Quadratic Chabauty for modular curves and modular forms of rank one

Netan Dogra[1] · **Samuel Le Fourn[2]**

Abstract

In this paper, we provide refined sufficient conditions for the quadratic Chabauty method on a curve X to produce an effective finite set of points containing the rational points $X(\mathbb{Q})$, with the condition on the rank of the Jacobian of X replaced by condition on the rank of a quotient of the Jacobian plus an associated space of Chow–Heegner points. We then apply this condition to prove the effective finiteness of $X(\mathbb{Q})$ for any modular curve $X = X_0^+(N)$ or $X_{\mathrm{ns}}^+(N)$ of genus at least 2 with N prime. The proof relies on the existence of a quotient of their Jacobians whose Mordell–Weil rank is equal to its dimension (and at least 2), which is proven via analytic estimates for orders of vanishing of L-functions of modular forms, thanks to a Kolyvagin–Logachev type result.

1 Introduction

The Chabauty–Kim method is a method for determining the set $X(\mathbb{Q})$ of rational points of a curve X over $\mathbb{Q}$ of genus bigger than 1. The idea is to locate $X(\mathbb{Q})$ inside $X(\mathbb{Q}_p)$ by finding an obstruction to a p-adic point being global. The method developed in [39,40] produces a tower of obstructions

$$X(\mathbb{Q}_p) \supset X(\mathbb{Q}_p)_1 \supset X(\mathbb{Q}_p)_2 \supset \ldots \supset X(\mathbb{Q})$$

✉ Netan Dogra
  netan.dogra@kcl.ac.uk

[1] Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK

[2] Université Grenoble Alpes, CNRS, Saint-Martin-d'Héres, IF 38000, France

In [5], it is conjectured that $X(\mathbb{Q}_p)_n = X(\mathbb{Q})$ for all $n \gg 0$, and in [40] it is proved that standard conjectures in arithmetic geometry imply $X(\mathbb{Q}_p)_n$ is finite for all $n \gg 0$, but in general these results are not known.

The first obstruction set $X(\mathbb{Q}_p)_1$ is the one produced by Chabauty's method. In situations when $X(\mathbb{Q}_p)_1$ is finite, it can often be used to determine $X(\mathbb{Q})$.

The main results of this paper concern the finiteness of the Chabauty–Kim set $X(\mathbb{Q}_p)_2$ when X is one of the modular curves $X_{\mathrm{ns}}^+(N)$ or $X_0^+(N)$ (N a prime different from p), whose definition and properties we now recall briefly (more details are given in Sect. 4).

The curve $X_0^+(N)$ is the quotient of $X_0(N)$ by the Atkin–Lehner involution w_N. The curve $X_{\mathrm{ns}}^+(N)$ is the quotient of $X(N)$ by the normalizer of a nonsplit Cartan subgroup. Determining the rational points of $X_{\mathrm{ns}}^+(N)$ would resolve Serre's uniformity question [58, §4.3]: is there an N_0 such that, for all $N > N_0$ and all elliptic curves E defined over $\mathbb{Q}$ without complex multiplication, the mod N Galois representation

$$\rho_{E,N} \colon \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(E[N])$$

is surjective? The Borel and normalizer of split Cartan subgroups of Serre's uniformity question have been given a positive answer respectively in the celebrated papers [50] and [11,12].

Mazur's proof may, very crudely, be described as having two stages.

1. Construct a non-constant map $f \colon X \to A$ from X to an abelian variety of rank zero over $\mathbb{Q}$.
2. Compute the finite set $A(\mathbb{Q})$, and the pre-image $f^{-1}(A(\mathbb{Q})) \supset X(\mathbb{Q})$.

As is explained in Sect. 4, in contrast to $X_0(N)$ and $X_s^+(N)$, for $X = X_0^+(N)$ or $X = X_{\mathrm{ns}}^+(N)$, the Birch–Swinnerton-Dyer conjecture implies that there are *no* non-constant maps from A to abelian varieties of rank zero over $\mathbb{Q}$. It is hence natural to ask whether we can attempt to mimic Mazur's strategy, with the set $A(\mathbb{Q})$ replaced by the set $X(\mathbb{Q}_p)_n$ for some n. In Sect. 4, we show that the Birch–Swinnerton-Dyer conjecture similarly implies $X(\mathbb{Q}_p)_1$ is infinite, hence we expect to need $n > 1$. The main result of this paper is to carry out the *first* stage of Mazur's strategy for $n = 2$.

Theorem 1 *1. For all prime N such that $g(X_0^+(N)) \geq 2$, $X_0^+(N)(\mathbb{Q}_p)_2$ is finite for any $p \neq N$.*

2. For all prime N such that $g(X_{\mathrm{ns}}^+(N)) \geq 2$ and $X_{\mathrm{ns}}^+(N)(\mathbb{Q}) \neq \emptyset$, $X_{\mathrm{ns}}^+(N)(\mathbb{Q}_p)_2$ is finite for any $p \neq N$.

Remark 1 – For all primes N for which one of the curves X above has genus 0 or 1, $X(\mathbb{Q})$ is infinite. Indeed, the prime numbers N such that $g(X_0^+(N)) \leq 1$ make up a finite list with maximal element 131 [27, Propositions 3.1 and 3.2], and the elliptic cases cases can then be checked on the LMFDB [48] by looking at the corresponding explicit elliptic curves sorted by conductor (the genus 0 case is automatic due to the rational cusp). For the nonsplit Cartan modular curve, the genus formula [55, Proposition 13] proves that $g(X_{\mathrm{ns}}^+(N)) \leq 1$ if and only if $N \leq 11$, and these 5 cases are sorted similarly, as one can always find a rational point associated to an elliptic curve with CM coming by one of the 9 class number one fields.

– The only reason for the assumption that $X_{ns}^+(N)(\mathbb{Q})$ is nonempty is that the definition of $X(\mathbb{Q}_p)_2$ currently assumes that X has a rational point (if Serre's uniformity question has a positive answer, then there are infinitely many N for which $X_{ns}^+(N)(\mathbb{Q})$ is empty). One can modify the definition of $X(\mathbb{Q}_p)_2$ - for example in a similar manner to [33] - to remove this assumption, and then $X_{ns}^+(N)(\mathbb{Q}_p)_2$ will be finite whenever the genus of $X_{ns}^+(N)$ is greater than 1. In particular, such a modification should in principle given a method to prove that $X_{ns}^+(N)(\mathbb{Q}_p)_2$, and hence $X_{ns}^+(N)(\mathbb{Q})$, is empty in these cases (although the large genera of such curves mean that in practice such curves are currently beyond the scope of existing computational methods for other reasons). As this involves several techniques not relevant to the proof of Theorem 1, we do not pursue this point in this paper.
– Finally, results of [3], together with Edixhoven and Parent's explicit models for $X_{ns}(N)$ [23], allow us to deduce from our result an explicit bound (polynomial in N) on the number of rational points on $X_0^+(N)$ and $X_{ns}^+(N)$, which we do in Sect. 3.1.

In this paper we say nothing about carrying out the second stage of Mazur's strategy (i.e. computing the finite set $X(\mathbb{Q}_p)_2$). However, as alluded to above, for a given X, if one can prove $X(\mathbb{Q}_p)_2$ is finite there has been significant recent progress in computing it, and $X(\mathbb{Q})$, in practice. For example, when $N = 13$, the rational points of $X_{ns}^+(N)$ are computed in [8], by computing $X(\mathbb{Q}_p)_2$. Similarly for $X = X_0^+(N)$, the rational points of all X of genus 2 are computed in [4], and in forthcoming work [9], the case of all X of genus three is handled.

The proof of Theorem 1 proceeds along the lines of the quadratic Chabauty method, which requires a precise inequality (namely (2)) in terms of invariants of the Jacobian J of X to hold (see Sect. 1.1). This inequality is expected to hold asymptotically for $X = X_0^+(N)$ or $X = X_{ns}^+(N)$ conditionally on Birch and Swinnerton-Dyer conjecture (see §4.1), but looks out of reach unconditionally for N in noncomputable range. There are thus two important steps obtained in the proof of Theorem 1:

- For p a prime of good reduction of a smooth projective geometrically irreducible curve X over $\mathbb{Q}$ with $X(\mathbb{Q}) \neq \emptyset$, $X(\mathbb{Q}_p)_2$ is finite under the condition that a similar inequality to (2) holds not for J but a quotient abelian variety A of J, *and* under an additional hypothesis (C) on X, J, A.
- For $X = X_0^+(N)$ or $X = X_{ns}^+(N)$, there is an abelian variety of A satisfying (2) and such that X, J, A satisfy (C), if for $M = N$ (resp. N^2) there are two distinct normalised eigenforms $f \in S_2(\Gamma_0(M))^{+,new}$ such that $L'(f, 1) \neq 0$.

The final input in the proof of Theorem 1 is the following Theorem.

Theorem 2 *For all $M = N$ or N^2 with N prime, if the space $S_2(\Gamma_0(M))^{+,new}$ is of dimension at least two, it contains two distinct normalised newforms f such that $L'(f, 1) \neq 0$.*

As explained in Remark 8, this result of nonvanishing is in fact quite weak compared to known or expected asymptotic estimates (giving a positive linear proportion of nonvanishing values) so the main difficulty in the proof of Theorem 2 lies in making such estimates effective enough to prove the result except for small enough N so that the remaining cases can be checked algorithmically.

1.1 Chow–Heegner points and quadratic Chabauty

In general, $X(\mathbb{Q}_p)_n$ cannot unconditionally be proved to be finite without some assumptions on the Jacobian of X (Kim showed that the Bloch–Kato conjectures imply that $X(\mathbb{Q}_p)_n$ is finite for all $n \gg 0$ [40, Observation 2]). In the case $n = 1$ (which reduces to the classical set-up of Chabauty's method) it is known that a sufficient condition is that

$$\mathrm{rk}(J) < \dim(J) \tag{1}$$

where $\mathrm{rk}(J)$ is the Mordell–Weil rank of $J(\mathbb{Q})$. The simplest instance extending Chabauty's method when finiteness of $X(\mathbb{Q}_p)_n$ can be proved for $n > 1$ is the following Lemma. To state the Lemma, let J denote the Jacobian of X, and recall that the Picard number $\rho(J)$ is defined to be the rank of the Néron–Severi group $\mathrm{NS}(J) := \mathrm{Pic}(J)/\mathrm{Pic}^0(J)$. By [51, Proposition 17.2], this is the same as the dimension of the subspace denoted by $\mathrm{End}^\dagger(J)$ of $\mathrm{End}^0(J) := \mathrm{End}(J) \otimes \mathbb{Q}$ consisting of endomorphisms that are symmetric, i.e. fixed by the Rosati involution.

Lemma 1 ([6], Lemma 3.2) *If*

$$\mathrm{rk}(J) < \dim(J) + \rho(J) - 1, \tag{2}$$

then $X(\mathbb{Q}_p)_2$ is finite. In particular, if $\mathrm{rk}(J) = \dim(J)$, then $X(\mathbb{Q}_p)_2$ is finite whenever $\rho(J) > 1$.

By Kolyvagin–Logachev type results due to Nekovář and Tian (see Proposition 8 and its Corollary 4), Theorem 2 implies that the Jacobians of $X_0^+(N)$ and $X_{\mathrm{ns}}^+(N)$, which we will henceforth denote by $J_0^+(N)$ and $J_{\mathrm{ns}}^+(N)$ respectively, do have $\mathbb{Q}$-isogeny factors A satisfying $\mathrm{rk}(A) < \dim(A) + \rho(A) - 1$, but it seems unattainable to prove unconditionally such a result for the full Jacobian. To deduce Theorem 1, we thus need a 'quadratic Chabauty for quotients' result, analogous to the well-known fact that Chabauty's method also works under the relaxed condition $\mathrm{rk}(A) < \dim(A)$, i.e. (1) for an isogeny factor A instead of J (in fact, for modular curves, Mazur–Kamienny's method refines this for factors A such that $\mathrm{rk}(A) = 0$, see e.g. [2]).

As explained below, in general such a result seems non-trivial. Fix a basepoint $b \in X(\mathbb{Q})$, and let $\mathrm{AJ} \colon X \to J$ be the corresponding Abel–Jacobi map. Let A, B be abelian varieties over $\mathbb{Q}$, satisfying $\mathrm{Hom}(A, B) = 0$, and suppose we have a surjection $(\pi_A, \pi_B) \colon J \to A \times B$.

A slight modification denoted by $\widetilde{\mathrm{AJ}}^*$ of the pullback by AJ (which basically amounts to considering the restriction of AJ^* on symmetric line bundles, see §2.1) vanishes on $\mathrm{Pic}^0(J)$, so it factors through $\mathrm{NS}(J)$ and $\widetilde{\mathrm{AJ}}^* \colon \mathrm{NS}(J) \to \mathrm{Pic}(X)$ will denote this factorisation by abuse of notation. It induces a map

$$d_{\pi_A} \colon \mathrm{NS}(A) \xrightarrow{\widetilde{\mathrm{AJ}}^* \circ \pi_A^*} \mathrm{Pic}(X) \xrightarrow{\deg} \mathbb{Z} \tag{3}$$

and therefore a map

$$\theta_{X,\pi_A,\pi_B} \colon \operatorname{Ker} d_{\pi_A} \xrightarrow{\widetilde{\operatorname{AJ}}^* \circ \pi_A^*} \operatorname{Pic}^0(X) \longrightarrow J(\mathbb{Q}) \xrightarrow{\pi_B \otimes \mathbb{Q}} B(\mathbb{Q}) \otimes \mathbb{Q}, \qquad (4)$$

which is called the *Chow-Heegner construction* (see Definition 3 for details).

Remark 2 As an alternative definition (useful for the proofs), for any correspondence $Z \subset X \times X$, we can associate a cycle $D_Z(b) \in \operatorname{Pic}^0(X)$ (see (16)), and this defines a homomorphism $\operatorname{NS}(X \times X) \to \operatorname{Pic}^0(X)$ so that the composition

$$\operatorname{NS}(J) \xrightarrow{(\operatorname{AJ}^{(2)})^*} \operatorname{NS}(X \times X) \longrightarrow \operatorname{Pic}^0(X),$$

where $\operatorname{AJ}^{(2)} \colon X \times X \to J$ is defined by $(x, y) \mapsto [x] + [y] - 2[b]$, is equal to $\widetilde{\operatorname{AJ}}^*$ on $(\widetilde{\operatorname{AJ}}^*)^{-1}(\operatorname{Pic}^0(X))$, which then allows us to retrieve θ_{X,π_A,π_B} on cycles Z coming from $\operatorname{Ker} d_{\pi_A}$.

The 'quadratic Chabauty for quotients' result that we prove in this paper says that we can replace J with A, but the price we pay is that we replace $\rho(J) - 1$ with the rank of $\operatorname{Ker}(\theta_{X,\pi_A,\pi_B})$, which can be smaller than $\rho(A) - 1$.

Proposition 1 *Let X be a curve as above. Suppose J admits an isogeny $(\pi_A, \pi_B) \colon J \to A \times B$, where $\operatorname{Hom}(A, B) = 0$. If*

$$\operatorname{rk}(A) < \dim(A) + \operatorname{rk}(\operatorname{Ker}(\theta_{X,\pi_A,\pi_B})), \qquad (C)$$

then $X(\mathbb{Q}_p)_2$ is finite.

In the case where $\operatorname{rk}(A) = \dim(A)$ which we will focus on, we can simplify this condition in terms of *nice correspondences*, defined in Sect. 2.1. More precisely, (π_A, π_B) induces an isomorphism $\operatorname{End}^0(J) \cong \operatorname{End}^0(A) \times \operatorname{End}^0(B)$, and $X(\mathbb{Q}_p)_2$ is finite whenever there exists a nontrivial nice correspondence Z on $X \times X$ whose corresponding endomorphism of J is zero in $\operatorname{End}^0(B)$, and whose corresponding Chow–Heegner point $D_Z(b) \in \operatorname{Pic}^0(X)$ is torsion when projected to B.

Remark 3 Note that, since $\operatorname{rk}(\operatorname{Ker}(\theta_{X,\pi_A,\pi_B})) \le \rho(A) - 1$, inequality (C) implies that A satisfies the naive analogue of Lemma 1

$$\operatorname{rk}(A) < \dim(A) + \rho(A) - 1. \qquad (5)$$

However, in general (C) is strictly stronger than (5). In fact, the trivial lower bound on $\operatorname{rk}(\operatorname{Ker}(\theta_{X,\pi_A,\pi_B})$ is $\rho(A) - 1 - \operatorname{rk}(B)$ and if the latter was positive, it would imply (2). This is why Proposition 2 looks quite particular to modular curves. Moreover, understanding the rank of $\operatorname{Ker}(\theta_{X,\pi_A,\pi_B})$ in general seems somewhat subtle - as becomes apparent in Example 1 and Sect. 3.2, this quantity is not an invariant of the pair (A, B), or even of the triple (X, A, B), and does not seem to behave so well functorially even under quite strong hypotheses. Finally, as explained in the first appendix, this quantity is also related to the Gross–Kudla–Schoen cycles constructed in [31].

The following proposition emphasises that in fact, the supplementary condition (C) can always be satisfied for our modular curves.

Proposition 2 *Let $X = X_0^+(N)$ or $X_{ns}^+(N)$, and $J = \text{Jac}(X)$. Assume Theorem 2 holds, and the genus of X is at least two. Then J admits an isogeny $(\pi_A, \pi_B) \colon J \to A \times B$ satisfying*

1. $\text{rk}(A) = \dim A \geq 2$.
2. $\rho(A) > 1$.
3. $\text{rk}(\text{Ker}(\theta_{X,\pi_A,\pi_B})) = \rho(A) - 1$.

As will become apparent in the proof, in fact we take A to be the *maximal* isogeny factor of J whose analytic rank is equal to its dimension and B its complement, otherwise we might not be able to ensure that the kernel of θ_{X,π_A,π_B} is nontrivial. This idea relies heavily on the use of (traces of) Heegner points on the modular curves $X_0(N), X_{ns}(N)$, which generate $A(\mathbb{Q})$ up to finite index, but will automatically be torsion in $B(\mathbb{Q})$, both situations being ultimately by-products of the generalised Gross–Zagier formula (see Sect. 4.2). Note that in this case the kernel of the theta morphism is not only nontrivial, but as large as it can be, which might indicate a deeper phenomenon at play.

The structure of the paper is as follows. In Sect. 2, we give some reminders on Néron–Severi groups, Chow groups and correspondences, and describe the map θ_{X,π_A,π_B} in terms of cycles. In Sect. 3 we prove Proposition 1. In Sect. 4, we prove Proposition 2 assuming Theorem 2, after some discussion on (C), and using generalised Gross–Zagier formulas. In Sect. 5, we prove Theorem 2. Finally, for sake of clarity and by lack of easily available references in the literature, we gather in Appendix 6 results about the Chow–Heegner construction above and explain in Appendix 7 the proof of the Kolyvagin–Logachev type result needed to translate Theorem 2 into an algebraic rank result.

1.2 Notation and conventions

Unless stated otherwise, we adopt the following conventions in this paper.

- X is a smooth projective geometrically irreducible curve of genus ≥ 2 over $\mathbb{Q}$. J is the Jacobian of X and $\text{AJ} \colon X \to J$ is the Albanese morphism with a fixed base point $b \in X(\mathbb{Q})$. The notation $\widetilde{\text{AJ}}^*$ refers to twice the pullback on symmetric line bundles of X to $\text{Pic}(X)$ (see (13)), and then factors through $\text{NS}(J)$ (this is not the same as just the pullback AJ^* from $\text{Pic}(J)$ to $\text{Pic}(X)$, which does not vanish on $\text{Pic}^0(J)$).
- For any n and any $S \subset \{1, \ldots, n\}$, the morphism

$$i_S(b) \colon X \to X^n \tag{6}$$

is defined so that the j-th coordinate of $i_S(b)(x)$ is x if $j \in S$ and b otherwise. When there is no ambiguity on b we denote it simply by i_S. Similarly, the morphism

$$\pi_S \colon X^n \to X^{\#S} \tag{7}$$

denotes the projection of $(x_1, \ldots, x_n)$ on the coordinates belonging to S.

● Morphisms between algebraic varieties over $\mathbb{Q}$ and their structures (line bundles, divisors, etc) are assumed to be defined over $\mathbb{Q}$.

● For a smooth projective algebraic variety Y over $\mathbb{Q}$, $\mathrm{NS}(Y)$ is the Néron-Severi group of Y, and $\rho(Y) := \mathrm{rk}\,\mathrm{NS}(J)$ is the Picard number of J (see §2.1).

● For any abelian variety A over $\mathbb{Q}$ (in particular for J), $\mathrm{rk}(A)$ is the rank of the finite type $\mathbb{Z}$-module $A(\mathbb{Q})$ and $\mathrm{End}^0(A) := (\mathrm{End}_{\mathbb{Q}} A) \otimes \mathbb{Q}$.

● N is a prime number (the level of our modular curves) and $M = N$ or N^2.

● $X_0(N)$ (resp. $X_s^+(N)$, $X_{\mathrm{ns}}^+(N)$) is the modular curve quotient of $X(N)$ corresponding to the Borel structure (resp. normaliser of split Cartan, normaliser of nonsplit Cartan), $X_0^+(N)$ is the quotient of $X_0(N)$ by the Atkin-Lehner w_N. Accordingly, the respective jacobians of these modular curves are denoted respectively by $J_0(N)$, $J_s^+(N)$, $J_{\mathrm{ns}}^+(N)$, $J_0^+(N)$ (see Sect. 4).

● For X a variety over a field $K \subset \mathbb{C}$, $H^k(X, \mathbb{Z})$ refers to the singular cohomology of $X(\mathbb{C})$.

● Given a unipotent group U, the central series filtration of U is defined by $U^{(1)} = U$ and $U^{(i+1)} = [U, U^{(i)}]$, and $\mathrm{gr}_i(U) := U^{(i)}/U^{(i+1)}$ (in particular $\mathrm{gr}_1(U) = U^{\mathrm{ab}}$). If a group G acts continuously on U, then G acts on the set of normal subgroups of U, and we say that a quotient U/H is G-stable if the normal subgroup H is stabilised by G. In this case there is a unique G-action on U/H making the surjection G-equivariant.

● The letter p denotes a prime number different from N which will be used (except in Appendix 7) only in the context of p-adic numbers.

2 The quadratic Chabauty condition (C) for a quotient

2.1 Reminders on Chow groups and Néron–Severi groups

We recall here the basic notions on correspondences of curves, and the Chow groups and Néron–Severi groups that we need. A good reference on correspondences is Smith's thesis [61, Chapter 3], and classical ones are [13, section 11.5] for the complex case and [26, Chapter 16] for the general case.

Definition 1 For any geometrically smooth and irreducible projective variety Y over $\mathbb{Q}$ and any $k \leq \dim Y$:

– The *Chow group* $\mathrm{CH}^k(Y)$ is the group of cycles of Y of codimension k up to rational equivalence.
– $c_k : \mathrm{CH}^k(Y) \to H^{2k}(Y, \mathbb{Z})$ is the *cycle map*, and $\mathrm{CH}_0^k(Y) := \mathrm{Ker}(c_k)$ is its subgroup of homologically trivial cycles (in $Y(\mathbb{C})$).

In particular, there are canonical isomorphisms

$$\mathrm{CH}^1(Y) \cong \mathrm{Pic}(Y), \quad \mathrm{CH}_0^1(Y) \cong \mathrm{Pic}^0(Y).$$

The Néron-Severi group $\mathrm{NS}(Y) := \mathrm{Pic}(Y)/\mathrm{Pic}^0(Y)$ is thus embedded in $H^2(Y(\mathbb{C}), \mathbb{Z})$.

