msp

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

Microlocal lifts and quantum unique ergodicity on $GL_2(\mathbb{Q}_p)$

Paul D. Nelson

We prove that arithmetic quantum unique ergodicity holds on compact arithmetic quotients of $\mathrm{GL}_2(\mathbb{Q}_p)$ for automorphic forms belonging to the principal series. We interpret this conclusion in terms of the equidistribution of eigenfunctions on covers of a fixed regular graph or along nested sequences of regular graphs.

Our results are the first of their kind on any p-adic arithmetic quotient. They may be understood as analogues of Lindenstrauss's theorem on the equidistribution of Maass forms on a compact arithmetic surface. The new ingredients here include the introduction of a representation-theoretic notion of "p-adic microlocal lifts" with favorable properties, such as diagonal invariance of limit measures; the proof of positive entropy of limit measures in a p-adic aspect, following the method of Bourgain–Lindenstrauss; and some analysis of local Rankin–Selberg integrals involving the microlocal lifts introduced here as well as classical newvectors. An important input is a measure-classification result of Einsiedler–Lindenstrauss.

1. Introduction

1.1. *Overview.* Let p be a prime number. This article is concerned with the limiting behavior of eigenfunctions on compact arithmetic quotients of the group $G := \mathrm{GL}_2(\mathbb{Q}_p)$. A rich class of such quotients is parametrized by the definite quaternion algebras B over $\mathbb{Q}$ that split at p. A maximal order R in such an algebra and an embedding $B \hookrightarrow M_2(\mathbb{Q}_p)$ give rise to a discrete cocompact subgroup $\Gamma := R[1/p]^\times$ of G. Fix one such Γ. The corresponding arithmetic quotient $X := \Gamma \backslash G$ is then compact; in interpreting this, it may help to note that the center of Γ is the discrete cocompact subgroup $\mathbb{Z}[1/p]^\times$ of $\mathbb{Q}_p^\times$. In adelic terms, we may identify X with $B^\times \backslash B_{\mathbb{A}}^\times / B_\infty^\times \prod_{\ell \neq p} R_\ell^\times$ (see Section 2.1 for notation).

The space X is a p-adic analogue of the cotangent bundle of an arithmetic hyperbolic surface, such as the modular surface $SL_2(\mathbb{Z})\backslash \mathbb{H}$. It comes with commuting families of Hecke correspondences T_ℓ indexed by the primes $\ell \neq p$ (see Section 3.1). To zeroth approximation, the space X is modeled by its minimal quotient $Y := X/K = \Gamma\backslash G/K$ by the maximal compact subgroup $K := GL_2(\mathbb{Z}_p)$ of G. That quotient Y comes with an additional Hecke correspondence T_p. To simplify the exposition of Section 1.1, it will be convenient to assume that

$$\text{(the torsion subgroup of } \Gamma) = \{\pm 1\}. \tag{1}$$

Then Y may be safely regarded as an undirected $(p+1)$-regular finite multigraph (see [Vignéras 1980; Serre 2003; Lindenstrauss 2006b, §8]), whose adjacency matrix is T_p. The simplifying assumption (1) holds when the underlying quaternion algebra has discriminant (say) 73, in which case the graph (Y, T_p) may be depicted as follows when $p = 2, 3$:[1]

Such graphs and their eigenfunctions appear naturally in several contexts, and have been extensively studied since the pioneering work of Brandt [1943] and Eichler [1955]; they specialize to the p-isogeny graphs of elliptic curves in finite characteristic [Gross 1987, §2], provide an important tool for constructing spaces of modular forms [Pizer 1980], and their remarkable expansion properties have been studied and applied in computer science following [Lubotzky et al. 1988].

To study the space X at a finer resolution than that of its minimal quotient Y, we introduce for each pair of integers m, m' the notation $m..m' := \{m, m+1, \ldots, m'\}$ and set

$$Y_{m..m'} := \left\{ \begin{array}{l} \text{nonbacktracking paths } x = (x_m \to x_{m+1} \to \cdots \to x_{m'}) \\ \hspace{3cm} \text{indexed by } m..m' \text{ on the graph } (Y, T_p) \end{array} \right\}. \tag{2}$$

We will recall in Definition 10 the standard group-theoretic realization of $Y_{m..m'}$ as a quotient of X. We may and shall identify $Y_{0..0}$ with Y. For $m..m' \supseteq n..n'$, we define compatible surjections $Y_{m..m'} \to Y_{n..n'}$ by forgetting part of the path. For example, if $N \geq 0$, then the map $Y_{-N..N} \to Y_{0..0} = Y$ sends a path x as in (2) to its central vertex x_0. We define $L^2(Y_{m..m'})$ with respect to the normalized counting measure, so that the maps $Y_{m..m'} \to Y_{n..n'}$ are measure-preserving.

We wish to study the asymptotic behavior of "eigenfunctions" in $L^2(Y_{m..m'})$ as $|m - m'| \to \infty$. From the arithmetic perspective, there is a distinguished collection of such eigenfunctions, whose definition is

---

[1] The images were produced using the "Graph" and "BrandtModule" functions in Sage [2015].

analogous to that of the set of normalized classical holomorphic newforms of some given weight and level:

Definition 1 (newvectors). Let $L^2_{\text{new}}(Y_{m..m'}) \subseteq L^2(Y_{m..m'})$ denote the space of functions $\varphi : Y_{m..m'} \to \mathbb{C}$ that are orthogonal to pullbacks from $Y_{n..n'}$ whenever $n..n' \subsetneq m..m'$. Let $\mathcal{F}_{m..m'} \subseteq L^2_{\text{new}}(Y_{m..m'})$ be an orthonormal basis consisting of φ for which:

- the pullback of φ to $X = \Gamma \backslash G$ generates an irreducible representation of $G = GL_2(\mathbb{Q}_p)$ under the right translation action, and

- φ is an eigenfunction of the Hecke operator T_ℓ (see Section 3.1) for all primes $\ell \neq p$.

It is known[2] then that $|\mathcal{F}_{m..m'}| \asymp |Y_{m..m'}| \asymp p^{|m-m'|}$ for $|m - m'|$ sufficiently large. To simplify the exposition of Section 1.1, we focus on the symmetric intervals $-N..N$. Fix $n \in \mathbb{Z}_{\geqslant 0}$. Let $N \geqslant n$ be an integral parameter tending off to ∞. Denote by $\text{pr} : Y_{-N..N} \twoheadrightarrow Y_{-n..n}$ the natural surjection. For $\varphi \in \mathcal{F}_{-N..N}$, we may define a probability measure μ_φ on $Y_{-n..n}$ by setting

$$\mu_\varphi(E) := \frac{1}{|Y_{-N..N}|} \sum_{x \in Y_{-N..N}:\text{pr}(x) \in E} |\varphi|^2(x).$$

For example, in the instructive special case $n = 0$, the measures μ_φ live on the base graph $Y_{0..0} = Y$ and assign to subsets $E \subseteq Y$ the number

$$\mu_\varphi(E) = \frac{1}{|Y_{-N..N}|} \sum_{x = (x_{-N} \to \cdots \to x_N) \in Y_{-N..N}:x_0 \in E} |\varphi|^2(x),$$

which quantifies how much mass $\varphi : Y_{-N..N} \to \mathbb{C}$ assigns to paths whose central vertex lies in E.

Question 2. Fix $n \in \mathbb{Z}_{\geqslant 0}$. Let $N \geqslant n$ traverse a sequence of positive integers tending to ∞. For each N, choose an element $\varphi_N \in \mathcal{F}_{-N..N}$. What are the possible limits of the sequence of measures μ_{φ_N} on the space $Y_{-n..n}$?

The following conjecture has not appeared explicitly in the literature, but may be regarded nowadays as a standard analogue of the arithmetic quantum unique ergodicity conjecture of Rudnick–Sarnak [1994] (see [Sarnak 2011; Nelson et al. 2014]).

Conjecture 3. *In the context of Question 2, the uniform measure on $Y_{-n..n}$ is the only possible weak limit. In other words, for any sequence $\varphi_N \in \mathcal{F}_{-N..N}$ and any $E \subseteq Y_{-n..n}$,*

$$\lim_{N \to \infty} \mu_{\varphi_N}(E) = \frac{|E|}{|Y_{-n..n}|}.$$

---

[2] One may verify this by applying the trace formula for $L^2(\Gamma \backslash G)$ to an element $f \in C_c^\infty(G)$, as in [Nelson 2017], that defines the orthogonal projection onto $L^2_{\text{new}}(Y_{m..m'})$, or alternatively by appealing to the Eichler/Jacquet–Langlands correspondence, which identifies $\mathcal{F}_{m..m'}$ with the set of normalized weight two newforms on $\Gamma_0(p^{|m-m'|}d_B)$, with d_B the discriminant of B, and appealing to standard formulas for dimensions of spaces of newforms.

Conjecture 3 predicts that for any sequence $\varphi_N \in \mathcal{F}_{-N..N}$, the corresponding sequence of L^2-masses μ_{φ_N} equidistributes under pushforward to any fixed space $Y_{-n..n}$. One can formulate this conclusion more concisely in terms of equidistribution on the compact space $\varprojlim Y_{-n..n}$ of infinite bidirectional nonbacktracking paths, or equivalently, on the space $X = \Gamma \backslash G$.

We note that the quantum unique ergodicity conjecture of Rudnick–Sarnak [1994] includes the case of *nonarithmetic* compact hyperbolic surfaces, while Conjecture 3, as formulated here, is specific to the arithmetic setting. We indicate in Remark 31 how one might formulate it more generally.

By explicating the triple product formula [Ichino and Ikeda 2010], one can show that Conjecture 3 follows from an open case of the subconvexity conjecture, which in turn follows from GRH; the latter can be shown to imply more precisely that

$$\mu_{\varphi_N}(E) = \frac{|E|}{|Y_{-n..n}|} + O(p^{-(1+o(1))N/2}) \tag{3}$$

for fixed n. There are nowadays well-developed techniques (see for instance [Nelson 2016, §1.4]) to establish that:

- the prediction (3) holds for φ_N outside a hypothetical exceptional subset of density $o(1)$,

- if (3) is true, it is essentially optimal, and

- Conjecture 3 holds for φ_N outside a hypothetical exceptional subset of extremely small density $|\mathcal{F}_{-N..N}|^{-1/2+o(1)}$. (This may be understood as a very strong form of "quantum ergodicity," which would assert the analogous conclusion with density $o(1)$; compare with [Anantharaman and Le Masson 2015; Le Masson and Sahlsten 2017].)

The problem of eliminating such exceptions entirely (in the present setting and related ones) has proven subtle.

For context, we recall some instances in which the difficulty indicated above has been overcome; notation and terminology should be clear by analogy.

Theorem 4 [Lindenstrauss 2006b]. *Let $\Gamma' \backslash \mathbb{H}$ be a compact hyperbolic surface attached to an order in a nonsplit indefinite quaternion algebra. Let φ traverse a sequence of L^2-normalized Hecke–Laplace eigenfunctions on $\Gamma' \backslash \mathbb{H}$ with Laplace eigenvalue tending to ∞. Then the L^2-masses μ_φ equidistribute.*

Theorem 5 (N, N–Pitale–Saha, Hu [Nelson 2011; Nelson et al. 2014; Hu 2018]). *Fix a natural number q_0. Let q traverse a sequence of natural numbers tending to ∞. Let φ be an L^2-normalized holomorphic Hecke newform on the standard congruence subgroup $\Gamma_0(q)$ of $\mathrm{SL}_2(\mathbb{Z})$. Then the pushforward to $\Gamma_0(q_0) \backslash \mathbb{H}$ of the L^2-mass of φ equidistributes.*

We may of course specialize Theorem 5 to powers of a fixed prime:

Theorem 6 (N, N–Pitale–Saha, Hu [Nelson 2011; Nelson et al. 2014; Hu 2018]). *Fix a prime p and a nonnegative integer n_0. Let n traverse a sequence of natural numbers tending to ∞. Let φ be an L^2-normalized holomorphic Hecke newform on $\Gamma_0(p^n)$. Then the pushforward to $\Gamma_0(p^{n_0}) \backslash \mathbb{H}$ of the L^2-mass of φ equidistributes.*

Conjecture 3 is in the spirit of Theorem 6, save a crucial distinction to be discussed in due course (see Remark 19). Unfortunately, the method underlying the proof of Theorem 6, due to Holowinsky–Soundararajan [2010], is fundamentally inapplicable to Conjecture 3 due to its reliance on parabolic Fourier expansions, which are unavailable on the compact quotient X. We will instead develop here a method more closely aligned with that underlying the proof of Theorem 4.

To describe our result, we must recall that the elements of $\mathcal{F}_{-N..N}$ may be partitioned according to the isomorphism class of the representation of $G = \text{GL}_2(\mathbb{Q}_p)$ that they generate. Any such representation has unramified central character,[3] and for N sufficiently large, is (isomorphic to) either:

- a (ramified) principal series representation (see Section 5.3), or
- a (supercuspidal) discrete series representation.

(See for instance [Schmidt 2002].) A (computable) positive proportion of elements of $\mathcal{F}_{-N..N}$ belongs to either category. The dichotomy here is analogous to that on $\text{SL}_2(\mathbb{Z}) \backslash \text{SL}_2(\mathbb{R})$ between Maass forms (principal series) and holomorphic forms (discrete series).

Theorem 7 (main result). *The conclusion of Conjecture 3 holds if φ_N belongs to the principal series.*

Theorem 7 represents the first genuine instance of arithmetic quantum unique ergodicity in the level aspect on a compact arithmetic quotient and also the first on any p-adic arithmetic quotient. It says that for a sequence $\varphi_N \in \mathcal{F}_{-N..N}$ belonging to the principal series, the corresponding L^2-masses equidistribute under pushforward to any fixed space $Y_{-n..n}$.

Remark 8. Our result might be described concisely as *arithmetic quantum unique ergodicity on the path space over the fixed regular graph* (Y, T_p) and as contributing to the growing literature concerning quantum chaos on regular graphs (see [Brooks and Lindenstrauss 2010; 2013; Anantharaman and Le Masson 2015]). Alternatively, one could fix an auxiliary split prime $\ell \neq p$, regard $(Y_{-N..N}, T_\ell)$ as traversing an inverse system of $(\ell + 1)$-regular graphs, and interpret Theorem 17 as a form of arithmetic quantum unique ergodicity for such a sequence of graphs.

Remark 9. Assuming the multiplicity hypothesis that an element $\varphi \in \mathcal{F}_{-N..N}$ generating an irreducible principal series representation of G is automatically an eigenfunction of the T_ℓ for $\ell \neq p$ (which is inspired by analogy from the conjectural simplicity of the spectrum of the Laplacian on $\text{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$), Theorem 17 may be understood as telling us something new about individual finite graphs (Y, T_p), such as those pictured above, together with their realization as $\Gamma \backslash G / K$.

As indicated already, the proof of Theorem 7 is patterned on that of Theorem 4. An important ingredient in the proof of Theorem 4 is the existence of a measure μ on $\Gamma' \backslash \text{SL}_2(\mathbb{R})$, called a *microlocal lift*, with the properties:

- μ lifts the measure $\lim_{j \to \infty} \mu_{\varphi_j}$ on $\Gamma' \backslash \mathbb{H}$.

---

[3] One may verify that "unramified central character" implies "trivial central character" in the present setup, but this special feature will not play an important role for us.

- μ is invariant under right translation by the diagonal subgroup of $\mathrm{SL}_2(\mathbb{R})$.

- $(\mu_{\varphi_j})_j \mapsto \mu$ is compatible with the Hecke operators (see [Silberman and Venkatesh 2007, Theorem 1.6] for details); this third property is that which is not obviously satisfied by the classical construction via charts and pseudodifferential calculus.

The known construction of μ with such properties, due to Zelditch and Wolpert (see [Zelditch 1987; Wolpert 2001; Lindenstrauss 2001]) and generalized by Silberman–Venkatesh [2007], relies heavily upon explicit calculation with raising and lowering operators in the Lie algebra of $\mathrm{SL}_2(\mathbb{R})$, which have no obvious p-adic analogue. One point of this paper is to introduce such an analogue and to investigate systematically its relationship to the classical theory of local newvectors. (The restriction to principal series in Theorem 7 then arises for the same reason that Lindenstrauss's argument does not apply to holomorphic forms of large weight: the absence of a "microlocal lift" invariant by a split torus.) The resulting construction may be of independent interest; for instance, it should have applications to the test vector problem (see Section 1.5 and Remark 50).

A curious subtlety of the argument, to be detailed further in Remark 26, is that the "lift" we construct is not a lift in the traditional sense (except against spherical observables, and even then only for $p \neq 2$). It instead satisfies a weaker "equidistribution implication" property which suffices for us. This subtlety is responsible for the most technical component of the argument (Section 6.3).

In the remainder of Section 1 we formulate our main result in a slightly more general setup (Section 1.2), introduce a key tool (Section 1.3), give an overview of the proof (Section 1.4), interpret our results in terms of L-functions (Section 1.5), and record some further remarks and open questions (Section 1.6).

1.2. *Main results: general form.* In this section we formulate a generalization of Theorem 4 in representation-theoretic language, which we adopt for the remainder of the paper.

Definition 10. Define the compact open subgroup

$$K_{m..m'} := \begin{bmatrix} \mathfrak{o} & \mathfrak{p}^{-m} \\ \mathfrak{p}^{m'} & \mathfrak{o} \end{bmatrix}^{\times}, \quad \mathfrak{o} := \mathbb{Z}_p, \ \mathfrak{p} := p\mathbb{Z}_p \tag{4}$$

of G. Each such subgroup is conjugate to $K_{0..n}$ for $n = m' - m \geqslant 0$, which is in turn analogous to the congruence subgroup $\Gamma_0(p^n)$ of $\mathrm{SL}_2(\mathbb{Z})$. Assuming (1), one has compatible bijections

$$X/K_{m..m'} = \Gamma\backslash G/K_{m..m'} \xrightarrow{\cong} Y_{m..m'},$$

$$\Gamma g K_{m..m'} \mapsto (x_m \to x_{m+1} \to \cdots \to x_{m'}) \text{ where } x_j := \Gamma g \begin{pmatrix} p^{-j} & \\ & 1 \end{pmatrix} K,$$

with $Y_{m..m'}$ as defined in (2).

Definition 11. The space $\mathcal{A}(X)$ of *smooth* functions on X consists of all functions $\varphi : X \to \mathbb{C}$ that are right-invariant under some open subgroup of G. An *eigenfunction* on X is an element $\varphi \in \mathcal{A}(X)$ that is a T_ℓ-eigenfunction for each ℓ and that generates an irreducible representation of G under the right translation action $g\varphi(x) := \rho_{\mathrm{reg}}(g)\varphi(x) := \varphi(xg)$. The *uniform measure* on X, denoted simply $\int_X$, is the probability

Haar coming from the G-action. An element $\varphi \in \mathcal{A}(X)$ is L^2-*normalized* if $\int_X |\varphi|^2 = 1$. In that case, the L^2-*mass* of φ is the probability measure μ_φ on X given by $\mu_\varphi(\Psi) := \int_X \Psi |\varphi|^2$. Convergence of measures always refers to the weak sense, i.e., $\lim_{n \to \infty} \mu_n = \mu$ if for each fixed $\Psi \in \mathcal{A}(X)$, $\lim_{n \to \infty} \mu_n(\Psi) = \mu(\Psi)$. A sequence of measures *equidistributes* if it converges to the uniform measure.

Definition 12. We denote by $\mathcal{H} \subseteq \mathrm{End}(\mathcal{A}(X))$ the ring generated by $\rho_{\mathrm{reg}}(G)$ and the T_ℓ, so that an eigenfunction in the sense of Definition 11 is an element of $\mathcal{A}(X)$ that generates an irreducible $\mathcal{H}$-submodule. We denote by $A(X)$ the set of irreducible $\mathcal{H}$-submodules of $\mathcal{A}(X)$, by $A_0(X) \subseteq A(X)$ the subset consisting of those that are not one-dimensional, and by $\mathcal{A}_0(X) \subseteq \mathcal{A}(X)$ the sum of the elements of $A_0(X)$, or equivalently, the orthogonal complement of the one-dimensional irreducible submodules.

A theorem of Eichler/Jacquet–Langlands implies that each $\pi \in A(X)$ occurs in $\mathcal{A}(X)$ with multiplicity one, so that $\mathcal{A}(X) = \bigoplus_{\pi \in A(X)} \pi$ and $\mathcal{A}_0(X) = \bigoplus_{\pi \in A_0(X)} \pi$. The one-dimensional elements of $A(X)$ are given by $\mathbb{C}(\chi \circ \det)$ for each character χ of the compact group $\mathbb{Q}_p^\times / \det(\Gamma)$, thus $A(X) = \{\mathbb{C}(\chi \circ \det)\} \bigsqcup A_0(X)$.

Definition 13. Let $\chi_\pi : \mathbb{Q}_p^\times \to \mathbb{C}^\times$ denote the central character of π. For $\pi \in A_0(X)$, the *conductor* of π has the form $C(\pi) = p^{c(\pi)}$, where $c(\pi)$ is the smallest nonnegative integer with the property that π contains a nonzero vector φ satisfying $g\varphi = \chi_\pi(d)g$ for all $g = \left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix} \right) \in K_{0..c(\pi)}$ [Casselman 1973a; Schmidt 2002].

Definition 14. Let $\pi \in A_0(X)$. For integers m, m', a vector $\varphi \in \pi$ will be called a *newvector of support* $m..m'$ if $m' - m = c(\pi)$ and $g\varphi = \chi_\pi(d)\varphi$ for all $g = \left(\begin{smallmatrix} * & * \\ * & d \end{smallmatrix} \right) \in K_{m..m'}$. Local newvector theory [Casselman 1973a; Schmidt 2002] implies that the space of such vectors is one-dimensional, so if φ is L^2-normalized, then the L^2-mass μ_φ depends only upon π and $m..m'$, not φ. A vector $\varphi \in \pi$ will be called a *generalized newvector* if it is a newvector of support $m..m'$ for some m, m'. (We include the adjective "generalized" only to indicate explicitly that we are not necessarily referring to the traditional case $m..m' = 0..c(\pi)$, which will play no distinguished role here.)

Remark 15. The newvectors of support $m..m'$ that generate representations with unramified central character may be characterized more simply as those eigenfunctions $\varphi \in \mathcal{A}(X)$ (in the sense of Definition 11) which:

 (1) are $K_{m..m'}$-invariant, or equivalently, descend to $\varphi : Y_{m..m'} \to \mathbb{C}$, and

 (2) are orthogonal to pullbacks from $Y_{n..n'}$ whenever $n..n' \subsetneq m..m'$.

(The proof of this characterization is the same as the proof that local newvector theory [Casselman 1973a] recovers classical Atkin–Lehner theory [1970].) Under the torsion-freeness assumption (1), "orthogonal" can be taken to mean with respect to the normalized counting measure on $Y_{m..m'}$; in general, one should take that induced by the uniform measure on X. In this sense, Definition 14 is consistent with Definition 1.

Definition 16. We say that $\pi \in A_0(X)$ *belongs to the principal series* if the corresponding representation of G does (see Section 5.3).

Theorem 17 (equidistribution of newvectors II). *Let $\pi_j \in A_0(X)$ ($j = 1, 2, 3, \dots$) be a sequence with $C(\bar{\pi}_j \times \pi_j) \to \infty$. Assume that π_j belongs to the principal series. Let $\varphi_j \in \pi_j$ be an L^2-normalized generalized newvector. Then μ_{φ_j} equidistributes as $j \to \infty$.*

Theorem 17 specializes to Theorem 7 upon requiring the central character of π_j to be unramified and restricting to newvectors of support $m..m' = -N..N$ for some N.

Remark 18. Unlike earlier works such as [Nelson 2011; Nelson et al. 2014; Hu 2018], we have allowed arbitrary central characters in Theorem 17. We note that the case of the argument in which the conductor of the central character is as large as possible relative to that of the representation is a bit more technically challenging than the others; see (26) and following.

Remark 19. Cases of Theorem 17 in which $m..m'$ is highly unbalanced, such as the most traditional case $m..m' = 0..n$ analogous to Theorem 6, are easier: they follow, sometimes with a power savings, from the triple product formula, the convexity bound for triple product L-functions, and nontrivial local estimates as in [Nelson et al. 2014; Hu 2018]. Cases in which $m..m'$ is balanced, such as the case $m..m' = -N..N$ illustrated in Section 1.1, do not follow from such local arguments and require the new ideas introduced here. This phenomenon is comparable to how the mass equidistribution on a hyperbolic surface $\Gamma' \backslash \mathbb{H}$ of a weight k vector in a principal series $\pi \hookrightarrow L^2(\Gamma' \backslash \mathrm{SL}_2(\mathbb{R}))$ of parameter $t \to \infty$ follows from essentially local means for $t/k = o(1)$ but not for $k = 0$, or even for $k \ll t$; see [Zelditch 1992; Reznikov 2001] for some discussion along such lines. See also Remark 30 and footnote 12.

1.3. *p-adic microlocal lifts.* We turn to the key definitions that power the proof of the above results. We develop them slightly more precisely and algebraically than is strictly necessary for the consequences indicated above.

Let k be a nonarchimedean local field with ring of integers $\mathfrak{o}$, maximal ideal $\mathfrak{p}$, normalized valuation $v : k \twoheadrightarrow \mathbb{Z} \cup \{+\infty\}$, and $q := \#\mathfrak{o}/\mathfrak{p}$. (The case $(k, \mathfrak{o}, \mathfrak{p}, q) = (\mathbb{Q}_p, \mathbb{Z}_p, p\mathbb{Z}_p, p)$ is relevant for the above application.)

To a generic irreducible representation π of $\mathrm{GL}_n(k)$ one may attach a *conductor* $C(\pi) = q^{c(\pi)}$, with $c(\pi) \in \mathbb{Z}_{\geqslant 0}$; we recall this assignment in the most relevant case $n = 2$ in Section 5.3 and Section 5.5. One also defines $c(\omega)$ for each character ω of $\mathfrak{o}^\times$; it is the smallest integer n for which ω has trivial restriction to $\mathfrak{o}^\times \cap 1 + \mathfrak{p}^n$.

For context, we record the local form of Definition 14:

Definition 20 (newvectors). A vector v in an irreducible generic representation π of $\mathrm{GL}_2(k)$ is a *newvector of support $m..m'$* if $m' - m = c(\pi)$ and

$$\pi(g)v = \chi_\pi(d)v \quad \text{for all } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathfrak{o}) \cap \begin{bmatrix} \mathfrak{o} & \mathfrak{p}^{-m} \\ \mathfrak{p}^{m'} & \mathfrak{o} \end{bmatrix}.$$

A *generalized newvector* is a newvector of some support.

Fix now for each nonnegative integer N a partition $N = N_1 + N_2$ into nonnegative integers N_1, N_2 with the property that $N_1, N_2 \to \infty$ as $N \to \infty$. The precise choice is unimportant; one might take

$N_1 := \lfloor N/2 \rfloor$, $N_2 := \lceil N/2 \rceil$ for concreteness. Using this choice, we introduce the following class of vectors:

Definition 21 (microlocal lifts). Let π be a $GL_2(k)$-module. A vector $v \in \pi$ shall be called a *microlocal lift* if:

- it is nonzero,

- it generates an irreducible admissible representation of $GL_2(k)$, and

- there is a positive integer N and characters ω_1, ω_2 of $\mathfrak{o}^\times$ so that $c(\omega_1/\omega_2) = N$ and

$$\pi(g)v = \omega_1(a)\omega_2(\det(g)/a)v \text{ for all } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathfrak{o}) \cap \begin{bmatrix} \mathfrak{o} & \mathfrak{p}^{N_1} \\ \mathfrak{p}^{N_2} & \mathfrak{o} \end{bmatrix}.$$

In that case, we refer to N as the *level* and (ω_1, ω_2) as the *orientation* of v.

The observation that the special case $\omega_1 = 1$ of Definition 21 is similar to Definition 20 leads easily to the following characterization of microlocal lifts as twists of generalized newvectors from "extremal principal series" representations "$1 \boxplus \chi$" (see Section 6.1 for the proof):

Lemma 22. *An irreducible admissible representation π of $GL_2(k)$ contains a microlocal lift if and only if π is an irreducible principal series representation $\pi \cong \chi_1 \boxplus \chi_2$ for which $N := c(\bar\pi \otimes \pi)/2 = c(\chi_1/\chi_2)$ is nonzero. In that case, the set of microlocal lifts is a disjoint union $\mathbb{C}^\times \varphi_+ \sqcup \mathbb{C}^\times \varphi_-$, where*

$$\mathbb{C}^\times \varphi_+ = \{\text{microlocal lifts in } \pi \text{ of level } N \text{ and orientation } (\omega_1, \omega_2)\},$$

$$\mathbb{C}^\times \varphi_- = \{ \text{microlocal lifts in } \pi \text{ of level } N \text{ and orientation } (\omega_2, \omega_1) \},$$

with $\omega_i := \chi_i|_{\mathfrak{o}^\times}$. Explicitly, $\mathbb{C}^\times \varphi_+$ is the inverse image under the nonequivariant twisting isomorphism $\pi \to \pi \otimes \chi_1^{-1} \cong 1 \boxplus \chi_1^{-1}\chi_2$ of the set of nonzero newvectors of support $-N_1..N_2$. The set $\mathbb{C}^\times \varphi_-$ is described similarly, with the roles of ω_1 and ω_2 reversed.

Remark 23. We briefly compare with the archimedean analogue inspiring Definition 21; a more complete exposition of this analogy seems beyond the scope of this article. Let π be a principal series representation of $PGL_2(\mathbb{R})$ of parameter $t \to \pm\infty$ with lowest weight vector φ_0 corresponding to a spherical Maass form of eigenvalue $\frac{1}{4} + t^2$ on some hyperbolic surface. The Zelditch–Wolpert construction[4] of a microlocal lift φ_1 of φ_0 is given up to normalizing factors in terms of standard raising/lowering operators X^n for $n \in \mathbb{Z}$ (see [Wolpert 2001; Lindenstrauss 2001]) by $\varphi_1 := \sum_{n:|n| \leqslant t_1} X^n \varphi_0$, where $|t| = t_1 t_2$ with $t_1, t_2 \to \infty$ as $|t| \to \infty$. The choice $\varphi_2 := \sum_{n:|n| \leqslant t_1} (-1)^n X^n \varphi_0$ also works. The analogue of $(|t|, \varphi_1, \varphi_2, |.|^{it}, |.|^{-it})$ in the notation of Definition 21 and Lemma 22 is $(q^N, v_1, v_2, \chi_1, \chi_2)$ with $q := \#\mathfrak{o}/\mathfrak{p}$ and $v_1, v_2 \in \pi$ microlocal lifts of respective orientations (ω_1, ω_2), (ω_2, ω_1). The analogy may be obtained by comparing how $GL_2(\mathfrak{o})$ acts on v_1, v_2 to how the Lie algebra of $PGL_2(\mathbb{R})$ acts on φ_1, φ_2. The factorization $|t| = t_1 t_2$ is roughly analogous to the partition $N = N_1 + N_2$. It is also instructive to compare the formulas for φ_1, φ_2 in their induced models with those of Section 6.2.

---

[4] We discuss here only the "positive measure" incarnation of that construction rather than the "distributional" one.

Remark 24. Le-Masson [2014] and Anantharaman–Le-Masson [2015] have introduced a notion of microlocal lifts on regular graphs and used that notion to prove some analogues of the quantum ergodicity theorem. Definition 21 serves different aims in that we do not explicitly vary the graph (except perhaps in the second sense indicated in Remark 8); it would be interesting to extend it further and compare the two notions on any domain of overlap.

For the remainder of Section 1.3, take $k = \mathbb{Q}_p$, so that $\mathrm{GL}_2(k) = G$. Definition 21 applies to $\pi \in A_0(X)$.

Theorem 25 (basic properties of microlocal lifts). *Let N traverse a sequence of positive integers tending to ∞, and let $\varphi \in \pi \in A_0(X)$ be an L^2-normalized microlocal lift of level N on X with L^2-mass μ_φ:*

- *__Diagonal invariance__: Any weak subsequential limit of the sequence of measures μ_φ is $a(\mathbb{Q}_p^\times)$-invariant.*

- *__Lifting property__: Suppose temporarily that $p \neq 2$, so that $v(2) = 0$. Let $\varphi' \in \pi$ be an L^2-normalized newvector of support $-N..N$, and let $\Psi \in \mathcal{A}(X)^K$ be independent of N and right-invariant by $K := \mathrm{GL}_2(\mathbb{Z}_p)$. Then*

$$\lim_{N \to \infty} (\mu_\varphi(\Psi) - \mu_{\varphi'}(\Psi)) = 0.$$

- *__Equidistribution implication__: Suppose that μ_φ equidistributes as $N \to \infty$. Let $\varphi' \in \pi$ be an L^2-normalized generalized newvector. Then $\mu_{\varphi'}$ equidistributes as $N \to \infty$.*

Theorem 25 is established in Section 7 after developing the necessary local preliminaries in Section 5 and Section 6. The proof involves uniqueness of invariant trilinear forms[5] on GL_2 and stationary phase analysis of local Rankin–Selberg integrals. Theorem 25 is essentially local, i.e., does not exploit the arithmeticity of $\Gamma \leqslant G$, and is stated here in a global setting only for convenience; see Theorem 49 for a local analogue.

Remark 26. The "lifting property" of Theorem 25 has been included only for the sake of illustration; it is not strictly necessary for the logical purposes of this paper. We have assumed $p \neq 2$ in its statement because the corresponding assertion is false when $p = 2$. For general p and nonspherical observables Ψ, there does not appear to be any simple relationship between the quantities $\mu_\varphi(\Psi)$ and $\mu_{\varphi'}(\Psi)$ except that convergence to $\int_X \Psi$ of the first implies that of the second (the "equidistribution implication"). The "lifting" relationship here is thus more subtle than that in [Lindenstrauss 2006b].

1.4. *Equidistribution of microlocal lifts.* Our core result (from which the others are ultimately derived) is the following:

Theorem 27 (equidistribution of microlocal lifts). *Let N traverse a sequence of positive integers tending to ∞. Let $\varphi \in \mathcal{A}(X)$ be an L^2-normalized microlocal lift of level N on X. Then μ_φ equidistributes.*

---

[5] It should be possible to avoid this comparatively deep fact in the proof of the first part of Theorem 25, but it is required by the application to subconvexity (Theorem 29), and the calculations required by that application already suffice here.

The proof depends upon an analogue of Lindenstrauss's celebrated result [2006b]. Here and throughout this article, "entropy" refers to the Kolmogorov–Sinai entropy of a measurable dynamical system (see, e.g., [Lindenstrauss 2006a, §8]).

Theorem 28 (measure classification). *Let μ be a probability measure on X, invariant by the center of G, with the properties*:

(1) *μ is $a(\mathbb{Q}_p^{\times})$-invariant.*

(2) *μ is T_ℓ-recurrent for some split prime $\ell \neq p$.*

(3) *The entropy of almost every ergodic component of μ is positive for the $a(\mathbb{Q}_p^{\times})$-action.*

Then μ is the uniform measure.

We explain in Section 2 the specialization of Theorem 28 from a result of Einsiedler–Lindenstrauss [2008, Theorem 1.5]. To deduce Theorem 27, we apply Theorem 28 with μ any weak limit of the L^2-masses of a sequence of L^2-normalized microlocal lifts of level tending to ∞. Since X is compact, μ is a probability measure. The invariance hypothesis follows from the diagonal invariance of Theorem 25, while the T_ℓ-recurrence and positive entropy hypotheses are verified below in Section 3 and Section 4. The proof of our main result Theorem 27 is then complete. Theorem 27 and the equidistribution implication of Theorem 25 imply Theorem 17.

1.5. *Estimates for L-functions.* For definitions of the L-functions and local distinguishedness, see [Piatetski-Shapiro and Rallis 1987; Ichino 2008]. We record the following because it provides an unambiguous benchmark of the strength of our results.

Theorem 29 (weakly subconvex bound). *Fix $\sigma \in A_0(X)$. Let $\pi \in A_0(X)$ traverse a sequence with $C(\bar{\pi} \times \pi) \to \infty$. Assume that π belongs to the principal series and that $\sigma \otimes \bar{\pi} \otimes \pi$ is locally distinguished. Then*

$$\frac{L(\sigma \times \bar{\pi} \times \pi, 1/2)}{L(\text{ad}\,\pi, 1)^2} = o(C(\sigma \times \bar{\pi} \times \pi)^{1/4}). \tag{5}$$

The previously best known estimate for the LHS of (5) is the general weakly subconvex estimate of Soundararajan [2010], specializing here to $L \ll C^{1/4}/(\log C)^{1-\varepsilon}$ with $L := L\left(\sigma \times \bar{\pi} \times \pi, \frac{1}{2}\right)$, $C := C(\sigma \times \bar{\pi} \times \pi)$. The bound (5) improves upon that estimate in the unlikely (but difficult to exclude) case that $L(\text{ad}\,\pi, 1)$ is exceptionally small, which turns out to be the most difficult one for equidistribution problems; see [Holowinsky and Soundararajan 2010] for further discussion.

Theorem 27 implies Theorem 29 after a local calculation with the triple product formula (see Section 7); in fact, the calculation shows that the two results are equivalent.

Remark 30. Theorem 29 implies Theorem 17, but the converse does not hold in general; a special case of the failure of that converse was noted and discussed at length in [Nelson et al. 2014, §1]. The present work may thus be understood as clarifying that discussion: the equivalence between subconvexity and equidistribution problems in the depth aspect is restored by working not with newvectors, but instead with the p-adic microlocal lifts introduced here.

1.6. *Further remarks.*

Remark 31. Theorems 17 and 27 apply only to sequences of vectors φ that generate irreducible $\mathcal{H}$-modules. One can ask whether the conclusion holds under the (hypothetically) weaker assumption that φ generates an irreducible G-module. The problem formulated this way makes sense for any finite volume quotient $\Gamma' \backslash G$, not necessarily arithmetic; an affirmative answer would represent a p-adic analogue of the Rudnick–Sarnak quantum unique ergodicity conjecture [1994]. In that direction, we note that the method of Brooks–Lindenstrauss [2014] should apply in our setting, allowing one to relax the hypothesis of irreducibility under the full Hecke algebra to that under a single auxiliary Hecke operator T_ℓ for some fixed split prime $\ell \neq p$.

An affirmative answer to the question raised above would, by the (proof of the) equidistribution implication of Theorem 25, imply that the conclusion of Conjecture 3 remains valid on possibly nonarithmetic quotients $\Gamma' \backslash G$ under the hypothesis that $\varphi_N \in L^2_{\text{new}}(Y_{-N..N})$ traverses a sequence of unit vectors that generate principal series representations of $\mathrm{GL}_2(\mathbb{Q}_p)$. (The analogous assertion for supercuspidal representations fails because such representations may be shown to occur with large multiplicity. A similar phenomenon is responsible for the subtlety in formulating holomorphic analogues of quantum unique ergodicity; see [Luo and Sarnak 2003; Holowinsky and Soundararajan 2010].)

Remark 32. Our results apply to principal series representations of conductor p^N with p fixed and $N \to \infty$. A natural question is whether one can establish analogous results for N fixed, such as $N = 100$, and $p \to \infty$. We highlight here the weaker question of whether one can establish equidistribution (in a balanced case, cf. Remark 19) as $N \to \infty$ for p satisfying $p \leqslant p_0(N)$ for some $p_0(N)$ tending *effectively* to ∞ as $N \to \infty$. Our results and a diagonalization argument imply an ineffective analogue.

Remark 33. The crucial local results of this article have been formulated and proved in generality, i.e., over any nonarchimedean local field. On the other hand, we have assumed in our global results that the subgroup Γ of G was constructed from a *maximal order* in a quaternion algebra *over* $\mathbb{Q}$. We expect that our results hold more generally:

(1) The statements and proofs of all our results except Theorem 29 extend straightforwardly to the case that Γ arises from a fixed *Eichler order* in a quaternion algebra over $\mathbb{Q}$. To extend Theorem 29 in that direction would require some local triple product estimates at the "uninteresting" primes $\ell \neq p$ which we do not pursue here.

(2) Our results should extend to Eichler orders in totally definite quaternion algebras over *totally real number fields*, but some mild care is required in formulating such extensions when the class group has nontrivial 2-torsion: as observed in a related context in [Nelson 2012], there are sequences of dihedral forms that fail to satisfy the most naive formulation of quantum unique ergodicity.

(3) We expect our results extend to automorphic forms on definite quaternion algebras having fixed nontrivial infinity type; such an extension would require a more careful study of the measure classification input in Section 2.

(4) Over function fields, analogues of our results should follow more directly and in quantitatively stronger forms from Deligne's theorem and extensions of the triple product formula to the function field setting.

We leave such extensions to the interested reader.

Organization of this paper. We verify the measure-classification (Theorem 28) and its hypotheses in Section 2, Section 3, and Section 4. We review the representation theory of $GL_2(k)$ in Section 5. In Section 6 and Section 7, we prove our core results, notably Theorem 25, and their applications. Some additional results of independent interest are recorded along the way.

2. Measure classification

The purpose of this section is to deduce Theorem 28 from the following specialization to $\mathbb{Q}_p$ of a result of Einsiedler–Lindenstrauss [2008, Theorem 1.5]:

Theorem 34. *Let $G = G_1 \times G_2$, where G_1 is a semisimple linear algebraic group over $\mathbb{Q}_p$ with $\mathbb{Q}_p$-rank 1 and G_2 is a characteristic zero S-algebraic group. Let $\Gamma' \subset G$ be a discrete subgroup. Let A_1 be a $\mathbb{Q}_p$-split torus of G_1 and let χ be a nontrivial $\mathbb{Q}_p$-character of A_1 that can be extended to $C_{G_1}(A_1)$. Let $M_1 = \{h \in C_{G_1}(A_1) : \chi(h) = 1\}$. Let ν be an A_1-invariant, G_2-recurrent probability measure on $\Gamma' \backslash G$ such that*:

(1) *almost every A_1-ergodic component of ν has positive entropy with respect to some $a \in A_1$ with $|\chi(a)| \neq 1$, and*

(2) *for ν-almost every $x \in \Gamma' \backslash G$, the group $\{h \in M_1 \times G_2 : xh = x\}$ is finite.*

Then ν is a convex combination of homogeneous measures, each of which is supported on an orbit of a subgroup H which contains a finite index subgroup of a semisimple algebraic subgroup of G_1 of $\mathbb{Q}_p$-rank one.

To deduce Theorem 28 from Theorem 34 requires no new ideas, but we record a complete verification for completeness.

2.1. ***Consequences of strong approximation.*** Recall that R is a maximal order in a definite quaternion algebra B. (For general background on quaternion algebras we mention [Vignéras 1980; Voight 2018; Nelson 2015, §2.2].)

For a prime p, we shall use the notations $B_p := B \otimes_{\mathbb{Q}} \mathbb{Q}_p$, $R_p := R \otimes_{\mathbb{Z}} \mathbb{Z}_p$. A superscripted (1) denotes "norm one elements," e.g., $B_p^{(1)} := \{b \in B_p^\times : \mathrm{nr}(b) = 1\}$. Denote by $\mathbb{A}_f$ the finite adele ring of $\mathbb{Q}$ and $\hat{B} := B \otimes_{\mathbb{Q}} \mathbb{A}_f$. (Thus $B_{\mathbb{A}} := B_\infty \times \hat{B}$ with $B_{\mathbb{A}} := B \otimes_{\mathbb{Q}} \mathbb{A}$, $B_\infty := B \otimes_{\mathbb{Q}} \mathbb{R}$, and $\mathbb{A}$ the adele ring of $\mathbb{Q}$.) Regard $B^\times$, $B_p^\times$, $R_p^\times$ as subsets of $\hat{B}^\times$ in the standard way.

Lemma 35. *Let U be a subgroup of $\hat{B}^\times$ for which*:

(i) *There is a prime p that splits B for which U contains an open subgroup of $\hat{B}^{(1)}$ containing $B_p^{(1)}$.*

(ii) *The image* $\mathrm{nr}(U)$ *of* U *under the reduced norm* $\mathrm{nr} : \hat{B}^\times \to \mathbb{A}_f^\times$ *satisfies* $\mathbb{Q}_+^\times \mathrm{nr}(U) = \mathbb{A}_f^\times$.

Then $B^\times U = \hat{B}^\times$.

Proof. It is known (e.g., by Hasse–Minkowski) that $\mathrm{nr} : B^\times \to \mathbb{Q}_+^\times$ is surjective. Let $b \in \hat{B}^\times$ be given. By (ii), there exists $\gamma \in B^\times$ and $h \in U$ for which $\gamma b h \in \hat{B}^{(1)}$. Let p be as in (i). The strong approximation theorem [Kneser 1966], applied to the simply connected semisimple algebraic group $B^{(1)}$ and its noncompact factor $B_p^{(1)}$, implies that $B^{(1)} B_p^{(1)}$ is dense in $\hat{B}^{(1)}$. By (i), we may write $\gamma b h = \delta h'$ for some $\delta \in B^{(1)}$ and $h' \in U$. Therefore $b = \gamma^{-1} \delta h' h^{-1}$ belongs to $B^\times U$, as required. $\qquad\square$

Let p be a split prime for B. For any prime ℓ, one has $\mathrm{nr}(B_\ell^\times) = \mathbb{Q}_\ell^\times$; because R is a maximal order (in particular, an Eichler order), one has moreover that $\mathrm{nr}(R_\ell^\times) = \mathbb{Z}_\ell^\times$. The hypotheses of Lemma 35 thus apply to $U = B_p^\times \prod_{\ell \neq p} R_\ell^\times$: (i) is clearly satisfied, while (ii) follows from the consequence $\mathbb{Q}_+^\times \mathbb{Q}_p^\times \prod_{\ell \neq p} \mathbb{Z}_\ell^\times = \mathbb{A}_f^\times$ of strong approximation for the ideles. For similar but simpler reasons, the hypotheses apply also to $U = B_p^\times B_\ell^\times \prod_{q \neq \ell, p} R_q^\times$. Thus

$$B^\times B_p^\times \prod_{\ell \neq p} R_\ell^\times = \hat{B}^\times = B^\times B_p^\times B_\ell^\times \prod_{q \neq \ell, p} R_q^\times.$$

We have $B^\times \cap \prod_{\ell \neq p} R_\ell^\times = R[1/p]^\times$ and $B^\times \cap \prod_{q \neq \ell, p} R_q^\times = R[1/p\ell]^\times$, whence the natural identifications

$$R[1/p]^\times \backslash B_p^\times / \mathbb{Q}_p^\times = B^\times \backslash \hat{B}^\times / \mathbb{Q}_p^\times \prod_{\ell \neq p} R_\ell^\times = R[1/p\ell]^\times \backslash B_p^\times B_\ell^\times / \mathbb{Q}_p^\times R_\ell^\times. \tag{6}$$

Since $\mathbb{Z}[1/p\ell]^\times \mathbb{Q}_p^\times \mathbb{Z}_\ell^\times = \mathbb{Q}_p^\times \mathbb{Q}_\ell^\times$, the RHS of (6) is unaffected by further reduction modulo $\mathbb{Q}_\ell^\times$, i.e.,

$$R[1/p]^\times \backslash B_p^\times / \mathbb{Q}_p^\times = R[1/p\ell]^\times \backslash B_p^\times B_\ell^\times / \mathbb{Q}_p^\times \mathbb{Q}_\ell^\times R_\ell^\times. \tag{7}$$

2.2. *Deduction of Theorem 28.* Let p be a split prime for B. Identify $B_p^\times = \mathrm{GL}_2(\mathbb{Q}_p)$ and $X = \Gamma \backslash \mathrm{GL}_2(\mathbb{Q}_p)$ as in Section 1. Let μ be a measure on X satisfying the hypotheses of Theorem 28. It is invariant under the diagonal torus of $\mathrm{GL}_2(\mathbb{Q}_p)$, which generates the latter modulo $\mathrm{SL}_2(\mathbb{Q}_p)$, so to prove that μ is the uniform measure, we need only verify that it is $\mathrm{SL}_2(\mathbb{Q}_p)$-invariant. To that end, we apply Theorem 34: Set $G_1 := \mathrm{PGL}_2(\mathbb{Q}_p) = B_p^\times / \mathbb{Q}_p^\times$, $G_2 := \mathrm{PGL}_2(\mathbb{Q}_\ell) = B_\ell^\times / \mathbb{Q}_\ell^\times$, $G := G_1 \times G_2$. Recall that $\Gamma = R[1/p]^\times$. Take for Γ' the image of $R[1/p\ell]^\times$ in G. By strong approximation in the form (7), we may identify $\Gamma \backslash \mathrm{GL}_2(\mathbb{Q}_p) / \mathbb{Q}_p^\times$ with $\Gamma' \backslash G / \mathrm{PGL}_2(\mathbb{Z}_\ell)$ and μ with a right $\mathrm{PGL}_2(\mathbb{Z}_\ell)$-invariant measure ν on $\Gamma' \backslash G$. Our task is then to verify that ν is invariant by the image of $\mathrm{SL}_2(\mathbb{Q}_p)$. Take for A_1 the diagonal torus in G_1 and for $\chi : A_1 \to \mathbb{Q}_p^\times$ the map $\chi(\mathrm{diag}(y_1, y_2)) := y_1/y_2$. We have $C_{G_1}(A_1) = A_1$. The group M_1 is trivial, hence each $\{h \in M_1 \times G_2 : xh = x\}$ is trivial. The hypotheses of Theorem 28 are satisfied, so ν is invariant by some finite index subgroup H_1 of some semisimple algebraic subgroup of G_1 (of $\mathbb{Q}_p$-rank one) that contains A_1. The smallest such H_1 is the image of $\mathrm{SL}_2(\mathbb{Q}_p)$, so we conclude.

3. Recurrence

In this section we formulate and verify the T_ℓ-recurrence hypothesis required by Theorem 28. The argument here is as in [Lindenstrauss 2006b, §8] except that we allow general central characters; for completeness, we record a proof of the key estimate in that case. The proof is simple; a key insight of Lindenstrauss [2006b] is that the condition enunciated here is useful for the present purposes.

3.1. *Hecke operators.*

3.1.1. *Summary of facts.* For a positive integer n coprime to p, the Hecke operator $T_n \in \text{End}(\mathcal{A}(X))$ is defined by $T_n\varphi(x) := \sum_{\alpha \in M_n/M_1} \varphi(\alpha^{-1}x)$, where $M_n := R[1/p] \cap \text{nr}^{-1}(n\mathbb{Z}[1/p]^\times)$, so that $M_1 = \Gamma$. These operators commute with one another and also with $\rho_{\text{reg}}(G)$. Given a scalar element m, let us introduce the general abbreviation $z(m)$ for the corresponding quaternion. For $m \in \mathbb{Q}^\times$, we abbreviate $z(m) := \rho_{\text{reg}}(z(m))$. If $\ell \mid \text{disc } B$, then the operator T_ℓ is an involution modulo the action of the center, namely $T_\ell^2 = T_{\ell^2} = z(\ell^{-1})$; otherwise, T_ℓ is induced by a correspondence of degree $\ell + 1$. The adjoint of T_n is $T_n^* = z(n)T_n$, and one has the composition formula

$$T_m T_n = \sum_{d \in \mathbb{Z}_{\geqslant 1}: d \mid \gcd(m,n),\, \gcd(d, \text{disc } B)=1} d \cdot z(d^{-1}) T_{mn/d^2}. \tag{8}$$

3.1.2. *Derivations.* Since we are unaware of a convenient reference for the facts recalled above, we briefly indicate how they fall out from the adelic picture and the structure of the local Hecke algebras. (The reader is strongly encouraged to skip this section, which we have included only for completeness.) With notation as in Section 2.1, let us abbreviate $H_\ell := B_\ell^\times$, $J_\ell := R_\ell^\times$ and $H := \prod_{\ell \neq p} H_\ell$ and $J := \prod_{\ell \neq p} J_\ell$, so that J is a compact open subgroup of H and $G \times H = \hat{B}^\times$. By strong approximation as in Section 2.1, the map $G \ni x \mapsto (x, 1) \in G \times H$ induces a bijection $X = \Gamma\backslash G \xrightarrow{\sim} B^\times\backslash(G \times H/J)$. In this way, we may identify each $\varphi \in \mathcal{A}(X)$ with a right-J-invariant function $\Phi : B^\times\backslash(G \times H) \to \mathbb{C}$, called the *lift* of φ. Equip H with the Haar measure assigning volume one to J. Then the algebra $\mathcal{H} := C_c^\infty(J\backslash H/J)$, under convolution, acts on $\mathcal{A}(X)$ by translating the corresponding lifts. The algebra $\mathcal{H}$ decomposes as a restricted tensor product of local Hecke algebras $\mathcal{H}_\ell = C_c^\infty(J_\ell\backslash H_\ell/J_\ell)$, where again we normalize so that J_ℓ has volume one. These local Hecke algebras may be described as follows:

- Suppose $\ell \mid \text{disc}(B)$, i.e., that ℓ does not split B, so that B_ℓ is a quaternion division algebra. Then J_ℓ is the kernel of the map $H_\ell \to \mathbb{Z}$ sending an element to the valuation of its reduced norm. This induces an isomorphism from $\mathcal{H}_\ell$ to the group algebra $\mathbb{C}[\mathbb{Z}]$. In other words, $\mathcal{H}_\ell$ has a basis given by the characteristic functions T_{ℓ^n} of those $x \in H_\ell$ with reduced norm of valuation n, and we have $T_{\ell^m} T_{\ell^n} = T_{\ell^{m+n}}$. We note that $T_{\ell^{2n}}$ is the characteristic function of $J\ell z(\ell^n)J_\ell$, where as usual $z(y) \in H_\ell$ denotes the scalar element corresponding to $y \in \mathbb{Q}_\ell^\times$.

- Suppose $\ell \nmid \text{disc}(B)$, i.e., that ℓ splits B. Then $H_\ell \cong \text{GL}_2(\mathbb{Q}_\ell)$ and $J_\ell \cong \text{GL}_2(\mathbb{Z}_\ell)$. Let $T_{\ell^n} \in \mathcal{H}_\ell$ denote the characteristic function of $H_\ell^{(\ell^n)}$, where $H_\ell^{(\ell^n)}$ denotes the set of all $k \in R_\ell$ with reduced norm of

valuation n. Then $T_{\ell^m} T_{\ell^n} = \sum_{j=0}^{|m-n|} \ell^j z(\ell^j) T_{\ell^{m+n-2j}}$, with $z(y)$ as before. The Hecke algebra $\mathcal{H}_j$ is generated by the T_{ℓ^n} together with the characteristic functions of $J_\ell z(y) J_\ell$ taken over $y \in \mathbb{Q}_\ell^\times / \mathbb{Z}_\ell^\times$.

In summary, the algebra $\mathcal{H}$ is generated by:

- For each $m \in \prod_{\ell \neq p} \mathbb{Q}_\ell^\times / \mathbb{Z}_\ell^\times$, the characteristic function of $J z(m^{-1}) J = J z(m^{-1}) = z(m^{-1}) J$.

- For each $n \in \prod_{\ell \neq p} (\mathbb{Q}_\ell^\times \cap \mathbb{Z}_\ell)/\mathbb{Z}_\ell^\times$, the characteristic function of the double-J-coset

$$H^{(n)} := \left\{ k \in \prod_{\ell \neq p} R_\ell : \mathrm{nr}(\ell) \in n \prod_{\ell \neq p} \mathbb{Z}_\ell^\times \right\}.$$

Let us denote the operators on $\mathcal{A}(X)$ obtained in the first case by $\widetilde{z(m)}$ and in the second by $\widetilde{T}_n$. Since $\mathbb{Q}^\times \prod_\ell \mathbb{Z}_\ell^\times = \prod_\ell \mathbb{Q}_\ell^\times$, we may assume in the first case that m is represented by an element of $\mathbb{Q}^\times$ coprime to p; we then verify readily, using the identity $\Phi(x, z(m^{-1})) = \Phi(z(m)x, 1)$, that $\widetilde{z(m)} = z(m)$ as defined above. In the second case, we note first that we may assume that n is a positive integer coprime to p. Using strong approximation as in Section 2.1, we see then that the natural map $M_n/M_1 \to H^{(n)}/J$ is bijective. Decomposing $H^{(n)}$ into right J-cosets, it follows readily that $\widetilde{T}_n = T_n$. Thus the operators T_n and $z(m)$ generate the same subalgebra of $\mathrm{End}(\mathcal{A}(X))$ as $\mathcal{H}$ does. The relations stated in Section 3.1.1 follow from the corresponding local relations given above.

3.2. *Spherical averaging operators.* Let n be a positive integer coprime to p. The operator T_n on $\mathcal{A}(X)$ is induced by the *correspondence* on X, denoted also by T_n, given for $x \in X$ by the multiset (i.e., formal sum) $T_n(x) := \sum_{s \in M_n/\Gamma} s^{-1}x$. Thus $T_n\varphi(x) = \sum_{y \in T_n(x)} \varphi(y)$. Denote by M_n^{prim} the set of all *primitive* elements of M_n, i.e., those that are not divisible inside $R[1/p]$ by any divisor $d > 1$ of n. Then M_n^{prim} is right-invariant by Γ, and one has $M_n = \bigsqcup_{d^2 \mid n} z(d) M_{n/d^2}^{\mathrm{prim}}$. Denote by S_n the "Hecke sphere" correspondence $S_n(x) := \sum_{s \in M_n^{\mathrm{prim}}/\Gamma} s^{-1}x$; it likewise induces an operator S_n on $\mathcal{A}(X)$ given by $S_n\varphi(x) := \sum_{s \in M_n^{\mathrm{prim}}/\Gamma} \varphi(s^{-1}x) = \sum_{y \in S_n(x)} \varphi(y)$, and one has

$$T_n(x) = \sum_{d^2 \mid n} z(d^{-1}) S_{n/d^2}(x). \tag{9}$$

3.3. *Recurrence.* Let $\ell \neq p$ be a *split prime*, that is to say, a prime that splits the quaternion algebra underlying the construction of Γ, so that the Hecke operator T_ℓ has degree $\ell + 1$.

Definition 36. Let Z denote the center of G. A finite Z-invariant measure μ on X is called T_ℓ-recurrent if for each Borel subset $E \subseteq X$ and μ-almost every $x \in E$, there exist infinitely many positive integers n for which $S_{\ell^n}(x) \cap E \neq \emptyset$.

Theorem 37 (Hecke recurrence). *Let μ be any subsequential limit of a sequence of L^2-masses μ_φ of L^2-normalized automorphic forms $\varphi \in \pi \in A_0(X)$. Then μ is T_ℓ-recurrent.*[6]

The proof of Theorem 37 reduces via measure-theoretic considerations as in [Lindenstrauss 2006b; Brooks and Lindenstrauss 2014] to that of the following:

---

[6]It suffices to assume only that φ is a T_ℓ-eigenfunction.

Lemma 38. *There exists $c_0 > 0$ and $n_0 \geqslant 1$ so that for each split prime ℓ and $\varphi \in \pi \in A_0(X)$ and $x \in X$, one has $\sum_{k \leqslant n} \sum_{y \in S_{\ell^k}(x)} |\varphi(y)|^2 \geqslant c_0 n |\varphi(x)|^2$ for all natural numbers $n \geqslant n_0$.*

Proof. By a theorem of Eichler, Shimura and Igusa, π is tempered,[7] hence there exist $\alpha, \beta \in \mathbb{C}^{(1)}$ (the Satake parameters) so that $\lambda_\pi(\ell) = \alpha + \beta$; one then has more generally for $n \in \mathbb{Z}_{\geqslant 1}$ that

$$\lambda_\pi(\ell^n) = \frac{\alpha^{n+1} - \beta^{n+1}}{\alpha - \beta}. \tag{10}$$

By (9), one has $T_{\ell^n} = \sum_{k \leqslant n : k \equiv n(2)} z(\ell^{(k-n)/2}) S_{\ell^k}$. Conversely, $S_{\ell^k} = T_{\ell^k} - 1_{k \geqslant 2} z(\ell^{-1}) T_{\ell^{k-2}}$. Since π has a unitary central character, there is $\theta \in \mathbb{C}^{(1)}$ so that $z(\ell^{-1})\varphi = \theta\varphi$ for all $\varphi \in \pi$. Thus, denoting by $\ell^{k/2}\sigma_k \in \mathbb{C}$ the scalar by which S_{ℓ^k} acts on π, one obtains $\sigma_k = \lambda(\ell^k) - 1_{k \geqslant 2} \theta \ell^{-1} \lambda(\ell^{k-2})$, which expands for $k \geqslant 2$ to

$$\sigma_k = \frac{\gamma_1 \alpha^k - \gamma_2 \beta^k}{\alpha - \beta}, \tag{11}$$

with $\gamma_1 := \alpha - \theta\ell^{-1}\alpha^{-1}$, $\gamma_2 := \beta - \theta\ell^{-1}\beta^{-1}$. Note that $|\gamma_1|, |\gamma_2| \geqslant \frac{1}{2}$.

