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$. This is a product of simple algebras, each of dimension m^2 over its centre. So every absolutely irreducible component of $\rho_\ell \otimes_{\mathbb{Q}_\ell} \bar{\mathbb{Q}}_\ell$ has multiplicity m, and it will suffice to show that m divides $u(G_\ell^\circ)$.

There are two cases: E is totally real or a CM field.

Case 1. E is totally real. In this case the Albert classification of endomorphism algebras of abelian varieties implies that $m \leq 2$, so it will suffice to show that 2 divides $u(G_\ell^\circ)$.

Let H_ℓ be the identity component of $\ker \det \rho_\ell \subset G_\ell$ (in other words, the ℓ-adic analogue of the Hodge group). By [Tan80] Lemma 1.4, the condition that E is totally real implies that the Hodge group of A is semisimple and by [SZ96] Theorem 3.2 this implies that H_ℓ is semisimple. Hence H_ℓ is the derived group of G_ℓ°.

Let μ be a weak Hodge cocharacter of G_ℓ in the sense of [Pin98] Definition 3.2 and let T be a maximal torus of G_ℓ containing the image of μ. Then $\rho_\ell \circ \mu$ has weights 0 and 1 each with multiplicity $\dim A$, so by Lemma 7.11,

$$\langle \omega, \mu_H \rangle = \pm \tfrac{1}{2}$$

for all characters ω in $\rho_{\ell|T}$, where μ_H is the component of μ in the quasi-cocharacter space of $T \cap H_\ell$.

Now $\langle -, \mu \rangle$ takes integer values on the roots of G_ℓ°. Since μ_0 is orthogonal to the roots, the same is true for $\langle -, \mu_H \rangle$. Because H_ℓ is semisimple, it is equal to the derived group of G_ℓ°, so μ_H is orthogonal to all characters which vanish on $T \cap G_\ell^{\circ\mathrm{der}}$. Hence $\langle -, \mu_H \rangle$ takes integer values on $\Lambda_0(G_\ell^\circ)$.

So in order for $\langle \omega, \mu_H \rangle$ to have denominator 2, the order of ω in $\Lambda(G_\ell)/\Lambda_0(G_\ell^\circ)$ must be even. Therefore $u(G_\ell^\circ)$ is divisible by 2.

Case 2. ***E* is a CM field.** For each place λ of E, let E_λ denote the completion of E at λ. Then $D_\lambda = D \otimes_E E_\lambda$ is a matrix ring over a division algebra with centre E_λ. Let m_λ be the order of $[D_\lambda]$ in $\operatorname{Br} E_\lambda$. By the Albert–Brauer–Hasse–Noether theorem ([Pie82] Theorem 18.5), the map $[D] \mapsto ([D_\lambda])$ is an injection

$$\operatorname{Br} E \to \bigoplus_\lambda \operatorname{Br} E_\lambda$$

so m is the lowest common multiple of the m_λ. So it suffices to show that m_λ divides $u(G_\ell^\circ)$ for every place λ.

Since E is a CM field, all its archimedean places have trivial Brauer group, so we need only consider non-archimedean places. Let λ be a non-archimedean place of E and ℓ' its residue characteristic. Then

$$\operatorname{End} \rho_{\ell'} = D \otimes_{\mathbb{Q}} \mathbb{Q}_{\ell'} = D \otimes_E \left(\prod_{\lambda' | \ell'} E_{\lambda'} \right) = \prod_{\lambda' | \ell'} D_{\lambda'}.$$

Hence $\rho_{\ell'}$ has a $\mathbb{Q}_{\ell'}$-irreducible subrepresentation with endomorphism algebra D_λ.

So by Lemma 7.7, m_λ divides $u(G_{\ell'}^\circ)$, and this is equal to $u(G_\ell^\circ)$ by Lemma 7.10. □

Theorem 7.2 follows from Corollary 7.5, Proposition 7.12 and the bounds for $u(G)$ in section 7.4.4.

7.4.4 Bounds for $u(G)$

Definition. Let $g(n)$ be the maximum value of $\operatorname{LCM}(a_i)$ where a_i are positive integers satisfying $\sum a_i = n$. (This is Landau's function.)