We can also define a geometric étale cycle map [21, Cycle]

$$c_k^{l,\text{ét}} \colon \mathrm{CH}^k(Y) \to H_{\text{ét}}^{2k}(Y_{\overline{\mathbb{Q}}}, \mathbb{Z}_l(k))$$

and an absolute étale cycle map

$$c_k^{\text{abs}} \colon \mathrm{CH}^k(Y) \to H_{\text{ét}}^{2k}(Y, \mathbb{Z}_l(k)).$$

By the Artin comparison theorem we have $\mathrm{Ker}(\prod_l c_k^{l,\text{ét}}) = \mathrm{CH}_0^k(Y)$. The étale Abel–Jacobi morphism is a homomorphism

$$\mathrm{AJ}_{\text{ét}} \colon \mathrm{CH}_0^k(Y) \to \mathrm{Ext}_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}^1(\mathbb{Q}_p, H_{\text{ét}}^{2k-1}(Y_{\overline{\mathbb{Q}}}, \mathbb{Q}_p(k)))$$

which may be defined using the Leray spectral sequence or (equivalently but more directly) by realising the extension class of a homologically trivial cycle Z inside $H^{2k-1}((X - Z)_{\overline{\mathbb{Q}}}, \mathbb{Q}_p(k))$ (see Jannsen [38, II.9] or Nekovar [53, 5.1]). By Poincaré duality, we may equivalently think of the target of $\mathrm{AJ}_{\text{ét}}$ as being

$$\mathrm{Ext}_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}^1(H_{\text{ét}}^{2(d-k)+1}(Y_{\overline{\mathbb{Q}}}, \mathbb{Q}_p(d)), \mathbb{Q}_p(k)) \quad (d = \dim Y).$$

In particular, when $Y = X$ is a curve, and for $k = 1$, the target of $\mathrm{AJ}_{\text{ét}}$ is

$$\mathrm{Ext}_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}^1(V_p(J), \mathbb{Q}_p(1)),$$

where J is the Jacobian of X and $V_p(J) = T_p(J) \otimes_{\mathbb{Z}_p} \mathbb{Q}_p$.

Let us now review the basic definitions of correspondences.

Definition 2 For two curves X_1, X_2 as before:

- A *correspondence* Z on X_1, X_2 is a divisor of $\mathrm{Div}(X_1 \times X_2)$, *prime* if the underlying divisor is. It is called *fibral* if its prime components are horizontal or vertical divisors.
- If Z is a nonfibral prime correspondence, the two projections $\pi_{1,Z}, \pi_{2,Z} \colon Z \to X_1$, X_2 are nonconstant so $\psi_Z := (\pi_{2,Z})_* \circ \pi_{1,Z}^*$ defines a morphism from $\mathrm{Div}(X_1)$ to $\mathrm{Div}(X_2)$, inducing a morphism between the Jacobians of X_1 and X_2, and two rationally equivalent divisors define the same morphism. This defines by linearity (extending to 0 for fibral prime divisors) a surjective morphism

$$\psi \colon \mathrm{Pic}(X_1 \times X_2) \to \mathrm{Hom}(\mathrm{Jac}(X_1), \mathrm{Jac}(X_2)), \tag{8}$$

with kernel $\pi_1^* \mathrm{Pic}(X_1) \oplus \pi_2^* \mathrm{Pic}(X_2)$ with notation (7) ([13, Theorem 11.5.1] or [61, Theorem 3.3.12]).

When $X = X_1 = X_2$, with the choice of a base point b, using notation from (6) and (7), we obtain from $\pi_1 \circ i_1 = \mathrm{Id}_X$ and similar relations the identities

$$\mathrm{Pic}(X \times X) = \pi_1^* \mathrm{Pic}(X) \oplus \pi_2^* \mathrm{Pic}(X) \oplus \mathrm{Ker}(i_1^* \oplus i_2^*) \tag{9}$$

$$\mathrm{Pic}^0(X \times X) = \pi_1^* \mathrm{Pic}^0(X) \oplus \pi_2^* \mathrm{Pic}^0(X), \tag{10}$$

(see [61, Proposition 3.3.8], as homologically trivial cycles are homomorphically trivial) which induces a decomposition

$$\mathrm{NS}(X \times X) = \pi_1^* \mathrm{NS}(X) \oplus \pi_2^* \mathrm{NS}(X) \oplus \mathrm{Ker}(i_1^* \oplus i_2^*), \qquad (11)$$

where the last direct factor then canonically identifies with $\mathrm{End}(J)$ via (8). By abuse of notation, we thus denote

$$\psi^{-1} \colon \mathrm{End}(J) \overset{\cong}{\to} \mathrm{Ker}(i_1^* \oplus i_2^*)$$

the inverse of this isomorphism. Now, the morphism $i_{1,2}^* - i_1^* - i_2^*$ is trivial when restricted to $\mathrm{Pic}^0(X \times X)$, hence induces a morphism

$$\varphi \colon \mathrm{NS}(X \times X) \to \mathrm{Pic}(X). \qquad (12)$$

Define

$$\mathrm{AJ}^{(2)} \colon \left| \begin{array}{l} X \times X \longrightarrow J \\ (x, y) \longmapsto [x] + [y] - 2[b] \end{array} \right., \quad \widetilde{\mathrm{AJ}}^* := \varphi \circ (\mathrm{AJ}^{(2)})^*.$$

We have $\widetilde{\mathrm{AJ}}^* = [2]^* \circ \mathrm{AJ}^* - 2\mathrm{AJ}^*$ so for $[\mathcal{L}] \in \mathrm{Pic}(J)$,

$$\widetilde{\mathrm{AJ}}^*([\mathcal{L}]) = \mathrm{AJ}^*([\mathcal{L}]) + \mathrm{AJ}^*([-1]^*[\mathcal{L}]). \qquad (13)$$

using the classical identity $[n]^*(\mathcal{L}) \simeq \mathcal{L}^{\otimes(\frac{n^2+n}{2})} \otimes [-1]^*(\mathcal{L}^{\otimes(\frac{n^2-n}{2})})$. In particular, $\widetilde{\mathrm{AJ}}^*$ is twice the usual pullback by AJ on symmetric line bundles.

For any divisor D of $X \times X$, the degree of $\varphi(D)$ is equal to the rational trace of $\psi(D)$ ([13, Proposition 11.5.2]). This induces a morphism

$$\widetilde{\theta}_{X,b} \colon \mathrm{End}(J)^{\mathrm{tr}=0} \overset{\varphi \circ \psi^{-1}}{\longrightarrow} \mathrm{Pic}^0(X).$$

By [52, IV.20], the rule $\mathcal{L} \mapsto \lambda_{\mathcal{L}}$ defined by $\lambda_{\mathcal{L}}(P) = T_P^* \mathcal{L} \otimes \mathcal{L}^{-1} \in \mathrm{Pic}^0(J)$ induces an isomorphism

$$\widetilde{\lambda} \colon \left| \begin{array}{l} \mathrm{NS}(J) \longrightarrow \mathrm{End}^\dagger(J) \\ [\mathcal{L}] \longmapsto \mathcal{P}^{-1} \circ \lambda_{\mathcal{L}} \end{array} \right. \qquad (14)$$

where $\mathcal{P} \colon J \overset{\cong}{\to} \widehat{J}$ is a natural principal polarisation given by a theta divisor. This the same as applying the composition $-\psi \circ (\mathrm{AJ}^{(2)})^*$. Indeed, via the natural morphisms $\widehat{J} \cong \mathrm{Pic}^0(J)$ and $\mathrm{Pic}^0(X) \cong J$, the inverse $\widehat{J} \to J$ of the principal polarisation given by a theta divisor on J is equal to $-\mathrm{AJ}^*$ from $\mathrm{Pic}^0(J)$ to $\mathrm{Pic}^0(X)$ [13, Proposition 11.3.5].

Now, in terms of line bundles, by definition, given a line bundle L on $X \times X$, the endomorphism of $\mathrm{Pic}(X)$ associated to it is given on points by $x \mapsto i_2^*(x)(L)$ with

notation (6). As $(\text{AJ}^{(2)} \circ i_2(x)) = T_{[x]-[b]} \circ \text{AJ}$, for a line bundle $\mathcal{L}$ on $\text{Pic}(J)$ and x, y points of X the endomorphism associated to $L = (\text{AJ}^{(2)})^* \mathcal{L}$ sends $[x] - [y]$ to

$$\text{AJ}^*(T_{[x]-[b]}^* \mathcal{L} - T_{[y]-[b]}^* \mathcal{L}) = \text{AJ}^*(T_{[x]-[y]}^* \mathcal{L} - \mathcal{L}) = \text{AJ}^* \lambda_{\mathcal{L}}([x] - [y]),$$

which gives the equality up to -1. Hence, if we define

$$\text{NS}(J)^0 := \text{Ker}(\text{NS}(J) \xrightarrow{\text{deg}} \text{NS}(X)),$$

and

$$\theta_{X,b} := \widetilde{\text{AJ}}^*_{|\text{NS}(J)^0} \colon \text{NS}(J)^0 \to \text{Pic}^0(X)$$

then we have the following commutative diagram to sum up all the previous properties. Every symbol $\circlearrowleft$ means that the diagram around it commutes, and every $\circlearrowleft_-$ means that one composition is equal to -1 times the other. Dashed arrows indicate that the morphisms are only defined on part of the domain or with small codomain, but in each case, it admits a natural extension. By abuse of notation, ψ and $(\text{AJ}^{(2)})^*$ are used both on Picard groups and Néron–Séveri groups.

Remark 4 In [8], an element of $\text{Pic}(X \times X)$ whose image under ψ lies in $\text{End}^\dagger(J)^{\text{tr}=0}$ is referred to as a 'nice correspondence'.

2.2 Chow–Heegner points and diagonal cycles

We recall an equivalent version of the morphism $\widetilde{\theta}_{X,b}$, which appears in [18] and [6]. As our discussion applies in fairly broad generality, we take X to be a smooth geometrically irreducible projective curve over a field K of characteristic zero. Fix $b \in X(K)$, and $S \subset \{1, \ldots n\}$, let X_S denote the image of X under the closed immersion $i_S(b)$ defined in (6). For any $Z \in \mathrm{Div}(X \times X)$, let $C_Z(b) := (i^*_{\{1,2\}}(b) - i^*_{\{1\}}(b) - i^*_{\{2\}}(b))(Z) = \varphi([Z])$ and

$$D_Z(b) := C_Z(b) - \deg(C_Z(b)) \cdot b \in \mathrm{Pic}^0(X). \tag{16}$$

We refer to $D_Z(b)$ and $C_Z(b)$ as *Chow–Heegner* points, following [19].

The map $Z \mapsto D_Z(b)$ factors through $\mathrm{Pic}(X \times X)$, and has the following relation to $\widetilde{\theta}_{X,b}$. The projection

$$\Pi : \mathrm{Pic}(X \times X) \to \mathrm{Ker}(i^*_1 \oplus i^*_2)$$

associated to (9) is given by $(1 - \pi^*_1 \circ i^*_1 - \pi^*_2 \circ i^*_2)$, giving the identities

$$\varphi \circ \Pi = i^*_{\{1,2\}} \circ \Pi = i^*_{\{1,2\}} - i^*_1 - i^*_2, \quad \psi^{-1} \circ \psi = \Pi.$$

Since $\deg(C_Z(b)) = \deg(\varphi(\Pi([Z])))$, for any Z in $\mathrm{Pic}(X \times X)$ which lies in the kernel of $\deg \varphi$, we have

$$D_Z(b) = C_Z(b) = \varphi([Z]) = \varphi(\Pi([Z])) = \tilde{\theta}_{X,b}(\psi([Z])). \tag{17}$$

These computations also prove the claims of Remark 2 using the diagram (15). We define $Z^t \in \mathrm{CH}^1(X \times X)$ to be the pull-back of Z under the involution

$$X \times X \to X \times X$$
$$(x, y) \mapsto (y, x).$$

Lemma 2 *In the notation of Definition 2, we have*

$$D_Z(b') - D_Z(b) = \psi_Z(b - b') + \psi_{Z^t}(b - b').$$

Proof We have $i_{\{1,2\}}(b) = i_{\{1,2\}}(b')$. Hence

$$C_Z(b') - C_Z(b) = i_{\{1\}}(b)^*(Z) - i_{\{1\}}(b')^*(Z) + i_{\{2\}}(b)^*(Z) - i_{\{2\}}(b')^*(Z).$$

By definition of the correspondences, we then have

$$(i_{\{1\}}(b)^* - i_{\{1\}}(b')^*)(Z) = \psi_Z(b - b')$$

and

$$(i_{\{2\}}(b)^* - i_{\{2\}}(b')^*)(Z) = \psi_{Z'}(b - b'),$$

which proves the equality for $C_Z(b') - C_Z(b)$, thus for $D_Z(b') - D_Z(b)$ as the degrees are then equal. $\qquad\square$

Definition 3 Given a surjective homomorphism $\pi_B \colon J \to B$ of abelian varieties, we obtain a homomorphism

$$\mathrm{Ker}(\mathrm{NS}(J) \xrightarrow{\deg \circ \widetilde{\mathrm{AJ}}^*} \mathbb{Z}) \xrightarrow{\widetilde{\mathrm{AJ}}^*} \mathrm{Pic}^0(X) \longrightarrow J \xrightarrow{\pi_B} B. \tag{18}$$

By Lemma 2 and (17), for a divisor Z on $X \times X$, if $\psi_{\Pi(Z)}$ has image contained in $\mathrm{Ker}(\pi_B)$, then the image of $[Z]$ in B via (18) is independent of the choice of basepoint. In particular, if we have a surjection $(\pi_A, \pi_B) \colon J \to A \times B$, and $\mathrm{Hom}(A, B) = 0$, then we obtain a homomorphism independent of b, which we will denote by

$$\theta_{X,\pi_A,\pi_B} \colon \left| \begin{array}{ll} \mathrm{Ker}(d_{\pi_A}) \subset \mathrm{NS}(A) & \longrightarrow B \\ [\mathcal{L}] & \longmapsto \pi_B \circ \theta_{X,b} \circ \pi_A^*([\mathcal{L}]) \end{array} \right. .$$

Remark 5 This construction also has a direct description in terms of line bundles, although this is not the one we use to calculate θ_{X,π_A,π_B} in examples. Given a line bundle $\mathcal{L}_A$ on A whose pull-back to X via $\mathrm{AJ}^* \circ \pi_A^*$ has degree zero, we may also consider the projection of $\mathrm{AJ}^* \circ \pi_A^*(\mathcal{L}_A)$ to B. Variants of this construction are studied in the thesis of Michael Daub [20] when $\mathrm{Hom}(A, B) = 0$. By (13) and because $\mathrm{Pic}^0(J)$ contains all classes of antisymmetric line bundles, we have the identity [20, Proposition 3.3.3]

$$\theta_{X,\pi_A,\pi_B} \circ p = [2] \circ \pi_B \circ \mathrm{AJ}^* \circ \pi_A^*;$$

where p is the projection from $\mathrm{Pic}(A)$ to $\mathrm{NS}(A)$ restricted to $p^{-1}(\mathrm{Ker}(d_{\pi_A}))$. In particular, the right-hand side does vanish on $\mathrm{Pic}^0(A)$ [20, Proposition 3.3.2].

Example 1 Note that θ_{X,π_A,π_B} is not an invariant of A and B, or even of X, A, B. For example, let A and B be distinct isogeny factors of $X_0(N)$, and let $X = X_0(N^2)$. Let $f_1, f_2 \colon X \to X_0(N)$ be the two natural morphisms, and let (π_{A_i}, π_{B_i}) be the morphisms $\mathrm{Jac}(X) \to A \times B$ obtained by composing the surjection $J_0(N) \to A \times B$ with f_{i*}. Then $\theta_{X,\pi_{A,i},\pi_{B,i}}$ can be nonzero (see [18] for examples), however if $i \neq j$, $\theta_{X,\pi_{A,i},\pi_{B,j}}$ is identically zero, since for any choice of line bundle $[\mathcal{L}]$ in $\mathrm{NS}(A)$, the associated point $D_{[\mathcal{L}]}(b)$ will lie in $f_i^* J_0(N)$, hence the projection to $f_{j*} J_0(N)$ will be torsion.

3 Proof of finiteness of the Chabauty–Kim set under (C)

The strategy of proof of Proposition 1 is very similar to that of [6, Lemma 3.2]. To explain this strategy, we need to establish some notation. X, A, B are as in the proposition. Define

$$V := T_p(J) \otimes \mathbb{Q}_p, \quad V_A := T_p(A) \otimes \mathbb{Q}_p, \quad V_B := T_p(B) \otimes \mathbb{Q}_p.$$

Let $U_n(b)$ denote the maximal n-unipotent quotient of the $\mathbb{Q}_p$-unipotent fundamental group of $\overline{X}$ at some basepoint b as defined in [22, §10]. Let U be a Galois-stable quotient of $U_n(b)$ (i.e. a quotient by a Galois-stable normal subgroup of $U_n(b)$). Let T_0 be the set of primes of bad reduction for X, and let $T = T_0 \cup \{p\}$. Denote the maximal quotient of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ unramified outside T by $G_{\mathbb{Q},T}$, and for $v \in T$ denote $\mathrm{Gal}(\overline{\mathbb{Q}}_v/\mathbb{Q}_v)$ by $G_{\mathbb{Q}_v}$. Then by [39,40], we have a commutative diagram

$$
\begin{array}{ccc}
X(\mathbb{Q}) & \xrightarrow{\quad j \quad} & H^1(G_{\mathbb{Q},T}, U) \\
\downarrow & & \downarrow{\scriptstyle \prod_{v \in T} \mathrm{loc}_v} \\
\prod_{v \in T} X(\mathbb{Q}_v) & \xrightarrow{\prod_{v \in T} j_v} & \prod_v H^1(G_{\mathbb{Q}_v}, U),
\end{array}
$$

with the following properties.

1. For $G = G_{\mathbb{Q},T}$ or $G_{\mathbb{Q}_v}$, and all $i < k$, the sets $H^1(G, U^{(i)}/U^{(k)})$ have the structure of $\mathbb{Q}_p$ points of an algebraic variety, so that the algebraic structure on $H^1(G, \mathrm{gr}_i U)$ is just the usual scheme structure on a vector space, and the maps

$$H^1(G, \mathrm{gr}_i U) \to H^1(G, U/U^{(i+1)}) \to H^1(G, U/U^{(i)})$$

come from morphisms of algebraic varieties. The maps loc_v are then algebraic for these structures.
2. For $v \in T_0$, the map j_v has finite image.
3. The image of the map j_p is contained inside the subvariety $H^1_f(G_{\mathbb{Q}_p}, U)$ of crystalline torsors.

The following Lemma is proved in [6, Lemma 3.1] (although the result is stated only in the case $A = J$, the proof generalises to the case where A is an arbitrary quotient of J).

Lemma 3 *Let U be a Galois-stable quotient of $U_2(b)$. Suppose U is an extension of V_A by $\mathbb{Q}_p(1)^n$, where A is some abelian variety over $\mathbb{Q}$ and $V_A = T_p(A) \otimes \mathbb{Q}_p$. If*

$$\mathrm{rk}(A(\mathbb{Q})) < n + \dim(A),$$

then $X(\mathbb{Q}_p)_2$ is finite. In particular, if $\mathrm{rk}(A(\mathbb{Q})) = \dim(A)$, then $X(\mathbb{Q}_p)_2$ is finite whenenever $n > 0$.

To prove Proposition 1, we construct a quotient U of $U_2(b)$ as in Lemma 3, with $n = \mathrm{rk}(\mathrm{Ker}\,\theta_{X,\pi_A,\pi_B})$. We again take X to be a smooth projective geometrically irreducible curve over a field K of characteristic zero.

The group $U_2(b)$ is an extension

$$1 \to \mathrm{Ker}(H^2(J_{\overline{\mathbb{Q}}}, \mathbb{Q}_p) \xrightarrow{\mathrm{AJ}^*} H^2(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_p))^* \to U_2(b) \to V \to 1. \qquad (19)$$

Hence for any $\xi \in \mathrm{Ker}(\mathrm{NS}(J) \overset{\widetilde{\mathrm{AJ}}^*}{\to} \mathrm{NS}(X))$, we may quotient by the kernel of the dual of the Chern class $c_p^{\text{ét}}(\xi) \in H^2(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_p(1))$ (see Sect. 1.1)

$$c_p^{\text{ét}}(\xi)^*(1) : \mathrm{Ker}(H^2(J_{\overline{\mathbb{Q}}}, \mathbb{Q}_p) \xrightarrow{\mathrm{AJ}^*} H^2(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_p))^* \to \mathbb{Q}_p(1)$$

to obtain a quotient U_Z of $U_2(b)$ which is an extension of V by $\mathbb{Q}_p(1)$. Similarly, for any nice correspondence on $X \times X$, we obtain a quotient of $U_2(b)$ which is an extension of V by $\mathbb{Q}_p(1)$.

Lemma 4 ([6],Theorem 6.3) *Let U be a Galois-stable quotient of $U_2(b)$ of the form*

$$1 \to \mathbb{Q}_p(1) \to U \to V_p(J) \to 1,$$

coming from a correspondence $Z \subset X \times X$ as above. Then the associated extension class of $\mathrm{Lie}(U)$ in $\mathrm{Ext}^1_{G_K}(V_p(J), \mathbb{Q}_p(1))$ is equal to the étale Abel–Jacobi class of the cycle $D_Z(b)$ (see Sect. 2.1).

Proof Let $\mathcal{E}(\mathrm{Lie}(U))$ be the universal enveloping algebra of $\mathrm{Lie}(U)$, and let $I(\mathrm{Lie}(U))$ be the kernel of the co-unit morphism $\mathcal{E}(\mathrm{Lie}(U)) \to \mathbb{Q}_p$. In [6, §6], a Galois representation E_Z is constructed as a quotient of $\mathcal{E}(\mathrm{Lie}(U))$. The image of $I(\mathrm{Lie}(U))$ in E_Z is an extension IE_Z of V by $\mathbb{Q}_p(1)$. By [6, Theorem 6.3], the extension class of IE_Z in $\mathrm{Ext}^1_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}(V_p(J), \mathbb{Q}_p(1))$ is the Abel–Jacobi class of $D_Z(b)$. The restriction of $I(\mathrm{Lie}(U)) \to IE_Z$ to $\mathrm{Lie}(U) \subset I(\mathrm{Lie}(U))$ is an isomorphism, and hence the extension class of $\mathrm{Lie}(U)$ is isomorphic to $D_Z(b)$. □

As explained in Appendix 6, Lemma 4 is really a consequence of Hain and Matsumoto's computation of the extension class of $\mathrm{Lie}(U_2)$ in terms of the Ceresa cycle. Hence to complete the proof of Proposition 1, it will be enough to prove the following Lemma.

Lemma 5 *Let U' denote the quotient of U_2 obtained from the surjection $\mathrm{gr}_2(U_2) \to \mathrm{Ker}(d_{\pi_A})^* \otimes \mathbb{Q}_p(1)$. There exists a Galois stable quotient U of U' which is an extension of V_A by $\mathrm{Ker}(\theta_{X,\pi_A,\pi_B})$:*

$$
\begin{array}{ccccccccc}
1 & \longrightarrow & \mathrm{Ker}(d_{\pi_A})^* \otimes \mathbb{Q}_p(1) & \longrightarrow & U' & \longrightarrow & V_A \oplus V_B & \longrightarrow & 1 \\
& & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})^* \otimes \mathbb{Q}_p(1) & \longrightarrow & U & \longrightarrow & V_A & \longrightarrow & 1.
\end{array}
$$

Proof It will be enough to prove the corresponding statement for the Lie algebra L' of U'. The commutator map

$$[\cdot, \cdot]_{U'} : (V_A \oplus V_B) \times (V_A \oplus V_B) \to \mathrm{Ker}(d_{\pi_A})^* \otimes \mathbb{Q}_p(1)$$

is the composite of the commutator on U_2, given by

$$(V_A \oplus V_B) \times (V_A \oplus V_B) \to \mathrm{Coker}(\mathbb{Q}_p(1) \xrightarrow{\cup^*} \wedge^2 V_A \oplus V_A \otimes V_B \oplus \wedge^2 V_B)$$

with the surjection

$$\mathrm{Coker}(\mathbb{Q}_p(1) \xrightarrow{\cup^*} \wedge^2 V_A \oplus V_A \otimes V_B \oplus \wedge^2 V_B) \to \mathrm{Ker}(d_{\pi_A})^* \otimes \mathbb{Q}_p(1)$$

Since the latter map factors through projection onto $\wedge^2 V_A / \mathbb{Q}_p(1)$, the composite map factors through projection onto $V_A \times V_A$. Hence for any quotient Q of $\mathrm{Ker}(d_{\pi_A})^* \otimes \mathbb{Q}_p(1)$, we can construct a Lie algebra quotient of L' which is an extension of V_A by Q. It remains to show that, when $Q = \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})$, we can make this quotient Galois stable. That is, we first quotient out by $(\mathrm{Ker}(d_{\pi_A}) / \mathrm{Ker}(\theta_{X,\pi_A,\pi_B}))^* \otimes \mathbb{Q}_p(1)$, to form an extension

$$0 \to \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})^* \otimes \mathbb{Q}_p(1) \to L'' \to V_A \oplus V_B \to 0.$$

The surjection $L'' \to V_B$ induces a Galois equivariant short exact sequence of Lie algebras

$$0 \to L' \to L'' \to V_B \to 0,$$

and to construct the quotient $U \to U'$, it is enough to show that this short exact sequence admits a Galois equivariant section. Here L' sits in a short exact sequence

$$0 \to \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})^* \otimes \mathbb{Q}_p(1) \to L' \to V_A \to 0,$$

and since $L'' / \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})^* \otimes \mathbb{Q}_p(1) = V_A \oplus V_B$, it is enough to show that image of $[L'']$ under the composite map

$$\mathrm{Ext}^1_{G_{\mathbb{Q}}}(V_A \oplus V_B, \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})^* \otimes \mathbb{Q}_p(1)) \to \mathrm{Ext}^1_{G_{\mathbb{Q}}}(V_B, \mathrm{Ker}(\theta_{X,\pi_A,\pi_B})^* \otimes \mathbb{Q}_p(1))$$

is zero.