We turn to the main argument. For $m, k \in \mathbb{Z}_{\geqslant 0}$, Cauchy–Schwarz gives

$$\ell^m |\lambda_\pi(\ell^m)\varphi(x)|^2 = |T_{\ell^m}\varphi(x)|^2 \leqslant (1 + \ell^{-1})\ell^m \sum_{y \in T_{\ell^m}(x)} |\varphi(y)|^2,$$

$$\ell^k |\sigma_k \varphi(x)|^2 = |S_{\ell^k}\varphi(x)|^2 \leqslant (1 + \ell^{-1})\ell^k \sum_{y \in S_{\ell^k}(x)} |\varphi(y)|^2,$$

whence by (9) that $\sum_{k \leqslant n} \sum_{y \in S_{\ell^k}(x)} |\varphi(y)|^2 \gg |\varphi(x)|^2 c_{\pi,\ell}(n)$ with

$$c_{\pi,\ell}(n) := \sum_{k \leqslant n} |\sigma_k|^2 + \max_{m \leqslant n} |\lambda_\pi(\ell^m)|^2. \tag{12}$$

Our task thereby reduces to verifying that $c_{\pi,\ell}(n) \gg n$, uniformly in π and (unimportantly) ℓ. Suppose this estimate fails. Then there is a sequence of integers $j \to \infty$ and tuples $(\pi, n, \ell) = (\pi_j, n_j, \ell_j)$ as above, depending upon j, so that $n \to \infty$ as $j \to \infty$ and $c_{\pi,\ell}(n) = o(n)$. Here asymptotic notation refers to the $j \to \infty$ limit, and for quantities $A, B = A_j, B_j$ depending (implicitly) upon j, we write $A \ll B$ for $\limsup_{j \to \infty} |A_j/B_j| < \infty$ and $A \lll B$ or $A = o(B)$ for $\limsup_{j \to \infty} |A_j/B_j| = 0$; the notations $A \gg B$ and $A \ggg B$ are defined symmetrically. We shall derive from this supposition a contradiction. By passing to subsequences, we may consider separately cases in which the Satake parameters α, β of π, as defined above, satisfy:

(i) $|\alpha - \beta| \ggg 1/n$, or

(ii) $|\alpha - \beta| \ll 1/n$.[8]

---

[7] As in the references, the nontempered case may be treated more simply.

[8] The standard argument considers cases for which $|\alpha - \beta| \gg 1/n$ and $|\alpha - \beta| \lll 1/n$. We have found the present division slightly more efficient.

In case (i), we have $|1 - \alpha\bar{\beta}|^{-1} \lll n$, and so upon expanding the square and summing the geometric series,

$$c_{\pi,\ell}(n) \geqslant \sum_{k \leqslant n} |\sigma_k|^2 = \frac{|\gamma_1|^2 n + |\gamma_2|^2 n + o(n)}{|\alpha - \beta|^2} \geqslant \frac{n/3}{|\alpha - \beta|^2} \gg n.$$

In case (ii), one has $|\alpha - \beta|^{-1}/10 \gg n$, so the largest positive integer $m \leqslant n$ for which $m|\alpha - \beta| < \frac{1}{10}$ satisfies $m \gg n$, and (10) gives $c_{\pi,\ell}(n) \geqslant |\lambda_\pi(\ell^m)|^2 \gg m^2 \gg n^2 \geqslant n$. In either case, we derive the required contradiction. □

4. Positive entropy

In this section we verify the entropy hypothesis required by Theorem 28. The basic ideas here are due to Bourgain–Lindenstrauss [2003] following earlier work of Rudnick–Sarnak [1994] and Lindenstrauss [2001] and followed by later developments of Silberman–Venkatesh [≥ 2018] and Brooks–Lindenstrauss [2014]. Those works dealt with archimedean aspects; the present p-adic adaptation is obtained by replacing the role played by the discreteness of $\mathbb{Z}$ in $\mathbb{R}$ with that of $\mathbb{Z}[1/p]$ in $\mathbb{R} \times \mathbb{Q}_p$. We also give a new formulation of the basic line of attack (Lemma 41) emphasizing convolution over covering arguments (compare with [Silberman and Venkatesh ≥ 2018, Lemma 3.4]), which may be of use in other contexts.

Call $\varepsilon > 0$ *admissible* if it belongs to the image of $|.| : \mathbb{Q}_p^\times \to \mathbb{R}_+^\times$. For a compact open subgroup C of $\mathbb{Q}_p^\times$ and admissible $\varepsilon > 0$ set

$$B(U, \varepsilon) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G : a, d \in U, |b|, |c| \leqslant \varepsilon \right\}.$$

We refer to [Lindenstrauss 2006a, §8] for definitions and basic facts concerning (Kolmogorov–Sinai) entropy. As in [Lindenstrauss 2006a, §8; Bourgain and Lindenstrauss 2003; Silberman and Venkatesh ≥ 2018, Theorem 6.4], the following criterion suffices:

Theorem 39 (positive entropy on almost every ergodic component). *For each compact subset Ω of G, there exists U as above and $C, c > 0$ so that for all admissible $\varepsilon \in (0, 1)$, all L^2-normalized $\varphi \in \pi \in A_0(X)$, and all $x \in \Omega$, one has $\mu_\varphi(x B(U, \varepsilon)) \leqslant C\varepsilon^c$.*

Let us henceforth fix Ω as in Theorem 39. We then take for U any open subgroup of $\mathfrak{o}^\times$ with the property that for small enough ε, one has

$$x B(U, \varepsilon) x^{-1} \subseteq K \text{ for all } x \in \Omega, \tag{13}$$

$$g B(U, \varepsilon) g^{-1} \cap \Gamma = \{1\} \text{ for all } g \in G. \tag{14}$$

(Let us recall why it is possible to do this. Since K is open, we may find for each $x \in \Omega$ a pair (U, ε) so that (13) holds. Since Ω is compact, we may find one pair that works for every x. Similarly, since Γ is discrete in G, we may find for each $g \in G$ a pair (U, ε) so that (14) holds. The validity of (14) depends only upon the class of g in the quotient $\Gamma \backslash G$, which is compact, so we may again find one pair that works every g.)

We now state two independent lemmas, prove Theorem 39 assuming them, and then prove the lemmas.

Lemma 40 (bounds for Hecke returns). *For all small enough admissible* $\varepsilon \in (0, 1)$, *all* $n \in \mathbb{Z}_{\geqslant 1}$ *coprime to* p *and satisfying* $n < \sqrt{1/2}\varepsilon^{-1}$, *all* $m \in \mathbb{Q}^{\times}$ *with numerator and denominator coprime to* p, *and all* $x \in \Omega$, *the set* $S := M_n \cap z(m)xB(U, \varepsilon)x^{-1}$ *has cardinality* $\#S \leqslant 6 \prod_{p^k \| n}(k + 1)$. *In particular,* $\#S \leqslant 2^{13}$ *if* n *has at most* 10 *prime divisors counted with multiplicity.*

Lemma 41 (geometric amplification). *Let* $(c_\ell)_{\ell \in \mathbb{Z}_{\geqslant 1}}$ *be a finitely supported sequence of scalars. Set* $T := \sum_\ell c_\ell T_\ell / \sqrt{\ell}$ *and* $T^a := \sum_\ell |c_\ell| T_\ell^* / \sqrt{\ell}$. *Let* $\varphi \in \mathcal{A}(X)$, $\psi, \nu \in C_c^\infty(G)$. *Define* $\Psi \in \mathcal{A}(X)$ *by* $\Psi(g) := \sum_{\gamma \in \Gamma} |\psi|(\gamma g)$ *and* $\psi * \nu \in C_c^\infty(G)$ *by* $\psi * \nu(x) := \int_{y \in G} \psi(xy)\nu(y)$. *Then*

$$\|T\varphi(\psi * \nu)\|_{L^2(G)} \leqslant \|\varphi\|_{L^2(X)} \|T^a\Psi\|_{L^2(X)} \|\nu\|_{L^2(G)}.$$

Proof of Theorem 39. We have $T\varphi = \lambda\varphi$ with $\lambda := \sum c_\ell \lambda_\pi(\ell)$, where $T_\ell\varphi = \sqrt{\ell}\lambda_\pi(\ell)\varphi$. Abbreviate $J := B(U, \varepsilon)$; it is a group. Let $x \in \Omega$. Take $\psi := 1_{xB(U,\varepsilon)} \geqslant 0$ and $\nu := e_J := \text{vol}(J)^{-1}1_J$. Then $1_{xB(U,\varepsilon)} = |\psi * \nu|^2$. By (14), we have $\mu_{T\varphi}(|\psi * \nu|^2) = \|T\varphi(\psi * \nu)\|_{L^2(G)}^2$, and so by Lemma 41, $\mu_\varphi(xB(U, \varepsilon)) \leqslant |\lambda|^{-1}\|T^a\Psi\|_{L^2(X)}\|\nu\|_{L^2(G)}$. The square $\|T^a\Psi\|_{L^2(X)}^2$ is a linear combination of terms $\langle T_\ell^*\Psi, T_{\ell'}^*\Psi \rangle = \langle T_{\ell'}T_\ell^*\Psi, \Psi \rangle$ to which we apply the Hecke multiplicativity (8) and the unfolding: for $m, n \in \mathbb{Z}_{\geqslant 1}$,

$$\langle z(m)T_n^*\Psi, \Psi \rangle \|\nu\|_{L^2(G)}^2 = \int_{g \in G} \sum_{s \in M_n} \psi(z(m)sg)\psi(g)\,\text{vol}(J)^{-1} = \#M_n \cap z(m^{-1})xJx^{-1}. \quad (15)$$

By Lemma 40, we thereby obtain

$$\mu_\varphi(xB(U, \varepsilon))^2 \leqslant 2^{13}|\lambda|^{-2}\sum_{\ell,\ell'}|c_\ell c_{\ell'}| \sum_{d \,|\, (\ell,\ell')} d/\sqrt{\ell\ell'}$$

provided that c_ℓ is supported on integers $\ell \leqslant 2^{-1/4}\varepsilon^{-1/2}$ having at most 5 prime factors counted with multiplicity. A standard choice of c_ℓ completes the proof. For completeness, we record a variant of the choice from [Venkatesh 2010, §4.1]: Set $L := (1/\varepsilon)^{0.1}$. Denote by $\mathcal{L}$ the set consisting of all $\ell = q$ or $\ell = q^2$ taken over primes $q \in [L, 2L]$; each such q splits B provided ε is small enough. Set $c_\ell := 0$ unless $\ell \in \mathcal{L}$, in which case $c_\ell := L^{-1}\log(L)\,\text{sgn}(\lambda_\pi(\ell))^{-1}$. We have $\sum_\ell |c_\ell| \asymp 1$ and $|c_\ell| \leqslant L^{-1}\log(L)$, while Iwaniec's trick $|\lambda_\pi(q)|^2 + |\lambda_\pi(q^2)| \geqslant 1$, a consequence of (8), implies $\lambda \asymp 1$. With trivial estimation we obtain $\mu_\varphi(xB(U, \varepsilon)) \ll L^{-1/2}(\log L)^{O(1)} \ll \varepsilon^{0.01}$, as required. $\square$

Proof of Lemma 40. Observe first, thanks to (13) and $n\mathbb{Z}[1/p]^\times \cap (\mathbb{Q}_+ \times \mathbb{Z}_p) = \{n\}$ and $z(m) \in K$, that $S \subseteq M_n \cap K = R(n) := \{\alpha \in R : \text{nr}(\alpha) = n\}$. Given $s, t \in S$, their commutator $u := sts^{-1}t^{-1}$ thus satisfies $\text{nr}(u) = 1$ and $n^2u = st s^\iota t^\iota \in R$, hence $\text{tr}(u) \in n^{-2}\mathbb{Z}$. Since S is conjugate to a subset of the preimage in $M_2(\mathfrak{o})$ of the upper-triangular Borel in $M_2(\mathfrak{o}/\mathfrak{q})$ with $\mathfrak{q} := \{x \in \mathfrak{o} : |x| \leqslant \varepsilon^2\}$, and the commutator of that preimage is contained in the preimage of the unipotent, one has $|\text{tr}(u) - 2|_p \leqslant \varepsilon^2$. Since B is definite, $|\text{tr}(u)|_\infty \leqslant 2|\text{nr}(u)|_\infty^{1/2} = 2$. The integer $a := n^2\text{tr}(u) - 2n^2$ thus satisfies $|a|_\infty|a|_p \leqslant 2n^2\varepsilon^2 < 1$ and so must be zero, i.e., $\text{tr}(u) = 2$; since B is nonsplit, $u = 1$. In summary, any two elements of S commute.

Since B is nonsplit and definite, S is contained in the set $\mathcal{O}(n)$ of norm n elements in some imaginary quadratic order $\mathcal{O} \subset R$. Thus $\#S \leqslant \#\mathcal{O}(n) \leqslant \#\mathcal{O}^\times \cdot \#\{I \subseteq \mathcal{O} : \mathrm{nr}(I) = n\} \leqslant 6 \prod_{p^k \| n}(k+1)$.   □

Proof of Lemma 41. Write $M := R[1/p]$. We may express the operator T by the formula $T\varphi(x) = \sum_{s \in M/\Gamma} h_s \varphi(s^{-1}x)$ for some finitely supported coefficients h_s; then $T^a\Psi(x) = \sum_{s \in M/\Gamma} |h_s|\Psi(sx)$. Abbreviate $I := \|T\varphi(\psi * \nu)\|_{L^2(G)}$. By the triangle inequality and a change of variables $x \mapsto sx$, we have

$$I \leqslant \sum_{s \in M/\Gamma} |h_s| \left(\int_{x \in G} |\varphi|^2(x) |\psi * \nu(sx)|^2 \right)^{1/2}.$$

By a change of variables, $\psi * \nu(sx) = \int_{y \in G} \psi(sy)\nu_y^*(x)$ with $\nu_y^*(x) := \nu(x^{-1}y)$. By the triangle inequality, $I \leqslant \int_{y \in G} \sum_{s \in M/\Gamma} |h_s| |\psi(sy)| \|\varphi\nu_y^*\|_{L^2(G)}$. We unfold $\int_{y \in G} \sum_{s \in M/\Gamma} = \int_{y \in X} \sum_{s \in \Gamma \backslash M} \sum_{\gamma \in \Gamma}$, giving $I \leqslant \int_{y \in X} T^a\Psi(y) \|\varphi\nu_y^*\|_{L^2(G)}$. We conclude via Cauchy–Schwartz and the identity $\int_{y \in X} \|\varphi\nu_y^*\|_{L^2(G)}^2 = \|\nu\|_{L^2(G)}^2 \|\varphi\|_{L^2(X)}^2$.   □

5. Representation-theoretic preliminaries

5.1. *Generalities.* Let k be a nonarchimedean local field with maximal order $\mathfrak{o}$, maximal ideal $\mathfrak{p}$, normalized valuation $v : k \to \mathbb{Z} \cup \{+\infty\}$, and $q := \#\mathfrak{o}/\mathfrak{p}$. Fix Haar measures $dx, d^\times y$ on $k, k^\times$ assigning volume one to maximal compact subgroups. Fix a nontrivial unramified additive character $\psi : k \to \mathbb{C}^{(1)}$. Set $G := \mathrm{GL}_2(k)$.

5.2. *Some notation and terminology.* For $x \in k$ and $y_1, y_2 \in k^\times$, set

$$n(x) := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \qquad n'(x) := \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix},$$

$$\mathrm{diag}(y_1, y_2) := \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}, \qquad w := \begin{pmatrix} & -1 \\ 1 & \end{pmatrix},$$

and $a(y) := \mathrm{diag}(y, 1), z(y) := \mathrm{diag}(y, y)$. Say that a vector v in some $\mathrm{GL}_2(k)$-module π is *supported on* $m..m'$, for integers m, m' with $m \leqslant m'$, if v is invariant by $n(\mathfrak{p}^{-m})$ and $n'(\mathfrak{p}^{m'})$, and that v has *orientation* (ω_1, ω_2), for characters ω_1, ω_2 of $\mathfrak{o}^\times$, if $\pi(\mathrm{diag}(y_1, y_2))v = \omega_1(y_1)\omega_2(y_2)v$ for all $y_1, y_2 \in \mathfrak{o}^\times$.

5.3. *Principal series representations.* For characters $\chi_1, \chi_2 : k^\times \to \mathbb{C}^\times$, denote by $\pi = \chi_1 \boxplus \chi_2$ the *principal series representation* of G realized in its *induced model* as a space of smooth functions $v : G \to \mathbb{C}$ satisfying $v(n(x) \mathrm{diag}(y_1, y_2)g) = |y_1/y_2|^{1/2}\chi_1(y_1)\chi_2(y_2)v(g)$ for all $x \in k$ and $y_1, y_2 \in k^\times$ and $g \in G$. A sufficient condition for π to be irreducible is that $c(\chi_1/\chi_2) \neq 0$ (see, e.g., [Schmidt 2002]). If χ_1, χ_2 are unitary, then π is unitary; an invariant norm is given by $\|v\|^2 := \int_{x \in k} |v(n'(x))|^2 \, dx$ (see, e.g., [Knapp 1986, (7.1)]). The log-conductor is $c(\pi) = c(\chi_1) + c(\chi_2)$ and the central character is $\chi_\pi = \chi_1\chi_2$ (see, e.g., [Schmidt 2002]).

The following "line model" parametrization of π shall be convenient: for suitable $f \in C^\infty(k)$, define $v_f \in \pi$ by

$$v_f(g) := f(c/d)|\det(g)/d^2|^{1/2}\chi_1(\det(g)/d)\chi_2(d), \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{16}$$

If χ_1, χ_2 are unitary, then $\|v_f\|^2 = \int_k |f|^2$.

5.4. *Generic representations.* Recall that an irreducible representation σ of G is *generic* if it is isomorphic to an irreducible subspace $\mathcal{W}(\sigma, \psi)$ of the space of smooth functions $W : G \to \mathbb{C}$ satisfying $W(n(x)g) = \psi(x)W(g)$ for all $x, g \in k, G$; in that case, $\mathcal{W}(\sigma, \psi)$ is called the *Whittaker model* of σ. It is known that every nongeneric irreducible representation of G is one-dimensional (see, e.g., [Schmidt 2002]).

For each $W \in \mathcal{W}(\sigma, \psi)$, denote also by W the function $W : k^\times \to \mathbb{C}$ defined by $W(y) := W(a(y))$. The space $\mathcal{K}(\sigma, \psi)$ of functions $W : k^\times \to \mathbb{C}$ arising in this way from some $W \in \mathcal{W}(\sigma, \psi)$ is called the *Kirillov model* of σ. It is known that the natural map $\mathcal{W}(\sigma, \psi) \to \mathcal{K}(\sigma, \psi)$ is an isomorphism and that $\mathcal{K}(\sigma, \psi) \supseteq C_c^\infty(k^\times)$ (see, e.g., [Schmidt 2002]).

An irreducible principal series representation $\pi = \chi_1 \boxplus \chi_2$ is generic (see, e.g., [Schmidt 2002]); the standard intertwining map from π to its ψ-Whittaker model $\mathcal{W}(\pi, \psi)$, denoted $\pi \ni v \mapsto W_v : \mathrm{GL}_2(k) \to \mathbb{C}$, is given by $W_v(g) := \int_{x \in k} v(wn(x)g)\psi(-x)\,dx$. In general, this integral fails to converge absolutely and must instead be interpreted via analytic continuation, regularization, or as a limit of integrals taken over the compact subgroups $\mathfrak{p}^{-n}$ of k as $n \to \infty$ (see, e.g., [Bump 1997, p. 485]); for the sake of presentation, we ignore such technicalities in what follows.

5.5. *Newvector theory.* Recall Definition 20. Recall also from Section 1.3 that we have fixed decompositions $N = N_1 + N_2$ of every nonnegative integer N, with $N_1, N_2 \to \infty$ as $N \to \infty$.

Theorem 42 (basic newvector theory). *Let π be a generic irreducible representation of $\mathrm{GL}_2(k)$ and let $m \leqslant m'$ be integers. Then the space of vectors in π supported on $m..m'$ and with orientation $(1, \chi_\pi|_{\mathfrak{o}^\times})$ has dimension $\max(0, 1 + |m - m'| - c(\pi))$.*

In particular, let π be any irreducible representation of $\mathrm{GL}_2(k)$ with ramified central character χ_π. Denote by V the space of vectors in π supported on $-N_1..N_2$ with orientation $(1, \chi_\pi|_{\mathfrak{o}^\times})$. Then $V = 0$ unless π is generic, in which case $\dim V = \max(0, 1 + N - c(\pi))$.

Proof. For the first assertion, see [Casselman 1973a]. The generic case of the second assertion follows from the first assertion, so suppose π is one-dimensional. Write $\pi = \chi \circ \det$ for some $\chi : k^\times \to \mathbb{C}^\times$. Since χ_π is ramified, the characters $(1, \chi_\pi|_{\mathfrak{o}^\times})$ and $(\chi|_{\mathfrak{o}^\times}, \chi|_{\mathfrak{o}^\times})$ of $\mathfrak{o}^\times \times \mathfrak{o}^\times$ are distinct, and so $V = 0$. $\square$

Lemma 43. *Let π be an irreducible generic representation of $\mathrm{GL}_2(k)$ with ramified central character χ_π. Then $c(\chi_\pi) \leqslant c(\pi)$ with equality precisely when π is isomorphic to an irreducible principal series representation $\chi_1 \boxplus \chi_2$ for which at least one of the inducing characters χ_1, χ_2 is unramified.*

Proof. This is well-known; see [Templier 2014, Lemma 3.1; Casselman 1973b, Proof of Proposition 2]. $\square$

Lemma 44. *Let $\pi = \chi_1 \boxplus \chi_2$ be an irreducible principal series representation of G. Let $v \in \pi$ be a newvector of some support $m..m'$:*

(1) *If χ_1 is ramified and χ_2 is ramified, then $v = v_f$ as in (16) for f a character multiple of the characteristic function of an $\mathfrak{o}^\times$-coset, thus $f = c\chi 1_{\varpi^n \mathfrak{o}^\times}$ for some $c \in \mathbb{C}$, $\chi : k^\times \to \mathbb{C}^\times$ and $n \in \mathbb{Z}$.*

(2) *If χ_1 is unramified and χ_2 is ramified, then $v = v_f$ for $f = c1_{\mathfrak{a}}$ for some scalar c and fractional $\mathfrak{o}$-ideal $\mathfrak{a} \subset k$.*

Proof. Both assertions are well-known in the special case $m = 0$ (see [Schmidt 2002]) and follow inductively in general using that $a(\varpi)$ bijectively maps newvectors of support $m..m'$ to those of support $m - 1..m' - 1$. □

5.6. *Local Rankin–Selberg integrals.* Let π be an irreducible unitary principal series representation of $G := \mathrm{GL}_2(k)$ and σ an irreducible generic unitary representation of $\mathrm{PGL}_2(k)$. We have the following special case of a theorem of D. Prasad:

Theorem 45 [Prasad 1990]. *The space $\mathrm{Hom}_G(\sigma \otimes \bar{\pi} \otimes \pi, \mathbb{C})$, consisting of trilinear functionals $\ell : \sigma \otimes \bar{\pi} \otimes \pi \to \mathbb{C}$ satisfying the diagonal invariance $\ell(\sigma(g)v_1, \bar{\pi}(g)v_2, \pi(g)v_3) = \ell(v_1, v_2, v_3)$ for all $g \in G$ and all vectors, is one-dimensional.*

We may fix a nonzero element $\ell_{\mathrm{RS}} \in \mathrm{Hom}_G(\sigma \otimes \bar{\pi} \otimes \pi, \mathbb{C})$ as follows: Denote by Z the center of G and $U := \{n(x) : x \in k\}$. Equip the right G-space $ZU \backslash G$ with the Haar measure for which

$$\int_{g \in ZU \backslash G} \phi(g) = \int_{y \in k^\times} \int_{x \in k} \phi(a(y)n'(x)) \frac{d^\times y}{|y|} \, dx \tag{17}$$

for $\phi \in C_c(ZU \backslash G)$ (see [Michel and Venkatesh 2010, §3.1.5]). Realize π in its induced model. For $W_1 \in \mathcal{W}(\sigma, \psi)$, $W_2 \in \mathcal{W}(\pi, \psi)$ and $v_3 \in \pi$, set $\ell_{\mathrm{RS}}(W_1, \overline{W}_2, v_3) := \int_{ZU \backslash G} W_1 \overline{W}_2 v_3$ (see [Michel and Venkatesh 2010, §3.4.1]). The definition applies in particular when W_2 is the image W_v of some $v \in \pi$ under the intertwiner from Section 5.4.

The trick encapsulated by the following lemma (a careful application of "nonarchimedean integration by parts") shall be exploited repeatedly in Section 6.3:

Lemma 46 (application of diagonal invariance). *Let $f \in C_c^\infty(k)$. Let U_1 be an open subgroup of $\mathfrak{o}^\times$ for which $\bar{f} \otimes f$ is U_1-invariant in the sense that $\bar{f}(ux)f(uy) = \bar{f}(x)f(y)$ for all $u, x, y \in U_1, k, k$. Let $W_1 \in \mathcal{W}(\sigma, \psi)$. Then*

$$\ell_{\mathrm{RS}}(W_1, \overline{W}_{v_f}, v_f) = \int_{x \in k, \, y \in k^\times, \, t \in k} f(x) \bar{f}\left(x + \frac{y}{t}\right) F(x, y, t; W_1, U_1) \frac{dt}{|t|} \, dx \, d^\times y,$$

where $F(x, y, t; W_1, U_1) := \mathbb{E}_{u \in U_1} W_1(a(y)n'(x/u)) \chi_1 \chi_2^{-1}(ut) \psi(ut)$ with $\mathbb{E}_{u \in U_1}$ denoting an integral with respect to the probability Haar.

Proof. Set $g := a(y)n'(x) = \left(\begin{smallmatrix} y & \\ x & 1 \end{smallmatrix}\right)$. For $t \in k$ one has $wn(t)g = \left(\begin{smallmatrix} -x & -1 \\ y+tx & t \end{smallmatrix}\right)$, hence

$$v_f(g) = f(x)|y|^{1/2}\chi_1(y),$$

$$\bar{v}_f(wn(t)g) = \bar{f}((y+tx)/t)|y/t^2|^{1/2}\chi_1^{-1}(y/t)\chi_2^{-1}(t),$$

$$v_f(g)\overline{W}_{v_f}(g) = \int_{t \in k} v_f(g)\overline{v_f(wn(t)g)\psi(-t)}\, dt$$

$$= |y|f(x)\int_{t \in k}\bar{f}\left(x+\frac{y}{t}\right)\chi_1\chi_2^{-1}(t)\psi(t)\,\frac{dt}{|t|}.$$

Integrating against $W_1(a(y)n'(x))|y|^{-1}\, dx\, d^\times y$ gives that $\ell_{RS}(W_1, \overline{W}_{v_f}, v_f)$ equals

$$\int_{x \in k,\, y \in k^\times,\, t \in k} f(x)\bar{f}\left(x+\frac{y}{t}\right)W_1(a(y)n'(x))\chi_1\chi_2^{-1}(t)\psi(t)\,\frac{dt}{|t|}\, dx\, d^\times y.$$

To obtain the claimed formula, we apply for $u \in U_1$ the substitutions $t \mapsto ut$, $x \mapsto x/u$, invoke the assumed U_1-invariance of $\bar{f} \otimes f$, and average over u. $\qquad\square$

5.7. *Gauss sums.* We shall repeatedly use the following without explicit mention:

Lemma 47. *Let $U_1 \leqslant \mathfrak{o}^\times$ be an open subgroup and ω a character of $\mathfrak{o}^\times$. For $t \in k^\times$, set $H(t) := H(t, \omega, U_1) := \mathbb{E}_{u \in U_1}\omega(ut)\psi(ut)$, where $\mathbb{E}$ denotes integration with respect to the probability Haar.*

(1) *For fixed U_1, one has $H(t) = 0$ unless $-v(t) = c(\omega) + O(1)$, in which case $H(t) \ll C(\omega)^{-1/2}$, with implied constants depending at most upon U_1.*

(2) *Suppose $U_1 = \mathfrak{o}^\times$ and $c(\omega) > 0$. Then $H(t) = 0$ unless $-v(t) = c(\omega)$, in which case $H(t)$ is independent of t and has magnitude $|H(t)| = cC(\omega)^{-1/2}$ for some $c > 0$ depending only upon k.*

Proof. For $U_1 = \mathfrak{o}^\times$, these are standard assertions concerning Gauss sums. The standard proof adapts to the general case (compare with [Michel and Venkatesh 2010, 3.1.14]). $\qquad\square$

6. Local study of nonarchimedean microlocal lifts

Recall Definition 21 and the statement of Lemma 22. Retain the notation of Section 5.

6.1. *Proof of Lemma 22: determination of microlocal lifts.* For any character $\chi : k^\times \to \mathbb{C}^\times$, the nonequivariant twisting isomorphism $\pi \to \pi' := \pi \otimes \chi$ induces nonequivariant linear isomorphisms

$$V := \{\text{microlocal lifts in } \pi \text{ of orientation } (\omega_1, \omega_2)\} \tag{18}$$

$$\cong \{\text{microlocal lifts in } \pi' \text{ of orientation } (\omega_1', \omega_2')\},$$

with $\omega_i' := \omega_i \cdot \chi|_{\mathfrak{o}^\times}$. We thereby reduce to verifying the conclusion in the special case $\omega_1 = 1$. Suppose $V \neq 0$. Write $\omega := \omega_2$. By the convention $N \geqslant 1$ of Definition 21, ω is ramified. The central character χ_π of π restricts to ω, hence is ramified; by Theorem 42, $\dim V = \max(0, 1 + c(\pi) - c(\chi_\pi))$, and so $V \neq 0$ only if $c(\pi) \geqslant c(\chi_\pi)$. By Lemma 43, the latter happens only if $c(\pi) = c(\chi_\pi)$ and π has the indicated form, in which case $\dim V = 1$. The explicit description of V now follows in general from (18).

6.2. *Explicit formulas.* Let $\pi := \chi_1 \boxplus \chi_2$ and $\omega_i := \chi_i|_{\mathfrak{o}^\times}$ with $N := c(\omega_1/\omega_2) \geqslant 1$.

Lemma 48. *Define $f_1, f_2 \in C^\infty(k)$ (as if in the "line model" of Section 5.3) by*

$$f_1(x) := 1_{\mathfrak{p}^{N_2}}(x), \quad f_2(x) := 1_{\mathfrak{p}^{N_1}}(1/x)|1/x|\chi_1^{-1}\chi_2(x)$$

and $v_1, v_2 \in \pi$ in the induced model on $g = \begin{pmatrix} * & * \\ c & d \end{pmatrix} \in \mathrm{GL}_2(k)$ by

$$v_1(g) := v_{f_1}(g) = 1_{\mathfrak{p}^{N_2}}(c/d)\left|\frac{\det g}{d^2}\right|^{1/2}\chi_1(\det(g)/d)\chi_2(d), \tag{19}$$

$$v_2(g) := v_{f_2}(g) = 1_{\mathfrak{p}^{N_1}}(d/c)\left|\frac{\det g}{c^2}\right|^{1/2}\chi_1(\det(g)/c)\chi_2(c), \tag{20}$$

and $W_1, W_2 \in \pi$ in the Kirillov model $\mathcal{K}(\pi, \psi)$ by[9]

$$W_1(y) := 1_{\mathfrak{p}^{-N_1}}(y)|y|^{1/2}\chi_1(y), \quad W_2(y) := 1_{\mathfrak{p}^{-N_1}}(y)|y|^{1/2}\chi_2(y). \tag{21}$$

Then v_1, W_1 and v_2, W_2 are microlocal lifts of orientations (ω_1, ω_2) and (ω_2, ω_1), respectively.

Proof. The formulas for W_1, v_1 in the case $\chi_1 = 1$ and those for W_2, v_2 in the case $\chi_2 = 1$ follow from known formulas for standard newvectors [Schmidt 2002]; the general case follows from the twisting isomorphisms (18). $\qquad\square$

6.3. *Stationary phase analysis of local Rankin–Selberg integrals.* In this section we apply stationary phase analysis to evaluate and estimate some local Rankin–Selberg integrals involving microlocal lifts and newvectors. We use these in Section 7 to prove Theorem 25 and Theorem 29. Retain the notation of Section 5.1. Let χ_1, χ_2 be unitary characters of $k^\times$ for which $N := c(\chi_1/\chi_2)$ is positive. Let $\pi = \chi_1 \boxplus \chi_2$ be the corresponding generic irreducible unitary principal series representation of $\mathrm{GL}_2(k)$, realized in its induced model and equipped with the norm given in Section 5.3. Equip the complex-conjugate representation $\bar{\pi}$ with the compatible unitary structure. Define the intertwiner $\pi \ni v \mapsto W_v \in \mathcal{W}(\pi, \psi)$ as in Section 5.4. Let σ be a generic irreducible unitary representation of $\mathrm{PGL}_2(k)$, realized in its ψ-Whittaker model $\sigma = \mathcal{W}(\sigma, \psi)$.

Theorem 49. *Let $v \in \pi$ be a microlocal lift of orientation $(\chi_1|_{\mathfrak{o}^\times}, \chi_2|_{\mathfrak{o}^\times})$, let $v' \in \pi$ be a generalized newvector, and let $W_1 \in \sigma$.*

(I) *If N is large enough in terms of W_1, then*

$$\ell_{\mathrm{RS}}(W_1, \overline{W}_v, v) = cq^{-N/2}\|v\|^2 \int_{y \in k^\times} W_1(y)\, d^\times y,$$

where[10] $c := q^{N/2}\int_{t \in k^\times}\chi_1\chi_2^{-1}(t)\psi(t)\, dt/|t| \asymp 1$ is a complex scalar which is independent of W_1 and whose magnitude depends only upon k.

---

[9]Recall that ψ is assumed unramified.

[10] The integral defining c should be interpreted in the usual way as (for instance) a limit of integrals over increasing finite unions of $\mathfrak{o}^\times$-cosets.

(II) *One has $\ell_{\text{RS}}(W_1 \otimes \overline{W}_{v'} \otimes v') \ll q^{-N/2}\|v'\|^2$ with the implied constant depending at most upon W_1.*

(III) *Suppose that $\nu(2) = 0$, χ_π is unramified, $\|v'\| = \|v\|$, the support of v' is $-N..N$, σ is unramified, and $W_1 \in \sigma$ is spherical, so that $N = c(\chi_1) = c(\chi_2)$ and $c(\pi) = 2N$. Then $\ell_{\text{RS}}(W_1, \overline{W}_v, v) = \ell_{\text{RS}}(W_1, \overline{W}_{v'}, v')$.*

The most difficult assertion is (II), which is used only to deduce the equidistribution of newvectors (Theorem 17). Assertion (III) serves only the purpose of illustration (see the discussion after Theorem 25). The other main results of this article (Theorems 29, 27) require only (I), whose proof is very short.

Proof of (I). Without loss of generality, let $v = v_f$ with $f(x) := 1_{\mathfrak{p}^{N_2}}(x)$. Because N_2 is large in enough in terms of W_1, we have whenever $f(x) \neq 0$ that $W_1(a(y)n'(x/u)) = W_1(y)$ for all $u \in \mathfrak{o}^\times$. Lemma 46 gives after the simplifications $f(x)\bar{f}(x + y/t) = 1_{\mathfrak{p}^{N_2}}(x)1_{\mathfrak{p}^{N_2}}(y/t)$ and $1_{\mathfrak{p}^{N_2}}(x)F(x, y, t; W_1, \mathfrak{o}^\times) = 1_{\mathfrak{p}^{N_2}}(x)W_1(y)H(t)$ with $H(t) := \mathbb{E}_{u \in \mathfrak{o}^\times} \chi_1\chi_2^{-1}(ut)\psi(ut)$ that

$$\ell_{\text{RS}}(W_1, \overline{W}_v, v) = \int_{y \in k^\times} W_1(y) \int_{x \in k} 1_{\mathfrak{p}^{N_2}}(x) \int_{t \in k} 1_{\mathfrak{p}^{N_2}}(y/t)H(t) \frac{dt}{|t|} \, dx \, d^\times y.$$

We have $W_1(y)H(t) = 0$ unless $|t| \asymp q^N$ and $|y| \ll 1$; because N_1 is large enough in terms of W_1, the factor $1_{\mathfrak{p}^{N_2}}(y/t) = 1$ is thus redundant. Since $\int_{x \in k} 1_{\mathfrak{p}^{N_2}}(x) \, dx = \int_k |f|^2 = \|v\|^2$, we obtain the required identity. $\square$

Proof of (II). Suppose first that χ_1 and χ_2 are both ramified. In that case, Lemma 44 says that $v' = v_f$ with f a character multiple of the characteristic function of some $\mathfrak{o}^\times$ coset. In particular,

$$f \text{ is supported on a coset of } \mathfrak{o}^\times, \text{ and } \bar{f} \otimes f \text{ is } \mathfrak{o}^\times\text{-invariant.} \tag{22}$$

From the mod-center identity $a(y)n'(x) \equiv n(y/x)a(y/x^2)wn(1/x)$, we have

$$W_1(a(y)n'(x)) = \psi(y/x)W_1(a(y/x^2)wn(1/x)). \tag{23}$$

From (23) and standard bounds on Whittaker functions, we have[11]

$$\sup_{x \in k} \int_{y \in k^\times} |W_1(a(y)n'(x))| \, d^\times y \ll 1. \tag{24}$$

By (23), there exists a fixed open subgroup $U_1 \leqslant \mathfrak{o}^\times$ for which

$$W_1(a(y)n'(x/u)) = W_1(a(y)n'(x)) \times \begin{cases} 1 & \text{for } |x| \leqslant 1, \\ \psi((u-1)y/x) & \text{for } |x| \geqslant 1. \end{cases} \tag{25}$$

Without loss of generality, suppose $\int_k |f|^2 = 1$. We apply Lemma 46, split the integral according as $|x| \leqslant 1$ or not, and appeal to (24) and (25); our task thereby reduces to showing with

$$H_1(t) := \mathbb{E}_{u \in U_1} \chi_1\chi_2^{-1}(ut)\psi(ut),$$

$$H_2(t, y/x) := \psi(-y/x)\mathbb{E}_{u \in U_1} \chi_1\chi_2^{-1}(ut)\psi(u(t + y/x))$$

---

[11] See [Michel and Venkatesh 2010, 3.2.3], and recall that σ is assumed generic and unitary.

that the quantities

$$I_1 := \sup_{y \in k^\times} \int_{t \in k^\times} \int_{x \in k : |x| \leqslant 1} |f(x)\bar{f}(x + y/t) H_1(t)| \, d^\times t \, dx \, d^\times y,$$

$$I_2 := \sup_{y \in k^\times} \int_{t \in k^\times} \int_{x \in k : |x| > 1} |f(x)\bar{f}(x + y/t) H_2(t, y/x)| \, d^\times t \, dx \, d^\times y$$

are $O(q^{-N/2})$. We have $H_1(t) = 0$ unless $|t| \asymp q^N$, in which case $H_1(t) \ll q^{-N/2}$; the set of such t has $d^\times t$-volume $O(1)$, so an adequate estimate for I_1 follows from Cauchy–Schwartz applied to the x-integral. Similarly, $H_2(t, y/x) = 0$ unless $|t + y/x| \asymp q^N$, in which case $H_2(t, y/x) \ll q^{-N/2}$; the support condition on f shows that $f(x)\bar{f}(x + y/t) = 0$ unless $|t + y/x| = |t|$, we may conclude once again by Cauchy–Schwartz.[12]

We turn to the case that one of χ_1, χ_2 is unramified. By the assumption $c(\chi_1/\chi_2) \neq 0$, the other one is ramified. By symmetry, we may suppose that χ_1 is unramified and χ_2 is ramified. By Lemma 44, we may suppose without loss of generality that $v' = v_f$ for $f = 1_\mathfrak{a}$ with $\mathfrak{a} \subset k$ a fractional $\mathfrak{o}$-ideal. Then $\bar{f} \otimes f$ is $\mathfrak{o}^\times$-invariant. We split the integral over $x \in k$ as above, and the same argument works for the range $|x| \leqslant 1$. The remaining range contributes

$$I_3 := \int_{x \in k : |x| > 1} \int_{y \in k^\times} \int_{t \in k^\times} 1_\mathfrak{a}(x) 1_\mathfrak{a}(x + y/t) H_3(t, y/x; x) \, dx \, d^\times y \, d^\times t, \qquad (26)$$

where

$$H_3(t, y/x; x) := W_1(a(y/x^2)wn(1/x)) \mathbb{E}_{u \in U_1} \chi_1 \chi_2^{-1}(ut) \psi(u(t + y/x)).$$

A bit more care is required than in the above argument, which gives now an upper bound of $+\infty$; the problem is that the nonvanishing of $H_3(t, y/x; x)$ no longer restricts t to a volume $O(1)$ subset of $k^\times$. We do better here by exploiting additional cancellation coming from the y-integral: Let C_1, C_2 be positive scalars, depending only upon W_1, U_1, so that

$$H_3(t, y/x; x) \neq 0 \implies C_1 q^N < |y/x + t| < C_2 q^N. \qquad (27)$$

If $|y/x| \geqslant C_2 q^N$, then $H_3(t, y/x; x) \neq 0$ only if $|t| = |y/x|$. If $|y/x| \leqslant C_1 q^N$, then $H_3(t, y/x; x) \neq 0$ only if $C_1 q^N < |t| < C_2 q^N$. Arguing as above, we reduce to considering the range $C_1 q^N < |y/x| < C_2 q^N$, in which $H_3(t, y/x; x) \neq 0$ only if $|t| < C_2 q^N$. The range $C_1 q^N \leqslant |t| < C_2 q^N$ may be treated as before, so we reduce to showing that

$$I_4 := \int_{\substack{x,y,t \in k, k^\times, k^\times : |x| > 1, \\ C_1 q^N < |y/x| < C_2 q^N, \, |t| < C_1 q^N}} 1_\mathfrak{a}(x) 1_\mathfrak{a}(x + y/t) H_3(t, y/x; x) \, dx \, d^\times y \, d^\times t = 0. \qquad (28)$$

---

[12] The estimate just derived is essentially sharp when f is supported in a fixed open subset of $k^\times$, but can be substantially sharpened when f is "unbalanced" in the sense that its support tends sufficiently rapidly with N either to zero or infinity. The possibility of such sharpening is the simplest case of the "weak subconvexity" phenomenon identified in [Nelson et al. 2014].

Note that the conditions defining the integrand imply that $|xt/y| < 1$. There is an open subgroup U_2 of $\mathfrak{o}^\times$, depending only upon W_1, U_1, so that

$$z \in U_2, |xt/y| < 1 \implies \frac{t+zy/x}{t+y/x} \in U_1, \tag{29}$$

$$|x| > 1, z \in U_2 \implies W_1(a(zy/x^2)wn(1/x)) = W_1(a(y/x^2)wn(1/x)), \tag{30}$$

$$|xt/y| < 1, z \in U_2 \implies 1_\mathfrak{a}(x + zy/t) = 1_\mathfrak{a}(x + y/t). \tag{31}$$

For N large enough in terms of W_1, U_1 and hence U_2, we have

$$|xt/y| < 1 \implies \mathbb{E}_{z \in U_2} \chi_1^{-1} \chi_2(t + zy/x) = 0. \tag{32}$$

In I_4, we substitute $y \mapsto yz$ with $z \in U_2$ and average over z; by (31) and (30), our task reduces to establishing for $|xt/y| < 1$ that

$$\mathbb{E}_{z \in U_2} \mathbb{E}_{u \in U_1} \chi_1 \chi_2^{-1}(ut) \psi(u(t + zy/x)) = 0,$$

which follows from (32) after the change of variables $u \mapsto u(t + y/x)/(t + zy/x)$ suggested by (29). $\square$

Proof of (III). By (I), our task reduces to showing that

$$\ell_{RS}(W_1, \overline{W}_{v'}, v') = cq^{-N/2} \|v'\|^2 \int_{y \in k^\times} W_1(y) \, d^\times y$$

with the same scalar c as in (I). Suppose without loss of generality that $v' = v_f$ with $f := \chi_2 1_{\mathfrak{o}^\times}$. Note that $\bar{f} \otimes f$ is $\mathfrak{o}^\times$-invariant. If $f(x) \neq 0$, then $W_1(a(y)n'(x/u)) = W_1(y)$ for all $u \in \mathfrak{o}^\times$. Lemma 46 gives after the simplification $f(x)F(x, y, t; W_1, \mathfrak{o}^\times) = f(x)W_1(y)H(t)$ with H as in the proof of (I) that

$$\ell_{RS}(W_1, \overline{W}_{v'}, v') = \int_{y \in k^\times} \int_{x \in k} \int_{t \in k} W_1(y) f(x) \bar{f}(x + y/t) H(t) \frac{dt}{|t|} \, dx \, d^\times y.$$

Because $\nu(2) = 0$, we have $c(\chi_2) = c(\chi_1 \chi_2^{-1}) = N$. Thus if $W_1(y)f(x)H(t) \neq 0$, then $y, x, t \in \mathfrak{o}, \mathfrak{o}^\times, \varpi^{-N}\mathfrak{o}^\times$ and so $f(x)\bar{f}(x + y/t) = 1$. From $\int_{x \in k} 1_{\mathfrak{o}^\times}(x) \, dx = \int_k |f|^2 = \|v'\|^2$, we conclude. $\square$

Remark 50. [Michel and Venkatesh 2010, 3.4.2; 2010, (3.25)] and Theorem 49(I) imply the following: Let $v_2, v_3 \in \pi$ be microlocal lifts of the same orientation and $v_1 \in \sigma$, realized in its Kirillov model $\mathcal{K}(\sigma, \psi)$. The formula $\|v_1\|^2 := \int_{y \in k^\times} |v_1(y)|^2 \, d^\times y$ is known to define an invariant norm on σ. Suppose that N is large enough in terms of v_1. Then

$$\int_{g \in Z \backslash G} \prod_{i=1,2,3} \langle \pi_i(g)v_i, v_i \rangle = cq^{-N} \|v_2\|^2 \|v_3\|^2 \int_{y \in k^\times} \langle a(y)v_1, v_1 \rangle \, d^\times y$$

for some positive scalar $c \asymp 1$ depending only upon k. This identity solves the problem of producing a subconvexity-critical test vector for the local triple product period in the QUE case when the varying representation is principal series. It would be interesting to verify whether the supercuspidal case follows

similarly using a modification of Definition 21 involving characters on an ε-neighborhood in $\mathrm{GL}_2(\mathfrak{o})$ of the points of a suitable nonsplit torus, where $\varepsilon \asymp C(\bar{\pi} \otimes \pi)^{-1/4}$.[13]

7. Completion of the proof

In this section, $\varphi \in \pi \in A_0(X)$ traverses a sequence of L^2-normalized microlocal lifts on X of level $N \to \infty$. Thus φ and π, like most objects to be considered in this section, depend upon N, but we omit this dependence from our notation. We use the abbreviations *fixed* to mean "independent of N" and *eventually* to mean "for large enough N." Asymptotic notation such as $o(1)$ refers to the $N \to \infty$ limit. Our aim is to verify the conclusions of Theorem 25 and Theorem 29.

As G-modules, $\pi \cong \chi_1 \boxplus \chi_2$ for some unitary characters χ_1, χ_2 of $\mathbb{Q}_p^\times$ for which $c(\chi_1/\chi_2) = N$.

Recall our simplifying assumption that R is a maximal order. This implies that for any irreducible $\mathcal{H}$-submodule π' of $\mathcal{A}(X)$, the vector space underlying π' is an irreducible admissible G-module. In other words, the local components at all places $v \neq p$ are one-dimensional.

The function φ has unitary central character, so the measure μ_φ is invariant by the center. Let ℓ be a prime dividing the discriminant of B. Recalling from Section 3.1.1 that T_ℓ is an involution modulo the center, we see that it acts on π by some scalar of magnitude one. Thus μ_φ is T_ℓ-invariant. The natural space of observables against which it suffices to test μ_φ is thus

$$\mathcal{A}^+(X) := \left\{ \Psi \in \mathcal{A}(X) : \begin{array}{l} T_\ell \Psi = \Psi \text{ for } \ell \mid \mathrm{disc}(B), \\ z\Psi = \Psi \text{ for } z \in Z := \text{center of } G, \end{array} \right\}.$$

That space decomposes further as $\mathcal{A}^+(X) = (\oplus_\chi \mathbb{C}(\chi \circ \det)) \oplus \mathcal{A}_0^+(X)$ where

- χ traverses the set of quadratic characters of the compact group $\mathbb{Q}_p^\times / \mathbb{Z}[1/p]^\times$ satisfying $\chi(\ell) = 1$ for $\ell \mid \mathrm{disc}(B)$, and

- $\mathcal{A}_0^+(X) := \mathcal{A}^+(X) \cap \mathcal{A}_0(X)$, which decomposes further as a countable direct sum $\mathcal{A}_0^+(X) = \oplus_{\sigma \in A_0^+(X)} \sigma$ where we substitute A for $\mathcal{A}$ to denote "irreducible submodules of."

Let $\sigma \in A^+(X)$ be fixed. It is either one-dimensional and of the form $\mathbb{C}(\chi \circ \det)$ for some χ as above, or belongs to $A_0^+(X)$ and is generic as a G-module. Denote by $\ell_{\mathrm{Aut}} : \sigma \otimes \bar{\pi} \otimes \pi \to \mathbb{C}$ the G-invariant functional defined by integration over X.

Lemma 51. *Suppose σ is one-dimensional and $\ell_{\mathrm{Aut}} \neq 0$. Then σ is trivial eventually.*

Proof. Write $\sigma = \mathbb{C}(\chi \circ \det)$ for some quadratic character χ. By Schur's lemma, $\pi \cong \chi_1 \boxplus \chi_2$ is isomorphic as a G-module to $\pi \otimes \chi \circ \det \cong \chi_1 \chi \boxplus \chi_2 \chi$, which happens (see, e.g., [Schmidt 2002]) only if either $\chi_1 = \chi_1 \chi$, in which case χ is trivial, or $\chi_1 = \chi_2 \chi$, in which case $c(\chi) = c(\chi_1/\chi_2) = N \to \infty$, which does not happen because χ is quadratic.[14]   □

---

[13] Added later: the recent work [Nelson et al. ≥ 2018] contains results in this direction.

[14] We use here that the local field $\mathbb{Q}_p$ is not a function field of characteristic 2.

We now prove Theorem 25. It suffices to verify that the various assertions hold for fixed $\Psi \in \sigma \in A^+(X)$. They are tautological if σ is trivial, so by Lemma 51, we reduce to the case that $\sigma \in A_0^+(X)$ is generic. Fix an unramified nontrivial character $\psi : \mathbb{Q}_p \to \mathbb{C}^{(1)}$ and G-equivariant isometric isomorphisms $\sigma \cong \mathcal{W}(\sigma, \psi)$, $\pi \cong \chi_1 \boxplus \chi_2$. Denote by $\ell_{\text{RS}} : \sigma \otimes \bar{\pi} \otimes \pi \to \mathbb{C}$ the trilinear form defined in Section 5.6. By Theorem 45 and the nonvanishing of ℓ_{RS}, there exists a complex scalar $\mathcal{L}^{1/2} \in \mathbb{C}$ so that

$$\ell_{\text{Aut}} = \mathcal{L}^{1/2}\ell_{\text{RS}}. \tag{33}$$

Theorem 49(I) implies that $\ell_{\text{RS}}(\sigma(a(y))\Psi, \bar{\varphi}, \varphi) = \ell_{\text{RS}}(\Psi, \bar{\varphi}, \varphi)$ holds eventually for fixed $y \in k^\times$; the required diagonal invariance then follows from (33). If $p \neq 2$ and φ' is an L^2-normalized newvector of support $-N..N$ and $\Psi \in \sigma^K$ is spherical, then Theorem 49(III) gives $\ell_{\text{RS}}(\Psi, \bar{\varphi}, \varphi) = \ell_{\text{RS}}(\Psi, \bar{\varphi}', \varphi')$ eventually; the lifting property then follows from (33). For the equidistribution application, we reduce by Lemma 51 and (33) and Theorem 49(II) to showing that $\mathcal{L}^{1/2} = o(p^{N/2})$ holds under the hypothesis that for each fixed $\Psi_0 \in \sigma$, one has $\ell_{\text{Aut}}(\Psi_0, \bar{\varphi}, \varphi) = o(1)$. Let $\Psi_0 \in \sigma \cong \mathcal{W}(\sigma, \psi)$ be given in the Kirillov model by the characteristic function of the unit group. By Theorem 49(I), $\ell_{\text{RS}}(\Psi_0, \bar{\varphi}, \varphi) \asymp p^{-N/2}$ eventually, so our hypothesis and (33) give the required estimate for $\mathcal{L}^{1/2}$.

We turn to the proof of Theorem 29. Our assumptions on π and σ imply that $\sigma \in A_0^+(X)$ and that the adelizations of σ, $\bar{\pi}$ and π at each $v \in S_B := \{\infty\} \cup \{\ell : \ell \mid \text{disc}(B)\}$ are one-dimensional and have trivial tensor product, hence that the product of their normalized matrix coefficients is one; by Ichino's formula [Ichino and Ikeda 2010] and [Michel and Venkatesh 2010, 3.4.2], it follows that $L \asymp |\mathcal{L}^{1/2}|^2$, where L denotes the LHS of (5) and $\mathcal{L}^{1/2}$ is as above (compare with Remark 50). By Theorem 27 and the argument of the previous paragraph, $\mathcal{L}^{1/2} = o(p^{N/2})$. Our goal is to show that $L = o(C^{1/4})$, where $C := C(\sigma \times \bar{\pi} \times \pi)$ is the global conductor; the contribution to C from $v \in S_B$ is bounded, while the contribution from p is

$$C(\sigma_p \otimes \chi_1^{-1}\chi_2)C(\sigma_p \otimes \chi_2^{-1}\chi_1)C(\sigma_p)^2 \asymp C(\chi_1^{-1}\chi_2)^4 = p^{4N}.$$

Thus $C \asymp p^{4N}$. The known estimate $\mathcal{L}^{1/2} = o(p^{N/2})$ thus translates to the goal $L = o(C^{1/4})$, as required.

Acknowledgements

References

[Anantharaman and Le Masson 2015] N. Anantharaman and E. Le Masson, "Quantum ergodicity on large regular graphs", *Duke Math. J.* **164**:4 (2015), 723–765. MR Zbl

[Atkin and Lehner 1970] A. O. L. Atkin and J. Lehner, "Hecke operators on $\Gamma_0(m)$", *Math. Ann.* **185** (1970), 134–160. MR Zbl

[Bourgain and Lindenstrauss 2003] J. Bourgain and E. Lindenstrauss, "Entropy of quantum limits", *Comm. Math. Phys.* **233**:1 (2003), 153–171. MR Zbl

[Brandt 1943] H. Brandt, "Zur Zahlentheorie der Quaternionen", *Jber. Deutsch. Math. Verein.* **53** (1943), 23–57. MR

[Brooks and Lindenstrauss 2010] S. Brooks and E. Lindenstrauss, "Graph eigenfunctions and quantum unique ergodicity", *C. R. Math. Acad. Sci. Paris* **348**:15-16 (2010), 829–834. MR Zbl

[Brooks and Lindenstrauss 2013] S. Brooks and E. Lindenstrauss, "Non-localization of eigenfunctions on large regular graphs", *Israel J. Math.* **193**:1 (2013), 1–14. MR Zbl

[Brooks and Lindenstrauss 2014] S. Brooks and E. Lindenstrauss, "Joint quasimodes, positive entropy, and quantum unique ergodicity", *Invent. Math.* **198**:1 (2014), 219–259. MR Zbl

[Bump 1997] D. Bump, *Automorphic forms and representations*, Cambridge Studies in Advanced Mathematics **55**, Cambridge University Press, 1997. MR Zbl

[Casselman 1973a] W. Casselman, "On some results of Atkin and Lehner", *Math. Ann.* **201** (1973), 301–314. MR Zbl

[Casselman 1973b] W. Casselman, "The restriction of a representation of $GL_2(k)$ to $GL_2(\mathfrak{o})$", *Math. Ann.* **206** (1973), 311–318. MR Zbl

[Eichler 1955] M. Eichler, "Zur Zahlentheorie der Quaternionen-Algebren", *J. Reine Angew. Math.* **195** (1955), 127–151. MR

[Einsiedler and Lindenstrauss 2008] M. Einsiedler and E. Lindenstrauss, "On measures invariant under diagonalizable actions: the rank-one case and the general low-entropy method", *J. Mod. Dyn.* **2**:1 (2008), 83–128. MR Zbl

[Gross 1987] B. H. Gross, "Heights and the special values of L-series", pp. 115–187 in *Number theory* (Montreal, Quebec, 1985), edited by H. Kisilevsky and J. Labute, CMS Conf. Proc. **7**, Amer. Math. Soc., Providence, RI, 1987. MR Zbl

[Holowinsky and Soundararajan 2010] R. Holowinsky and K. Soundararajan, "Mass equidistribution for Hecke eigenforms", *Ann. of Math.* (2) **172**:2 (2010), 1517–1528. MR Zbl

[Hu 2018] Y. Hu, "Triple product formula and mass equidistribution on modular curves of level N", *Int. Math. Res. Not.* **2018**:9 (2018), 2899–2943. MR

[Ichino 2008] A. Ichino, "Trilinear forms and the central values of triple product L-functions", *Duke Math. J.* **145**:2 (2008), 281–307. MR Zbl

[Ichino and Ikeda 2010] A. Ichino and T. Ikeda, "On the periods of automorphic forms on special orthogonal groups and the Gross–Prasad conjecture", *Geom. Funct. Anal.* **19**:5 (2010), 1378–1425. MR Zbl

[Knapp 1986] A. W. Knapp, *Representation theory of semisimple groups*, Princeton Mathematical Series **36**, Princeton University Press, 1986. MR Zbl

[Kneser 1966] M. Kneser, "Strong approximation", pp. 187–196 in *Algebraic Groups and Discontinuous Subgroups* (Boulder, Colorado, 1965), edited by A. Borel and G. D. Mostow, Proc. Sympos. Pure Math. **9**, Amer. Math. Soc., Providence, RI, 1966. MR Zbl

[Le Masson 2014] E. Le Masson, "Pseudo-differential calculus on homogeneous trees", *Ann. Henri Poincaré* **15**:9 (2014), 1697–1732. MR Zbl

[Le Masson and Sahlsten 2017] E. Le Masson and T. Sahlsten, "Quantum ergodicity and Benjamini–Schramm convergence of hyperbolic surfaces", *Duke Math. J.* **166**:18 (2017), 3425–3460. MR Zbl

[Lindenstrauss 2001] E. Lindenstrauss, "On quantum unique ergodicity for $\Gamma \backslash \mathbb{H} \times \mathbb{H}$", *Internat. Math. Res. Notices* 17 (2001), 913–933. MR Zbl

[Lindenstrauss 2006a] E. Lindenstrauss, "Adelic dynamics and arithmetic quantum unique ergodicity", pp. 111–139 in *Current developments in mathematics, 2004*, edited by D. Jerison et al., Int. Press, Somerville, MA, 2006. MR Zbl

[Lindenstrauss 2006b] E. Lindenstrauss, "Invariant measures and arithmetic quantum unique ergodicity", *Ann. of Math.* (2) **163**:1 (2006), 165–219. MR Zbl

[Lubotzky et al. 1988] A. Lubotzky, R. Phillips, and P. Sarnak, "Ramanujan graphs", *Combinatorica* **8**:3 (1988), 261–277. MR Zbl

[Luo and Sarnak 2003] W. Luo and P. Sarnak, "Mass equidistribution for Hecke eigenforms", *Comm. Pure Appl. Math.* **56**:7 (2003), 874–891. MR Zbl

[Michel and Venkatesh 2010] P. Michel and A. Venkatesh, "The subconvexity problem for GL_2", *Publ. Math. Inst. Hautes Études Sci.* 111 (2010), 171–271. MR Zbl

[Nelson 2011] P. D. Nelson, "Equidistribution of cusp forms in the level aspect", *Duke Math. J.* **160**:3 (2011), 467–501. MR Zbl

[Nelson 2012] P. D. Nelson, "Mass equidistribution of Hilbert modular eigenforms", *Ramanujan J.* **27**:2 (2012), 235–284. MR Zbl

[Nelson 2015] P. D. Nelson, "Evaluating modular forms on Shimura curves", *Math. Comp.* **84**:295 (2015), 2471–2503. MR Zbl

[Nelson 2016] P. D. Nelson, "Quantum variance on quaternion algebras, I", preprint, 2016. arXiv

[Nelson 2017] P. D. Nelson, "Analytic isolation of newforms of given level", *Arch. Math.* (*Basel*) **108**:6 (2017), 555–568. MR Zbl

[Nelson et al. 2014] P. D. Nelson, A. Pitale, and A. Saha, "Bounds for Rankin–Selberg integrals and quantum unique ergodicity for powerful levels", *J. Amer. Math. Soc.* **27**:1 (2014), 147–191. MR Zbl

[Nelson et al. ≥ 2018] P. D. Nelson, A. Saha, and Y. Hu, "Some analytic aspects of automorphic forms on GL(2) of minimal type", To appear in *Comm. Math. Helv.*

[Piatetski-Shapiro and Rallis 1987] I. Piatetski-Shapiro and S. Rallis, "Rankin triple *L* functions", *Compositio Math.* **64**:1 (1987), 31–115. MR Zbl

[Pizer 1980] A. Pizer, "An algorithm for computing modular forms on $\Gamma_0(N)$", *J. Algebra* **64**:2 (1980), 340–390. MR Zbl

[Prasad 1990] D. Prasad, "Trilinear forms for representations of GL(2) and local ϵ-factors", *Compositio Math.* **75**:1 (1990), 1–46. MR Zbl

[Reznikov 2001] A. Reznikov, "Laplace–Beltrami operator on a Riemann surface and equidistribution of measures", *Comm. Math. Phys.* **222**:2 (2001), 249–267. MR Zbl

[Rudnick and Sarnak 1994] Z. Rudnick and P. Sarnak, "The behaviour of eigenstates of arithmetic hyperbolic manifolds", *Comm. Math. Phys.* **161**:1 (1994), 195–213. MR Zbl

[Sage 2015] W. A. Stein et al., "Sage Mathematics Software", 2015, available at http://www.sagemath.org. Version 6.7.