Let $g_1(n)$ be the maximum value of $\operatorname{LCM}(a_i)$ where a_i are integers greater than 1 satisfying $\sum (a_i - 1) = n$.

For $n \geq 2$, let

$$\alpha(n) = \frac{\log_2 g_1(n)}{\sqrt{n \log n}}.$$

Lemma 7.13. *For any reductive group G, $u(G) \leq g_1(\operatorname{rk} G)$.*

Proof. Let Φ_i (for $i \in I$) be the simple components of the root system of G.

The group Λ/Λ_0 is a subgroup of the product of the fundamental groups of the Φ_i. So $u(G)$ divides the lowest common multiple of the exponents of these fundamental groups.

Let e_i be the exponent of the fundamental group of Φ_i. Then $e_i \leq \operatorname{rk} \Phi_i + 1$ for all i (by the classification of simple root systems), and so $\sum_i (e_i - 1) \leq \operatorname{rk} G$.

By the definition of g_1,

$$u(G) \leq g_1\left(\sum_{i\in I}(e_i - 1)\right)$$

and this is less than or equal to $g_1(\operatorname{rk} G)$ because g_1 is nondecreasing. □

Corollary 7.14. *$\alpha(n) \to \frac{1}{\log 2}$ as $n \to \infty$ and $\alpha(n) < 2$ for all $n \geq 2$.*

Proof. We use two results on the size of $g(n)$: Landau's asymptotic result ([Lan09] section 61)

$$\frac{\log_e g(n)}{\sqrt{n \log n}} \to 1 \text{ as } n \to \infty$$

and Massias' bound [Mas84b]

$$\log_e g(n) < 1.05314\sqrt{n \log n} \text{ for all } n \geq 2.$$

We note that $g(n) \leq g_1(n) \leq g(n + \lfloor\sqrt{2n}\rfloor)$ since any set of distinct positive integers satisfying $\sum_i (a_i - 1) = n$ will satisfy $\sum_i a_i \leq n + \lfloor\sqrt{2n}\rfloor$.

Let

$$f(x) = \frac{(x + \sqrt{2x})\log(x + \sqrt{2x})}{x \log x}.$$

Since $f(x) \to 1$ as $x \to \infty$, we conclude that $\alpha(n) \to \frac{1}{\log 2}$.

Likewise by Massias' bound

$$\alpha(n) \leq \frac{\log_e g(n + \lfloor\sqrt{2n}\rfloor)}{\log 2\sqrt{n \log n}} < \frac{1.05314\sqrt{f(n)}}{\log 2} \leq \frac{1.05314\sqrt{f(9)}}{\log 2} < 2$$

for $n \geq 9$ since $f(x)$ is decreasing for $x > 1$.

Manual calculation shows that $\alpha(n) < 2$ for $2 \leq n \leq 8$. □

7.5 Some examples

In this section, we will give three examples of families of abelian varieties with commutative endomorphism ring for which Theorem 7.1 is sharp. Note that any abelian variety for which the rank of the Mumford–Tate group is equal to the bound of Theorem 7.1 necessarily satisfies the Mumford–Tate conjecture, because the rank of the ℓ-adic monodromy groups are less than or equal to that of the Mumford–Tate group but satisfy the same lower bound. Hence the examples we give show that the bound is sharp for the ℓ-adic monodromy group as well as for the Mumford–Tate group.

We also give one family of simple abelian varieties with noncommutative endomorphism ring for which the Mumford–Tate group has rank n and the dimension g satisfies $\log_2 g = n + \frac{1}{2}\log_2 n + \mathrm{O}(1)$. This shows that the bound in Theorem 7.2 cannot be improved to $n \geq \log_2 g + \mathrm{O}(1)$. Because we have not calculated the exact lower bound in the noncommutative case we cannot deduce that these varieties satisfy the Mumford–Tate conjecture purely from the rank bound. But for the examples constructed here, we can show that they satisfy the Mumford–Tate conjecture by using [Pin98] Proposition 4.3.

7.5.1 Examples with commutative endomorphism ring

Example 1: Complex multiplication. Let F be a totally real field such that $[F : \mathbb{Q}] = n - 1$. By [Shi70] Theorem 1.10, there is an imaginary quadratic extension K of F such that for every CM type (K, Φ), the reflex type (K', Φ') satisfies $[K' : \mathbb{Q}] = 2^{n-1}$. Such a CM type is primitive.