Equivalently, we want to show that $\mathrm{Ker}(\theta_{X,\pi_A,\pi_B})$ is contained in the kernel of the homomorphism

$$\mathrm{Ker}(d_{\pi_A}) \to \mathrm{Ext}^1_{G_{\mathbb{Q}}}(V_B, \mathbb{Q}_p(1))$$

sending $\xi \in \mathrm{Ker}(d_{\pi_A})$ to the V_B component of the extension class in $\mathrm{Ext}^1(V_A \oplus V_B, \mathbb{Q}_p(1))$ associated to the quotient of L' defined by $c_p^{\mathrm{\acute{e}t}}(\xi)$:

$$
\begin{array}{ccccccccc}
0 & \longrightarrow & \mathrm{Ker}(d_{\pi_A})^* \otimes \mathbb{Q}_p(1) & \longrightarrow & L' & \longrightarrow & V_A \oplus V_B & \longrightarrow & 0 \\
& & \downarrow{\scriptstyle c_p^{\mathrm{\acute{e}t}}(\xi)^* \otimes \mathbb{Q}_p(1)} & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathbb{Q}_p(1) & \longrightarrow & c_p^{\mathrm{\acute{e}t}}(\xi)^*(L) & \longrightarrow & V_A \oplus V_B & \longrightarrow & 0.
\end{array}
$$

By Lemma 4, this extension class is equal to the étale Abel–Jacobi class of $D_{c_p^{\mathrm{\acute{e}t}}(\xi)}(b)$, and hence its V_B component is equal to the étale Abel–Jacobi class of $\theta_{X,\pi_A,\pi_B}(c_p^{\mathrm{\acute{e}t}}(\xi))$.

Under the hypothesis, the latter is 0 so the extension class is trivial, which concludes the proof of Proposition 1. □

3.1 Bounding the number of rational points on curves satisfying (C)

Following [3], the proof of finiteness of $X(\mathbb{Q}_p)_2$ may be used to prove an explicit upper bound on $\#X(\mathbb{Q}_p)_2$. To explain this, we introduce some notation. By [41, Corollary 1], for all $v \neq p$, the size of the image of $X(\mathbb{Q}_v)$ in $H^1(G_{\mathbb{Q}_v}, U_2)$ is finite, and is equal to one for all primes of good reduction for X. Let T_0 denote the set of primes of bad reduction for X, and for $v \in T_0$ let n_v denote the size of the image of $X(\mathbb{Q}_v)$ in $H^1(G_{\mathbb{Q}_v}, U_2)$.

Corollary 1 *Suppose X satisfies the hypotheses of Proposition 1, and furthermore that the rank of $A(\mathbb{Q})$ is equal to its dimension, and the p-adic closure of A has finite index in $A(\mathbb{Q}_p)$. Let $n := \prod_{v \in T_0} n_v$. Let D be an effective divisor on X, let $Y \subset X_{\mathbb{Z}_p}$ be the complement of the support of a normal crossings divisor on Y with generic fibre D, and let $\{\omega_0, \ldots, \omega_{2g-1}\}$ be a set of differentials in $H^0(X, \Omega(D))$ forming a basis of $H^1_{dR}(X)$. Then there are $a_{ij}, a_i \in \mathbb{Q}_p$, $\eta \in H^0(X, \Omega(D))$ and $g \in H^0(X, \Omega(2D))$, and $\alpha_1, \ldots, \alpha_n$ in $\mathbb{Q}_p$, such that*

$$
X(\mathbb{Q}_p)_2 \cap Y(\mathbb{Z}_p) \subset \bigcup_{i=1}^n \{x \in Y(\mathbb{Z}_p) : \sum a_{ij} \int_b^x \omega_i \omega_j
$$
$$
+ \sum a_i \int_b^x \omega_i + \int_b^x \eta + g(x) = \alpha_i\}. \tag{20}
$$

Proof The argument is identical to the proof of [7, Proposition 6.4], however as the hypotheses are different we explain the steps. Arguing as in loc. cit, there are b_{ij}, b_i in $\mathbb{Q}_p$ such that $X(\mathbb{Q}_p)_2 \cap Y(\mathbb{Z}_p)$ is contained in the finite set of $x \in Y(\mathbb{Z}_p)$ satisfying

$$
h_p(A_Z(x)) - \sum b_{ij} \left(\int_b^x \omega_i \right) \left(\int_b^x \omega_j \right) - \sum \int_b^x \omega_i = - \sum_{v \in T_0} h(A_Z(b)^{\phi_v}),
$$

for some (ϕ_v) in $\prod_{v \in T_0} j_v(X(\mathbb{Q}_v))$. Here $A_Z(b)^{(\phi_v)}$ denotes the twist of $A_Z(b)$ by ϕ_v.

Hence we deduce (20) from the formula for $h_p(A_Z(x))$ given in [7, Lemma 6.7], and the formula

$$
\left(\int_b^x \omega_i \right) \left(\int_b^x \omega_j \right) = \int_b^x \omega_i \omega_j + \int_b^x \omega_j \omega_i.
$$

□

Corollary 2 *Suppose X satisfies the hypotheses of Proposition 1, and furthermore that the rank of $A(\mathbb{Q})$ is equal to its dimension. Then*

$$\#X(\mathbb{Q}) < \kappa_p \left(\prod_{v \in T_0} n_v \right) \#X(\mathbb{F}_p)(16g^3 + 15g^2 - 16g + 10),$$

where $\kappa_p := 1 + \frac{p-1}{p-2} \frac{1}{\log(p)}$.

Proof It is enough to prove that, for all $x_0 \in X(\mathbb{Z}_p)$, we can choose D, ω_i such that $\overline{x} := \mathrm{red}(x_0)$ lies in $Y(\mathbb{F}_p)$, and

$$\#\{x \in \mathrm{red}^{-1}(\{\overline{x}\}) \subset X(\mathbb{Q}_p) : \sum a_{ij} \int_b^x \omega_i \omega_j + \sum_i a_i \int_b^x \omega_i + \int_b^x \eta + g(x) = 0\}$$

$$< \kappa_p(16g^3 + 15g^2 - 16g + 10).$$

This follows from [3, Proposition 3.2] together with [3, §4, below Lemma 4.4.]. □

Remark 6 In [10], it is proved that the size of $j_{2,v}(X(\mathbb{Q}_v))$ can be bounded by the number of irreducible components of a regular semistable model of X over a finite extension of $\mathbb{Q}_v$. Hence using work of Edixhoven and Parent on stable models of $X_{\mathrm{ns}}^+(N)$ [23], one can use the above corollary, together with Theorem 1, to give explicit bounds on the size of $X_{\mathrm{ns}}^+(N)$ and $X_0^+(N)$.

3.2 Functoriality properties of (C)

The heart of the proof of Proposition 3 is an interpretation of diagonal cycles on $X_0(N)$ and $X_{\mathrm{ns}}(N)$ in terms of Heegner points. The following Lemma allows us to use this to deduce something about diagonal cycles on $X_0^+(N)$ and $X_{\mathrm{ns}}^+(N)$. This lemma is a special case of a theorem of Daub [20, Proposition 3.3.5].

Lemma 6 *1. Let $f : X' \to X$ be a non-constant morphism of curves over a field K. Suppose $b' \in X'(K)$ maps to $b \in X(K)$ under f, and let Z be an element of $\mathrm{CH}^1(X \times X)$. Then*

$$D_{(f,f)^*Z}(b') = f^*(D_Z(b)).$$

2. Let $f : X' \to X$ and b' be as above, and let f_* denote the induced surjection $J' := \mathrm{Jac}(X') \to J := \mathrm{Jac}(X)$. Let (π_A, π_B) be a surjective homomorphism from J to $A \times B$. Then

$$\mathrm{Ker}(\theta_{X,\pi_A,\pi_B}) = \mathrm{Ker}(\theta_{X',\pi_A \circ f_*, \pi_B \circ f_*}).$$

Proof For $* = \{1\}$, $\{2\}$ or $\{1, 2\}$, the diagram

$$
\begin{array}{ccc}
X' & \xrightarrow{\ i_*(b')\ } & X' \times X' \\
\Big\downarrow{\scriptstyle f} & & \Big\downarrow{\scriptstyle (f,\,f)} \\
X & \xrightarrow[\ i_*(b)\ ]{} & X \times X
\end{array}
$$

commutes. Hence we obtain, in $\mathrm{CH}^1(X')$,

$$
\begin{aligned}
f^*(C_Z(b)) &= (f^* \circ i_{\{1,2\}}(b)^* - f^* \circ i_{\{1\}}(b)^* - f^* \circ i_{\{2\}}(b)^*)(Z) \\
&= (i_{\{1,2\}}(b')^* \circ (f,f)^* - i_{\{1\}}(b')^* \circ (f,f)^* - i_{\{2\}}(b)^* \circ (f,f)^*)(Z) \\
&= C_{(f,f)^*(Z)}(b')
\end{aligned}
$$

and the result follows for $D_Z(b)$. The second item follows from the first, as we now prove. Let $\mathcal{L}$ be a line bundle on A belonging to $\mathrm{Ker}\, d_{\pi_A}$. By definition of θ_{X,π_A,π_B} and the right part of diagram (15), we fix some cycle Z on $X \times X$ such that $[Z] = (\mathrm{AJ}^{(2)})^* \circ \pi_A^*([\mathcal{L}])$, and then by (17)

$$
\theta_{X,\pi_A,\pi_B}([\mathcal{L}]) = (\pi_B \otimes \mathbb{Q}) \circ \varphi([Z]) = (\pi_B \otimes \mathbb{Q})(D_Z(b)).
$$

Now, considering the morphism $f : X' \to X$ with those choices of base points, we have $f_* \circ \mathrm{AJ}_{X'}^{(2)} = \mathrm{AJ}_X^{(2)} \circ (f,f)$. Consequently, with the same $\mathcal{L}$ and Z, $[(f,f)^*Z] = (\mathrm{AJ}_{X'}^{(2)})^* \circ (\pi_A \circ f_*)^*([\mathcal{L}])$, so the kernels of d_{π_A} and $d_{\pi_A \circ f_*}$ are the same, and on this common kernel,

$$
\begin{aligned}
\theta_{X',\pi_A \circ f_*, \pi_B \circ f_*}([\mathcal{L}]) &= ((\pi_B \circ f_*) \otimes \mathbb{Q})(D_{(f,f)^*Z}(b)) \\
&= ((\pi_B \circ f_*) \otimes \mathbb{Q})(f^*(D_Z(b))) \\
&= [\deg f](\pi_B \otimes \mathbb{Q})(D_Z(b)).
\end{aligned}
$$

In particular, θ_{X,π_A,π_B} and $\theta_{X',\pi_A \circ f_*, \pi_B \circ f_*}$ have the same kernel. $\qquad\square$

Note that while the behaviour of diagonal cycles under pull-backs is tautological, their behaviour under push-forwards is not. For this reason it seems difficult to deduce statements about diagonal cycles on $X_{\mathrm{ns}}(N)$ from results on $X_{\mathrm{s}}(N)$, in spite of the explicit isogeny relating their Jacobians explained below.

4 Proof of (C) for $X_0^+(N)$ and $X_{\mathrm{ns}}^+(N)$

Given Proposition 1, it will be enough to prove Theorem 2, and the following.

Proposition 3 *Assume Theorem 2. Then, for $X = X_0^+(N)$ or $X_{\mathrm{ns}}^+(N)$ of genus at least 2, there exists an isogeny*

$$
(\pi_A, \pi_B) : J \to A \times B,
$$

where $\mathrm{rk}(A) = \dim(A) = \rho(A) \geq 2$ *and such that, for all* $\mathcal{L}$ *in* $\mathrm{Ker}(d_{\pi_A})$, $\theta_{X,\pi_A,\pi_B}(\mathcal{L}) = 0$ *is torsion (see Definition 4 for the choices of A and B).*

We recall the definitions of some of the modular curves which appear, for example, in [16]. Define $C_{\mathrm{ns}}^+(N)$, $C_{\mathrm{s}}^+(N)$ to be normalisers in $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ of fixed choices of non-split Cartan $C_{\mathrm{ns}}(N)$ and split Cartan subgroups $C_{\mathrm{s}}(N)$ of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. The (normaliser of) split and nonsplit Cartan modular curves are defined by

$$X_{\mathrm{ns}}^+(N) := X(N)/C_{\mathrm{ns}}^+(N), \quad X_{\mathrm{s}}^+(N) = X(N)/C_{\mathrm{s}}^+(N).$$

Similarly we define $X_{\mathrm{ns}}(N)$ and $X_{\mathrm{s}}(N)$ to be the quotients of $X(N)$ by $C_{\mathrm{ns}}(N)$ and $C_{\mathrm{s}}(N)$ respectively. Since $C_{\mathrm{ns}}(N)$ and $C_{\mathrm{s}}(N)$ contain the centre of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ and their determinant goes through all $(\mathbb{Z}/N\mathbb{Z})^*$, all $X_{\mathrm{ns}}(N)$, $X_{\mathrm{s}}(N)$ and their Atkin–Lehner quotients are geometrically connected and defined over $\mathbb{Q}$.

Non-cuspidal points of $X_{\mathrm{s}}(N)$ (in characteristic not dividing N) correspond to elliptic curves E together with a pair C_1, C_2 of cyclic subgroups of E of order N generating $E[N]$. We have an isomorphism

$$X_0(N^2) \simeq X_{\mathrm{s}}(N), \tag{21}$$

which sends a point $(f : E \to E')$ to (E'', C_1, C_2), where $E'' := E/(N \cdot \mathrm{Ker}(f))$, C_1 is the image of $\mathrm{Ker}(f)$ in E'', and C_2 is the image of $E[N]$ in E''.

The curve $X_{\mathrm{s}}(N)$ is naturally a degree two cover of $X_{\mathrm{s}}^+(N)$, and there is an isomorphism $X_{\mathrm{s}}^+(N) \simeq X_0^+(N^2)$ compatible with (21).

4.1 Jacobians of modular curves and the asymptotics of the quadratic Chabauty condition

We recall a formula for the Picard numbers and ranks of modular Jacobians and their quotients, due to Siksek [59]. Let $\mathcal{B}_{N^k}$ denote a normalised eigenbasis for the space of newforms in $S_2(\Gamma_0(N^k))$. Let $\mathcal{B}_{N^k}/\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ denote a choice of representatives of the orbits of $\mathcal{B}_{N^k}$ under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. We denote by $\mathcal{B}_{N^k}^+$ the subset of $\mathcal{B}_{N^k}$ with Atkin–Lehner eigenvalue 1 for w_{N^k}. The Jacobians $J_0(N^k)^{\mathrm{new}}$ and $J_0^+(N^k)^{\mathrm{new}}$ admit $\mathbb{Q}$-isogenies

$$J_0(N^k)^{\mathrm{new}} \sim \prod_{f \in \mathcal{B}_{N^k}/\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})} A_f, \quad J_0^+(N^k)^{\mathrm{new}} \sim \prod_{f \in \mathcal{B}_{N^k}^+/\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})} A_f,$$

where A_f denotes the $\mathbb{Q}$-simple abelian variety associated to f by the Eichler–Shimura correspondence (which is independent of the choice of representative of the orbit). Because $X_{\mathrm{s}}^+(N)$ is isomorphic to $X_0^+(N^2)$ as we have seen above,

$$J_s^+(N) \cong J_0^+(N^2) \sim J_0(N) \times J_0^+(N^2)^{\mathrm{new}}$$

and by a theorem of Chen [16, Theorem 1], we also have a $\mathbb{Q}$-isogeny

$$J_{\text{ns}}^+(N) \sim J_0^+(N^2)^{\text{new}}. \tag{22}$$

The following lemma says that one would not expect to be able to use Chabauty's method to understand $X(\mathbb{Q})$.

Lemma 7 *Let* $X = X_0^+(N)$ *or* $X_{\text{ns}}(N)$. *Then the weak Birch–Swinnerton-Dyer conjecture implies* $X(\mathbb{Q}_p)_1 = X(\mathbb{Q}_p)$.

Proof The weak Birch–Swinnerton-Dyer conjecture implies that, for $f \in \mathcal{B}_{N^k}$, A_f will have positive rank whenever f has positive analytic rank. Since $f \in \mathcal{B}_{N^k}$ has odd analytic rank whenever $w_{N^k}(f) = 1$, and A_f is simple over $\mathbb{Q}$, the Birch–Swinnerton-Dyer conjecture hence implies that every isogeny factor of $\text{Jac}(X)$ (over $\mathbb{Q}$) has positive rank.

Since $\text{End}(A_f)$ is an order in the totally real field K_f, every isogeny factor of $\text{Jac}(X)$ has rank at least equal to its dimension. To prove the lemma, we must show that the image of $A_f(\mathbb{Q})$ in $\text{Lie}(A_f)_{\mathbb{Q}_p}$ under the p-adic logarithm map generates $\text{Lie}(A_f)_{\mathbb{Q}_p}$ as a $\mathbb{Q}_p$-vector space. This is equivalent to the statement that the image of $A_f(\mathbb{Q})$ in $\text{Lie}(A_f)_{\mathbb{C}_p}$ generates the latter as a $\mathbb{C}_p$-vector space. Since $\text{Lie}(A_f)_{\overline{\mathbb{Q}}}$ decomposes as a sum of one-dimensional isotypic components $\text{Lie}(A_f)_{\overline{\mathbb{Q}}, g}$, for g conjugate to f, and the p-adic logarithm is $\text{End}(A_f)$-equivariant, we deduce that if the image of $A_f(\mathbb{Q})$ does not span $\text{Lie}(A_f)_{\mathbb{C}_p}$ then there is a g conjugate to f such that the image of $A_f(\mathbb{Q})$ in $\text{Lie}(A_f)_{\mathbb{C}_p, g}$ is zero. By the p-adic analytic subgroup theorem [49, Theorem 1], [25, Theorem 2.2] if $P \in A_f(\overline{\mathbb{Q}})$ has the property that $\log(P) \in \text{Lie}(A_f)_{\mathbb{C}_p}$ lies in a proper subspace defined over $\overline{\mathbb{Q}}$, then P lies in a proper commutative sub-variety $B \subset A_{f, \overline{\mathbb{Q}}}$. Hence we deduce that if $A_f(\mathbb{Q})$ does not generate $\text{Lie}(A_f)_{\mathbb{Q}_p}$, then $A_f(\mathbb{Q})$ lies in a proper commutative subvariety of $A_{f, \overline{\mathbb{Q}}}$, since the isotypic components of $\text{Lie}(A_f)_{\mathbb{C}_p}$ are defined over $\overline{\mathbb{Q}}$.

We claim that this contradicts the Birch–Swinnerton-Dyer conjecture. More generally, if A is a simple abelian variety over $\mathbb{Q}$ and $\pi : A_K \to B$ is a non-zero morphism of abelian varieties over a finite Galois extension $K/\mathbb{Q}$, we claim that $P \in A(\mathbb{Q})$ is torsion if and only if its image in $B(K)$ is torsion (in particular, when $A = A_f$ and B is an isogeny factor, we deduce that A_f has rank zero over $\mathbb{Q}$ if and only if there is as isogeny factor B of $A_{f, \overline{\mathbb{Q}}}$ such that the image of $A_f(\mathbb{Q})$ in B is torsion). To see this claim, for $\sigma \in \text{Gal}(K/\mathbb{Q})$ let π^σ denote the conjugate homomorphism $A_K \to B^\sigma$. If $\pi(P)$ is torsion then $\pi^\sigma(P) = \pi(P)^\sigma$ is torsion for all σ, hence the image of P under the map

$$\prod_{\sigma \in \text{Gal}(K|\mathbb{Q})} \pi^\sigma : A_K \to \prod_\sigma B^\sigma$$

is torsion. However, this map descends to a non-zero morphism of $\mathbb{Q}$, and hence by simplicity of A, if $\pi(P)$ is torsion then P is torsion. $\qquad\square$

Moreover, two abelian varieties A_f, A_g for $f, g \in \mathcal{B}_{N^k}$ are non-isogenous unless f and g are conjugate by $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, and $\text{End}^\dagger(A_f)$ is always totally real of rank

$\dim(A_f)$, which proves that each of the Jacobians $J = J_0^+(N)$, $J_s^+(N)$, $J_{\mathrm{ns}}^+(N)$ satisfies $\rho(J) = \dim J$, and hence the condition (2) becomes

$$\mathrm{rk}(J) < 2 \cdot \dim(J) - 1 \tag{23}$$

(for a more general such condition for modular curves, see the main result of [59]). Using the isogenies above, the Birch–Swinnerton-Dyer conjecture implies

$$\mathrm{rk}(J_0^+(N)) = \sum_{f \in \mathcal{B}_N^+} \mathrm{ord}_{s=1} L(f, s), \quad \mathrm{rk}(J_{\mathrm{ns}}^+(N)) = \sum_{f \in \mathcal{B}_{N^2}^+} \mathrm{ord}_{s=1} L(f, s).$$

There is a whole literature on analytic estimates for these types of analytic ranks. In particular, using [45, Theorem 1.4] one can show that the Birch–Swinnerton-Dyer conjecture implies that

$$\limsup_N \frac{\mathrm{rk}(J_0^+(N))}{\dim J_0^+(N)} \leq 1.3782,$$

and in particular asymptotically that (2) is always satisfied. It is likely that the same result can be obtained for $J_{\mathrm{ns}}^+(N)$, but the square level (we are looking at $J_0^+(N^2)^{\mathrm{new}}$) raises serious technical difficulties for analytic estimates of second moments used there.

On the other hand, by Corollary 4, Theorem 2 implies that we have an isogeny factor A of J satisfying $\rho(A) > 1$ and $\mathrm{rk}(A) = \dim(A)$, hence to prove Proposition 3 it suffices to construct a nonzero $[L] \in \mathrm{Ker}(\mathrm{NS}(A) \to \mathrm{NS}(X))$ satisfying $\theta_{X,\pi_A,\pi_B}([L]) = 0$, where B is the isogeny factor consisting of modular abelian varieties associated to modular forms whose analytic rank of L-functions is greater than 1. It will be shown that for any L, its image $\theta_{X,\pi_A,\pi_B}(L)$ can be represented by a divisor supported on cusps and Heegner points, and hence is torsion by the generalised Gross–Zagier formula ([67, Theorem 6.1]) This motivates the following definition.

Definition 4 (Heegner quotient) Let $M = N$ or N^2. The *Heegner quotient* A of $J_0(M)^{\mathrm{new}}$ is the product

$$A := \prod_{\substack{f \in \mathcal{B}_M^{+,\mathrm{new}}/\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \\ L'(f,1) \neq 0}} A_f,$$

and its *complement* is

$$B := \prod_{\substack{f \in \mathcal{B}_M^{+,\mathrm{new}}/\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \\ L'(f,1) = 0}} A_f$$

(so that $A \times B$ is isogenous to $J_0^+(M)^{\mathrm{new}}$, not the full $J_0(M)^{\mathrm{new}}$).

In particular, Corollary 4 implies that $\mathrm{rk}(A) = \dim(A)$ (assuming the Birch–Swinnerton-Dyer conjecture, it is the largest factor of $J_0^+(M)$ with this property) and the generalised Gross–Zagier formula implies that all images of traces of Heegner points on $X_0(N)$ in B are torsion (see Sects. 4.2 and 4.3). In the case of $X_{\mathrm{ns}}(N)$, there is also a notion of Heegner point due to Kohen and Pacetti, inspired by the points used in Zhang's Gross–Zagier formula for $X_{\mathrm{ns}}(N)$ (and more general Shimura curves).