[Sarnak 2011] P. Sarnak, "Recent progress on the quantum unique ergodicity conjecture", *Bull. Amer. Math. Soc.* (*N.S.*) **48**:2 (2011), 211–228. MR Zbl

[Schmidt 2002] R. Schmidt, "Some remarks on local newforms for GL(2)", *J. Ramanujan Math. Soc.* **17**:2 (2002), 115–147. MR Zbl

[Serre 2003] J.-P. Serre, *Trees*, Springer, Berlin, 2003. MR Zbl

[Silberman and Venkatesh 2007] L. Silberman and A. Venkatesh, "On quantum unique ergodicity for locally symmetric spaces", *Geom. Funct. Anal.* **17**:3 (2007), 960–998. MR Zbl

[Silberman and Venkatesh ≥ 2018] L. Silberman and A. Venkatesh, "Quantum unique ergodicity for locally symmetric spaces II", available at http://www.math.ubc.ca/~lior/work/AQUE-nov6.pdf. Zbl

[Soundararajan 2010] K. Soundararajan, "Weak subconvexity for central values of *L*-functions", *Ann. of Math.* (2) **172**:2 (2010), 1469–1498. MR Zbl

[Templier 2014] N. Templier, "Large values of modular forms", *Camb. J. Math.* **2**:1 (2014), 91–116. MR Zbl

[Venkatesh 2010] A. Venkatesh, "Sparse equidistribution problems, period bounds and subconvexity", *Ann. of Math.* (2) **172**:2 (2010), 989–1094. MR Zbl

[Vignéras 1980] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics **800**, Springer, Berlin, 1980. MR Zbl

[Voight 2018] J. Voight, "Quaternion algebras", Dartmouth College, 2018, available at https://math.dartmouth.edu/~jvoight/quat.html. Zbl

[Wolpert 2001]   S. A. Wolpert, "The modulus of continuity for $\Gamma_0(m)\backslash\mathbb{H}$ semi-classical limits", *Comm. Math. Phys.* **216**:2 (2001), 313–323. MR Zbl

[Zelditch 1987]   S. Zelditch, "Uniform distribution of eigenfunctions on compact hyperbolic surfaces", *Duke Math. J.* **55**:4 (1987), 919–941. MR Zbl

[Zelditch 1992]   S. Zelditch, "On a "quantum chaos" theorem of R. Schrader and M. Taylor", *J. Funct. Anal.* **109**:1 (1992), 1–21. MR Zbl

paul.nelson@math.ethz.ch   *Departement Mathematik, ETH, Zurich, Switzerland*

Heights on squares of modular curves

Pierre Parent
Appendix by Pascal Autissier

We develop a strategy for bounding from above the height of rational points of modular curves with values in number fields, by functions which are polynomial in the curve's level. Our main technical tools come from effective Arakelov descriptions of modular curves and jacobians. We then fulfill this program in the following particular case:

If p is a not-too-small prime number, let $X_0(p)$ be the classical modular curve of level p over $\mathbb{Q}$. Assume Brumer's conjecture on the dimension of winding quotients of $J_0(p)$. We prove that there is a function $b(p) = O(p^5 \log p)$ (depending only on p) such that, for any quadratic number field K, the j-height of points in $X_0(p)(K)$ which are not lifts of elements of $X_0^+(p)(\mathbb{Q})$ is less or equal to $b(p)$.

1. Introduction

Let N be an integer, Γ_N a level-N congruence subgroup of $\mathrm{GL}_2(\mathbb{Z})$, and X_{Γ_N} the associated modular curve over some subfield of $\mathbb{Q}(\mu_N)$ which, to simplify the discussion, we assume from now on to be $\mathbb{Q}$. The genus g_N of X_{Γ_N} grows roughly as a polynomial function of N. So, if N is not too small, X_{Γ_N} has only a finite number of rational points with values in any given number field, by Mordell–Faltings. If one is interested in explicitly determining the set of rational points, however, finiteness is of course not sufficient; a much more desirable control would be provided by upper bounds, for some handy height, on those points. Proving such an "effective Mordell" is known to be an extremely hard problem for arbitrary algebraic curves on number fields.

In the case of modular curves, however, the situation is much better. Indeed, whereas the jacobian of a random algebraic curve should be a somewhat equally random simple abelian variety, it is well-known that the jacobian J_{Γ_N} of X_{Γ_N} decomposes up to isogeny into a product of quotient abelian varieties defined by Galois orbits of newforms for Γ_N. Moreover, in many cases, a nontrivial part of those factors happen to have rank zero over $\mathbb{Q}$. Our rustic starting observation is therefore the following: if $J_{\Gamma_N,e}$ is the "winding quotient" of J_{Γ_N}, that is the largest quotient $J_{\Gamma_N,e}$ with trivial $\mathbb{Q}$-rank, and

$$X_{\Gamma_N} \overset{\iota}{\hookrightarrow} J_{\Gamma_N} \overset{\pi_e}{\twoheadrightarrow} J_{\Gamma_N,e}$$

is some Albanese map from the curve to its jacobian followed by the projection to $J_{\Gamma_N,e}$, then any rational point on X_{Γ_N} has an image which is a torsion point (because rational) on $J_{\Gamma_N,e}$, hence has 0 normalized height. The pull-back of some invertible sheaf defining the (say) theta height on $J_{\Gamma_N,e}$ therefore defines a height on X_{Γ_N} which is trivial on rational points. That height in turn necessarily compares to any other natural one, for instance the modular j-height. Therefore the j-height of any rational point on X_{Γ_N} is also zero "up to error terms". Making those error terms explicit would give us the desired upper bound for the height of rational points on X_{Γ_N}.

That approach can in principle be generalized to degree-d number fields, by considering rational points on symmetric powers $X_{\Gamma_N}^{(d)}$ of X_{Γ_N} (at least if $\dim J_{\Gamma_N,e} \geq d$). To be a little bit more precise in the present case of symmetric squares, let us associate to a quadratic point P in $X_0(p)$ the $\mathbb{Q}$-point $Q := (P, {}^\sigma P)$ of $X_0(p)^{(2)}$. Its image $\iota(Q)$ via some appropriate Albanese embedding in $J_0(p)$ lies above a torsion point a in J_e: assume for simplicity $a = 0$. We therefore know $\iota(Q)$ belongs to the intersection of $\iota(X_0(p)^{(2)})$ with the kernel $\tilde{J}_e^\perp$ of the projection

$$\pi_e : J_0(p) \twoheadrightarrow J_e.$$

To improve the situation we can further remark that $\iota(Q)$ actually lies at the intersection of $\iota(X_0(p)^{(2)})$ with the "projection", in some appropriate sense, of the latter surface on $\tilde{J}_e^\perp$. Then one can show that this intersection is 0-dimensional (but here we need to assume Brumer's conjecture, see below) so that its theta height is controlled, via some arithmetic Bézout theorem, in terms of the degree and height of the two surfaces we intersect. Using an appropriate version of Mumford's repulsion principle one derives a bound for the height of $\iota(P)$ too (and not only for its sum $\iota(Q)$ with its Galois conjugate). Then one makes the translation again from theta height to j-height on $X_0(p)$.

Nontrivial technical work is of course necessary to give sense to the straightforward strategy sketched above. The aim of this article is thus to show the possibility of that approach, by making it work in what we feel to be the simplest nontrivial case: that of quadratic points of the classical modular curve $X_0(p)$ as above (or $X_0(p^2)$, for technical reasons), for p a prime number.[1] In the course of the proof we

---

[1]Larson and Vaintrob [2014, Corollary 6.5] have proven, under the generalized Riemann hypothesis, the asymptotic triviality of rational points on $X_0(p)$ with values in any given number field which does not contain the Hilbert class field of some quadratic imaginary field. Independently of any conjecture, Momose [1995] had already proven the same result in the case where K is a given quadratic number field. Our method however provides bounds which do not depend on the field, and should generalize to some other congruence subgroups.

are led to assume the already mentioned conjecture of Brumer, which asserts that the winding quotient of $J_0(p) := J_{\Gamma_0(p)}$ has dimension roughly half that of $J_0(p)$. That hypothesis is actually used in only one, technical, but crucial place, where we prove that a morphism between two curves is a generic isomorphism (see last point of Lemma 7.2). Note that a lower bound of $\frac{1}{4}$ (instead of the desired $\frac{1}{2}$) for the asymptotic ratio $\dim J_e / \dim J_0(p)$ has been proven by Iwaniec and Sarnak [2000] and Kowalski, Michel and Vanderkam [Kowalski et al. 2000]. (Actually, $\frac{1}{3} + \varepsilon$ would be sufficient for us; see Lemma 7.2 and the proof of Theorem 7.5 below.) In any case we cannot at the moment get rid of this assumption—note it can in principle be numerically checked in all specific cases. In this setting, our main result is the following (see Theorem 7.5).

Theorem 1.1. *For w_p the Fricke involution, set $X_0^+(p) = X_0(p)/w_p$. Assume Brumer's conjecture (see Section 2, (21)).[2] Then the quadratic points of $X_0(p)$, which are not lifts of elements of $X_0^+(p)(\mathbb{Q})$, have j-height bounded from above by $O(p^5 \log p)$.*

The same holds true for quadratic points of $X_0(p^2)$, without the restriction about $X_0^+(p)$.

Needless to say, this result cries for both sharpening and generalization. Yet it should be possible to immediately use avatars of Theorem 1.1 to prove that rational points are only cusps and CM points, for some specific modular curves of arithmetic interest. If combined with lower bounds for heights furnished by isogeny theorems as in [Bilu et al. 2013], the above theorem already has consequences on rational points (see Corollary 7.6).

Regarding past works about rational points on modular curves, one can notice that most of them use, at least in parts, some variants of Mazur's method, which can very roughly be divided into two steps: first, map modular curves to winding quotients as described above; then prove some quite delicate properties about completions of that map to J_e (formal immersion criteria). The second step is probably the most difficult to carry over to great generality. Therefore, the method we propose here allows one to use only the first and crucial fact: the mere existence of nontrivial winding quotients. In many cases, the existence of such quotients is known by a deep result of Kolyvagin, Logachev and Kato, à la Birch–Swinnerton-Dyer conjecture, which, again, seems to reflect, from the arithmetic point of view, the special properties of the image locus (in the moduli space of principally polarized abelian varieties) of modular curves, among all algebraic curves, under Torelli's map.

The methods used in this paper are mainly explicit Arakelov techniques for modular curves and abelian varieties. Such techniques and results have been pioneered, as far as we know, by Abbes, Michel and Ullmo at the end of the 1990s (see in particular [Abbes and Ullmo 1995; Michel and Ullmo 1998; Ullmo 2000], whose results we here eagerly use). They have subsequently been revisited and extended in the work developed by Edixhoven and his school, as mainly (but not exhaustively) presented in the orange book [Edixhoven and Couveignes 2011]. That work was motivated by algorithmic Galois-representation issues, but its tools are well suited to our rational points questions, as we wish to show here. We similarly

---

[2]The weak version of that conjecture we actually need is stated in (22).

hope that the effective Arakelov results about modular curves and jacobians we work out in the present article shall prove useful in other contexts.[3]

The layout of this article is as follows. In Section 2 we start gathering classical instrumental facts on quotients of modular jacobians and regular models of $X_0(p)$ over rings of algebraic integers. In Section 3 we make a precise description of the arithmetic Chow group of $X_0(p)$. Section 4 provides an explicit comparison theorem between j-heights and pull-back of normalized theta height on the jacobian. Section 5 computes the degree and Faltings height of the image of symmetric products within modular jacobians. In Section 6 we prove our arithmetic Bézout theorem (in the sense of [Bost et al. 1994]) for cycles in $J_0(p)$, relative to cubist metrics (instead of the more usual Fubini–Study metrics). This seems more natural and has the advantage of being quantitatively more efficient; that constitutes the technical heart of the present paper. Then we apply that arithmetic Bézout to our modular jacobian after technical computations on metric comparisons. Section 7 concludes the computations of the height bounds for quadratic rational points on $X_0(p)$ by making various intersections, projections and manipulations for which to refer to [loc. cit.].

Convention. In order to avoid numerical troubles, we safely assume in all of what follows that primes are by definition strictly larger than 17.

2. Curves, jacobians, their quotients and subvarieties

2A. *Abelian varieties.*

2A1. *Decompositions.* Let K be a field, J an abelian variety of dimension g over K and $\mathcal{L}$ an ample invertible sheaf defining a polarization of J. Assume J is K-isogenous to a product of two (nonzero) subvarieties, that is, there are abelian subvarieties

$$\iota_A : A \hookrightarrow J, \quad \iota_B : B \hookrightarrow J \tag{1}$$

endowed with polarizations $\iota_A^*(\mathcal{L})$ and $\iota_B^*(\mathcal{L})$, respectively, such that $\iota_A + \iota_B : A \times B \to J$ is an isogeny. (Recall that by convention, all abelian (sub)varieties are assumed to be connected.) Then $\pi_A : J \to A' := J \bmod B$, and similarly $\pi_B : J \to B'$, are called *optimal quotients* of J.

To simplify things we also assume from now on that $\mathrm{End}_{\overline{K}}(A, B) = \{0\}$. The product isogeny $\pi := \pi_A \times \pi_B : J \to A' \times B'$ induces isogenies $A \to A'$ and $B \to B'$. We write

$$\Phi : A \times B \to J \to A' \times B'$$

for the obvious composition. Taking for instance dual isogenies of $A \to A'$ and $B \to B'$, we also define an endomorphism

$$\Psi : J \to A' \times B' \to A \times B \to J. \tag{2}$$

---

[3]For recent investigations related to more general questions of effective bounds of algebraic points on curves, one can check [Checcoli et al. 2016].

When $K = \mathbb{C}$, the above constructions are transparent. There is a $\mathbb{Z}$-lattice Λ in $\mathbb{C}^g$, endowed with a symplectic pairing, such that $J(\mathbb{C}) \simeq \mathbb{C}^g / \Lambda$ and one can find a direct sum decomposition $\mathbb{C}^g = \mathbb{C}^{g_A} \oplus \mathbb{C}^{g_B}$ such that if $\Lambda_A = \Lambda \cap \mathbb{C}^{g_A}$ and $\Lambda_B = \Lambda \cap \mathbb{C}^{g_B}$, then

$$A(\mathbb{C}) \simeq \mathbb{C}^{g_A} / \Lambda_A \quad \text{and} \quad B(\mathbb{C}) \simeq \mathbb{C}^{g_B} / \Lambda_B.$$

If $p_A : \mathbb{C}^g \to \mathbb{C}^{g_A}$ and $p_B : \mathbb{C}^g \to \mathbb{C}^{g_B}$ are the $\mathbb{C}$-linear projections relative to that decomposition, the analytic description of $\pi_{A,\mathbb{C}} : J(\mathbb{C}) \to A'(\mathbb{C})$ is then

$$z \bmod \Lambda \mapsto z \bmod (\Lambda + \Lambda_B \otimes \mathbb{R}) = p_A(z) \bmod (p_A(\Lambda)).$$

Summing up, we have lattice inclusions $\Lambda_A \subseteq p_A(\Lambda)$ and $\Lambda_B \subseteq p_B(\Lambda)$, with finite indices, in $\mathbb{C}^g$ such that our isogenies are induced by

$$\Lambda_A \oplus \Lambda_B \subseteq \Lambda \subseteq p_A(\Lambda) \oplus p_B(\Lambda).$$

The isogeny $I'_A : A \to A'$ deduced from the inclusion $\Lambda_A \subseteq p_A(\Lambda)$ has degree $\mathrm{card}(p_A(\Lambda)/\Lambda_A)$. If N_A is a multiple of the exponent of the quotient $p_A(\Lambda)/\Lambda_A$, there is an isogeny $I_{A,N_A} : A' \to A$ such that $I_{A,N_A} \circ I'_A$ and $I'_A \circ I_{A,N_A}$ both are multiplication by N_A. The analytic descriptions of the above clearly are:

$$
\begin{array}{ccc}
A(\mathbb{C}) \simeq \mathbb{C}^{g_A}/\Lambda_A \xrightarrow{I'_A} A'(\mathbb{C}) \simeq \mathbb{C}^{g_A}/p_A(\Lambda) & \text{and} & \mathbb{C}^{g_A}/p_A(\Lambda) \xrightarrow{I_{A,N_A}} \mathbb{C}^{g_A}/\Lambda_A \\
z \longmapsto z & & z \longmapsto N_A z.
\end{array}
\tag{3}
$$

Remark 2.1. Instead of considering two immersions as in (1), suppose only $A \hookrightarrow J$ is given, and K is a number field. One might apply [Gaudron and Rémond 2014a, Théorème 1.3] to deduce the existence of an abelian variety B over K such that, with our previous notations, the degree of $A \times B \xrightarrow{+} J$,

$$|A \cap B| = |\Lambda / \Lambda_A \oplus \Lambda_B|,$$

is bounded from above by an explicit function $\kappa(J)$ of the stable Faltings' height $h_F(J)$,

$$\kappa(J) = ((14g)^{64g^2}[K : \mathbb{Q}] \max(h_F(J), \log[K : \mathbb{Q}], 1)^2)^{2^{10} g^3},$$

and this does not depend on the choice of the embedding $K \hookrightarrow \mathbb{C}$. Note that when A and $J \bmod A$ are not isogenous (which will be the case for us), then there is actually no choice for that $B \hookrightarrow J$: it has to be the Poincaré complement to A. The isogeny $J \to A' \times B'$ given by the two projections has degree $|p_A(\Lambda) \oplus p_B(\Lambda)/\Lambda|$, which also is $|A \cap B| := N$. One can therefore take the N_A appearing in (3) as equal to N, and

$$N \leq \kappa(J).$$

Making the same for $B' \to B$, the above morphism Ψ (see (2)) is then simply the multiplication $J \xrightarrow{[N \cdot]} J$ by the integer N. Although we will not need numerical estimates for those quantities in what follows, it is straightforward, using [Ullmo 2000], to make them explicit in our setting of modular curves and jacobians.

2A2. *Polarizations and heights.* Keeping the above notations and hypothesis, consider in addition now an ample sheaf Θ on J and let $I_A := I_{A,N} : A' \to A$ (respectively, $I_{B,N}$) be as in (3). We pull-back Θ along the composed morphism

$$\varphi_A : J \xrightarrow{\pi_A} A' \xrightarrow{I_A} A \xrightarrow{\iota_A} J \tag{4}$$

so that the immersion $\iota_A : A \hookrightarrow J$ defines a polarization $\Theta_A := \iota_A^*(\Theta)$ on A, whence a polarization $\Theta_{A'} := I_A^*(\Theta_A)$ on A', and finally an invertible sheaf $\Theta_{J,A} := \pi_A^*(\Theta_{A'})$ on J. Composing the morphisms

$$J \xrightarrow{\pi_A \times \pi_B} A' \times B' \xrightarrow{I_A \times I_B} A \times B \xrightarrow{\iota_A + \iota_B} J \tag{5}$$

gives the multiplication-by-N map $J \xrightarrow{[\cdot N]} J$. Assuming for simplicity Θ is symmetric one therefore has

$$[\cdot N]^* \Theta \simeq \Theta^{\cdot \otimes N^2} \simeq \Theta_{J,A} \otimes_{\mathcal{O}_J} \Theta_{J,B}. \tag{6}$$

If K is a number field, the Néron–Tate normalization process associates with Θ a system of compatible Euclidean norms $h_\Theta = \|\cdot\|_\Theta^2$ on the finite-dimensional $\mathbb{Q}$-vector spaces $J(F) \otimes_{\mathbb{Z}} \mathbb{Q}$, for F/K running through the number field extensions of K, and similarly Euclidean norms

$$h_{\Theta_A} := \|\cdot\|^2_{\Theta_A^{\cdot \otimes \frac{1}{N^2}}} := \frac{1}{N^2} \|\cdot\|_{\Theta_A}^2 \quad \text{and} \quad h_{\Theta_B} := \frac{1}{N^2} \|\cdot\|_{\Theta_B}^2$$

on $A(F) \otimes_{\mathbb{Z}} \mathbb{Q}$ and $B(F) \otimes_{\mathbb{Z}} \mathbb{Q}$, respectively, such that, under the isomorphisms $J(F) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq (A(F) \otimes_{\mathbb{Z}} \mathbb{Q}) \oplus (B(F) \otimes_{\mathbb{Z}} \mathbb{Q})$, one has

$$h_\Theta = h_{\Theta_A} + h_{\Theta_B}. \tag{7}$$

Recall from (3) the definition of N_A, that of the maps $A' \xrightarrow{I_{A,N_A}} A$ and $A \xhookrightarrow{\iota_A} J$. Denote by $[N_A]_A$ the multiplication by N_A restricted to A. If V is a closed algebraic subvariety of J, define

$$\mathcal{P}_A(V) := (\iota_A [N_A]_A^{-1} I_{A,N_A} \pi_A)(V) \tag{8}$$

as the reduced closed subscheme with relevant support. The map $\mathcal{P}_A$ would simply be the projection of V on A if J were *isomorphic* to the product $A \times B$ of subvarieties and is the best approximation to that projection in our case when J is only isogenous to $A \times B$.

Note that $\mathcal{P}_A(V)$ is a priori highly nonconnected. All its irreducible geometric components are however obtained from each other by translation by an N_A-torsion point of $A(\overline{\mathbb{Q}})$. For our later purposes (see the proof of Theorem 7.5), we will have the possibility to replace $\mathcal{P}_A(V)$ by one of its components containing a specific point, say P_0: we shall denote that component by $\mathcal{P}_A(V)_{P_0}$ and refer to it as the "pseudoprojection" of V on A containing P_0.

Suppose now $J \sim A \times B$ as above is the jacobian of an algebraic curve X on K with positive genus g. For P_0 a point of $X(K)$ (or more generally a K-divisor of degree 1 on X) let

$$\iota_{P_0} : X \hookrightarrow J, \quad P \mapsto (P) - (P_0), \tag{9}$$

be the Albanese embedding associated with P_0. We define the classical theta divisor θ on J which is the image of $\iota_{P_0}^{g-1} : X^{g-1} \to J$ and its symmetric version

$$\Theta := (\theta \otimes_{\mathcal{O}_J} [-1]^* \theta)^{\cdot \otimes 1/2} \tag{10}$$

(which is a translate of θ obtained as $\iota_{\kappa_0}^{g-1}(X^{g-1})$, where $\iota_{\kappa_0} = t_{\kappa_0}^* \iota_{P_0}$ for t_{κ_0} the translation by some κ_0 with $(2g-2)\kappa_0 = \kappa$, the canonical divisor on X; of course Θ does not need to be defined over K). Our first aim will be to compare the height functions $\|\iota_{P_0}(\cdot)\|_{\Theta_A \cdot \otimes 1/N^2}$ on $X(F)$, when X is a modular curve, with another natural height given by the modular j-function.

We will discuss in [Section 3](#) an Arakelov description of Néron–Tate height. We conclude this paragraph by a few remarks as a preparation. Let $B_2 := \{\omega_1, \ldots, \omega_g\}$ be a basis of $H^0(X(\mathbb{C}), \Omega_{X/\mathbb{C}}^1) \simeq H^0(J(\mathbb{C}), \Omega_{J/\mathbb{C}}^1)$, which is orthogonal with respect to the norm

$$\|\omega\|^2 = \frac{i}{2} \int_{X(\mathbb{C})} \omega \wedge \bar{\omega}.$$

The transcendent writing-up of the Abel–Jacobi map $\iota_{P_0} : P \mapsto \left(\int_{P_0}^P \omega_i \right)_{1 \leq i \leq g}$ shows that the pull-back to $X(\mathbb{C})$ of the translation-invariant measure on $J(\mathbb{C})$, normalized to have total mass 1, is

$$\mu_0 = \frac{i}{2g} \sum_{B_2} \frac{\omega \wedge \bar{\omega}}{\|\omega\|^2}. \tag{11}$$

More generally, $\pi_A \circ \iota_{P_0}$ is, over $\mathbb{C}$, the map $P \mapsto \left(\int_{P_0}^P \omega \right)_{\omega \in B_2^A}$, where B_2^A is some orthogonal basis of $H^0(A'(\mathbb{C}), \Omega_{A'/\mathbb{C}}^1) \simeq H^0(J(\mathbb{C}), \pi_A^*(\Omega_{A'/\mathbb{C}}^1)) \subseteq H^0(J(\mathbb{C}), \Omega_{J/\mathbb{C}}^1)$. Therefore, writing $g_A := \dim(A') = \dim(A)$ (we assume $A \neq 0$), the pull-back to $X(\mathbb{C})$ of the translation-invariant measure on $A'(\mathbb{C})$ (normalized so to have total mass 1 on the curve again) is

$$\mu_A = \frac{i}{2g_A} \sum_{B_2^A} \frac{\omega \wedge \bar{\omega}}{\|\omega\|^2}. \tag{12}$$

2B. *Modular curves.* Here we recall a few classical facts on the minimal regular model of the modular curve $X_0(p)$, for p a prime number, over a ring of algebraic integers. The first general reference on this topic is [Deligne and Rapoport 1973]; see also [Edixhoven and Couveignes 2011; Menares 2008; 2011].

2B1. *The j-height.* The quotient of the completed Poincaré upper half-plane $\mathscr{H} \cup \mathbb{P}^1(\mathbb{Q})$ by the classical congruence subgroup $\Gamma_0(p)$ defines a Riemann surface $X_0(p)(\mathbb{C})$ which is known to have a geometrically connected smooth and proper model over $\mathbb{Q}$. All through this paper, we denote its genus by g.

The first technical theme of this article is the explicit comparison of various heights on $X_0(p)(\bar{\mathbb{Q}})$. When V is an algebraic variety over a number field K, any finite K-map $\varphi : V \to \mathbb{P}_K^N$ to some projective space defines a naive Weil height on $V(\bar{K})$. This applies in particular when V is a curve and φ is the finite morphism defined by an element of the function field of V, and in the case of a modular curve X_Γ associated with some congruence subgroup Γ, say, a natural height to choose on $X_\Gamma(\bar{\mathbb{Q}})$ is precisely Weil's height $\mathrm{h}(P) = \mathrm{h}(j(P))$ relative to the classical j-function. The degree of the associated map $X_\Gamma \to X(1) \simeq \mathbb{P}^1$

is $[\mathrm{PSL}_2(\mathbb{Z}) : \Gamma]$, so that number is the class of our Weil height in the Néron–Severi group $\mathrm{NS}(X_\Gamma)$ identified with $\mathbb{Z}$. More explicitly if $X = X_\Gamma$ is defined over the number field K, say, the j-morphism is

$$X \xrightarrow{\ j\ } \mathbb{P}^1_K = \mathrm{Proj}(K[X_0, X_1]) \hookleftarrow \mathbb{A}^1_K = \mathrm{Spec}(K[X_1/X_0])$$

$$P \longmapsto (1, j(P)) = (1/j(P), 1) \leftarrow j(P) = \frac{X_1}{X_0}(P),$$

and the Weil height of a point $P \in X(K)$ is therefore the naive height of its j-invariant as an algebraic number

$$\mathrm{h}(P) = \mathrm{h}(j(P)) = \frac{1}{[K : \mathbb{Q}]} \sum_{v \in M_K} [K_v : \mathbb{Q}_v] \log(\max(1, |j(P)|_v))$$

which is also Weil's projective height $\mathrm{h}(j(P))$ with respect to the above basis $(X_0, X_1 = X_0 j)$ of global sections of $\mathcal{O}_{\mathbb{P}^1_K}(1)$. Our Weil height on X is associated with the linear equivalence classes of divisors D corresponding to $j^*(\mathcal{O}_{\mathbb{P}^1_K}(1))$, so that

$$D \sim (\text{poles of } j \text{ on } X)(\sim (\text{zeroes of } j)) \sim \sum_{c \in \{\text{cusps of } X\}} e_c . c$$

where each e_c is the ramification index of c via j.

Those considerations lead to explicit comparisons with other heights. Indeed, a more intrinsic way to define heights on algebraic varieties is provided by Arakelov theory. Defining this properly in the case of our modular curves demands a precise description of regular models for them, which we now recall.

2B2. *Regular models.* The normalization of the j-map $X_0(p) \to X(1)_{/\mathbb{Z}} \simeq \mathbb{P}^1_{/\mathbb{Z}}$ over $\mathbb{Z}$ defines a model for $X_0(p)$ that we call the modular model, it is smooth over $\mathbb{Z}[1/p]$.

We fix a number field K, write $\mathcal{O}_K$ for its ring of integers, and deduce by base change a model for $X_0(p)$ over $\mathcal{O}_K$. We know its only singularities are normal crossing, so after a few blow-ups, if necessary, we obtain a regular model of $X_0(p)$ over $\mathcal{O}_K$; see Theorem 1.1.d of the Appendix of [Mazur 1977]. We denote it from now on by $\mathcal{X}_0(p)_{/\mathcal{O}_K}$, or simply $\mathcal{X}_0(p)$ if the context prevents confusion. We stress here that for F/K a field extension, $\mathcal{X}_0(p)_{/\mathcal{O}_F}$ is *not* the base change to $\mathcal{O}_F$ of $\mathcal{X}_0(p)_{/\mathcal{O}_K}$ if F/K ramifies above p. Let v be a place of $\mathcal{O}_K$ above p, with residue field $k(v)$. The dual graph of $\mathcal{X}_0(p)$ at v is made of two extremal vertices, which we label C_0 and C_∞, containing the cusps 0 and ∞ respectively (see Figure 1). Those two vertices, which correspond to irreducible components of genus 0, are linked by

$$s := g + 1$$

branches. Each branch corresponds to a singular point S in $\mathcal{X}_0(p)(\mathbb{F}_{p^2})$, which in turn parametrizes an isomorphism class of supersingular elliptic curve E_S in characteristic p.

The Fricke involution w_p acts on the dual graph as the continuous isomorphism which exchanges C_0 and C_∞ and acts on the branches as a generator of $\mathrm{Gal}(\mathbb{F}_{p^2}/\mathbb{F}_p)$.

We list the supersingular points as $S(1), \ldots, S(s)$ and for each one define

$$w_n := \# \mathrm{Aut}(S(n))/\langle \pm 1 \rangle := \# \mathrm{Aut}_{\mathbb{F}_{p^2}}(E_{S(n)})/\langle \pm 1 \rangle \tag{13}$$
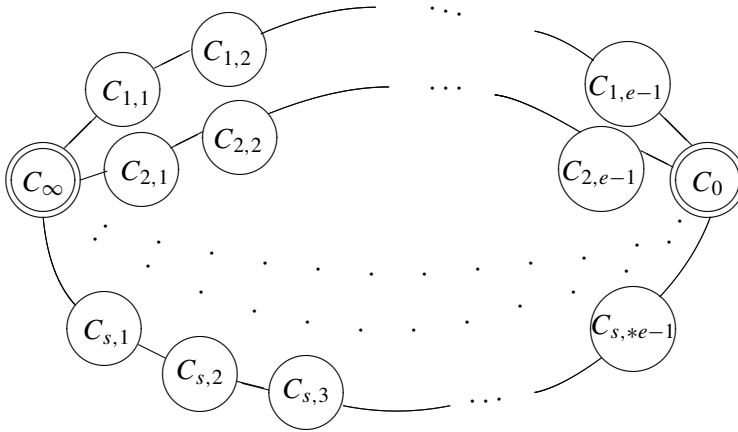
Figure 1. Dual graph of $\mathcal{X}_0(p)_{/\mathcal{O}_K}$ at v.

which is equal to 1 except in the (at most two) cases when the underlying supersingular elliptic curve has j-invariant 1728 or 0, where it is equal to 2 or 3 respectively. Now each path, or branch, on our dual graph at v passes through $(w_n e - 1)$ vertices (for e the ramification index of K at v), that is, again, equal to $e - 1$ except for at most two branches: one of length $2e - 1$ (obtained by blowing-up the supersingular point of moduli $j \equiv 1728 \bmod v$, if it exists) and a path of length $3e - 1$ (obtained by blowing-up, if needed, at the supersingular point of moduli $j \equiv 0 \bmod v$). We enumerate the vertices $\{C_{n,m}\}_{1 \le m \le w_n e - 1}$ in the n-th path. We also denote by $w(\mathrm{Eis})$ the familiar quantity $\sum 1/w_n$, the sum being taken over the set of all supersingular points of $\mathcal{X}_0(p)_{/\mathcal{O}_{K,v}}$. The well-known Eichler mass formula says that

$$w(\mathrm{Eis}) = \sum_{1 \le n \le s} \frac{1}{w_n} = \frac{p-1}{12} \tag{14}$$

(see for instance [Gross 1987b, p. 117]). Recall that this implies the genus g of $X_0(p)$ is asymptotically equivalent to $p/12$ (the exact formula depending on the residue class of $p \bmod 12$) and in any case

$$\frac{p-13}{12} \le g \le \frac{p+1}{12} \tag{15}$$

(see for instance [Gross 1987b, p. 117], again).

Abusing notation a bit, C_∞ will sometimes also be denoted as $C_{n,0}$ and similarly C_0 might be written as $C_{n,w_n e}$. We choose as a basis for $\oplus_C \mathbb{Z} \cdot C$ the ordered set

$$\mathcal{B} = (C_\infty, (C_{1,1}, C_{1,2}, \ldots, C_{1,e-1}), (C_{2,1}, \ldots, C_{2,e-1}), \ldots, (C_{s,1}, \ldots, C_{s,w_s e-1}), C_0) \tag{16}$$

(that is, we enumerate the vertices by running through each branch successively, and put the possible branches of length twice or thrice the generic length at the end). At bad places v the intersection matrix restricted to each submodule $\oplus_{m=1}^{w_n e-1} \mathbb{Z} \cdot C_{n,m}$ (for some fixed branch of index n) is then $(\log(\#k(v)) \cdot \mathcal{M}_0,$

where

$$\mathcal{M}_0 = \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & 1 & -2 & 1 \\ 0 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix} \tag{17}$$

whose only dependence on n is that its type is $(w_n e - 1) \times (w_n e - 1)$. That matrix has determinant $(-1)^{w_n e - 1} w_n e$. Define the row vectors

$$L := \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad L' := \begin{pmatrix} 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

(with length implicitly defined by the next lines) and the transpose column vectors

$$V := L^t, \quad V' := L'^t.$$

The intersection matrix on the whole space $\mathbb{Z}^{\mathcal{B}}$ is finally $(\log(\#k(v)) \cdot \mathcal{M})$ for

$$\mathcal{M} = \begin{pmatrix} -s & L & L & \cdots & L & 0 \\ V & \mathcal{M}_0 & 0 & \cdots & 0 & V' \\ V & 0 & \mathcal{M}_0 & \cdots & 0 & V' \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ V & 0 & 0 & \cdots & \mathcal{M}_0 & V' \\ 0 & L' & L' & \cdots & L' & -s \end{pmatrix}. \tag{18}$$

(This has to be modified in the obvious way when $e_v = 1$.)

2B3. *Winding quotients, their dimension.* We denote as usual the jacobian of $X_0(p)_{\mathbb{Q}}$ by $J_0(p)$. As follows from Section 2B2, $\mathcal{X}_0(p)$ is semistable over $\mathbb{Z}$ and the neutral component of the Néron model $\mathcal{J}_0(p)$ of $J_0(p)$ is a semiabelian scheme over $\mathbb{Z}$ (and an abelian scheme over $\mathbb{Z}[1/p]$). Its neutral component represents the neutral component $\mathrm{Pic}^0_{\mathbb{Z}}(\mathcal{X}_0(p))$ of the relative Picard functor of $\mathcal{X}_0(p)$ over $\mathbb{Z}$.

We know from Shimura's theory that the natural decomposition of cotangent spaces into Hecke eigenspaces induces a corresponding decomposition over $\mathbb{Q}$ of abelian varieties up to isogenies:

$$J_0(p) \sim \prod_{f \in B_2 / \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})} J_f \tag{19}$$

indexed by Galois orbits in some set B_2 of newforms. A first useful sorting of this decomposition comes from the sign of the functional equations for the L-functions of eigenforms f, that is, whether $w_p(f)$ equals f or $-f$. One accordingly writes $J_0(p)^-$ for the optimal quotient abelian variety associated with $\prod_{f, w_p(f)=-f} J_f$ in (19), and similarly $J_0(p)^+$, so that $J_0(p)^- = J_0(p)/(1 + w_p)J_0(p)$ and $J_0(p)^+ = J_0(p)/(1 - w_p)J_0(p)$. One knows that

$$\dim J_0(p)^- = \left(\tfrac{1}{2} + o(1) \right) \dim J_0(p)$$

(see, e.g., [Royer 2001, Lemme 3.2]).

A more subtle object is the *winding quotient* J_e, defined as the optimal quotient of $J_0(p)$ corresponding to $\prod_{f,L(f,1)\neq 0} J_f$ in decomposition (19). One can write

$$J_e = J_0(p)/I_e J_0(p) \tag{20}$$

for some ideal I_e of the Hecke algebra $\mathbb{T}_{\Gamma_0(p)}$. Similarly, $J_e^\perp = J_0(p)/I_e^\perp J_0(p)$ will denote the optimal quotient corresponding to $\prod_{f,L(f,1)=0} J_f$. For obvious reasons regarding signs of functional equations, J_e is contained in $J_0(p)^-$. But more is expected: in line with the principle that "the vanishing order of a (modular) L functions at the critical point should generically be as small as allowed by parity", Brumer [1995] conjectured that, as p tends to infinity,

$$(?) \qquad \dim J_e = (1 - o(1)) \dim J_0(p)^-. \qquad \text{(Brumer)} \tag{21}$$

Equivalently, it is conjectured that $\dim J_e = (\frac{1}{2} + o(1)) \dim J_0(p)$, or that the dimensions of J_e and $J_e^\perp$ should be, asymptotically in p, of equal size. Note that (21) above is also implied by the "density conjecture" of [Iwaniec et al. 2000], p. 56 et seq., see also Remark F on p. 65.[4] Actually, what we eventually need in this article (see Section 7) is a weaker form of (21), which is

$$(?) \qquad \dim J_e > \frac{\dim J_0(p)}{3} + \frac{2}{3} \tag{22}$$

for large enough p. An important theorem of Iwaniec and Sarnak [2000, Corollary 13] and Kowalski, Michel and Vanderkam [Kowalski et al. 2000] asserts something nearly as good, namely

$$\left(\tfrac{1}{4} - o(1)\right) \dim J_0(p) \leq \dim J_e \left(\leq \left(\tfrac{1}{2} + o(1)\right) \dim J_0(p)\right) \tag{23}$$

as p goes to infinity $\left(\text{so that } \left(\tfrac{1}{2} - o(1)\right) \dim J_0(p) \leq \dim J_e^\perp \leq \left(\tfrac{3}{4} + o(1)\right) \dim J_0(p)\right)$. Breaking that $\frac{1}{4}$ is known to be closely linked to the Landau–Siegel zero problem. Assuming the generalized Riemann hypothesis for L-functions of modular forms, Iwaniec, Luo and Sarnak [2000, Corollary 1.6, (1.54)] prove one can improve $\frac{1}{4}$ to $\frac{9}{32}$. That seems to be all for the moment.

The central object of this paper will eventually be the maps

$$X_0(p)^{(d)} \to J_0(p) \to J_e$$

from symmetric products of $X_0(p)$ (mainly the curve itself and its square) to the winding quotient.

3. Arithmetic Chow group of modular curves

We now give a description of the Arakelov geometry of $X_0(p)$, relying on the work of many people; that topic has been pioneered by Abbes and Ullmo [1995], Michel and Ullmo [1998] and Ullmo [2000] and notably developed by Edixhoven and Couveignes [2011] and their coauthors. We shall also use the work

---

[4]Quoting Olga Balkanova (private communication), *"Theorem 1.1 in [Iwaniec et al. 2000] is proved for the test function ϕ, whose Fourier transform is supported on the interval $[-2, 2]$. The density conjecture claims that the same results are true without restriction on Fourier transform of ϕ; see formula 1.9 of [loc. cit.]."*

of Bruin [2014], Jorgenson and Kramer [2006] and Menares [2008; 2011] among others. We refer to those articles for general facts on Arakelov theory (see [Chinburg 1986; Edixhoven and de Jong 2011c]).

Let $\mathcal{X}$ be any regular and proper arithmetic surface over the integer ring $\mathcal{O}_K$ of a number field K. Fixing in general smooth hermitian metrics μ on the base changes of $\mathcal{X}$ to $\mathbb{C}$, it follows from the basics of Arakelov theory that for any horizontal divisor D on $\mathcal{X}$ over $\mathcal{O}_K$ there are Green functions $g_{\mu,D}$ on each Archimedean completion $\mathcal{X}(\mathbb{C})$ satisfying the differential equation

$$\Delta g_{\mu,D} = -\delta_D + \deg(D)\mu$$

for $\Delta = 1/(i\pi)\partial\overline{\partial}$ the Laplace operator and δ_D the Dirac distribution relative to $D_{\mathbb{C}}$ on $\mathcal{X}(\mathbb{C})$. The function $g_{\mu,D}$ is integrable on the compact Riemann surface $\mathcal{X}(\mathbb{C})$ endowed with its measure μ, and uniquely determined up to an additive constant which is often fixed by imposing the normalizing condition that

$$\int_{\mathcal{X}(\mathbb{C})} g_{\mu,D}\mu = 0. \tag{24}$$

When the horizontal divisor D is a section P_0 in $\mathcal{X}(\mathcal{O}_K)$, one will sometimes also use the notation $g_\mu(P_0, z)$ for $g_{\mu,P_0}(z)$. The Green functions relative to fixed smooth $(1,1)$-forms μ allows one to define an Arakelov intersection product relative to the μ, which will be denoted by $[\cdot, \cdot]_\mu$ or $[\cdot, \cdot]$ if there is no ambiguity about the implicit form. In particular the index will often be dropped for divisors intersections of which one at least is vertical, where the choice of μ does not intervene.

We shall denote by μ_0 the canonical Arakelov $(1,1)$-form on the Riemann surface $\mathcal{X}(\mathbb{C})$ (assumed to have positive genus), inducing the "flat metric". It corresponds to the pullback, by any Albanese morphism $\mathcal{X}(\mathbb{C}) \to \mathrm{Jac}(\mathcal{X}_K)(\mathbb{C})$, of the "cubist" metric in the sense of Moret-Bailly [1985a] (more about this shortly) on the jacobian $\mathrm{Jac}(\mathcal{X}_K)$, associated with the Néron–Tate normalized height h_Θ.

We now specialize to the case of $\mathcal{X}_0(p)$ as in Section 2B. If f is a modular form of weight 2 for $\Gamma_0(p)$, let $\|f\|^2$ be its Petersson norm. Because newforms are orthogonal in prime level we have, as in (11),

$$\mu_0 := \frac{i}{2 \dim(J_0(p))} \sum_{f \in B_2} \frac{f\frac{dq}{q} \wedge \overline{f\frac{dq}{q}}}{\|f\|^2}. \tag{25}$$

We shall also need to consider Néron–Tate heights h_A for subabelian varieties $A \hookrightarrow J_0(p)$ as in Section 2A2 (recall $A \neq 0$). The associated $(1,1)$-form μ_A is given by (12). More specifically, we focus on h_{Θ_e} on J_e (as in (7) and around, for $A' = J_e$) which induces a height $h_{\Theta_e} \circ \iota_{e,P_0}$ on $X_0(p)$ via the map $\iota_{e,P_0} : X_0(p) \hookrightarrow J \twoheadrightarrow J_e$. The curvature form of the hermitian sheaf on $X_0(p)$ defining the Arakelov height associated with $h_{\Theta_e} \circ \iota_{e,P_0}$ is

$$\mu_e := \frac{i}{2 \dim(J_e)} \sum_{f \in B_2[I_e]} \frac{f\frac{dq}{q} \wedge \overline{f\frac{dq}{q}}}{\|f\|^2}, \tag{26}$$

where $B_2[I_e]$ stands for the set of newforms killed by the ideal I_e defining J_e as in (20).

Remark 3.1. Notice that both μ_0 and μ_e, or any μ_A above, are invariant by pull-back w_p^* by the Fricke involution. In particular the Arakelov intersection products $[\cdot, \cdot]_{\mu_0}$ and $[\cdot, \cdot]_{\mu_e}$, relative to μ_0 and μ_e respectively, are w_p-invariant. The latter was clear already from the fact that, more generally, w_p is an orthogonal symmetry on $J_0(p)$ endowed with its quadratic form h_Θ, which respects the orthogonal decomposition $\prod_f J_f$ of (19).

One can now specialize the Hodge index theorem to our modular setting (see [Menares 2011, Theorem 4.16; 2008, Theorem 3.26] or more generally [Moret-Bailly 1985a, p. 85 et seq.]).

Theorem 3.2. *Let K be a number field, μ be a smooth nonzero $(1, 1)$-form on $X_0(p)(\mathbb{C})$ as given in (12), and $\widehat{\mathrm{CH}}(p)_{\mathbb{R},\mu}^{\mathrm{num}}$ be the arithmetic Chow group with real coefficients up to numerical equivalence of $\mathcal{X}_0(p)$ over $\mathcal{O}_K$, relative to μ. Denote by ∞ the horizontal divisor defined by the ∞-cusp on $\mathcal{X}_0(p)$ over $\mathbb{Z}$ (which is the Zariski closure of the $\mathbb{Q}$-point ∞ in $X_0(p)(\mathbb{Q})$), compactified with the normalizing condition (24). Write $\mathbb{R} \cdot X_\infty$ for the line of divisors with real coefficients supported on some fixed full vertical fiber X_∞. Define, for all $v \in \mathrm{Spec}(\mathcal{O}_K)$ above p, the $\mathbb{R}$-vector space*

$$G_v := \bigoplus_{C \neq C_\infty} \mathbb{R} \cdot C$$

where the sum runs through all the irreducible components of $\mathcal{X}_0(p) \times_{\mathcal{O}_K} k(v)$ except C_∞ (the one containing $\infty(k(v))$). Identify finally $J_0(p)(K)/$ torsion with the subgroup of divisor classes D_0 which are compactified under the normalizing condition $g_{D_0}(\infty) = 0$ (which is therefore different *from (24)). One has a decomposition:*

$$\widehat{\mathrm{CH}}(p)_{\mathbb{R},\mu}^{\mathrm{num}} = (\mathbb{R} \cdot \infty \oplus \mathbb{R} \cdot X_\infty) \oplus_{v \mid p}^\perp G_v \oplus^\perp (J_0(p)(K) \otimes \mathbb{R}) \qquad (27)$$

where the "$\oplus^\perp$" means that the direct factors are mutually orthogonal with respect to the Arakelov intersection product. Moreover, the restriction of the self-intersection product to $J_0(p)(K) \otimes \mathbb{R}$ coincides with twice the opposite of the Néron–Tate pairing.

Proof. The proof can be immediately adapted from that of [Menares 2011, Theorem 4.16] for L_2^1-admissible measures (a setting allowing to define convenient actions of the Hecke algebra on the Chow group). For further computational use we recall how one decomposes divisors in practice. Take D in $\widehat{\mathrm{CH}}(p)_{\mathbb{R},\mu}^{\mathrm{num}}$, with degree d on the generic fiber. There is a vertical divisor Φ_D, with support in fibers above places of bad reduction (that is, of characteristic p), such that $(D - d\infty - \Phi_D)$ has a real multiple which belongs to the neutral component $\mathrm{Pic}^0(\mathcal{J}_0(p))_{/\mathcal{O}_K}$. That Φ_D is well-defined up to multiple of full vertical fibers, so we can assume Φ_D belongs to $\oplus^\perp G_p$ (and is then unambiguously defined). One associates to $(D - d\infty - \Phi_D) \in \mathbb{R} \cdot \mathcal{J}_0^0(p)(\mathcal{O}_K)$ an element δ in $\widehat{\mathrm{CH}}(p)_{\mathbb{R},\mu}^{\mathrm{num}}$ by imposing a compactification such that $[\infty, \delta]_\mu = 0$. The general Hodge index theorem (see for instance [Moret-Bailly 1985a]) then finally asserts that $(D - d\infty - \Phi_D - \delta)$ can be written as an element in $\mathbb{R} \cdot X_\infty$. $\qquad \square$

In order to later on interpret the Néron–Tate height (associated with some given (symmetric) invertible sheaf) as an Arakelov height in a suitable sense (see [Abbes 1997] paragraph 3, or [Moret-Bailly 1985b]),

we will need to compute explicitly, given $P \in X_0(p)(K)$, the vertical divisor $\Phi_P = \oplus_{v \mid p} \Phi_{P,v}$ such that

$$[C, P - \infty - \Phi_P] = 0 \tag{28}$$

for any irreducible component of any fiber of $\mathcal{X}_0(p) \to \operatorname{Spec}(\mathcal{O}_K)$, as in the proof of Theorem 3.2.

Lemma 3.3. *Consider a bad fiber $\mathcal{X}_0(p)_{k(v)}$, with e_v the absolute ramification index of v, and write $\#k(v) = p^{f_v}$. Let $P \in X_0(p)(K)$ and let $C_{P,v}$ be the irreducible component of $\mathcal{X}_0(p)_{k(v)}$ which contains $P(k(v))$. As $\mathcal{X}_0(p)$ is assumed to be regular, the section P hits each fiber on its smooth locus, so that the component P belongs to is unambiguously defined in each bad fiber. Write*

$$\Phi_{P,v} = \sum_{n,m} a_{n,m}[C_{n,m}]$$

with notations as in (16). Recall that, by our convention, $a_{C_\infty} = a_{*,0} = 0$.

(a) *If $C_{P,v} = C_0$ then for all n and m,*

$$a_{n,m} = \frac{-12}{(p-1) \cdot w_n} \cdot m.$$

   (Recall (see (13)) that $w_n := \#\operatorname{Aut}(S(n))/\langle \pm 1 \rangle \in \{1, 2, 3\}$, with $S(n)$ the supersingular point corresponding to the branch $\{C_{n,.}\}$.)

   For further use we henceforth write Φ_{C_0} for the above vector $\Phi_{P,v} \in \mathbb{Z}^{\mathcal{B}}$.

(b) *If $C_{P,v} = C_{n_0,m_0} \neq C_0, C_\infty$ then:*

   • *For $n = n_0$ and $m \in \{0, m_0\}$, one has*

$$a_{n,m} = \left(\frac{m_0}{w_{n_0} e_v} \left(1 - \frac{12}{(p-1)w_{n_0}} \right) - 1 \right) \cdot m.$$

   • *For $n = n_0$ and $m \in \{m_0, w_{n_0} e_v\}$, one has*

$$a_{n,m} = \left(\frac{m_0}{w_{n_0} e_v} \left(1 - \frac{12}{(p-1)w_{n_0}} \right) \right) \cdot m - m_0.$$

   • *For $n \neq n_0$ and all $m \in \{0, w_n e_v\}$, one has*

$$a_{n,m} = \frac{-12 m_0}{(p-1)w_{n_0} e_v} \cdot \frac{m}{w_n}.$$

(c) *Of course if $C_{P,v} = C_\infty$ then $\Phi_{P,v} = 0$.*

Remark 3.4. We have distinguished different cases above because the proof naturally leads to doing so, and it will be of interest below to have the simpler case (a) explicitly displayed. Note however that all outputs are actually covered by the formulae of case (b). Notice also that, in case (a), all coefficients of $\Phi_{P,v}$ satisfy

$$0 \geq a_{n,m} \geq a_0 := a_{C_0} = a_{n,w_n m} = \frac{-12 e_v}{(p-1)}.$$

As for case (b), all coefficients of $\Phi_{P,v}$ satisfy

$$0 \geq a_{n,m} \geq a_{n_0,m_0} = \left(\frac{m_0}{w_{n_0} e_v} \left(1 - \frac{12}{(p-1) w_{n_0}} \right) - 1 \right) \cdot m_0$$

(remember $0 \leq m \leq w_n e_v$ for all m). Computing the minimum of the above right-hand as a polynomial in m_0 gives

$$0 \geq a_{n,m} \geq \frac{-e_v w_{n_0}}{4 \left(1 - \frac{12}{(p-1) w_{n_0}} \right)} \geq \frac{-e_v w_{n_0}}{4 - \frac{3}{w_{n_0}}} \geq -3 e_v \tag{29}$$

(recalling we always assume $p \geq 17$).

Proof. Given the intersection matrix (18) and condition (28), $[C, P - \infty - \Phi_{P,v}] = 0$ for all C in the fiber at v, gives the matrix equation

$$\log(\#k(v)) \mathcal{M} \cdot \Phi_{P,v} = \log(\#k(v))(-1, 0, \ldots, 1, 0, \ldots, 0)^t \tag{30}$$

where the coefficient 1 (respectively -1) in the right-hand column vector is at the place corresponding to $C_{P,v} = C_{n,m}$ (respectively to $C_\infty = C_{n,0}$) in the ordering of our component basis (16). That is however more easily solved by running through the dual graph of $\mathcal{X}_0(p)_{k(v)}$ "branch by branch" as follows. Suppose first that $C_{P,v} = C_0$, and recall $a_{C_\infty} = 0$ by convention. Equation (28) translates into:

- $-1 - \sum_{n=1}^{s} a_{n,1} = 0$, for $C = C_\infty$.
- $1 + s a_0 - \sum_{n=1}^{s} a_{n, w_n e_v - 1} = 0$, for $C = C_0$.
- $a_{n,m-1} - 2 a_{n,m} + a_{n,m+1} = 0$, for all others $C = C_{n,m}$.

The equations of the third line in turn define, for each branch (that is, for fixed n), a sequence defined by linear double induction with solution $a_{n,m} = m \cdot \alpha_n$ for some α_n which is easily computed to be $-1/(w(\text{Eis}) \cdot w_n) = -12/((p-1) w_n)$ (see (14)). (Note this is true even for $e_v = 1$.)

For case (b), the intersection equations become:

- $-1 - \sum_{n=1}^{s} a_{n,1} = 0$, for $C = C_\infty$.
- $s a_0 - \sum_{n=1}^{s} a_{n, w_n e_v - 1} = 0$, for $C = C_0$.
- $1 - a_{n_0,m_0-1} + 2 a_{n_0,m_0} - a_{n_0,m_0+1} = 0$, for $C = C_{P,v} = C_{n_0,m_0}$.
- $a_{n,m-1} - 2 a_{n,m} + a_{n,m+1} = 0$, for all others $C = C_{n,m}$.

As above, solving these equations in all branches not containing $C_{P,v}$ gives $a_{n,m} = m \beta_n$ and the same is true in the branch containing $C_{P,v}$ for $m \in \{0, \ldots, m_0\}$. We also see that $a_{n_0,m_0+1} = (m_0 + 1) \beta_{n_0} + 1$, and then $a_{n_0,m} = m(\beta_{n_0} + 1) - m_0$ for $m \in \{m_0 + 1, w_n e_v\}$. We have $a_0 = w_n e_v \beta_n$ for all $n \neq n_0$, so let β be the common value of the β_n for $n \neq n_0$ with $w_n = 1$. (There is always such an n as we assumed $p > 13$. Note also those computations still cover the case $e_v = 1$.) From $\beta = a_0/e_v$ and $a_0 = w_{n_0} e_v (\beta_{n_0} + 1) - m_0$ we derive

$$\beta_{n_0} = \frac{a_0 + m_0 - w_{n_0} e_v}{w_{n_0} e_v} = \frac{\beta}{w_{n_0}} + \frac{m_0}{w_{n_0} e_v} - 1.$$

Hence, because of the first equation $\left(-1 - \sum_{n=1}^{s} a_{n,1} = 0\right)$,

$$0 = -1 - \beta_{n_0} - \sum_{1 \leq n \leq s, n \neq n_0} \frac{\beta}{w_n} = -\beta w(\text{Eis}) - \frac{m_0}{w_{n_0} e_v}$$

so that

$$\beta = \frac{-m_0}{w(\text{Eis}) w_{n_0} e_v} = \frac{-12\, m_0}{(p-1) w_{n_0} e_v}. \qquad \square$$

Lemma 3.5. *Let μ be some $(1,1)$-form on $X_0(p)(\mathbb{C})$ as in Theorem 3.2.*

(a) *The class in $\widehat{\text{CH}}(p)_{\mathbb{R},\mu}^{\text{num}}$ of the cuspidal divisor $(0) - (\infty)$ satisfies*

$$(0) - (\infty) \equiv \Phi_{C_0}^0 := \Phi_{C_0} + \sum_{v \mid p} \frac{6 e_v}{p-1} \left(\sum_C [C] \right) = \sum_{v \mid p} \sum_{n,m} \frac{6}{(p-1)} \left(e_v - \frac{2m}{w_n} \right) [C_{n,m}] \qquad (31)$$

with notations as in Lemma 3.3 (a). This is an eigenvector of the Fricke $\mathbb{Z}$-automorphism w_p with eigenvalue -1.

(b) *One has $[\infty, \infty]_\mu = [0, 0]_\mu = [0, \infty]_\mu - 6 \log p / (p-1)$. If μ is the Green–Arakelov measure μ_0 then $0 \geq [\infty, \infty]_{\mu_0} = O(\log p / p)$ and similarly $[0, \infty]_{\mu_0} = O(\log p / p)$ with $[0, \infty]_{\mu_0}$ nonpositive too, at least for large enough p. If $\mu = \mu_e$ (see (26))—or more generally any submeasure of μ_0— then $[0, \infty]_{\mu_e} = O(p \log p)$.*

Proof. By the Manin–Drinfeld theorem, $(0) - (\infty)$ is torsion as a divisor in the generic fiber $\mathcal{X}_0(p) \times_{\mathbb{Z}} \mathbb{Q}$. One therefore has

$$(0) - (\infty) \equiv \Phi + c X_\infty$$

in the decomposition (27) of $\widehat{\text{CH}}(p)_{\mathbb{R},\mu}^{\text{num}}$, for Φ some vertical divisor with support in the fibers above p. This divisor is determined by the same equations (28) as Φ_{C_0} in Lemma 3.3(a). For each $v \mid p$ the full v-fiber $\sum_C [C]$ is numerically equivalent to some real multiple of the archimedean fiber X_∞; there is therefore a real number a such that

$$\Phi_{C_0}^0 := \Phi_{C_0} + \sum_{v \mid p} \frac{6 e_v}{p-1} \left(\sum_C [C] \right) \equiv \Phi_{C_0} + a X_\infty.$$

Now w_p switches the cusps 0 and ∞ so the divisor $(0) - (\infty)$ is antisymmetric for w_p:

$$w_p^*((0) - (\infty)) = -((0) - (\infty))$$

and clearly $w_p^*(\Phi_{C_0}^0) = -\Phi_{C_0}^0$. The fact that w_p preserves the archimedean fiber concludes the proof of (a).

To prove (b) we compute

$$0 = [0 - \infty - \Phi_{C_0}^0, \infty]_\mu = [0, \infty]_\mu - [\infty, \infty]_\mu - \frac{6}{p-1} \log p$$

and

$$0 = [0 - \infty - \Phi_{C_0}^0, 0]_\mu = [0, 0]_\mu - [0, \infty]_\mu + \frac{6}{p-1} \log p$$

so that $[\infty, \infty]_\mu = [0, 0]_\mu = [0, \infty]_\mu - 6 \log p/(p-1)$. The cusps 0 and ∞ are known not to intersect on $\mathcal{X}_0(p)_{/\mathbb{Z}}$ so that $[0, \infty]_\mu = -g_\mu(0, \infty)$. When $\mu = \mu_0$, this special value of the Arakelov–Green function has been computed by Michel and Ullmo; it satisfies, by [Michel and Ullmo 1998, (12), p. 650],

$$g_{\mu_0}(0, \infty) = \frac{1}{2g} \log p \left(1 + O\left(\frac{\log \log p}{\log p} \right) \right) = O\left(\frac{\log p}{p} \right).$$

Finally, using [Bruin 2014, Theorem 7.1(c) and paragraph 8] and plugging into Bruin's method the estimates of [Michel and Ullmo 1998] regarding the comparison function $F(z) = O((\log p)/p)$ between Green–Arakelov and Poincaré measures, we obtain a bound of shape $O(p \log p)$ for $|g_{\mu_e}(0, \infty)|$ (see also Remark 4.5). This completes the proof of (b). □

Instrumental in the sequel will be the explicit decomposition of the relative dualizing sheaf ω in the arithmetic Chow group.