Let A be a complex abelian variety corresponding to the CM type (K', Φ'). Then the Mumford–Tate group M is a torus, isomorphic to the image of the homomorphism $\mathrm{Res}_{K/\mathbb{Q}} \mathbb{G}_m \to \mathrm{Res}_{K'/\mathbb{Q}} \mathbb{G}_m$ induced by the reflex norm $K^\times \to K'^\times$.

This image has rank at most $[K : \mathbb{Q}] + 1 = n + 1$. But $\dim A = 2^{n-1}$ so by Theorem 7.1, $\mathrm{rk}\, M \geq n + 1$. So in fact $\mathrm{rk}\, M = n + 1 = \log_2 \dim A + 2$.

The endomorphism ring of A is the field K'.

Example 2: Spin group. This example generalises the Kuga-Satake construction of an abelian variety attached to a polarised $K3$ surface [KS67].

Let n be a positive integer congruent to 1 or 2 mod 4. Let W be a $\mathbb{Q}$-vector space of dimension $2n + 1$, and let Q be the quadratic form

$$Q(x) = x_1^2 + x_2^2 - x_3^2 - \cdots - x_{2n+1}^2$$

of signature $(2, 2n - 1)$. The even Clifford algebra $C^+(W, Q)$ is isomorphic to $\mathrm{M}_{2^n}(\mathbb{Q})$, and so it has a unique faithful irreducible $\mathbb{Q}$-representation of dimension 2^n, called the spin representation.

Let M be the Clifford group

$$\mathrm{GSpin}(W, Q) = \{x \in C^+(W, Q) \mid xWx^{-1} \subset W\}.$$

This is a reductive group of rank $n + 1$, with root system B_n and centre $\mathbb{G}_m$. Let $\rho : M \to \mathrm{GL}(V)$ be the spin representation of M. This is an absolutely irreducible representation of dimension 2^n.

Let $\{e_1, e_2\}$ be an orthonormal basis for the positive definite subspace of W. The homomorphism $\varphi : \mathbb{C}^\times \to M(\mathbb{R})$ given by

$$\varphi(a + ib) = a + be_1e_2$$

defines a Hodge structure on V of type $\{(0,-1),(-1,0)\}$. The conditions on n mod 4 and on the signature of W ensure that this Hodge structure is polarisable.

Because M^{der} is almost simple, replacing φ by a generic $M(\mathbb{R})$-conjugate gives a Hodge structure whose Mumford–Tate group is M. Let A be a complex abelian variety corresponding to such a Hodge structure. It has dimension 2^{n-1} and endomorphism algebra $\mathbb{Q}$, and its Mumford–Tate group has rank $n+1$.

Example 3: Product of copies of SL_2. This example generalises the example of Mumford [Mum69] of a family of abelian varieties of dimension 4 with Mumford–Tate group M such that $M_{\mathbb{C}}$ is isogenous to $\mathbb{G}_m \times (\mathrm{SL}_2)^3$.

Let n be an odd positive integer, and F a totally real number field of degree n. Let D be a quaternion algebra over F such that:

(i) $\mathrm{Cor}_{F/\mathbb{Q}} D$ is split over $\mathbb{Q}$, i.e. is isomorphic to $\mathrm{M}_{2^n}(\mathbb{Q})$.

(ii) D is split at exactly one real place of F.

Let M be the $\mathbb{Q}$-algebraic group $M(A) = \{x \in (D \otimes A)^\times | x\bar{x} \in A^\times\}$ (where $\bar{x}$ is the standard involution of D). By condition (ii), $M_{\mathbb{R}}$ is isomorphic to

$$\left(\mathbb{G}_{m,\mathbb{R}} \times \mathrm{SL}_{2,\mathbb{R}} \times \mathrm{SU}_2^{n-1}\right) / \{(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n) \mid \varepsilon_i \in \{\pm 1\}, \varepsilon_0\varepsilon_1 \cdots \varepsilon_n = 1\}.$$

By condition (i), M has a faithful irreducible $\mathbb{Q}$-representation ρ of dimension 2^n. Then $\rho \otimes_{\mathbb{Q}} \mathbb{C}$ is isomorphic to the tensor product of the standard 1-dimensional representation of $\mathbb{G}_{m,\mathbb{C}}$ with the standard 2-dimensional representation of each factor $\mathrm{SL}_{2,\mathbb{C}}$.