The main result of the next section is the following lemma, which refers to $X_0(N)$ and $X_{\mathrm{ns}}(N)$ rather than their Atkin–Lehner quotients. However, by Lemma 6 it implies Proposition 3.

Lemma 8 *Let $X = X_0(N)$ or $X_{\mathrm{ns}}(N)$, and A, B the Heegner quotient and its complement as defined above, endowed with the natural projections $(\pi_A, \pi_B) : \mathrm{Jac}(X) \to A \times B$. Then for all $[L]$ in $\mathrm{Ker}(d_{\pi_A})$, $\theta_{X,\pi_A,\pi_B}([L])$ is torsion. In particular the rank of the kernel of θ_{X,π_A,π_B} is maximal (in particular at least 1 if $\dim A \geq 2$).*

4.2 How to prove (C) using Heegner points under the analytic hypothesis: $X = X_0(N)$

In this section we prove Lemma 8. We will deduce it from the Gross–Zagier–Zhang theorem. In the case of $X_0(N)$, as explained in [20] or [19], we could also deduce it from the Yuan–Zhang–Zhang formula for the height of diagonal cycles (see Sect. 4.4). By a *Heegner point* on $X_0(N)$ we will mean a point

$$E \to E'$$

on $Y_0(N)$ such that E and E' have CM by the same order of an imaginary quadratic field K, not necessarily maximal but assumed to be with conductor prime to N (see [28] for a review of their properties, in particular N has to be split or ramified in K).

An eigenform $f \in S_2(\Gamma_0(N))^{+,\mathrm{new}}$ defines by Eichler-Shimura theory a $\mathbb{Q}$-simple quotient $\pi : J_0(N) \to A_f$ of $J_0(N)$ (in fact of $J_0^+(N)$) and the Heegner points behave on A_f in the following way.

Lemma 9 *1. If $L'(f, 1) \neq 0$, then $\mathrm{rk}(A_f) = \dim(A_f)$ (and $A_f(\mathbb{Q})$ is generated by the projection of a trace of a suitable choice of Heegner point).*
2. If $L'(f, 1) = 0$, then for any P in $\mathrm{Div}^0(X_0(N))(\overline{\mathbb{Q}})^{\mathrm{Gal}(\overline{\mathbb{Q}}|\mathbb{Q})}$ supported on the set of Heegner points, the image $\pi(P)$ is torsion in $A_f(\mathbb{Q})$.

Remark 7 The original Gross–Zagier formula [32, Theorem I.6.3] is not sufficient for the second part of the Lemma, as it only deals with Heegner points for which the discriminant of the order is squarefree (in particular, the order is maximal) and prime to N, which we cannot afford to assume here. This is why we need Zhang's formula and the ensuing technical interpretation.

Proof The first part is given by Proposition 8. The second part is a consequence of the generalised Gross–Zagier formula of Zhang [67, Theorem 6.1] which for this case is made completely explicit in [15, Theorem 1.1], see also [15, Example after Theorem 1.5]. We use the following notation: $f \in S_2(\Gamma_0(N))$ is a normalised eigenform, K

an imaginary quadratic field number field in which N is not inert, c prime to N, $\mathcal{O}_c = \mathbb{Z} + c\mathcal{O}_K$, and 1_c the trivial ring class character on $\mathrm{Pic}(\mathcal{O}_c)$. We denote by H_c the ray class field of K with conductor c. If P is a Heegner point on $X_0(N)$ with CM by $\mathcal{O}_c$, it belongs to $X_0(N)(H_c)$, and we define

$$P_{1_c} = \sum_{\sigma \in \mathrm{Gal}(H_c/K)} (P^\sigma - [\infty]) \in J_0(N)(K) \subset J_0(N)(H_c).$$

On the other hand, if $J(H_c) \otimes \mathbb{C}$ denotes the extension of scalars of $J(H_c)$ endowed with the extended Néron-Tate height, we have the decomposition into isotypical components

$$J_0(N)(H_c) \otimes \mathbb{C} = \bigoplus_g J_0(N)_g,$$

where g goes through all eigenforms of weight 2 of $J_0(N)$, so that $J_0(N)_g$ is exactly the isotypical part where T_n acts by multiplication by $a_n(g)$. We denote by $P_{1_c}^f$ the projection of P_{1_c} on the f-isotypical component. The statement of [15, Theorem 1.1] then tells (which is sufficient for us) that $L'(f, 1_c, 1)$ as defined there is proportional (by an explicit nonzero factor) to the extended Néron-Tate height of $P_{1_c}^f$.

We have the equality of L-functions

$$L(f, 1_K, s) = L(f, s) L(f \otimes \chi_K, s),$$

with 1_K the trivial class character on $\mathrm{Pic}(\mathcal{O}_K)$ and χ_K the Dirichlet character associated to K. In particular (and given the signs of functional equations on the right), our hypothesis $L'(f, 1) = 0$ guarantees that $L(f, 1_K, s)$ vanishes with order at least 2 at 1, so the left-hand side of [15, Theorem 1.1] is zero for $c = 1$. This also holds for any c prime to N, because by construction $L(f, 1_c, s)$ is a multiple of $L(f, 1_K, s)$ around 1 (given the definition again). We have thus proved that $P_{1_c}^g$ is zero in $J_0(N)(H_c) \otimes \mathbb{C}$.

Now, the group $\mathrm{Aut}(\mathbb{C})$ acts on $J_0(N)(H_c) \otimes \mathbb{C}$ by the identity on the left and the natural action on the right, and for every $\alpha \in \mathrm{Aut}(\mathbb{C})$ acting as such, we have $P_{1_c}^\alpha = P_{1_c}$ and then for every $\alpha \in \mathrm{Aut}(\mathbb{C})$, we obtain $(P_{1_c}^g)^\alpha = P_{1_c}^{\alpha(g)}$ where $\alpha(g)$ is the eigenform obtained by conjugating the coefficients of g (see [32, Corollary V.1.2]). Now, as we also have the decomposition

$$J_0(N)(H_c) \otimes \mathbb{C} \cong \prod_{f \in \mathcal{B}_N/\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})} A_f(H_c) \otimes \mathbb{C}$$

in subrepresentations of the Hecke algebra, the sum of all $P_{1_c}^g$ for g conjugate to f is proportional to the projection π of the trace of $P - (\infty)$ (belonging to $J_0(N)(K)$) in $A_f(K) \otimes \mathbb{C}$, so we have proven that this projection in $A_f(K)$ is torsion. $\qquad\square$

We now explain how to deduce Lemma 8 from this result. Let m be an integer coprime to N. Define the Hecke correspondence $\widetilde{C}_m$ to be the image of $X_0(mN)$ in $X_0(N) \times$

$X_0(N)$ under the product of the two natural maps $X_0(mN) \to X_0(N)$. We define

$$C_m = (1 - \pi_1^* i_1^* - \pi_2^* i_2^*)\widetilde{C}_m$$

to be the projection of $\widetilde{C}_m$ onto the $\mathrm{End}(J_0(N))$ component of $\mathrm{Pic}(X_0(N) \times X_0(N))$ (see (8)). Then C_m lands in the subspace $\mathrm{NS}(J_0(N)) \subset \mathrm{End}(J_0(N))$ of endomorphisms symmetric with respect to the Rosati involution. When m is square-free, C_m is the Hecke operator T_m. In general, C_m is a linear combination of $T_{m/d}$ for d divisors of m.

Recall that $i_{1,2} \colon X_0(N) \hookrightarrow X_0(N) \times X_0(N)$ denotes the diagonal morphism. A non-cuspidal point in the support of $i_{1,2}^*(\widetilde{C}_m)$ is a cyclic N-isogeny $f \colon E_1 \to E_2$, together with cyclic subgroups G_i of E_i of order m such that $f(G_1) = G_2$, and isomorphisms

$$E_i \xrightarrow{\simeq} E_i/G_i$$

which commute with f and the induced isogeny $E_1/G_1 \to E_2/G_2$. In particular, the ring of endomorphisms of each E_i, of discriminant denoted by D_i, thus contains an element of norm m so there exist A_i, B_i in $\mathbb{Z}$ for which

$$A_i^2 + D_i B_i^2 = 4m. \tag{24}$$

The isogeny being cyclic, A_i and B_i must be coprime here. The point $E_1 \to E_2$ is a Heegner point of $Y_0(N)$ if and only if $D_1 = D_2$.

Lemma 10 *Let $X = X_0(N)$, let m be prime to N, and let $\widetilde{C}_m$ be the Hecke correspondence defined above. Then the divisor $i_{1,2}^* \widetilde{C}_m$ is supported on the set of Heegner points whenever m is less than $N/4$.*

Proof Let $(E_1 \to E_2)$ be a non-cuspidal point in the support of $i_{1,2}^* \widetilde{C}_m$ as above. Suppose the point is not Heegner. Since E_1 and E_2 are N-isogenous, $D_2 = \lambda^2 D_1$ for some rational $\lambda > 0$ a power of N. Since $\lambda \neq 1$, we must have D_i divisible by N^2 for some i, and hence $m > N^2/4$, by (24). Finally, if the conductor of the order was not prime to N, we would also have $N^2 | D_i$ which leads to the same inequality. $\quad\square$

By the following Lemma (essentially just the Sturm bound) we have enough Hecke operators C_m for which $i_{1,2}^* C_m$ is supported on cusps and Heegner points to complete the proof of the first part of Lemma 8.

Lemma 11 *Let N be a prime. Then, any element of $\mathrm{End}^\dagger(J_0^+(N))^{\mathrm{tr}=0}$, viewed as a subspace of $\mathrm{End}^\dagger(J_0(N))^{\mathrm{tr}=0}$, can be written as a $\mathbb{Z}$-linear combination of endomorphisms associated to the Hecke correspondences C_m, for $m < N^2/4$ prime to N.*

Proof By the Sturm bound ([62] Theorem 9.18), the set of Hecke operators T_m for $m < N^2/4$ spans the Hecke algebra of endomorphisms of $J_0(N)$. Since $a_N(f) = -1$ on newforms such that $f_{|w_N} = -f$, the set of Hecke operators T_m for $m < N^2/4$ prime to N spans the Hecke algebra of endomorphisms of $J_0^+(N)$ (which is the full endomorphism algebra over $\mathbb{Q}$). $\quad\square$

This completes the proof of case (1) of Proposition 3. Indeed, Lemma 11 implies that any nice correspondence Z on $X_0(N)$ can be written as a linear combination of the C_m for $m < N^2/4$ prime to N. By Lemma 10, for any such Z, $D_Z(b)$ is supported on Heegner points and cusps, so by Lemma 9 (part 2), its image by π_B is torsion.

4.3 How to prove (C) using Heegner points under the analytic hypothesis: $X = X_{ns}^+(N)$

The second case is similar to the first, but we must replace the classical notion of Heegner point with Heegner points on non-split Cartan modular curves in the sense of Zhang/Kohen–Pacetti, and replace Gross–Zagier–Zhang on $X_0(N)$ with Zhang's Gross–Zagier theorem on $X_{ns}(N)$.

To make results easier to state, we use the moduli interpretation of $X_{ns}(N)$ and $X_{ns}^+(N)$ given in [42] and its consequences. To do so, one fixes an $\varepsilon \in \mathbb{F}_N$ which is not a square. A pair (E, ϕ_ε) is then an elliptic curve E together with an endomorphism ϕ_ε of $E[N]$ whose square is multiplication by ε. Such an endomorphism has eigenvalues in $\mathbb{F}_{N^2} \backslash \mathbb{F}_N$, and two pairs (E, ϕ_ε) and (E', ϕ'_ε) are isomorphic if there is an isomorphism $\psi : E \to E'$ such that on $E[N]$, $\psi \circ \phi_\varepsilon = \phi'_\varepsilon \circ \psi$.

$X_{ns}(N)$ is the compactified moduli space of such pairs up to isomorphism [42, §1.2]. Furthermore, the natural involution on this modular curve is given by $(E, \phi_\varepsilon) \mapsto (E, -\phi_\varepsilon)$.

First, we define Hecke correspondences $\widetilde{C_m} \subset X_{ns}(N) \times X_{ns}(N)$ (for m prime to N) as follows. We have a curve $X_{ns}(N, m) = X_{ns}(N) \times_{X(1)} X_0(m)$ given by adding an auxiliary $\Gamma_0(m)$ structure. We have two maps $X_{ns}(N, m) \to X_{ns}(N)$, the forgetful one, and the one sending (E, ϕ_ε, C) to $(E/C, \overline{\pi_C} \circ \phi_\varepsilon \circ \overline{\pi_C}^{-1})$ where C is a cyclic subgroup of order m, $\pi_C : E \to E/C$ the natural projection, and $\overline{\pi_C}$ the induced map $E[N] \to (E/C)[N]$. Furthermore, Chen morphisms between $J_{ns}(N)$ and $J_0(N^2)$ are equivariant with respect to the Hecke actions [42, Theorem 1.11].

We will again use the generalised Gross–Zagier formula from Zhang from [67], in a slightly different context here. We follow the notation of [67, §6]. Let $K/\mathbb{Q}$ be an imaginary quadratic field inert at N (instead of split or ramified in the previous case), and let $K \hookrightarrow M_2(\mathbb{Q})$ be an embedding associated to an integral basis of $\mathcal{O}_K$. For a choice of order $\mathcal{O}_c$ of K of conductor c prime to N, define

$$R_c = \mathcal{O}_c + N \cdot M_2(\mathbb{Z})$$

(notice the index of $N\mathcal{O}_K$ is N^2). The Shimura variety M_{U_c} is then uniformised as

$$M_{U_c}(\mathbb{C}) = \mathrm{GL}_2(\mathbb{Q})_+ \backslash \mathcal{H} \times \mathrm{GL}_2(\mathbb{A}_f)/U_c,$$

where U_c can be defined as $\mathrm{GL}_2(\mathbb{Z}_v)$ for places v not dividing N, and $(R_c \otimes \mathbb{Z}_N)^* \subset \mathrm{GL}_2(\mathbb{Z}_N)$ at N (seen in $\mathrm{GL}_2(\mathbb{Z}_N)$). Note that $\mathrm{GL}_2(\mathbb{Q})^+ \cdot U_c = \mathrm{GL}_2(\mathbb{A}_f)$ and $\mathrm{GL}_2(\mathbb{Q})_+ \cap U_c \subset \mathrm{SL}_2(\mathbb{Z})$ contains the subgroup $\Gamma(N)$ of $\mathrm{SL}_2(\mathbb{Z})$ of all matrices congruent to the identity modulo N, and the quotient is a conjugate of $C_{ns}(N) \cap \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, where the precise choice of $C_{ns}(N)$ comes from the reduction

modulo N of $\mathcal{O}_c$ inside $M_2(\mathbb{Z}/N\mathbb{Z})$ given by the embedding (it is nonsplit precisely because N is inert in $\mathcal{O}_c$). This gives an isomorphism

$$M_{U_c}(\mathbb{C}) \simeq Y_{ns}(N)_{\mathbb{C}}.$$

The CM points on M_{U_c} in the sense of Zhang are then the double cosets of pairs (h_0, i_c), where h_0 is fixed by the image T of the torus $K^\times$. and i_c has the property that

$$i_c U_c i_c^{-1} \cap T(\mathbb{A}_f) \simeq \widehat{\mathcal{O}}_c^\times / \widehat{\mathcal{O}}_F^\times,$$

in other words the nonsplit Cartan structure of level N is the one determined by the endomorphism ring of the CM elliptic curve.

On the other hand, we say that $(E, \phi_\varepsilon) \in Y_{ns}(N)$ is a Heegner point (in the sense of Kohen–Pacetti) with multiplication by $\mathcal{O}_c$ if $\mathrm{End}(E) \cong \mathcal{O}_c$ (with c prime to N) and ϕ_ε comes from an endomorphism β of E. Note that this implies that N is inert in $\mathcal{O}_c$, since the minimal polynomial of β modulo N is then irreducible.

This discussion thus implies the following equivalence of definitions.

Lemma 12 *Under the identification $M_{U_c} \simeq Y_{ns}(N)$ for every order $\mathcal{O}_c$ of conductor c prime to N, Zhang's CM points correspond to Heegner points with CM by $\mathcal{O}_c$ in $Y_{ns}(N)$ in the sense of Kohen–Pacetti.*

Let f be an eigenform in $S_2(\Gamma_0(N^2))^{+,\mathrm{new}}$. It can be seen as an automorphic form on an M_{U_c} as above, using the isomorphism of Hecke modules $S_2(\Gamma_0(N^2))^{+,\mathrm{new}} \cong S_2(\Gamma_{ns}^+(N))$ and the isomorphism $M_{U_c}(\mathbb{C}) \cong Y_{ns}(N)_{\mathbb{C}}$ and we again have by Eichler-Shimura theory a $\mathbb{Q}$-simple quotient A_f of $J_{ns}^+(N)$.

The consequence of Zhang's result that we will use is the following.

Theorem 3 ([67], Theorem 6.1) *With notation as above, let 1_c be the trivial character of $\mathrm{Gal}(H_c/K)$ and P a Heegner point on $Y_{ns}(N)$ with CM by $\mathcal{O}_c$ in the sense of Kohen-Pacetti. Denote by P_{1_c} be the projection of $P - \xi$ (ξ the Hodge class) in $J_{ns}(N)(K) = J_{ns}(N)(H_c)^{1_c}$. Let $P_{1_c}^f$ be the projection of P_{1_c} onto the f-isotypical component of $J_{ns}(N)(H_c) \otimes \mathbb{C}$.*

If $L'(f, 1) = 0$, then $P_{1_c}^f = 0$ and $\pi_f(P_{1_c})$ is torsion in $A_f(H_c)$.

Proof Using the previous lemmas and discussion, we can translate everything in terms of the Shimura curve M_{U_c}: the Heegner point P becomes a CM point in the sense of Zhang and f becomes an automorphic representation ϕ. These changes are compatible with Hecke operators and Galois actions, so they preserve the decompositions into isotypical components above. We can then proceed along the same lines as the proof of Lemma 9 part 2 to deduce the conclusion from Zhang's theorem. □

We are now ready to prove the analogue of Lemma 10 with $X_0(N)$ replaced by $X_{ns}^+(N)$.

Lemma 13 *Let $X = X_{ns}(N)$, let m be prime to N, and let $\widetilde{C}_m$ be the Hecke correspondence defined above. Then the divisor $i_{1,2}^* \widetilde{C}_m$ is supported on Heegner points in the sense of Kohen-Pacetti and cusps whenever m is less than $N^2/4$.*

Proof By the moduli interpretation of $X_{ns}(N)$ and the Hecke correspondences, a non-cuspidal point in the support of $i_{1,2}^* \widetilde{C}_m$ is a pair (E, ϕ_ε) such that there exists an endomorphism α of E of norm m with cyclic kernel (of order m) such that if $\overline{\alpha}$ is the induced endomorphism of $E[N]$, $\overline{\alpha} \circ \phi_\varepsilon \circ \overline{\alpha}^{-1} = \phi_\varepsilon$. This implies that $\overline{\alpha}$ belongs to the nonsplit Cartan subgroup associated to ϕ_ε (which is also the group of invertible elements of $\mathbb{Z}[\phi_\varepsilon]$). We claim $\overline{\alpha}$ is not scalar: if it were, we could write $\alpha = k + N\beta, k \in \mathbb{Z} \beta \in \mathrm{End}(E)$ and then the norm of α being $m < N^2/4$ forces β to be an integer as well, contradicting the assumption that α has cyclic kernel.

From this, we deduce that $\mathbb{Z}[\overline{\alpha}] = \mathbb{Z}[\phi_\varepsilon]$, as both are $\mathbb{Z}/N\mathbb{Z}$-vector spaces of dimension 2 and the former is included in the latter. This implies that ϕ_ε is induced by the action of an element of $\mathbb{Z}[\alpha] \subset \mathrm{End}(E)$ on $E[N]$, and the ring of endomorphisms has conductor prime to N for the same reasons as in $X_0(N)$, and its discriminant is automatically prime to N as discussed after defining Heegner points in the sense of Kohen-Pacetti. $\qquad\square$

By the compatibility with Hecke correspondences on $X_0(N^2)$ (which is a consequence of Chen's theorem without quotient by Atkin-Lehner involutions, e.g. [60, Théorème 2]), Lemma 11 implies that any nice correspondence Z on $X_{ns}^+(N)$ can be written as a linear combination of C_m for $m < N^2/4$ prime to N. By Lemma 13, for any such Z, $D_Z(b)$ is supported on Heegner points (in the sense of Kohen–Pacetti) and cusps. Hence, Zhang's Gross–Zagier theorem (together with Manin–Drinfeld) implies $\pi_B(D_Z(b))$ is torsion. Assuming the conclusions of Theorem 2 hold for M, the Heegner quotient A of $J_0^+(M)^{\mathrm{new}}$ is of dimension at least 2 so $\rho(A) \geq 2$. This completes the proof of case (2) of Proposition 3.

4.4 An alternative approach

In this subsection, we sketch an alternative and less ad hoc approach for proving Proposition 3 in the case $X = X_0^+(N)$, using the Theorem of Yuan–Zhang–Zhang on the heights of diagonal cycles.

Theorem 4 (Darmon–Rotger–Sols [19], Theorem 3.7) *Let $X = X_0(N)$, and let f, g be non-conjugate eigenforms in $S_2(\Gamma_0(N))$. Let $Z \in \mathrm{NS}(J_0(N))$ lie in the image of $\mathrm{NS}(A_g)$. Suppose $\epsilon(f) = -1$ and $\epsilon(\mathrm{Sym}^2(g) \otimes f) = 1$. If the projection of $D_Z(b)$ to A_f is non-torsion, then $L'(f, 1) \neq 0$.*

The result above holds for arbitrary N, but is most useful when N is prime, since in this case we have $\epsilon(f \otimes g \otimes g) = -a_N(f)a_N(g)^2 = -a_N(f)$ (see e.g. [30]). Hence in this case Theorem 4 implies that the image of $D_Z(b)$ in A_f is torsion for all eigenforms f in $S_2^+(\Gamma_0(N))$., which implies that we get an alternative proof for $X_0^+(N)$. One way to view Proposition 3 is that it shows that it is easier to prove diagonal cycles are torsion than it is to prove they are non-torsion. On the other hand, one can show directly that the image of $D_Z(b)$ in A_f is torsion for all eigenforms f satisfying $w_N(f) = -f$, as explained in [20, Theorem 3.3.8]: by Lemma 6, we have

$$w_N^*(D_Z(b)) = D_{w_N^*(Z)}(b).$$

Since $w_N^*(Z) = Z$, and w_N^* acts as (-1) on A_f, we deduce $\pi_{f*}(D_Z(b))$ is torsion.

5 Proof of the analytic part

In this section, we prove Theorem 2 using analytic weighted averages techniques, following guiding principles e.g. from [37] and [24]. For convenience and consistency, the notation below is as close as possible to that from [47].

Notation

– N is a prime number and $M = N$ or N^2 in all of the following.
– If $f, g \in S_2(\Gamma_0(M))$, we denote their Petersson scalar product by

$$\langle f, g \rangle_M = \int_{\mathcal{D}} \overline{f(x + iy)} g(x + iy) dx dy,$$

where $\mathcal{D}$ is a fundamental domain of $\Gamma_0(M)$, and the associated Petersson norm by $\| \cdot \|_M$.
– For $\varepsilon = \pm 1$, the space $S_2(\Gamma_0(M))^\varepsilon$ refers to the subspace of modular forms f of $S_2(\Gamma_0(M))$ such that $f_{|w_M} = \varepsilon \cdot f$, where w_M is the Fricke involution of $S_2(\Gamma_0(M))$. Note that in weight 2, this is the space of modular forms f such that $L(f, s)$ has root number $-\varepsilon$.
– For A, B linear forms on $S_2(\Gamma_0(M))$ (resp. on a subspace indicated by superscripts), we write

$$\langle A, B \rangle_M = \sum_f \frac{\overline{A(f)} B(f)}{\|f\|_M^2},$$

where f goes through an orthogonal basis of $S_2(\Gamma_0(M))$ (it is readily checked not to depend on this choice of basis), resp. of the prescribed subspace. We will add superscripts $\{+, -, \mathrm{new}, \mathrm{old}\}$ to refer to the sum restricted to an orthogonal basis of the corresponding subspaces of $S_2(\Gamma_0(M))$.
– We denote by a_m (for $m \in \mathbb{N}_{\geq 1}$) and L' the linear forms on $S_2(\Gamma_0(M))$ which to f associate respectively the m-th coefficient of the q-expansion of f, and $L'(f, 1)$ (defined properly in the next paragraph).
– The (positive) greatest common divisors of integers a, b or integers a, b, c are respectively denoted by (a, b) and (a, b, c).
– For any positive number B, $O_1(B)$ refers to a complex number of absolute value $\leq B$.