Proposition 3.6. *The relative dualizing sheaf ω of the minimal regular model $\mathcal{X}_0(p) \to \mathcal{O}_K$ can be written, in the decomposition (27) of $\widehat{\mathrm{CH}}(p)_{\mathbb{R},\mu_0}^{\mathrm{num}}$ relative to the canonical Green–Arakelov $(1,1)$-form μ_0, as*

$$\omega = (2g - 2)\infty + \sum_{v \mid p} \Phi_{\omega,v} + \omega^0 + [K : \mathbb{Q}]c_\omega X_\infty, \tag{32}$$

where the above components satisfy the following properties:

- *The number c_ω is equal to $\frac{(1-2g)}{[K:\mathbb{Q}]}[\infty, \infty]_{\mu_0}$, so that $0 \leq c_\omega \leq O(\log p)$.*

- *Set*

$$H_4 := \frac{1}{2} \sum_{P \in \mathcal{H}_4} \left(P - \frac{1}{2}(0 + \infty) \right), \quad H_3 := \frac{2}{3} \sum_{p \in \mathcal{H}_3} \left(P - \frac{1}{2}(0 + \infty) \right)$$

 where the sums run over the sets $\mathcal{H}_4$ and $\mathcal{H}_3$, whose number of elements can be 0 or 2, of Heegner points of $X_0(p)$ with j-invariant 1728 and 0 respectively. Define

$$H_4^0 := H_4 + [K : \mathbb{Q}]c_4 X_\infty \quad and \quad H_3^0 := H_3 + [K : \mathbb{Q}]c_3 X_\infty$$

 for two numbers c_3 and c_4 with $c_3 = O(\log p)$, and the same for c_4. (Recall this means the H_* are compactified with the normalizing condition (24), whereas the H_*^0 are the orthogonal projections on $(J_0(p)(K) \otimes \mathbb{R}) \subseteq \widehat{\mathrm{CH}}(p)_{\mathbb{R},\mu_0}^{\mathrm{num}}$ of the H_*, so that $[\infty, H_*^0]_{\mu_0} = 0$, for $* = 3$ or 4.) One sets $\omega^0 := -H_4^0 - H_3^0$, which can be chosen in $J_0(p)^0(\overline{\mathbb{Q}})$.

- *Finally, the component $\Phi_{\omega,v}$ in each G_v for $v \mid p$ is*

$$\Phi_{\omega,v} = -12 \frac{(g-1)}{(p-1)} \sum_{n,m} \frac{m}{w_n} C_{n,m} \tag{33}$$

 with notations as in (16). We therefore have $\Phi_{\omega,v} = (g-1)\Phi_{C_0}$ using notations of Lemma 3.3. In particular, recalling e_v is the ramification index of $K/\mathbb{Q}$ at v, the coefficients $\omega_{n,m}$ of $\Phi_{\omega,v}$ in (33) satisfy

$$0 \geq \omega_{n,m} \geq -e_v. \tag{34}$$

Proof. Many parts of those statements are deduced from [Michel and Ullmo 1998, Section 6] and results of Edixhoven and de Jong [2011b]. See also [Menares 2011, Section 4.4].

We start by estimating c_ω. By Arakelov's adjunction formula,

$$-[\infty, \infty]_{\mu_0} = [\infty, \omega]_{\mu_0} = (2g - 2)[\infty, \infty]_{\mu_0} + [K : \mathbb{Q}]c_\omega$$

because of the orthogonality of the decomposition (27). Lemma 3.5 therefore implies

$$0 \le c_\omega = \frac{(1 - 2g)}{[K : \mathbb{Q}]}[\infty, \infty]_{\mu_0} = O(\log p).$$

The computations of the $J_0(p)$-part $\omega^0 := -(H_3^0 + H_4^0)$ follows from the Hurwitz formula, as explained in [Michel and Ullmo 1998, paragraph 6, p. 670]. One indeed checks that, on the generic fiber $X_0(p)_{/\mathbb{Q}} = \mathcal{X}_0(p) \times_{\mathbb{Z}} \mathbb{Q}$, the canonical divisor is linearly equivalent to

$$(2g - 2)\infty - \left(\frac{1}{2} \sum_{j(P)=e^{i\pi/2}}{}'(P - \infty) + \frac{2}{3} \sum_{j(P)=e^{2i\pi/3}}{}'(P - \infty) \right)$$

where the sums $\sum'$ are here restricted to points P at which $X_0(p) \to X(1)$ is unramified (these are the Heegner points alluded to in our statement). It follows from the modular interpretation that in each of those sums there are two Heegner points (if any), which are then ordinary at p (recall we assume $p > 13 > 3$). This proves that the $J_0(p)(K) \otimes_{\mathbb{Z}} \mathbb{R}$-part of ω is indeed $-(H_4^0 + H_3^0)$ with $H_4^0 = H_4 + [K : \mathbb{Q}]c_4 X_\infty$ and $H_3^0 = H_3 + [K : \mathbb{Q}]c_3 X_\infty$ for some real numbers c_3 and c_4. (Note that, as Heegner points are preserved by the Atkin–Lehner involution [Gross 1984, paragraph 5, p. 90] their specializations above p share themselves between the two components C_0 and C_∞ of $\mathcal{X}_0(p)_{/\mathbb{F}_p}$, so that $2H_3^0 = \sum_{j(P)=e^{i\pi/2}}{}'(P - \infty)$ and $\frac{2}{3}H_4^0 = \sum_{j(P)=e^{2i\pi/3}}{}'(P - \infty)$ belong to the neutral component $J_0(p)^0(\mathcal{O}_K)$.) The estimates on c_3 and c_4 will be justified at the end of the proof.

The bad fibers divisors $\Phi_{\omega,v} := \sum_{n,m} \omega_{n,m}[C_{n,m}]$ can be computed with the "vertical" adjunction formula [Liu 2002, Chapter 9, Theorem 1.37] as in [Menares 2011, Lemma 4.22]. Indeed, for each irreducible component C in the v-fiber having genus 0, one has

$$[C, C + \omega] = -2 \log(\#k(v)).$$

If $\mathcal{M}$ is the intersection matrix displayed in (18), and $\delta_{*,*}$ is Kronecker's symbol, we therefore have

$$C \cdot \mathcal{M} \cdot \Phi_{\omega,v} = -2 - \frac{1}{\log(\#k(v))}[C, C] - (2g - 2)\delta_{C,C_\infty} = \begin{cases} 0 & \text{if } C \ne C_\infty, C_0, \\ s - 2g & \text{if } C = C_\infty, \\ s - 2 & \text{if } C = C_0, \end{cases} \tag{35}$$

that is, as $s = g + 1$,

$$\mathcal{M} \cdot \Phi_{\omega,v} = (g - 1)(-1, 0, \ldots, 0, 1)^t.$$

That equation is (30) (up to a multiplicative scalar), which has been solved in the first case of Lemma 3.3. Therefore

$$\Phi_{\omega,v} = (g-1)\Phi_{C_0}, \quad \text{that is } \omega_{n,m} = \frac{12(1-g)}{(p-1)} \cdot \frac{m}{w_n}. \tag{36}$$

As noted in Remark 3.4 and using (15), this implies the coefficients $\omega_{n,m}$ of $\Phi_{\omega,v}$ satisfy

$$0 \geq \omega_{n,m} \geq \frac{12(1-g)}{p-1} e_v > -e_v.$$

We finally estimate the intersection products

$$c_3 = \frac{-1}{[K:\mathbb{Q}]}[\infty, H_3]_{\mu_0} \quad \text{and} \quad c_4 = \frac{-1}{[K:\mathbb{Q}]}[\infty, H_4]_{\mu_0}.$$

By the adjunction formula and Hriljac–Faltings' theorem [Chinburg 1986, Theorem 5.1(ii)] we compute that for any $P \in X_0(p)(K)$,

$$-2[K:\mathbb{Q}]\mathrm{h}_\Theta(P - \frac{1}{2g-2}\omega) = \left[P - \frac{1}{2g-2}\omega - \Phi_\omega(P), P - \frac{1}{2g-2}\omega - \Phi_\omega(P)\right]_{\mu_0}$$

$$= \frac{1}{(2g-2)^2}[\omega,\omega]_{\mu_0} + \frac{g}{g-1}[P,P]_{\mu_0} - \Phi_\omega(P)^2$$

where here $\Phi_\omega(P)$ is a vertical divisor supported at bad fibers such that

$$\left[C, P - \frac{1}{2g-2}\omega - \Phi_\omega(P)\right] = 0 \tag{37}$$

for any irreducible component C of any bad fiber of $\mathcal{X}_0(p)_{/\mathcal{O}_K}$. Hence

$$\frac{1}{(2g-2)^2}\omega^2 + \frac{g}{g-1}[P,P]_{\mu_0} - \Phi_\omega(P)^2 = -2[K:\mathbb{Q}]\mathrm{h}_\Theta((P-\infty) + \frac{1}{2g-2}(H_3 + H_4)). \tag{38}$$

We specialize to the case when $P = P_*^*$ (where the upper star is 1 or 2 and the lower star is 4 or 3) is one of the Heegner points occurring in H_4 or H_3, respectively. We replace for now the base field K by $F := \mathbb{Q}(P_*^*) = \mathbb{Q}(\sqrt{-1})$ (respectively, $\mathbb{Q}(\sqrt{-3})$). The right-hand of (38), if nonzero, is then

$$-8\log(p)(1 + o(1)) \quad \text{or} \quad -12\log(p)(1 + o(1)), \quad \text{respectively,} \tag{39}$$

by [Michel and Ullmo 1998, p. 673]. If those Heegner points occur we know that p splits in F, so there are two bad primes v and v' on $\mathcal{O}_F$ (therefore two bad fibers on $\mathcal{X}_0(p)_{/\mathcal{O}_F}$ and two G_v, $G_{v'}$) to take into account. We compute $\Phi_\omega(P_*^*)$ and $\Phi_\omega(P_*^*)^2$. As mentioned at the beginning of the proof, P_*^* specializes to the component C_0 at a place, say v, of F above p, and to C_∞ at the conjugate place v'. Condition (37) therefore gives that, for any irreducible component C of the fiber at v,

$$0 = \left[C, P_*^* - \frac{1}{2g-2}\omega - \Phi_\omega(P_*^*)_v\right] = \left[C, 0 - \infty - \frac{1}{2g-2}\Phi_{\omega,v} - \Phi_\omega(P_*^*)_v\right]$$

and using Lemma 3.3, 3.5 and (36) one obtains

$$\Phi_\omega(P_*^*)_v = -\frac{1}{2g-2}\Phi_{\omega,v} + \Phi_{C_0,v} = \tfrac{1}{2}\Phi_{C_0,v}$$

whereas, at v'

$$\Phi_\omega(P_*^*)_{v'} = -\frac{1}{2g-2}\Phi_{\omega,v'} = -\tfrac{1}{2}\Phi_{C_0,v'}.$$

Using Lemma 3.3 and 3.5 again we therefore have

$$\Phi_\omega(P_*^*)^2 = \sum_{w\mid p} \tfrac{1}{4}\Phi_{C_0,w}^2 = \sum_{w\mid p}\tfrac{1}{4}[\Phi_{C_0,w}, 0-\infty] = \tfrac{1}{2}a_0\log p = -\frac{6\log(p)}{p-1}. \tag{40}$$

As for the self-intersection of ω one knows from [Ullmo 2000, Introduction] that

$$\omega_{\mathcal{X}_0(p)/\mathbb{Z}}^2 = 3g\log(p)(1+o(1)).$$

As the quantity $\frac{1}{[F:K]}[\omega]^2$ is known to be independent from the number field extension F/K, the dualizing sheaf $\omega_{\mathcal{X}_0(p)/\mathcal{O}_F}$ of $\mathcal{X}_0(p)$ over $\mathcal{O}_F$ (instead of $\mathbb{Z}$) satisfies $\omega^2 = 6g\log(p)(1+o(1))$. Summing-up, (38) implies that

$$[P_*^*, P_*^*]_{\mu_0} = O(\log(p)) \tag{41}$$

for each Heegner point P_*^*. Now, on the other hand, the vertical divisor $\Phi_{P_*^*}$ in the sense of (28) and Lemma 3.3 is $\Phi_{P_*^*} = \Phi_{C_0,v}$ for the place v of F where P_*^* specializes on C_0 and not C_∞. Therefore

$$-4h_\Theta(P_*^* - \infty) = [P_*^* - \infty - \Phi_{P_*^*}, P_*^* - \infty - \Phi_{P_*^*}]_{\mu_0}$$
$$= -2[P_*^*, \infty]_{\mu_0} + [P_*^*, P_*^*]_{\mu_0} + [\infty, \infty]_{\mu_0} - (\Phi_{P_*^*})^2 \tag{42}$$

whence, using (39), (40), (41) and 3.5(b),

$$[P_*^*, \infty]_{\mu_0} = \tfrac{1}{2}([P_*^*, P_*^*]_{\mu_0} + [\infty, \infty]_{\mu_0} - (\Phi_{C_0,v})^2 + 4h_\Theta(P_*^* - \infty)) = O(\log p).$$

Putting everything together and using 3.5 once more we conclude that

$$c_4 = -\frac{1}{[K:\mathbb{Q}]}[\infty, H_4]_{\mu_0} = \frac{1}{2[K:\mathbb{Q}]}(-[\infty, P_4^1 + P_4^2]_{\mu_0} + [\infty, 0+\infty]_{\mu_0}) = O(\log p) \tag{43}$$

and similarly for c_3. (Note that the Arakelov intersection products, in the computations around (42), were performed over $F = \mathbb{Q}(P_*^*)$ and not K, although we did not indicate this in the notation in order to keep it from becoming too heavy. We however want quantities over K for the statement of the theorem, so we need considering Arakelov products over K in (43) above.) $\qquad\square$

Remark 3.7. It may be convenient to write, with notations as in (32), a more symmetric ω as

$$\omega = (g-1)(\infty+0) + (-H_4^0 - H_3^0) + [K:\mathbb{Q}]c_\omega X_\infty \tag{44}$$

which yields an element with no vertical component at bad fibers.

4. j-height and Θ-height

In this section we compare two natural heights on $X_0(p)(\overline{\mathbb{Q}})$, namely the j-height and the one induced from the Néron–Tate Θ-height on $J_0(p)(\overline{\mathbb{Q}})$. We start with an explicit description of the latter, for which it is actually convenient to use a bit of Zhang's language [1993] about "adelic metrics" which, in our modular setting, has a very concrete form.

Using notations and results from Section 2B2 we therefore consider the limit, as e_v goes to ∞, of the dual graph of the special fiber of $\mathcal{X}_0(p)$ at a place v of a p-adic local field with ramification index e_v at p (see Figure 1). Here we normalize the length of the $s = g + 1$ edges from C_∞ to C_0 to be 1, so that the vertex $C_{n,m}$ corresponds to the point of the n-th edge with distance $m/(e_v w_n)$ from the origin C_∞. Now associate to any edge $n \in \{1, \ldots, s\}$ the quadratic polynomial function

$$g_n(x) : [0, 1] \to \mathbb{R}, \quad x \mapsto \frac{x}{2}\left(\left(w_n - \frac{12}{(p-1)}\right)x - w_n - 12\frac{(g-1)}{(p-1)}\right). \tag{45}$$

For K any number field, P in $X_0(p)(K)$, and v a place of K whose ramification degree and residual degree are still denoted by e_v and f_v respectively, let

$$G(P(K_v)) = e_v f_v \log(p) \cdot g_n(C_{P(k(v))}) \tag{46}$$

where $C_{P(k(v))}$ is the component to which the specialization of P belongs at v, identified to a point of the n-th edge where it lives.

Theorem 4.1. *For any number field K, there is an element*

$$\tilde{\omega}_{\Theta,K} = (g \cdot \infty + \Phi_{\Theta,K} + c_{\Theta,K} X_\infty) \tag{47}$$

of $\widehat{\mathrm{CH}}(p)^{\mathrm{num}}_{\mathbb{R},\mu_0}$ *such that for any $P \in X_0(p)(K)$ one has, with notations as in Proposition 3.6,*

$$h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) = \frac{1}{[K:\mathbb{Q}]}[P, \tilde{\omega}_{\Theta,K}]_{\mu_0} \tag{48}$$

and the terms of (47) *satisfy*

$$0 \geq [P, \Phi_{\Theta,K}] \geq -2[K:\mathbb{Q}]\log(p) \quad \text{and} \quad c_{\Theta,K} = [K:\mathbb{Q}]O(\log p). \tag{49}$$

Passing to the limit on all number fields, the height induced on $X_0(p)(\overline{\mathbb{Q}})$ by pulling-back Néron–Tate's Θ-height on $J_0(p)(\overline{\mathbb{Q}})$ via the embedding $P \mapsto P - \infty + \tfrac{1}{2}\omega^0$ can be written as

$$h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) = \frac{1}{[K:\mathbb{Q}]}\left(g[P, \infty]_{\mu_0} + \sum_{v \in M_K, v \mid p} G(P(K_v)) + c_{\Theta,K}\right) \tag{50}$$

where Zhang's Green function G at bad fibers is defined in (45) *and* (46).

In any case one has that the height satisfies

$$h_\Theta\left(P - \infty + \tfrac{\omega^0}{2}\right) = \frac{1}{[K:\mathbb{Q}]}[P, g \cdot \infty]_{\mu_0} + O(\log p). \tag{51}$$

Proof. We prove (48) and (49); from there reformulation (50) and (51) are straightforward.

Recall $\mathcal{X}_0(p)$ denotes the minimal regular model of $X_0(p)$ on $\mathrm{Spec}(\mathcal{O}_K)$, that $\mathcal{J}_0(p)$ is the Néron model of $J_0(p)$ on the same base, and $\mathcal{J}_0(p)^0$ stands for its neutral component. Let δ be an element of $J_0(p)(K)$, seen as a degree 0 divisor on $X_0(p)$. Up to making a base extension we can assume δ is linearly equivalent to a sum of points in $X_0(p)(K)$. We shall denote by $\tilde{\delta} = \delta + \Phi_\delta$ (for Φ_δ some vertical divisor on $\mathcal{X}_0(p)$, with multiplicity 0 on the component containing ∞, following our running conventions) the associated element of the neutral component $\mathcal{J}_0(p)^0(\mathcal{O}_K)$ (that is, the one whose associated divisor has degree zero on each irreducible component, in any fiber, of $\mathcal{X}_0(p)$, and therefore defines a point of $\mathcal{J}_0(p)^0(\mathcal{O}_K)$). For any point P in $X_0(p)(K) \hookrightarrow \mathcal{X}_0(p)(\mathcal{O}_K)$ let similarly Φ_P be the vertical divisor on $\mathcal{X}_0(p)$, with support on the bad fibers, such that $(P - \infty - \Phi_P)$ has divisor class belonging to the neutral component $\mathcal{J}_0(p)^0(\mathcal{O}_K)$ and, again, Φ_P has everywhere trivial ∞-component, see (28). Recall we can compute Φ_P explicitly by Lemma 3.3. We write $\Phi_P = \sum_{v \in M_K, v \mid p} \sum_{C_v} a_{C_v}[C_v]$ where the sum is taken on irreducible components C_v of vertical bad fibers of $\mathcal{X}_0(p)$. Using notations of Lemma 3.3 (b) we also define the following new vertical divisor at bad fibers:

$$\Phi_{\vartheta, K} := \sum_{v \in M_K, v \mid p} \sum_{Q_v} a_{C_{Q_v}} C_{Q_v} = \sum_{v \mid p} \sum_{(n_0, m_0)} a^v_{n_0, m_0} C_{n_0, m_0} \tag{52}$$

so that

$$a^v_{n_0, m_0} = \left(\frac{m_0}{w_{n_0} e_v} \left(1 - \frac{12}{(p-1)w_{n_0}} \right) - 1 \right) \cdot m_0.$$

Our very definitions imply

$$\Phi_P^2 = [P, \Phi_P] = [P, \Phi_{\vartheta, K}] \tag{53}$$

for any $P \in X_0(p)(K)$. Using Faltings' Hodge index theorem we can write the Néron–Tate height $h_\Theta(P - \infty + \delta)$ as

$$h_\Theta(P - \infty + \delta)$$
$$= \frac{-1}{2[K:\mathbb{Q}]}[P - \infty + \tilde{\delta} - \Phi_P, P - \infty + \tilde{\delta} - \Phi_P]_{\mu_0}$$
$$= \frac{1}{2[K:\mathbb{Q}]}([P, \omega + 2\infty - 2\tilde{\delta}]_{\mu_0} + 2[P, \Phi_P]_{\mu_0} - [\Phi_P, \Phi_P]_{\mu_0} + [\tilde{\delta}, 2\infty - \tilde{\delta}]_{\mu_0} - [\infty, \infty]_{\mu_0})$$
$$= \frac{1}{2[K:\mathbb{Q}]}([P, \omega + 2\infty - 2\tilde{\delta} + \Phi_{\vartheta, K}]_{\mu_0} + [\tilde{\delta}, 2\infty - \tilde{\delta}]_{\mu_0} - [\infty, \infty]_{\mu_0})$$
$$= \frac{1}{[K:\mathbb{Q}]}[P, \tilde{\omega}_\delta]_{\mu_0} \tag{54}$$

with

$$\tilde{\omega}_\delta := \left(\tfrac{1}{2}(\omega + \Phi_{\vartheta, K}) + \infty - \tilde{\delta} \right) + c_\delta X_\infty \tag{55}$$

for X_∞ some fixed archimedean fiber of $\mathcal{X}_0(p)$ and c_δ is the real number

$$c_\delta = \tfrac{1}{2}(-[\infty, \infty]_{\mu_0} + [\tilde{\delta}, 2\infty - \tilde{\delta}]_{\mu_0}). \tag{56}$$

Note that $\tilde{\omega}_\delta$ does not depend on P (as $\Phi_{\vartheta,K}$ was introduced to that aim).

Let us now take $\delta = \frac{1}{2}\omega^0 = -\frac{1}{2}(H_3 + H_4) \in \frac{1}{12} \cdot J_0(p)^0(\mathbb{Q})$, as defined in Proposition 3.6. (This is Riemann's characteristic (the "κ" of [Hindry and Silverman 2000, p. 138] for instance, that is the generic fiber of the $J_0(p)(\mathbb{Q}) \otimes \mathbb{R}$-part of ω in the decomposition (32).) Set $\Phi_{\Theta,K} := \frac{1}{2}(\Phi_\omega + \Phi_{\vartheta,K})$. Then

$$\tilde{\omega}_\Theta := \tilde{\omega}_\delta = (g \cdot \infty + \Phi_{\Theta,K} + c_{\Theta,K} X_\infty) \tag{57}$$

for $c_{\Theta,K}$ which, still using notations of Proposition 3.6 and its proof, is explicitly given by

$$\frac{1}{[K:\mathbb{Q}]} c_{\Theta,K} = \frac{1}{2}\left(c_\omega - c_4 - c_3 + \frac{1}{2}h_\Theta(H_3 + H_4) - \frac{1}{[K:\mathbb{Q}]}([\infty]_{\mu_0}^2 + [\infty, H_3 + H_4]_{\mu_0}) \right)$$
$$= \frac{1}{2}\left(c_\omega - \frac{1}{[K:\mathbb{Q}]}[\infty]_{\mu_0}^2 + \frac{1}{2}h_\Theta(H_3 + H_4) \right).$$

As in the proof of Proposition 3.6 we invoke p. 673 of [Michel and Ullmo 1998] to assert $h_\Theta(H_3 + H_4) = O(\log(p))$. We moreover know from the same proposition and from Lemma 3.5 that both $|c_\omega| = O(\log p)$ and $[\infty, \infty]_{\mu_0} = [K:\mathbb{Q}]O(\log p/p)$, so that

$$c_{\Theta,K} = [K:\mathbb{Q}]O(\log p). \tag{58}$$

The contribution of $\Phi_{\Theta,K}$ is controlled by Lemma 3.3 and Remark 3.4. On one hand

$$0 \geq [P, \Phi_{\vartheta,K}] = [P, \Phi_P] = \sum_{v \in M_K, v \mid p} a_{C_P, v} \log(\#k_v) \geq \sum_{v \in M_K, v \mid p} -3e_v \log(p^{f_v}) \geq -3[K:\mathbb{Q}]\log(p). \tag{59}$$

On the other hand, by (34), the coefficients of the vertical components $\Phi_{\omega,v}$ satisfy $0 \geq \omega_{n,m} \geq -e_v$, so writing $\omega_{n_P, m_P, v}$ for the coefficient in $\Phi_{\omega,v}$ of the component containing $P(k(v))$ we have

$$0 \geq [P, \Phi_\omega] = \sum_{v \mid p} \omega_{n_P, m_P, v} \log(\#k(v)) \geq \sum_{v \mid p} -e_v \log(p^{f_v}) = -[K:\mathbb{Q}]\log(p). \tag{60}$$

Putting (58), (59) and (60) together completes the proof of (48) and (49) and the proof. $\square$

Remark 4.2. Estimates on the Green–Zhang function on $X_0(p)$ as in the above theorem will be extended below to the Néron model over $\overline{\mathbb{Z}}$ of the whole jacobian $J_0(p)$, see Proposition 5.8.

Remark 4.3. As already noticed, the involution w_p acts as an isometry (actually, an orthogonal symmetry) with respect to the quadratic form h_Θ on $J_0(p)(K) \otimes_\mathbb{Z} \mathbb{R}$. Indeed w_p acts as multiplication by ± 1 on each factor of Shimura's decomposition up to isogeny

$$J_0(p) \sim \prod_{f \in G_\mathbb{Q} \cdot S_2(\Gamma_0(p))^{\text{new}}} J_f$$

whose factors are h_Θ-orthogonal subspaces. (See also [Menares 2008, Corollaire 4.3] or [Menares 2011, Theorem 4.5(3)].) As $w_p(\omega^0) = \omega^0$ (see the proof of Proposition 3.6) this implies

$$h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) = h_\Theta\left(w_p\left(P - \infty + \tfrac{1}{2}\omega^0\right)\right) = h_\Theta\left(w_p(P) - 0 + \tfrac{1}{2}\omega^0\right) = h_\Theta\left(w_p(P) - \infty + \tfrac{1}{2}\omega^0\right)$$

using once more that $(0) - (\infty)$ is torsion, so that

$$[P, \tilde{\omega}_\Theta]_{\mu_0} = [w_p(P), \tilde{\omega}_\Theta]_{\mu_0} = [P, w_p^*(\tilde{\omega}_\Theta)]_{w_p^*(\mu_0)} = [P, w_p^*(\tilde{\omega}_\Theta)]_{\mu_0} \tag{61}$$

(see Remark 3.1). This suggests it could sometimes be convenient to write $\tilde{\omega}_\Theta$ in a w_p-eigenbasis of $\widehat{\mathrm{CH}}(p)^{\mathrm{num}}_{\mathbb{R},\mu}$ instead of that of Theorem 3.2, for instance

$$\widehat{\mathrm{CH}}(p)^{\mathrm{num}}_{\mathbb{R},\mu_0} = \mathbb{R} \cdot \tfrac{1}{2}(0 + \infty) \oplus \mathbb{R} \cdot X_\infty \oplus_{v \mid p} \Gamma_v \oplus (J_0(p)(K) \otimes \mathbb{R}) \tag{62}$$

where now the Γ_v decompose as the direct sum of eigenspaces $\Gamma_v^{w_p=-1}$ and $\Gamma_v^{w_p=+1}$, with bases

$$\{C_{n,m}^- := C_{n,m} - w_p(C_{n,m})\}_{\substack{1 \le n \le s \\ 0 \le m \le ew_n/2}} \quad \text{and} \quad \{C_{n,m}^+ := C_{n,m} + w_p(C_{n,m}) - C_0 - C_\infty\}_{\substack{1 \le n \le s \\ 1 \le m \le ew_n/2}} \tag{63}$$

respectively. Using 3.5 and Proposition 3.6, a lengthy but easy computation allows one to check that

$$\tilde{\omega}_\Theta = g \cdot \tfrac{1}{2}(0 + \infty) + \Phi_\Theta^+ + \gamma_\Theta X_\infty$$

where Φ_Θ^+ is an explicit vertical divisor above p with $w_p^*(\Phi_\Theta^+) = \Phi_\Theta^+$, so that indeed

$$w_p^*(\tilde{\omega}_\Theta) = \tilde{\omega}_\Theta$$

thus recovering (61).

Consider for instance the case of $\mathcal{X}_0(p)$ over $\mathbb{Z}$, for $p \equiv 1 \bmod 12$ (that is, $\mathcal{X}_0(p)_{/\mathbb{Z}}$ is regular, so that there is no need to blow-up singular points of width larger than 1). Here $\Gamma_v = \Gamma_v^- = \mathbb{R} \cdot C_0^- = \mathbb{R} \cdot ([C_\infty] - [C_0]))$ and one readily checks that

$$\tilde{\omega}_\Theta = \frac{g}{2}(0 + \infty) + \gamma_\Theta X_\infty \tag{64}$$

that is, there is no Γ_v-component at all in that case. Evaluating $\mathrm{h}_\Theta\left(\tfrac{1}{2}\omega^0\right)$ as in the proof of Proposition 3.6 and using 3.5,

$$\gamma_\Theta = -\frac{g}{2}[\infty, 0 + \infty]_{\mu_0} + \mathrm{h}_\Theta\left(\tfrac{1}{2}\omega^0\right) = g\,O(\log p/p) + O(\log p) = O(\log p).$$

We then turn to the j-height, first making a comparison of h_j with the "degree component" (in the sense of Theorem 3.2) of the hermitian sheaf ω.

Proposition 4.4. *Let h_j be Weil's j-height on $X_0(p)$ as defined in Section 2B, and let μ_0 and μ_e be the $(1, 1)$-forms defined in (25) and (26). Recall $\sup_{X_0(p)(\mathbb{C})} g_\mu$ stands for the upper bound for all Green functions $g_{\mu,a}$ relative to some point a of $X_0(p)(\mathbb{C})$ and to the measure μ.*

If p is a prime number, K is a number field, and P belongs to $X_0(p)(K)$, then

$$\mathrm{h}_j(P) \le (p+1)\left(\frac{1}{[K:\mathbb{Q}]}[P, \infty]_{\mu_0} + \sup_{X_0(p)(\mathbb{C})} g_{\mu_0} + O(1)\right) \le \frac{(p+1)}{[K:\mathbb{Q}]}[P, \infty]_{\mu_0} + O(p^2 \log p) \tag{65}$$

and similarly

$$\mathrm{h}_j(P) \le (p+1)\left(\frac{1}{[K:\mathbb{Q}]}[P, \infty]_{\mu_e} + \sup_{X_0(p)(\mathbb{C})} g_{\mu_e} + O(1)\right) \le \frac{(p+1)}{[K:\mathbb{Q}]}[P, \infty]_{\mu_e} + O(p^3). \tag{66}$$

Remark 4.5. As explained in the proof below, the function $O(p^2 \log p)$ of (65) comes from [Wilms 2017, Corollary 1.5] together with [Ullmo 2000, Corollaire 1.3] for the estimate of Faltings' δ invariant for $X_0(p)$, which imply the suprema of our functions verify

$$\sup_{X_0(p)(\mathbb{C})} g_{\mu_0} \leq O(p \log p). \tag{67}$$

The function $O(p^3)$ of (66) in turns follows from the main result of [Bruin 2014]. Indeed this states explicitly that $\sup_{X_0(p)(\mathbb{C})} g_{\mu_0} \leq 0.088 \cdot p^2 + 7.7 \cdot p + 1.6 \cdot 10^4$ [loc. cit., Theorem 1.2]. It follows from measures comparison (see (74) below) and the method of P. Bruin that this holds for $\sup_{X_0(p)(\mathbb{C})} g_{\mu_e}$ too, so that

$$\sup_{X_0(p)(\mathbb{C})} g_{\mu_e} \leq O(p^2). \tag{68}$$

It seems that, at least in the case of $X_0(p)$, if we plug into Bruin's method the estimates of [Michel and Ullmo 1998] regarding the comparison function $F(z)$ between Green–Arakelov and Poincaré measures, we recover bounds of shape $O(p \log p)$ instead of $O(p^2)$ (see [Bruin 2014], p. 263 and §8 (Theorem 7.1 in particular)), and the same again holds true for the Green function g_{μ_e}. One should therefore be able to obtain the same error term $O(p^2 \log p)$ for (66) as for (65).

Note that the main theorems of [Jorgenson and Kramer 2006; Aryasomayajula 2013] might even yield that the above functions $O(p^2)$ or $O(p \log p)$ could be replaced by a uniform bound $O(1)$.

Proof. This is essentially a question of measure comparisons on $X_0(p)(\mathbb{C})$ between $j^*(\mu_{FS})$ on one hand (where μ_{FS} is the Fubini–Study $(1, 1)$-form on $X(1)(\mathbb{C}) \simeq \mathbb{P}^1(\mathbb{C})$) and the Green–Arakelov form μ_0 (respectively, μ_e) on the other hand. We adapt the main result of [Edixhoven and de Jong 2011a].

We define first a somewhat canonical Arakelov intersection product $[\cdot, \cdot]_{\mu_{FS}}$ on the projective line using μ_{FS}. Write $\mathbb{P}^1_{/\mathcal{O}_K} = \text{Proj}(\mathcal{O}_K[x_0, x_1]) = \overline{\text{Spec}}^{\text{Zar}}(\mathcal{O}_K[j])$ (with $j = x_1/x_0$), so that the horizontal divisor $\infty(\mathcal{O}_K)$ is $V(x_0)$ and, for any $P = [x_0 : x_1]$, let the associated Green function be

$$g_{\mu_{FS}, \infty}(P) = g_{\mu_{FS}, \infty}(j(P)) = \tfrac{1}{2} \log\left(\frac{|x_0|^2}{|x_0|^2 + |x_1|^2} \right) = -\tfrac{1}{2} \log(1 + |j(P)|^2)$$

at any point different from $\infty = [0 : 1]$. (We note in passing this ad hoc Green function does not need to fulfill the normalization condition (24).) Then for any P in $X(1)(K)$ one easily checks that

$$\left| h_j(P) - \frac{1}{[K : \mathbb{Q}]} [j(P), \infty]_{\mu_{FS}} \right| \leq \tfrac{1}{2} \log(2). \tag{69}$$

Applying [Edixhoven and de Jong 2011a], Theorem 9.1.3 and its proof to the setting described above gives, for any P in $X_0(p)(K)$,

$$[j(P), \infty]_{\mu_{FS}} \leq [P, j^*(\infty)]_{\mu_0} + (p + 1) \sum_\sigma \sup_{X_0(p)_\sigma} g_{\mu_0} + \tfrac{1}{2} \sum_\sigma \int_{X_0(p)_\sigma} \log(|j|^2 + 1) \mu_0 \tag{70}$$

where σ runs through the infinite places of K and $X_0(p)_\sigma := X_0(p) \times_{\mathcal{O}_K, \sigma} \mathbb{C}$.

We estimate the right-hand terms of (70). As for the last integrals we recall that, on the union of disks of ray $|q| < r$ around the cusps (that is, on the image in $X_0(p)(\mathbb{C})$ of the open subset $D_r := \{z \in \mathcal{H} : \Im(z) > -(\log r)/2\pi\}$ in Poincaré upper half-plane $\mathcal{H}$) for some fixed r in $]0, 1[$, one has

$$\left| \frac{f(q)}{q} \right| \leq \frac{2}{(1-r)^2}$$

for any newform f in $S_2(\Gamma_0(p))$. (See for instance [Edixhoven and de Jong 2011b], Lemma 11.3.7 and its proof.) We also know that the Petersson norm of such an f satisfies $\|f\|^2 \geq \pi e^{-4\pi}$ [Edixhoven and de Jong 2011b, Lemma 11.1.2]. Choose $r = \frac{1}{2}$ to fix ideas. On $D_{1/2}$, we have (see (25)):

$$\mu_0 = \frac{i}{2 \dim(J)} \sum_{f \in B_2} \frac{f \frac{dq}{q} \wedge \overline{f \frac{dq}{q}}}{\|f\|^2} \leq \frac{64 e^{4\pi}}{\pi} \frac{i}{2} dq \wedge \overline{dq}.$$

(Sharper bounds should be achievable, but the one above is good enough for our present purpose.) It follows that there exists some real A such that, in the decomposition

$$\int_{X_0(p)(\mathbb{C})} \log(|j|^2 + 1)\mu_0 = \int_{X_0(p)(\mathbb{C}) \cap D_{1/2}} \log(|j|^2 + 1)\mu_0 + \int_{X_0(p)(\mathbb{C}) \setminus D_{1/2}} \log(|j|^2 + 1)\mu_0 \qquad (71)$$

the first term of the right-hand side satisfies

$$\int_{X_0(p)(\mathbb{C}) \cap D_{1/2}} \log(|j|^2 + 1)\mu_0 \leq \frac{64 e^{4\pi}}{\pi} [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(p)] \int_{X(1)(\mathbb{C}) \cap D_{1/2}} \log(|j|^2 + 1)\frac{i}{2} dq \wedge \overline{dq} \leq (p+1)A.$$

As for the second term, remembering that μ_0 has total mass 1 on $X_0(p)(\mathbb{C})$ we check that

$$\int_{X_0(p)(\mathbb{C}) \setminus D_{1/2}} \log(|j|^2 + 1)\mu_0 \leq M_{1/2} := \max_{X(1)(\mathbb{C}) \setminus D_{1/2}} (\log(|j|^2 + 1))$$

whence the existence of some absolute real number A_0 such that

$$\int_{X_0(p)(\mathbb{C})} \log(|j|^2 + 1)\mu_0 \leq (p+1)A_0. \qquad (72)$$

Putting this together with (70) we obtain a constant C for which (69) reads

$$\mathrm{h}_j(P) \leq \frac{1}{[K : \mathbb{Q}]} [P, j^*(\infty)]_{\mu_0} + (p+1)(\sup_{X_0(p)(\mathbb{C})} g_{\mu_0} + A_0).$$

With notations of 3.5, one further has

$$j^*(\infty) = p(0) + (\infty) \equiv (p+1)\infty + p \cdot \Phi_{C_0}^0 \qquad (73)$$

as elements of $\widehat{\mathrm{CH}}(p)_{\mathbb{R}, \mu_0}^{\mathrm{num}}$. Using 3.5(a) we get

$$|[P, \Phi_{C_0}^0]| \leq [K : \mathbb{Q}] \frac{6 \log p}{p - 1}$$

so that, with (67),

$$h_j(P) \le \frac{1}{[K:\mathbb{Q}]}[P, (p+1)\infty]_{\mu_0} + (p+1)(\sup_{X_0(p)(\mathbb{C})} g_{\mu_0} + A_0) + O(\log p)$$
$$\le \frac{1}{[K:\mathbb{Q}]}[P, (p+1)\infty]_{\mu_0} + C_0 \cdot p^2 \log p,$$

which is (65).

The proof of (66) proceeds along the same lines, with one more ingredient. Applying Theorem 9.1.3 of [Edixhoven and de Jong 2011a] with the measure μ_e instead of μ_0 gives the corresponding version of (70). To obtain an upper bound for $\sup_{X_0(p)(\mathbb{C})} g_{\mu_e}$ we recall that the theorem of Kowalski, Michel and Vanderkam asserts that $\dim(J_e) \ge \dim(J_0(p))/5$ for large enough p; see (23). Our measure $\mu_e := \frac{1}{\dim(J_e)} \sum_{S_e} \frac{i}{2}(f\frac{dq}{q} \wedge f\frac{dq}{q})/\|f\|^2$ (see (26)) therefore satisfies

$$0 \le \mu_e \le \frac{g}{\dim(J_e)}\mu_0 \le 5\mu_0. \tag{74}$$

This shows that, as in (68), Bruin's theorem [2014, Theorem 7.1] provides a universal c_e such that

$$\sup_{X_0(p)(\mathbb{C})} g_{\mu_e} \le c_e \, p^2. \tag{75}$$

Using (72) we obtain

$$\int_{X_0(p)(\mathbb{C})} \log(|j|^2 + 1)\mu_e \le (p+1)A_e. \tag{76}$$

Finally, equivalence (73) remains naturally true in the Chow group $\widehat{\mathrm{CH}}(p)^{\mathrm{num}}_{\mathbb{R},\mu_e}$ relative to the measure μ_e instead of μ_0, as remarked in 3.5(a). This completes the proof of (66). $\qquad\square$

We can finally relate h_j and the Néron–Tate height h_Θ relative to the Θ-divisor (see (10)).

Theorem 4.6. *There are real numbers γ and γ_1 such that the following holds. Let K a number field and p a prime number. Let $\omega^0 := -(H_4 + H_3)$ be the 0-component of the canonical sheaf ω on $X_0(p)$ over K (as in Proposition 3.6 and Theorem 4.1). If P is a point of $X_0(p)(K)$ then*

$$h_j(P) \le (12 + o(1)) \cdot h_\Theta\big(P - \infty + \tfrac{1}{2}\omega^0\big) + \gamma \cdot p^2 \log p, \tag{77}$$
$$h_j(P) \le (24 + o(1)) \cdot h_\Theta(P - \infty) + \gamma_1 \cdot p^2 \log p. \tag{78}$$

Remark 4.7. Theorem 4.6 offers only one direction of inequality between j-height and Θ-height; with our method of proof, it is harder to give an effective form to the reverse inequality, because of the metrics comparisons we use (see below).

Notice also that going through the above proofs using the estimate $\sup_{X_0(p)(\mathbb{C})} g_{\mu_0} = O(1)$ of [Jorgenson and Kramer 2006] and [Aryasomayajula 2013] (see Remark 4.5) would even give an error term of shape $O(p)$ instead of $O(p^2 \log p)$ in (78).

Those results are in some sense (hopefully sharp) special cases of the main results of [Pazuki 2012], after rewriting the j-function in terms of classical Θ.

Proof. Using Theorem 4.1, (51), Proposition 4.4 and (15) we obtain

$$\mathrm{h}_j(P) \le 12 \frac{p+1}{p-13} \mathrm{h}_\Theta\big(P - \infty + \tfrac{1}{2}\omega^0\big) + O(p^2 \log p).$$

The last estimate (78) of the theorem comes from the fact that h_Θ is a quadratic form and that

$$\mathrm{h}_\Theta(\omega^0) = O(\log p) \tag{79}$$

by the results of [Michel and Ullmo 1998] now many times mentioned. $\qquad\square$

5. Height of modular curves and the various W_d

We prove in this section a certain number of technical results about heights of cycles in the modular jacobian, which will be useful in the sequel. For applications of the explicit arithmetic Bézout theorem displayed in next section (Proposition 6.1), we indeed first need estimates for the degree and height of the image of $X_0(p)$, together with its various d-th symmetric-products (usually called "W_d"), within either $J_0(p)$ or its quotient J_e, relative to the Θ-polarization. (For more general considerations on this topic, we also refer to [de Jong 2018].) We estimate those heights both in the normalized Néron–Tate sense and for some good ("Moret-Bailly") projective models, to be defined shortly.

Let us first define the height of cycles relative to some hermitian bundle. For further details on this we refer to [Zhang 1995], or to [Abbes 1997, Section 2] for a more informal introduction.

Definition 5.1. Let K be a number field and $\mathcal{O}_K$ its ring of integers. Let $\mathcal{X}$ be an arithmetic scheme over $\mathcal{O}_K$, that is an integral scheme which is projective and flat over $\mathcal{O}_K$, having smooth generic fiber X over K. Let $\mathcal{F}$ be a generically ample and relatively semiample hermitian sheaf with smooth metric, see [Zhang 1995, Section 5]. We denote by $\hat{c}_1(\mathcal{F})$ the first arithmetic Chern class of $\mathcal{F}$, and similarly by $c_1(F)$ the first Chern class of F.

Such a pair $(\mathcal{X}, \mathcal{F})$ will be called a model, in the sense of Zhang, of its pull-back $(X, F) = (\mathcal{X}_K, \mathcal{F}_K)$ to the generic fiber.

Consider a model $(\mathcal{X}, \mathcal{F})$ as in Definition 5.1, and let Y be a d-dimensional subvariety of X. The degree of Y with respect to F is as usual the nonnegative integer given by the d-th power self-intersection of $c_1(F)$ with Y, that is

$$\deg_F(Y) = (c_1(F)^d|Y).$$

We shall sometimes also write that quantity as $\deg_{\mathcal{F}}(Y)$.

Now let $\mathcal{Y} \to \mathcal{X}$ be some "generic resolution of singularities" of Y (that is, some good integral model for some desingularization of Y, see Section 1 of [Zhang 1995]). The height of Y with respect to $\mathcal{F}$ will similarly be the real number obtained by taking the $(\dim \mathcal{Y})$-th power self-intersection of $\hat{c}_1(\mathcal{F})$ with $\mathcal{Y}$, divided by the degree of Y and normalized so that

$$\mathrm{h}_{\mathcal{F}}(Y) = \frac{(\hat{c}_1(\mathcal{F})^{d+1}|\mathcal{Y})}{[K : \mathbb{Q}](d+1) \deg_F(Y)}. \tag{80}$$

One can check that definition[5] does not depend on the desingularization $\mathcal{Y} \to \mathcal{X}$.

Instrumental to us will here be Zhang's control of heights in terms of essential minima. Recall that the (first) essential minimum $\mu_{\mathcal{F}}^{\mathrm{ess}}(Y)$ of Y is the minimum of the set of real numbers μ such that there is a sequence of points (x_n) in $Y(\overline{\mathbb{Q}})$ which is Zariski dense in Y and $\mathrm{h}_{\mathcal{F}}(x_n) \le \mu$ for all n. Zhang's theorem [1995, (5.2)] then asserts that

$$\mathrm{h}_{\mathcal{F}}(Y) \le \mu_{\mathcal{F}}^{\mathrm{ess}}(Y). \tag{81}$$

Note that if $\mathrm{h}_{\mathcal{F}} \ge 0$ on $Y(\overline{\mathbb{Q}})$ one also knows from [Zhang 1995, Theorem 5.2] the reverse inequality

$$\mathrm{h}_{\mathcal{F}}(Y) \ge \frac{\mu_{\mathcal{F}}^{\mathrm{ess}}(Y)}{d+1}. \tag{82}$$

If $(\mathcal{X}, \mathcal{F})$ is a model over $\mathcal{O}_K$, in the sense of Definition 5.1, of a polarized abelian variety (X, F) over $K = \mathrm{Frac}(\mathcal{O}_K)$, and Y again is a d-dimensional subvariety of the generic fiber X, we still define its normalized Néron–Tate height relative to F as the limit

$$\mathrm{h}_F(Y) := \lim_{n \to \infty} \frac{1}{N^{2n}} \mathrm{h}_{\mathcal{F}}([N^n]Y)$$

where N is any fixed integer larger than 1 and $[N^n]Y$ is the image of Y under multiplication by N^n in X. This normalized height, which is a direct generalization of the classical notion of Néron–Tate height for points, is known not to depend neither on the model $\mathcal{X}$ of X, nor the extension $\mathcal{F}$ of F, nor its hermitian structure (and not on N), so that the notation $\mathrm{h}_F(\cdot)$ is finally unambiguous. We refer to [Abbes 1997], Proposition-Définition 3.2 of Section 3 for more details. We will actually use the extension of the two inequalities (81) and (82) to the case where the heights and essential minima are those given by the limit process defining Néron–Tate height (which is known to be nonnegative on points) that is, with obvious notations

$$\frac{\mu_F^{\mathrm{ess}}(Y)}{d+1} \le \mathrm{h}_F(Y) \le \mu_F^{\mathrm{ess}}(Y), \tag{83}$$

see Théorème 3.4 of [Abbes 1997]. As we will see in Section 5C and below, Moret-Bailly theory allows, under certain conditions, to interpret Néron–Tate heights as Arakelov projective heights (that is, without going through limit process).

5A. *Néron–Tate heights.* We shall apply the above to cycles in modular abelian varieties endowed with their symmetric theta divisor: the notation h_Θ will always stand for normalized Néron–Tate height of cycles.

Proposition 5.2. *Let X be the image via $\pi_A \circ \iota_\infty : X_0(p) \to A$ of the modular curve $X_0(p)$ mapped to a nonzero quotient $\pi_A : J_0(p) \to A$ of its jacobian, endowed with the polarization Θ_A induced by the*

---

[5]It could have been simpler to systematically use the definition of height of [Bost et al. 1994, Section 3.1] which does not demand desingularization, as we do in the proof of Proposition 6.1 at the end of Section 6. We could not find references however for Zhang's inequality (see (81)) in that setting, so we stick to the above definitions.

Θ-*divisor* (*see* (4), (9) *and around*). *The degree and normalized Néron–Tate height of* X *satisfy*

$$\deg_{\Theta_A}(X) = \dim(A) = O(p) \quad and \quad h_{\Theta_A}(X) = O(\log p).$$

Proof. If $(A, \Theta_A) = (\mathrm{Jac}(X_0(p)), \Theta)$, it is well-known that the Θ-degree of $X_0(p)$ (or in fact any curve) embedded in its jacobian via some Albanese embedding, equals its genus. That can be seen in many ways, among which one can invoke Wirtinger's theorem [Griffiths and Harris 1978, p. 171], which yields in fact the desired result for any quotient (A, Θ_A); using the notation before (12) we have

$$\deg_{\Theta_A}(X) = \int_{X_0(p)} \sum_{f \in B_2^A} \frac{i}{2} \frac{f \frac{dq}{q} \wedge \overline{f \frac{dq}{q}}}{\|f\|^2} = \dim A \leq g(X_0(p)).$$

We then apply once more the fact (15) that the genus $g(X_0(p))$ is roughly $p/12$. (We could also more simply say that the degree is decreasing by projection, as in the argument below.)

As for the height, the main result of [Michel and Ullmo 1998] gives that the essential minimum of the normalized Néron–Tate height $\mu_\Theta^{\mathrm{ess}}(X_0(p))$ is $O(\log p)$. As the height of points decreases by projection (see Section 2A2 and in particular (7)) the same is true for $\mu_{\Theta_A}^{\mathrm{ess}}(X)$ and we conclude with Zhang's (83). $\square$

Now for the Néron–Tate normalized height of symmetric squares and variants.

Proposition 5.3. *Assume* $X := X_0(p)$ *has gonality strictly larger than* 2 (*which is true as soon as* $p > 71$, *see* [Ogg 1974]). *Let* $\iota := \iota_\infty : X_0(p) \hookrightarrow J_0(p)$ *be the Albanese embedding as in* Proposition 5.2. *Let* $X^{(2)}$ *be the symmetric square* $X_0(p)^{(2)}$ *embedded in* $J_0(p)$ *via* $(P_1, P_2) \mapsto \iota(P_1) + \iota(P_2)$, *and similarly let* $X^{(2),-}$ *be the image of* $(P_1, P_2) \mapsto \iota(P_1) - \iota(P_2)$. *Let* $X_{e^\perp}^{(2)}$ *and* $X_{e^\perp}^{(2),-}$ *be the projections of* $X^{(2)}$ *and* $X^{(2),-}$, *respectively, to* $J_e^\perp$ (*the "orthogonal complement" to the winding quotient* J_e; *see* Section 2B3). *Then with notations as in* Proposition 5.2 *taking* $A = J_0(p)$ *and* $A = J_e^\perp$ *respectively one has*

$$\deg_\Theta(X^{(2)}) = O(p^2) = \deg_\Theta(X^{(2),-}), \quad h_\Theta(X^{(2)}) = O(\log p) = h_\Theta(X^{(2),-})$$

and the same holds for the quotient objects

$$\deg_{\Theta_e^\perp}(X_{e^\perp}^{(2)}) = O(p^2) = \deg_{\Theta_e^\perp}(X_{e^\perp}^{(2),-}), \quad h_{\Theta_e^\perp}(X_{e^\perp}^{(2)}) = O(\log p) = h_{\Theta_e^\perp}(X_{e^\perp}^{(2),-}).$$

Proof. Denoting by p_1 and p_2 the obvious projections below we factor in the common way (see [Mumford 1966], paragraph 3, Proposition 1 on p. 320) our maps over $\mathbb{Q}$ as follows:

$$
\begin{array}{c}
X_0(p) \times X_0(p) \xrightarrow{\pi_A\iota \times \pi_A\iota} A \times A \xrightarrow{M} A \times A \overset{p_2}{\nearrow} A \\
(x, y) \longmapsto (x+y, x-y) \underset{p_1}{\searrow} A
\end{array}
\tag{84}
$$

so $X^{(2)} = p_1 \circ M \circ (\pi_A\iota \times \pi_A\iota)(X_0(p) \times X_0(p))$ and $X^{(2),-} = p_2 \circ M \circ (\pi_A\iota \times \pi_A\iota)(X_0(p) \times X_0(p))$ when $A = J_0(p)$, and the same with $X_{e^\perp}^{(2)}$ and $X_{e^\perp}^{(2),-}$ with $A = J_e^\perp$. We endow $A \times A$ with the hermitian sheaf

$\Theta_A{}^{\boxtimes 2} := p_1^* \Theta_A \otimes p_2^* \Theta_A$. Then $M^*(\Theta_A{}^{\boxtimes 2}) \simeq (\Theta_A{}^{\boxtimes 2})^{\otimes 2}$ [Mumford 1966, p. 320]. Therefore, writing X for $\pi_A \iota(X_0(p))$ in short and using Proposition 5.2,

$$\deg_{\Theta_A{}^{\boxtimes 2}}(M(X \times X)) = 4 \deg_{\Theta_A{}^{\boxtimes 2}}(X \times X) = 8(\deg_{\Theta_A}(X))^2 = O(g^2).$$

As degree decreases by our projections and $O(g^2) = O(p^2)$, $\deg_{\Theta_A}(X^{(2)})$ and $\deg_{\Theta_A}(X^{(2),-})$ are $O(p^2)$.

By definition of essential minima,

$$\mu^{\text{ess}}_{\Theta_A{}^{\boxtimes 2}}(X \times X) \leq 2\mu^{\text{ess}}_{\Theta_A}(X).$$

This implies that $\mu^{\text{ess}}_{\Theta_A{}^{\boxtimes 2}}(M(X \times X)) \leq 4\mu^{\text{ess}}_{\Theta_A}(X)$. Invoking (83) again and Proposition 5.2 together with the fact that the height of points also decreases by projection,

$$\mu^{\text{ess}}_{\Theta_A}(X^{(2)}) \leq \mu^{\text{ess}}_{\Theta_A{}^{\boxtimes 2}}(M(X \times X)) \leq 4\mu^{\text{ess}}_{\Theta_A}(X) \leq 8\text{h}_{\Theta_A}(X) \leq O(\log p).$$

Therefore

$$\text{h}_{\Theta_A}(X^{(2)}) = O(\log p). \qquad \square$$

Note that this proof applies more generally to any subquotient of $J_0(p)$.

5B. *Moret-Bailly models and associated projective heights.* To build-up the projective models of the jacobian (over $\mathbb{Z}$, or finite extensions), and associated heights, that we shall need for our arithmetic Bézout, we use Moret-Bailly theory, in the sense of [Moret-Bailly 1985b], as follows. For more about similar constructions in the general setting of abelian varieties we refer to [Bost 1996, 2.4 and 4.3]; see also [Pazuki 2012].

Let therefore $(J, L(\Theta))$ stand for the principally polarized abelian variety $J_0(p)$ endowed with the invertible sheaf associated with its symmetric theta divisor, defined over some small extension of $\mathbb{Q}$ (see (89) below and around for more details). Endow the complex base-changes of the associated invertible sheaf $L(\Theta)$ with its cubist hermitian metric. If $\mathcal{N}_{J,\mathcal{O}_K}$ is the Néron model of J over the ring of integers $\mathcal{O}_K$ of a number field K, we know it is a semistable scheme over $\mathcal{O}_K$, whose only nonproper fibers are above primes $\mathfrak{P}$ of characteristic p, where it then is purely toric. At any such $\mathfrak{P}$, with ramification index $e_{\mathfrak{P}}$, the group scheme $\mathcal{N}_{J,\mathcal{O}_K}$ has components group

$$\Phi_{\mathfrak{P}} \simeq (\mathbb{Z}/N_0 e_{\mathfrak{P}} \mathbb{Z}) \times (\mathbb{Z}/e_{\mathfrak{P}} \mathbb{Z})^{g-1} \tag{85}$$

for $g := \dim J$ and $N_0 := \text{num}((p-1)/12)$ (see, e.g., [Le Fourn 2016, Proposition 2.11]).

We choose and fix an integer $N > 0$ and a number field $K \supseteq \mathbb{Q}(J[2N])$, for all this paragraph, so that all the $2N$-torsion points in J have values in K. One then observes from (85) that $2N$ divides all the ramification indices $e_{\mathfrak{P}}$, and Proposition II.1.2.2 on p. 45 of [Moret-Bailly 1985b] asserts that $L(\Theta)$ has a cubist extension, let us denote it by $\mathcal{L}(\Theta)$, to the open subgroup scheme $\mathcal{N}_{J,N}$ of the Néron model $\mathcal{N}_{J,\mathcal{O}_K}$ over $\mathcal{O}_K$ whose fibers have component group killed by N.

Such an extension $\mathcal{L}(\Theta)$ is actually symmetric [Moret-Bailly 1985b, Remarque II.1.2.6.2] and unique (see Théorème II.1.1.i on p. 40 of [loc. cit.]). Moreover $\mathcal{L}(\Theta)$ is ample on $\mathcal{N}_{J,N}$ [loc. cit., Proposition VI.2.1

on p. 134]. Its powers $\mathcal{L}(\Theta)^{\otimes r}$ are even very ample on $\mathcal{N}_{J,N} \times_{\mathcal{O}_K} \mathcal{O}_K[1/2p]$ as soon as $r \geq 3$, as follows from the general theory of theta functions. Provided $N > 1$, the sheaf $\mathcal{L}(\Theta)^{\otimes N}$ is spanned by its global sections on the whole of $\mathcal{N}_{J,N}$ [loc. cit., Proposition VI.2.2], although we shall not use that last fact as such.

Picking-up a basis of *generic* global sections in $H^0(J_0(p)_K, L(\Theta)^{\otimes N})$, with $N \geq 3$, we thus defines a map $J_0(p)_K \xrightarrow{J_N} \mathbb{P}^n_K$, for $n = N^g - 1$. Assume our generic global sections extend to a set $\mathcal{S}$ in $H^0(\mathcal{N}_{J,N}, \mathcal{L}(\Theta)^{\otimes N})$. Let $\mathcal{J} \xhookrightarrow{J} \mathbb{P}^n_{\mathcal{O}_K}$ be the schematic closure in $\mathbb{P}^n_{\mathcal{O}_K}$ of the generic fiber $(\mathcal{N}_{J,N})_K = J_K$ via the associated composed embedding $J_K \hookrightarrow \mathbb{P}^n_K \hookrightarrow \mathbb{P}^n_{\mathcal{O}_K}$. Define $\mathcal{M} = J^*\mathcal{O}_{\mathbb{P}^n_{\mathcal{O}_K}}(1)$ on $\mathcal{J}$. Let on the other hand $\mathcal{M}_{\mathcal{N}_{J,N}} := \left(\sum_{s \in \mathcal{S}} \mathcal{O}_K \cdot s\right)$ be the subsheaf of $\mathcal{L}(\Theta)^{\otimes N}$ on $\mathcal{N}_{J,N}$ spanned by $\mathcal{S}$. Write $\nu : \widetilde{\mathcal{N}}_{J,N} \to \mathcal{N}_{J,N}$ for the blowup at base points for $\mathcal{M}_{\mathcal{N}_{J,N}}$ on $\mathcal{N}_{J,N}$, that is, the blowup along the closed subscheme of $\mathcal{N}_{J,N}$ defined by the sheaf $\mathcal{L}(\Theta)^{\otimes N}/\mathcal{M}_{\mathcal{N}_{J,N}}$. We have a commutative diagram

$$
\begin{array}{ccc}
 & \widetilde{\mathcal{N}}_{J,N} & \\
 {}_{\iota_{\mathcal{N}}}\swarrow \quad \downarrow \quad \searrow^{J_{\mathcal{N}}} & \\
J_K \hookrightarrow \quad \mathcal{J} \xhookrightarrow{J} \mathbb{P}^n_{\mathcal{O}_K} &
\end{array}
\tag{86}
$$

where the only nontrivial map $J_{\mathcal{N}}$ (whence $\iota_{\mathcal{N}}$) is deduced from the fundamental properties of blowups. Considering the complex base-changes of the generic fiber we note that $\mathcal{M}$ is automatically endowed with a cubist hermitian structure induced by that of $L(\Theta)_{\mathbb{C}}$ (see [Bost 1996], (4.3.3) and following lines).

Definition 5.4. Given an integer $N \geq 3$, and a number field K containing $\mathbb{Q}(J_0(p)[2N])$, we define the "good model" for $(J_0(p), L(\Theta)^{\otimes N})$ relative to some finite set $\mathcal{S}$ in $H^0(\mathcal{N}_{J,N}, \mathcal{L}(\Theta)^{\otimes N})$, which spans $H^0(J_0(p), L(\Theta)^{\otimes N})$, as the projective scheme $\mathcal{J}$ over $\mathrm{Spec}(\mathcal{O}_K)$ enhanced with the hermitian sheaf $\mathcal{M}$ constructed above, and $h_{\mathcal{M}}$ the associated height.

Outside base points for $\mathcal{M}_{\mathcal{N}_{J,N}}$ on $\mathcal{N}_{J,N}$ the blowup $\nu : \widetilde{\mathcal{N}}_{J,N} \to \mathcal{N}_{J,N}$ is an isomorphism and on that open locus we have

$$
\mathcal{L}(\Theta)^{\otimes N} \simeq \mathcal{M}_{\mathcal{N}_{J,N}} \simeq \iota_{\mathcal{N}}^* \mathcal{M} = J_{\mathcal{N}}^* \mathcal{O}_{\mathbb{P}^n_{\mathcal{O}_K}}(1)
\tag{87}
$$

so we dwell on the fact that the height $h_{\mathcal{M}}$ of our "good models" for $(J_0(p), L(\Theta)^{\otimes N})$ will indeed compute (N times) the Néron–Tate height of *certain* $\overline{\mathbb{Q}}$-points (those whose closure factorizes through $\mathcal{N}_{J,N}$ deprived from the base points for $\mathcal{S}$), but definitely *not all*. For arbitrary points, still, one can deduce from the work of Bost ([Bost 1996], 4.3) the following inequality.