Let $\varphi : \mathbb{C}^\times \to M(\mathbb{R})$ be the homomorphism

$$\varphi(a+ib) = \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \text{ in } \mathrm{GL}_2 \cong (\mathbb{G}_m \times \mathrm{SL}_2)/\{\pm 1\}$$
$$\text{and trivial in the } \mathrm{SU}_2 \text{ factors.}$$

Then $\rho \circ \phi$ defines a Hodge structure of type $\{(0,-1),(-1,0)\}$. By condition (ii), this Hodge structure is polarisable.

Again M^{der} is almost $\mathbb{Q}$-simple, so replacing φ by a generic element of its $M(\mathbb{R})$-conjugacy class gives a Hodge structure with Mumford–Tate group equal to M. An abelian variety corresponding to such a Hodge structure will have dimension 2^{n-1}, endomorphism algebra $\mathbb{Q}$ and Mumford–Tate group of rank $n+1$.

7.5.2 An example with large multiplicity

Let n be an odd integer and $r = (n-1)/2$. We will construct a simple abelian variety of dimension $g(n) = n\binom{n}{r}$ whose Mumford–Tate group is a $\mathbb{Q}$-form of GL_n.

The Mumford–Tate representation is isomorphic over $\mathbb{C}$ to the sum of $2n$ copies of the r-th exterior power of the standard representation. By Stirling's formula $\log_2 g(n) = n + \frac{1}{2}\log_2 n + \mathrm{O}(1)$.

Let K be an imaginary quadratic field, and D a central division algebra over K of dimension n^2 with an involution $*$ of the second kind. The $\mathbb{Q}$-algebraic groups

$$H(A) = \{d \in (D \otimes_{\mathbb{Q}} A)^\times \mid dd^* = 1\},$$
$$G(A) = \{d \in (D \otimes_{\mathbb{Q}} A)^\times \mid dd^* \in A^\times\}$$

are $\mathbb{Q}$-forms of SL_n and GL_n. By choosing $*$ appropriately, we may suppose that $H_\mathbb{R}$ is the unitary group of a Hermitian form of signature $(1, n-1)$.

We can view D as a K-irreducible representation of H_K. Over $\mathbb{C}$, $D_\mathbb{C}$ is isomorphic to the sum of n copies of the standard representation of SL_n, so its highest weight is ϖ_1. The endomorphism ring of this representation is D^{op}, so

$$\alpha_{H,K}(\varpi_1) = [D]$$

for Tits' homomorphism $\alpha_{H,K} : \Lambda^\Gamma \to \mathrm{Br}\, K$.

Let $r = (n-1)/2$ and let $\tilde{D}$ be the central division algebra over K such that $[\tilde{D}] = [D]^r$ in $\mathrm{Br}\, K$. Now $[D]$ has order n in $\mathrm{Br}\, K$. Since r and n are coprime, $[\tilde{D}]$ also has order n and $\tilde{D} \otimes_K \mathbb{C} \cong M_n(\mathbb{C})$.

Let $\tilde{\rho}$ be the K-irreducible representation of H_K with highest weight ϖ_r. We know that $\varpi_r \equiv r\varpi_1$ modulo the roots of H_K, so $\alpha_{H,K}(\varpi_r) = [D]^r = [\tilde{D}]$. Hence $\tilde{\rho}$ has endomorphism ring $\tilde{D}^{\mathrm{op}}$, so $\tilde{\rho}_\mathbb{C}$ is the sum of n copies of an irreducible representation of SL_n. This irreducible representation is the r-th exterior power of the standard representation, so $\dim_K \tilde{\rho} = n\binom{n}{r}$.

If λI is a scalar matrix in $H(\mathbb{C})$, then $\tilde{\rho}_\mathbb{C}(\lambda I)$ is multiplication by λ^r. So we can extend $\tilde{\rho}$ to a representation of G_K by letting each scalar matrix λI act as multiplication by λ^r.