The proof of Theorem 2 relies on the following lemma.

Lemma 14 *Theorem 2 holds for M if*

$$\langle a_1, L' \rangle_M^{+,\mathrm{new}} \neq 0 \quad and \quad \frac{\langle a_2, L' \rangle_M^{+,\mathrm{new}}}{\langle a_1, L' \rangle_M^{+,\mathrm{new}}} \in \,]0, 1[.$$

Proof If $\langle a_1, L' \rangle_M^{+,\text{new}} \neq 0$, by definition of this sum, there must be at least one normalised newform $f \in S_2(\Gamma_0(M))^{+,\text{new}}$ such that $L'(f, 1) \neq 0$. As a byproduct of the Gross–Zagier formula ([32], Corollary V.1.3), this implies that $L'(g, 1) \neq 0$ for all normalised newforms g which are conjugates of f by $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, thus Theorem 2 holds for M unless the field of coefficients of f is $\mathbb{Q}$ and this f is unique, which we assume now. As f is normalised, those coefficients are algebraic integers hence belong to $\mathbb{Z}$. Now, one has

$$\frac{\langle a_2, L' \rangle_M^{+,\text{new}}}{\langle a_1, L' \rangle_M^{+,\text{new}}} = \frac{\overline{a_2(f)}L'(f, 1)\|f\|_M^2}{\overline{a_1(f)}L'(f, 1)\|f\|_M^2} = a_2(f) \in]0, 1[$$

by hypothesis, so $a_2(f) \notin \mathbb{Z}$ which leads to a contradiction and Theorem 2 holds. $\square$

Remark 8 The statement of this lemma appears quite *ad hoc* so let us explain the main motivations behind it.

– As we will see later, as long as m is small compared to $\sqrt{M}$, one has

$$\frac{\langle a_m, L' \rangle_M^{+,\text{new}}}{4\pi} = \ln(\sqrt{M}) + C - \ln(m) + O(m/\sqrt{M})$$

with explicit implied constants. This proves that the hypotheses of the lemma are indeed satisfied for large M.
– The error terms of the estimate above are smaller when the m's are smaller, hence the choices of $m = 1$ and 2 for the ratio.
– There are far better asymptotic estimates on the number of newforms f in $S_2(\Gamma_0(M))^{+,\text{new}}$ such that $L'(f, 1) \neq 0$, e.g. : by [45] (at least for $M = N$ prime), the proportion of such forms is asymptotically at least 7/8, in particular there are far more than just 2 for M large). These techniques, using also estimates of second moments and of the norms $\|f\|_M$, are harder to make explicit, and we suspect the effective bounds obtained by following step-by-step the arguments would be huge. Lemma 14, while very crude (and giving a weaker result) is tailor-made to be efficient enough for precise estimates and approachable bounds.

5.1 Splitting of the terms to estimate the first moments

The starting point to estimate the weighted averages $\langle a_m, L' \rangle_N^{\text{new}}$ is the following trace formula of Petersson adapted by Akbary (and proven in greater generality in [47]).

Proposition 4 *Let m, n, M be three positive integers, and $\varepsilon = \pm 1$. Then, we have*

$$\frac{1}{2\pi\sqrt{mn}}\langle a_m, a_n \rangle_M^{\varepsilon} = \delta_{mn} - 2\pi \sum_{\substack{c>0 \\ M|c}} \frac{S(m, n; c)}{c} J_1\left(\frac{4\pi\sqrt{mn}}{c}\right)$$

$$-2\pi\varepsilon \sum_{\substack{d>0 \\ (d,M)=1}} \frac{S(m, nM^{-1}; d)}{d\sqrt{M}} J_1\left(\frac{4\pi\sqrt{mn}}{d\sqrt{M}}\right), \quad (25)$$

where S is the notation for Kloosterman sums

$$S(m, n; c) = \sum_{k \in (\mathbb{Z}/c\mathbb{Z})^*} e^{2i\pi(mk+nk^{-1})/c}$$

(except for $c = 1$ where its value is 1 by convention), Q^{-1} means the inverse of Q modulo d in the Kloosterman sums and J_1 is the Bessel function of the first kind and order 1.

The sums on the right-hand side are absolutely convergent thanks to the following well-known uniform bounds: $|J_1(x)| \leq |x|/2$ for all x, and the Weil bounds

$$|S(m, n; c)| \leq (m, n, c)^{1/2}\tau(c)\sqrt{c}, \tag{26}$$

with τ the divisor-counting function, which improves, if M is a prime power dividing c, in

$$|S(m, n; c)| \leq 2(m, n, c)^{1/2}\tau(c/M)\sqrt{c}$$

([36], (3.2), (3.3), Theorem 11.11 and Corollary 11.12).

Now, our normalisation of the L-function associated to a form $f \in S_2(\Gamma_0(M))$ is given by

$$L(f, s) = \sum_{n=1} \frac{a_n(f)}{n^s},$$

and this L-series converges uniformly on any compact subset of $\{\text{Re}(s) > 2\}$.

One can express $L'(f, 1)$ itself in terms of the Fourier coefficients of f in the following way.

Lemma 15 *For any $M \geq 1$ and any $f \in S_2(\Gamma_0(M))^+$, one has*

$$L'(f, 1) = 2\sum_{n=1}^{+\infty} \frac{a_n(f)}{n} E_1\left(\frac{2\pi n}{\sqrt{M}}\right)$$

where E_1 is the exponential integral function, defined on $]0, +\infty[$ by

$$E_1(y) = \int_y^{+\infty} \frac{e^{-t}}{t} dt.$$

Proof We define the completed L-function Λ associated to L by

$$\Lambda(f, s) := \left(\frac{\sqrt{M}}{2\pi}\right)^s \Gamma(s)L(f, s). \tag{27}$$

By standard arguments(e.g. [14], section 1.5), this function extends to an holomorphic function on $\mathbb{C}$ and satisfies the functional equation

$$\Lambda(f, 2 - s) = -\Lambda(f_{|w_M}, s). \tag{28}$$

The expression of $L'(f, 1)$ is then deduced from the functional equation of Λ by integration of residues on vertical axes and Mellin transform (see e.g. [36] (26.10) where the definition of L is translated by $1/2$). □

With this formula and by uniform convergence of the terms involved, we obtain:

$$\frac{\langle a_m, L' \rangle_M^+}{4\pi} = E_1\left(\frac{2\pi m}{\sqrt{M}}\right) - 2\pi\sqrt{m}\left(\sum_{M|c}\frac{\mathcal{S}(c)}{c} + \sum_{(d,M)=1}\frac{\mathcal{T}(d)}{d\sqrt{M}}\right), \tag{29}$$

where

$$\mathcal{S}(c) = \sum_{n=1}^{+\infty}\frac{S(m, n; c)}{\sqrt{n}}J_1\left(\frac{4\pi\sqrt{mn}}{c}\right)E_1\left(\frac{2\pi n}{\sqrt{M}}\right) \tag{30}$$

and

$$\mathcal{T}(d) = \sum_{n=1}^{+\infty}\frac{S(m, nM^{-1}; d)}{\sqrt{n}}J_1\left(\frac{4\pi\sqrt{mn}}{d\sqrt{M}}\right)E_1\left(\frac{2\pi n}{\sqrt{M}}\right). \tag{31}$$

The main term in (29) will be $E_1(2\pi m/\sqrt{M})$ as long as $m \ll \sqrt{M}$.

The trace formula does not separate the old and new spaces, which we need for $M = N^2$. This is taken care of in the following lemma.

Lemma 16 *For N prime and $m \geq 1$ not divisible by N,*

$$\langle a_m, L' \rangle_{N^2}^{+,\text{new}} = \langle a_m, L' \rangle_{N^2}^+ - \frac{1}{N-1}\left(\langle a_m, L' \rangle_N^+ + \frac{\ln(N)}{2}\langle a_m, L \rangle_N^-\right).$$

Proof By orthogonality of the new and old subspaces,

$$\langle a_m, L' \rangle_{N^2}^{+,\text{new}} = \langle a_m, L' \rangle_{N^2}^+ - \langle a_m, L' \rangle_{N^2}^{+,\text{old}}.$$

To prove the formula on the oldpart, we need to be a bit careful with the definitions of completed L-functions: although the definition of $L(f, s)$ does not depend on the ambient space of modular forms, the definition of the completed L-function $\Lambda(f, s)$ in (27) does. The degeneracy operators are denoted by A_n as in the original article [1]. Let

$$A_1 = I_2, \quad A_N = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}, \quad W_N = \begin{pmatrix} 0 & 1 \\ -N & 0 \end{pmatrix}, \quad W_{N^2} = \begin{pmatrix} 0 & 1 \\ -N^2 & 0 \end{pmatrix}.$$

Notice that $(A_N W_{N^2} W_N^{-1})/N$ belongs to $\Gamma_0(N)$, thus for $f \in S_2(\Gamma_0(N))$ such that $f_{|W_N} = \varepsilon_f \cdot f$, one has

$$(f_{|A_N})_{|W_{N^2}} = (f_{|W_N})_{|A_1} = \varepsilon_f \cdot f_{|A_1}, \tag{32}$$

hence also

$$(f_{|A_1})_{|W_{N^2}} = \varepsilon_f \cdot f_{|A_N}.$$

Consequently, an orthogonal (see the computations of section 4 of [47] for example) basis of $S_2(\Gamma_0(N^2))^{+,\text{old}}$ is given by the $f_{|A_1} + (f_{|A_1})_{|W_{N^2}}$, where f goes through an eigenbasis of $S_2(\Gamma_0(N))$. The aforementioned computations also prove with (32) that if $f_{|W_N} = \varepsilon_f \cdot f$, then

$$\langle f_{|A_1} + (f_{|A_1})_{|W_{N^2}}, f_{|A_1} + (f_{|A_1})_{|W_{N^2}} \rangle_{N^2} = 2(N-1)\langle f, f \rangle_N.$$

If N does not divide m (so that $a_m(f_{|A_N}) = 0$), this implies that

$$\langle a_m, L' \rangle_{N^2}^{+,\text{old}} = \frac{1}{2(N-1)} \sum_f \overline{a_m(f)} L'(f_{|A_1} + (f_{|A_1})_{|W_{N^2}}, 1)$$

where f goes through an orthonormal basis of $S_2(\Gamma_0(N))$. Now, by the functional equation of $\Lambda(f, s)$ in (28), $\Lambda'(f_{|A_1}, 1) = \Lambda'((f_{|A_1})_{|W_{N^2}}, 1)$ but

$$\Lambda'(f_{|A_1}, 1) = \frac{N}{2\pi}(L'(f_{|A_1}, 1) + (\ln(N/2\pi) + \gamma)L(f, 1))$$

$$\Lambda'((f_{|A_1})_{|W_{N^2}}, 1) = \frac{N}{2\pi}(L'((f_{|A_1})_{|W_{N^2}}, 1) + (\ln(N/2\pi) + \gamma)\varepsilon_f L(f, 1)).$$

The first equality is a direct application of the definition of Λ, the second one uses that $L(f_{|A_N}, 1) = L(f, 1)$ (easy to show by the integral formula of $L(f, 1)$) and the results above. Thus, to compute $L'(f_{|A_1} + (f_{|A_1})_{|W_{N^2}}, 1)$, it is enough to know the sum of the two right-hand terms which is the sum of the two left-hand terms, which equal one another. Now, if $\varepsilon_f = 1$ then $L(f, 1) = 0$ by sign of the functional equation of $\Lambda(f, s)$ (in level N here !), and if $\varepsilon_f = -1$, $\Lambda'(f, 1) = 0$. We thus obtain in this case

$$L'(f, 1) = -(\ln(\sqrt{N}/(2\pi)) + \gamma)L(f, 1),$$

and get the lemma by summation on those forms f's gathered by sign of ε_f. □

5.2 First estimates

We recall that $M = N$ or N^2.

Lemma 17 *Using the Weil bounds, we get for every c multiple of M and d prime to M:*

$$|\mathcal{S}(c)| \leq 2\sqrt{mM}\tau(c/M)\frac{f((m,c))}{\sqrt{c}}, \quad |\mathcal{T}(d)| \leq \tau(d)\sqrt{m}\frac{f((m,d))}{\sqrt{d}}$$

where for every integer k, $f(k) = \sum_{k'|k} \frac{1}{\sqrt{k'}}$. For $m = 2$ and c, d even, these estimates are improved to

$$|\mathcal{S}(c)| \leq (\sqrt{2}+2)\frac{\sqrt{M}\tau(c/M)}{\sqrt{c}}, \quad |\mathcal{T}(d)| \leq (1+1/\sqrt{2})\frac{\tau(d)}{\sqrt{d}}. \tag{33}$$

Proof In the definitions of $\mathcal{S}(c)$ (and similarly for $\mathcal{T}(d)$), we separate the terms in n depending on the values of $(m, n, c) = m'$ which is a divisor of (m, c). Then, using $|J_1(x)| \leq |x|/2$, it only remains to control the sum of the $E_1(2\pi m'n/\sqrt{M})$ for n from 1 to $+\infty$, which after sum-integral comparison and variable change is smaller than $\sqrt{M}/(2\pi m')$.

In the specific case where $m = 2$ and c or d even, the cases are made from the beginning on the values of $(m, c)^{1/2}$ instead of bounding by $(m, c)^{1/2}$, and a careful computation gives those bounds. □

This allows us to bound the sum of the $\mathcal{S}(c)/c$ for all multiples c of M. By multiplicativity of τ,

$$\left| \sum_{M|c} \frac{\mathcal{S}(c)}{c} \right| \leq \frac{2\sqrt{m}}{M} \sum_{m'|m} \frac{f(m')\tau(m')}{(m')^{3/2}} \sum_{c=1}^{+\infty} \frac{\tau(c)}{c^{3/2}}$$

$$\leq \frac{2\sqrt{m}}{M} \sum_{m'|m} \frac{\tau(m')}{m'} \sum_{c=1}^{+\infty} \frac{\tau(c)}{c^{3/2}},$$

the sum on c being exactly $\zeta(3/2)^2$. We denote

$$g(m) = \sum_{m'|m} \frac{f(m')\tau(m')}{(m')^{3/2}}$$

hence (and similarly for $\mathcal{T}$):

$$2\pi\sqrt{m}\left| \sum_{M|c} \frac{\mathcal{S}(c)}{c} \right| \leq \frac{86m}{M}g(m), \quad 2\pi\sqrt{m}\left| \sum_{(d,M)=1} \frac{\mathcal{T}(d)}{d\sqrt{M}} \right| \leq \frac{43m}{\sqrt{M}}g(m) \tag{34}$$

which gives

$$\frac{\langle a_m, L'\rangle_M^+}{4\pi} = E_1(2\pi m/\sqrt{M}) + g(m)m\left(O_1\left(\frac{86}{M}\right) + O_1\left(\frac{43}{\sqrt{M}}\right)\right). \tag{35}$$

For $m = 2$, the previous refinements can be exploited and we get instead

$$2\pi\sqrt{2}\left|\sum_{M|c}\frac{S(c)}{c}\right| \leq \frac{213}{M}, \quad 2\pi\sqrt{2}\left|\sum_{(d,M)=1}\frac{T(d)}{d\sqrt{M}}\right| \leq \frac{97}{\sqrt{M}}$$

hence

$$\frac{\langle a_2, L'\rangle_M^+}{4\pi} = E_1(4\pi/\sqrt{M}) + O_1\left(\frac{213}{M}\right) + O_1\left(\frac{97}{\sqrt{M}}\right). \qquad (36)$$

Identical bounds are found for

$$S_0(c) = \sum_{n=1}^{+\infty}\frac{S(m, n; c)}{\sqrt{n}}J_1\left(\frac{4\pi\sqrt{mn}}{c}\right)\exp\left(-\frac{2\pi n}{\sqrt{M}}\right)$$

$$T_0(d) = \sum_{n=1}^{+\infty}\frac{S(m, nM^{-1}; d)}{\sqrt{n}}J_1\left(\frac{4\pi\sqrt{mn}}{c\sqrt{M}}\right)\exp\left(-\frac{2\pi n}{\sqrt{M}}\right)$$

as the integral of e^{-t} on $[0, +\infty[$ is equal to 1 like the one of E_1. Thus, by similar computations,

$$\frac{\langle a_m, L\rangle_N^-}{4\pi} = e^{-2\pi m/\sqrt{N}} + mg(m)\left(O_1\left(\frac{86}{N} + \frac{43}{\sqrt{N}}\right)\right).$$

Gathering those bounds, we get for all m prime to N,

$$\frac{\langle a_m, L'\rangle_{N^2}^{+,\text{new}}}{4\pi} = E_1\left(\frac{2\pi m}{N}\right) - \frac{E_1\left(\frac{2\pi m}{\sqrt{N}}\right)}{N-1} - \frac{\ln(N)e^{-2\pi m/\sqrt{N}}}{2(N-1)} \qquad (37)$$

$$+mg(m)O_1\left(\frac{86}{N^2} + \frac{43}{N} + \frac{\ln(N)/2+1}{N-1}\left(\frac{86}{N} + \frac{43}{\sqrt{N}}\right)\right) \qquad (38)$$

and slightly better ones for $m = 2$ coming from refinements above (it suffices to replace $86mg(m)$ by 213 and $43mg(m)$ by 97 above).

By computations on Sage, we deduce the following first estimates.

Proposition 5 *With the previous estimates, one finds*

$\langle a_1, L'\rangle_N^+ > 0$ for $N \geq 1213$	$\langle a_1, L'\rangle_{N^2}^{+,\text{new}} > 0$ for $N \geq 47$
$\langle a_2, L'\rangle_N^+ > 0$ for $N \geq 5437$	$\langle a_2, L'\rangle_{N^2}^{+,\text{new}} > 0$ for $N \geq 97$
$\frac{\langle a_2, L'\rangle_N^+}{\langle a_1, L'\rangle_N^+} \in]0, 1[$ for $N \geq 45341$	$\frac{\langle a_2, L'\rangle_{N^2}^{+,\text{new}}}{\langle a_1, L'\rangle_{N^2}^{+,\text{new}}} \in]0, 1[$ for $N \geq 269$.

Hence, Lemma 14 applies and Theorem 2 is true for $N \geq 45341$ for $X_0^+(N)$ and for $N \geq 269$ for $X_{\text{ns}}^+(N)$.

For $M = N$, the estimates of $\langle a_m, L' \rangle_N$ are readily obtained, but the slowness of convergence is much more visible. This is mainly due to the fact that the error term is in $m/\sqrt{N}$ instead of m/N.

5.3 Improving the estimates for prime level

To attain from $N \geq 45341$ a range where all remaining primes can be checked by a different method, one needs to improve upon the worst error term appearing in $\langle a_m, L' \rangle_N^+$, which is in $m/\sqrt{N}$ and comes from the estimates of $\mathcal{T}(d)$ after looking at (33).

The following arguments rely on cancellations of Kloosterman sums not exploited by the Weil bounds. For $d = 1$, the Kloosterman sum is always 1 (see the convention) so this case has to be dealt with separately. A careful analysis proves that

$$0.4\sqrt{m} \leq \mathcal{T}(1) \leq \sqrt{m},$$

which will slightly improve the bounds later.

Assume now that $d \geq 2$. The main term contributing to the bound is $E_1(2\pi n/\sqrt{N})$, hence we write

$$\mathcal{T}(d) = \mathcal{T}_M(d) + \mathcal{T}_R(d),$$

where $\mathcal{T}_M(d)$ is the sum of terms for which $n \leq 3\sqrt{N}/\pi$ and $\mathcal{T}_R(d)$ is the remainder.

By the Weil bounds, using the fact that the integral of E_1 on $[5, +\infty[$ is less than 10^{-4}, we obtain

$$2\pi\sqrt{m} \sum_{d \geq 2} \left| \frac{\mathcal{T}_R(d)}{d\sqrt{N}} \right| \leq 10^{-4} \frac{\lambda_m}{\sqrt{N}}$$

where $\lambda_m = 43$ for $m = 1$ and 97 for $m = 2$ as before, so this contribution will be very small. For $\mathcal{T}_M(d)$, we will exploit Polyà-Vinogradov-type estimates ([46], Lemma 5.9).

Proposition 6 *For every $d > 1$, every k invertible modulo d and every $m, K, K' \in \mathbb{N}$,*

$$\left| \sum_{n=K}^{K'} S(m, nk; d) \right| \leq \frac{4d}{\pi^2} (\log(d) + 1.5).$$

Now, assume $N \geq 1000$, so that for $m = 1$ or 2 and $n \leq 5\sqrt{N}/(2\pi)$, $4\pi\sqrt{mn}/(d\sqrt{N}) \leq 1.5$. This implies that in the considered range for n, the function $t \mapsto J_1(4\pi\sqrt{mt}/(d\sqrt{N}))/\sqrt{t}E_1(2\pi t/\sqrt{N})$ is decreasing and positive (as the product of two such functions). Its total variation on $[1, 5\sqrt{N}/2\pi]$ is then bounded by its first value (itself controlled by $E_1(2\pi/\sqrt{N})/2$).

By Abel transform and the previous proposition, we thus obtain

$$|\mathcal{T}_M(d)| \leq \frac{8}{\pi} \frac{\sqrt{m}}{\sqrt{N}} (\log(d) + 1.5) E_1 \left(\frac{2\pi}{\sqrt{N}} \right).$$

Compared to Weil bounds in Lemma 17, the new bound is approximately the best for $d \leq f(N) = \lfloor N/(2.5^2 E_1(2\pi/\sqrt{N})^2) \rfloor$. We then obtain

$$2\pi \sqrt{m} \left| \sum_{d=2}^{f(N)} \frac{\mathcal{T}_M(d)}{d\sqrt{N}} \right| \leq \frac{16m}{N} E_1 \left(\frac{2\pi}{\sqrt{N}} \right) \sum_{d=2}^{f(N)} \frac{\log(d) + 1.5}{d}$$

$$\leq \frac{8m}{N} E_1 \left(\frac{2\pi}{\sqrt{N}} \right) \left(\log(f(N))^2 + 3\log(f(N)) + 1 \right)$$

with lemma 5.11 of [46]. By Weil bounds and the same lemma, for $m = 1$,

$$2\pi \left| \sum_{d=f(N)+1}^{+\infty} \frac{\mathcal{T}_M(d)}{d\sqrt{N}} \right| \leq \frac{4\pi}{\sqrt{Nf(N)}} (\log(f(N)) + 4) \tag{39}$$

and for $m = 2$,

$$2\pi \sqrt{2} \left| \sum_{d=f(N)+1}^{+\infty} \frac{\mathcal{T}_M(d)}{d\sqrt{N}} \right| \leq \frac{8\pi(2 - 1/\sqrt{2})}{\sqrt{Nf(N)}} (\log(f(N)) + 4). \tag{40}$$

Combining these arguments, we get, for $N \geq 1000$,

$$\frac{\langle a_1, L' \rangle_N^+}{4\pi} \geq E_1 \left(\frac{2\pi}{\sqrt{N}} \right) - \frac{6.3}{\sqrt{N}} - \frac{86}{N} - 2\pi \left| \sum_{d=2}^{+\infty} \frac{\mathcal{T}_M(d)}{d\sqrt{N}} \right|$$

and

$$\frac{\langle a_2, L' \rangle_N^+}{4\pi} \geq E_1 \left(\frac{4\pi}{\sqrt{N}} \right) - \frac{6.3\sqrt{2}}{\sqrt{N}} - \frac{213}{N} - 2\pi\sqrt{2} \left| \sum_{d=2}^{+\infty} \frac{\mathcal{T}_M(d)}{d\sqrt{N}} \right|$$

and finally

$$\langle a_1, L' \rangle_N^+ > 0 \quad \text{and} \quad \frac{\langle a_2, L' \rangle_N^+}{\langle a_1, L' \rangle_N^+} \in]0, 1[$$

for $N \geq 8641$, which is much more reasonable than 45341.