Proposition 5.5. *For any point P in $J_0(p)(\overline{\mathbb{Q}})$, the height $h_{\mathcal{M}}(P)$ of Definition 5.4 satisfies*

$$
h_{\mathcal{M}}(P) \leq N h_{\Theta}(P).
$$

Proof. We briefly adapt [Bost 1996, 2.4 and 4.3] using our above notations. Of course this statement has nothing to see with modular jacobians, and holds for any abelian variety over a number field. Let N' be some integer such that P defines a section of $\mathcal{N}_{J,N'}(\mathcal{O}_F)$ for some ring of integers $\mathcal{O}_F$. Up to replacing

$\mathcal{O}_F$ by a sufficiently ramified finite extension, we can assume $L(\Theta)^{\otimes N}$ has a cubist extension $\mathcal{L}(\Theta)^{\otimes N}$ to all of $\mathcal{N}_{J,N'}$ over $\mathcal{O}_F$ [Moret-Bailly 1985b, Proposition II.1.2.2]. One has

$$h_\Theta(P) = \frac{1}{N}\frac{1}{[F:\mathbb{Q}]}\widehat{\deg}(P^*(\mathcal{L}(\Theta)^{\otimes N})).$$

As in (86) however we see that there is no well-defined map from $\mathcal{N}_{J,N'}$ to $\mathbb{P}^n_{\mathcal{O}_F}$ because $\mathcal{L}(\Theta)^{\otimes N}$ needs not be spanned by elements of $\mathcal{S}$ on all of $\mathcal{N}_{J,N'}$ (even though it is, by hypothesis, on the generic fiber). To remedy this we adapt the construction (86).

If $\pi': \mathcal{N}_{J,N'} \to \mathrm{Spec}(\mathcal{O}_F)$ is the structural morphism, we define now $\mathcal{M}'_\mathcal{N} := \left(\sum_{s\in\mathcal{S}}\mathcal{O}_F \cdot s\right)$ as the subsheaf of $\mathcal{L}(\Theta)^{\otimes N}$ on $\mathcal{N}_{J,N'}$ spanned by $\mathcal{S}$, still endowed with the metric induced by that of $\mathcal{L}(\Theta)^{\otimes N}$. One checks (see [Bost 1996, (4.3.8)]) that the projective model $\mathcal{J}_{\mathcal{O}_F}$ of $(\mathcal{N}_{J,N'})_F \simeq J_F$ in $\mathbb{P}^n_{\mathcal{O}_F}$ defined as in (86) yields a sheaf $\mathcal{M}'$ on $\mathcal{J}_{\mathcal{O}_F}$, whence a height $h_{\mathcal{M}'}$, which coincides with the height $h_\mathcal{M}$ on the base change of the good model $\mathcal{J}_{\mathcal{O}_K}$.

Replacing $\mathcal{N}_{J,N'}$ by its blowup $\nu': \widetilde{\mathcal{N}}_{J,N'} \to \mathcal{N}_{J,N'}$ at base points for $\mathcal{M}'_\mathcal{N}$ in $\mathcal{L}(\Theta)^{\otimes N}$ on $\mathcal{N}_{J,N'}$, we keep on following construction (86) to obtain maps $\iota'_\mathcal{N}: \widetilde{\mathcal{N}}_{J,N'} \to \mathcal{J}_{\mathcal{O}_F}$ and $\jmath'_\mathcal{N}: \widetilde{\mathcal{N}}_{J,N'} \to \mathbb{P}^n_{\mathcal{O}_F}$ such that the Zariski closure of $\jmath'_\mathcal{N}(\widetilde{\mathcal{N}}_{J,N'})$ identifies with $\mathcal{J}_{\mathcal{O}_F}$. We moreover have

$$\iota'^*_\mathcal{N}(\mathcal{M}') = \nu'^*(\mathcal{L}(\Theta)^{\otimes N}) \otimes \mathcal{O}(-E)$$

where E is the exceptional divisor of the blowup which is by definition effective. The section P of $\mathcal{N}_{J,N'}(\mathcal{O}_F)$ lifts to some $\tilde{P}$ of $\widetilde{\mathcal{N}}_{J,N'}(\mathcal{O}_F)$. Let ε_P be the section of $\mathcal{J}(\mathcal{O}_F)$ defined by the Zariski closure of $P(F)$ in $\mathcal{J}$. One can finally compute

$$h_\mathcal{M}(P) = h_{\mathcal{M}'}(P) = \frac{1}{[F:\mathbb{Q}]}\widehat{\deg}(\varepsilon_P^*(\mathcal{M}')) = \frac{1}{[F:\mathbb{Q}]}\widehat{\deg}(\tilde{P}^*(\iota'^*_\mathcal{N}(\mathcal{M}')))$$

$$\leq \frac{1}{[F:\mathbb{Q}]}\widehat{\deg}(\tilde{P}^*(\nu'^*(\mathcal{L}(\Theta)^{\otimes N}))) = \frac{1}{[F:\mathbb{Q}]}\widehat{\deg}(P^*(\mathcal{L}(\Theta)^{\otimes N})) = N\,h_\Theta(P). \quad \square$$

The following straightforward generalization to higher dimension will be useful in next section.

Corollary 5.6. *If Y is a d-dimensional irreducible subvariety of $J_0(p)$ then*

$$h_\mathcal{M}(Y) \leq (d+1)\,N\,h_\Theta(Y).$$

Proof. Combine Zhang's formulas (81) and (83) with Proposition 5.5. $\square$

Recall from (8) that one can define the "pseudoprojection" $\mathcal{P}_{\tilde{J}_{e^\perp}}(\iota_\infty(X_0(p)))$ of the image of $X_0(p) \xrightarrow{\iota_\infty} J_0(p)$ on the subabelian variety $\tilde{J}_{e^\perp} \subseteq J_0(p)$. Let $X_{e^\perp}$ be any of its irreducible components. Define similarly $X^{(2)}$, $X^{(2),-}$, $X^{(2)}_{e^\perp}$ and $X^{(2),-}_{e^\perp}$ as in Proposition 5.3. Note that, by construction, the degree and normalized Néron–Tate height of $X_{e^\perp}$ (and other similar pseudoprojections: $X^{(2)}_{e^\perp}$ etc.), as an irreducible subvariety of $J_0(p)$ endowed with h_Θ, are those of $\pi_{J_e^\perp}(X_0(p)) = X^{(2),-}_{e^\perp}$ relative to the only natural hermitian sheaf of $J_e^\perp$, that is, the $\Theta_e^\perp = \Theta_{J_e^\perp}$ described in paragraph 2A2 and estimated in Proposition 5.2.

Corollary 5.7. *For any fixed integer $N \geq 3$, and any number field K containing $\mathbb{Q}(J_0(p)[2N])$, let $(\mathcal{J}, \mathcal{M})$ be the good model for $(J_0(p), L(\Theta)^{\otimes N})$, and $h_\mathcal{M}$ the associated projective height, given in*

Definition 5.4. Let X be the image of $X_0(p) \overset{\iota_\infty}{\hookrightarrow} J_0(p)$, *and more generally* $X^{(2)}$, $X^{(2),-}$, $X^{(2)}_{e^\perp}$ *and* $X^{(2),-}_{e^\perp}$ *be the objects* $X^{(2)}, \ldots$ *defined in Proposition 5.3 (or their pseudoprojections). Then their* $\mathcal{M}^{\otimes \frac{1}{N}}$*-heights are bounded from above by similar functions as their Néron–Tate height (Proposition 5.3). Explicitly,* $\mathrm{h}_{\mathcal{M}^{\otimes \frac{1}{N}}}(X_0(p))$ *is less than* $O(\log p)$*, and* $\mathrm{h}_{\mathcal{M}^{\otimes \frac{1}{N}}} X^{(2)}$*, etc., are all less than* $O(\log p)$*. Similarly the* $\mathcal{M}^{\otimes \frac{1}{N}}$*-degree of* $X_0(p)$ *is* $O(p)$*, and the* $\mathcal{M}^{\otimes \frac{1}{N}}$*-degrees of* $X^{(2)}$*, etc., are all* $O(p^2)$*.*

Proof. Combine Zhang's formulas (81) and (83) with Propositions 5.2, 5.3 and 5.5. $\qquad \square$

5C. *Estimates on Green–Zhang functions for* $J_0(p)$. We shall later on need some control on the p-adic Néron–Tate metric of Θ as alluded to in Remark 4.3. (Those statements can probably be best formulated in the setting of Berkovich theory, for which one might check in particular [Ducros 2007, Proposition 2.12] and [Thuillier 2005]. A useful point of view is also proposed by that of "tropical jacobians", see [Mikhalkin and Zharkov 2008; de Jong and Shokrieh 2018]. We will content ourselves here with our down-to-earth point of view). We therefore define

$$\hat{\Phi}_p := \varinjlim_{K_{\mathfrak{P}} \supseteq \mathbb{Q}_p} \Phi_{\mathfrak{P}}$$

as the direct limit, on a tower of totally ramified extensions $K_{\mathfrak{P}}/\mathbb{Q}_p$, of the component groups $\Phi_{\mathfrak{P}}$ of the Néron models of $J_0(p)$ at $\mathfrak{P}$, see (85). The compatible embeddings

$$Z := \langle C_0 - C_\infty \rangle \simeq \langle (0) - (\infty) \rangle \simeq \mathbb{Z}/N_0 \mathbb{Z} \hookrightarrow \Phi_{\mathfrak{P}}$$

for each $\mathfrak{P}$ induce an exact sequence $0 \to Z \to \hat{\Phi}_p \to \varinjlim_{e_{\mathfrak{P}}}(\mathbb{Z}/e_{\mathfrak{P}}\mathbb{Z})^g \simeq (\mathbb{Q}/\mathbb{Z})^g \to 0$. Passing to the real completion yields a presentation:

$$0 \to Z \simeq \mathbb{Z}/N_0 \mathbb{Z} \to \hat{\Phi}_{p,\mathbb{R}} \to (\mathbb{R}/\mathbb{Z})^g \to 0 \tag{88}$$

(where $\hat{\Phi}_{p,\mathbb{R}}$ must be the "skeleton", in the sense of Berkovich, of the Néron model over $\overline{\mathbb{Z}}_p$ of $J_0(p)$, and the tropical jacobian, see [de Jong and Shokrieh 2018], of the curve $X_0(p)$ above p). The right-hand side of (88) is more canonically written $(\mathbb{R}/\mathbb{Z})^g \simeq (\mathbb{R}/\mathbb{Z})^s / \Delta(\mathbb{R})$, for Δ the almost diagonal map [Le Fourn 2016, Proposition 2.11.(c)]

$$\Delta(z) \mapsto \left(\tfrac{1}{w_i} z \right)_{1 \le i \le g+1}.$$

We then sum up useful properties about theta divisors and theta functions "over $\overline{\mathbb{Z}}$".

As $J_0(p)$ is principally polarized over $\mathbb{Q}$, the complex extension of scalars $J_0(p)(\mathbb{C})$ can be given a classical complex uniformization $\mathbb{C}^g / (\mathbb{Z}^g + \tau \mathbb{Z}^g)$ for some τ in Siegel's upper half-plane. The associated Riemann theta function

$$\theta(z) = \sum_{m \in \mathbb{Z}^g} \exp(i\pi \,{}^t m \cdot \tau \cdot m + 2i\pi \,{}^t m \cdot z) \tag{89}$$

defines the tautological global section 1 of a trivialization of $\mathcal{O}_{J_0(p)}(\Theta_{\mathbb{C}})(= \mathcal{M}_{\mathbb{C}}^{\otimes 1/N})$ for $\Theta_{\mathbb{C}}$ the image W_{g-1} of some $(g-1)$-st power of $X_0(p)$ in $J_0(p)$. More precisely, Riemann's classical results (e.g., [Griffiths and Harris 1978], Theorem on p. 338) assert that $\mathrm{div}(\theta(z)) = \Theta_{\mathbb{C}}$ is the divisor with support

$\{\kappa_{P_0} + \sum_{i=1}^{g-1} \iota_{P_0}(P_i), P_i \in X_0(p)(\mathbb{C})\}$, where for any $P_0 \in X_0(p)(\mathbb{C})$ we write $\iota_{P_0} : X_0(p) \hookrightarrow J_0(p)$ for the Albanese morphism with base point P_0, and $\kappa = \kappa_{P_0} = \text{``}\frac{1}{2}(\iota_{P_0}(K_{X_0(p)}))\text{''}$ for the image of Riemann's characteristic, which is some preimage under duplication in $J_0(p)$ of the image of some canonical divisor: $\omega^0 = \iota_{P_0}(K_{X_0(p)})$ (see Theorem 4.6 above).

Among the translates $\Theta_D = t_D^* \Theta$, for $D \in J_0(p)(\mathbb{C})$, of the above symmetric Θ, the divisor $\Theta_\kappa = t_\kappa^* \Theta = \sum_{i=1}^{g-1} \iota_\infty(X_0(p)_\mathbb{Q})$ defines an invertible sheaf $L(\Theta_\kappa)$ on $J_0(p)$ over $\mathbb{Q}$. If $\mathcal{N}_{J,1}$ denotes the neutral component of the Néron model of J over $\mathbb{Z}$ and $\mathcal{L}(\Theta_\kappa)$ is the cubist extension of $L(\Theta_\kappa)$ to $\mathcal{N}_{J,1}$ (compare [Moret-Bailly 1985b, Proposition II.1.2.2], as in Section 5B above), we know that $H^0(\mathcal{N}_{J,1}, \mathcal{L}(\Theta_\kappa))$ is a (locally) free $\mathbb{Z}$-module of rank 1, so that the complex base-change $H^0(J_0(p)(\mathbb{C}), L(\Theta_{\kappa,\mathbb{C}}))$ is similarly a complex line. This means that if s_θ is a generator of the former space, whose image in the later we denote by $s_{\theta,\mathbb{C}}$, there is a nonzero complex number C_ϑ such that

$$s_{\theta,\mathbb{C}}(z) = C_\vartheta \cdot \theta(z + \kappa). \tag{90}$$

Up to making some base-change from $\mathbb{Z}$ to some $\mathcal{O}_K$ we can now forget about κ and come back to the symmetric Θ; we define a global section

$$s_{\mathcal{J}^0} := (t_{-\kappa}^*)s_\theta \in H^0(\mathcal{N}_{J,1}, \mathcal{L}(\Theta)_{\mathcal{O}_K}) \quad \text{so that} \quad s_{\mathcal{J}^0,\mathbb{C}}(z) = C_\vartheta \cdot \theta(z). \tag{91}$$

If one replaces $\mathcal{N}_{J,1}$ by the Néron model, say $\mathcal{N}_{\mathcal{O}_{K_1}}$, of $J_0(p)$ over any extension K_1 of K, then [Moret-Bailly 1985b, Proposition II.1.2.2] insures that up to making some further field extension K_2/K_1 the sheaf $L(\Theta)_{K_2}$ has a cubist extension $\mathcal{L}(\Theta)_{\mathcal{O}_{K_2}}$ to $\mathcal{N}_{\mathcal{O}_{K_1}} \times_{\mathcal{O}_{K_1}} \mathcal{O}_{K_2}$. Therefore $s_{\mathcal{J}^0}$ extends to a *rational* section (we shall sometimes write *meromorphic* section) of $\mathcal{L}(\Theta)_{\mathcal{O}_{K_2}}$ on $\mathcal{N}_{\mathcal{O}_{K_1}} \times_{\mathcal{O}_{K_1}} \mathcal{O}_{K_2}$. Abusing notations we still denote that extended section by $s_{\mathcal{J}^0}$, and write accordingly Θ for its divisor $\text{div}(s_{\mathcal{J}^0})$ on $\mathcal{N}_{\mathcal{O}_{K_1}} \times_{\mathcal{O}_{K_1}} \mathcal{O}_{K_2}$. Because $s_{\mathcal{J}^0}$ is well defined (and nonzero) on the neutral component of the Néron model, its poles on $\mathcal{N}_{\mathcal{O}_{K_1}} \times_{\mathcal{O}_{K_1}} \mathcal{O}_{K_2}$ can only show-up at places of bad reduction.

Proposition 5.8. *The multiplicity of the Θ-divisor at any component of the Néron model of $J_0(p)$ over $\overline{\mathbb{Z}}$, normalized to be 0 along the neutral component, is $O(p)$.*

Proof. We start by the following observations. Let us write $s_{\mathcal{J}^0,\mathbb{C}}(z) = C_\vartheta \cdot \theta(z)$ as in (91). Take D in $J_0(p)(\mathbb{C})$ which can written as the linear equivalence class of some divisor

$$D = \sum_{i=1}^{g} -(Q_i - \infty)$$

for points Q_i in $X_0(p)(\mathbb{C})$. We associate to D the embedding

$$\iota_{\kappa+D} : X_0(p) \hookrightarrow J_0(p), \quad P \mapsto \text{cl}(P - \infty + \kappa + D)$$

where κ is Riemann's characteristic (see just before (57)). For such a D whose Q_i are assumed to belong to $X_0(p)(\overline{\mathbb{Q}})$, we know from the proof of Theorem 4.1 (see (54)) that

$$\text{h}_\Theta(\iota_{\kappa+D}(P)) = \frac{1}{[K(P,D):\mathbb{Q}]}[P, \tilde{\omega}_D]_{\mu_0} \tag{92}$$

with

$$\tilde{\omega}_D = \sum_i Q_i + \Phi_D + c_D X_\infty \tag{93}$$

and Φ_D is the explicit vertical divisor

$$\Phi_D = \tfrac{1}{2}(\Phi_\omega + \Phi_\vartheta) - \sum_{i=1}^{g} \Phi_{Q_i} \tag{94}$$

at each bad place, with notations as those of the proof of Theorem 4.1; see (55).

Moreover, it is well known that there is a subset of $J_0(p)(\mathbb{C})$ which is open for the complex topology, and even the Zariski topology, in which all points $D = \sum_1^g -(Q_i - \infty)$ as above are such that

$$\dim_{\mathbb{C}} H^0(X_0(p)(\mathbb{C}), L(-D + g \cdot \infty)_{\mathbb{C}}) = \dim_{\mathbb{C}} H^0(X_0(p)(\mathbb{C}), \iota_{\kappa+D}^* L(\Theta_{\mathbb{C}})) = 1 \tag{95}$$

so that $\iota_{\kappa+D}^*(\Theta_{\mathbb{C}}) = \sum_i Q_{i,\mathbb{C}}$, the latter being an equality between effective divisors, not just a linear equivalence [Griffiths and Harris 1978, pp. 336–340]. As the height h_Θ, in the Néron model of $J_0(p)$, can be understood as the Arakelov intersection with $\Theta = \mathrm{div}(s_{\mathcal{J}^0})$ it follows that, on the curve $X_0(p)$, $\mathrm{div}(s_{\mathcal{J}^0,\mathbb{C}}) \cap \iota_{\kappa+D}(X_0(p))(\mathbb{C}) = \bigcup_i \iota_{\kappa+D}(Q_{i,\mathbb{C}})$ or $\mathrm{div}(\iota_{\kappa+D}^*(s_{\mathcal{J}^0,\mathbb{C}})) = \sum_i Q_i$ over $\mathbb{C}$. More precisely, extending base to some ring of integers $\mathcal{O}_K$ so that the Q_i define sections of the minimal regular model $\mathcal{X}_0(p)_{\mathcal{O}_K}$ of $X_0(p)$ over $\mathcal{O}_K$, and making if necessary a further base extension such that $\mathcal{L}(\Theta)$ has a cubist extension on the whole Néron model of $J_0(p)$ over $\mathcal{O}_K$ (as after (91)), one sees that $s_{\mathcal{J}^0}$ defines a meromorphic section of $\mathcal{L}(\Theta)_{\mathcal{O}_K}$ and the restriction to the generic fiber $X_0(p)_K$ of $\mathrm{div}(\iota_{\kappa+D}^*(s_{\mathcal{J}^0}))$ has to be equal (and not merely linearly equivalent) to $\sum_i Q_i$. Now in such a situation, the multiplicity of $\mathrm{div}(s_{\mathcal{J}^0})$ on a component of the Néron model to which $\mathcal{X}_0(p)_{\mathcal{O}_K}^{\mathrm{smooth}}$ is mapped via $\iota_{\kappa+D}$, can be read on the multiplicity of $\iota_{\kappa+D}^*(s_{\mathcal{J}^0})$ along that component of $\mathcal{X}_0(p)_{\mathcal{O}_K}^{\mathrm{smooth}}$. In turn, because of decompositions of the arithmetic Chow group similar to that of Theorem 3.2, multiplicities of $\mathrm{div}(s_{\mathcal{J}^0})$ are determined by the Φ_D of (93), up to constant addition of vertical fibers. The property that $\mathrm{div}(s_{\mathcal{J}^0})$ has multiplicity 0 along the neutral component of the Néron model (see (91)) fixes that last indetermination. Now if $\mathfrak{P}$ is a place of bad reduction for $\mathcal{X}_0(p)_{\mathcal{O}_K}$, and if the Q_i move slightly in the $\mathfrak{P}$-adic topology (without modifying their specialization component at $\mathfrak{P}$), the vertical divisor Φ_D does not change either at $\mathfrak{P}$, and the above reasoning regarding the components values of Θ is actually independent from the fact that condition (95) holds true or not (provided, we insist, that the specialization components of the Q_i at $\mathfrak{P}$ do not vary).

We shall gain some flexibility with a last preliminary remark. If k is any integer between 0 and $N_0 - 1$ (recall N_0 is the order of the Eisenstein element $(0 - \infty)$), the divisor $\tilde{\omega}_D$ of (93) can still be written as

$$\tilde{\omega}_D = \left(k \cdot 0 + (g - k) \cdot \infty - k\Phi_{C_0} + \tfrac{1}{2}(\Phi_\omega + \Phi_\vartheta) - \tilde{D}\right) + c_D X_\infty$$

so that if

$$D = \left(\sum_{i=1}^{g} -(Q_i - \infty)\right) + k(0 - \infty) = \sum_{i=1}^{k} -(Q_i - 0) + \sum_{i=k+1}^{g} -(Q_i - \infty)$$

then $\tilde{\omega}_D = \sum_{i=1}^{g} Q_i + \Phi_D + c_D X_\infty$ where Φ_D is still

$$\Phi_D = \tfrac{1}{2}(\Phi_\omega + \Phi_\vartheta) - \sum_{i=1}^{g} \Phi_{Q_i}. \tag{96}$$

Coming back to the proof of the present Proposition 5.8, and assuming first $D = 0$, it follows from what we have just discussed that the multiplicity of the Θ-divisor on the components of the jacobian to which the components of $\mathcal{X}_0(p)^{\text{smooth}}_{\mathcal{O}_K}$ map under ι_κ is given by the functions g_n and G of (45) and (46); see Theorem 4.1. To obtain the multiplicity of the Θ-divisor on *all* components of the jacobian we shall shift our Albanese embeddings $\iota_{\kappa+D}$ in order to explore all of $J_0(p)/J_0(p)^0$ with successive translations of $\mathcal{X}_0(p)^{\text{smooth}}_{\mathcal{O}_K}$ inside $J_0(p)$.

To be more explicit, let $\mathfrak{C}$ be an element of the component group $J_0(p)/J_0(p)^0$ at $\mathfrak{P}$, and $D = \sum_{i=1}^{g}(P_i - \infty)$ be a divisor, with all P_i in $X_0(p)(K)$, which reduces to $\mathfrak{C}$ at $\mathfrak{P}$. For all r in $\{1, \ldots, g\}$, set $D_r = \sum_{i=1}^{r}(P_i - \infty)$ and let also k_r in $\{1, \ldots, N_0 - 1\}$ and $Q_{i,r}$ be g associated points on the curve such that one can write both

$$D_r = \sum_{i=1}^{r}(P_i - \infty) \quad \text{and} \quad D_r = \sum_{i=1}^{g} -(Q_{i,r} - \infty) + k_r(0 - \infty).$$

As always in this proof, up to making a finite base-field extension one can assume all points have values in K. Recall also from the discussion above that one can move slightly the Q_i in the $\mathfrak{P}$-adic topology, as all that interests us here is the component $\mathfrak{C}_r$, $1 \le r \le g$, of $(J_0(p)/J_0(p)^0)_{\mathfrak{P}}$ to which D_r maps. One can therefore assume if one wishes that $\iota^*_{\kappa+D_r}(\Theta_{\mathbb{C}}) = \sum_i Q_{i,\mathbb{C}}$ (equality, not just linear equivalence). The presentation of $\Phi_{\mathfrak{P}}$ given in (88) and above also shows one can assume that the specialization components at $\mathfrak{P}$ of the $Q_{i,r}$, in $\mathcal{X}_0(p)^{\text{smooth}}_{\mathcal{O}_K}$, which are not C_∞, are all different (see Figure 1).

Taking first $D = 0$, that is, using the map ι_κ, we already remarked that (94) implies the value V_1 of $\text{div}(s_{\mathcal{J}^0})$ on $\mathfrak{C}_1$ is $V_1 = \left[\tfrac{1}{2}(\Phi_\vartheta + \Phi_\omega), P_1\right] = \tfrac{1}{2}([\Phi_\omega, P_1] + [\Phi_{P_1}]^2)$ (see (53)). By Remark 3.4 and (34), $|V_1| \le 2$.

Going one step further we reach $\mathfrak{C}_2$ by considering the Albanese image $\iota_{\kappa+D_1}(\mathcal{X}_0(p)^{\text{smooth}}_{\mathcal{O}_K})$ and looking at the image of P_2. Here we need not to forget that the ∞-cusp in $X_0(p)$ now maps to $\mathfrak{C}_1$, so the normalization of components-divisor on the *curve* $\mathcal{X}_0(p)^{\text{smooth}}_{\mathcal{O}_K}$ at $\mathfrak{P}$ cannot be fixed to be 0 along the ∞-component any longer; it needs to take the value V_1 found above, in order to match with the normalization of the theta divisor on the jacobian. Applying the same reasoning as before with formula (96) gives that the value of Θ on $\mathfrak{C}_2$ is

$$V_2 = \left[P_2, \tfrac{1}{2}(\Phi_\omega + \Phi_\vartheta) - \sum_{i=1}^{g} \Phi_{Q_{i,1}} + V_1\right] = \tfrac{1}{2}([\Phi_\omega, P_2] + [\Phi_{P_2}]^2) - \sum_{i=1}^{g}[\Phi_{Q_{i,1}}, P_2] + V_1$$

so that $|V_2| \le 9$ invoking Remark 3.4 again, and recalling the $Q_{i,1}$ specialize to different branches of Figure 1.

From there the inductive process is clear which yields that the value of Θ on $\mathfrak{C}_r$ has absolute value less or equal to $7r$, whence the proof of Proposition 5.8. $\qquad\square$

5D. *Explicit modular version of Mumford's repulsion principle.* We conclude this section by writing-down, for later use, an explicit version of Mumford's well-known "repulsion principle" for points, in the case of modular curves.

Proposition 5.9. *For P and Q two different points of $X_0(p)(\overline{\mathbb{Q}})$ one has*

$$h_\Theta(P - Q) \geq \frac{g-2}{4g}(h_\Theta(P - \infty) + h_\Theta(Q - \infty)) - O(p \log p). \tag{97}$$

Proof. Let K be a number field such that both P and Q have values in K. Using notations of Section 3, the adjunction formula and Hodge index theorem give

$$
\begin{aligned}
2[K:\mathbb{Q}]h_\Theta(P - Q) &= -[P - Q - \Phi_P + \Phi_Q, P - Q - \Phi_P + \Phi_Q]_{\mu_0} \\
&= [P + Q, \omega]_{\mu_0} + 2[P, Q]_{\mu_0} + [\Phi_P - \Phi_Q]^2 \\
&\geq [P + Q, \omega]_{\mu_0} - 2[K:\mathbb{Q}]\sup g_{\mu_0} + [\Phi_P - \Phi_Q]^2.
\end{aligned}
$$

In the same way,

$$
\begin{aligned}
[P, \omega]_{\mu_0} &= 2[K:\mathbb{Q}]h_\Theta(P - \infty) - 2[P, \infty]_{\mu_0} + [\infty]^2_{\mu_0} - [\Phi_P]^2 \\
&\geq [K:\mathbb{Q}]h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) - 2[P, \infty]_{\mu_0} + [\infty]^2_{\mu_0} - [\Phi_P]^2
\end{aligned}
$$

where the last inequality comes from the quadratic nature of h_Θ, plus the fact that the error term of (97) allows us to assume $h_\Theta(P - \infty) \geq 1/(12 - 8\sqrt{2})h_\Theta(\omega^0) = O(\log p)$ (see (79) and the end of proof of Theorem 4.6). Now by (51),

$$h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) = \frac{1}{[K:\mathbb{Q}]}[P, g \cdot \infty]_{\mu_0} + O(\log p)$$

and using Remark 3.4 and 3.5 gives

$$[P, \omega]_{\mu_0} \geq \frac{g-2}{g}[K:\mathbb{Q}]h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) + [K:\mathbb{Q}]O(\log p).$$

As $[\Phi_P, \Phi_Q] = [P, \Phi_Q] = [Q, \Phi_P]$, we have $|[\Phi_P, \Phi_Q]| \leq 3[K:\mathbb{Q}]\log p$ using Remark 3.4 again. Putting everything together with Remark 4.5 about $\sup g_{\mu_0}$ we obtain

$$h_\Theta(P - Q) \geq \frac{g-2}{2g}\left(h_\Theta\left(P - \infty + \tfrac{1}{2}\omega^0\right) + h_\Theta\left(Q - \infty + \tfrac{1}{2}\omega^0\right)\right) - O(p \log p)$$

which, by our previous remarks, can again be written as

$$h_\Theta(P - Q) \geq \frac{g-2}{4g}(h_\Theta(P - \infty) + h_\Theta(Q - \infty)) - O(p \log p). \qquad\square$$

(For large p, the angle between two points of equal large enough height is here therefore at least $\arccos \frac{3}{4} - \varepsilon > \frac{\pi}{6}$. Of course the natural value is $\frac{\pi}{2}$, to which one tends when sharpening the computations.)

6. Arithmetic Bézout theorem with cubist metric

We display in this section an explicit version of Bézout arithmetic theorem, in the sense of Philippon [1994] or Bost, Gillet and Soulé [1994], for intersections of cycles in our modular abelian varieties over number fields, with the following variants: we use Arakelov heights (as in Section 5 above, see (80)) on higher-dimensional cycles and we endow the implicit hermitian sheaf for this height with its cubist metric (instead of Fubini–Study).

It indeed seems that one generally uses Fubini–Study metrics for arithmetic Bézout because they are the only natural explicit ones available on a general projective space (a necessary frame for the approach we follow for Bézout-like statements). They moreover have the pleasant feature that the relevant projective embeddings have tautological basis of global sections with sup-norm less than 1 which, for instance, allows for proving that the induced Faltings height is nonnegative on effective cycles [Faltings 1991, Proposition 2.6]. For our present purposes however, we need bounds for the Néron–Tate heights of points, that is, Arakelov heights induced by cubist metrics. One could in principle have tried working with Fubini–Study metrics as in [Bost et al. 1994] and then directly compare with Néron–Tate heights, but comparison terms tend to be huge. In the case of rational points, for instance (that is, horizontal cycles of relative dimension 0), within jacobians, those error terms are bounded by Manin and Zarhin [1972] linearly in the ambient projective dimension, that is exponential in the dimension of the abelian variety. In other words, for our modular curves, the error terms would be exponential in the level p. It is therefore much preferable to stick to cubist metrics. This implies we avoid the use of joins as in [Bost et al. 1994], as those need a sheaf metrization on the whole of the ambient projective spaces, and we instead use plain Segre embeddings. The extra numerical cost essentially consists of the appearance of modest binomial coefficients, which do not significantly alter the quantitative bounds we eventually obtain.

We also need to work with projective models which are "almost" compactifications of relevant Néron models of our jacobians. This we do with the help of Moret-Bailly theory as introduced in Section 5.

Let us recall that there still is another approach for such arithmetic Bézout theorems which uses Chow forms [Philippon 1994; Rémond 2000]. That is however known to amount to working again with Faltings' height relative to the Fubini–Study metrics [Philippon 1994; Soulé 1991] that we said we cannot afford.

Finally, regarding generality, it would of course be desirable to have a proof available for arbitrary abelian varieties. Many of the present arguments are however quite particular to our application to $J_0(p)$. We therefore prefer working in our concrete setting from the beginning, instead of considering a somewhat artificial generality.

Proposition 6.1 (arithmetic Bézout theorem for $J_0(p)$). *Let $(J_0(p), \Theta)$ be defined over some number field K, endowed with the principal and symmetric polarization Θ. Let V and W be two irreducible K-subvarieties of $J_0(p)$, of dimension $d_V := \dim_K V$ and $d_W := \dim_K W$, respectively, such that*

$$d_V + d_W \leq g = \dim J_0(p)$$

and assume $V \cap W$ has dimension 0.

If P is an element of $(V \cap W)(K)$ then its Néron–Tate Θ-height satisfies

$$\mathrm{h}_\Theta(P) \le \frac{4^{d_V+d_W}}{2} \frac{(d_V + d_W + 1)!}{d_V! \, d_W!} \deg_\Theta(V) \deg_\Theta(W)[(d_W + 1)\mathrm{h}_\Theta(W) + (d_V + 1)\mathrm{h}_\Theta(V) + O(p \log p)].$$

$$(98)$$

Remark 6.2. The general aspect of the above release of arithmetic Bézout might look a bit different from the original ones, as can be found in [Bost et al. 1994]; this is due to the fact that our definition of the height of some cycle Y (see Section 5, (80)) amounts to dividing its height in the sense of [loc. cit.] by the product of the degree and absolute dimension of Y.

Let us first sketch the strategy of proof, which occupies the rest of this Section 6. We henceforth fix a prime number p and some perfect square integer $N := r^2$. (We shall eventually take $r = 2$.) We write $(\mathcal{J}, \mathcal{M})$ for the Moret-Bailly projective model of $(J_0(p), L(\Theta)^{\otimes N})$ given by Definition 5.4, relative to some given set of global sections $\mathcal{S}$ in $H^0(\mathcal{N}_{J,N}, \mathcal{L}(\Theta)^{\otimes N})$, of size N^g, to be described later (Lemma 6.5). That model is defined over some ring of integers $\mathcal{O}_K$. Consider the morphisms

$$
\begin{array}{ccc}
\mathcal{J} & \xrightarrow{\ \Delta\ } & \mathcal{J} \times \mathcal{J} \\
{\scriptstyle \mathcal{P}}\Big\downarrow & \searrow{\scriptstyle \iota} & \\
\mathbb{P}^n_{\mathcal{O}_K} \times \mathbb{P}^n_{\mathcal{O}_K} & \xrightarrow{\ S\ } & \mathbb{P}^{n^2+2n}_{\mathcal{O}_K}
\end{array}
\tag{99}
$$

where Δ is the diagonal map, $n = N^g - 1$, $\mathcal{P}$ is the product of two $\mathcal{S}$-embeddings $\mathcal{J} \overset{J}{\hookrightarrow} \mathbb{P}^n = \mathbb{P}^n_{\mathcal{O}_K}$ and the application $\iota : \mathcal{J} \times \mathcal{J} \to \mathbb{P}^{n^2+2n}$ is the composition of the Segre embedding S with $\mathcal{P}$. As sheaves,

$$S^*(\mathcal{O}_{\mathbb{P}^{n^2+2n}}(1)) = \mathcal{O}_{\mathbb{P}^n}(1) \otimes_{\mathcal{O}_K} \mathcal{O}_{\mathbb{P}^n}(1) \quad \text{and} \quad \mathcal{P}^*(\mathcal{O}_{\mathbb{P}^n}(1) \otimes_{\mathcal{O}_K} \mathcal{O}_{\mathbb{P}^n}(1)) = \mathcal{M} \otimes_{\mathcal{O}_K} \mathcal{M} =: \mathcal{M}^{\boxtimes 2}$$

so that

$$\iota^*(\mathcal{O}_{\mathbb{P}^{n^2+2n}}(1)) = \mathcal{M}^{\boxtimes 2}$$

and

$$\Delta^* \iota^* \mathcal{O}_{\mathbb{P}^{n^2+2n}}(1) = \mathcal{M} \otimes_{\mathcal{O}_{\mathcal{J}}} \mathcal{M} = \mathcal{M}^{\otimes 2}. \tag{100}$$

We naturally endow the sheaves $\mathcal{M}^{\boxtimes 2}$, $\mathcal{M}^{\otimes 2}$, and so on with the hermitian structures induced by the cubist metric on the various $\mathcal{M}_\sigma$ for $\sigma : K \hookrightarrow \mathbb{C}$, denoted by $\|\cdot\|_{\mathrm{cub}}$.

We then pick two copies $(x_i)_{0 \le i \le n}$ and $(y_j)_{0 \le j \le n}$ of the canonical basis of global sections for each $\mathcal{O}_{\mathbb{P}^n}(1)$ on the two factors of $\mathbb{P}^n_{\mathcal{O}_K} \times \mathbb{P}^n_{\mathcal{O}_K}$ of (99), which give our basis $\mathcal{S}$ by restriction to $\mathcal{J}$. Then we provide the sheaf $\mathcal{O}_{\mathbb{P}^{n^2+2n}}(1)$ on $\mathbb{P}^{n^2+2n}_{\mathcal{O}_K}$ with the basis of global sections $(z_{i,j})_{0 \le i,j \le n}$, each of which is mapped to $x_i \otimes_{\mathcal{O}_K} y_j$ under S^*. Define $\mathcal{D}$ as the diagonal linear subspace of $\mathbb{P}^{n^2+2n}_{\mathcal{O}_K}$ defined by the *linear equations* $z_{i,j} = z_{j,i}$ for all i and j.

Let $V, W \subseteq J = \mathcal{J}_K$ be two closed subvarieties over K. The support of $V \cap W$ is the same as that of $(\iota \circ \Delta)^{-1}(\mathcal{D} \cap \iota(V \times W))$. To bound from above the height of points in $V \cap W$ it is therefore sufficient to estimate Faltings' height of $\mathcal{D} \cap \iota(V \times W)$, relative to the hermitian line bundle $\mathcal{O}_{\mathbb{P}^{n^2+2n}}(1)_{|\iota(J \times J)}$ endowed with the cubist metric. As $\mathcal{D}$ is a linear subspace that height is essentially the same as that of $(V \times W)$,

up to an explicit error term which depends on the degree. In turn this error term is a priori linear in the number of (relevant) equations for $\mathcal{D}$, and this is way too high. But if one knows $V \cap W$ has dimension 0, it is enough to choose $(\dim V + \dim W)$ equations (up to perhaps increasing a bit the size of the set whose height we estimate), which makes the error term much smaller.

That is the basic strategy of proof for Proposition 6.1. To make it effective however we must control the "error terms" alluded to in the preceding lines, and those crucially depend on the supremum, on the set $\mathcal{S}$, of values for the cubist metric of global sections defining the projective embedding $\mathcal{J} \hookrightarrow \mathbb{P}^n_{\mathcal{O}_K}$. We shall build that $\mathcal{S}$ using theta functions as follows.

Recall Riemann's theta function on $J_0(p)$ introduced in Section 5C; see (89). Its usual analytic norm is

$$\|\theta(z)\|_{\mathrm{an}} := \det(\Im(\tau))^{1/4} \exp(-\pi y \Im(\tau)^{-1} y) |\theta(z)| \tag{101}$$

for $z = x + iy \in \mathbb{C}^g$ (see [Moret-Bailly 1990, (3.2.2)]). That analytic metric will have to be compared to the cubist one, about which we recall the following basic facts.

Let A be an abelian variety over a number field K, which extends to a semiabelian scheme $\mathcal{A}$ over the ring of integers $\mathcal{O}_K$. We endow $\mathcal{A}$ with a symmetric ample invertible sheaf $\mathcal{L}$. Define, for $I \subseteq \{1, 2, 3\}$, the projection $p_I : \mathcal{A}^3 \to \mathcal{A}$, $p_I(x_1, x_2, x_3) = \sum_{i \in I} x_i$. It is known to follow from the theorem of the cube [Moret-Bailly 1985b] that the sheaf $\mathcal{D}_3(\mathcal{L}) := \bigotimes_{I \subseteq \{1,2,3\}} p_I^* \mathcal{L}^{\otimes (-1)^{|I|}}$ is trivial on $\mathcal{A}^3$. Let us therefore fix an isomorphism $\phi : \mathcal{O}_{\mathcal{A}^3} \to \mathcal{D}_3(\mathcal{L})$. For every complex place σ of $\mathcal{O}_K$ one can endow $\mathcal{L}_\sigma$ with some cubist metric $\|\cdot\|_\sigma$ such that one obtains through ϕ the trivial metric on $\mathcal{O}_{\mathcal{A}^3}$. Each cubist metric $\|\cdot\|_\sigma$ is determined only up to multiplication by some constant factor so we perform the following rigidification to remove that ambiguity. If $0_{\mathcal{A}} : \operatorname{Spec}(\mathcal{O}_K) \to \mathcal{A}$ denotes the zero section, we replace $\mathcal{L}$ by $\mathcal{L} \otimes_{\mathcal{O}_K} (\pi^* 0_{\mathcal{A}}^* \mathcal{L}^{\otimes -1})$ on $\mathcal{A}$. Then

$$0_{\mathcal{A}}^*(\mathcal{L}) \simeq \mathcal{O}_K$$

and we demand that the $\|\cdot\|_\sigma$ be adjusted so that the above sheaf isomorphism is an isometry at each σ, where $\mathcal{O}_K$ is endowed with the trivial metric so that $\|1\| = 1$. This uniquely determines our cubist metrics $\|\cdot\|_\sigma$. Now by construction the hermitian sheaf $\overline{\mathcal{L}}$ on $\mathcal{A}$ defines a height h verifying the expected normalization condition h(0) = 0.

Having the same curvature form, the analytic and cubist metrics are known to differ by constant factors, at each complex place, on the theta sheaf, as we shall use in the proof of Lemma 6.4 below.

Recall we also defined in (91) a "meromorphic theta function $s_{\mathcal{J}^0}$ over $\overline{\overline{\mathbb{Z}}}$", which can be generalized; we have $[r]^* \mathcal{L}(\Theta)_{|\mathcal{N}_{J,1}} \simeq \mathcal{L}(\Theta)^{\otimes r^2}$ on $\mathcal{N}_{J,r}$ [Pazuki 2012, Proposition 5.1], so we define a global section

$$s_{\mathcal{M}} := ([r]^* t_{-\kappa}^*) s_{\mathcal{J}^0} \in H^0(\mathcal{N}_{J,r}, [r]^* \mathcal{L}(\Theta)_{\mathcal{O}_K}). \tag{102}$$

We will shortly show how to control the supremum of $\|s_{\mathcal{J}^0}\|_{\mathrm{cub}}$, therefore of $\|s_{\mathcal{M}}\|_{\mathrm{cub}}$, on $J_0(p)(\mathbb{C})$ (see Lemma 6.4). Writing $N = r^2$, we shall moreover fix the morphism $j_{\mathcal{M}} : \widetilde{\mathcal{N}}_{J,N} \to \mathcal{J} \hookrightarrow \mathbb{P}^n_{\mathcal{O}_K}$ of (86) by mapping the canonical coordinates $(x_i)_{0 \le i \le n}$ to sections (s_i) which will be translates by r-torsion points of a multiple of the above $s_{\mathcal{M}}$ by some constant, as explained in Lemma 6.5 and its proof.

This will allow us to control as well the supremum of those s_i, relative to the cubist metrics, on the complex base change of our abelian varieties, as is required by the proof of arithmetic Bézout theorems.

We now start the technical preparation for the proof of Proposition 6.1, for which we need some lemmas on the behavior of heights and degree under Segre maps, comparison between cubist and analytic metrics on theta functions, and estimates for all.

Lemma 6.3. *There is an infinite sequence* $(P_i)_{i \in \mathbb{N}}$ *of points in* $X_0(p)(\overline{\mathbb{Q}})$ *which are ordinary at all places dividing* p *and have everywhere integral* j*-invariant. Moreover their normalized theta height satisfies* $\mathrm{h}_\Theta(P_i - \infty + \frac{1}{2}\omega^0) = O(p^3)$, *with notations of Theorem 4.1.*

Proof. Let $(\zeta_i)_{\mathbb{N}}$ be a infinite sequence of roots of unity. One can assume none are congruent to some supersingular j-invariant in characteristic p, modulo any place of $\overline{\mathbb{Q}}$ above p. (Indeed, as the supersingular j-invariants are quadratic over $\mathbb{F}_p$, it is enough for instance to choose for the ζ_i some primitive ℓ_i-roots of unity, with ℓ_i running through the set of primes larger than $p^2 - 1$.) Lift each j-invariant equal to ζ_i to some point P_i in $X_0(p)(\overline{\mathbb{Q}})$. By construction, this makes a sequence of points with j-height $\mathrm{h}_j(P_i)$ equal to 0. As for their (normalized) theta height one sees from Theorem 4.1 that

$$\mathrm{h}_\Theta\left(P_i - \infty + \tfrac{1}{2}\omega^0\right) = \frac{1}{[K(P_i):\mathbb{Q}]}[P_i, \tilde{\omega}_\Theta]_{\mu_0} = \frac{-1}{[K(P_i):\mathbb{Q}]} \sum_{\sigma:K(P_i) \hookrightarrow \mathbb{C}} g \cdot g_{\mu_0}(\infty, \sigma(P_i)) + O(\log p)$$

as the contribution at finite places of $[P_i, \infty]$ is 0. It is therefore enough to bound the $|g_{\mu_0}(\infty, \sigma(P_i))|$.

Now $|j(P_i)|_\sigma = 1$ for all $\sigma : K(P_i) \hookrightarrow \mathbb{C}$, so the corresponding elements τ in the usual fundamental domain in Poincaré upper half-plane for $X_0(p)$ or $X(p)$ are absolutely bounded, and the same for the absolute values of $q_\tau = e^{2i\pi\tau}$. (For a useless explicit estimate of this bound one can check Corollary 2.2 of [Bilu and Parent 2011] which proposes $|q_\tau| \geq e^{-2500}$.) From this, running through the proof of Theorem 11.3.1 of [Edixhoven and de Jong 2011b], and adapting it to the case of $X_0(p)$ instead of $X_1(pl)$, we deduce that the $\sigma(P_i)$ do not belong to the open neighborhood, in the atlas of [loc. cit.], of the cusp ∞ in $X_0(p)(\mathbb{C})$. Therefore Proposition 10.13 of [Merkl 2011] applies and gives, with notations of that work,

$$|g_{\mu_0}(\infty, \sigma(P_i))| = |g_{\mu_0}(\infty, \sigma(P_i)) - h_\infty(\sigma(P_i))| = O(p^2) \tag{103}$$

(see Theorem 11.3.1 of [Edixhoven and de Jong 2011b] and its proof). □

Lemma 6.4. *Let* s_θ *be the "theta function over* $\mathbb{Z}$*", that is, the global section introduced just before* (90). *One has*

$$\sup_{J_0(p)(\mathbb{C})} (\log \|s_\theta\|_{\mathrm{cub}}) \leq O(p \log p). \tag{104}$$

Proof. Writing $s_{\theta,\mathbb{C}}(z) = C_\vartheta \cdot \theta(z + \kappa)$ as in (90), we shall bound from above both $|C_\vartheta|$ and the contribution of the difference between cubist and analytic metrics. Then we will use upper bounds for the analytic norm of the theta function due to P. Autissier and proven in the Appendix of the present paper.

We invoke again some key arguments of the proof of Proposition 5.8. For D in $J_0(p)(\mathbb{C})$, written as the linear equivalence class of some divisor $\sum_{i=1}^{g}(P_i - \infty)$ on $X_0(p)(\mathbb{C})$, we indeed once more consider

the embedding

$$\iota_{\kappa-D} : X_0(p) \hookrightarrow J_0(p), \quad P \mapsto \mathrm{cl}(P - \infty + \kappa - D)$$

as in Proposition 5.8. For such a D whose P_i are assumed to belong to $X_0(p)(\overline{\mathbb{Q}})$, we recall (92) that

$$\mathrm{h}_\Theta(\iota_{\kappa-D}(P)) = \frac{1}{[K(P, D) : \mathbb{Q}]} \left[P, \sum_i P_i + \Phi_D + c_D X_\infty \right]_{\mu_0}.$$

If the P_i all have everywhere ordinary reduction, as will be the case in (105) below, the vertical divisor Φ_D will contribute at most $O(\log p)$ to the height of points (see Remark 3.4).

Note that we can fulfill condition (95) considering only points P_i of same type as occurring in Lemma 6.3 (which, in particular, are ordinary and have integral j-invariants), because those P_i make a Zariski-dense subset of $X_0(p)(\overline{\mathbb{Q}})$ (and the onto-ness of the map $X_0(p)^{(g)} \xrightarrow{\iota_\infty^g} J_0(p)$). We therefore conclude as in the proof of Proposition 5.8 that $\mathrm{div}(\iota_{\kappa-D}^*(s_\theta))$ has indeed to be $\left(\sum_i P_i + \Phi_D \right)$ on $X_0(p)^{\mathrm{smooth}}_{\mathcal{O}_K}$.[6]

On the other hand, for some of those choices of $(P_i)_{1 \le i \le g}$, our $\mathbb{Z}$-theta function s_θ does not vanish at $\iota_{\kappa-D}(\infty)(\mathbb{C})$, so $\mathrm{h}_\Theta(\iota_{\kappa-D}(\infty))$ can also be computed as the Arakelov degree:

$$\mathrm{h}_\Theta(\iota_{\kappa-D}(\infty)) = \widehat{\deg}(\infty^* \iota_{\kappa-D}^*(\mathcal{L}(\Theta))).$$

Integrality of the P_i shows the intersection numbers $[\infty, P_i]$ have trivial nonarchimedean contribution. The only finite contribution to our Arakelov degree therefore comes from intersection with vertical components, that is, if K_D is a sufficiently large field over which D is defined, then for a set of elements $(z_\sigma)_{\sigma:K_D \hookrightarrow \overline{\mathbb{Q}}}$ which lift $\sigma(-D)$ in the complex tangent space of $J_0(p)$ to 0 one has

$$\mathrm{h}_\Theta(\iota_{\kappa-D}(\infty)) = \widehat{\deg}(0^*_{\mathcal{J}_0(p)}(t^*_{\kappa-D}\mathcal{L}(\Theta))) = \widehat{\deg}(0^*_{\mathcal{J}_0(p)}(t^*_{-D}\mathcal{L}(\Theta_\kappa)))$$

$$= -\frac{1}{[K_D : \mathbb{Q}]} \sum_{K_D \xrightarrow{\sigma} \mathbb{C}} \log\|s_\theta(z_\sigma)\|_{\mathrm{cub}} + O(\log p),$$

whence, as $s_{\theta,\mathbb{C}}(z) = C_\vartheta \cdot \theta(z + \kappa)$,

$$\log|C_\vartheta| = -\mathrm{h}_\Theta(\iota_{\kappa-D}(\infty)) - \frac{1}{[K_D(\kappa) : \mathbb{Q}]} \sum_{K_D(\kappa) \xrightarrow{\sigma} \mathbb{C}} \log\|\theta((z+\kappa)_\sigma)\|_{\mathrm{cub}} + O(\log p). \tag{105}$$

Following [Gaudron and Rémond 2014b, paragraph 8] we now write $J_0(p)(\mathbb{C}) = \mathbb{C}^g/(\mathbb{Z}^g + \tau\mathbb{Z}^g)$ for τ in Siegel's fundamental domain, write $z \in \mathbb{C}^g$ as $z = \tau \cdot p + q$ for $p, q \in \mathbb{R}^g$, and introduce the function $F : \mathbb{C}^g \to \mathbb{C}$ defined as

$$F(z) = \det(2\Im(z))^{1/4} \sum_{n \in \mathbb{Z}^g} \exp(i\pi {}^t(n+p)\tau(n+p) + 2i\pi {}^t nq).$$

One then has $|F(z)| = 2^{g/4}\|\theta(z)\|_{\mathrm{an}}$. Indeed there is a constant $A \in \mathbb{R}^*_+$ such that $|F(z)| = A \cdot \|\theta(z)\|_{\mathrm{an}}$ (see the end of proof of Lemma 8.3 of [loc. cit.]), $\int_{J_0(p)(\mathbb{C})} |F|^2 \, d\nu = 1$ (where $d\nu$ is the probability

---

[6]Although we shall not use this, one can check that $\mathrm{h}_\Theta(\iota_{\kappa-D}(\infty)) = \|-(\sum_i P_i - \infty) + \frac{1}{2}\omega^0\|^2_\Theta = O(p^5)$ by Lemma 6.3 and (79).

Haar measure on $J_0(p)(\mathbb{C})$, see [loc. cit., Lemma 8.2(1)]), and $\int_{J_0(p)(\mathbb{C})} \|\theta(z)\|_{\mathrm{an}}^2 \, d\nu = 2^{-g/2}$ (see, e.g., [Moret-Bailly 1990, (3.2.1) and (3.2.2)]). Therefore Lemme 8.3 of [Gaudron and Rémond 2014b] gives, using definitions of [loc. cit., Théorème 8.1],

$$-\frac{1}{[K_D(\kappa):\mathbb{Q}]} \sum_{K_D(\kappa)\overset{\sigma}{\hookrightarrow}\mathbb{C}} \left(\log\|\theta((z+\kappa)_\sigma)\|_{\mathrm{an}} + \tfrac{g}{4}\log 2\right) \le \mathrm{h}_\Theta(\iota_{\kappa-D}(\infty)) + \tfrac{1}{2}\mathrm{h}_F(J_0(p)) + \tfrac{g}{4}\log 2\pi.$$

Remember Faltings' height of $J_0(p)$ is known to satisfy $\mathrm{h}_F(J_0(p)) = O(p\log p)$ by [Ullmo 2000, Théorème 1.2]. (We remark that Ullmo's normalization of Faltings' height differs from that of Gaudron and Rémond, but the difference term is linear in $g = O(p)$ so the bound $O(p\log p)$ remains valid for the above $\mathrm{h}_F(J_0(p))$).) Writing $\|\cdot\|_{\mathrm{cub}} = e^\varphi\|\cdot\|_{\mathrm{an}}$ we therefore see that (105) implies

$$\log|C_\vartheta| + \varphi \le \tfrac{1}{2}\mathrm{h}_F(J_0(p)) + O(p) \le O(p\log p).$$

Given this upper bound for $e^\varphi|C_\vartheta|$ we can now go the other way round to derive an upper bound for $\|s_\theta\|_{\mathrm{cub}} = C_\vartheta \cdot \|\theta(z+\kappa)\|_{\mathrm{cub}}$, by using estimates for analytic theta functions. For any principally polarized complex abelian variety whose complex invariant τ is chosen within Siegel's fundamental domain F_g, Autissier's result in the Appendix (Proposition A.1 below) indeed gives, with notations as in (101), that

$$\frac{1}{\det(\Im(\tau))^{1/4}}\|\theta(z)\|_{\mathrm{an}} = \exp(-\pi y\Im(\tau)^{-1}y)|\theta(z)| \le g^{g/2}. \tag{106}$$

We refer to the Appendix for a bound which is slightly sharper.[7]

As for the factor $\det(\Im(\tau))^{1/4}$, Lemma 11.2.2 of [Edixhoven and de Jong 2011b] gives the general result

$$\det(\Im(z))^{1/2} \le \frac{(2g)!V_{2g}}{2^g V_g} \prod_{g+1\le i\le 2g} \lambda_i,$$

where for any k we write V_k for the volume of the unit ball in $\mathbb{R}^k$ endowed with its standard Euclidean structure, and the λ_r are the successive minima, relative to the Riemann form, of the lattice $\Lambda = \mathbb{Z}^g + \tau\cdot\mathbb{Z}^g$. To bound the λ_i we need to invoke an avatar of [loc. cit., Lemma 11.2.3]. But the very same proof shows that for any integer N, the group $\Gamma_0(N)$ has a set of generators having entries of absolute value less or equal to the very same bound $N^6/4$. (That term could be improved, but this would have an invisible impact on the final bounds so we here content ourselves with it.) We can therefore rewrite the proof of Lemma 11.2.4 verbatim. This gives that Λ is generated by elements having naive hermitian norm $\|x\|_E^2$ less or equal to gp^{46}. Finally, in our case the Gram matrix is diagonal (no 2×2-blocks, at the difference of Lemma 11.1.4 of [loc. cit.]) so Lemma 11.2.5 a fortiori holds: if $\|\cdot\|_P$ denotes the hermitian product on $\mathbb{C}^g$ induced by the polarization, $\|\cdot\|_P^2 \le e^{4\pi}/\pi\|\cdot\|_E^2$. This allows to conclude as in p. 228 of [loc. cit.]:

$$\left(\prod_{i=g+1}^{2g} \lambda_i\right)^2 \le \left(\frac{e^{4\pi}}{\pi}gp^{46}\right)^g$$

---

[7]Works of Igusa and Edixhoven and de Jong [2011b, pp. 231–232] give $1/\det(\Im(\tau))^{1/4}\|\theta(z)\|_{\mathrm{an}} \le 2^{3g^3+5g}$.

so that

$$\log(\det(\Im(\tau))) \leq O(p \log p)$$

and combining with (106),

$$\log\|\theta(z)\|_{\mathrm{an}} \leq O(p \log p).$$

Putting everything together finally yields

$$\sup_{z \in J_0(p)(\mathbb{C})} \log\|s_{\theta,\mathbb{C}}(z)\|_{\mathrm{cub}} = \sup_{z \in J_0(p)(\mathbb{C})} \log\|C_\vartheta \cdot \theta(z+\kappa)\|_{\mathrm{cub}}$$

$$= (\log|C_\vartheta| + \varphi) + \sup_{z \in J_0(p)(\mathbb{C})} \log\|\theta(z+\kappa)\|_{\mathrm{an}}$$

$$\leq O(p \log p). \qquad \square$$

Lemma 6.5. *Assume the same hypothesis and notations as in Definition 5.4. After possibly making some finite base extension one can pick a set $\mathcal{S}$ in $H^0(\mathcal{N}_{J,4}, \mathcal{L}(\Theta)^{\otimes 4})$ of 4^g global sections $(s_i)_{1 \leq i \leq 4^g}$, which span $\mathcal{L}(\Theta)^{\otimes 4}$ on $\mathcal{N}_{J,4}[1/2p]$, and verify*

$$\sup_{J_0(p)} (\log\|s_i\|_{\mathrm{cub}}) \leq O(p \log p). \tag{107}$$

Proof. We fix $N = r^2 = 4$ for the construction of a good model as in Definition 5.4. Up to making a base extension, we can assume $L(\Theta)^{\otimes 4}$ and $[2]^* L(\Theta)$ have cubist extensions $\mathcal{L}(\Theta)^{\otimes 4}$ and $[2]^* \mathcal{L}(\Theta))$ on $\mathcal{N}_{J,4}$, respectively. As Θ is symmetric one knows there is an isomorphism $[2]^* \mathcal{L}(\Theta) \to \mathcal{L}(\Theta)^{\otimes 4}$ which actually is an isometry [Pazuki 2012, Proposition 5.1], by which we identify those two objects from now on. On the other hand, every element x of $J_0(p)[4](\overline{\mathbb{Q}}) = J_0(p)[4](K)$ defines a section $\tilde{x}$ in $\mathcal{N}_{J,4}(\mathrm{Spec}(\mathcal{O}_K))$. Letting $t_{\tilde{x}}$ denote the translation by $\tilde{x}$ on $\mathcal{N}_{J,4}$ we have

$$t_{\tilde{x}}^* \mathcal{L}(\Theta)^{\otimes 4} \simeq \mathcal{L}(\Theta)^{\otimes 4}. \tag{108}$$

(This is indeed true over $\mathbb{C}$ by Lemma 2.4.7.c of [Birkenhake and Lange 2004], hence over K, then over $\mathrm{Spec}(\mathcal{O}_K)$ by uniqueness of cubist extensions.) The interpretation as Néron–Tate heights shows that as $\mathcal{L}(\Theta)$ is endowed with its cubist metric, this isomorphism even is an isometry. Recall the section $s_\mathcal{M}$ defined in (102), belonging to $H^0(\mathcal{N}_{J,2}, [2]^* \mathcal{L}(\Theta))$. Up to making an extension to some larger base ring of integer, we may assume $s_\mathcal{M}$ extends as a meromorphic section on $\mathcal{N}_{J,4}$ and Proposition 5.8, which gives estimates on the poles of $s_{\mathcal{J}^0}$ at bad components, implies that $s_\mathcal{M}$ is actually holomorphic (has no pole on the new components) after multiplication by some power C_1 of p with $\log C_1 = O(p \log p)$. We can therefore define a set $(s_i)_{1 \leq i \leq 4^g}$ in $H^0(\mathcal{N}_{J,4}, [2]^* \mathcal{L}(\Theta))$ made of 4^g elements of shape

$$s_i := t_{\tilde{x}_i}^* C_1 \cdot s_\mathcal{M} \tag{109}$$

for $\tilde{x}_i$ running through a set of representatives, in $J_0(p)[4](K)$, of $J_0(p)[4]/J_0(p)[2]$. Note that one can explicitly lift $s_\mathcal{M}$ on the complex tangent space at 0 of $J_0(p)(\mathbb{C})$ as

$$s_{\mathcal{M},\mathbb{C}}(z) = C_\vartheta \cdot \theta(2 \cdot z) \tag{110}$$

where C_ϑ is defined in the proof of Lemma 6.4 and the $s_{i,\mathbb{C}}$ are constant multiple of the basis denoted by $h_{\vec{a},\vec{b}}(\vec{z})$ in [Mumford 1984], Proposition II.1.3.iii on p. 124.[8] From here, Lemma 6.4 and Proposition 5.8 give (107).