Let $\rho = \mathrm{Res}_{K/\mathbb{Q}}\, \tilde{\rho}$. This is a $\mathbb{Q}$-irreducible representation of G of dimension $2n\binom{n}{r}$. We have $\ker \rho = \mu_r$ so ρ factorises through $M = G/\mu_r$, and the resulting representation of M is faithful.

In order to specify the Hodge structure, we will first define $\varphi' : \mathbb{C}^\times \to G(\mathbb{R})$ as follows: recall that $H_\mathbb{R}$ is the unitary group of a Hermitian form Ψ of signature $(1, n-1)$. Then let $\phi'(z)$ act as $z^r/\bar{z}^{r-1}$ on the 1-dimensional space where h is positive definite and as $\bar{z}$ on the $(n-1)$-dimensional space where h is negative definite.

Then $\rho \circ \varphi'$ has weights z^r and $\bar{z}^r$. Because ρ is faithful as a representation of M, it follows that there is a homomorphism $\varphi : \mathbb{C}^\times \to M(\mathbb{R})$ whose r-th power is φ'. Then (M, ρ, φ) defines a $\mathbb{Q}$-Hodge structure of type $\{(-1,0),(0,-1)\}$. The Hermitian form Ψ induces a polarisation of this Hodge structure.

Once again, M^{der} is almost simple, so replacing φ by a generic $M(\mathbb{R})$-conjugate gives a Hodge structure with Mumford–Tate group M. A corresponding abelian variety will have endomorphism algebra $\tilde{D}^{\mathrm{op}}$ and dimension $g = n\binom{n}{r}$.

We shall confirm that this variety satisfies the Mumford–Tate conjecture. Let σ be an absolutely irreducible component of $\rho \otimes_{\mathbb{Q}} \bar{\mathbb{Q}}_\ell$. Then $(G \times_{\mathbb{Q}} \bar{\mathbb{Q}}_\ell, \sigma)$, with a suitable set of cocharacters, form a weak Mumford–Tate triple of weights $\{0, 1\}$. By Faltings' theorem, the restriction of σ to $G_{\ell, \bar{\mathbb{Q}}_\ell}$ must remain irreducible, where G_ℓ is the ℓ-adic monodromy group. It also is part of a weak Mumford–Tate triple of weights $\{0, 1\}$. But our $(G \times_{\mathbb{Q}} \bar{\mathbb{Q}}_\ell, \sigma)$ is in the fourth column of [Pin98] Table 4.2: type A with σ not the standard representation. Hence according to Pink's Proposition 4.3, $G_\ell = G \times_{\mathbb{Q}} \mathbb{Q}_\ell$.

References

[And89] Y. André, *G-functions and geometry*, Aspects of Mathematics, E13, Friedr. Vieweg & Sohn, Braunschweig, 1989.

[And92] ———, *Mumford-Tate groups of mixed Hodge structures and the theorem of the fixed part*, Compositio Math. **82** (1992), no. 1, 1–24.

[Ax71] J. Ax, *On Schanuel's conjectures*, Ann. of Math. (2) **93** (1971), 252–268.

[Ax72] ———, *Some topics in differential algebraic geometry. I. Analytic subgroups of algebraic groups*, Amer. J. Math. **94** (1972), 1195–1204.

[Bai58] W. L. Baily, Jr., *Satake's compactification of V_n*, Amer. J. Math. **80** (1958), 348–364.

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

[BO07] Y. Benoist and H. Oh, *Polar decomposition for p-adic symmetric spaces*, Int. Math. Res. Not. IMRN (2007), no. 24, Art. ID rnm121, 20.

[Bog81] F. A. Bogomolov, *Points of finite order on an abelian variety*, Izv. Math. **17** (1981), no. 1, 55–72.

[Bor69] A. Borel, *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.

[Bor72] ———, *Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem*, J. Differential Geometry **6** (1972), 543–560.

[Car57] H. Cartan, *Quotient d'un espace analytique par un groupe d'automorphismes*, Algebraic geometry and topology, Princeton University Press, Princeton, N. J., 1957, pp. 90–102.