The same improvements for the bounds apply exactly for $M = N^2 \geq 1000$, thus allowing us to replace the estimate in $43/N$ in (37) by the same expressions as above with $f(M)$ instead of $f(N)$.

One gets that $\langle a_2, L'\rangle_{N^2}^{+,\text{new}} > 0$ for $N \geq 71$ instead of 97, and that

$$\frac{\langle a_2, L'\rangle_{N^2}^{+,\text{new}}}{\langle a_1, L'\rangle_{N^2}^{+,\text{new}}} \in \,]0, 1[$$

for $N \geq 151$.

We now discuss how to deal with the remaining cases, namely those for which $N \leq 8641$ and $g(X_0^+(N)) \geq 2$, and those for which $N \leq 151$ and $g(X_{\text{ns}}^+(N)) \geq 2$.

The most natural approach is the following: for any small N, compute a basis of eigenforms for $S_2(\Gamma_0(M))^{+,\text{new}}$, and for every f (normalised) in this basis, compute $L'(f, 1)$ up to sufficient precision to ensure that $L'(f, 1) \neq 0$.

Recall that by ([32], Corollary V.1.3), if $L'(f, 1) \neq 0$ under the same assumptions, the same is true for the Galois conjugate eigenforms, so only one check needs to be performed for the Galois orbit. Theorem 2 requires exactly that the sum of sizes of those Galois orbits is at least 2, so we only need to check that for two Galois orbits of size 1 (or one of size at least 2), one has $L'(f, 1) \neq 0$.

We have performed these verifications in MAGMA, and obtained the following.

• For any prime $N \leq 2000$ such that $X_0^+(N)$ is of genus at least two, there are at least two distincts normalised newforms such that $L'(f, 1) \neq 0$, hence Theorem 2 holds. In fact, we have also checked that for all such N, $L'(f, 1) \neq 0$ for *all* the eigenforms in $S_2(\Gamma_0(N))^+$, therefore by Proposition 8, rank $J_0^+(N)(\mathbb{Q}) = \dim J_0^+(N)$ unconditionally for all those small primes.

• Similarly, for any prime $N \leq 53$ such that $X_{\text{ns}}^+(N)$ is of genus at least two, $L'(f, 1) \neq 0$ for *all* the eigenforms in $S_2(\Gamma_0(N^2))^{+,\text{new}}$, therefore by the same arguments, rank $\text{Jac}(X_{\text{ns}}^+(N))(\mathbb{Q}) = \dim \text{Jac}(X_{\text{ns}}^+(N))$ for all those small primes.

Unfortunately, these algorithms require explicit embeddings of the fields of coefficients K_f of f into $\mathbb{C}$, which makes them very slow when N becomes larger than 2000 (then, the degree of K_f can be larger than 100). We thus could not complete the argument by using only this method, let us explain how to deal with the intermediary range $N \in [2000, 9000]$ for $X_0^+(N)$ and $N \in [59, 151]$ for $X_{\text{ns}}^+(N)$.

The idea is to look at the simple quotients of the two relevant Jacobians which are elliptic curves. If there are none, in this range, we have proved that $\langle a_1, L'\rangle_M^{+,\text{new}} \neq 0$ so we must have f such that $L'(f, 1) \neq 0$, and it generates a simple quotient of dimension at least 2 by hypothesis, so we are done.

Now, if there *are* elliptic curves in there, it is sufficient to find two of them of rank 1 for the same reasons. Quotients of $J_0(M)^{+,\text{new}}$ of dimension 1 are in one-to-one correspondence with isogeny classes of elliptic curves of conductor N and root number -1 (the fact that this correspondence is surjective is a consequence of Cremona's tables in this range but also a particular case of modularity theorems).

One can thus eliminate all levels N except the ones for which there exists exactly one (up to isogeny) elliptic curve E of analytic rank 1 and conductor N. Using Cremona's tables, we obtain a list of respectively 70 ($M = N$) and 7 ($M = N^2$) possible exceptions, namely N in $\{61, 67, 73, 101, 109, 113\}$ for the latter.

Now, we use a last argument: if the modular form f_E associated to E is really the only one such that $L'(f, 1) \neq 0$ in the space, one should have

$$\langle a_1, L' \rangle_M^{+,\text{new}} = \frac{L'(E, 1)}{\| f_E \|^2}$$

(the fact that this equality holds without a normalisation factor comes from the Manin constant being equal to 1 here, which is true in this range by results of Cremona).

Now, the left-hand side is larger than 4/5 for $M = N$, $N \geq 2000$ and than 1/2 for $M = N^2$, $N \geq 53$ by the (optimised) lower bounds given above, and the right-hand side is computable in terms of periods of E. Using this idea turns out to eliminate all remaining possible exceptions in both cases of M, which concludes the proof.

Remark 9 In some sense, this heuristic is natural: all terms in the sum defined by $\langle a_1, L' \rangle_M^{+,\text{new}}$ are positive (another consequence of Gross–Zagier formula), hence there is no cancellation among those, and the idea is that one of them alone cannot be enough to approach the estimates given for the sum.

6 Appendix: Chow–Heegner points and Ceresa cycles

In this appendix we explain how Lemma 4 is a consequence of Hain and Matsumoto's work relating the extension [Lie(U_2)] to the Ceresa cycle.

6.1 Ceresa cycles and Gross–Kudla–Schoen cycles

We recall some properties of modified diagonal cycles studied in [17,31] and [19]. As our discussion applies in fairly broad generality, we take X to be a smooth geometrically irreducible projective curve over a field K of characteristic zero. Let π_S denote the projection

$$X^n \to X^{\#S}$$

defined by projecting onto the coordinates in S as in (7). The *Gross–Kudla–Schoen* cycle is defined to be

$$\Delta_{GKS} := \sum_{\emptyset \neq S \subset \{1,2,3\}} (-1)^{\#S-1} X_S,$$

where X_S is as defined in section 2.2.

It defines an element of the group $\mathrm{CH}^2(X^3)$ of codimension two cycles in the triple product $X \times X \times X$. By [31, Proposition 3.1], the class of Δ_{GKS} lies in the subspace $\mathrm{CH}_0^2(X^3)$ of homologically trivial cycles.

Now let $Z \subset X \times X$ be a correspondence, and let

$$\Pi_Z : \mathrm{CH}^2(X^3) \to \mathrm{CH}^1(X)$$

be the composite map

$$\mathrm{CH}^2(X^3) \xrightarrow{\pi^*_{\{1,2,3\}}} \mathrm{CH}^2(X^4) \xrightarrow{\cdot(Z \times X^2)} \mathrm{CH}^4(X^4) \xrightarrow{(\pi_4)_*} \mathrm{CH}^1(X),$$

where the second map is the intersection product with $Z \times X^2 \subset X^4$.

Lemma 18 ([19] Lemma 2.1) *We have*

$$D_Z(b) = \Pi_Z(\Delta_{GKS}).$$

6.2 The Gross–Kudla–Schoen cycle and the Ceresa cycle

Since $[\Delta_{GKS}]$ is homologically trivial, it has (Sect. 2.1) an étale Abel–Jacobi class

$$\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}]) \in H^1(G_K, H^3_{\text{ét}}(X^3_{\overline{K}}, \mathbb{Q}_p(2))).$$

By [31, Corollary 2.6], the cycle class $\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}])$ lies in the image of the Kunneth projector

$$\begin{aligned} P_{e*} : H^1(G_K, H^3_{\text{ét}}(X^3_{\overline{K}}, \mathbb{Q}_p(2))) &\to H^1(G_K, H^1_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p)^{\otimes 3}(2)) \\ &\simeq H^1(G_K, V^{\otimes 3}(-1)) \\ &\hookrightarrow H^1(G_K, H^3_{\text{ét}}(X^3_{\overline{K}}, \mathbb{Q}_p(2))), \end{aligned}$$

and hence may be thought of as an element of $H^1(G_K, V^{\otimes 3}(-1))$ (here $V := H^1_{\text{ét}}(X_{\overline{K}}, \mathbb{Q}_p(1))$). The action of S_3 on X^3 induces an action on $V^{\otimes 3}(-1)$, which is given by $\epsilon \otimes \sigma$, where ϵ is the sign of a permutation and σ is the natural action of S_3 on $V^{\otimes 3}$. Since Δ_{GKS} is invariant under the S_3 action, it lies in the image of

$H^1(G_K, \wedge^3 V(-1))$ under the map induced by the inclusion

$$\iota : \wedge^3 V \rightarrow V^{\otimes 3}$$

$$v_1 \wedge v_2 \wedge v_3 \mapsto \frac{1}{6} \sum_{\tau \in S_3} \epsilon(\tau) v_{\tau(1)} \otimes v_{\tau(2)} \otimes v_{\tau(3)}. \tag{41}$$

For the relations to fundamental groups, it will be helpful to recall the relation between Δ_{GKS} and the *Ceresa cycle*. By [31, Proposition 5.3], the image of Δ_{GKS} in $\mathrm{CH}^{g-1}(J)$ under the map

$$\mu : X^3 \rightarrow J$$

$$(x_i) \mapsto \sum [x_i] - 3[b]$$

is rationally equivalent to

$$([3]_* - 3[2]_* + 3[1] * -3[0]_*)\mathrm{AJ}(X).$$

The *Ceresa cycle* C_b is defined to be

$$\mathrm{AJ}(X) - [-1]_* \mathrm{AJ}(X) \in \mathrm{CH}^{g-1}(J).$$

Proposition 7 (Colombo–van Geemen, [17], Proposition 2.9) *We have*

$$\mathrm{AJ}_{\text{ét}}(\mu_*(\Delta_{GKS})) = 3\mathrm{AJ}_{\text{ét}}([C_b])$$

in $H^1(G_K, \wedge^3 V(-1))$.

We first recall Hain and Matsumoto's description of the Galois action on U_2. We again take X to be a smooth projective geometrically irreducible curve over a field K of characteristic zero. The group U_2 is an extension

$$1 \rightarrow \mathrm{Ker}(H^2(J_{\overline{\mathbb{Q}}}, \mathbb{Q}_p) \xrightarrow{\mathrm{AJ}^*} H^2(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_p))^* \rightarrow U_2 \rightarrow V \rightarrow 1, \tag{42}$$

with $V = T_p J \otimes \mathbb{Q}_p$ again. We define

$$\overline{\wedge^2 V} := \mathrm{Ker}(H^2(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_p) \xrightarrow{\mathrm{AJ}^*} H^2(J_{\overline{\mathbb{Q}}}, \mathbb{Q}_p))^*,$$

and write the image of $v_1 \wedge v_2$ in $\overline{\wedge^2 V}$ as $\overline{v_1 \wedge v_2}$. Taking the Lie algebra L_2 of U_2, we obtain an element $[L_2] \in \mathrm{Ext}^1_{G_K}(V, \overline{\wedge^2 V})$, or equivalently an element of $H^1(G_K, V(-1) \otimes \overline{\wedge^2 V})$. The following theorem of Hain and Matsumoto characterises this extension class in terms of the Gross–Kudla–Schoen cycle.

Theorem 5 (Hain–Matsumoto [34], Theorem 3) *Let* $\alpha : \wedge^3 V \to V \otimes \overline{\wedge^2 V}$ *be the injective homomorphism*

$$v_1 \wedge v_2 \wedge v_3 \mapsto v_1 \otimes (\overline{v_2 \wedge v_3}) + v_2 \otimes (\overline{v_3 \wedge v_1}) + v_3 \otimes (\overline{v_1 \wedge v_2}).$$

Then $[L_2] \in H^1(G_K, V(-1) \otimes \overline{\wedge^2 V})$ *is equal to* $\alpha(-1)_*(\mathrm{AJ}_{\text{ét}}[C_b])$, *where* $[C_b]$ *is the class of the Ceresa cycle in* $\mathrm{CH}^{g-1}(J)$, *and* $\mathrm{AJ}_{\text{ét}}([C_b])$ *is its image in* $H^1(G_K, \wedge^3 V(-1))$.

Via the relation between the Ceresa cycle and the Gross–Kudla–Schoen cycle, this has the following corollary.

Corollary 3 *The extension class* $[L_2] \in H^1(G_K, V(-1) \otimes \overline{\wedge^2 V})$ *is equal to the image of* $\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}])$ *under the map*

$$H^1(G_K, V^{\otimes 3}) \to H^1(G_K, V \otimes \overline{\wedge^2 V})$$

induced by the quotient

$$\tau : V^{\otimes 3} \to V \otimes \overline{\wedge^2 V}$$
$$v_1 \otimes v_2 \otimes v_3 \mapsto v_1 \otimes \overline{v_2 \wedge v_3}.$$

Proof Let $\iota : \wedge^3 V \to V^{\otimes 3}$ be the inclusion (41), and $\tau' : V^{\otimes 3} \to \wedge^3 V$ the quotient map $v_1 \otimes v_2 \otimes v_3 \mapsto v_1 \wedge v_3 \wedge v_3$. By Proposition 7, the image of $\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}])$ in $H^1(G_K, \wedge^3 V(-1))$ under τ'_* is equal to $\frac{1}{3}\mathrm{AJ}_{\text{ét}}([C_b])$. Since $\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}])$ lies in the image of ι_*, and

$$\alpha = 3\tau \circ \iota,$$

we have

$$\alpha_* \circ \tau'_*[\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}])] = 3\tau_*[\mathrm{AJ}_{\text{ét}}([\Delta_{GKS}])] \in H^1(G_K, V(-1) \otimes \overline{\wedge^2 V}).$$

Hence we deduce from Theorem 5 that

$$[L_2] = \frac{1}{3}\alpha_* \circ \tau'_*[\Delta_{GKS}] = \tau_*[\Delta_{GKS}].$$

$\square$

We now return to the case where $K = \mathbb{Q}$. Via the commutative diagram

$$
\begin{array}{ccc}
\mathrm{NS}(J_{\mathbb{Q}}) & \xrightarrow{\ \mathrm{AJ}^*\ } & \mathrm{NS}(X_{\mathbb{Q}}) \\
\downarrow{\scriptstyle c} & & \downarrow{\scriptstyle c} \\
H^2_{\text{ét}}(J_{\overline{\mathbb{Q}}}, \mathbb{Q}_p(1)) & \xrightarrow{\ \mathrm{AJ}^*\ } & H^2_{\text{ét}}(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_p(1)),
\end{array}
$$

(where c denotes the Chern class), we hence obtain a homomorphism

$$\mathrm{Ker}(\mathrm{NS}(J_{\mathbb{Q}}) \to \mathrm{NS}(X_{\mathbb{Q}})) \to \mathrm{Ext}^1(V, \mathbb{Q}_p(1)).$$
$$[\mathcal{L}] \mapsto [c([\mathcal{L}])^*([L_2])],$$

where $L_2 := \mathrm{Lie}(U_2)$. The extensions obtained come from points on J. They can be related to the Gross–Kudla–Schoen cycle via the theorem of Hain and Matusmoto (the argument given below follows Darmon, Rotger and Sols [19], who prove a Hodge theoretic analogue of the Lemma below using, using the theorems of Harris and Pulte, which are Hodge theoretic analogues of the Hain–Matsumoto theorem).

Lemma 19 *Let $Z \subset X \times X$ be a codimension 1 cycle. Let $i_1, i_2, i_3 \colon X \hookrightarrow X \times X$ be the closed immersions defined by the subschemes $\{b\} \times X$, $X \times \{b\}$ and the diagonal Δ_X of $X \times X$ respectively. For $j = 1, 2, \{1, 2\}$, let i_j^* denote the pull-back morphism*

$$\mathrm{CH}^1(X \times X) \to \mathrm{CH}^1(X).$$

Then the extension class in $H^1(G_K, V)$ associated to the Lie algebra L_Z is given by $\mathrm{AJ}_{\text{ét}}(D_Z(b))$, with $D_Z(b)$ as in (16).

Proof The class $[L_Z]$ is the image of $[L_2]$ under the morphism

$$\mathrm{Ext}^1_{G_K}(V, \overline{\wedge^2 V}) \to \mathrm{Ext}^1_{G_{\mathbb{Q}}}(V, \mathbb{Q}_p(1))$$

induced by $\pi_Z \colon \overline{\wedge^2 V} \to \mathbb{Q}_p(1)$. We have a commutative diagram

$$
\begin{array}{ccc}
\mathrm{CH}^2(X^3)_0 & \xrightarrow{\mathrm{AJ}_{\text{ét}}} & H^1(G_{\mathbb{Q}}, V^{\otimes 3}(-1)) \\
\downarrow{\scriptstyle \Pi_Z} & & \downarrow{\scriptstyle \Pi_{Z*}} \\
\mathrm{Pic}^0(X) & \xrightarrow{\mathrm{AJ}_{\text{ét}}} & H^1(G_{\mathbb{Q}}, V).
\end{array}
$$

By Theorem 5, the extension class $[L_2]$ is given by $\mathrm{AJ}_{\text{ét}}(\Delta_{GKS})$, hence

$$[L_Z] = \Pi_{Z*}([L_2]) = \mathrm{AJ}_{\text{ét}}(D_Z(b)),$$

by Lemma 18. $\qquad\square$

7 Appendix: Proof of the Kolyvagin–Logachev type result

In this appendix, we fix the following notation:
- M is a fixed odd level (which for our applications will be N or N^2)
- $f \in S_2(\Gamma_0(M))^{+,\mathrm{new}}$ is a normalised eigenform.
- $A = A_f$ is its associated quotient of $J_0(M)$, together with the canonical projection $\pi \colon J_0(M) \to A$ (independent of the choice of f in its Galois orbit).

We explain here the following result, attributed to Kolyvagin and Logachev.

Proposition 8 (Rank 1 BSD for modular abelian varieties) *If $L'(f, 1) \neq 0$, the rank of $A(\mathbb{Q})$ is exactly $g := \dim A$.*

Corollary 4 *If $L'(f, 1) \neq 0$ for at least two distinct newforms f, for the Heegner quotient A of $J_0(M)^{+,new}$ (Definition 4),*

$$\mathrm{rk}(A) = \dim(A) = \rho(A) \geq 2.$$

Proof of the Corollary By Proposition 8 the rank of A is equal to its dimension as it is true for each of its factors A_f. Now, we recall that all endomorphisms of an A_f are symmetric and the latter is of GL_2-type, in particular $\mathrm{End}^\dagger(A_f)$ is of rank $\dim A_f$ (see Sect. 4.1). Finally, for f, g non Galois conjugates, there is no morphism between A_f and A_g (by multiplicity one in the newpart) so the endomorphism ring splits and we get the last equality. □

Remark 10 This result is well-known if $\dim A = 1$ ([43] for the original reference, [29] for a survey), and proven in much greater generality in [54], all these along the lines of a stronger result in the rank zero case proved in [44]. It is also (a slightly weaker version of) the main result in Tian's thesis [64] and of a paper of Tian and Zhang in preparation [65] for which we could not find quotable material. In any case, we felt it sufficiently different from the former references (to which we borrow constantly) to deserve a proof for the nonexperts. For the same reasons, we will simply refer to those papers for parts of the proofs which generalise seamlessly and focus on the more technical points.

Convention We use a well-chosen prime number p to obtain Proposition 8. As we only need one such p, in all this appendix, when a property holds when p is large enough, we then automatically assume it is without further mention.

We will prove Proposition 8 by reducing it successively to other statements which will be emphasized.

Notation Throughout this text, τ denotes the usual complex conjugation and when it acts on an $\mathbb{Z}$-module $\mathcal{M}$, $\mathcal{M}^+$ and $\mathcal{M}^-$ denote the spaces of $m \in \mathcal{M}$ respectively fixed and reversed by τ. If $\mathcal{M}$ is finite of odd order, $\mathcal{M} = \mathcal{M}^+ \oplus \mathcal{M}^-$, which we will frequently use implicitly.

Given an Galois extension L/K of number fields and $\mathfrak{P}$ a prime ideal of L unramified over $\mathfrak{p}$, $(\mathfrak{P}, L/K)$ denotes the Frobenius of $\mathfrak{P}$ for this extension, and $(\mathfrak{p}, L/K)$ the conjugacy class of such Frobenius's in $\mathrm{Gal}(L/K)$.

7.1 Structure of the *p*-torsion and reduction to Selmer groups

Let K_f be the number field of coefficients of f. By [44, section 2.1], there is an isomorphism $[\cdot] : K_f \to \mathrm{End}_\mathbb{Q} A \otimes \mathbb{Q}$ such that for every prime $\ell \nmid N$, $[a_\ell(f)] \in \mathrm{End}_\mathbb{Q} A$ and

$$[a_\ell(f)] \circ \pi = \pi \circ T_\ell. \tag{43}$$

The inverse image of $\mathrm{End}_{\mathbb{Q}}\, A$ is thus an order in K_f denoted by $\mathcal{O}$, and A is endowed with a structure of $\mathcal{O}$-module.

We now fix p an odd prime totally split in K_f and prime to the conductor of $\mathcal{O}$ (there are infinitely many such primes by Cebotarev density theorem), so that $p\mathcal{O} = \mathfrak{P}_1 \ldots \mathfrak{P}_g$ as a decomposition into prime ideals. In all the following, the notation $\mathfrak{P}$ will run through $\mathfrak{P}_1, \ldots, \mathfrak{P}_g$.

Remark 11 It is likely the proof still holds for any type of decomposition of p but this hypothesis makes the exposition much more symmetric (and there are infinitely many of such p's so we can choose it as large as necessary). In the opposite situation, if there is an inert prime in K_f, the proof should be a bit simpler.

One of the key ideas to get closer to the case of elliptic curves is decomposing every structure of $\mathcal{O}/(p)$-modules using those prime ideals. Our tool is the following Lemma, often used without mention.

Lemma 20 *By the Chinese remainder theorem, $\mathcal{O}/(p) \cong \bigoplus_{\mathfrak{P}} \mathcal{O}/\mathfrak{P}$, in particular each $\mathcal{O}/\mathfrak{P}$ is projective and flat over $\mathcal{O}/(p)$. Every $\mathcal{O}/(p)$-module $\mathcal{M}$ splits canonically into sub-$\mathcal{O}/(p)$-modules*

$$\mathcal{M} = \bigoplus_{\mathfrak{P}} \mathcal{M}[\mathfrak{P}], \quad \mathcal{M}[\mathfrak{P}] = \{m \in \mathcal{M}, \ \mathfrak{P} \cdot m = 0\} \cong \mathcal{M}/\mathfrak{P}\mathcal{M},$$

and projections are given by elements of $\mathcal{O}$. All these isomorphisms are canonical, and for every $m \in \mathcal{M}$, we will denote by $m_{\mathfrak{P}}$ its projection onto $\mathcal{M}[\mathfrak{P}]$ (or in $\mathcal{M}/\mathfrak{P}\mathcal{M}$ depending on the context).