By the theory of theta functions [Pazuki 2012, Proposition 2.5 and its proof; Mumford 1966; Moret-Bailly 1985b, Chapitre VI] the s_i make a generic basis of global sections, which span $\mathcal{L}(\Theta)^{\otimes 4}$ over $\mathrm{Spec}(\mathcal{O}_K[1/2p])$. $\qquad\square$

Lemma 6.6. *Let V and W be two closed K-subvarieties, with dimension d_V and d_W respectively, of a smooth projective variety A over a number field K, endowed with an ample sheaf M. Assume the flat projective scheme $(\mathcal{A}, \mathcal{M})$ over $\mathrm{Spec}(\mathcal{O}_K)$, with $\mathcal{M}$ an hermitian sheaf on $\mathcal{A}$, is a model for (A, M). Let $\mathcal{V}$ and $\mathcal{W}$ be the Zariski closure in $\mathcal{A}$ of V and W respectively. Then, with definitions as in [Bost et al. 1994, §3.1],*

$$(c_1(M^{\boxtimes 2})^{d_V+d_W} \mid (V \times W)) = \binom{d_V+d_W}{d_V}(c_1(M)^{d_V} \mid V)(c_1(M)^{d_W} \mid W) \qquad (111)$$

and

$$(\hat{c}_1(\mathcal{M}^{\boxtimes 2})^{d_V+d_W+1} \mid \mathcal{V} \times \mathcal{W})$$
$$= \binom{d_V+d_W+1}{d_V}(c_1(M)^{d_V} \mid V)(\hat{c}_1(\mathcal{M})^{d_W+1} \mid \mathcal{W}) + \binom{d_V+d_W+1}{d_W}(\hat{c}_1(\mathcal{M})^{d_V+1} \mid \mathcal{V})(c_1(M)^{d_W} \mid W). \quad (112)$$

Remark 6.7. Equation (111) can be read as

$$\deg_{M^{\boxtimes 2}}(V \times W) = \binom{d_V+d_W}{d_V}\deg_M(V)\deg_M(W).$$

Equation (112) in turn fits with Zhang's interpretation (83) in terms of essential minima, compare the proof of Proposition 6.1 below.

Proof of Lemma 6.6. For (111), one can realize it is elementary, or refer to Lemme 2.2 of [Rémond 2010], or proceed as follows. Using (2.3.18), (2.3.19), and Proposition 3.2.1(iii) of [Bost et al. 1994], and noticing

$$c_1(M^{\boxtimes 2}) = c_1(M) \times \mathbf{1} + \mathbf{1} \times c_1(M)$$

(and same with $\hat{c}_1(\mathcal{M})$ and $\hat{c}_1(\mathcal{M}^{\boxtimes 2})$ instead) one computes

---

[8]where it seems by the way that the expression "$h_{\vec{a},\vec{b}}(\vec{z}) = \vartheta\begin{bmatrix}\vec{a}/k \\ \vec{b}/k\end{bmatrix}(\ell \cdot \vec{z}, \Omega)$" should read "$\cdots = \vartheta\begin{bmatrix}\vec{a}/k \\ \vec{b}/k\end{bmatrix}(k \cdot \vec{z}, \Omega)$" (notations of [loc. cit.]).

$$(c_1(M^{\boxtimes 2})^{d_V+d_W} \mid (V \times W)) = \left(\sum_{k=0}^{d_V+d_W} \binom{d_V+d_W}{k} c_1(M)^k \times c_1(M)^{d_V+d_W-k} \mid V \times W \right)$$

$$= \sum_{k=0}^{d_V+d_W} \binom{d_V+d_W}{k} (c_1(M)^k \times c_1(M)^{d_V+d_W-k} \mid V \times W)$$

$$= \sum_{k=0}^{d_V+d_W} \binom{d_V+d_W}{k} (c_1(M)^k \mid V)(c_1(M)^{d_V+d_W-k} \mid W)$$

$$= \binom{d_V+d_W}{k} (c_1(M)^{d_V} \mid V)(c_1(M)^{d_W} \mid W),$$

where the last equality comes from the fact that the only nonzero term in the line before occurs for $k = d_V$.

An analogous computation, using [Bost et al. 1994, (2.3.19)], can be used for the arithmetic degree:

$$(\hat{c}_1(\mathcal{M}^{\boxtimes 2})^{d_V+d_W+1} \mid \mathcal{V} \times \mathcal{W})$$

$$= \sum_{k=0}^{d_V+d_W+1} \binom{d_V+d_W+1}{k} (\hat{c}_1(\mathcal{M})^k \times \hat{c}_1(\mathcal{M})^{d_V+d_W+1-k} \mid \mathcal{V} \times \mathcal{W})$$

$$= \binom{d_V+d_W+1}{d_V} (c_1(M)^{d_V} \mid V)(\hat{c}_1(\mathcal{M})^{d_W+1} \mid \mathcal{W}) \binom{d_V+d_W+1}{d_W} (\hat{c}_1(\mathcal{M})^{d_V+1} \mid \mathcal{V})(c_1(M)^{d_W} \mid W). \quad \square$$

For the rest of this Section we fix the model $(\mathcal{J}, \mathcal{M})$ for $(J_0(p), \Theta)$ (see (99)) as the one built with the set $\mathcal{S}$ of $N^g = 4^g$ sections provided by Lemma 6.5. Before settling the proof of the arithmetic Bézout theorem, we need a last lemma on the comparison between the projective height on $(\mathcal{J}, \mathcal{M})$ and its Néron–Tate avatar.

Lemma 6.8. *Up to translation by torsion points, the projective height $h_\mathcal{M}$ on points in $J_0(p)(\overline{\mathbb{Q}})$ (associated with the good model $(\mathcal{J}, \mathcal{M})$) differs from the Néron–Tate theta-height $4h_\Theta$ by an error term of shape $O(p \log p)$.*

Proof. Lemma 6.5 implies that the elements of $\mathcal{S}$ extend as holomorphic sections to *any* component of the Néron model $\overline{\mathcal{N}}$ of $J_0(p)$ over $\overline{\mathbb{Z}}$ (see (109)). As remarked in the proof of Lemma 6.5, Mumford's algebraic theory of theta-functions implies that the sections in $\mathcal{S}$ do define a projective embedding of $\overline{\mathcal{N}}$ over $\overline{\mathbb{Z}}[1/2p]$; the only fibers of $\overline{\mathcal{N}}$ over $\overline{\mathbb{Z}}$ where base points for $\mathcal{S}$ can show up are above 2 and p. If one seeks to approximate the Néron–Tate height of a given point P in $J_0(p)(\overline{\mathbb{Q}})$ by the projective height of our good model $(\mathcal{J}, \mathcal{M})$, one needs the section of the Néron model $\overline{\mathcal{N}}$ defined by P to avoid those base points, or at least control their length.

Given P in $J_0(p)(\overline{\mathbb{Q}})$, we claim one can translate P by some torsion point in $J_0(p)(\overline{\mathbb{Q}})$ so that the translated new point $P + t$ does avoid base points in characteristic 2. Indeed, choose a Galois extension $F/\mathbb{Q}$ such that the base locus is defined over $\mathrm{Spec}(\mathcal{O}_F \otimes \mathbb{F}_2)$. Summing-up, as divisors, all the Galois conjugates of that base locus in each fiber of characteristic 2, one obtains a constant cycle C_κ, in each fiber at κ, which is defined over $\mathbb{F}_2$. (In our case one actually could have taken $F = \mathbb{Q}$.) Density of torsion points then shows that one can replace our point P by $P + t$, for some torsion point t, such that $P + t$

does not belong to C_{κ_0} for some κ_0, then for all κ of characteristic 2 because C_κ is constant. This proves our claim. Now in characteristic p, we know from Proposition 5.8 again that possible base points have length at most $O(p)$, which gives an estimate of size $O(p \log p)$ for the difference error term between projective height on $\mathcal{J}$ and Néron–Tate height [Pazuki 2012, Proposition 4.1]. □

Proof of Proposition 6.1. Before proceeding we will allow ourselves for this proof only, in order to not overcomplicate the computations, to work with heights defined as in [Bost et al. 1994, §3.1]. Namely, for $\mathcal{Y}$ a cycle of dimension $(d+1)$ in a regular arithmetic variety endowed with a hermitian sheaf $\mathcal{F}$, we multiply our definition (80) of its height by degree and absolute dimension and we set

$$\mathrm{h}'_{\mathcal{F}}(\mathcal{Y}) = \frac{(\hat{c}_1(\mathcal{F})^{d+1} \mid \mathcal{Y})}{[K : \mathbb{Q}]}.$$

Note that h and h' coincide on K-rational points, in which case we might use either notation.

Construction (99) gives a $\mathbb{Q}$-embedding $V \times W \overset{\iota}{\hookrightarrow} \mathbb{P}^{n^2+2n}$ via a Segre map. We set

$$s_{\underline{i},\underline{j}} := \iota^*(z_{\underline{i},\underline{j}} - z_{\underline{j},\underline{i}})$$

for all $(\underline{i}, \underline{j})$, and denote by $\mathcal{O}_N$ the ambient line bundle $\iota^*(\mathcal{O}_{\mathbb{P}^{n^2+2n}}(1)) = \mathcal{M}^{\boxtimes 2}$ as before (100). (Recall we will eventually specialize to $N = 4$.) Set also $O_N := \mathcal{O}_N \otimes \mathbb{Q}$. We intersect $\iota(V \times W)$ with one of the $\mathrm{div}(z_{\underline{i_0},\underline{j_0}} - z_{\underline{j_0},\underline{i_0}})_{\mathbb{Q}}$ such that the two cycles meet properly; define

$$J_1 = \mathrm{div}(s_{\underline{i_0},\underline{j_0}}{}_{\mathbb{Q}}) \cap (V \times W)$$

in the generic fiber $(J_0(p) \times J_0(p))_{\mathbb{Q}}$. As $\mathrm{div}(z_{\underline{i_0},\underline{j_0}} - z_{\underline{j_0},\underline{i_0}})$ is a projective hyperplane we have by definition

$$\deg_{O_N}(J_1) = \deg_{O_N}(V \times W).$$

For the same linearity reason, a similar statement is true for heights. Indeed, let $\mathcal{V}$ and $\mathcal{W}$ denote the schematic closure in $\mathcal{J}$ of V and W respectively, and $\mathcal{J}_1$ the schematic closure of J_1 in $\mathcal{J} \times \mathcal{J}$, which satisfies

$$\mathrm{h}'_{\mathcal{O}_N}(\mathcal{J}_1) \leq \mathrm{h}'_{\mathcal{O}_N}(\mathrm{div}(s_{\underline{i_0},\underline{j_0}}) \cap (\mathcal{V} \times \mathcal{W}))$$

(as there might be vertical components in the intersection of the right-hand side which do not intervene in the left, and contribute positively to the height).

Proposition 3.2.1(iv) of [Bost et al. 1994] gives, with notations of [loc. cit.], that

$$\mathrm{h}'_{\mathcal{O}_N}(\mathrm{div}(s_{\underline{i_0},\underline{j_0}}) \cap (\mathcal{V} \times \mathcal{W})) = \mathrm{h}'_{\mathcal{O}_N}(\mathcal{V} \times \mathcal{W}) + \frac{1}{[K : \mathbb{Q}]} \sum_{\sigma:K \hookrightarrow \mathbb{C}} \int_{(V \times W)_\sigma(\mathbb{C})} \log\|s_{\underline{i_0},\underline{j_0}}{}_{\mathbb{C}}\| c_1(\mathcal{O}_N)^{d_V + d_W} \quad (113)$$

where $\|\cdot\| = \|\cdot\|_{\mathrm{cub}}$ shall denote the cubist metric, or the metric induced by the cubist metric on products or powers of relevant sheaves. To estimate the last integral we note that at any point of $(V \times W)_\sigma(\mathbb{C})$ and

for any $(\underline{i}, \underline{j})$,

$$\|s_{\underline{i},\underline{j}}\| = \|z_{\underline{i},\underline{j}} - z_{\underline{j},\underline{i}}\|_{\mathcal{M}^{\boxtimes 2}} \leq \|z_{\underline{i},\underline{j}}\|_{\mathcal{M}^{\boxtimes 2}} + \|z_{\underline{j},\underline{i}}\|_{\mathcal{M}^{\boxtimes 2}}$$

$$\leq \|x_{\underline{i}}\|_{\mathcal{M}}\|y_{\underline{j}}\|_{\mathcal{M}} + \|x_{\underline{j}}\|_{\mathcal{M}}\|y_{\underline{i}}\|_{\mathcal{M}} \leq 2(\sup_i \|x_i\|_{\mathcal{M}})^2$$

$$\leq \exp(2\log(\sup\|s_i\|_{\mathrm{cub}}) + \log 2)$$

with notations of Lemma 6.5. Setting $M_{\mathcal{J},\mathcal{M}} = \log(\sup\|s_i\|_{\mathrm{cub}})$ we obtain

$$\mathrm{h}'_{\mathcal{O}_N}(\mathcal{J}_1) \leq \mathrm{h}'_{\mathcal{O}_{N_0}}(\mathcal{V} \times \mathcal{W}) + (2M_{\mathcal{J},\mathcal{M}} + \log 2)\deg_{\mathcal{O}_N}(V \times W).$$

Call I_1 one of the reduced irreducible components of J_1 containing the point $\iota(\Delta(P))$ of $V \cap W$ considered in the statement of Proposition 6.1 and let $\mathcal{I}_1$ denote its Zariski closure in $\mathcal{J}$. It has $\mathcal{O}_N$-height (and degree) less than or equal to those of $\mathcal{J}_1$, so that again

$$\mathrm{h}'_{\mathcal{O}_N}(\mathcal{I}_1) \leq \mathrm{h}'_{\mathcal{O}_N}(\mathcal{V} \times \mathcal{W}) + (2M_{\mathcal{J},\mathcal{M}} + \log 2)\deg_{\mathcal{O}_N}(V \times W)$$

and we can iterate the process with I_1 in place of $V \times W$: we obtain some $J_2, \mathcal{J}_2, I_2, \mathcal{I}_2$ such that

$$\mathrm{h}'_{\mathcal{O}_N}(\mathcal{I}_2) \leq \mathrm{h}'_{\mathcal{O}_N}(\mathcal{I}_1) + (2M_{\mathcal{J},\mathcal{M}} + \log 2)\deg_{\mathcal{O}_N}(I_1)$$

$$\leq \mathrm{h}'_{\mathcal{O}_N}(\mathcal{V} \times \mathcal{W}) + 2(2M_{\mathcal{J},\mathcal{M}} + \log 2)\deg_{\mathcal{O}_N}(V \times W).$$

(The only obstruction to this step is if all the $s_{\underline{k},\underline{l}}$ vanish on I_1, which implies it is contained in the diagonal of $J_0(p) \times J_0(p)$ - so that $I_1 = \iota(\Delta(P))$ by construction and that means we are already done.) Processing, one builds a sequence $(\mathcal{I}_k)$ of integral closed subschemes of $\mathcal{J} \times \mathcal{J}$, with decreasing dimension, such that the last step gives

$$\mathrm{h}'_{\mathcal{O}_N}(\mathcal{I}_{d_V+d_W}) \leq \mathrm{h}'_{\mathcal{O}_N}(\mathcal{V} \times \mathcal{W}) + (d_V + d_W)(2M_{\mathcal{J},\mathcal{M}} + \log 2)\deg_{O_N}(V \times W).$$

Now

$$\mathrm{h}'_{\mathcal{O}_N}(\mathcal{I}_{d_V+d_W}) \geq \mathrm{h}'_{\mathcal{O}_N}(\Delta(P, P)) = \mathrm{h}'_{\mathcal{M}^{\boxtimes 2}}(P) = \mathrm{h}_{\mathcal{L}^{\otimes 2N}}(P) = 2N\,\mathrm{h}_\Theta(P) + O(p\log p),$$

for $\mathrm{h}_\Theta(P)$ the Néron–Tate normalized theta height. Indeed the statement of the present Proposition 6.1 is invariant by translation of every object by some fixed torsion point, so that one can apply Lemma 6.8.

Using Lemma 6.6 and Corollary 5.6 and writing $\mathrm{h}'_\Theta(Y) = (\dim(Y) + 1)\deg_\Theta(Y)\mathrm{h}_\Theta(Y)$ we therefore obtain

$$2N\mathrm{h}_\Theta(P)$$

$$\leq N^{d_v+d_W+1}\left[(d_W+1)\binom{d_V+d_W+1}{d_V}\mathrm{h}'_\Theta(W)\deg_\Theta(V) + (d_V+1)\binom{d_V+d_W+1}{d_W}\mathrm{h}'_\Theta(V)\deg_\Theta(W)\right]$$

$$+ N^{d_V+d_W}(d_V+d_W)(2M_{\mathcal{J},\mathcal{M}} + \log 2)\binom{d_V+d_W}{d_V}\deg_\Theta(V)\deg_\Theta(W) + O(p\log p).$$

From here, fixing $N = 4$, the bound $M_{\mathcal{J},\mathcal{M}} \leq O(p\log p)$ (Lemma 6.5) concludes the proof, after expressing quantities h'_Θ back into h_Θ. $\qquad\square$

That arithmetic Bézout theorem will be our principal tool in the sequel.

7. Height bounds for quadratic points on $X_0(p)$

Proposition 7.1. *Let $\iota : X \hookrightarrow J$ be some Albanese map from a curve (of positive genus) over some field K to its jacobian J. Let $\pi : J \to A$ be some quotient of J, with $\dim(A) > 1$, and X' be the normalization of the image $\pi \circ \iota(X)$ of X in A. Then the map $\pi' : X \to X'$ induced by $\pi \circ \iota$ verifies*

$$\deg(\pi') \leq \frac{\dim(J) - 1}{\dim(A) - 1}.$$

Proof. The map $\pi \circ \iota$ induces an inclusion of function fields which defines the map $\pi' : X \to X'$. If J' is the jacobian of X', Albanese functoriality says that π factorizes through surjective morphisms $J \to J' \to A$. Hurwitz formula writes

$$\deg(\pi') = \frac{\dim(J) - 1 - \frac{1}{2} \deg R}{\dim(J') - 1}$$

for R the ramification divisor of π', whence the result. □

Lemma 7.2. *For all large enough prime p, let $X := X_0(p)$ and $\pi_e : J_0(p) \twoheadrightarrow J_e$ be the projection. Let*

$$\iota_{P_0} : X_0(p) \hookrightarrow J_0(p), \quad P \mapsto \mathrm{cl}(P - P_0)$$

for some P_0 in $X_0(p)(\overline{\mathbb{Q}})$ such that $w_p(P_0) = P_0$ (there are roughly $\sqrt{p}$ such points, [Gross 1987a, Proposition 3.1]) and set $\varphi_e := \pi_e \circ \iota_{P_0}$. Then:

- *If $a \in J_e(\mathbb{Q})$ is some (necessarily torsion) point, the equality $\varphi_e(X_0(p)) = a - \varphi_e(X_0(p))$ implies*

$$\varphi_e(X_0(p)) = a + \varphi_e(X_0(p)) \tag{114}$$

  and $a = 0$.

- *If d is the degree of the map $X_0(p) \to \widetilde{\varphi_e(X_0(p))}$ to the normalization of $\varphi_e(X_0(p))$, then d is either 1, 3 or 4.*

- *Assuming moreover Brumer's conjecture (see (21) and (22)) equality (114) implies $d = 1$ for large enough p.*

Proof. Notice first that, by our choice of P_0 (whence ι), and because J_e belongs to the w_p-minus part of $J_0(p)$, one has

$$\varphi_e(w_p(P)) = w_p(\varphi_e(P)) = -\varphi_e(P),$$

for all $P \in X_0(p)(\mathbb{C})$, whence equality (114). So let n be the order of a, which also is that of the automorphism "translation by a restricted to $\varphi_e(X_0(p))$" . We remark that the degree d cannot equal 2, as otherwise the extension of fraction fields $K(X_0(p))/K(\varphi_e(X_0(p)))$ would be Galois and $X_0(p)$ would possess an involution different from w_p, which it does not by Ogg's theorem [1977] (or even [Kenku and Momose 1988]). If $d = 1$, the same reason that $\mathrm{Aut}(X_0(p)) = \langle w_p \rangle$ implies that $n = 1$. Let now X' be the normalization of the quotient of $\varphi_e(X_0(p))$ by the automorphism $P \mapsto P + a$, that is, the image of $\varphi_e(X_0(p))$ by the quotient morphism $J_e \twoheadrightarrow J_e/\langle a \rangle$. Let π be the composed map $J_0(p) \xrightarrow{\varphi_e} J_e \to J_e/\langle a \rangle$.

The degree of $X_0(p) \to X'$ is $d \cdot n$ and Proposition 7.1 together with the left part of inequalities (23) implies

$$d \cdot n \leq \frac{g-1}{\left(\frac{1}{4} - o(1)\right)g - 1} \leq 4 + o(1)$$

for large enough p. This shows that if $d = 3$ or 4 one still has $a = 0$, whence the proposition's first two statements. Assuming (22) we have $d \cdot n < 3$, so that $d = 1$ and $a = 0$ by previous arguments. $\qquad\square$

Remark 7.3. Replace, in Lemma 7.2, the map $X_0(p) \to J_e$ by $X_0(p) \xrightarrow{\varphi} J_0(p)^-$ (by which the former factorizes, by the way). The above proof shows that the map $X_0(p) \to \varphi(X_0(p))$ is of generic degree 1 (independently on any conjecture), but of course it need not be injective on points: a finite number of points can be mapped together to singular points on $\varphi(X_0(p))$. In our case one checks those are among the Heegner points P such that $P = w_p(P)$ (for which we again refer to [Gross 1987a, Proposition 3.1]). Indeed, the endomorphism of $J_0(p)$ defined by multiplication by $(1 - w_p)$ factorizes through φ and $\cdot(1 - w_p)$ is the map considered in (4) and what follows, inducing multiplication by 2 on tangent spaces. Therefore, if P maps to a multiple point of $\varphi(X_0(p))$, it also maps to a multiple point of $(1 - w_p) \circ \iota(X_0(p))$. Now assuming $X_0(p)$ has gonality larger than 2 (which is true as soon as $p > 71$ [Ogg 1974, Theorem 2]) the equality $\mathrm{cl}((1 - w_p)P) = \mathrm{cl}((1 - w_p)P')$ in $J_0(p)$, for some P' on $X_0(p)$ different from P, implies $P = w_p P$ and $P' = w_p P'$. That is, P and P' are Heegner points.

Lemma 7.4. *Suppose P belongs to $X_0(p^2)(K)$ for some quadratic number field K, and P is not a complex multiplication point. Then for one of the two natural degeneracy morphisms π from $X_0(p^2)$ to $X_0(p)$, the point $Q := \pi(P)$ in $X_0(p)(K)$ does not define a $\mathbb{Q}$-valued point of the quotient curve $X_0^+(p) := X_0(p)/w_p$.*

Proof. Using the modular interpretation, we write $P = (E, C_{p^2})$ for E an elliptic curve over K and C_{p^2} a cyclic K-isogeny of degree p^2, from which we obtain the two points $Q_1 := (E, p \cdot C_{p^2})$ and $Q_2 := (E/p \cdot C_{p^2}, C_{p^2} \bmod p \cdot C_{p^2})$ in $X_0(p)(K)$. Assume both Q_1 and Q_2 do define elements of $X_0^+(p)(\mathbb{Q})$. If σ denotes a generator of $\mathrm{Gal}(K/\mathbb{Q})$ we then have

$$w_p(Q_1) = (E/p \cdot C_{p^2}, E[p] \bmod p \cdot C_{p^2}) \simeq \sigma(Q_1)$$

and

$$w_p(Q_2) = (E/C_{p^2}, E[p] + C_{p^2} \bmod C_{p^2}) \simeq \sigma(Q_2).$$

Therefore $E \simeq {}^\sigma(E/p \cdot C_{p^2}) \simeq E/C_{p^2}$, which means E has complex multiplication. $\qquad\square$

We can now conclude with the main result of this paper.

Theorem 7.5. *There is an integer C such that the following holds. If p is a prime number such that (22), the weak form of Brumer's conjecture, holds and P is a quadratic point of $X_0(p)$ (that is, P is an element of $X_0(p)(K)$ for some quadratic number field K) which does not come from $X_0(p)^+(\mathbb{Q})$, then its j-height satisfies*

$$\mathrm{h}_j(P) < C \cdot p^5 \log p. \tag{115}$$

If P is a quadratic point of $X_0(p^2)$ then the same conclusion holds without further assumption apart from (22).

Proof. In the case P is a quadratic point of $X_0(p^2)$, by Lemma 7.4, one can deduce from P a point P' in $X_0(p)(K)$ which does not induce an element of $X_0^+(p)(\mathbb{Q})$ and whose j-height, say, is equal to $h_j(P) + O(\log p)$ for an explicit function $O(\log p)$ (see, e.g., [Pellarin 2001, inequality (51) on p. 240; Bilu et al. 2013, Proposition 4.4(i)]). Replace P by P' if necessary. By Theorem 4.6 it is now sufficient to prove that $h_\Theta(P - \infty) = O(p^5 \log p)$.

Keep the notation of Lemma 7.2. By construction, the point

$$a := \varphi_e(P) + \varphi_e(^\sigma P) = \varphi_e(P) - \varphi_e(w_p(^\sigma P)) = \varphi_e(P - w_p(^\sigma P))$$

is torsion. First assume $a = 0$. Set $X^{(2),-} := \{\iota_\infty(x) - \iota_\infty(y), (x, y) \in X_0(p)^2\}$ as in Proposition 5.3. Recall from Section 2 that $\tilde{I}_{J_e^\perp, N_e^\perp} : J_e^\perp \to \tilde{J}_e^\perp$ is the map defined as in (3), that $\iota_{\tilde{J}_e^\perp, N_e^\perp}$ is the embedding $\tilde{J}_e^\perp \hookrightarrow J_0(p)$, and denote by $[N_{\tilde{J}_e^\perp}]_{\tilde{J}_e^\perp}$ the multiplication by $N_{\tilde{J}_e^\perp}$ restricted to $\tilde{J}_e^\perp$. As in (8) and before Corollary 5.7 we use our pseudoprojections and define

$$\tilde{X}^{(2),-} := \iota_{\tilde{J}_e^\perp, N_e^\perp}[N_{\tilde{J}_e^\perp}]_{\tilde{J}_e^\perp}^{-1} \tilde{I}_{J_e^\perp, N_e^\perp} \pi_{J_e^\perp}(X^{(2),-}).$$

Then $P - w_p(^\sigma P)$ belongs to $X^{(2),-} \cap \tilde{J}_e^\perp$, and even to the intersection of surfaces (in the generic fiber)

$$X^{(2),-} \cap \tilde{X}^{(2),-}.$$

Recall (see (8)) that $\tilde{X}^{(2),-}$ is a priori highly nonconnected, being the inverse image of multiplication by $N_{\tilde{J}_e^\perp}$ in $\tilde{J}_e^\perp$ of the (irreducible) surface $\tilde{I}_{J_e^\perp, N_e^\perp} \pi_{J_e^\perp}(X^{(2),-})$. However, in what follows we can replace $\tilde{X}^{(2),-}$ by one of its connected components containing $P - w_p(^\sigma P)$. Denote that component by $\tilde{X}_P^{(2),-}$.

By construction, the theta degree and height of $\tilde{X}_P^{(2),-}$, as an irreducible subvariety of $J_0(p)$ endowed with Θ, are those of $\pi_{J_e^\perp}(X^{(2),-}) = X_{e^\perp}^{(2),-}$ relative to the only natural hermitian sheaf of $J_e^\perp$, that is, the $\Theta_e^\perp = \Theta_{J_e^\perp}$ described in paragraph 2A2. One can therefore apply Proposition 5.3 to obtain that all theta degrees are $O(p^2)$, all Néron–Tate theta heights are $O(\log p)$. We claim the dimension of $(X^{(2),-} \cap \tilde{X}_P^{(2),-})$ is zero. That intersection indeed corresponds to pairs of distinct points on $X_0(p)$ having same image (0) under φ_e. On the other hand, Brumer's conjecture implies $X_0(p) \to \varphi_e(X_0(p))$ has generic degree one (see Lemma 7.2), so our intersection points correspond to singular points on $\varphi_e(X_0(p))$, which of course make a finite set.

We therefore are in position to apply our arithmetic Bézout theorem (Proposition 6.1), which yields $h_\Theta(P - w_p(^\sigma P)) \le O(p^5 \log p)$. The two points $(P - \infty)$ and $(w_p(^\sigma P) - \infty)$ have same Θ-height (recall w_p is an isometry on $J_0(p)$ for h_Θ, compare the end of Remark 4.3), and are by hypothesis different, so one can apply them Mumford's repulsion principle (Proposition 5.9) to obtain

$$h_\Theta(P - \infty) \le O(p^5 \log p). \tag{116}$$

Let us finally deal with the case when the torsion point $a = \varphi_e(P) + \varphi_e(^\sigma P)$ is nonzero. We adapt the previous argument: pick a lift $\tilde{a} \in J_0(p)(\overline{\mathbb{Q}})$ of a by $\pi_e^\perp$ which also is torsion, and let $t_{\tilde{a}}$ be the translation by $\tilde{a}$ in $J_0(p)$. Replace $(P - w_p(^\sigma P))$ by $t_{\tilde{a}}^*(P - w_p(^\sigma P))$, $X^{(2),-}$ by $t_{\tilde{a}}^* X^{(2),-}$ and $\tilde{X}^{(2),-}$ by

$$t_{\tilde{a}}^* \widetilde{X}^{(2),-} = \iota_{\tilde{J}_e^\perp, N_e^\perp} [N_{\tilde{J}_e^\perp}]^{-1}_{\tilde{J}_e^\perp} \tilde{I}_{J_e^\perp, N_e^\perp} \pi_{J_e^\perp} (t_{\tilde{a}}^* X^{(2),-}).$$

Now $t_{\tilde{a}}^*(P - w_p(^\sigma P))$ belongs to $(t_{\tilde{a}}^* X^{(2),-} \cap t_{\tilde{a}}^* \widetilde{X}^{(2),-})$. The theta degree and height of $t_{\tilde{a}}^* X^{(2),-}$ and $t_{\tilde{a}}^* \widetilde{X}^{(2),-}$ (or rather, as above, some connected component $t_{\tilde{a}}^* \widetilde{X}_P^{(2),-}$ of it containing $t_{\tilde{a}}^*(P - w_p(^\sigma P)))$ are the same as for the former objects in the case $a = 0$. The fact that the intersection

$$t_{\tilde{a}}^* X^{(2),-} \cap t_{\tilde{a}}^* \widetilde{X}_P^{(2),-}$$

is zero-dimensional comes from the fact that otherwise, we would have $\varphi_e(X_0(p)) = a - \varphi_e(X_0(p))$, a contradiction with our present hypothesis $a \neq 0$ by Lemma 7.2. The height bound for P is therefore the same as (116). $\qquad \square$

Corollary 7.6. *Under the assumptions of Theorem 7.5, if p is a large enough prime number and P is a quadratic point of $X_0(p^\gamma)$ for some integer γ, such that P is not a cusp nor a complex multiplication point, then $\gamma \leq 10$.*

Proof. Let P be a point in $X_0(p^\gamma)(K)$, which is not a cusp nor a CM point, for some quadratic number field K. Then the isogeny bounds of [Gaudron and Rémond 2014b, Theorem 1.4] imply there is some real κ with

$$p^\gamma < \kappa (\mathrm{h}_j(P))^2.$$

Now Theorem 7.5 gives that there is some absolute real constant B such that, if $p \geq B$ then $\gamma \leq 10$. $\quad \square$

Remark 7.7. A similar (but technically simpler) approach for the morphism $X_0(p) \to J_e$ over $\mathbb{Q}$ should give (independently of any conjecture) a bound of shape $O(p^3 \log p)$ for the j-height of $\mathbb{Q}$-rational (noncuspidal) points of $X_0(p)$ (which are known not to exist for $p > 163$ by Mazur's theorem). The same should apply for $\mathbb{Q}$-points of $X_{\mathrm{split}}(p)$ (and here again, we obtain a weak version of known results).

Actually, sharpening results directly coming from Section 4 (that is, avoiding the use of Bézout) might even yield the full strength of the above results about $X_0(p)(\mathbb{Q})$ and $X_{\mathrm{split}}(p)(\mathbb{Q})$, with more straightforward (unconditional) proofs.

Appendix: An upper bound for the theta function
by Pascal Autissier

In this appendix, I give a new upper bound for the norm of the classical theta function on any complex abelian variety. This result, apart from its role in the present paper (see Section 6), has been used by Wilms [2017] to bound the Green–Arakelov function on curves.

Result. Let g be a positive integer. Write $\mathbb{H}_g$ for the Siegel space of symmetric matrices $Z \in M_g(\mathbb{C})$ such that $\mathrm{Im}\, Z$ is positive definite. To every $Z \in \mathbb{H}_g$ is associated the theta function defined by,

$$\theta_Z(z) = \sum_{m \in \mathbb{Z}^g} \exp(i\pi\, {}^t m Z m + 2i\pi\, {}^t m z), \quad \forall z \in \mathbb{C}^g,$$

and its norm defined by,

$$\|\theta_Z(z)\| = \sqrt[4]{\det Y} \exp(-\pi\, {}^t y Y^{-1} y)|\theta_Z(z)|, \quad \forall z = x + iy \in \mathbb{C}^g,$$

where $Y = \mathrm{Im}\, Z$.

My contribution here is the following:

Proposition A.1. *Let $Z \in \mathbb{H}_g$ and assume that Z is Siegel-reduced. Put $c_g = (g+2)/2$ if $g \leq 3$ and $c_g = \big((g+2)/(\pi\sqrt{3})\big)^{g/2}(g+2)/2$ if $g \geq 4$. The upper bound $\|\theta_Z(z)\| \leq c_g(\det \mathrm{Im}\, Z)^{1/4}$ holds for every $z \in \mathbb{C}^g$.*

Let us remark that $c_g \leq g^{g/2}$ for every $g \geq 2$. In comparison, Edixhoven and de Jong [2011b, p. 231] obtained the statement of Proposition A.1 with c_g replaced by 2^{3g^3+5g}.

Proof. Fix a positive integer g. Denote by $\mathbb{S}_g$ the set of symmetric matrices $Y \in M_g(\mathbb{R})$ that are positive definite. Let us recall a special case of the functional equation for the theta function (see [Mumford 1983, (5.6), p. 195]): for every $Y \in \mathbb{S}_g$ and every $z \in \mathbb{C}^g$, one has

$$\theta_{iY^{-1}}(-iY^{-1}z) = \sqrt{\det Y} \exp(\pi\, {}^t z Y^{-1} z)\theta_{iY}(z). \tag{117}$$

Lemma A.2. *Let $Z \in \mathbb{H}_g$ and $z \in \mathbb{C}^g$. Putting $Y = \mathrm{Im}\, Z$, one has the inequality*

$$\|\theta_Z(z)\| \leq \|\theta_{iY}(0)\| = \theta_{iY}(0)\sqrt[4]{\det Y}.$$

Proof. Put $y = \mathrm{Im}\, z$. One has

$$|\theta_Z(z)| = \left| \sum_{m \in \mathbb{Z}^g} \exp(i\pi\, {}^t m Z m + 2i\pi\, {}^t m z) \right| \leq \sum_{m \in \mathbb{Z}^g} |\exp(i\pi\, {}^t m Z m + 2i\pi\, {}^t m z)| = \theta_{iY}(iy),$$

that is, $\|\theta_Z(z)\| \leq \|\theta_{iY}(iy)\|$. The functional equation (117) gives $\|\theta_{iY^{-1}}(Y^{-1}y)\| = \|\theta_{iY}(iy)\|$, and one deduces

$$\|\theta_Z(z)\| \leq \|\theta_{iY^{-1}}(Y^{-1}y)\|. \tag{118}$$

Applying again (118) with Z replaced by iY^{-1} and z by $Y^{-1}y$, one gets

$$\|\theta_{iY^{-1}}(Y^{-1}y)\| \leq \|\theta_{iY}(0)\|.$$

Whence the result.   □

Let $Y \in \mathbb{S}_g$. Define $\lambda(Y) = \min_{m \in \mathbb{Z}^g - \{0\}} {}^t m Y m$. For every $t \in \mathbb{R}_+^*$, put

$$f_Y(t) = \theta_{itY}(0) = \sum_{m \in \mathbb{Z}^g} \exp(-\pi t\, {}^t m Y m).$$

Lemma A.3. *Let* $Y \in \mathbb{S}_g$ *and put* $\lambda = \lambda(Y)$. *The following properties hold*:

(a) *The function* $\mathbb{R}_+^* \to \mathbb{R}$ *that maps* t *to* $t^{g/2} f_Y(t)$ *is increasing.*

(b) *One has the estimate* $f_Y((g+2)/(2\pi\lambda)) \leq (g+2)/2$.

Proof.

(a) The functional equation (117) implies $\sqrt{\det Y} t^{g/2} f_Y(t) = f_{Y^{-1}}(1/t)$ for every $t \in \mathbb{R}_+^*$; conclude by remarking that $f_{Y^{-1}}$ is decreasing.

(b) Part (a) gives $\frac{d}{dt}[t^{g/2} f_Y(t)] \geq 0$, that is, $\frac{g}{2t} f_Y(t) \geq -f_Y'(t)$ for every $t > 0$. On the other hand,

$$-\frac{1}{\pi} f_Y'(t) = \sum_{m \in \mathbb{Z}^g} {}^t m Y m \exp(-\pi t \, {}^t m Y m) \geq \sum_{m \in \mathbb{Z}^g - \{0\}} \lambda \exp(-\pi t \, {}^t m Y m) = \lambda[f_Y(t) - 1].$$

One infers $\frac{g}{2t} f_Y(t) \geq \pi \lambda[f_Y(t) - 1]$. Choosing $t = (g+2)/(2\pi\lambda)$, one obtains the result. $\qquad\square$

Proposition A.4. *Let* $Y \in \mathbb{S}_g$. *Putting* $\lambda = \lambda(Y)$, *one has the upper bound*

$$\theta_{iY}(0) \leq \frac{g+2}{2} \max\left[\left(\frac{g+2}{2\pi\lambda} \right)^{g/2}, 1 \right].$$

Proof. Put $t = (g+2)/(2\pi\lambda)$. If $t \geq 1$, then Lemma A.3(a) implies the inequality $f_Y(1) \leq t^{g/2} f_Y(t)$. If $t \leq 1$, then $f_Y(1) \leq f_Y(t)$ since f_Y is decreasing. In any case, one obtains

$$\theta_{iY}(0) = f_Y(1) \leq \max(t^{g/2}, 1) f_Y(t).$$

Conclude by applying Lemma A.3(b). $\qquad\square$

Now, to prove Proposition A.1 from Lemma A.2 and Proposition A.4, it suffices to observe that if $Z \in \mathbb{H}_g$ is Siegel-reduced, then $\lambda(\operatorname{Im} Z) \geq \frac{\sqrt{3}}{2}$ (see [Igusa 1972, Lemma 15, p. 195]).

Acknowledgments

---

[9] Although, as goes without saying, he bears no responsibility for the mistakes which remain.

[10] They have already been used by R. Wilms [2017]; see the introduction to the Appendix.

curves led there by Bas Edixhoven, Jean-Marc Couveignes and their coauthors. We also benefited from a visit to Leiden University in June of 2015, where we had very enlightening discussions with Bas, Peter Bruin, Robin de Jong and David Holmes.

Olga Balkanova, Samuel Le Fourn and Guillaume Ricotta helped a lot with references and explanations about some results of analytic number theory, and Jean-Benoît Bost kindly answered some questions about his own arithmetic Bézout theorem.

Finally, many thanks are due to the referee for her or his substantial and helpful work.

References

[Abbes 1997] A. Abbes, "Hauteurs et discrétude (d'après L. Szpiro, E. Ullmo et S. Zhang)", exposé 825, 4, pp. 141–166 in *Séminaire Bourbaki,* 1996/1997, Astérisque **245**, Soc. Mat. de France, Paris, 1997. MR Zbl

[Abbes and Ullmo 1995] A. Abbes and E. Ullmo, "Comparaison des métriques d'Arakelov et de Poincaré sur $X_0(N)$", *Duke Math. J.* **80**:2 (1995), 295–307. MR Zbl

[Aryasomayajula 2013] A. Aryasomayajula, *Bounds for Green's functions on hyperbolic Riemann surfaces of finite volume*, Ph.D. thesis, Humboldt-Universität zu Berlin, 2013, Available at https://edoc.hu-berlin.de/handle/18452/17480. Zbl

[Bilu and Parent 2011] Y. Bilu and P. Parent, "Runge's method and modular curves", *Int. Math. Res. Not.* **2011**:9 (2011), 1997–2027. MR Zbl

[Bilu et al. 2013] Y. Bilu, P. Parent, and M. Rebolledo, "Rational points on $X_0^+(p^r)$", *Ann. Inst. Fourier* (*Grenoble*) **63**:3 (2013), 957–984. MR Zbl

[Birkenhake and Lange 2004] C. Birkenhake and H. Lange, *Complex abelian varieties*, 2nd ed., Grundlehren der Mathematischen Wissenschaften **302**, Springer, 2004. MR Zbl

[Bost 1996] J.-B. Bost, "Intrinsic heights of stable varieties and abelian varieties", *Duke Math. J.* **82**:1 (1996), 21–70. MR Zbl

[Bost et al. 1994] J.-B. Bost, H. Gillet, and C. Soulé, "Heights of projective varieties and positive Green forms", *J. Amer. Math. Soc.* **7**:4 (1994), 903–1027. MR Zbl

[Bruin 2014] P. Bruin, "Explicit bounds on automorphic and canonical Green functions of Fuchsian groups", *Mathematika* **60**:2 (2014), 257–306. MR Zbl

[Brumer 1995] A. Brumer, "The rank of $J_0(N)$", pp. 3, 41–68 in *Columbia University Number Theory Seminar* (New York, 1992), Astérisque **228**, Soc. Math. France, Paris, 1995. MR Zbl

[Checcoli et al. 2016] S. Checcoli, F. Veneziano, and E. Viada, "The explicit Mordell conjecture for families of curves", 2016. arXiv

[Chinburg 1986] T. Chinburg, "An introduction to Arakelov intersection theory", pp. 289–307 in *Arithmetic geometry* (Storrs, Conneticut, 1984), edited by G. Cornell and J. H. Silverman, Springer, 1986. MR Zbl

[Deligne and Rapoport 1973] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques", pp. 143–316 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **349**, Springer, Berlin, 1973. MR Zbl

[Ducros 2007] A. Ducros, "Espaces analytiques p-adiques au sens de Berkovich", exposé 958, viii, pp. 137–176 in *Séminaire Bourbaki,* 2005/2006, Astérisque **311**, Soc. Mat. de France, Paris, 2007. MR Zbl

[Edixhoven and Couveignes 2011] B. Edixhoven and J.-M. Couveignes (editors), *Computational aspects of modular forms and Galois representations*, Annals of Mathematics Studies **176**, Princeton University Press, 2011. MR Zbl

[Edixhoven and de Jong 2011a] B. Edixhoven and R. de Jong, "Applying Arakelov theory", pp. 187–201 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR

[Edixhoven and de Jong 2011b] B. Edixhoven and R. de Jong, "Bounds for Arakelov invariants of modular curves", pp. 217–256 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR

[Edixhoven and de Jong 2011c] B. Edixhoven and R. de Jong, "Short introduction to heights and Arakelov theory", pp. 79–94 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR

[Faltings 1991] G. Faltings, "Diophantine approximation on abelian varieties", *Ann. of Math.* (2) **133**:3 (1991), 549–576. MR Zbl

[Gaudron and Rémond 2014a] E. Gaudron and G. Rémond, "Polarisations et isogénies", *Duke Math. J.* **163**:11 (2014), 2057–2108. MR Zbl

[Gaudron and Rémond 2014b] E. Gaudron and G. Rémond, "Théorème des périodes et degrés minimaux d'isogénies", *Comment. Math. Helv.* **89**:2 (2014), 343–403. MR Zbl

[Griffiths and Harris 1978] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley, New York, 1978. MR Zbl

[Gross 1984] B. H. Gross, "Heegner points on $X_0(N)$", pp. 87–105 in *Modular forms* (Durham, 1983), edited by R. A. Rankin, Horwood, Chichester, New York, 1984. MR Zbl

[Gross 1987a] B. H. Gross, "Heegner points and the modular curve of prime level", *J. Math. Soc. Japan* **39**:2 (1987), 345–362. MR Zbl

[Gross 1987b] B. H. Gross, "Heights and the special values of L-series", pp. 115–187 in *Number theory* (Montreal, Que., 1985), edited by H. Kisilevsky and J. Labute, CMS Conf. Proc. **7**, Amer. Math. Soc., Providence, RI, 1987. MR Zbl

[Hindry and Silverman 2000] M. Hindry and J. H. Silverman, *Diophantine geometry*, Graduate Texts in Mathematics **201**, Springer, 2000. MR Zbl

[Igusa 1972] J.-i. Igusa, *Theta functions*, Springer, 1972. MR Zbl

[Iwaniec and Sarnak 2000] H. Iwaniec and P. Sarnak, "The non-vanishing of central values of automorphic L-functions and Landau–Siegel zeros", *Israel J. Math.* **120**:part A (2000), 155–177. MR Zbl

[Iwaniec et al. 2000] H. Iwaniec, W. Luo, and P. Sarnak, "Low lying zeros of families of L-functions", *Inst. Hautes Études Sci. Publ. Math.* 91 (2000), 55–131. MR Zbl

[de Jong 2018] R. de Jong, "Néron–Tate heights of cycles on Jacobians", *J. Algebraic Geom.* **27**:2 (2018), 339–381. MR Zbl

[de Jong and Shokrieh 2018] R. de Jong and F. Shokrieh, "Tropical moments of tropical Jacobians", preprint, 2018. arXiv

[Jorgenson and Kramer 2006] J. Jorgenson and J. Kramer, "Bounds on canonical Green's functions", *Compos. Math.* **142**:3 (2006), 679–700. MR Zbl

[Kenku and Momose 1988] M. A. Kenku and F. Momose, "Automorphism groups of the modular curves $X_0(N)$", *Compositio Math.* **65**:1 (1988), 51–80. MR Zbl

[Kowalski et al. 2000] E. Kowalski, P. Michel, and J. Vanderkam, "Non-vanishing of high derivatives of automorphic L-functions at the center of the critical strip", *J. Reine Angew. Math.* **526** (2000), 1–34. MR Zbl

[Larson and Vaintrob 2014] E. Larson and D. Vaintrob, "Determinants of subquotients of Galois representations associated with abelian varieties", *J. Inst. Math. Jussieu* **13**:3 (2014), 517–559. MR Zbl

[Le Fourn 2016] S. Le Fourn, "Surjectivity of Galois representations associated with quadratic ℚ-curves", *Math. Ann.* **365**:1-2 (2016), 173–214. MR Zbl

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, 2002. MR Zbl

[Mazur 1977] B. Mazur, "Modular curves and the Eisenstein ideal", *Inst. Hautes Études Sci. Publ. Math.* 47 (1977), 33–186. MR Zbl

[Menares 2008] R. Menares, *Nombres d'intersection arithmétiques et opérateurs de Hecke sur les courbes modulaires $X_0(N)$*, Ph.D. thesis, l'Université de Paris-Sud-Orsay, 2008, Available at http://tel.archives-ouvertes.fr/tel-00360171.

[Menares 2011] R. Menares, "Correspondences in Arakelov geometry and applications to the case of Hecke operators on modular curves", *Manuscripta Math.* **136**:3-4 (2011), 501–543. MR Zbl

[Merkl 2011] F. Merkl, "An upper bound for Green functions on Riemann surfaces", pp. 203–215 in *Computational aspects of modular forms and Galois representations*, edited by B. Edixhoven and J.-M. Couveignes, Ann. of Math. Stud. **176**, Princeton Univ. Press, 2011. MR

[Michel and Ullmo 1998] P. Michel and E. Ullmo, "Points de petite hauteur sur les courbes modulaires $X_0(N)$", *Invent. Math.* **131**:3 (1998), 645–674. MR Zbl

[Mikhalkin and Zharkov 2008] G. Mikhalkin and I. Zharkov, "Tropical curves, their Jacobians and theta functions", pp. 203–230 in *Curves and abelian varieties*, edited by V. Alexeev et al., Contemp. Math. **465**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Momose 1995] F. Momose, "Isogenies of prime degree over number fields", *Compositio Math.* **97**:3 (1995), 329–348. MR Zbl

[Moret-Bailly 1985a] L. Moret-Bailly, "Métriques permises", pp. 29–87 in *Seminar on arithmetic bundles: the Mordell conjecture* (Paris, 1983/84), Astérisque **127**, Soc. Mat. de France, Paris, 1985. MR Zbl

[Moret-Bailly 1985b] L. Moret-Bailly, "Pinceaux de variétés abéliennes", pp. 266 Astérisque **129**, Soc. Mat. de France, Paris, 1985. MR Zbl

[Moret-Bailly 1990] L. Moret-Bailly, "Sur l'équation fonctionnelle de la fonction thêta de Riemann", *Compositio Math.* **75**:2 (1990), 203–217. MR Zbl

[Mumford 1966] D. Mumford, "On the equations defining abelian varieties, I", *Invent. Math.* **1** (1966), 287–354. MR Zbl

[Mumford 1983] D. Mumford, *Tata lectures on theta, I*, Progress in Mathematics **28**, Birkhäuser Boston, 1983. MR Zbl

[Mumford 1984] D. Mumford, *Tata lectures on theta, II*, Progress in Mathematics **43**, Birkhäuser Boston, 1984. MR Zbl

[Ogg 1974] A. P. Ogg, "Hyperelliptic modular curves", *Bull. Soc. Math. France* **102** (1974), 449–462. MR Zbl

[Ogg 1977] A. P. Ogg, "Über die Automorphismengruppe von $X_0(N)$", *Math. Ann.* **228**:3 (1977), 279–292. MR Zbl

[Pazuki 2012] F. Pazuki, "Theta height and Faltings height", *Bull. Soc. Math. France* **140**:1 (2012), 19–49. MR Zbl

[Pellarin 2001] F. Pellarin, "Sur une majoration explicite pour un degré d'isogénie liant deux courbes elliptiques", *Acta Arith.* **100**:3 (2001), 203–243. MR Zbl

[Philippon 1994] P. Philippon, "Sur des hauteurs alternatives, II", *Ann. Inst. Fourier* (*Grenoble*) **44**:4 (1994), 1043–1065. MR Zbl

[Rémond 2000] G. Rémond, "Décompte dans une conjecture de Lang", *Invent. Math.* **142**:3 (2000), 513–545. MR Zbl

[Rémond 2010] G. Rémond, "Nombre de points rationnels des courbes", *Proc. Lond. Math. Soc.* (3) **101**:3 (2010), 759–794. MR Zbl

[Royer 2001] E. Royer, "Petits zéros de fonctions L de formes modulaires", *Acta Arith.* **99**:2 (2001), 147–172. MR Zbl

[Soulé 1991] C. Soulé, "Géométrie d'Arakelov et théorie des nombres transcendants", pp. 355–371 in *Journées Arithmétiques* (Luminy, 1989), Astérisque **198-200**, Soc. Mat. de France, 1991. MR Zbl

[Thuillier 2005] A. Thuillier, *Théorie du potentiel sur les courbes en géométrie non archimédienne. Applications à la théorie d'Arakelov*, Ph.D. thesis, Université de Rennes 1, 2005, Available at https://tel.archives-ouvertes.fr/file/index/docid/48750/filename/tel-00010990.pdf.

[Ullmo 2000] E. Ullmo, "Hauteur de Faltings de quotients de $J_0(N)$, discriminants d'algèbres de Hecke et congruences entre formes modulaires", *Amer. J. Math.* **122**:1 (2000), 83–115. MR Zbl

[Wilms 2017] R. Wilms, "New explicit formulas for Faltings' delta-invariant", *Invent. Math.* **209**:2 (2017), 481–539. MR Zbl

[Zarhin and Manin 1972] J. G. Zarhin and J. I. Manin, "Height on families of abelian varieties", *Mat. Sb.* (*N.S.*) **89(131)** (1972), 171–181, 349. In Russian; translated in *Math. USSR-Sb* **18** (1972), 169–179. MR Zbl

[Zhang 1993] S. Zhang, "Admissible pairing on a curve", *Invent. Math.* **112**:1 (1993), 171–193. MR Zbl

[Zhang 1995] S. Zhang, "Positive line bundles on arithmetic varieties", *J. Amer. Math. Soc.* **8**:1 (1995), 187–221. MR Zbl

pierre.parent@math.u-bordeaux.fr   *Institut de Mathématiques de Bordeaux, Université Bordeaux, Talence, France*

pascal.autissier@math.u-bordeaux.fr   *Institut de Mathématiques de Bordeaux, Université Bordeaux, Talence, France*

A formula for the Jacobian of a genus one curve of arbitrary degree

Tom Fisher

We extend the formulae of classical invariant theory for the Jacobian of a genus one curve of degree $n \leq 4$ to curves of arbitrary degree. To do this, we associate to each genus one normal curve of degree n, an $n \times n$ alternating matrix of quadratic forms in n variables, that represents the invariant differential. We then exhibit the invariants we need as homogeneous polynomials of degrees 4 and 6 in the coefficients of the entries of this matrix.

Introduction

Let C be a smooth curve of genus one defined over a field K. Its Jacobian is an elliptic curve E defined over the same field K. However it is only if C has a K-rational point that C and E are isomorphic over K. Starting with equations for C we would like to compute a Weierstrass equation for E.

Let D be a K-rational divisor on C of degree $n \geq 1$. It is natural to split into cases according to the value of n. If $n = 1$ then C has a K-rational point, and our task is that of writing an elliptic curve in Weierstrass form. If $n \geq 2$ then the complete linear system $|D|$ defines a morphism $C \to \mathbb{P}^{n-1}$. Explicitly, the map is given by $(f_1 : \cdots : f_n)$, where $f_1, \ldots, f_n$ are a basis for the Riemann–Roch space $\mathcal{L}(D)$. If $n = 2$ then C is a double cover of $\mathbb{P}^1$ and is given by an equation of the form $y^2 = F(x_1, x_2)$, where F is a binary quartic. In this case Weil [1954; 1983] showed that the classical invariants of the binary quartic F give a formula for the Jacobian.

If $n \geq 3$ then the morphism $C \to \mathbb{P}^{n-1}$ is an embedding. The image is a *genus one normal curve* of degree n. The word *normal* refers to the fact C is projectively normal (see for example [Hulek 1986, Proposition IV.1.2]), i.e., if H is the divisor of a hyperplane section then the natural map

$$S^d \mathcal{L}(H) \to \mathcal{L}(dH) \tag{1}$$

is surjective for all $d \geq 1$. If $n = 3$ then $C \subset \mathbb{P}^2$ is a smooth plane cubic, say with equation $F(x_1, x_2, x_3) = 0$. The invariants of a ternary cubic F were computed by Aronhold [1858], and again Weil (in the notes to [Weil 1954] in his collected papers) showed that these give a formula for the Jacobian. If $n = 4$ then $C \subset \mathbb{P}^3$ is the complete intersection of two quadrics. If we represent these quadrics by 4×4 symmetric matrices A and B, then $F(x_1, x_2) = \det(Ax_1 + Bx_2)$ is a binary quartic. The invariants of this binary

quartic again give a formula for the Jacobian. For further details of these formulae in the cases $n = 2, 3, 4$, see [An et al. 2001; Artin et al. 2005; Fisher 2008].

If $n = 5$ then $C \subset \mathbb{P}^4$ is no longer a complete intersection, and indeed the homogeneous ideal is generated by 5 quadrics. The Buchsbaum–Eisenbud structure theorem [1982; 1977] shows that these quadrics may be written as the 4×4 Pfaffians of a 5×5 alternating matrix of linear forms. The space of all such matrices is a 50-dimensional affine space, with a natural action of $\mathrm{GL}_5 \times \mathrm{GL}_5$. In [Fisher 2008] we computed generators for the ring of invariants and showed that they again give a formula for the Jacobian. In fact the invariants are too large to write down as explicit polynomials, so instead we gave a practical algorithm for evaluating them (based in part on the case $n = 5$ of Proposition 9.3). More recently, B. Gross [2011] gave a uniform description of the invariants in the cases $n = 2, 3, 4, 5$, using results of Vinberg, although this does not appear to give any way of evaluating the invariants in the case $n = 5$.

In this paper we extend these formulae for the Jacobian to genus one normal curves of arbitrary degree.

Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree $n \geq 3$. Since C has genus one, the space of regular differentials on C has dimension 1, say spanned by ω. We call ω an *invariant differential*, since geometrically it is invariant under all translation maps. There is a linear map

$$\wedge^2 \mathcal{L}(H) \to \mathcal{L}(2H); \quad f \wedge g \mapsto \frac{f\,dg - g\,df}{\omega}. \tag{2}$$

Since (1) is surjective for $d = 2$, we may represent this map by an $n \times n$ alternating matrix of quadratic forms in $x_1, \ldots, x_n$. This matrix Ω represents ω in the sense that

$$\omega = \frac{x_j^2 d(x_i/x_j)}{\Omega_{ij}(x_1, \ldots, x_n)} \quad \text{for all } i \neq j.$$

However if $n \geq 4$ then there are quadrics vanishing on $C \subset \mathbb{P}^{n-1}$ and so this description does not determine Ω uniquely. Nonetheless we show, by proving [Fisher 2013b, Conjecture 7.4], that there is a canonical choice of Ω. We then define polynomials c_4 and c_6 of degrees 4 and 6 in the coefficients of the entries of Ω, and show that the Jacobian has Weierstrass equation

$$y^2 = x^3 - 27c_4(\Omega)x - 54c_6(\Omega).$$

These main results are stated in Section 1. In the next two sections we show that c_4 and c_6 are invariants for the appropriate action of GL_n, and that they reduce to the previously known formulae for $n \leq 5$. At this point the proof of our results for any given value of n is a finite calculation. However finding a proof that works for all n is more challenging.

In Section 4 we show that if we can find a matrix Ω satisfying some apparently weaker hypotheses, then it will satisfy the properties claimed in Theorem 1.1. For the actual construction of Ω in Section 5 we reduce to the case where C is an elliptic curve E embedded in $\mathbb{P}^{n-1}$ via the complete linear system $|n.0_E|$. At first we specify Ω as a linear map $\wedge^2 \mathcal{L}(n.0_E) \to S^2 \mathcal{L}(n.0_E)$, and use this in Section 6 to complete the proof of Theorem 1.1. Then in Section 7 we make a specific choice of basis for $\mathcal{L}(n.0_E)$, so that Ω

becomes an alternating matrix of quadratic forms. We compute this matrix explicitly and, in Section 8, prove the formula for the Jacobian by computing $c_4(\Omega)$ and $c_6(\Omega)$. Much of the work here is in checking that the invariants c_4 and c_6 are scaled correctly for all n.

The description of Ω in Theorem 1.1 involves higher secant varieties. We quote any general results we need about these as required. Proofs, or references to the literature, are then given in Section 9.

In future work we plan to study the space of all matrices Ω. This appears to be defined by $d_1 + d_2$ quadrics in $\mathbb{P}^{N-1}$, where $N = (n^2 - 1)(n^2 - 4)/4$ and

$$d_1 = (n^2 - 1)(n^2 - 4)(n^2 - 9)/36, \quad d_2 = (n^2 - 1)^2(n^2 - 9)/9.$$

The numbers N, d_1, and d_2 are dimensions of irreducible representations for GL_n. Moreover, as suggested by Manjul Bhargava, we expect that d_2 of the quadrics can be explained by an associative law, similar to that used in [Bhargava 2008, §4].

We work throughout over a field K of characteristic 0, although it would in fact be sufficient that the characteristic is not too small compared to n. Except at the end of Section 1, where we give the application to computing Jacobians, we will assume that K is algebraically closed. For a projective variety X we write $I(X)$ for its homogeneous ideal, and $T_P X$ for the tangent space at $P \in X$. A Magma script containing some of the formulae in this paper is available from the author's website.

1. Statement of results

Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree $n \geq 3$. For any integer $r \geq 1$ the r-th *higher secant variety* $\mathrm{Sec}^r C$ is the Zariski closure of the locus of all $(r-1)$-planes through r points on C. For example, if $r = 1$ then $\mathrm{Sec}^1 C = C$. The codimension of $\mathrm{Sec}^r C$ in $\mathbb{P}^{n-1}$ is $\max(n - 2r, 0)$. So according as n is odd or even there is a higher secant variety of codimension 1 or 2. If $n = 2r + 1$ then $\mathrm{Sec}^r C$ is a hypersurface of degree n, whereas if $n = 2r + 2$ then $\mathrm{Sec}^r C$ is the complete intersection of two forms of degree $r + 1$. In Section 9 we give references for these facts about higher secant varieties, and also explain how to compute equations for $\mathrm{Sec}^r C$ from equations for C.