[Coh96] P. B. Cohen, *Humbert surfaces and transcendence properties of automorphic functions*, Rocky Mountain J. Math. **26** (1996), no. 3, 987–1001.

[Del71] P. Deligne, *Travaux de Shimura*, Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389, Springer, Berlin, 1971, pp. 123–165. Lecture Notes in Math., Vol. 244.

[Del79] ———, *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, Amer. Math. Soc., Providence, R.I., 1979, pp. 247–289.

[Del82] ———, *Hodge cycles on abelian varieties (notes by J. S. Milne)*, Hodge cycles, motives and Shimura varieties, Lecture Notes in Mathematics, vol. 900, Springer-Verlag, Berlin, 1982, pp. 9–100.

[Dod87] B. Dodson, *On the Mumford-Tate group of an abelian variety with complex multiplication*, J. Algebra **111** (1987), no. 1, 49–73.

[EY03] B. Edixhoven and A. Yafaev, *Subvarieties of Shimura varieties*, Ann. of Math. (2) **157** (2003), no. 2, 621–645.

[Fal83a] G. Faltings, *Arakelov's theorem for abelian varieties*, Invent. Math. **73** (1983), no. 3, 337–347.

[Fal83b] ———, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366.

[GR12] É. Gaudron and G. Rémond, *Polarisations et isogénies*, Preprint, available at `http://math.univ-bpclermont.fr/~gaudron/art13.pdf`, 2012.

[Har77] R. Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977.

[HP12] P. Habegger and J. Pila, *Some unlikely intersections beyond André-Oort*, Compos. Math. **148** (2012), no. 1, 1–27.

[HR10] M. Hindry and N. Ratazzi, *Torsion dans un produit de courbes elliptiques*, J. Ramanujan Math. Soc. **25** (2010), no. 1, 81–111.

[HW93] A. G. Helminck and S. P. Wang, *On rationality properties of involutions of reductive groups*, Adv. Math. **99** (1993), no. 1, 26–96.

[KMRT98] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, American Mathematical Society Colloquium Publications, vol. 44, American Mathematical Society, Providence, RI, 1998.

[KS67] M. Kuga and I. Satake, *Abelian varieties attached to polarized K_3-surfaces*, Math. Ann. **169** (1967), 239–242.

[KY13] B. Klingler and A. Yafaev, *The André–Oort conjecture*, Ann. of Math. (2) (to appear) (2013), available at `http://arxiv.org/abs/1209.0936`.

[Lan09] E. Landau, *Handbuch der lehre von der verteilung der primzahlen*, vol. I, 1909.

[Lew82] D. W. Lewis, *Quaternionic skew-Hermitian forms over a number field*, J. Algebra **74** (1982), no. 1, 232–240.

[LP90] M. Larsen and R. Pink, *Determining representations from invariant dimensions*, Invent. Math. **102** (1990), no. 2, 377–398.

[Mar91] G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 17, Springer-Verlag, Berlin, 1991.

[Mas84a] D. W. Masser, *Small values of the quadratic part of the Néron-Tate height on an abelian variety*, Compositio Math. **53** (1984), no. 2, 153–170.

[Mas84b] J.-P. Massias, *Majoration explicite de l'ordre maximum d'un élément du groupe symétrique*, Ann. Fac. Sci. Toulouse Math. (5) **6** (1984), no. 3-4, 269–281 (1985).

[Mil86] J. S. Milne, *Abelian varieties*, Arithmetic geometry (Storrs, Conn., 1984), Springer, New York, 1986, pp. 103–150.

[Mil05] ______, *Introduction to Shimura varieties*, Harmonic analysis, the trace formula, and Shimura varieties, Clay Math. Proc., vol. 4, Amer. Math. Soc., Providence, RI, 2005, pp. 265–378.

[Mil08] ______, *Abelian varieties (v2.00)*, Available at `http://www.jmilne.org/math/CourseNotes/av.html`, 2008.

[Moo98] B. Moonen, *Linearity properties of Shimura varieties. I*, J. Algebraic Geom. **7** (1998), no. 3, 539–567.

[Mum65] D. Mumford, *Geometric invariant theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 34, Springer-Verlag, Berlin, 1965.

[Mum69] ______, *A note of Shimura's paper "Discontinuous groups and abelian varieties"*, Math. Ann. **181** (1969), 345–351.