Proof The $\mathfrak{P}$ are pairwise coprime so the Chinese remainder theorems holds, and tensoring $\mathcal{M}$ by $\mathcal{O}/(p)$ on one hand fixes it and the other one decomposes it canonically into $\bigoplus_{\mathfrak{P}} \mathcal{M}/\mathfrak{P}\mathcal{M}$. The latter clearly identifies each $\mathcal{M}/\mathfrak{P}\mathcal{M}$ with the $\mathfrak{P}$-torsion part of $\mathcal{M}$, and the other statements follow. $\square$

The $\mathcal{O}$-linear representation $A[p]$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ thus splits into $\bigoplus_{\mathfrak{P}} A[\mathfrak{P}]$ and for any extension L of $\mathbb{Q}$, we have canonical isomorphisms of $\mathcal{O}/(p)$-modules

$$(A(L)/pA(L))[\mathfrak{P}] \cong A(L)/\mathfrak{P}A(L) \quad H^1(L, A[p])[\mathfrak{P}] \cong H^1(L, A[\mathfrak{P}]). \quad (44)$$

If L is a number field, for every place v of L, the natural localisation maps loc_v give rise to a commutative diagram

$$
\begin{array}{ccccccccc}
0 & \longrightarrow & A(L)/\mathfrak{P}A(L) & \overset{\delta}{\longrightarrow} & H^1(L, A[\mathfrak{P}]) & \longrightarrow & H^1(L, A)[\mathfrak{P}] & \longrightarrow & 0 \\
& & \downarrow{\scriptstyle \mathrm{loc}_v} & & \downarrow{\scriptstyle \mathrm{loc}_v} & & \downarrow{\scriptstyle \mathrm{loc}_v} & & \\
0 & \longrightarrow & A(L_v)/\mathfrak{P}A(L_v) & \overset{\delta_v}{\longrightarrow} & H^1(L_v, A[\mathfrak{P}]) & \longrightarrow & H^1(L_v, A)[\mathfrak{P}] & \longrightarrow & 0,
\end{array} \quad (45)
$$

inherited by flatness from the commonly known analogous diagram for the ideal (p) (for references on those facts and the Selmer groups, see [35, Appendix C.4]). Let us

define the $\mathfrak{P}$-Selmer group as

$$\mathrm{Sel}_{\mathfrak{P}}(L, A) := \{s \in H^1(L, A[\mathfrak{P}]), \forall v, \mathrm{loc}_v\, s \in \delta_v(A(K_v)/\mathfrak{P}A(K_v))\}, \quad (46)$$

again canonically identified to $\mathrm{Sel}_p(L, A)[\mathfrak{P}]$ hence fitting by the same arguments into the exact sequence

$$0 \longrightarrow A(L)/\mathfrak{P}A(L) \overset{\delta}{\longrightarrow} \mathrm{Sel}_{\mathfrak{P}}(L, A) \longrightarrow \mathrm{III}(L, A)[\mathfrak{P}] \longrightarrow 0. \quad (47)$$

Now, consider an imaginary quadratic field K whose discriminant $D_K < -4$ is squarefree, prime to the level M and a square modulo M. These conditions guarantee that there is a Heegner point (we fix definitively $\mathfrak{n}$ and $[\mathfrak{a}_0]$)

$$x = (\mathcal{O}_K, \mathfrak{n}, [\mathfrak{a}_0]) \in X_0(M)(H) \quad (48)$$

in the notation of [28], where H is the Hilbert class field of K. As $f_{|w_M} = f$, $\pi \circ w_M = \pi$ therefore by elementary properties of Heegner points [28, formulas (4.1) to (5.2)], for $y_1 = \pi((x) - (\infty)) \in A(H)$, one has

$$y_K := \mathrm{Tr}_{H/K}\, y_1 = \pi \left(\sum_{[\mathfrak{a}] \in \mathrm{Cl}(K)} (\mathcal{O}_K, \mathfrak{n}, [\mathfrak{a}]) - h_K(\infty) \right) \in A(K), \quad (49)$$

$$\tau(y_K) = \pi \left(\sum_{[\mathfrak{a}] \in \mathrm{Cl}(K)} w_M \cdot (\mathcal{O}_K, \mathfrak{n}, [\mathfrak{a}]) - h_K(\infty) \right) \in y_K + A(\mathbb{Q})_{\mathrm{tors}}, \quad (50)$$

Now, using a theorem of Waldspurger [66, Théorème 2.3], let us fix once and for all a K such that $L(f \otimes \varepsilon_K, 1) \neq 0$ where ε_K is the Dirichlet character associated to K. By Gross–Zagier formula ([32], Theorem I.6.3), the point y_K is then nontorsion in $A(K)$ and has an integer multiple in $A(\mathbb{Q})$ by (50). The subgroup $\mathcal{O} \cdot y_K$ is thus a subgroup of $A(K)$ of rank g (as nonzero elements of $\mathcal{O}$ act by isogenies), which leads us to the following.

Reduction 1 *Prove that $\mathcal{O} \cdot y_K$ is of finite index in $A(K)$.*

Now, for p large enough,

$$y_K \notin \mathfrak{P}A(K) \text{ for all } \mathfrak{P}, \quad (51)$$

which further leads by (47) to

Reduction 2 *Prove that for all $\mathfrak{P}$, $\delta(\overline{y_K})$ generates $\mathrm{Sel}_{\mathfrak{P}}(K, A)$.*

Proof If this claim holds, every $\mathrm{Sel}_{\mathfrak{P}}(K, A)$ is an $\mathcal{O}/\mathfrak{P} \cong \mathbb{F}_p$-vector space of dimension 1, so $A(K)/\mathfrak{P}A(K)$ is of dimension at most 1 by (47), and

$$A(K)/pA(K) \cong \bigoplus_{\mathfrak{P}} A(K)/\mathfrak{P}A(K)$$

is of dimension at most g over $\mathbb{F}_p$. This imposes that the Mordell–Weil rank of $A(K)$ over $\mathbb{Z}$ is at most g, hence the equality using $\mathcal{O} \cdot y_K$. □

To conclude this paragraph, τ acts naturally on $A(\overline{\mathbb{Q}})$, $A[\mathfrak{P}]$, $H^1(K, A[\mathfrak{P}])$ and $\mathrm{Sel}_{\mathfrak{P}}(K, A)$, and the action of $\mathcal{O}$ and the morphisms between those in (44) and (45) are τ-equivariant. We fix from now on a polarisation $A \to \widehat{A}$ of degree prime to p (otherwise choose a larger prime p), which thus defines a Weil pairing $A[p] \times A[p] \to \mu_p$. Its elementary properties [51, Lemma 16.2] then imply the following structural result, crucial for our understanding.

Lemma 21 *For every $\mathfrak{P}$ and $\varepsilon = \pm 1$, the following are true.*
- *The $2g$ spaces $A[\mathfrak{P}]^\varepsilon$ are pairwise orthogonal for the Weil pairing, except the $A[\mathfrak{P}]^\varepsilon$ with the same $\mathfrak{P}$ and opposite sign.*
- *The two spaces $A[p]^\varepsilon$ are isotropic for the Weil pairing.*
- *Each $A[\mathfrak{P}]^\varepsilon$ is then of dimension 1 over $\mathbb{F}_p$ and $\dim_{\mathbb{F}_p} A[\mathfrak{P}] = 2$.*

7.2 Pairing the Galois group and Selmer groups, and Kolyvagin primes

Throughout this appendix, we fix

$$L := K(A[\mathfrak{P}]), \quad G := \mathrm{Gal}(L/K)$$

(notice L is Galois over $\mathbb{Q}$).

Proposition 9 *For p large enough:*
(a) $A[\mathfrak{P}]$ is (absolutely) irreducible as a representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.
(b) The canonical restriction morphism

$$H^1(K, A[\mathfrak{P}]) \overset{\mathrm{res}}{\to} H^1(L, A[\mathfrak{P}])^G = \mathrm{Hom}_G(\mathrm{Gal}(L^{\mathrm{ab}}/L), A[\mathfrak{P}])$$

is injective, with the action of G on $\mathrm{Gal}(L^{\mathrm{ab}}/L)$ defined by conjugation in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

Remark 12 Here is an important difference with the $\dim A = 1$ case: the Galois representation $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(A[\mathfrak{P}]) \cong \mathrm{GL}_2(\mathbb{F}_p)$ is not proven to be surjective ([57] does not cover the square M case), but we will manage with (a) and (b) although it introduces significant changes compared to some arguments in [29].

Proof (a) is Lemma 3.7 of [56] and (b) is Proposition 6.1.2 of [54]. □

We now choose S a finite sub-$\mathcal{O}$-module of $H^1(K, A[\mathfrak{P}])$, stable by τ (this will be first $\mathrm{Sel}_{\mathfrak{P}}(K, A)$ and then an auxiliary module for the proof). By Proposition 9 (b), there is a pairing

$$\begin{aligned} S \times \mathrm{Gal}(L^{\mathrm{ab}}/L) &\longrightarrow A[\mathfrak{P}] \\ (s, \sigma) &\longmapsto \mathrm{res}(s)(\sigma) \end{aligned}$$

which is injective on the left. We define L_S the extension of L whose absolute Galois group is the orthogonal of S, and thus obtain a nondegenerate pairing between finite

abelian p-torsion groups

$$[\cdot, \cdot]_S \colon S \times H_S \to A[\mathfrak{P}], \quad H_S := \mathrm{Gal}(L_S/L).$$

Keeping track of the actions of τ and the $\sigma \in G$, we have that

$$\tau[s, \rho]_S = [\tau(s), \tau\rho\tau^{-1}]_S, \quad \sigma[s, \rho] = [s, \sigma\rho\sigma^{-1}]. \tag{52}$$

In particular, the extension $L_S/\mathbb{Q}$ is Galois.

Lemma 22 *This pairing induces a perfect bilinear pairing from $S^\varepsilon \times H_S^+$ to $A[\mathfrak{P}]^\varepsilon \cong \mathbb{F}_p$, hence a duality between S^ε and H_S^+.*

Proof By (52), these two pairings (for $\varepsilon = \pm 1$) are well-defined, let us prove they are injective on the left and on the right, they will then be perfect as everything is finite(-dimensional). For $s \in S^\varepsilon$, if $[s, H_S^+]_S = 0$, then

$$[s, H_S]_S = [s, H_S^-]_S \subset A[\mathfrak{P}]^{-\varepsilon}$$

by the same arguments, but $[s, H_S]_S$ is stable by $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ by (52) again. As $A[\mathfrak{P}]$ is irreducible by Proposition 9 (*a*), it imposes $[s, H_S]_S = 0$ therefore $s = 0$ by nondegeneracy. Now, assume $[S^\varepsilon, h]_S = 0$ for some $h \in H_S^+$. This holds for all conjugates $\sigma h \sigma^{-1}$ of h in H_S by (52), so on the group $H' \subset H_S$ they generate. Again, this forces $[S, H']_S \subset A[\mathfrak{P}]^{-\varepsilon}$, but this group is stable by $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ hence $H' = 0$. $\square$

Lemma 23 *Fix $\varepsilon = \pm 1$ and I_S^+ a proper subgroup of H_S^+. Then, $s \in S^\varepsilon$ is 0 if for all $\rho \in H_S^+ \backslash I_S^+$, $[s, \rho]_S = 0$.*

Proof It is a trivial consequence of the perfect duality above, knowing that the sub-$\mathbb{F}_p$-vector space generated by $H_0^+ \backslash I_0^+$ is H_0^+ itself, e. g. by a counting argument. $\square$

    Reduction 3 *For all $\mathfrak{P}$, apply Lemma 23 to $(s_0 = 0, \varepsilon = -1)$ (resp. $\delta\overline{y_K}, \varepsilon = 1$) to prove that $\mathrm{Sel}_{\mathfrak{P}}(K, A)^- = 0$ (resp. $\mathrm{Sel}_{\mathfrak{P}}(K, A)^+ = \langle \delta\overline{y_K} \rangle$).*
    The next subsection will show us how to compute the pairing $[\cdot, \cdot]_S$.

7.3 Kolyvagin primes

Definition 5   – A *Kolyvagin prime* ℓ is a prime number such that:

    – ℓ does not divide $D_K M p$ (or the conductor of $\mathcal{O}$), so is unramified in L.
    – The conjugacy class of $(\ell, L/\mathbb{Q})$ is the one of τ in $\mathrm{Gal}(L/\mathbb{Q})$. In particular, $\ell\mathcal{O}_K =: \lambda_\ell$ is inert over ℓ. We will often shorten it to λ if ℓ is nonambiguous, and for any extension K' of K, $\lambda_{K'}$ will be a choice of prime ideal of $\mathcal{O}_{K'}$ above λ (in a consistent fashion if multiple extensions are considered).

– A *Kolyvagin number* n is a squarefree product of Kolyvagin primes ℓ.

In the same fashion as in [28, (3.3)], Kolyvagin primes have many strong properties.

Proposition 10 *For a Kolyvagin prime ℓ, λ splits completely in L. Furthermore:*

$$p|a_\ell(f), \quad p|\ell + 1$$

in $\mathcal{O}$, and all the points of $A[\mathfrak{P}]$ are defined over K_λ. Moreover, the two eigenspaces $(A(K_\lambda)/\mathfrak{P}A(K_\lambda))^\pm$ for the action of $\mathrm{Frob}(\ell)$ are of dimension 1 over $\mathbb{F}_p$.

Proof Up to conjugation, $(\lambda_L, L/K) = (\lambda_L, L/\mathbb{Q})^{f(\lambda/\ell)} = \tau^2 = \mathrm{Id}$ so λ_L/λ is totally split. Now, by Eichler-Shimura theory [44, formula (2.1.8)], the characteristic polynomial of the Frobenius endomorphism $\mathrm{Frob}(\ell)$ on the reduction $\widetilde{A}$ of A modulo ℓ (as an $\mathcal{O}$-linear endomorphism) is $X^2 - a_\ell(f)X + \ell$ and the one of the complex conjugation is $X^2 - 1$, and they must agree on $\widetilde{A}[p]$. In particular, $\mathrm{Frob}(\ell)^2$ acts trivially on $A[\mathfrak{P}]$ so $\widetilde{A}[\mathfrak{P}] = \widetilde{A}[\mathfrak{P}](\mathbb{F}_\lambda)$ and we can lift those points to K_λ. By the same arguments, on also has the decomposition

$$\widetilde{A}[\mathfrak{P}](\mathbb{F}_\lambda) = \widetilde{A}[\mathfrak{P}](\mathbb{F}_\lambda)^+ \oplus \widetilde{A}[\mathfrak{P}](\mathbb{F}_\lambda)^-$$

in two nontrivial spaces, given the characteristic polynomial of $\mathrm{Frob}(\ell)$, so each of the two spaces on the right-hand side is of dimension 1 over $\mathbb{F}_p$. We deduce immediately by the structure of finite abelian groups that as groups,

$$(\widetilde{A}(\mathbb{F}_\lambda)^\varepsilon/\mathfrak{P}\widetilde{A}(\mathbb{F}_\lambda)^\varepsilon) \cong \widetilde{A}(\mathbb{F}_\lambda)^\varepsilon[\mathfrak{P}],$$

which proves that each $(\widetilde{A}(\mathbb{F}_\lambda)/\mathfrak{P}\widetilde{A}(\mathbb{F}_\lambda))^\varepsilon$ must be of dimension 1 over $\mathbb{F}_p$, and this also lifts to K_λ (without increasing the dimension as the group of elements reducing to 0 modulo λ is p-divisible). □

To state the next result, recall that for a finite place $v \nmid p$ of good reduction of A, the image of $A(K_v)/pA(K_v)$ in $H^1(K_v, A[p])$ is precisely the inflation of $H^1(K_v^{\mathrm{unr}}/K_v, A[p])$, called the unramified part. The latter is isomorphic to $A[p]$ when all the p-torsion is defined over K_v, via the evaluation of the cocycles at $\mathrm{Frob}(v)$ the topological generator of $\mathrm{Gal}(K_v^{\mathrm{unr}}/K_v)$. The same argument translates for $A[\mathfrak{P}]$ by tensoring by $\mathcal{O}/\mathfrak{P}$ again.

Proposition 11 *Let $\mathcal{L}$ be an unramified prime ideal of L_S whose Frobenius in $\mathrm{Gal}(L_S/\mathbb{Q})$ is τh for $h \in H_S$. It is above a Kolyvagin prime ℓ and for every $s \in S$ whose localisation at $\lambda = \ell\mathcal{O}_K$ is unramified,*

$$[s, (\tau h)^2]_S = \mathrm{ev}_\lambda(s) := (\mathrm{loc}_\lambda s)(\mathrm{Frob}(\lambda)) \in A[\mathfrak{P}].$$

through the identification described above, as all $A[\mathfrak{P}]$ is defined over K_λ.

Proof By hypothesis, $(\mathcal{L}, L/\mathbb{Q})_{|L} = \tau$ so $\lambda_L = \mathcal{L} \cap \mathcal{O}$ is indeed above a Kolyvagin prime ℓ. On the other hand, $(\mathcal{L}, L_S/L) = (\mathcal{L}, L_S/\mathbb{Q})^2 = (\tau h)^2$ as the inertia does not

change between K and L_S. Now, the diagram

$$
\begin{array}{ccc}
S \subset H^1(K, A[\mathfrak{P}]) & \xrightarrow{\mathrm{res}} & \mathrm{Hom}_G(H_S, A[\mathfrak{P}]) \\
\downarrow{\scriptstyle \mathrm{loc}_\lambda} & & \downarrow{\scriptstyle \mathrm{ev}_{(\mathcal{L}, L_S/L)}} \\
H^1_{\mathrm{unr}}(K_\lambda, A[\mathfrak{P}]) & \xrightarrow{\mathrm{ev}_{\mathrm{Frob}(\lambda)}} & A[\mathfrak{P}]
\end{array}
$$

is clearly commutative, which establishes the equality by definition. $\qquad\square$

Remark 13 The set of all $(\tau h)^2$ thus obtained is exactly H_S^+, by Cebotarev density theorem.

Now, for any place v of K, we can construct ([63], section 2) a canonical bilinear pairing obtained from Tate duality

$$
\langle \cdot, \cdot \rangle_{K_v} : A(K_v)/pA(K_v) \times H^1(K_v, A)[p] \to \mathrm{Br}(K_v)[p] \cong \mathbb{Z}/p\mathbb{Z}. \tag{53}
$$

The key use of Tate duality is the following Proposition, which is a slight generalisation of [29, Proposition 8.2].

Proposition 12 *If for a prime λ of K (above a Kolyvagin prime) and a $\gamma \in H^1(K, A)^\varepsilon[\mathfrak{P}]$, one has $\mathrm{loc}_v \gamma = 0$ for all $v \neq \lambda$ but $\mathrm{loc}_\lambda \gamma \neq 0$, then for every $s \in \mathrm{Sel}_{\mathfrak{P}}(K, A)^\varepsilon$, $\mathrm{loc}_\lambda s = 0$.*

Proof By its definition, (53) comes from the Weil pairing in the sense that the latter induces a cup product

$$
(\cdot, \cdot)_{K_v} : H^1(K_v, A[p]) \times H^1(K_v, A[p]) \to H^2(K_v, \mu_p) = \mathrm{Br}(K_v)[p],
$$

for which $\delta_v(A(K_v)/pA(K_v))$ is isotropic, and the resulting quotiented pairing is exactly $\langle \cdot, \cdot \rangle_{K_v}$. Now, the so-called *global Tate duality* states that for any $s \in \mathrm{Sel}_p(K, A)$, $\gamma \in H^1(K, A)[p]$,

$$
\sum_{v \in M_K} \mathrm{inv}_v \langle \delta_v^{-1} \mathrm{loc}_v s, \mathrm{loc}_v \gamma \rangle_{K_v} = 0 \in \mathbb{Q}/\mathbb{Z},
$$

where $\mathrm{inv}_v : \mathrm{Br}(K_v) \to \mathbb{Q}/\mathbb{Z}$ is the Brauer invariant isomorphism for all v. Indeed, let us lift γ to $\widetilde{\gamma} \in H^1(K_v, A[p])$, so that for every $v \in M_K$,

$$
\langle \delta_v^{-1} \mathrm{loc}_v s, \mathrm{loc}_v \gamma \rangle_{K_v} = (\mathrm{loc}_v s, \mathrm{loc}_v \widetilde{\gamma})_{K_v} = \mathrm{loc}_{v, \mathrm{Br}}(s, \widetilde{\gamma})_K
$$

with the analogous definition of $(\cdot, \cdot)_K$ on K, and $\mathrm{loc}_{v, \mathrm{Br}} : \mathrm{Br}(K) \to \mathrm{Br}(K_v)$ the usual localisation. Now, by properties of Brauer groups, the sum of $\mathrm{inv}_v \circ \mathrm{loc}_v$ is 0 on $\mathrm{Br}(K)$ hence the formula.

Under our assumptions on γ and s, we thus have $\langle \delta_\lambda^{-1} \mathrm{loc}_\lambda s, \mathrm{loc}_\lambda \gamma \rangle_{K_\lambda} = 0$ but $\mathrm{loc}_\lambda \gamma \neq 0$, let us show how this implies that $\mathrm{loc}_\lambda s = 0$.

By the original arguments of [63], the pairing $\langle \cdot, \cdot, \rangle_{K_\lambda}$ is a perfect pairing. Being inherited from the Weil pairing, the $\mathfrak{P}$ and $\mathfrak{P}'$-parts for $\mathfrak{P} \neq \mathfrak{P}'$ are orthogonal, so it induces a duality

$$A(K_\lambda)/\mathfrak{P}A(K_\lambda) \times H^1(K_\lambda, A)[\mathfrak{P}] \to \mathbb{Z}/p\mathbb{Z}.$$

Now, it is also invariant by $\mathrm{Gal}(K_\lambda/\mathbb{Q}_\ell)$-action (there is a difference with the Weil pairing here, but it is also inherited from the cup product $(\cdot, \cdot)_{K_\lambda}$), so the $+$ and $-$ spaces on each side are orthogonal. We thus have for $\varepsilon = \pm 1$ a duality

$$(A(K_\lambda)/\mathfrak{P}A(K_\lambda))^\varepsilon \times H^1(K_\lambda, A)^\varepsilon[\mathfrak{P}] \to \mathbb{Z}/p\mathbb{Z},$$

but making use of the fact that λ is above a Kolyvagin prime, each space of the duality is thus of dimension 1 over $\mathbb{F}_p$ (Proposition 10), and so the pairing can be 0 only if one of the terms is 0, hence $\mathrm{loc}_\lambda s = 0$. $\qquad \square$
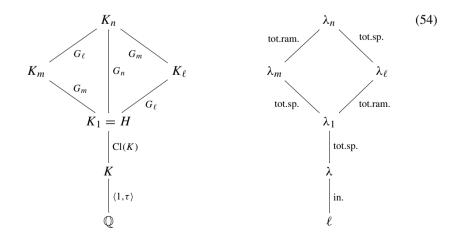
7.4 Construction of the Kolyvagin classes

Following [44], one takes the classes $[\mathfrak{a}]$ and prime ideal $\mathfrak{n}$ induced by the choices made in (48) on orders of $\mathcal{O}_K$, and for any Kolyvagin number n, we get Heegner points

$$x_n = (\mathbb{Z} + n\mathcal{O}_K, \mathfrak{n} \cap (\mathbb{Z} + n\mathcal{O}_K), [\mathfrak{a}]), \quad y_n = \pi((x_n) - (\infty)) \in A(K_n),$$

where by class field theory, K_n is the class ring field of conductor n ($K_1 = H$).

The notation $\lambda_{n,\ell}$ will refer to a choice of prime ideal of K_n above ℓ a Kolyvagin prime, consistent in case of towers of extensions, shortened to λ_n if there is no doubt on ℓ. One has that $G_n := \mathrm{Gal}(K_n/K_1) \cong (\mathcal{O}_K/n\mathcal{O}_K)^*/(\mathbb{Z}/n\mathbb{Z})^*$ and the following diagrams for $n = \ell m$ by class field theory.

$$(54)$$

In particular, $\mathbb{F}_{\lambda_n} = \mathbb{F}_{\lambda_m} = \mathbb{F}_\lambda$, a fact which will be ubiquitous and used without further mention in the end of the argument.

The crucial properties of these points (making them a *Kolyvagin system*) are the following, $\widetilde{A}$ denoting the (good) reduction of A modulo ℓ and $\mathrm{Frob}(\ell)$ the associated Frobenius endomorphism on $\widetilde{A}$.

Proposition 13 *For $n = \ell m$ a Kolyvagin number,*

$$\mathrm{Tr}_{K_n/K_m} \, y_n = [a_\ell(f)] y_m \in A(K_m) \tag{55}$$

$$y_n \bmod \lambda_n = \mathrm{Frob}(\ell) \cdot y_m \text{ in } \widetilde{A}(\mathbb{F}_{\lambda_n}) = \widetilde{A}(\mathbb{F}_\lambda) \tag{56}$$

$$\tau(y_n) \in \sigma(y_n) + A(K_n)_{\mathrm{tors}} \tag{57}$$

for some $\sigma \in \mathcal{G}_n := \mathrm{Gal}(K_n/K)$.

Proof By classical properties of Heegner points ([28], paragraphs 4 and 5) and class field theory for K_n/K_m,

$$\mathrm{Tr}_{K_n/K_m} \, x_n = T_\ell \cdot x_m \tag{58}$$

as divisors on $X_0(N)$, which proves (55) when combined with (43). We obtain (57) with the same properties.