We give the polynomial ring $R = K[x_1, \ldots, x_n]$ its usual grading by degree, say $R = \bigoplus_d R_d$, and write $R(d)$ for the graded R-module with e-th graded piece R_{d+e}. Maps between graded free R-modules are required to have relative degree 0, and are labelled by the matrices of forms that represent them. Our first main result is

Theorem 1.1. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree $n \geq 3$:*

(i) *If n is odd, say $n = 2r + 1$, and $\mathrm{Sec}^r C = \{F = 0\}$ then there is a minimal free resolution*

$$0 \to R(-2n) \overset{\nabla^T}{\to} R(-n-1)^n \overset{\Omega}{\to} R(-n+1)^n \overset{\nabla}{\to} R,$$

where Ω is an $n \times n$ alternating matrix of quadratic forms and

$$\nabla = \nabla(F) = \left(\frac{\partial F}{\partial x_1} \cdots \frac{\partial F}{\partial x_n} \right).$$

(ii) *If n is even, say $n = 2r + 2$, and $\mathrm{Sec}^r C = \{F_1 = F_2 = 0\}$ then there is a minimal free resolution*

$$0 \to R(-n)^2 \xrightarrow{\nabla^T} R\left(\tfrac{1}{2}(-n-2)\right)^n \xrightarrow{\Omega} R\left(\tfrac{1}{2}(-n+2)\right)^n \xrightarrow{\nabla} R^2,$$

where Ω is an $n \times n$ alternating matrix of quadratic forms and

$$\nabla = \nabla(F_1, F_2) = \begin{pmatrix} \partial F_1/\partial x_1 & \cdots & \partial F_1/\partial x_n \\ \partial F_2/\partial x_1 & \cdots & \partial F_2/\partial x_n \end{pmatrix}.$$

We remarked in [Fisher 2013b, §7] that Theorem 1.1(i) follows from the Buchsbaum–Eisenbud structure theorem for Gorenstein ideals of codimension 3. In this paper we give a different proof, not only so that it runs in parallel with our proof of Theorem 1.1(ii), but also because this is needed for the proof of Theorem 1.2.

If the matrix Ω exists then, by the uniqueness of minimal free resolutions (see for example [Eisenbud 1995, §20.1; Peeva 2011, §7]), it is uniquely determined up to scalars. Moreover starting from equations for $\mathrm{Sec}^r C$ we can solve for Ω by linear algebra. The details are very similar to those in [Fisher 2013a, §4].

Let $\Omega = (\Omega_{ij})$ be as specified in Theorem 1.1. We put

$$M_{ij} = \sum_{r,s=1}^{n} \frac{\partial \Omega_{ir}}{\partial x_s} \frac{\partial \Omega_{js}}{\partial x_r} \quad \text{and} \quad N_{ijk} = \sum_{r=1}^{n} \frac{\partial M_{ij}}{\partial x_r} \Omega_{rk}. \tag{3}$$

We then define

$$c_4(\Omega) = \frac{3(n-2)^2}{2^4 n \binom{n+3}{5}} \sum_{i,j,r,s=1}^{n} \frac{\partial^2 M_{ij}}{\partial x_r \partial x_s} \frac{\partial^2 M_{rs}}{\partial x_i \partial x_j} \tag{4}$$

and

$$c_6(\Omega) = \frac{-(n-2)^3}{2^6 n \binom{n+5}{7}} \sum_{i,j,k,r,s,t=1}^{n} \frac{\partial^3 N_{ijk}}{\partial x_r \partial x_s \partial x_t} \frac{\partial^3 N_{rst}}{\partial x_i \partial x_j \partial x_k}. \tag{5}$$

Let C_1 and C_2 be genus one curves with invariant differentials ω_1 and ω_2. An isomorphism $\gamma : (C_1, \omega_1) \to (C_2, \omega_2)$ is an isomorphism of curves $\gamma : C_1 \to C_2$ with $\gamma^* \omega_2 = \omega_1$.

Theorem 1.2. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree $n \geq 3$, and let Ω be an alternating matrix of quadratic forms as specified in Theorem 1.1. Then:*

(i) *There is an invariant differential ω on C such that*

$$\omega = \frac{x_j^2 d(x_i/x_j)}{\Omega_{ij}(x_1, \ldots, x_n)} \quad \text{for all } i \neq j.$$

(ii) *The pair (C, ω) is isomorphic (over $K = \bar{K}$) to*

$$(y^2 = x^3 - 27 c_4(\Omega) x - 54 c_6(\Omega), 3 dx/y).$$

The following corollary gives the application of Theorem 1.2 to computing Jacobians. For this result only we drop our assumption that K is algebraically closed.

Corollary 1.3. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve defined over a field K. Suppose we scale the matrix Ω in Theorem 1.1 so that the coefficients of its entries are in K. Then C has Jacobian elliptic curve $y^2 = x^3 - 27c_4(\Omega)x - 54c_6(\Omega)$.*

Proof. Let E be the elliptic curve $y^2 = x^3 - 27c_4(\Omega)x - 54c_6(\Omega)$. By Theorem 1.2 there is an isomorphism $\gamma : C \to E$ with $\gamma^*(3dx/y) = \omega$. Let $\xi_\sigma = \sigma(\gamma)\gamma^{-1}$ for $\sigma \in \text{Gal}(\bar{K}/K)$. Since $3dx/y$ and ω are both K-rational it follows that $\xi_\sigma^*(3dx/y) = 3dx/y$. This implies, as explained for example in [Fisher 2008, Lemma 2.4], that $\xi_\sigma : E \to E$ is a translation map. Then C is the twist of E by the class of $\{\xi_\sigma\}$ in $H^1(K, E)$. It follows by Theorems 3.6 and 3.8 in [Silverman 2009, Chapter X] that C is a principal homogeneous space under E, and E is the Jacobian of C. $\qquad\square$

Remark 1.4. Although we will not need it for the proofs of Theorems 1.1 and 1.2, it is natural to ask whether $C \subset \mathbb{P}^{n-1}$ is uniquely determined by Ω. The answer is that it is. Indeed by the minimal free resolutions in Theorem 1.1 we can recover ∇ from Ω. Then by Euler's identity we obtain equations for $\text{Sec}^r C$ where $n - 2r = 1$ or 2. This then determines $\text{Sec}^1 C = C$ by Theorem 9.1(v).

2. Changes of coordinates

We show that the constructions in Section 1 behave well under all changes of coordinates. First we define an action of GL_n on the space of all $n \times n$ alternating matrices of quadratic forms in $x_1, \ldots x_n$. For $g \in \text{GL}_n$ we put

$$g \star \Omega = g^{-T}\left(\Omega\left(\sum_{i=1}^{n} g_{i1}x_i, \ldots, \sum_{i=1}^{n} g_{in}x_i\right)\right)g^{-1},$$

where g^{-T} is the inverse transpose of g. Since the scalar matrices act trivially, this could equally be viewed as an action of PGL_n.

Lemma 2.1. *Let $C \subset \mathbb{P}^{n-1}$ and $C' \subset \mathbb{P}^{n-1}$ be genus one normal curves. Let Ω and Ω' be alternating matrices of quadratic forms that satisfy the conclusions of Theorem 1.1, and define invariant differentials ω and ω' on C and C'. If $\gamma : C' \to C$ is an isomorphism given by*

$$(x_1 : \ldots : x_n) \mapsto \left(\sum_{i=1}^{n} g_{i1}x_i : \cdots : \sum_{i=1}^{n} g_{in}x_i\right)$$

for some $g \in \text{GL}_n$ then there exists $\lambda \in K^\times$ such that $g \star \Omega = \lambda\Omega'$ and $\gamma^*\omega = \lambda^{-1}\omega'$.

Proof. Suppose n is odd, say $n = 2r + 1$ and $\text{Sec}^r C = \{F = 0\}$. Then $\text{Sec}^r C'$ is defined by

$$F'(x_1, \ldots, x_n) = F(y_1, \ldots, y_n)$$

where $y_j = \sum_{i=1}^{n} g_{ij}x_i$. By the chain rule

$$\nabla(F')(x_1, \ldots, x_n) = \nabla(F)(y_1, \ldots, y_n)\ g^T.$$

Then

$$\nabla(F)\Omega = 0 \implies \nabla(F')(g \star \Omega) = 0.$$

It follows by the uniqueness of minimal free resolutions that $g \star \Omega = \lambda \Omega'$ for some $\lambda \in K^\times$. The case n is even is similar.

We also have $\gamma^* \omega = \mu \omega'$ for some $\mu \in K^\times$. If $y_j = \sum_{i=1}^n g_{ij} x_i$ then

$$y_s^2 d(y_r/y_s) = \sum_{i,j=1}^n g_{ir} g_{js} x_j^2 d(x_i/x_j).$$

Dividing by $\gamma^* \omega = \mu \omega'$ gives

$$\Omega(y_1, \ldots, y_n) = \mu^{-1} g^T \Omega'(x_1, \ldots, x_n) g.$$

Hence $g \star \Omega = \mu^{-1} \Omega'$ and so $\mu = \lambda^{-1}$.   □

Lemma 2.2. *The polynomials c_4 and c_6 are invariants for the action of GL_n, i.e., $c_4(g \star \Omega) = c_4(\Omega)$ and $c_6(g \star \Omega) = c_6(\Omega)$ for all $g \in \mathrm{GL}_n$.*

Proof. Let $\Omega' = g \star \Omega$, i.e.,

$$\Omega'_{ij}(x_1, \ldots, x_n) = \sum_{a,b=1}^n (g^{-1})_{ai} (g^{-1})_{bj} \Omega_{ab}(y_1, \ldots, y_n),$$

where $y_j = \sum_{i=1}^n g_{ij} x_i$. Direct calculation using (3) shows that

$$M'_{ij}(x_1, \ldots, x_n) = \sum_{a,b=1}^n (g^{-1})_{ai} (g^{-1})_{bj} M_{ab}(y_1, \ldots, y_n),$$

$$N'_{ijk}(x_1, \ldots, x_n) = \sum_{a,b,c=1}^n (g^{-1})_{ai} (g^{-1})_{bj} (g^{-1})_{ck} N_{abc}(y_1, \ldots, y_n).$$

Then

$$\frac{\partial^2 M'_{ij}}{\partial x_r \partial x_s} = \sum_{a,b,c,d=1}^n (g^{-1})_{ai} (g^{-1})_{bj} g_{rc} g_{sd} \frac{\partial^2 M_{ab}}{\partial x_c \partial x_d},$$

$$\frac{\partial^2 M'_{rs}}{\partial x_i \partial x_j} = \sum_{A,B,C,D=1}^n (g^{-1})_{Cr} (g^{-1})_{Ds} g_{iA} g_{jB} \frac{\partial^2 M_{CD}}{\partial x_A \partial x_B}.$$

Multiplying these together and summing gives

$$\sum_{i,j,r,s=1}^n \frac{\partial^2 M'_{ij}}{\partial x_r \partial x_s} \frac{\partial^2 M'_{rs}}{\partial x_i \partial x_j} = \sum_{a,b,c,d=1}^n \frac{\partial^2 M_{ab}}{\partial x_c \partial x_d} \frac{\partial^2 M_{cd}}{\partial x_a \partial x_b}.$$

Thus $c_4(\Omega') = c_4(\Omega)$. A similar argument shows that $c_6(\Omega') = c_6(\Omega)$.   □

The following corollary shows that to prove Theorems 1.1 and 1.2 for a fixed value of n, it suffices to prove them for a family of curves covering the j-line.

Corollary 2.3. *Let Ω_1 and Ω_2 correspond to pairs (C_1, ω_1) and (C_2, ω_2). If there is an isomorphism $\gamma : C_1 \to C_2$ with $\gamma^* \omega_2 = \lambda \omega_1$ then $c_4(\Omega_1) = \lambda^4 c_4(\Omega_2)$ and $c_6(\Omega_1) = \lambda^6 c_6(\Omega_2)$.*

Proof. Let C_1 and C_2 have hyperplane sections H_1 and H_2. Then H_1 and $\gamma^* H_2$ are degree n divisors on C_1. After composing the isomorphism γ with a translation map, we may suppose (see [Silverman 2009, III.3.5]) that $H_1 \sim \gamma^* H_2$. Then γ is given by a change of coordinates on $\mathbb{P}^{n-1}$. The case $\lambda = 1$ is immediate from Lemmas 2.1 and 2.2. In general we use that c_4 and c_6 are homogeneous polynomials of degrees 4 and 6. $\qquad\square$

3. Curves of small degree

We compare our general formula for the Jacobian with the formulae previously known for genus one normal curves of degrees 3, 4, and 5.

For curves of degrees 3 and 4 it is easy to write down a matrix Ω satisfying the conclusions of Theorems 1.1 and 1.2(i). Indeed for $C = \{F(x_1, x_2, x_3) = 0\} \subset \mathbb{P}^2$ a plane cubic we put

$$\Omega = \begin{pmatrix} 0 & \partial F / \partial x_3 & -\partial F / \partial x_2 \\ -\partial F / \partial x_3 & 0 & \partial F / \partial x_1 \\ \partial F / \partial x_2 & -\partial F / \partial x_1 & 0 \end{pmatrix},$$

and for $C = \{F_1 = F_2 = 0\} \subset \mathbb{P}^3$ a quadric intersection we let Ω be the 4×4 alternating matrix with entries

$$\Omega_{ij} = \frac{\partial F_1}{\partial x_k} \frac{\partial F_2}{\partial x_l} - \frac{\partial F_1}{\partial x_l} \frac{\partial F_2}{\partial x_k},$$

where (i, j, k, l) is an even permutation of $(1, 2, 3, 4)$. To prove Theorem 1.2(ii) in these cases we may check by direct computation that $c_4(\Omega)$ and $c_6(\Omega)$ are the classical invariants of a ternary cubic or quadric intersection, as scaled in [Fisher 2008, §7]. We note that these are polynomials of degrees 4 and 6 in the coefficients of F, respectively of degrees 8 and 12 in the coefficients of F_1 and F_2.

As described for example in [Fisher 2013a, §4], a genus one normal curve of degree $n = 5$ is defined by the 4×4 Pfaffians $p_1, \ldots, p_5$ of a 5×5 alternating matrix of linear forms on $\mathbb{P}^4$. We call the matrix of linear forms Φ a *genus one model* of degree 5, and note that there is a natural action of $\mathrm{GL}_5 \times \mathrm{GL}_5$ on the space of all such models. It is shown in [Hulek 1986, Proposition VIII.2.5] that the secant variety $\mathrm{Sec}^2 C$ is a hypersurface of degree 5 with equation $F = 0$, where F is the determinant of the Jacobian matrix of $p_1, \ldots, p_5$. In [Fisher 2013b, §7] we proved that there is a degree 5 covariant Ω satisfying the conclusions of Theorems 1.1 and 1.2(i). We gave an explicit formula for this covariant in [Fisher and Sadek 2016, §2].

We claim that $c_4(\Omega)$ and $c_6(\Omega)$ are invariants for the action of $\mathrm{SL}_5 \times \mathrm{SL}_5$. For the action of SL_5 via changes of coordinates on $\mathbb{P}^4$ this follows from Lemma 2.2. For the action of SL_5 via $\Phi \mapsto A \Phi A^T$ it turns out that the coefficients of the entries of Ω are already invariants. Since Ω is a covariant of degree 5, the invariants $c_4(\Omega)$ and $c_6(\Omega)$ have degrees 20 and 30 in the coefficients of the entries of Φ. Computing a

single numerical example (to check the scaling) shows that $c_4(\Omega)$ and $c_6(\Omega)$ are the same as the invariants $c_4(\Phi)$ and $c_6(\Phi)$ constructed in [Fisher 2008].

4. Minimal free resolutions

Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree $n \geq 3$. Let Ω be an $n \times n$ alternating matrix of quadratic forms in $x_1, \ldots, x_n$. In Sections 5 and 6 we exhibit Ω satisfying the following three hypotheses:

(H1) If $n - 2r \geq 1$ and $f \in I(\mathrm{Sec}^r C)$ then $\sum_{i=1}^{n} \dfrac{\partial f}{\partial x_i} \Omega_{ij} \in I(\mathrm{Sec}^r C)$ for all $1 \leq j \leq n$.

(H2) If $n - 2r = 2$ and $\mathrm{Sec}^r C = \{F_1 = F_2 = 0\}$ then $\sum_{i,j=1}^{n} \dfrac{\partial F_1}{\partial x_i} \Omega_{ij} \dfrac{\partial F_2}{\partial x_j} = 0.$

(H3) If $n - 2r \geq 1$ then there exists $P \in \mathrm{Sec}^r C$ with rank $\Omega(P) = 2r$.

In this section we prove:

Theorem 4.1. *Let Ω be an $n \times n$ alternating matrix of quadratic forms, satisfying the hypotheses* **(H1)**, **(H2)**, *and* **(H3)**. *Then there is a minimal free resolution as described in Theorem 1.1.*

The next two propositions are proved in Section 9. By abuse of notation we write P both for a point in $\mathbb{P}^{n-1}$ and for a vector of length n representing this point.

Proposition 4.2. *If $n - 2r \geq 1$ and $P = \sum_{i=1}^{r} \xi_i P_i$ for some $P_1, \ldots, P_r \in C$ distinct and $\xi_1, \ldots, \xi_r \neq 0$ then the tangent space $T_P \mathrm{Sec}^r C$ is the linear span of the tangent lines $T_{P_1} C, \ldots, T_{P_r} C$.*

Proposition 4.3. *Let $\nabla(F)$ and $\nabla(F_1, F_2)$ be as defined in Theorem 1.1:*

  (i) *If $n - 2r = 1$ and $\mathrm{Sec}^r C = \{F = 0\}$ then the entries of $\nabla(F)$ define a variety in $\mathbb{P}^{n-1}$ of codimension 3.*

  (ii) *If $n - 2r = 2$ and $\mathrm{Sec}^r C = \{F_1 = F_2 = 0\}$ then the 2×2 minors of $\nabla(F_1, F_2)$ define a variety in $\mathbb{P}^{n-1}$ of codimension 3.*

Proof. (i) Theorem 9.1 tells us that $\mathrm{Sec}^r C$ has singular locus $\mathrm{Sec}^{r-1} C$, and that this has codimension 3.

(ii) This is proved in Section 9.3.   □

We start the proof of Theorem 4.1 with the following lemma.

Lemma 4.4. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve. Suppose that $n - 2r \geq 1$ and $\ell_1, \ldots, \ell_n$ are linear forms in $x_1, \ldots, x_n$ such that*

$$\sum_{i=1}^{n} \ell_i \frac{\partial f}{\partial x_i} \in I(\mathrm{Sec}^r C) \quad \text{for all } f \in I(\mathrm{Sec}^r C). \tag{6}$$

Then there exists $\lambda \in K$ such that $\ell_i = \lambda x_i$ for all $1 \leq i \leq n$.

Proof. The coefficients of $\ell_1, \ldots, \ell_n$ form an $n \times n$ matrix. Let $V \subset \mathrm{Mat}_n(K)$ be the subspace of all solutions to (6). We must show that V consists only of scalar matrices. Let E be the Jacobian of C. Translation by $T \in E[n]$ is an automorphism of C that extends to an automorphism of $\mathbb{P}^{n-1}$, say given by

a matrix M_T. Now V is stable under conjugation by each M_T. By considering the standard representation of the Heisenberg group (see for example [Fisher 2010, §3]) it follows that V has a basis $\{M_T : T \in X\}$ for some subset $X \subset E[n]$.

We suppose for a contradiction that $M_T \in V$ for some $0_E \neq T \in E[n]$. Then translation by T on C extends to an automorphism of $\mathbb{P}^{n-1}$ that sends each point $P \in \operatorname{Sec}^r C$ to a point in the tangent space $T_P \operatorname{Sec}^r C$. Let H be the divisor of a hyperplane section on C. For D an effective divisor on C we write $\bar{D} \subset \mathbb{P}^{n-1}$ for the linear subspace cut out by $\mathcal{L}(H - D) \subset \mathcal{L}(H)$. For example, if D is a sum of distinct points on C then $\bar{D}$ is the linear span of these points. We also write D_T for D translated by T. We choose $D = P_1 + \cdots + P_r$ an effective divisor of degree r such that:

(i) $P_1, \ldots, P_r \in C$ are distinct,

(ii) D and D_T have disjoint support, and

(iii) $2D + D_T \not\sim H$.

Proposition 4.2 shows that for generic $P \in \bar{D}$ we have $T_P \operatorname{Sec}^r C = \overline{2D}$. It follows from our assumption $M_T \in V$ that $\overline{D_T} \subset \overline{2D}$, equivalently $\mathcal{L}(H - 2D) \subset \mathcal{L}(H - D_T)$. Then by (ii) we have

$$\mathcal{L}(H - 2D) = \mathcal{L}(H - 2D) \cap \mathcal{L}(H - D_T) = \mathcal{L}(H - 2D - D_T).$$

However by (iii) and the Riemann–Roch theorem these spaces do not have the same dimension. Indeed, since $r \geq 1$ and $n - 2r \geq 1$ we have

$$\dim \mathcal{L}(H - 2D) = n - 2r \neq \max(n - 3r, 0) = \dim \mathcal{L}(H - 2D - D_T).$$

This is the required contradiction. $\qquad\square$

We show that the resolution in Theorem 1.1 is a complex.

Lemma 4.5. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve, and let Ω be an alternating matrix of quadratic forms satisfying the hypotheses* (**H1**) *and* (**H2**):

(i) *If $n = 2r + 1$ and $\operatorname{Sec}^r C = \{F = 0\}$ then*

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i} \Omega_{ij} = 0 \quad \text{for all } 1 \leq j \leq n.$$

(ii) *If $n = 2r + 2$ and $\operatorname{Sec}^r C = \{F_1 = F_2 = 0\}$ then*

$$\sum_{i=1}^{n} \frac{\partial F_1}{\partial x_i} \Omega_{ij} = \sum_{i=1}^{n} \frac{\partial F_2}{\partial x_i} \Omega_{ij} = 0 \quad \text{for all } 1 \leq j \leq n.$$

Proof. (*i*) By the hypothesis (**H1**) we have

$$\sum_{i=1}^{n} \frac{\partial F}{\partial x_i} \Omega_{ij} = \ell_j F \quad \text{for all } 1 \leq j \leq n,$$

for some linear forms $\ell_1, \ldots, \ell_n$. We multiply by $\partial F/\partial x_j$ and sum over j. Since Ω is alternating the left-hand side is zero. Therefore

$$\sum_{j=1}^{n} \ell_j \frac{\partial F}{\partial x_j} = 0.$$

By Lemma 4.4 and Euler's identity it follows that $\ell_1 = \cdots = \ell_n = 0$ as required.

(ii) By the hypothesis (**H1**) we have

$$\sum_{i=1}^{n} \frac{\partial F_1}{\partial x_i} \Omega_{ij} = \ell_j F_1 + m_j F_2 \quad \text{for all } 1 \le j \le n, \tag{7}$$

for some linear forms $\ell_1, \ldots, \ell_n$ and $m_1, \ldots, m_n$. We multiply by $\partial F_1/\partial x_j$ and sum over j. Since Ω is alternating the left-hand side is zero. Since F_1 and F_2 are forms defining a variety of codimension 2 they must be coprime. Therefore

$$\sum_{j=1}^{n} \ell_j \frac{\partial F_1}{\partial x_j} = \xi F_2 \quad \text{and} \quad \sum_{j=1}^{n} m_j \frac{\partial F_1}{\partial x_j} = -\xi F_1$$

for some $\xi \in K$. If instead we multiply (7) by $\partial F_2/\partial x_j$ and sum over j then using the hypothesis (**H2**) we find that

$$\sum_{j=1}^{n} \ell_j \frac{\partial F_2}{\partial x_j} = \eta F_2 \quad \text{and} \quad \sum_{j=1}^{n} m_j \frac{\partial F_2}{\partial x_j} = -\eta F_1$$

for some $\eta \in K$.

By Lemma 4.4 there exist $\lambda, \mu \in K$ such that $\ell_i = \lambda x_i$ and $m_i = \mu x_i$ for all $1 \le i \le n$. By Euler's identity and the linear independence of F_1 and F_2 it follows that $\lambda = \mu = 0$. Therefore

$$\sum_{i=1}^{n} \frac{\partial F_1}{\partial x_i} \Omega_{ij} = 0 \quad \text{for all } 1 \le j \le n.$$

The corresponding result for F_2 follows by symmetry. $\qquad\qquad\square$

To complete the proof of Theorem 4.1 we must show that the complex is exact. First we need some linear algebra. If B is an $n \times n$ matrix and $S \subset \{1, \ldots, n\}$ then we write B^S for the $(n - |S|) \times (n - |S|)$ matrix obtained by deleting the rows and columns indexed by S. The Pfaffian $\mathrm{pf}(M)$ of an alternating matrix M is a polynomial in the matrix entries with the property that $\det(M) = \mathrm{pf}(M)^2$.

Lemma 4.6. (i) *Let $A = (a_i)$ be a $1 \times n$ matrix and B an $n \times n$ alternating matrix over a field K. Suppose that $\mathrm{rank}\, A = 1$, $\mathrm{rank}\, B = n - 1$, and $AB = 0$. Then there exists $\lambda \in K^\times$ such that*

$$(-1)^i \, \mathrm{pf}(B^{\{i\}}) = \lambda a_i$$

for all $1 \le i \le n$.

(ii) *Let $A = (a_{ij})$ be a $2 \times n$ matrix and B an $n \times n$ alternating matrix over a field K. Suppose that rank $A = 2$, rank $B = n - 2$ and $AB = 0$. Then there exists $\lambda \in K^{\times}$ such that*

$$(-1)^{i+j} \operatorname{pf}(B^{\{i,j\}}) = \lambda(a_{1i}a_{2j} - a_{1j}a_{2i})$$

for all $1 \leq i < j \leq n$.

Proof. (*i*) It is well known that the vector with i-th entry $(-1)^i \operatorname{pf}(B^{\{i\}})$ belongs to the kernel of B. See for example [Bruns and Herzog 1993, §3.4]. Since rank $B = n - 1$, this vector is nonzero and the kernel is 1-dimensional. The result follows.

(*ii*) We first claim there exist $\lambda_1, \ldots, \lambda_n \in K$ such that

$$(-1)^{i+j} \operatorname{pf}(B^{\{i,j\}}) = \begin{cases} \lambda_i(a_{1i}a_{2j} - a_{1j}a_{2i}) & \text{if } i < j, \\ -\lambda_i(a_{1i}a_{2j} - a_{1j}a_{2i}) & \text{if } i > j. \end{cases}$$

Indeed taking a_{2i} times the first row of A minus a_{1i} times the second row of A gives a nonzero vector in the kernel of $B^{\{i\}}$. If rank $B^{\{i\}} = n - 2$ then we argue as in (i). Otherwise we can simply take $\lambda_i = 0$. This proves the claim.

Now let $C = (a_{1i}a_{2j} - a_{1j}a_{2i})_{i,j=1,\ldots,n}$ and let D be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$. We must show that if $CD = DC$ then CD is a scalar multiple of C. More generally this is true for any rank 2 alternating matrix C and diagonal matrix D. Indeed we may reorder the rows and columns so that the diagonal entries of D which are equal are grouped together. Then C is in block diagonal form. Since C is alternating of rank 2, exactly one of these blocks is nonzero. The result is then clear. $\square$

Lemma 4.7. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve, and let Ω be an alternating matrix of quadratic forms satisfying the hypotheses* (**H1**), (**H2**), *and* (**H3**):

(i) *If $n = 2r + 1$ and $\operatorname{Sec}^r C = \{F = 0\}$ then the $(n-1) \times (n-1)$ Pfaffians of Ω are (scalar multiples of) the partial derivatives of F.*

(ii) *If $n = 2r + 2$ and $\operatorname{Sec}^r C = \{F_1 = F_2 = 0\}$ then the $(n-2) \times (n-2)$ Pfaffians of Ω are (scalar multiples of) the 2×2 minors of $\nabla(F_1, F_2)$.*

Proof. We apply Lemma 4.6 over the function field $K(x_1, \ldots, x_n)$.

(*i*) By Lemma 4.5 we have $\sum_{i=1}^n \partial F / \partial x_i \, \Omega_{ij} = 0$. By the hypothesis (**H3**) the generic rank of Ω is $n - 1$. So by Lemma 4.6(i) there exists $\lambda \in K(x_1, \ldots, x_n)$ such that

$$(-1)^i \operatorname{pf}(\Omega^{\{i\}}) = \lambda \frac{\partial F}{\partial x_i} \quad \text{for all } 1 \leq i \leq n.$$

Since $\operatorname{pf}(\Omega^{\{i\}})$ and $\partial F / \partial x_i$ are forms of degree $n - 1$, we can write $\lambda = u/v$ where u and v are coprime forms of the same degree. Then v divides $\partial F / \partial x_i$ for all i, and so must be a constant by Proposition 4.3(i). Therefore λ is a constant.

(ii) By Lemma 4.5 we have $\sum_{i=1}^{n} \partial F_1/\partial x_i \, \Omega_{ij} = \sum_{i=1}^{n} \partial F_2/\partial x_i \, \Omega_{ij} = 0$. By the hypothesis **(H3)** the generic rank of Ω is $n-2$. So by Lemma 4.6(ii) there exists $\lambda \in K(x_1, \ldots, x_n)$ such that

$$(-1)^{i+j} \, \mathrm{pf}(\Omega^{\{i,j\}}) = \lambda \frac{\partial(F_1, F_2)}{\partial(x_i, x_j)} \quad \text{for all } 1 \le i < j \le n.$$

Since $\mathrm{pf}(\Omega^{\{i,j\}})$ and $\partial(F_1, F_2)/\partial(x_i, x_j)$ are forms of degree $n-2$, we can write $\lambda = u/v$ where u and v are coprime forms of the same degree. Then v divides $\partial(F_1, F_2)/\partial(x_i, x_j)$ for all i, j, and so must be a constant by Proposition 4.3(ii). Therefore λ is a constant. $\qquad\square$

Let $R = K[x_1, \ldots, x_n]$. Consider a complex of graded free R-modules

$$0 \to F_m \xrightarrow{\varphi_m} F_{m-1} \to \cdots \to F_1 \xrightarrow{\varphi_1} F_0. \tag{8}$$

We write $V_k \subset \mathbb{P}^{n-1}$ for the subvariety defined by the $r_k \times r_k$ minors of φ_k where $r_k = \mathrm{rank}(\varphi_k)$. The Buchsbaum–Eisenbud acyclicity criterion (see [Bruns and Herzog 1993, Theorem 1.4.13; Eisenbud 1995, Theorem 20.9]) states that (8) is exact if and only if rank $F_k = \mathrm{rank}\,\varphi_k + \mathrm{rank}\,\varphi_{k+1}$ and codim $V_k \ge k$ for all $1 \le k \le m$.

Proof of Theorem 4.1. We already saw in Lemma 4.5 that the resolution in Theorem 1.1 is a complex. We must prove it is exact. If n is odd then the free R-modules have ranks $1, n, n, 1$ and the maps have ranks $1, n-1, 1$. If n is even then the free R-modules have ranks $2, n, n, 2$ and the maps have ranks $2, n-2, 2$. By Lemma 4.7 we have $V_1 = V_2 = V_3$ and Proposition 4.3 shows that this variety has codimension 3. We now apply the Buchsbaum–Eisenbud acyclicity criterion. $\qquad\square$

5. A basis-free construction

The results of Section 2 show that for the proof of Theorems 1.1 and 1.2 we are free to make changes of coordinates on $\mathbb{P}^{n-1}$. Since we are working over an algebraically closed field we can therefore reduce to the following situation. Let E be the elliptic curve

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$

with point at infinity 0_E and invariant differential

$$\omega = dx/(2y + a_1 x + a_3) = dy/(3x^2 + 2a_2 x + a_4 - a_1 y).$$

Let $C \subset \mathbb{P}^{n-1}$ be the image of E embedded via the complete linear system $|n.0_E|$. The embedding depends on a choice of basis for the Riemann–Roch space $\mathcal{L}(n.0_E)$, but the only effect of changing this is to make a change of coordinates on $\mathbb{P}^{n-1}$. In this section we define a linear map $\Omega : \wedge^2 \mathcal{L}(n.0_E) \to S^2 \mathcal{L}(n.0_E)$. In the next section we show that the corresponding alternating matrix of quadratic forms satisfies the hypotheses **(H1)**, **(H2)**, and **(H3)**.

For $f \in \mathcal{L}(n.0_E)$ we put $\dot{f} = df/\omega \in \mathcal{L}((n+1).0_E)$. Motivated by (2) we define a linear map

$$A : \wedge^2 \mathcal{L}(n.0_E) \to S^2 \mathcal{L}((n+1).0_E); \quad f \wedge g \mapsto f \otimes \dot{g} - g \otimes \dot{f}.$$

Lemma 5.1. *Let $f, g \in \mathcal{L}(n.0_E)$. Then the rational function on $E \times E$ given by*

$$(P, Q) \mapsto \frac{y_P + y_Q + a_1 x_Q + a_3}{x_P - x_Q}(f(Q)g(P) - f(P)g(Q))$$

belongs to $\mathcal{L}((n+1).0_E) \otimes \mathcal{L}((n+1).0_E)$.

Proof. (i) If we fix $Q = (x_Q, y_Q)$ then as rational functions of $P = (x, y)$,

$$\frac{y + y_Q + a_1 x_Q + a_3}{x - x_Q} \in \mathcal{L}(0_E + Q) \quad \text{and} \quad f(Q)g - g(Q)f \in \mathcal{L}(n.0_E - Q).$$

Therefore the product belongs to $\mathcal{L}((n+1).0_E)$.

(ii) If we fix $P = (x_P, y_P)$ then as rational functions of $Q = (x, y)$,

$$\frac{y_P + y + a_1 x + a_3}{x_P - x} \in \mathcal{L}(0_E + P) \quad \text{and} \quad g(P)f - f(P)g \in \mathcal{L}(n.0_E - P).$$

Therefore the product belongs to $\mathcal{L}((n+1).0_E)$. $\square$

We define a second linear map

$$B : \wedge^2 \mathcal{L}(n.0_E) \to S^2 \mathcal{L}((n+1).0_E)$$

$$f \wedge g \mapsto \frac{y_P + y_Q + a_1 x_Q + a_3}{x_P - x_Q}(f(Q)g(P) - f(P)g(Q))\big|_{P=Q},$$

where $|_{P=Q}$ is our notation for the natural map

$$\mathcal{L}((n+1).0_E) \otimes \mathcal{L}((n+1).0_E) \to S^2 \mathcal{L}((n+1).0_E).$$

We show that A and B both represent the invariant differential ω, in the sense of [Theorem 1.2(i)].

Lemma 5.2. *As rational functions on E we have*

$$A(f \wedge g) = B(f \wedge g) = f\dot{g} - g\dot{f} = \frac{f\,dg - g\,df}{\omega}.$$

Proof. This is clear for A. For B we apply l'Hôpital's rule to get

$$\frac{f(Q)g - g(Q)f}{x - x_Q}\bigg|_{P=Q} = \frac{f(Q)\dot{g} - g(Q)\dot{f}}{\dot{x}}\bigg|_{P=Q},$$

and then use that $\dot{x} = 2y + a_1 x + a_3$. $\square$

If we pick bases for $\mathcal{L}(n.0_E)$ and $\mathcal{L}((n+1).0_E)$ then A and B are (represented by) $n \times n$ alternating matrices of quadratic forms in $n + 1$ variables. However the matrix Ω we seek is an $n \times n$ alternating matrix of quadratic forms in n variables. It turns out that the correct choice of Ω is a linear combination of A and B.

We may expand rational functions on E as Laurent power series in the local parameter $t = x/y$ at 0_E. Let ϕ be the linear map that reads off the coefficient of t^{-n-1}. There are exact sequences

$$0 \to \mathcal{L}(n.0_E) \to \mathcal{L}((n+1).0_E) \xrightarrow{\phi} K \to 0$$

and

$$0 \to S^2 \mathcal{L}(n.0_E) \to S^2 \mathcal{L}((n+1).0_E) \overset{\phi_2}{\to} \mathcal{L}((n+1).0_E) \to 0 \tag{9}$$

where $\phi_2(f \otimes g) = \phi(f)g + \phi(g)f$.

Lemma 5.3. *Let $f, g \in \mathcal{L}(n.0_E)$ be rational functions whose coefficients of t^{-n} (when expanded as Laurent power series in t) are $0, 1$ respectively. Then*

$$\phi_2(A(f \wedge g)) = nf \quad and \quad \phi_2(B(f \wedge g)) = 2f.$$

Proof. (i) We have $x = t^{-2} + \cdots$ and $y = t^{-3} + \cdots$. Then $\dot{x} = 2y + a_1 x + a_3 = 2t^{-3} + \cdots$ and $\dot{y} = 3x^2 + 2a_2 x + a_4 - a_1 y = 3t^{-4} + \cdots$. Writing g as a polynomial in x and y it follows that $g = t^{-n} + \cdots$ and $\dot{g} = nt^{-n-1} + \cdots$. Therefore $\phi(f) = \phi(g) = \phi(\dot{f}) = 0$ and $\phi(\dot{g}) = n$. We compute

$$\phi_2(A(f \wedge g)) = \phi_2(f \otimes \dot{g} - g \otimes \dot{f}) = nf.$$

(ii) If we fix $Q = (x_Q, y_Q)$ then as rational functions of $P = (x, y)$,

$$\frac{y + y_Q + a_1 x_Q + a_3}{x - x_Q} = t^{-1} + \cdots \quad and \quad f(Q)g - g(Q)f = f(Q)t^{-n} + \cdots$$

with product $f(Q)t^{-n-1} + \cdots$.

If we fix $P = (x_P, y_P)$ then as rational functions of $Q = (x, y)$,

$$\frac{y_P + y + a_1 x + a_3}{x_P - x} = -t^{-1} + \cdots \quad and \quad g(P)f - f(P)g = -f(P)t^{-n} + \cdots$$

with product $f(P)t^{-n-1} + \cdots$. In both cases the leading coefficient is f. Adding these together gives $\phi_2(B(f \wedge g)) = 2f$. $\square$

Corollary 5.4. *Let $\Omega = nB - 2A$. Then Ω is a linear map $\wedge^2 \mathcal{L}(n.0_E) \to S^2 \mathcal{L}(n.0_E)$.*

Proof. This follows from Lemma 5.3 and the exact sequence (9). $\square$

6. Proof of Theorem 1.1

If we pick a basis for $\mathcal{L}(n.0_E)$ then the linear map defined in Corollary 5.4 is represented by an $n \times n$ alternating matrix of quadratic forms in n variables. In this section we complete the proof of Theorem 1.1 by showing that this matrix Ω satisfies the hypotheses **(H1)**, **(H2)**, and **(H3)**, as stated at the start of Section 4.

For $0_E \neq P \in E$ we write $\boldsymbol{P}$ and $d\boldsymbol{P}$ for the linear maps $f \mapsto f(P)$ and $f \mapsto \dot{f}(P)$ in the dual space $\mathcal{L}(n.0_E)^*$. For example, if $\mathcal{L}(n.0_E)$ has basis $1, x, y, x^2, xy, \ldots$ then

$$\boldsymbol{P} = (1, x_P, y_P, x_P^2, x_P y_P, \ldots), \quad d\boldsymbol{P} = (0, 2y_P + a_1 x_P + a_3, 3x_P^2 + 2a_2 x_P + a_4 - a_1 y_P, \ldots).$$

We note that $[\boldsymbol{P}]$ is a point on $C \subset \mathbb{P}^{n-1} = \mathbb{P}(\mathcal{L}(n.0_E)^*)$, with tangent line passing through $[d\boldsymbol{P}]$. The square brackets indicate that we are taking the 1-dimensional subspaces spanned by these vectors, i.e., the corresponding points in projective space. For $0_E \neq Q \in E$ we likewise define $\boldsymbol{Q}$ and $d\boldsymbol{Q}$.

For $P, Q \in E$ let $\lambda_{P,Q}$ be the slope of the chord (or tangent line if $P = Q$) joining P and Q. In the following lemma the vectors $\boldsymbol{P}, \boldsymbol{Q}, d\boldsymbol{P}, d\boldsymbol{Q}$ in $\mathcal{L}(n.0_E)^*$ are extended to $\mathcal{L}((n+1).0_E)^*$ using exactly the same definition. Evaluating A or B at a linear combination $\xi \boldsymbol{P} + \eta \boldsymbol{Q}$ gives an element of $(\wedge^2 \mathcal{L}(n.0_E))^* = \wedge^2(\mathcal{L}(n.0_E)^*)$.

Lemma 6.1. *Let* $0_E \neq P, Q \in E$, *and* $\xi, \eta \in K$. *Then*

(i) $\qquad A(\xi \boldsymbol{P} + \eta \boldsymbol{Q}) = \xi^2 (\boldsymbol{P} \wedge d\boldsymbol{P}) + \xi \eta (\boldsymbol{P} \wedge d\boldsymbol{Q} + \boldsymbol{Q} \wedge d\boldsymbol{P}) + \eta^2 (\boldsymbol{Q} \wedge d\boldsymbol{Q})$,

(ii) $\qquad B(\xi \boldsymbol{P} + \eta \boldsymbol{Q}) = \xi^2 (\boldsymbol{P} \wedge d\boldsymbol{P}) + \xi \eta (\lambda_{Q,-P} - \lambda_{P,-Q})(\boldsymbol{P} \wedge \boldsymbol{Q}) + \eta^2 (\boldsymbol{Q} \wedge d\boldsymbol{Q})$.

Proof. (i) For $f, g \in \mathcal{L}(n.0_E)$ we compute

$$A(\boldsymbol{P})(f \wedge g) = (f \dot{g} - g \dot{f})(P) = (\boldsymbol{P} \wedge d\boldsymbol{P})(f \wedge g).$$

The formula for $A(\xi \boldsymbol{P} + \eta \boldsymbol{Q})$ follows by bilinearity.

(ii) For $f, g \in \mathcal{L}(n.0_E)$ we write

$$B(\xi \boldsymbol{P} + \eta \boldsymbol{Q})(f \wedge g) = \xi^2 B_0 + \xi \eta B_1 + \eta^2 B_2.$$

By Lemma 5.2 we have

$$B_0 = (f \dot{g} - g \dot{f})(P) = (\boldsymbol{P} \wedge d\boldsymbol{P})(f \wedge g), \quad B_2 = (f \dot{g} - g \dot{f})(Q) = (\boldsymbol{Q} \wedge d\boldsymbol{Q})(f \wedge g).$$

Since for $s, t \in \mathcal{L}((n+1).0_E)$ we have

$$\begin{aligned}
(s \otimes t)(\xi \boldsymbol{P} + \eta \boldsymbol{Q}) &= s(\xi \boldsymbol{P} + \eta \boldsymbol{Q})t(\xi \boldsymbol{P} + \eta \boldsymbol{Q}) \\
&= \xi^2 s(P)t(P) + \xi \eta (s(P)t(Q) + s(Q)t(P)) + \eta^2 s(Q)t(Q),
\end{aligned}$$

it follows from the definition of B that

$$\begin{aligned}
B_1 &= \lambda_{P,-Q}(f(Q)g(P) - f(P)g(Q)) + \lambda_{Q,-P}(f(P)g(Q) - f(Q)g(P)) \\
&= (\lambda_{Q,-P} - \lambda_{P,-Q})(\boldsymbol{P} \wedge \boldsymbol{Q})(f \wedge g). \qquad \square
\end{aligned}$$

We pick a basis for $\mathcal{L}(n.0_E)$, so that now $\Omega(\boldsymbol{P})$ is an $n \times n$ alternating matrix, and $\boldsymbol{P}, \boldsymbol{Q}, d\boldsymbol{P}, d\boldsymbol{Q}$ are column vectors.

Lemma 6.2. *Let* $0_E \neq P_1, \ldots, P_r \in E$ *distinct and* $\xi_1, \ldots, \xi_r \in K$. *Then*

$$\Omega\left(\sum_{i=1}^r \xi_i P_i\right) = \Pi \begin{pmatrix} * & \Xi \\ -\Xi & 0 \end{pmatrix} \Pi^T,$$

where

$$\Xi = \begin{pmatrix} (n-2)\xi_1^2 & -2\xi_1\xi_2 & \cdots & -2\xi_1\xi_r \\ -2\xi_1\xi_2 & (n-2)\xi_2^2 & \cdots & -2\xi_2\xi_r \\ \vdots & \vdots & \ddots & \vdots \\ -2\xi_1\xi_r & -2\xi_2\xi_r & \cdots & (n-2)\xi_r^2 \end{pmatrix} \tag{10}$$

and Π *is the* $n \times 2r$ *matrix with columns* $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_r, d\boldsymbol{P}_1, \ldots, d\boldsymbol{P}_r$.

Proof. Recall that $\Omega = nB - 2A$. The case $r = 2$ is immediate from Lemma 6.1. Since the entries of Ω are quadratic forms the general case follows. $\qquad\square$

We now check the hypotheses (H1), (H2), and (H3).

Proof of (H1) *and* (H3). Suppose $n - 2r \geq 1$. A generic point $P \in \mathrm{Sec}^r C$ may be written $P = \left[\sum_{i=1}^r \xi_i P_i \right]$ for some $0_E \neq P_1, \ldots, P_r \in E$ distinct and $\xi_1, \ldots, \xi_r \neq 0$. By Proposition 4.2 the tangent space $T_P \mathrm{Sec}^r C \subset \mathbb{P}^{n-1}$ is spanned by $P_1, \ldots, P_r, dP_1, \ldots dP_r$. In particular these $2r$ vectors are linearly independent.

For $f \in I(\mathrm{Sec}^r C)$ we have $\sum_{i=1}^n \partial f / \partial x_i(P) v_i = 0$ for any v in the linear span of $P_1, \ldots, P_r, dP_1, \ldots, dP_r$. By Lemma 6.2 the columns of Ω are linear combinations of these vectors. So for each $1 \leq j \leq n$ the form $\sum_{i=1}^n \partial f / \partial x_i \, \Omega_{ij}$ vanishes at P. Since $P \in \mathrm{Sec}^r C$ is generic, this proves (H1). Since $n \notin \{0, 2r\}$ and $\xi_1, \ldots, \xi_r \neq 0$, the matrix (10) is nonsingular. Therefore rank $\Omega(P) = 2r$ and this proves (H3). $\qquad\square$

Proof of (H2). We write $n = 2r$ and $\mathrm{Sec}^{r-1} C = \{F_1 = F_2 = 0\}$, where F_1 and F_2 are forms of degree r. We must show that the form

$$\sum_{i,j=1}^n \frac{\partial F_1}{\partial x_i} \Omega_{ij} \frac{\partial F_2}{\partial x_j} \tag{11}$$

is identically zero. A generic point $P \in \mathrm{Sec}^r C = \mathbb{P}^{n-1}$ may be written $P = \left[\sum_{i=1}^r \xi_i P_i \right]$ for some $0_E \neq P_1, \ldots, P_r \in E$ distinct and $\xi_1, \ldots, \xi_r \neq 0$. In addition we may assume that $2(P_1 + \cdots + P_r) \not\sim H$, where H is the hyperplane section. This ensures that the vectors $P_1, \ldots, P_r, dP_1, \ldots dP_r$ are linearly independent. We choose coordinates on $\mathbb{P}^{n-1}$ so that $[P_1] = (1 : 0 : \cdots : 0)$, $[P_2] = (0 : 1 : 0 : \cdots : 0)$, $\ldots$, $dP_r = (0 : \cdots : 0 : 1)$. Since F_1 and F_2 vanish on $\mathrm{Sec}^{r-1} C$ they vanish on the linear span of any $r - 1$ of the $[P_i]$. Replacing F_1 and F_2 by suitable linear combinations we may assume

$$F_1(x_1, \ldots, x_r, 0, \ldots, 0) = 0, \quad F_2(x_1, \ldots, x_r, 0, \ldots, 0) = x_1 x_2 \ldots x_r.$$

Therefore at $P = (\xi_1 : \cdots : \xi_r : 0 : \cdots : 0)$ we have

$$\left(\frac{\partial F_1}{\partial x_1}(P), \ldots, \frac{\partial F_1}{\partial x_n}(P) \right) = (0, \ldots, 0, *, \ldots, *),$$

$$\left(\frac{\partial F_2}{\partial x_1}(P), \ldots, \frac{\partial F_2}{\partial x_n}(P) \right) = \left(\prod_{i \neq 1} \xi_i, \ldots, \prod_{i \neq r} \xi_i, *, \ldots, * \right).$$

By Lemma 6.2 we have

$$\Omega(P) = \begin{pmatrix} * & \Xi \\ -\Xi & 0 \end{pmatrix},$$

where Ξ is the matrix (10). Since $n = 2r$, the coefficients in each row and column of Ξ sum to zero. Therefore the form (11) vanishes at P. Since $P \in \mathbb{P}^{n-1}$ is generic, this shows that the form is identically zero. $\qquad\square$

This completes the proof of Theorem 1.1.

7. Explicit formulae

In this section we give an explicit formula for the matrix Ω defined in Section 5. As before E is the elliptic curve

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$

with invariant differential $\omega = dx/(2y + a_1 x + a_3)$. We embed E in $\mathbb{P}^{n-1}$ via

$$(x_0 : x_2 : x_3 : \cdots : x_n) = (1, x, y, x^2, xy, x^3, x^2 y, x^4, \ldots). \tag{12}$$

Notice there is no x_1. The indicator function of a set X is denoted $\mathbf{1}_X$. We define linear forms in indeterminates $\{x_m : m \in \mathbb{Z}\}$ as follows:

$$\dot{x}_m = \tfrac{1}{2} m (2x_{m+1} + a_1 x_m + a_3 x_{m-2}) + \mathbf{1}_{\text{odd}}(m) \sum_{i=1}^{6} (-1)^i \big(m - \tfrac{1}{2} i\big) a_i x_{m+1-i},$$

$$\bar{x}_m = \tfrac{1}{2} (2x_{m+1} + a_1 x_m + a_3 x_{m-2}) + \mathbf{1}_{\text{odd}}(m) \sum_{i=1}^{6} (-1)^i a_i x_{m+1-i},$$

where by convention $a_5 = 0$. The relation to the notation $\dot{f} = df/\omega$ used in Section 5 will be explained below. For $x \in \mathbb{R}$ we let $\text{sign}(x) = -1, 0, 1$ according as x is negative, zero, or positive. For $r, s \in \mathbb{Z}$ we define

$$A_{rs} = x_r \dot{x}_s - x_s \dot{x}_r, \quad B_{rs} = \sum_{k=-\infty}^{\infty} \text{sign}\big(k + \tfrac{1}{2}\big)(x_{r+2k} \bar{x}_{s-2k} - x_{s+2k} \bar{x}_{r-2k}).$$

Theorem 7.1. *Let $C \subset \mathbb{P}^{n-1}$ be the image of E under the embedding* (12):

(i) *$A = (A_{rs})_{r,s=0,2,\ldots,n}$ and $B = (B_{rs})_{r,s=0,2,\ldots,n}$ are $n \times n$ alternating matrices of quadratic forms in $x_0, x_2, \ldots, x_{n+1}$.*

(ii) *$\Omega = nB - 2A$ is an $n \times n$ alternating matrix of quadratic forms in $x_0, x_2, \ldots, x_n$. It satisfies the conclusions of Theorem 1.1 and*

$$(n - 2)\omega = \frac{x_j^2 d(x_i/x_j)}{\Omega_{ij}(x_1, \ldots, x_n)} \quad \text{for all } i \neq j. \tag{13}$$

Proof. It is part of the theorem that the indeterminates x_m for $m \notin \{0, 2, 3, \ldots, n\}$ cancel from the formula for Ω. So when applying the theorem we simply set them to be zero. However we will not do this in the proof. Since $\bar{x}_m$ is a linear combination of $x_{m+1}, x_m, \ldots, x_{m-5}$, each B_{rs} is of the form $\sum_{ij} c_{ij} x_i x_j$, where each c_{ij} is a finite sum. But it is not immediately clear that the B_{rs} are polynomials, i.e., that $c_{ij} = 0$ for all but finitely many pairs (i, j). We check this first.

If $r \equiv s \pmod 2$ and $r < s$ then

$$B_{rs} = 2(x_r \bar{x}_s + x_{r+2} \bar{x}_{s-2} + \ldots + x_{s-2} \bar{x}_{r+2}), \tag{14}$$

whereas if r is even and s is odd then

$$B_{rs} = -a_1 x_r x_s + Q_{r,s+1} + a_2 Q_{r,s-1} + a_4 Q_{r,s-3} + a_6 Q_{r,s-5} - Q_{s,r+1}, \tag{15}$$

where

$$Q_{ij} = \begin{cases} x_i x_j + x_{i+2} x_{j-2} + \cdots + x_j x_i & \text{if } i < j+2, \\ 0 & \text{if } i = j+2, \\ -(x_{i-2} x_{j+2} + x_{i-4} x_{j+4} + \ldots + x_{j+2} x_{i-2}) & \text{if } i > j+2. \end{cases}$$

Since $B_{sr} = -B_{rs}$ this proves that the B_{rs} are polynomials.

We show that the matrices A and B defined in the statement of the theorem represent the linear maps A and B defined in Section 5. The theorem then follows from the results of Sections 4, 5, and 6. In particular (13) follows from Lemma 5.2.

In the statement of the theorem the $\{x_m : m \in \mathbb{Z}\}$ are indeterminates. However for the proof they will be the following rational functions on E,

$$x_m = \begin{cases} x^{m/2} & \text{if } m \text{ is even,} \\ x^{(m-3)/2} y & \text{if } m \text{ is odd.} \end{cases}$$

As rational functions on E, we claim that $\dot{x}_m = dx_m / \omega$ (in agreement with the notation in Section 5) and $\bar{x}_m = \frac{1}{2} x_{m-2}(2y + a_1 x + a_3)$. In checking these claims, we start with the right-hand sides, since this also serves to motivate the definitions of $\dot{x}_m$ and $\bar{x}_m$. For m even we have

$$\begin{aligned}
dx_m / \omega &= \tfrac{1}{2} m x^{(m-2)/2} (dx/\omega) \\
&= \tfrac{1}{2} m x^{(m-2)/2} (2y + a_1 x + a_3) \\
&= \tfrac{1}{2} m (2x_{m+1} + a_1 x_m + a_3 x_{m-2}),
\end{aligned}$$

and

$$\tfrac{1}{2} x_{m-2}(2y + a_1 x + a_3) = \tfrac{1}{2}(2x_{m+1} + a_1 x_m + a_3 x_{m-2}).$$

For m odd we have

$$\begin{aligned}
dx_m / \omega &= \tfrac{1}{2}(m-3)x^{(m-5)/2} y (dx/\omega) + x^{(m-3)/2}(dy/\omega) \\
&= \tfrac{1}{2}(m-3)x^{(m-5)/2}(2y^2 + a_1 xy + a_3 y) + x^{(m-3)/2}(3x^2 + 2a_2 x + a_4 - a_1 y) \\
&= \tfrac{1}{2}(m-3)x^{(m-5)/2}(-a_1 xy - a_3 y + 2x^3 + 2a_2 x^2 + 2a_4 x + 2a_6) \\
&\qquad\qquad\qquad\qquad + x^{(m-5)/2}(3x^3 + 2a_2 x^2 + a_4 x - a_1 xy) \\
&= x^{(m-5)/2}\Big(mx^3 - \tfrac{1}{2}(m-1)a_1 xy - \tfrac{1}{2}(m-3)a_3 y + \sum_{i=1}^{3}(m-i)a_{2i} x^{3-i}\Big) \\
&= \tfrac{1}{2} m(2x_{m+1} + a_1 x_m + a_3 x_{m-2}) + \sum_{i=1}^{6}(-1)^i \big(m - \tfrac{1}{2}i\big) a_i x_{m+1-i},
\end{aligned}$$

and

$$\tfrac{1}{2}x_{m-2}(2y+a_1x+a_3) = \tfrac{1}{2}x^{(m-5)/2}(2y^2+a_1xy+a_3y)$$
$$= \tfrac{1}{2}x^{(m-5)/2}(-a_1xy-a_3y+2x^3+2a_2x^2+2a_4x+2a_6)$$
$$= \tfrac{1}{2}(2x_{m+1}+a_1x_m+a_3x_{m-2}) + \sum_{i=1}^{6}(-1)^i a_i x_{m+1-i}.$$

It is now clear that $A(x_r \wedge x_s) = A_{rs}$ for all $r, s \in \mathbb{Z}$. It remains to prove the same for B. By definition of B we have

$$B(x_r \wedge x_s) = \frac{y_P + y_Q + a_1x_Q + a_3}{x_P - x_Q}(x_r(Q)x_s(P) - x_r(P)x_s(Q))\big|_{P=Q},$$

where P, Q are points on E. Since $\bar{x}_m = \tfrac{1}{2}x_{m-2}(2y+a_1x+a_3)$ we have

$$2x_r(P)\bar{x}_s(Q) = (2y_Q+a_1x_Q+a_3)x_r(P)x_{s-2}(Q)$$
$$= \frac{2y_Q+a_1x_Q+a_3}{x_P-x_Q}(x_{r+2}(P)x_{s-2}(Q) - x_r(P)x_s(Q)).$$

Adding this to the same expression with (r, s) replaced by $(s - 2, r + 2)$ and then setting $P = Q$ gives

$$B_{rs} - B_{r+2,s-2} = 2(x_r\bar{x}_s + x_{s-2}\bar{x}_{r+2}) = B(x_r \wedge x_s) - B(x_{r+2} \wedge x_{s-2}). \tag{16}$$

Rather more obviously, replacing (r, s) by $(r + 2, s + 2)$ changes B_{rs} and $B(x_r \wedge x_s)$ in the same way, that is, by shifting the subscripts up by 2. So to prove $B(x_r \wedge x_s) = B_{rs}$ for all $r, s \in \mathbb{Z}$ it suffices to prove it for all $r \in \{0, 1\}$ and $s \in \{0, 1, 2, 3\}$. This is a finite calculation. We give two examples:

$$B(x_0 \wedge x_3) = \frac{y_P + y_Q + a_1x_Q + a_3}{x_P - x_Q}(y_P - y_Q)\big|_{P=Q}$$
$$= \frac{(y_P^2 + a_1x_Py_P + a_3y_P) - (y_Q^2 + a_1x_Qy_Q + a_3y_Q)}{x_P - x_Q} - a_1y_P\big|_{P=Q}$$
$$= (x_P^2 + x_Px_Q + x_Q^2 - a_1y_P + a_2(x_P + x_Q) + a_4)\big|_{P=Q}$$
$$= 2x_0x_4 + x_2^2 - a_1x_0x_3 + 2a_2x_0x_2 + a_4x_0^2,$$

and

$$B(x_2 \wedge x_3) = \frac{y_P + y_Q + a_1x_Q + a_3}{x_P - x_Q}(y_P(x_Q - x_P) + x_P(y_P - y_Q))\big|_{P=Q}$$
$$= (-y_P(y_P + y_Q + a_1x_Q + a_3) + x_P(x_P^2 + x_Px_Q + \cdots + a_4))\big|_{P=Q}$$
$$= (x_P^2x_Q + x_Px_Q^2 - y_Py_Q - a_1x_Qy_P + a_2x_Px_Q - a_6x_0^2)\big|_{P=Q}$$
$$= 2x_2x_4 - x_3^2 - a_1x_2x_3 + a_2x_2^2 - a_6x_0^2.$$

It is easy to check using (15) that these are equal to B_{03} and B_{23}. The other cases we need can then be checked using (16) and the fact that B is alternating. $\qquad\square$

8. Proof of Theorem 1.2

Let $\Omega = nB - 2A$ be as in Theorem 7.1. Then $c_4(\Omega) = f_n(a_1, \ldots, a_6)$ and $c_6(\Omega) = g_n(a_1, \ldots, a_6)$ for some polynomials f_n and g_n. We consider the effect of a change of Weierstrass equation, with notation as in [Silverman 2009, Chapter III].

Lemma 8.1. *Let $a_1, \ldots, a_6$ and $a_1', \ldots, a_6'$ be the coefficients of two Weierstrass equations related by $x = u^2 x' + r$ and $y = u^3 y' + u^2 s x' + t$. Then*

$$f_n(a_1, \ldots, a_6) = u^4 f_n(a_1', \ldots, a_6'), \quad g_n(a_1, \ldots, a_6) = u^6 g_n(a_1', \ldots, a_6').$$

Proof. This follows from Corollary 2.3 and $u^{-1}\omega' = \omega$. □

It follows by Lemma 8.1, and the standard procedure for converting a Weierstrass equation to the shorter form $y^2 = x^3 + ax + b$, that f_n and g_n are scalar multiples of the usual polynomials c_4 and c_6 in $a_1, \ldots, a_6$. Explicitly,

$$
\begin{aligned}
f_n(a_1, \ldots, a_6) &= \xi_n(b_2^2 - 24b_4) = \xi_n(a_1^4 + \cdots), \\
g_n(a_1, \ldots, a_6) &= \eta_n(-b_2^3 + 36b_2b_4 - 216b_6) = \eta_n(-a_1^6 + \cdots),
\end{aligned}
\tag{17}
$$

where $b_2 = a_1^2 + 4a_2$, $b_4 = 2a_4 + a_1a_3$ and $b_6 = a_3^2 + 4a_6$.

To complete the proof of Theorem 1.2 we must compute the constants ξ_n and η_n. For any given value of n these can be read off from a single numerical example. However we need to compute these constants for all n. We write

$$\Omega = \Omega^{(0)} + a_1\Omega^{(1)} + a_2\Omega^{(2)} + a_3\Omega^{(3)} + a_4\Omega^{(4)} + a_6\Omega^{(6)}.$$

Since $c_4(\Omega)$ and $c_6(\Omega)$ have degrees 4 and 6 in the coefficients of the entries of Ω, we see by (17) that it suffices to compute the invariants of $\Omega^{(1)}$.