[MW93a] D. Masser and G. Wüstholz, *Isogeny estimates for abelian varieties, and finiteness theorems*, Ann. of Math. (2) **137** (1993), no. 3, 459–472.

[MW93b] ______, *Periods and minimal abelian subvarieties*, Ann. of Math. (2) **137** (1993), no. 2, 407–458.

[Noo95] R. Noot, *Abelian varieties—Galois representation and properties of ordinary reduction*, Compositio Math. **97** (1995), no. 1-2, 161–171.

[O'M63] O. T. O'Meara, *Introduction to quadratic forms*, Die Grundlehren der mathematischen Wissenschaften, Bd. 117, Academic Press Inc., Publishers, New York, 1963.

[Oor97] F. Oort, *Canonical liftings and dense sets of CM-points*, Arithmetic geometry (Cortona, 1994), Sympos. Math., XXXVII, Cambridge Univ. Press, Cambridge, 1997, pp. 228–234.

[Pie82] R. S. Pierce, *Associative algebras*, Graduate Texts in Mathematics, vol. 88, Springer-Verlag, New York, 1982.

[Pil09a] J. Pila, *On the algebraic points of a definable set*, Selecta Math. (N.S.) **15** (2009), no. 1, 151–170.

[Pil09b] ———, *Rational points of definable sets and results of André–Oort–Manin–Mumford type*, Int. Math. Res. Not. IMRN (2009), no. 13, 2476–2507.

[Pil11] ———, *O-minimality and the André–Oort conjecture for $\mathbb{C}^n$*, Ann. of Math. (2) **173** (2011), no. 3, 1779–1840.

[Pin89] R. Pink, *Arithmetical compactification of mixed Shimura varieties*, Bonner Mathematische Schriften **209** (1989), Thesis.

[Pin98] ———, *ℓ-adic algebraic monodromy groups, cocharacters, and the Mumford-Tate conjecture*, J. Reine Angew. Math. **495** (1998), 187–237.

[Pin05a] ———, *A combination of the conjectures of Mordell-Lang and André-Oort*, Geometric methods in algebra and number theory, Progr. Math., vol. 235, Birkhäuser Boston, Boston, MA, 2005, pp. 251–282.

[Pin05b] ———, *A common generalization of the conjectures of André–Oort, Manin-Mumford, and Mordell–Lang*, Preprint, available at `http://www.math.ethz.ch/~pink/ftp/AOMMML.pdf`, 2005.

[PR94] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press Inc., Boston, MA, 1994, Translated from the 1991 Russian original by Rachel Rowen.

[PS13] Y. Peterzil and S. Starchenko, *Definability of restricted theta functions and families of abelian varieties*, Duke Math. J. **162** (2013), no. 4, 731–765.

[PT12] J. Pila and J. Tsimerman, *Ax–Lindemann for $\mathcal{A}_g$*, Preprint, available at `http://arxiv.org/abs/1206.2663`, 2012.

[PT13] ———, *The André–Oort conjecture for the moduli space of abelian surfaces*, Compositio Math. **149** (2013), no. 2, 204–216.

[PW06] J. Pila and A. J. Wilkie, *The rational points of a definable set*, Duke Math. J. **133** (2006), no. 3, 591–616.

[PZ08] J. Pila and U. Zannier, *Rational points in periodic analytic sets and the Manin–Mumford conjecture*, Rend. Lincei (9) Mat. Appl. **19** (2008), no. 2, 149–162.

[Rag72] M. S. Raghunathan, *Discrete subgroups of Lie groups*, Springer-Verlag, New York, 1972.

[Ray83a] M. Raynaud, *Courbes sur une variété abélienne et points de torsion*, Invent. Math. **71** (1983), no. 1, 207–233.

[Ray83b] ———, *Sous-variétés d'une variété abélienne et points de torsion*, Arithmetic and geometry, Vol. I, Progr. Math., vol. 35, Birkhäuser Boston, Boston, MA, 1983, pp. 327–352.

[Rib81] K. A. Ribet, *Division fields of abelian varieties with complex multiplication*, Mém. Soc. Math. France (N.S.) (1980/81), no. 2, 75–94.