Looking at the diagrams (54), as λ_n/λ_m is totally ramified, the reduction of the left-hand side of (58) is $(\ell+1) x_n \bmod \lambda_n$, and the one of the right-hand side has one term equal to $\mathrm{Frob}(\ell) x_m$ by the Eichler-Shimura relation $T_\ell = \mathrm{Frob}(\ell) + \widehat{\mathrm{Frob}(\ell)}$, so there exists $\sigma \in \mathrm{Gal}(K_n/K_m)$ such that the reduction of σx_n is $\mathrm{Frob}(\ell) \widetilde{x_m}$, but every σ reduces to the identity on $\widetilde{A}(\mathbb{F}_\lambda)$ so the equality is true term by term hence (56). See also [44, Corollaries 2.3.3 and 2.3.4] for the $n = \ell$ case. $\qquad\square$

Proposition 14 *For every Kolyvagin number n, one can define in successive order (using the Heegner points y_m for $m|n$):*
- *A point $P_n \in A(K_n)$ whose class $[P_n] \in A(K_n)/pA(K_n)$ is fixed by $\mathcal{G}_n$ (and $P_1 = y_K$).*
- *The unique class $c(n) \in H^1(K, A[p])$ whose restriction to $H^1(K_n, A[p])^{\mathcal{G}_n}$ comes from $[P_n]$, and its image $d(n)$ in $H^1(K, A)[p]$. They correspond to one another in the following commutative diagram with exact rows and columns*

$$\tag{59}$$

$$
\begin{array}{ccccccccc}
& & & & & & \overset{\widehat{d(n)}}{H^1(K_n/K, A)[p]} & & \\
& & & & & & \downarrow {\scriptstyle \mathrm{inf}} & & \\
& & & & \overset{c(n)}{} & & \overset{d(n)}{} & & \\
0 & \longrightarrow & A(K)/pA(K) & \overset{\delta}{\longrightarrow} & H^1(K, A[p]) & \longrightarrow & H^1(K, A)[p] & \longrightarrow & 0 \\
& & \downarrow & & \sim \downarrow {\scriptstyle \mathrm{Res}} & & \downarrow {\scriptstyle \mathrm{Res}} & & \\
& & {\scriptstyle [P_n]} & & & & & & \\
0 & \longrightarrow & (A(K_n)/pA(K_n))^{\mathcal{G}_n} & \overset{\delta_n}{\longrightarrow} & H^1(K_n, A[p])^{\mathcal{G}_n} & \longrightarrow & H^1(K_n, A)^{\mathcal{G}_n}[p] & &
\end{array}
$$

Proof The construction and properties of P_n proceeds exactly as in ([29], (3.5) to (4.1)). The only nontrivial thing to prove (to define $c(n)$ from $[P_n]$)is that the central row of (59) is an isomorphism. The extension $K_n/\mathbb{Q}$ is unramified outside primes dividing $D_K n$, and the extension $\mathbb{Q}(A[p])/\mathbb{Q}$ is unramified outside primes dividing Mp, so as $D_K n$ and pM are coprime by construction, these extensions are linearly disjoint. In particular, $K_n(A[p])/K_n$ has Galois group isomorphic to $\mathrm{Gal}(\mathbb{Q}(A[p])/\mathbb{Q})$ and thus no fixed point in $A[p]$ by Proposition 9 (a). The isomorphism follows by [29, (4.2)]. $\qquad\square$

These points enjoy a wealth of very strong properties detailed below.

Proposition 15 *For every Kolyvagin number n:*

(a) $[P_n]$ *(resp.* $c(n), d(n)$*) lives in the $\mu(n)$-eigenspace of $A(K_n)/pA(K_n)$ (resp.* $H^1(K, A[p])$, $H^1(K, A)[p]$*), where $\mu(n)$ is the Moebius function.*

(b) The class $c(n)_{\mathfrak{P}} \in H^1(K, A[\mathfrak{P}])$ (resp. $d(n)_{\mathfrak{P}} \in H^1(K, A)[\mathfrak{P}]$) is trivial if and only if $P_n \in \mathfrak{P}A(K_n)$ (resp. $\mathfrak{P}A(K_n) + A(K)^{\mu(n)}$).

(c) For every place v of K, the class $\mathrm{loc}_v\, d(n)$ is trivial except if $v|n$.

(d) If $n = \ell m$ and $\lambda = \ell\mathcal{O}_K$, the class $\mathrm{loc}_\lambda\, d(n)_{\mathfrak{P}}$ is trivial if and only if $P_m \in \mathfrak{P}A(K_{\lambda_m})$ if and only if $\mathrm{loc}_\lambda\, c(m)_{\mathfrak{P}} = 0$.

Proof (a) for $[P_n]$ is inherited from (57) by the construction of P_n (see Proposition 5.4 of [29]), and deduced for $c(n), d(n)$ by τ-equivariance of the morphisms of (59).

(b) is obtained by tensoring (59) by $\mathcal{O}/\mathfrak{P}$, which preserves exactness by flatness and $[P_n]$ seen in $A(K_n)/pA(K_n) \otimes \mathcal{O}/\mathfrak{P}$ is exactly the image of P_n in $A(K_n)/\mathfrak{P}A(K_n)$. The proof of (c) is given by Proposition 6.2 of [29].

For (d), define $D = \mathrm{Gal}((K_n)_{\lambda_n}/K_\lambda)$, which is cyclic generated by some σ_ℓ. We thus have injective arrows (defined below)

$$H^1(D, A)[p] \overset{\mathrm{red}}{\hookrightarrow} \widetilde{A}(\mathbb{F}_\lambda)[p] \cong H^1(\mathbb{F}_\lambda, \widetilde{A}[p]) \overset{\iota}{\hookleftarrow} \widetilde{A}(\mathbb{F}_\lambda)/p\widetilde{A}(\mathbb{F}_\lambda) \qquad (60)$$

where for a cocycle $c \in Z^1(D, A)$, $\mathrm{red}(c) = c(\sigma_\ell) \bmod \lambda_n$, and invariant up to coboundary because K_n/K_m is totally ramified at λ_m, so red is well-defined. As $A^1((K_n)_{\lambda_n})$ is a pro-ℓ-group, $H^1(D, A^1)[p] = 0$ which proves that red is injective. The map ι is the quotiented connecting homomorphism, automatically injective. As $\widetilde{A}(\mathbb{F}_\lambda)$ is a finite abelian group, the orders of $\widetilde{A}(\mathbb{F}_\lambda)[p]$ and $\widetilde{A}(\mathbb{F}_\lambda)/p\widetilde{A}(\mathbb{F}_\lambda)$ are readily seen to be equal so ι is also an isomorphism. By [29, Proposition 6.2(2)], the image of $\mathrm{loc}_\lambda\, d(n)$ in $\widetilde{A}(\mathbb{F}_\lambda)[p]$ by red is

$$((\ell + 1)\,\mathrm{Frob}(\ell) - [a_\ell(f)]) \cdot \widetilde{R_m},$$

where $\widetilde{R_m}$ is any choice of p-th root of $\widetilde{P_m}$ in $\widetilde{A}$. By the proof of Proposition 10, its image by $\mathrm{Frob}(\ell)$ is then

$$\ell(\mathrm{Frob}(\ell)^2 - \mathrm{Id})\widetilde{R_m} = -(\mathrm{Frob}(\ell)^2 - \mathrm{Id})\widetilde{R_m},$$

but the injection ι from (60) is explicitly given by taking a p-th root and applying $(\mathrm{Frob}(\ell)^2 - \mathrm{Id})$, as $\mathrm{Frob}(\ell)^2 = \mathrm{Frob}(\lambda)$ ([44], Lemma 3.4.2 for details). The image

of $\mathrm{loc}_\lambda\, d(n)$ in $\widetilde{A}(\mathbb{F}_\lambda)/p\widetilde{A}(\mathbb{F}_\lambda)$ via (60) is thus exactly $-\mathrm{Frob}(\ell)^{-1} \cdot \widetilde{P_m}$, and its $\mathfrak{P}$-part is trivial if and only if the $\mathfrak{P}$-part of $\widetilde{P_m}$ is. Finally, $A^1(K_{\lambda_m})$ is p-divisible hence the equality of $\mathcal{O}/(p)$-modules $A(K_{\lambda_m})/pA(K_{\lambda_m}) \cong \widetilde{A}(\mathbb{F}_\lambda)/p\widetilde{A}(\mathbb{F}_\lambda)$, so finally $\mathrm{loc}_\lambda\, d(n)_\mathfrak{P}$ is trivial if and only if $[P_m] \in A(K_{\lambda_m})/pA(K_{\lambda_m})[\mathfrak{P}]$, which is equivalent to $P_m \in \mathfrak{P}A(K_{\lambda_m})$ and the equivalence in terms of $c(m)$ is straightforward. $\qquad\square$

7.5 End of the proof

Let $S = \mathrm{Sel}_\mathfrak{P}(K, A)$. By (51), $P_1 = y_K \notin \mathfrak{P}A(K)$, hence it defines a nonzero $s_K := c(1) \in S^+$ (Proposition 15 (a)). Fixing $s \in S$, for every $h \in H_S$, by Cebotarev density theorem, there is a prime ideal $\mathcal{L}$ such that $(\mathcal{L}, L_S/\mathbb{Q}) = \tau h$, and by Proposition 11,

$$[s, (\tau h)^2]_S = \mathrm{loc}_\lambda\, s(\mathrm{Frob}(\lambda))$$

where λ is the prime ideal of K below $\mathcal{L}$, and above ℓ which is a Kolyvagin prime. Outside of I_S^+ (defined as the $+$-part of the orthogonal of s_K), this formula proves that $\mathrm{loc}_\lambda\, s_K \neq 0$, so $\mathrm{loc}_\lambda\, d(\ell)_\mathfrak{P} \neq 0$ and all other localisations of $d(\ell)_\mathfrak{P}$ are trivial by Proposition 15. By Proposition 12, if $s \in S^-$, $\mathrm{loc}_\lambda\, s = 0$ so $[s, (\tau h)^2]_S = 0$, hence $S^- = 0$ by Lemma 23.

Now, consider $s \in S^+$ such that for some $\mathcal{L}$ as above (fixed, so it fixes λ and h above), $\mathrm{loc}_\lambda\, s = 0$. We have $\mathrm{loc}_\lambda\, s_K \neq 0$ by hypothesis on h, so in turn $\mathrm{loc}_\lambda\, d(\ell)_\mathfrak{P} \neq 0$ by Proposition 15 (d) and $c(\ell)_\mathfrak{P}$ does not belong to S. By the perfect pairing result of Lemma 22 applied to $\langle S, c(\ell)\rangle$ if $(\tau h)^2 \notin I_S^+$, the extensions L_S and $L_{\langle c(\ell)\rangle}$ are linearly disjoint over L, which allows us, for any $h' \in H_S$, to choose $\mathcal{L}'$ a prime ideal of $L_S L_{\langle c(\ell)\rangle}$ whose Frobenius restricted to L_S is $\tau h'$ and whose Frobenius restricted to $L_{\langle c(\ell)\rangle}$ is of the shape τh_0 and *not* orthogonal to $c(\ell)_\mathfrak{P}$. Denoting ℓ' the corresponding Kolyvagin prime and λ' the ideal of $\mathcal{O}_K$, we thus have

$$[c(\ell)_\mathfrak{P}, (\tau h_0)^2] = \mathrm{loc}_{\lambda'}\, c(\ell)_\mathfrak{P}(\mathrm{Frob}(\lambda')),$$

this formula being legitimate because $\mathrm{loc}_{\lambda'}(d(\ell)_\mathfrak{P}) = 0$ by Proposition 15 (c). All this proves that $\mathrm{loc}_{\lambda'}\, c(\ell)_\mathfrak{P} \neq 0$ so $\mathrm{loc}_{\lambda'}\, d(\ell\ell')_\mathfrak{P} \neq 0$ by Proposition 15 (d), and it belongs to $H^1(K, A)^+[\mathfrak{P}]$. Now, for our s above, the global Tate duality between s and $d(\ell\ell')$ in the proof of Proposition 12 has two possible nonzero terms (in λ and λ'), but by hypothesis $\mathrm{loc}_\lambda\, s = 0$ so the λ'-term is alone, therefore 0 as well. This implies by Proposition 12 that $\mathrm{loc}_{\lambda'}\, s = 0$ for all such λ', therefore $s = 0$ in this case by Lemma 23.

Finally, for $s \in S^+$, as $\mathrm{loc}_\lambda\, s_K \neq 0$ and the space $(A(K_\lambda)/\mathfrak{P}A(K_\lambda))^+$ is one-dimensional (Proposition 10), there is $k \in \mathbb{Z}$ such that $s - ks_K$ satisfies the previous hypothesis and then $s = ks_K$, so we have proved that $S^+ = \langle s_K\rangle$.

References

1. Atkin, A., Lehner, J.: Hecke operators on $\Gamma_0(m)$. Math. Ann. **185**, 134–160 (1970)

2. Baker, M.H.: Kamienny's criterion and the method of Coleman and Chabauty. Proc. Am. Math. Soc. **127**(10), 2851–2856 (1999)
3. Balakrishnan, J., Dogra, N.: An effective Chabauty–Kim theorem. Compos. Math. **155**(6), 1057–1075 (2019). https://doi.org/10.1112/s0010437x19007243
4. Balakrishnan, J.S., Best, A.J., Bianchi, F., Lawrence, B., Müller, J.S., Triantafillou, N., Vonk, J.: Two recent p-adic approaches towards the (effective) Mordell conjecture. arXiv preprint arXiv:1910.12755 (2019)
5. Balakrishnan, J.S., Dan-Cohen, I., Kim, M., Wewers, S.: A non-abelian conjecture of Tate–Shafarevich type for hyperbolic curves. Math. Ann. **372**(1–2), 369–428 (2018)
6. Balakrishnan, J.S., Dogra, N.: Quadratic Chabauty and rational points, I: p-adic heights. Duke Math. J. **167**(11), 1981–2038 (2018). https://doi.org/10.1215/00127094-2018-0013
7. Balakrishnan, J.S., Dogra, N.: Quadratic Chabauty and rational points II: generalised height functions on selmer varieties. Int. Math. Res. Not. (2020). https://doi.org/10.1093/imrn/rnz362
8. Balakrishnan, J.S., Dogra, N., Müller, J.S., Tuitman, J., Vonk, J.: Explicit Chabauty–Kim for the split Cartan modular curve of level 13. Ann. Math. (2) **189**(3), 885–944 (2019). https://doi.org/10.4007/annals.2019.189.3.6
9. Balakrishnan, J.S., Dogra, N., Müller, J.S., Tuitman, J., Vonk, J.: Quadratic Chabauty for modular curves: algorithms and examples. In preparation (2020)
10. Betts, L.A., Dogra, N.: Ramification of étale path torsors and harmonic analysis on graphs. arXiv preprint arXiv:1909.05734 (2019)
11. Bilu, Y., Parent, P.: Serre's uniformity problem in the split Cartan case. Ann. Math. **2**(173), 569–584 (2011)
12. Bilu, Y., Parent, P., Rebolledo, M.: Rational points on $X_0^+(p^r)$. Annales de l'Institut Fourier **63**, (2013)
13. Birkenhake, C., Lange, H.: Complex abelian varieties, 2nd edn. Grundlehren der Mathematischen Wissenschaften , vol. 302. Springer, Berlin, Heidelberg (2004)
14. Bump, D.: Automorphic forms and representations. Cambridge University Press, Cambridge (1996)
15. Cai, L., Shu, J., Tian, Y.: Explicit Gross-Zagier and Waldspurger formulae. Algebra Number Theory **8**(10), 2523–2572 (2014). https://doi.org/10.2140/ant.2014.8.2523
16. Chen, I.: On relations between Jacobians of certain modular curves. J. Algebra **231**(1), 414–448 (2000)
17. Colombo, E., van Geemen, B.: Note on curves in a Jacobian. Compos. Math. **88**(3), 333–353 (1993)
18. Darmon, H., Rotger, V.: Diagonal cycles and Euler systems I: A p-adic Gross-Zagier formula. Ann. Sci. Éc. Norm. Supér. (4) **47**(4), 779–832 (2014)
19. Darmon, H., Rotger, V., Sols, I.: Iterated integrals, diagonal cycles and rational points on elliptic curves. Publications Mathématiques de Besançon **2**, 19–46 (2012). https://doi.org/10.5802/pmb.a-145
20. Daub, M.: Complex and p-adic computations of Chow–Heegner points. PhD Thesis, Berkeley (2013)
21. Deligne, P.: Cohomologie étale. Lecture Notes in Mathematics, vol. 569. Springer-Verlag, Berlin (1977)
22. Deligne, P.: Le groupe fondamental de la droite projective moins trois points. In: Galois groups over **Q** (Berkeley, CA, 1987), *Math. Sci. Res. Inst. Publ.*, vol. 16, pp. 79–297. Springer, New York (1989)
23. Edixhoven, B., Parent, P.: Semistable reduction of modular curves associated with maximal subgroups in prime level. arXiv preprint arXiv:1907.02418 (2019)
24. Ellenberg, J.S.: Galois representations attached to $\mathbb{Q}$-curves and the generalized Fermat equation $A^4 + B^2 = C^p$. Am. J. Math. **126**(4), 763–787 (2004)
25. Fuchs, C., Pham, D.H.: The p-adic analytic subgroup theorem revisited. P-adic Numbers Ultrametr Anal Appl **7**(2), 143–156 (2015)
26. Fulton, W.: Intersection theory, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 2, second edn. Springer-Verlag, Berlin (1998). https://doi.org/10.1007/978-1-4612-1700-8
27. González, J., Lario, J.C.: Rational and Elliptic parametrizations of $\mathbb{Q}$-curves. J. Number Theory **72**(1), 13–31 (1998)
28. Gross, B.: Heegner points on $X_0(N)$. In: Modular forms (Durham, 1983), Ellis Horwood Ser. Math. Appl.: Statist. Oper. Res., pp. 87–105. Horwood, Chichester (1984)
29. Gross, B.H.: Kolyvagin's work on modular elliptic curves. In: *L*-functions and arithmetic (Durham, 1989), *London Math. Soc. Lecture Note Ser.*, vol. 153, pp. 235–256. Cambridge Univ. Press, Cambridge (1991). https://doi.org/10.1017/CBO9780511526053.009
30. Gross, B.H., Kudla, S.S.: Heights and the central critical values of triple product *L*-functions. Compos. Math. **81**(2), 143–209 (1992)
31. Gross, B.H., Schoen, C.: The modified diagonal cycle on the triple product of a pointed curve. Ann. Inst. Fourier **45**(3), 649–679 (1995)

32. Gross, B.H., Zagier, D.B.: Heegner points and derivatives of *L*-series. Invent. Math. **84**(2), 225–320 (1986). https://doi.org/10.1007/BF01388809

33. Hain, R.: Rational points of universal curves. J. Am. Math. Soc. **24**(3), 709–769 (2011). https://doi.org/10.1090/S0894-0347-2011-00693-0

34. Hain, R., Matsumoto, M.: Galois actions of fundamental groups of curves and the cycle $C - C^-$. J. Inst. Math. Jussieu **4**(3), 363–403 (2005)

35. Hindry, M., Silverman, J.H.: Diophantine geometry. An introduction., vol. 201. New York, NY: Springer (2000)

36. Iwaniec, H., Kowalski, E.: Analytic number theory., vol. 53. Providence, RI: American Mathematical Society (AMS) (2004)

37. Iwaniec, H., Sarnak, P.: The non-vanishing of central values of automorphic *L*-functions and Landau-Siegel zeros. Isr. J. Math. **120**, 155–177 (2000)

38. Jannsen, U.: Mixed motives and algebraic *K*-theory. Lecture Notes in Mathematics, vol. 1400. Springer-Verlag, Berlin (1990)

39. Kim, M.: The motivic fundamental group of $\mathbf{P}^1 - \{0, 1, \infty\}$ and the theorem of Siegel. Invent. Math. **161**(3), 629–656 (2005)

40. Kim, M.: The unipotent Albanese map and Selmer varieties for curves. Publ. Res. Inst. Math. Sci. **45**(1), 89–133 (2009)

41. Kim, M., Tamagawa, A.: The *l*-component of the unipotent Albanese map. Math. Ann. **340**(1), 223–235 (2008). https://doi.org/10.1007/s00208-007-0151-x

42. Kohen, D., Pacetti, A.: Heegner points on Cartan non-split curves. Canad. J. Math. **68**(2), 422–444 (2016). https://doi.org/10.4153/CJM-2015-047-6

43. Kolyvagin, V.A.: Euler systems. In: The Grothendieck Festschrift, Vol. II, *Progr. Math.*, vol. 87, pp. 435–483. Birkhäuser Boston, Boston, MA (1990)

44. Kolyvagin, V.A., Logachev, D.Y.: Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties. Leningr. Math. J. **1**(5), 1229–1253 (1990)

45. Kowalski, E., Michel, P., VanderKam, J.: Non-vanishing of high derivatives of automorphic *L*-functions at the center of the critical strip. J. Reine Angew. Math. **526**, 1–34 (2000)

46. Le Fourn, S.: Surjectivity of Galois representations associated with quadratic $\mathbb{Q}$-curves. Math. Ann. **365**(1), 173–214 (2016)

47. Le Fourn, S.: Nonvanishing of central values of *L*-functions of newforms in $S_2(\Gamma_0(dp^2))$ twisted by quadratic characters. Canad. Math. Bull. **60**(2), 329–349 (2017)

48. LMFDB Collaboration, T.: The L-functions and modular forms database. http://www.lmfdb.org (2019)

49. Matev, T.: The *p*-adic analytic subgroup theorem and applications. arXiv preprint arXiv:1010.3156 (2010)

50. Mazur, B.: Rational isogenies of prime degree (with an appendix by D. Goldfeld). Invent. Math. **44**(2), 129–162 (1978)

51. Milne, J.S.: Abelian varieties. In: Arithmetic geometry (Storrs, Conn., 1984), pp. 103–150. Springer, New York (1986)

52. Mumford, D.: Abelian varieties. With appendices by C. P. Ramanujam and Yuri Manin. Corrected reprint of the 2nd ed. 1974., corrected reprint of the 2nd ed. 1974 edn. New Delhi: Hindustan Book Agency/distrib. by American Mathematical Society (AMS); Bombay: Tata Institute of Fundamental Research (2008)

53. Nekovář, J.: On *p*-adic height pairings. In: Séminaire de Théorie des Nombres, Paris, 1990–91, *Progr. Math.*, vol. 108, pp. 127–202. Birkhäuser Boston, Boston, MA (1993). https://doi.org/10.1007/s10107-005-0696-y

54. Nekovář, J.: The Euler system method for CM points on Shimura curves. In: *L*-functions and Galois representations, *London Math. Soc. Lecture Note Ser.*, vol. 320, pp. 471–547. Cambridge Univ. Press, Cambridge (2007). https://doi.org/10.1017/CBO9780511721267.014

55. Rebolledo, M., Wuthrich, C.: A moduli interpretation for the non-split Cartan modular curve (2017). ArXiv:1402.3498

56. Ribet, K.: Abelian Varieties over $\mathbb{Q}$ and Modular Forms. In: Modular Curves and Abelian Varieties, pp. 241–261. Birkhäuser (2004)

57. Ribet, K.A.: Galois action on division points of Abelian varieties with real multiplications. Am. J. Math. **98**(3), 751–804 (1976). https://doi.org/10.2307/2373815

58. Serre, J.P.: Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. Invent. Math. **15**(4), 259–331 (1972)

59. Siksek, S.: Quadratic Chabauty for modular curves (2017). ArXiv:1704.00473
60. de Smit, B., Edixhoven, B.: Sur un résultat d'Imin Chen. Mat. Res. Lett. **7**, 147–153 (2000)
61. Smith, B.: Explicit endomorphisms and correspondences (2005). PhD thesis
62. Stein, W.: Modular forms, a computational approach, *Graduate Studies in Mathematics*, vol. 79. American Mathematical Society, Providence, RI (2007). https://doi.org/10.1090/gsm/079. With an appendix by Paul E. Gunnells
63. Tate, J.: WC-groups over p-adic fields. Séminaire Bourbaki 10e année, Textes des Conférences, Exposé No. 156, 13 p. (1958). (1958)
64. Tian, Y.: Euler systems of CM points on Shimura curves (2003). PhD Thesis, Columbia University
65. Tian, Y., Zhang, S.W.: Euler systems of CM points on Shimura curves. In preparation
66. Vignéras, M.F.: Valeur au centre de symétrie des fonctions L associées aux formes modulaires. In: Séminaire de Théorie des Nombres, Paris 1979-1980, Progress in Mathematics, vol .12, pp. 331–356. Birkhaüser, Boston (1981)
67. Zhang, S.W.: Gross-Zagier formula for GL(2). II. In: Heegner points and Rankin L-series, *Math. Sci. Res. Inst. Publ.*, vol. 49, pp. 191–214. Cambridge Univ. Press, Cambridge (2004). https://doi.org/10.1017/CBO9780511756375.008