[Ric67] R. W. Richardson, Jr., *Conjugacy classes in Lie algebras and algebraic groups*, Ann. of Math. (2) **86** (1967), 1–15.

[Sai93] M.-H. Saitō, *Classification of nonrigid families of abelian varieties*, Tohoku Math. J. (2) **45** (1993), no. 2, 159–189.

[Sat56] I. Satake, *On the compactification of the Siegel space*, J. Indian Math. Soc. (N.S.) **20** (1956), 259–281.

[Sca12] T. Scanlon, *Counting special points: Logic, diophantine geometry, and transcendence theory*, Bull. Amer. Math. Soc. (N.S.) **49** (2012), no. 1, 51–71.

[Sch85] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften, vol. 270, Springer-Verlag, Berlin, 1985.

[Ser79] J.-P. Serre, *Groupes algébriques associés aux modules de Hodge-Tate*, Journées de Géométrie Algébrique de Rennes. (Rennes, 1978), Vol. III, Astérisque, vol. 65, Soc. Math. France, Paris, 1979, pp. 155–188.

[Ser81] ———, *Lettre à Ken Ribet de 1/1/1981*, Œuvres. Collected papers. IV, 1981, pp. 6–17.

[Ser85] ———, *Résumés des cours au Collège de France, 1984-85*, Œuvres. Collected papers. IV, 1985.

[Shi63] G. Shimura, *Arithmetic of alternating forms and quaternion hermitian forms*, J. Math. Soc. Japan **15** (1963), 33–65.

[Shi70] ———, *On canonical models of arithmetic quotients of bounded symmetric domains*, Ann. of Math. (2) **91** (1970), 144–222.

[Shi97] ———, *Euler products and Eisenstein series*, CBMS Regional Conference Series in Mathematics, vol. 93, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[SW95] H. Shiga and J. Wolfart, *Criteria for complex multiplication and transcendence properties of automorphic functions*, J. Reine Angew. Math. **463** (1995), 1–25.

[SZ96] A. Silverberg and Yu. G. Zarhin, *Hodge groups of abelian varieties with purely multiplicative reduction*, Izv. Math. **60** (1996), no. 2, 379–389.

[Tan80] S. G. Tankeev, *Algebraic cycles on abelian varieties. II*, Izv. Math. **14** (1980), no. 2, 383–394.

[Tit71] J. Tits, *Représentations linéaires irréductibles d'un groupe réductif sur un corps quelconque*, J. Reine Angew. Math. **247** (1971), 196–220.

[Ull12] E. Ullmo, *Quelques applications du théorème de Ax–Lindemann hyperbolique*, Compositio Math. (to appear) (2012), available at `http://www.math.u-psud.fr/~ullmo/Prebublications/cordAxLind.pdf`.

[UY11] E. Ullmo and A. Yafaev, *A characterization of special subvarieties*, Mathematika **57** (2011), no. 2, 263–273.

[UY13a] ———, *Galois orbits and equidistribution of special subvarieties: towards the André–Oort conjecture*, Ann. of Math. (2) (to appear) (2013), available at `http://annals.math.princeton.edu/articles/8222`.

[UY13b] ———, *Hyperbolic Ax–Lindemann theorem in the cocompact case*, Duke Math. J. (to appear) (2013), available at `http://arxiv.org/abs/1209.0939`.

[vdD98] L. van den Dries, *Tame topology and o-minimal structures*, London Mathematical Society Lecture Note Series, vol. 248, Cambridge University Press, Cambridge, 1998.

[vdDM94] L. van den Dries and C. Miller, *On the real exponential field with restricted analytic functions*, Israel J. Math. **85** (1994), no. 1-3, 19–56.

[Win86] J.-P. Wintenberger, *Groupes algébriques associés à certaines representations p-adiques*, Amer. J. Math. **108** (1986), no. 6, 1425–1466.

[Zar84] Yu. G. Zarkhin, *Weights of simple Lie algebras in the cohomology of algebraic varieties*, Izv. Math. **24** (1984), no. 2, 245–281.

[Zil02] B. Zilber, *Exponential sums equations and the Schanuel conjecture*, J. London Math. Soc. (2) **65** (2002), no. 1, 27–44.