We put

$$\gamma_{rs} = (-1)^{\max(r,s)} \operatorname{sign}(s-r)n - 2\left((-1)^s \left\lfloor \tfrac{1}{2}s \right\rfloor - (-1)^r \left\lfloor \tfrac{1}{2}r \right\rfloor\right).$$

Lemma 8.2. *The alternating matrix $\Omega^{(1)}$ has entries above the diagonal*

$$\gamma_{rs}x_rx_s + (-1)^s n \mathbf{1}_{\text{even}}(r+s) \sum_{k=1}^{(s-r)/2-1} x_{r+2k}x_{s-2k}.
\tag{18}$$

Proof. Since $\Omega = nB - 2A$ we have $\Omega^{(1)} = nB^{(1)} - 2A^{(1)}$, where the superscripts indicate that we are taking the coefficient of a_1. Then $A^{(1)}$ has (r, s) entry

$$\left((-1)^s \left\lfloor \tfrac{1}{2}s \right\rfloor - (-1)^r \left\lfloor \tfrac{1}{2}r \right\rfloor\right)x_rx_s,$$

whereas (14) and (15) show that if $r < s$ then $B^{(1)}$ has (r, s) entry

$$
\begin{cases}
(-1)^s(x_rx_s + x_{r+2}x_{s-2} + \ldots x_{s-2}x_{r+2}) & \text{if } r \equiv s \pmod 2, \\
(-1)^s x_rx_s & \text{if } r \not\equiv s \pmod 2.
\end{cases}
$$
□

Lemma 8.3. *The matrices* $\Omega^{(1)}$, $\Omega' = (\gamma_{rs} x_r x_s)_{r,s=0,2,3,\ldots,n}$ *and*

$$\Lambda = \left((\text{sign}(j-i)n - 2(j-i))x_i x_j\right)_{i,j=0,1,\ldots,n-1}$$

all have the same invariants c_4 *and* c_6.

Proof. We first explain why $\Omega^{(1)}$ and Ω' have the same invariants, despite the "extra terms" in (18). We start with $\Omega^{(1)}$. The only entries involving x_0 are in the first row and column. We replace x_0 by $\lambda^{-1}x_0$ and multiply the first row and column by λ. By Lemma 2.2 this does not change the invariants, but setting $\lambda = 0$ removes the extra terms from the first row and column. Now the only entries involving x_2 are in the second row and column. We replace x_2 by $\lambda^{-1}x_2$ and multiply the second row and column by λ. This does not change the invariants, but setting $\lambda = 0$ removes the extra terms from the second row and column. We repeat this procedure for all subsequent rows and columns. In the end we remove all the extra terms, and are left with the matrix Ω'.

We define a bijection $\pi : \{0, 1, \ldots, n-1\} \to \{0, 2, 3, \ldots, n\}$ by

$$\pi(i) = \begin{cases} 2i & \text{if } i \leq n/2, \\ 2(n-i)+1 & \text{if } i > n/2. \end{cases}$$

We then compute

$$\gamma_{\pi(i),\pi(j)} = \begin{cases} \text{sign}(j-i)n - 2(j-i) & \text{if } i \leq n/2 \text{ and } j \leq n/2, \\ -n - 2(-(n-j)-i) & \text{if } i \leq n/2 \text{ and } j > n/2, \\ n - 2(j+(n-i)) & \text{if } i > n/2 \text{ and } j \leq n/2, \\ \text{sign}(j-i)n - 2(-(n-j)+(n-i)) & \text{if } i > n/2 \text{ and } j > n/2. \end{cases}$$

In all cases we have $\gamma_{\pi(i),\pi(j)} = \text{sign}(j-i)n - 2(j-i)$. Therefore Ω' and Λ are related by a permutation matrix. It follows by Lemma 2.2 that they have the same invariants. $\qquad\square$

Lemma 8.4. *The alternating matrix of quadratic forms*

$$\Lambda = \begin{pmatrix} 0 & (n-2)x_1x_2 & (n-4)x_1x_3 & (n-6)x_1x_4 & \cdots & (2-n)x_1x_n \\ & 0 & (n-2)x_2x_3 & (n-4)x_2x_4 & \cdots & (4-n)x_2x_n \\ & & 0 & (n-2)x_3x_4 & \cdots & (6-n)x_3x_n \\ & - & & \ddots & \ddots & \vdots \\ & & & & & (n-2)x_{n-1}x_n \\ & & & & & 0 \end{pmatrix}$$

has invariants $c_4(\Lambda) = (n-2)^4$ *and* $c_6(\Lambda) = -(n-2)^6$.

Proof. We have $\Lambda = (\lambda_{rs}x_r x_s)_{r,s=1,\ldots,n}$, where $\lambda_{rs} = \text{sign}(s-r)n - 2(s-r)$. Following the definitions of c_4 and c_6 in Section 1 we put

$$M_{ij} = \sum_{r,s=1}^{n} \frac{\partial \Lambda_{ir}}{\partial x_s} \frac{\partial \Lambda_{js}}{\partial x_r} = \mu_{ij}x_i x_j, \quad N_{ijk} = \sum_{r=1}^{n} \frac{\partial M_{ij}}{\partial x_r} \Lambda_{rk} = \nu_{ijk}x_i x_j x_k,$$

where $\mu_{ij} = \left(\sum_{r=1}^{n} \lambda_{ir} \lambda_{jr} \right) - \lambda_{ij}^2$ and $v_{ijk} = \mu_{ij}(\lambda_{ik} + \lambda_{jk})$. It is not hard to show that

$$\sum_{r=1}^{n} \text{sign}(i-r) \, \text{sign}(j-r) = n - 2|i-j| - \delta_{ij},$$

$$\sum_{r=1}^{n} (i-r) \, \text{sign}(j-r) = 2ij - j^2 - (n+1)i + n(n+1)/2,$$

$$\sum_{r=1}^{n} (i-r)(j-r) = nij - (i+j)n(n+1)/2 + n(n+1)(2n+1)/6.$$

We use these to compute

$$\sum_{r=1}^{n} \lambda_{ir} \lambda_{jr} = 2n|i-j|^2 - 2n^2|i-j| - \delta_{ij}n^2 + (n^3 + 2n)/3$$

and then subtract off

$$\lambda_{ij}^2 = 4|i-j|^2 - 4n|i-j| + (1 - \delta_{ij})n^2$$

to get

$$\mu_{ij} = 2(n-2)(|i-j|^2 - n|i-j|) + n(n-1)(n-2)/3.$$

Noting the symmetries $\mu_{ij} = \mu_{ji}$ and $v_{ijk} = v_{jik}$, and using computer algebra to check our calculations, we find

$$\sum_{i,j,r,s=1}^{n} \frac{\partial^2 M_{ij}}{\partial x_r \partial x_s} \frac{\partial^2 M_{rs}}{\partial x_i \partial x_j} = 4 \sum_{i \leq j} \mu_{ij}^2 = \left(\tfrac{16}{3} \right) n(n-2)^2 \binom{n+3}{5}$$

and

$$\sum_{i,j,k,r,s,t=1}^{n} \frac{\partial^3 N_{ijk}}{\partial x_r \partial x_s \partial x_t} \frac{\partial^3 N_{rst}}{\partial x_i \partial x_j \partial x_k}$$

$$= 4 \sum_{i \leq j \leq k} (v_{ijk} + v_{jki} + v_{kij})^2$$

$$= 4 \sum_{i \leq j \leq k} (\lambda_{ij}(\mu_{ik} - \mu_{jk}) + \lambda_{jk}(\mu_{ij} - \mu_{ik}) + \lambda_{ik}(\mu_{ij} - \mu_{jk}))^2$$

$$= 64(n-2)^2 \sum_{i \leq j \leq k} (i - 2j + k)^2 (n+i+j-2k)^2 (n+2i-j-k)^2$$

$$= 64n(n-2)^3 \binom{n+5}{7}.$$

The final sums are evaluated using the standard formulae for $\sum_{i=1}^{n} i$, $\sum_{i=1}^{n} i^2$, etc. In practice it is simpler to observe that the answer is a polynomial in n, say of degree at most d, and then check the result for $d+1$ distinct values of n.

Finally scaling by the constants included in the definitions (4) and (5) it follows that $c_4(\Lambda) = (n-2)^4$ and $c_6(\Lambda) = -(n-2)^6$. □

The last two lemmas show that $\xi_n = (n-2)^4$ and $\eta_n = (n-2)^6$. Therefore $c_4(\Omega) = (n-2)^4 c_4(E)$ and $c_6(\Omega) = (n-2)^6 c_6(E)$. Let $\omega = dx/(2y + a_1 x + a_3)$. By the formulae in [Silverman 2009, Chapter III] we have

$$(E, \omega) \cong (y^2 = x^3 - 27c_4(E)x - 54c_6(E), 3dx/y).$$

Therefore

$$(E, (n-2)\omega) \cong (y^2 = x^3 - 27c_4(\Omega)x - 54c_6(\Omega), 3dx/y).$$

Recalling from Theorem 7.1 that $\Omega = nB - 2A$ represents the invariant differential $(n-2)\omega$, this completes the proof of Theorem 1.2.

9. Higher secant varieties

In this final section we give references and proofs for the facts about higher secant varieties we used earlier in the paper.

Theorem 9.1. *Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree $n \geq 3$:*

(i) *$\operatorname{Sec}^r C \subset \mathbb{P}^{n-1}$ is an irreducible variety of codimension $\max(n - 2r, 0)$.*

(ii) *The vector space of forms of degree $r + 1$ vanishing on $\operatorname{Sec}^r C$ has dimension $\beta(r + 1, n)$, where*

$$\beta(r, n) = \binom{n-r}{r} + \binom{n-r-1}{r-1}$$

is the number of ways of choosing r elements from $\mathbb{Z}/n\mathbb{Z}$ such that no two elements are adjacent.

(iii) *If $n - 2r \geq 2$ then the homogeneous ideal $I(\operatorname{Sec}^r C)$ is generated by forms of degree $r + 1$.*

(iv) *If $n - 2r = 1$ then $\operatorname{Sec}^r C$ is a hypersurface of degree n.*

(v) *If $n - 2r \geq 1$ then $\operatorname{Sec}^r C$ has singular locus $\operatorname{Sec}^{r-1} C$.*

Proof. (i) This is a general fact about curves. See for example [Lange 1984, §1].

(ii), (iii), (iv) More generally the minimal free resolution for $I(\operatorname{Sec}^r C)$ was computed in [Graf v. Bothmer and Hulek 2004, §8]. See [Gross and Popescu 1998, §5] for the cases $r = 1, 2$, and [Fisher 2010, §4] for further discussion.

(v) This is [Graf v. Bothmer and Hulek 2004, Proposition 8.15]. □

9.1. *Computing equations for higher secant varieties.* The following two propositions may be used to compute equations for $\operatorname{Sec}^r C$ from equations for C. We say that a form f vanishes on C with multiplicity r if (passing to affine coordinates) the Taylor expansion of f at each point $P \in C$ begins with terms of order greater than or equal to r.

Proposition 9.2. *Let $C \subset \mathbb{P}^{n-1}$ be a variety contained in no hyperplane. Let f be a form of degree $r + 1$:*

(i) *If $r \geq 1$ then*

$$f \in I(\operatorname{Sec}^r C) \iff f \text{ vanishes on } C \text{ with multiplicity } r.$$

(ii) *If $r \geq 2$ then*

$$f \in I(\mathrm{Sec}^r C) \Longleftrightarrow \frac{\partial f}{\partial x_i} \in I(\mathrm{Sec}^{r-1} C) \quad \textit{for all } i = 1, \ldots, n.$$

Proof. (i) We choose $P_1, \ldots, P_n \in C$ spanning $\mathbb{P}^{n-1}$. By a change of coordinates we may assume $P_1 = (1 : 0 : \cdots : 0)$, $P_2 = (0 : 1 : 0 : \cdots : 0), \ldots, P_n = (0 : 0 : \cdots : 1)$. If $f \in I(\mathrm{Sec}^r C)$ then it vanishes on the linear span of any r of the P_i. Therefore the monomials appearing in f involve at least $r + 1$ of the x_i, and since f has degree $r + 1$ must be squarefree. But then f vanishes at P_1 with multiplicity r. Since $P_1 \in C$ was arbitrary it follows that f vanishes on C with multiplicity r.

Conversely, suppose f vanishes on C with multiplicity r. Let Π be an $(r - 1)$-plane spanned by points $P_1, \ldots, P_r \in C$. By a change of coordinates we may assume $P_1 = (1 : 0 : \ldots : 0)$, $P_2 = (0 : 1 : 0 : \cdots : 0), \ldots$ Then $f(x_1, \ldots, x_r, 0, \ldots, 0)$ has total degree $r + 1$, but has degree at most 1 in each of the variables. It follows that f vanishes on Π. By definition $\mathrm{Sec}^r C$ is the Zariski closure of the union of all such $(r - 1)$-planes. Therefore $f \in I(\mathrm{Sec}^r C)$ as required.

(ii) Since $\mathrm{char}(K) = 0$ this follows from (i). $\qquad\square$

Now let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve. Taking $r = 1$ in Theorem 9.1 shows that the homogeneous ideal $I(C)$ is generated by a vector space of quadrics of dimension $n(n - 3)/2$. Suppose we know a basis for this space. Then by repeatedly applying Proposition 9.2(ii) we can find a basis for the space of forms of degree $r + 1$ vanishing on $\mathrm{Sec}^r C$. Theorem 9.1(iii) tells us that if $n - 2r \geq 2$ then these forms define $\mathrm{Sec}^r C$. The following proposition covers the remaining case:

Proposition 9.3. *Suppose $n - 2r = 1$. Let f be a form of degree n. If $r \geq 2$ then*

$$f \in I(\mathrm{Sec}^r C) \Longleftrightarrow \frac{\partial f}{\partial x_i} \in I(\mathrm{Sec}^{r-1} C)^2 \quad \textit{for all } i = 1, \ldots, n.$$

Proof. For $\Rightarrow$: Let H be the divisor of a hyperplane section, and let $P \in C$ be any point. Let $C_+ \subset \mathbb{P}^n$ and $C_- \subset \mathbb{P}^{n-2}$ be the images of C embedded via the linear systems $|H \pm P|$. We choose coordinates so that the isomorphisms $C_+ \to C \to C_-$ are given by

$$(x_1 : \cdots : x_{n+1}) \mapsto (x_1 : \cdots : x_n) \mapsto (x_1 : \cdots : x_{n-1}).$$

In particular P is the point $(x_1 : \cdots : x_n) = (0 : \cdots : 0 : 1)$. By Theorem 9.1 we know that $I(\mathrm{Sec}^{r-1} C_-)$ is generated by forms $g_1, g_2 \in K[x_1, \ldots, x_{n-1}]$ of degree r. By [Fisher 2010, Corollary 2.3] there exist forms $h_1, h_2 \in K[x_1, \ldots, x_n]$ of degree $r + 1$ such that $f_i = x_{n+1} g_i + h_i \in I(\mathrm{Sec}^r C_+)$ for $i = 1, 2$. Then $F = g_1 h_2 - g_2 h_1$ belongs to

$$I(\mathrm{Sec}^r C_+) \cap K[x_1, \ldots, x_n] = I(\mathrm{Sec}^r C).$$

Since g_1, g_2 are coprime and f_1, f_2 are irreducible it is clear that F is nonzero. By Theorem 9.1(iv) we have $I(\mathrm{Sec}^r C) = (F)$. We compute

$$\frac{\partial F}{\partial x_n} = \frac{\partial f_1}{\partial x_{n+1}} \frac{\partial f_2}{\partial x_n} - \frac{\partial f_1}{\partial x_n} \frac{\partial f_2}{\partial x_{n+1}}.$$

On the other hand, for $i = 1, 2$ and $j = n, n + 1$ we have

$$\frac{\partial f_i}{\partial x_j} \in I(\operatorname{Sec}^{r-1} C_+) \cap K[x_1, \ldots, x_n] = I(\operatorname{Sec}^{r-1} C).$$

Therefore $\partial F / \partial x_n \in I(\operatorname{Sec}^{r-1} C)^2$. Since $P \in C$ was arbitrary, and C spans $\mathbb{P}^{n-1}$, the result follows.

For $\Leftarrow$: Let $P_1, \ldots, P_r$ be r distinct points on C. By a change of coordinates we may assume $P_1 = (1 : 0 : \cdots : 0)$, $P_2 = (0 : 1 : 0 : \cdots : 0)$, ... By Proposition 9.2 we know that f vanishes on C with multiplicity $2(r - 1) + 1 = n - 2$. Therefore $f(x_1, \ldots, x_r, 0, \ldots, 0)$ has total degree n, but has degree at most 2 in each of the variables. Since $2r < n$ it follows that f vanishes on the linear span of $P_1, \ldots, P_r$. By definition $\operatorname{Sec}^r C$ is the Zariski closure of the union of all such $(r - 1)$-planes. Therefore $f \in I(\operatorname{Sec}^r C)$ as required. $\qquad\square$

9.2. *Proof of Proposition 4.2.* Let $C \subset \mathbb{P}^{n-1}$ be a genus one normal curve of degree n. Let H be the divisor of a hyperplane section. We identify $\mathcal{L}(H)$ with the space of linear forms on $\mathbb{P}^{n-1}$. For D an effective divisor on C we write $\bar{D} \subset \mathbb{P}^{n-1}$ for the linear subspace cut out by $\mathcal{L}(H - D) \subset \mathcal{L}(H)$. We have

$$\operatorname{Sec}^r C = \bigcup_{\deg D = r} \bar{D}.$$

We also put $D^\circ = \bar{D} \setminus \cup_{D' < D} \bar{D'}$. The gcd and lcm of divisors $\sum m_P P$ and $\sum m'_P P$ are $\sum \min(m_P, m'_P) P$ and $\sum \max(m_P, m'_P) P$.

Lemma 9.4. *Let D, D_1, D_2 be effective divisors on C:*

  (i) *If $\deg D < n$ then $\dim \bar{D} = \deg D - 1$.*

 (ii) *The linear span of $\bar{D_1}$ and $\bar{D_2}$ is $\overline{\operatorname{lcm}(D_1, D_2)}$.*

(iii) *If $\deg(\operatorname{lcm}(D_1, D_2)) < n$ then $\bar{D_1} \cap \bar{D_2} = \overline{\gcd(D_1, D_2)}$.*

Proof. (i) By Riemann–Roch we have $\dim \mathcal{L}(H - D) = n - \deg D$.

(ii) We have $\mathcal{L}(H - D_1) \cap \mathcal{L}(H - D_2) = \mathcal{L}(H - \operatorname{lcm}(D_1, D_2))$.

(iii) The inclusion "$\supset$" is clear. Equality follows by counting dimensions using (i) and (ii). $\qquad\square$

With the above notation, Proposition 4.2 becomes

Proposition 9.5. *Suppose $n - 2r \geq 1$. Let $D = P_1 + \cdots + P_r$ be an effective divisor of degree r with $P_1, \ldots, P_r \in C$ distinct. Then for any $P \in D^\circ$ we have $T_P \operatorname{Sec}^r C = \overline{2D}$.*

Proof. If $P \in \bar{D'}$ for D' an effective divisor of degree at most r, then by Lemma 9.4(iii) we have $D = D'$. In particular $P \notin \operatorname{Sec}^{r-1} C$. It follows by Theorem 9.1(v) that P is a smooth point on $\operatorname{Sec}^r C$. The next lemma shows that $\overline{2D} \subset T_P \operatorname{Sec}^r C$, and equality follows by comparing dimensions, using Lemma 9.4(i) and Theorem 9.1(i). $\qquad\square$

Lemma 9.6. *Let X be an affine variety and $P_1, \ldots, P_r \in X$. Let $P = \sum \xi_i P_i$, where $\sum \xi_i = 1$. If $\xi_i \neq 0$ then $T_{P_i} X \subset T_P(\mathrm{Sec}^r X)$.*

Proof. There is a morphism $X \times \cdots \times X \to \mathrm{Sec}^r X$; $(a_1, \ldots, a_r) \mapsto \sum \xi_i a_i$ with derivative $T_{P_1} X \times \cdots \times T_{P_r} X \to T_P(\mathrm{Sec}^r X)$; $(b_1, \ldots, b_r) \mapsto \sum \xi_i b_i$. $\square$

In fact Proposition 9.5 is true without the hypothesis that $P_1, \ldots, P_r$ are distinct. However, since we do not need this, we omit the details.

9.3. *Proof of Proposition 4.3.* We must prove the following:

Proposition 9.7. *Suppose $n - 2r = 2$ and write $\mathrm{Sec}^r C = \{F_1 = F_2 = 0\}$. Then the variety $X \subset \mathbb{P}^{n-1}$ defined by*

$$\mathrm{rank} \begin{pmatrix} \partial F_1/\partial x_1 & \cdots & \partial F_1/\partial x_n \\ \partial F_2/\partial x_1 & \cdots & \partial F_2/\partial x_n \end{pmatrix} \leq 1$$

has codimension 3.

If $n = 4$ then $C = \{F_1 = F_2 = 0\} \subset \mathbb{P}^3$ is the intersection of two quadrics. There are four singular quadrics in the pencil spanned by F_1 and F_2, and each is singular at just one point. Then X is the union of these four singular points, and so has codimension 3.

We now generalise this argument. Let H be the divisor of a hyperplane section. We identify $\mathcal{L}(H)$ with the space of linear forms on $\mathbb{P}^{n-1}$. Let D_1 and D_2 be divisors on C of degree $r + 1$ with $D_1 + D_2 = H$. Let $\Phi(D_1, D_2)$ be the $(r + 1) \times (r + 1)$ matrix of linear forms representing the multiplication map

$$\mathcal{L}(D_1) \times \mathcal{L}(D_2) \to \mathcal{L}(H).$$

Since $\Phi(D_1, D_2)$ has rank at most 1 on C, it has rank at most r on $\mathrm{Sec}^r C$. Therefore $\det \Phi(D_1, D_2)$ is a form of degree $r + 1$ vanishing on $\mathrm{Sec}^r C$. In particular it belongs to the pencil spanned by F_1 and F_2.

Lemma 9.8. *Every linear combination of F_1 and F_2 arises in this way. Moreover there are exactly four forms in the pencil arising as $\det \Phi(D_1, D_2)$ with $D_1 \sim D_2$.*

Proof. We say that divisor pairs (D_1, D_2) and (D_1', D_2') are *equivalent* if $D_1 \sim D_1'$ or $D_1 \sim D_2'$. It is shown in [Fisher 2010, Lemma 2.9] that if (D_1, D_2) and (D_1', D_2') are inequivalent then $\mathrm{Sec}^r C = \{\det \Phi(D_1, D_2) = \det \Phi(D_1', D_2') = 0\} \subset \mathbb{P}^{n-1}$. In particular these two forms are linearly independent.

We claim that the map $(D_1, D_2) \mapsto \Phi(D_1, D_2)$ is a bijection between the equivalence classes of divisor pairs and the pencil of forms spanned by F_1 and F_2. To prove this let C be the image of an elliptic curve E embedded in $\mathbb{P}^{n-1}$ by $|n.0_E|$. Then writing

$$\det \Phi(r.0_E + P, (r + 2).0_E - P) = s(P) F_1 + t(P) F_2,$$

for $P \in E$, we can see that s/t is a rational function on E. It therefore defines a morphism $(s : t) : E \to \mathbb{P}^1$. By the previous paragraph, this morphism is nonconstant, and indeed has fibres of the form $\{P, -P\}$. It must therefore be surjective. This proves the claim.

For the final statement we note that $r.0_E + P \sim (r + 2).0_E - P$ if and only if $P \in E[2]$. $\square$

Lemma 9.9. *Let S be the singular locus of $V = \{\det \Phi(D_1, D_2) = 0\} \subset \mathbb{P}^{n-1}$. Then S contains $\operatorname{Sec}^{r-1} C$. Moreover*:

(i) *If $D_1 \not\sim D_2$ then $S = \operatorname{Sec}^{r-1} C$.*

(ii) *If $D_1 \sim D_2$ then S has codimension* 3.

Proof. Since C spans $\mathbb{P}^{n-1}$ it is clear that for each $P \in \operatorname{Sec}^{r-1} C$ we have $T_P \operatorname{Sec}^r C = \mathbb{P}^{n-1}$. Therefore S contains $\operatorname{Sec}^{r-1} C$.

(i) Let $P \in V \setminus \operatorname{Sec}^{r-1} C$ be any point. According to [Fisher 2010, Theorem 1.3] the $r \times r$ minors of $\Phi(D_1, D_2)$ generate $I(\operatorname{Sec}^{r-1} C)$. Therefore evaluating $\Phi(D_1, D_2)$ at P gives a matrix of rank r. Moving P to $(1 : 0 : \cdots : 0)$ and picking suitable bases for $\mathcal{L}(D_1)$ and $\mathcal{L}(D_2)$ we have

$$\Phi(D_1, D_2) = x_1 \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} + \Phi',$$

where Φ' is an $(r+1) \times (r+1)$ matrix of linear forms in $x_2, \ldots, x_n$. Now the top left entry of $\Phi(D_1, D_2)$ is an equation for $T_P V$. Since the product of nonzero rational functions on C is again nonzero, the entries of $\Phi(D_1, D_2)$ are nonzero. Therefore $P \in V$ is a smooth point.

(ii) Picking suitable bases for $\mathcal{L}(D_1)$ and $\mathcal{L}(D_2)$ we may suppose that $\Phi(D_1, D_2)$ is symmetric. Since $\{\operatorname{rank} \Phi(D_1, D_2) \leq r - 1\} \subset S$, and the quadratic forms of rank at most $m - 2$ have codimension 3 in the space of all quadratic forms in m variables, it follows that S has codimension at most 3. Suppose for a contradiction that S has codimension at most 2. Then its intersection with $\operatorname{Sec}^r C = \{F_1 = F_2 = 0\}$ has codimension at most 3. But this intersection is contained in the singular locus of $\operatorname{Sec}^r C$, which by Theorem 9.1 has codimension 4. This is the required contradiction. $\square$

To complete the proof of Proposition 9.7, we note that X is the union of the singular loci of the hypersurfaces defined by linear combinations of F_1 and F_2. It follows by Lemmas 9.8 and 9.9 that X has codimension 3.

References

[An et al. 2001] S. Y. An, S. Y. Kim, D. C. Marshall, S. H. Marshall, W. G. McCallum, and A. R. Perlis, "Jacobians of genus one curves", *J. Number Theory* **90**:2 (2001), 304–315. MR

[Aronhold 1858] S. Aronhold, "Theorie der homogenen Functionen dritten Grades von drei Veränderlichen", *J. Reine Angew. Math.* **55** (1858), 97–191. MR

[Artin et al. 2005] M. Artin, F. Rodriguez-Villegas, and J. Tate, "On the Jacobians of plane cubics", *Adv. Math.* **198**:1 (2005), 366–382. MR Zbl

[Bhargava 2008] M. Bhargava, "Higher composition laws, IV: The parametrization of quintic rings", *Ann. of Math.* (2) **167**:1 (2008), 53–94. MR Zbl

[Graf v. Bothmer and Hulek 2004] H.-C. Graf v. Bothmer and K. Hulek, "Geometric syzygies of elliptic normal curves and their secant varieties", *Manuscripta Math.* **113**:1 (2004), 35–68. MR Zbl

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR

[Buchsbaum and Eisenbud 1977] D. A. Buchsbaum and D. Eisenbud, "Algebra structures for finite free resolutions, and some structure theorems for ideals of codimension 3", *Amer. J. Math.* **99**:3 (1977), 447–485. MR Zbl

[Buchsbaum and Eisenbud 1982] D. A. Buchsbaum and D. Eisenbud, "Gorenstein ideals of height 3", pp. 30–48 in *Seminar D. Eisenbud/B. Singh/W. Vogel*, vol. 2, Teubner-Texte zur Math. **48**, Teubner, Leipzig, Germany, 1982. MR Zbl

[Eisenbud 1995] D. Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. MR Zbl

[Fisher 2008] T. Fisher, "The invariants of a genus one curve", *Proc. Lond. Math. Soc.* (3) **97**:3 (2008), 753–782. MR Zbl

[Fisher 2010] T. Fisher, "Pfaffian presentations of elliptic normal curves", *Trans. Amer. Math. Soc.* **362**:5 (2010), 2525–2540. MR Zbl

[Fisher 2013a] T. Fisher, "Explicit 5-descent on elliptic curves", pp. 395–411 in *ANTS X — Proceedings of the Tenth Algorithmic Number Theory Symposium*, edited by E. W. Howe and K. S. Kedlaya, Open Book Ser. **1**, Mathematical Sciences Publishers, Berkeley, CA, 2013. MR Zbl

[Fisher 2013b] T. Fisher, "Invariant theory for the elliptic normal quintic, I: Twists of $X(5)$", *Math. Ann.* **356**:2 (2013), 589–616. MR Zbl

[Fisher and Sadek 2016] T. Fisher and M. Sadek, "On genus one curves of degree 5 with square-free discriminant", *J. Ramanujan Math. Soc.* **31**:4 (2016), 359–383. MR

[Gross 2011] B. H. Gross, "On Bhargava's representation and Vinberg's invariant theory", pp. 317–321 in *Frontiers of mathematical sciences*, edited by B. Gu and S.-T. Yau, Int. Press, Somerville, MA, 2011. MR

[Gross and Popescu 1998] M. Gross and S. Popescu, "Equations of $(1, d)$-polarized abelian surfaces", *Math. Ann.* **310**:2 (1998), 333–377. MR Zbl

[Hulek 1986] K. Hulek, "Projective geometry of elliptic curves", pp. 143 pp. Astérisque **137**, Société Mathématique de France, Paris, 1986. MR Zbl

[Lange 1984] H. Lange, "Higher secant varieties of curves and the theorem of Nagata on ruled surfaces", *Manuscripta Math.* **47**:1-3 (1984), 263–269. MR Zbl

[Peeva 2011] I. Peeva, *Graded syzygies*, Algebra and Applications **14**, Springer, London, 2011. MR Zbl

[Silverman 2009] J. H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics **106**, Springer, Dordrecht, The Netherlands, 2009. MR Zbl

[Weil 1954] A. Weil, "Remarques sur un mémoire d'Hermite", *Arch. Math.* (*Basel*) **5** (1954), 197–202. MR Zbl

[Weil 1983] A. Weil, "Euler and the Jacobians of elliptic curves", pp. 353–359 in *Arithmetic and geometry*, vol. 35, edited by M. Artin and J. Tate, Birkhäuser, Boston, MA, 1983. MR Zbl

T.A.Fisher@dpmms.cam.ac.uk   *Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdom*

msp

Random flag complexes and
asymptotic syzygies

Daniel Erman and Jay Yang

We use the probabilistic method to construct examples of conjectured phenomena about asymptotic syzygies. In particular, we use Stanley–Reisner ideals of random flag complexes to construct new examples of Ein and Lazarsfeld's nonvanishing for asymptotic syzygies and of Ein, Erman, and Lazarsfeld's conjecture on how asymptotic Betti numbers behave like binomial coefficients.

Using the probabilistic method, we produce examples of conjectured behavior on asymptotic syzygies. One of these provides the first known example of a phenomenon conjectured by Ein, Erman, and Lazarsfeld.

Our central construction involves random flag complexes. We use $G \sim G(n, p)$ to denote an Erdős–Rényi random graph on n vertices, where each edge is attached with probability p. We turn G into a flag complex by adjoining a k-simplex to every $(k + 1)$-clique in the graph, and $\Delta \sim \Delta(n, p)$ denotes a flag complex chosen with respect to this distribution. The properties of random flag complexes have been studied extensively in recent years; see [Kahle 2014b] for a survey of recent results. From Δ, Stanley–Reisner theory yields a squarefree monomial ideal $I_\Delta \subseteq k[x_1, x_2, \ldots, x_n]$ [Bruns and Herzog 1993, Chapter 5], and we analyze the Betti numbers of I_Δ.

A recent paper by De Loera, Petrović, Silverstein, Stasi, and Wilburne [Loera et al. 2017] also produces random monomial ideals via a construction similar to Erdős–Rényi random graphs, and one of their constructions specializes to ours. They study thresholds and the distribution of algebraic invariants in this framework, and they provide an array of results and conjectures.

We are motivated by questions and conjectures about asymptotic syzygies. These questions are generally outside of the range computable in Macaulay2 or elsewhere, and so there is a lack of known examples. By contrast, results on random flag complexes are asymptotic in nature. By using probabilistic techniques to analyze the syzygies of I_Δ, we produce new examples of behaviors conjectured in [Ein and Lazarsfeld 2012; Ein et al. 2015].

We now summarize Ein and Lazarsfeld's central result on asymptotic syzygies. For a graded module M over a polynomial ring, we recall that $\beta_{i,j}(M)$ denotes the number of minimal generators of degree j of the i-th syzygy module of M; see [Eisenbud 2005, §1B] for a review. We define $\rho_k(M)$ as the ratio of

.

Figure 1. Each dot represents a known nonzero entry in the Betti table of $\mathbb{P}^3$ embedded by $\mathcal{O}(n)$ for $n = 10$. By Ein and Lazarsfeld's Theorem 1.1, the density of the dots in rows 1, 2, and 3 will approach 1 as $n \to \infty$. Theorem 1.3 shows a similar phenomenon holds for ideals of random flag complexes.

nonzero entries in the k-th row of the Betti table:

$$\rho_k(M) := \frac{\#\{i \in [0, \mathrm{pdim}(M)] \text{ where } \beta_{i,i+k}(M) \neq 0\}}{\mathrm{pdim}(M) + 1}.$$

Under increasingly positive embeddings, [Ein and Lazarsfeld 2012] shows that these densities approach 1.

Theorem 1.1 (Ein and Lazarsfeld 2012). *Let X be a smooth, d-dimensional projective variety and let A be a very ample divisor on X. For any $n \geq 1$, let S_n be the homogeneous coordinate ring of X embedded by nA. For each $1 \leq k \leq d$, $\rho_k(S_n) \to 1$ as $n \to \infty$.*

See [Ein and Lazarsfeld 2012, Theorem A] for the sharper result and Figure 1 for an illustration. A similar nonvanishing phenomenon was shown to hold for integral varieties [Zhou 2014, Theorem, p. 2256], arithmetically Cohen–Macaulay varieties [Ein et al. 2016, Theorem 3.1], and certain iterated subdivisions of Stanley–Reisner rings [Conca et al. 2018]. Moreover, experiments in Macaulay2 with different asymptotic families of ideals (graph curves, unions of linear spaces, etc.) suggest that this asymptotic nonvanishing behavior occurs in a broad range of examples. This motivates the following question:

Question 1.2. Let $\{I_n\}$ be a family of ideals where $\mathrm{pdim}(I_n) \to \infty$. Fix some k. Under what conditions will $\rho_k(S/I_n) \to 1$ as $n \to \infty$?

One way to understand these asymptotic nonvanishing results is by considering the overlaps between the nonzero entries in the rows of the Betti table. The Hilbert function of a graded module will determine the alternating sum of the entries along the slope one diagonals of the Betti table. We define *overlapping Betti numbers* as Betti numbers that are not determined by the Hilbert function: e.g., when $\beta_{i,j}$ and $\beta_{i+1,j}$ are both nonzero. Theorem 1.1 and the related followup results show that such overlapping Betti numbers are the norm in many different families of examples.

While Question 1.2 addresses qualitative expectations about asymptotic syzygies, the corresponding quantitative behavior of asymptotic syzygies was raised in [Ein et al. 2015]. They introduce a random Betti table model to provide a heuristic for the asymptotic behavior of certain families of Betti tables. Their analysis suggests that, roughly speaking, each row of the Betti table of any very positive embedding displays the pattern of a large Koszul complex [Ein et al. 2015, Conjecture B and Theorem C]. Yet despite

the expectation that this behavior should be common, the only known occurrence is for a smooth curve of high degree [Ein et al. 2015, Proposition A].

Our main results provide new families whose Betti tables exhibit the conjectured behaviors described above. We write $f(n) \ll g(n)$ if $\lim_{n \to \infty} f(n)/g(n) = 0$.

Theorem 1.3. *Fix some $r \geq 1$. Let $\Delta \sim \Delta(n, p)$ with $1/n^{1/r} \ll p \ll 1$. For each $1 \leq k \leq r + 1$, we have $\rho_k(S/I_\Delta) \to 1$ in probability.*

Saying that $\rho_k(S/I_\Delta) \to 1$ in probability is equivalent to asking that for any $\epsilon > 0$, the probability that $\rho_k(S/I_\Delta) \geq 1 - \epsilon$ goes to 1 as $n \to \infty$. In particular, for the given parameter range, random flag complexes in the $\Delta(n, p)$ model provide a positive answer to Question 1.2, similar to Theorem 1.1. See Example 5.1.

The proof of Theorem 1.3 uses randomness to find particular subcomplexes of Δ. As we will review in Section 2, the boundary complex of the $(s + 1)$-dimensional octahedron has the minimal number of edges possible for a flag complex with $(s + 1)$-th homology, and it is thus the most likely subcomplex to contribute to the $(s + 1)$-th row of the Betti table of S/I_Δ. The main step of the proof comes from Theorem 1.6 below, where we show that the bound $1/n^{1/s} \ll p$ is the threshold for the existence of this particular subcomplex. Once we have crossed this threshold, we can find this particular subcomplex, and minor variants of it, yielding nonzero Betti numbers throughout nearly the entire $(s + 1)$-th row.

Next we construct examples whose Betti tables exhibit the more detailed asymptotics suggested in [Ein et al. 2015]. For any I_Δ, the Hilbert function of S/I_Δ will have the form $(1, n, \dots)$, and thus as $n \to \infty$, the Betti table will necessarily scale with n. To account for this growth, we normalize the Betti table, defining $\overline{\beta}(S/I_\Delta) := (1/n)\beta(S/I_\Delta)$.[1]

Theorem 1.4. *Fix a constant $0 < c < 1$ and let $\Delta \sim \Delta(n, c/n)$ be a random flag complex. If $\{i_n\}$ is an integer sequence satisfying $i_n = n/2 + o(n)$, and if $C := (1 - c)/2$, then*

$$\frac{\overline{\beta}_{i_n, i_n + 1}(S/I_\Delta)}{C\binom{n}{i_n}} \to 1$$

in probability.

Theorem 1.4 is a local limit theorem, in the sense that it is a pointwise convergence rather than a global result about the whole distribution. Moreover, the theorem is entirely focused on Betti numbers near the middle of the first row. Yet, by a standard change of variables, this suffices to provide an example of the behavior predicted by [Ein et al. 2015, Conjecture B].

Corollary 1.5. *Fix a constant $0 < c < 1$ and let $\Delta \sim \Delta(n, c/n)$ be a random flag complex. If $\{i_n\}$ is a sequence of integers converging to $n/2 + a\sqrt{n}/2$, then*

$$\frac{\sqrt{2\pi}}{(1 - c)2^n \sqrt{n}} \cdot \beta_{i_n, i_n + 1}(S/I_\Delta) \to e^{-a^2/2}$$

in probability.

---

[1]For a similar reason, [Ein et al. 2015, Conjecture B] also allows for a rescaling function.
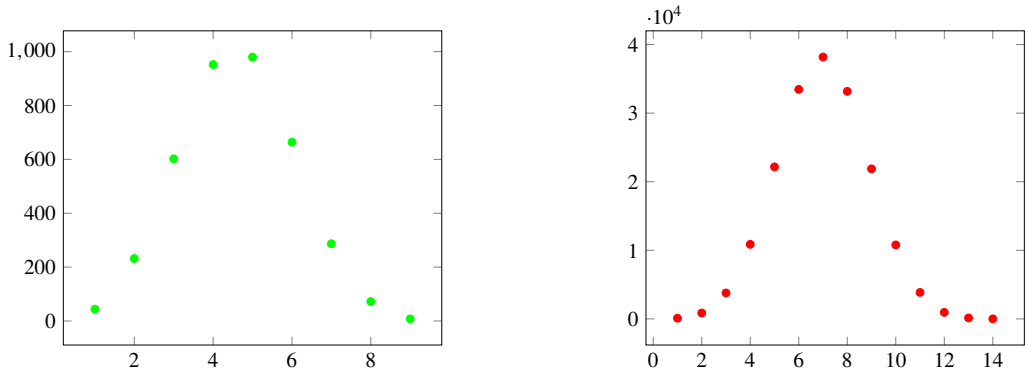
Figure 2. We plot the function $i \mapsto \beta_{i,i+1}(S/I_\Delta)$ for a random $\Delta \sim \Delta\left(10, \frac{1}{20}\right)$ and $\Delta \sim \Delta\left(15, \frac{1}{30}\right)$, respectively. These appear consistent with the appearance of binomial coefficients, as in the heuristic of [Ein et al. 2015] and in Theorem 1.4.

See Figure 2 for a couple of examples.

The only previously known example of this kind comes from smooth curves [Ein et al. 2015, Theorem A]. However, that example avoids the complexity of overlapping Betti numbers. By contrast, for the family of ideals in Theorem 1.4, the Betti numbers are not always clustered in a single row (see Remark 6.1). Thus, Theorem 1.4 produces the first known families of ideals which exhibit overlapping Betti numbers and behave like [Ein et al. 2015, Conjecture B].

The following simple computation suggests why the Betti numbers of random flag complexes should behave like rescaled binomial coefficients. For a subset α of the vertices, we write $\Delta|_\alpha$ for the restricted flag complex. Hochster's formula [Bruns and Herzog 1993, Theorem 5.5.1] shows that $\beta_{i,i+1}(S/I_\Delta)$ is the sum over all $\alpha \in \binom{[n]}{i}$ of $\dim \widetilde{H}_0(\Delta|_\alpha)$. By linearity of expectations, the expected value of $\beta_{i,i+1}(S/I_\Delta)$ is

$$E[\beta_{i,i+1}(S/I_\Delta)] = \sum_{\alpha \in \binom{[n]}{i}} \dim \widetilde{H}_0(\Delta|_\alpha) = \binom{n}{i} E[\widetilde{H}_0(\Delta')],$$

where $\Delta' \sim \Delta(i, c/n)$ is a random flag complex. So it suffices to control how the expectation $E[\widetilde{H}_0(\Delta')]$ varies with i. The main issue in proving Theorem 1.4 thus arises in showing convergence in probability, stemming from the fact that $\beta_{i,i+1}(S/I_\Delta)$ is a sum of dependent random variables.

Not coincidentally, the choice $p = c/n$ (as in Theorem 1.4) is a much-studied regime in the random graph literature. See [Alon and Spencer 2016, §11; Frieze and Karoński 2016, §2.1], among other references. We rely on some of those structural results about random graphs in this regime for our proofs of Proposition 6.2 and Theorem 1.4.

We also prove some results on the algebraic invariants of S/I_Δ. For instance, we prove the following threshold result for individual Betti numbers:

Theorem 1.6 (Betti number thresholds). *Fix i, v with $1 \leq i$ and $i + 1 \leq v \leq 2i$ and let $s := v - i - 1$. Fix some constant $0 < \epsilon \leq \frac{1}{2}$ and let $\Delta \sim \Delta(n, p)$:*

(1) *If $1/n^{1/s} \ll p \leq \epsilon$ then $\mathbf{P}[\beta_{i,v}(S/I_\Delta) \neq 0] \to 1$.*

(2) *If $p \ll 1/n^{1/s}$ then $\mathbf{P}[\beta_{i,v}(S/I_\Delta) = 0] \to 1$.*

We use this to bound the Castelnuovo–Mumford regularity of S/I_Δ in Corollary 5.2. Corollary 7.1 also shows that while S/I_Δ is almost never Cohen–Macaulay, the depth and codimension of S/I_Δ converge as $n \to \infty$.

This paper is organized as follows. Section 2 provides some essential definitions. Section 4 provides a threshold for the vanishing/nonvanishing of individual Betti numbers, the nonvanishing half of which relies on a variance bound proven Section 3. In Section 5 we use the Betti number threshold to prove Theorem 1.3. In Section 6 we prove Theorem 1.4 and Corollary 1.5. Section 7 contains estimates on the projective dimension of the ideal I_Δ.

2. Background and notation

We work over an arbitrary field k. We write $\mathbf{P}[-]$ for the probability of an event and $\mathbf{E}[-]$ for the expected value of a random variable.

A flag complex is a simplicial complex obtained from a graph by adjoining a k-simplex to every $(k+1)$-clique in the graph. We use $G \sim G(n, p)$ to denote an Erdős–Rényi random graph on n vertices, where each edge is attached with probability p, and we use $\Delta \sim \Delta(n, p)$ to denote the corresponding random flag complex. If H is a subset of the n vertices, then we use $\Delta|_H$ for the induced flag complex.

The generators of I_Δ correspond to the maximal nonfaces of Δ [Bruns and Herzog 1993, Chapter 5], and since Δ is flag this means that I_Δ is generated by quadrics. Hochster's formula (Theorem 5.5.1 in the same reference), which relates the Betti table of S/I_Δ to topological properties of Δ, is our key tool for studying the syzygies of S/I_Δ.

Remark 2.1. As discussed in the introduction, our goal is to use the I_Δ to model asymptotic syzygies. The ideals of high degree Veroneses always admit a quadratic Gröbner basis [Eisenbud et al. 1994], and this is one reason why we chose to use random flag complexes. By contrast, models in [Loera et al. 2017] often produce ideals with generators in different degrees, and those would thus provide better models for other families of examples.

Example 2.2. Hochster's formula implies that $\beta_{r+1,2r+2}(S/I_\Delta)$ is the number of subcomplexes $\Delta|_H \subseteq \Delta$, where H has $2r+2$ vertices and where $\widetilde{H}_r(\Delta|_H) \neq 0$. For instance $\beta_{1,2}(S/I_\Delta)$ is the number of pairs of disjoint vertices in Δ, or equivalently it is the number of nonedges of the Δ, and $\beta_{2,4}(S/I_\Delta)$ is the number of squares in Δ. On the other hand, $\beta_{2,5}(S/I_\Delta)$ counts subcomplexes on five vertices with nonzero $\widetilde{H}_1$. There are several different types of examples, such as

Lemma 2.3. *If Δ is a flag complex, then $\beta_{i,j}(S/I_\Delta) = 0$ for all $j > 2i$.*

Proof. Since Δ is flag, I_Δ is a monomial ideal generated by quadrics. The Taylor resolution of S/I_Δ thus involves monomials of degree 0, 1, or 2 [Peeva 2011, Construction 26.5]. $\qquad\square$

The boundary complex of the $(r + 1)$-dimensional octahedron plays a key role in our results (for instance, see Remark 3.1), and we denote this flag complex by $\Diamond_r$. We note that $\Diamond_r$ is also the r-fold suspension of 2 points. See Figure 3. Since a pair of points is disconnected, we have $\widetilde{H}_0(\Diamond_0) \cong \mathbb{Z}$, and since taking suspensions shifts reduced homology groups up by one degree, we have that $\widetilde{H}_r(\Diamond_r) \cong \mathbb{Z}$. We now observe that any flag complex with nonzero r-th homology will have at least as many vertices and edges as $\Diamond_r$.

Lemma 2.4. *Let Δ be a flag complex with $\widetilde{H}_r(\Delta) \neq 0$:*

(1) *Then Δ has at least $2r + 2$ vertices.*

(2) *If $v \in \Delta$ is a vertex such that $\widetilde{H}_r(\Delta_{\Delta - v}) = 0$, then $\deg(v) \geq 2r$.*

(3) *Δ has at least $2r(r + 1)$ edges.*

Proof. This result is folklore. Part (1) is proven in [Conca et al. 2018, Lemma 3.6]. Parts (2) and (3) follow easily by standard topological arguments. $\qquad\square$

Remark 2.5. The complex $\Diamond_r$ shows that the bounds in Lemma 2.4 are sharp.

3. Variance bound

In this section we prove a variance bound that is used in our convergence results. The proof is similar to those in [Bollobás and Erdős 1976, Theorem 1; Kahle 2014a, Lemma 2.2].

Remark 3.1. We are particularly interested in the appearance of subcomplexes of the form $\Diamond_s$, as by Lemma 2.4 these are the flag complexes with the fewest edges and nonzero s-th homology. Since in our models p goes to 0 as $n \to \infty$, subcomplexes with fewer edges are more likely to appear, and so we expect these $\Diamond_s$ to control the $(s + 1)$-th row of $\beta(S/I_\Delta)$.

Remark 3.2. In $\Diamond_s$, every vertex has a unique antipodal vertex, and thus as a subgraph of Δ, $\Diamond_s$ is determined by $s + 1$ pairs of vertices, all distinct. In particular, given a set of vertices $V \in \binom{[n]}{2(s+1)}$, there
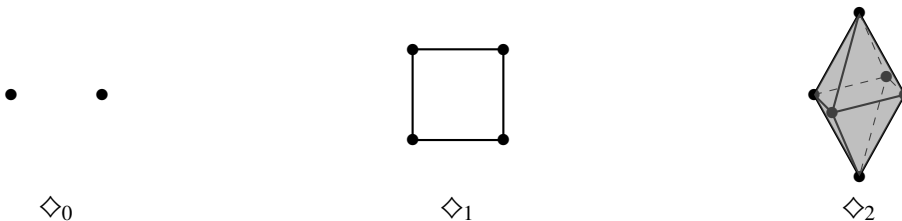
Figure 3. Among flag complexes with nonzero r-th homology, the boundary complex of the $(r + 1)$-dimensional octahedron, which we denote $\Diamond_r$, has the fewest edges.

are multiple ways that $\Delta|_V$ could be an $\diamondsuit_s$-subcomplex; to simplify the computations in this section, it will be useful to parametrize each potential $\diamondsuit_s$ separately, even those that involve the same vertices. We define Λ_s as vertex sets $V \in \binom{[n]}{2(s+1)}$ of size $2(s+1)$ together with an unordered decomposition $V = P_0 \cup \cdots \cup P_s$, where each P_i is an unordered pair of vertices. With this definition, there is then a bijection between elements of Λ_s and potential subcomplexes $\diamondsuit_s \subseteq \Delta$. Thus, given any $H \in \Lambda_s$, the probability that $\Delta|_H$ is $\diamondsuit_s$ is given precisely by the probability that $\Delta|_H$ has exactly the specified edges, which is $p^{2s(s+1)}(1-p)^{\binom{2(s+1)}{2}-2s(s+1)}$.

Definition 3.3. Let $X_s = X_s(n, p)$ denote the random variable for the number of copies of $\diamondsuit_s$ appearing as a subgraph of a random graph $G \sim G(n, p)$. Given $H \in \Lambda_s$ we then define X_H as the indicator random variable for whether the subgraph on H has the form $\diamondsuit_s$.

Thus we have $X_s = \sum_{H \in \Lambda_s} X_H$. We will now use this to bound the variance $\mathrm{Var}[X_s]$.

Lemma 3.4 (variance bound). *If $np^{(s+1/2)} \to \infty$ and $p \leq (1-p)$, then $\mathrm{Var}[X_s]/E[X_s]^2 \to 0$.*

Proof. We start by computing

$$E[X_s^2] = \sum_{H,J \in \Lambda_s} E[X_H X_J] = \sum_{H,J \in \Lambda_s} P[X_J = 1 \mid X_H = 1]P[X_H = 1]$$

$$= \sum_{H \in \Lambda_s} P[X_H = 1] \sum_{J \in \Lambda_s} P[X_J = 1 \mid X_H = 1].$$

Since $\sum_{J \in \Lambda_s} P[X_J = 1 \mid X_H = 1]$ is independent of the choice of H, we may fix an H' to decouple the factors, yielding

$$= \left(\sum_{H \in \Lambda_s} P[X_H = 1] \right) \sum_{J \in \Lambda_s} P[X_J = 1 \mid X_{H'} = 1] = E[X_s]E[X_s \mid X_{H'} = 1].$$

Since $\mathrm{Var}[X_s] = E[X_s^2] - E[X_s]^2$, the above computation allows us to compute

$$\mathrm{Var}(X_s)/E[X_s]^2 = \frac{E[X_s \mid X_H = 1] - E[X_s]}{E[X_s]} = \frac{\sum_{m=0}^{2s+2} \sum_{|J \cap H|=m} P[X_J = 1 \mid X_H = 1] - P[X_J = 1]}{E[X_s]}.$$

If J and H are disjoint or intersect in only a single vertex, then $P[X_J = 1 \mid X_H = 1] = P[X_J = 1]$. We can thus ignore the terms with $m = 0$ or $m = 1$ in this sum:

$$= \frac{\sum_{m=2}^{2s+2} \sum_{|J \cap H|=m} P[X_J = 1 \mid X_H = 1] - P[X_J = 1]}{E[X_s]}.$$

By Lemma 3.5, we obtain the bound

$$\leq \frac{\sum_{m=2}^{2s+2} \sum_{|J \cap H|=m} p^{-m(m-1)/2} P[X_J = 1] - P[X_J = 1]}{E[X_s]}.$$

Since the probability $P[X_J = 1]$ does not depend on J, we can use the bound from Lemma 3.6 to pull $P[X_J = 1]/E[X_s]$ outside, and simplify the expression, where C is a constant:

$$\leq Cn^{-2(s+1)} \sum_{m=2}^{2s+2} \sum_{|J \cap H|=m} p^{-m(m-1)/2} - 1.$$

Up to a constant, for a fixed H there are $n^{2(s+1)-m}$ choices of J where $|J \cap H| = m$. Absorbing those constants into our C we get

$$\leq Cn^{-2(s+1)} \sum_{m=2}^{2s+2} n^{2(s+1)-m}(p^{-m(m-1)/2} - 1) = C \sum_{m=2}^{2s+2} n^{-m}(p^{-m(m-1)/2} - 1) \leq C \sum_{m=2}^{2s+2} (np^{(m-1)/2})^{-m}.$$

Since $0 < (m-1)/2 \leq s + \frac{1}{2}$ we have $np^{(m-1)/2} \to \infty$ by hypothesis. It follows that all of the finitely many terms in the sum go to 0, and thus $\mathrm{Var}(X_s)/E[X_s]^2 \to 0$. $\square$

Lemma 3.5. *Given* $J, H \in \Lambda_s$ *such that* $|J \cap H| = m$,

$$P[X_J = 1 \mid X_H = 1] \leq p^{-m(m-1)/2} P[X_J = 1].$$

Proof. If $X_H = 1$ then the edges in $J \cap H$ are completely determined. If those edges do not match the required edges for J, then $P[X_J = 1 \mid X_H = 1] = 0$. If they do match the required edges, then since the probability of any edge existing or not existing is p or $1 - p$, and since $p \leq 1 - p$, we get that $P[X_J = 1 \mid X_H = 1] \leq p^{-m(m-1)/2} P[X_J = 1]$. $\square$

Lemma 3.6. *For any fixed* $H \in \Lambda_s$, *we have* $P[X_H = 1]/E[X_s] \leq Cn^{-2(s+1)}$ *for some constant C.*

Proof. Since $X_s = \sum_H X_H$ we have $E[X_s] = \sum_H P[X_H = 1]$. But since $P[X_H = 1]$ does not depend on H, this amounts to counting the number of possible choices of H, which is the cardinality of Λ_s. Each element of Λ_s corresponds to $s + 1$ pairs of vertices in Δ, of which there $(1/(s+1)!)\binom{n}{2,2,2,...,n-2(s+1)}$ choices. It follows that, for an appropriate constant C, we have $P[X_H = 1]/E[X_s] \leq Cn^{-2(s+1)}$. $\square$

4. Betti number thresholds

In this section, we determine thresholds of nonvanishing for individual Betti numbers. Lemma 2.3 shows that $\beta_{i,v}(S/I_\Delta) = 0$ whenever $v \leq i$ or $v \geq 2i$, and Theorem 1.6 computes thresholds in the remaining cases. To prove that theorem, we first bound the expected values of the Betti numbers. For $\Delta \sim \Delta(n, p)$ we define $B_{i,v}$ where $B_{i,v}(\Delta) := \beta_{i,v}(S/I_\Delta)$. By convention, when $s = 0$ we interpret $1/n^{1/s} \ll p$ as a trivial bound.

Lemma 4.1. *Fix any constant* $0 < \epsilon < 1$. *Let* $1/n^{1/s} \ll p \leq \epsilon$ *and* $\Delta \sim \Delta(n, p)$. *We have* $E[B_{s+1,2s+2}] \to \infty$ *as* $n \to \infty$.

Proof. By Hochster's formula [Bruns and Herzog 1993, Theorem 5.5.1], since $\widetilde{H}_s(\diamondsuit_s) \neq 0$, we have $E[B_{s+1,2s+2}] \geq \sum_H E[X_H]$, where as in Definition 3.3, H is a set of $s + 1$ pairs of vertices, all distinct.

Since any $\diamondsuit_s$ involves $s(2s+2)$ edges and $s+1$ nonedges, we have

$$E[X_H] = P[X_H = 1] = p^{s(2s+2)}(1-p)^{s+1}.$$

As in the proof of Lemma 3.6, the number of choices for H is at least Cn^{2s+2} for some positive constant C, and thus

$$E[B_{s+1,2s+2}] = \sum_H E[X_H] \geq Cn^{2s+2}p^{s(2s+2)}(1-p)^{s+1} \geq C'(np^s)^{2s+2},$$

where $C' = C(1-\epsilon)^{s+1}$. Since $np^s \to \infty$ it follows that $E[B_{s+1,2s+2}] \to \infty$. $\qquad\square$

To prove the other threshold, we introduce new random variables.

Definition 4.2. Let $Y_v^s = Y_v^s(n, p)$ be the number of subgraphs with $m \leq v$ vertices and at least ms edges. If K is a subset of m vertices, we let Y_K^s be the indicator random variable for whether the subgraph on K has at least ms edges.

Lemma 4.3. *If $p \ll 1/n^{1/s}$ then $E[B_{i,v}] \to 0$.*

Proof. Lemma 2.4 shows that if K is a minimal subset of vertices of Δ such that $\widetilde{H}_s(\Delta|_K) \neq 0$, then each vertex in $\Delta|_K$ has degree $\geq 2s$. In particular, if $\beta_{i,v}(S/I_\Delta) \neq 0$, then there must exist some subgraph K of size at most v (and with at least $2s+2$ vertices) where every vertex has degree $\geq 2s$. It thus suffices to prove that $E[Y_v^s] \to 0$.

We have $Y_v^s = \sum_{K,|K| \leq v} Y_K^s$. For a fixed K with $|K| = m$, we want to compute the probability that $\Delta|_K$ has at least ms edges. We use $M := \binom{m}{2}$ to denote the maximal number of possible edges. We thus have

$$P[Y_K^s = 1] = \sum_{e=ms}^M \binom{M}{e} p^e (1-p)^{M-e}.$$

We then compute

$$E[Y_v^s] = \sum_{m=2s+2}^v \sum_{K,|K|=m} P[Y_K^s = 1] = \sum_{m=2s+2}^v \binom{n}{m} \sum_{e=ms}^M \binom{M}{e} p^e (1-p)^{M-e}$$

$$\leq \sum_{m=2s+2}^v \binom{n}{m} \sum_{e=ms}^M \binom{M}{e} p^e \leq \sum_{m=2s+2}^v \binom{n}{m} p^{ms} \sum_{e=ms}^M \binom{M}{e} p^{e-ms}.$$

However, we can bound $\sum_{e=ms}^M \binom{M}{e} p^{e-ms}$ by a constant $C_{s,m}$ depending only on s and m, and we can bound $\binom{n}{m}$ by n^m. This yields

$$\leq \sum_{m=2s}^v n^m p^{ms} C_{s,m} = \sum_{m=2s}^v (np^s)^m C_{s,m}.$$

Finally, since $np^s \to 0$ by assumption, we conclude that $E[Y_v^s] \to 0$. $\qquad\square$

Proof of Theorem 1.6. For statement (1), we first consider the case where $v = 2i = 2s + 2$. Lemma 4.1 implies that $E[\mathrm{B}_{s+1,2s+2}] \to \infty$. Thus to prove that $P[\mathrm{B}_{s+1,2s+2} \neq 0] \to 1$, we may bound the variance of $\mathrm{B}_{s+1,2s+2}$. This is done in Lemma 3.4 since $\mathrm{B}_{s+1,2s+2} = \mathrm{X}_s$. There we show that

$$\frac{\mathrm{Var}[\mathrm{B}_{s+1,2s+2}]}{E[\mathrm{B}_{s+1,2s+2}]^2} \to 0.$$

Thus we can apply Chebyshev's inequality to say the following:

$$P[\mathrm{B}_{s+1,2s+2} = 0] \leq P[|E[\mathrm{B}_{s+1,2s+2}] - \mathrm{B}_{s+1,2s+2}| \geq E[\mathrm{B}_{s+1,2s+2}]] \leq \frac{\mathrm{Var}[B_{s+1,2s+2}]}{(E[\mathrm{B}_{s+1,2s+2}] - 1)^2} \to 0.$$

We now let $v < 2i$. The case $v = 2s + 2$ implies the existence of some $\Diamond_s \subseteq \Delta$ with probability $1 - o(1)$. Fix some vertex $u \in \Diamond_s$. Let J be the set of vertices $w \in \Delta$ which don't lie in $\Diamond_s$ and which are not connected with u. Since the complement of $\Diamond_s$ consists of $n - (2s + 2)$ vertices, the expected number of vertices in J is $(n - (2s + 2))(1 - p) = n - o(n)$. Moreover, since those conditions are independent, the weak law of large numbers implies that this happens with high probability. Let $J' \subseteq J$ be any subset of cardinality $v - (2s + 2)$. Since the only edges in $\Diamond_s \cup J'$ through the vertex u are the ones from $\Diamond_s$, it follows $\widetilde{\mathrm{H}}_s(\Diamond_s \cup J')$ is still nonzero. Hence $\mathrm{B}_{i,v} \neq 0$ with high probability as desired.

For (2), we must show that $\mathrm{B}_{i,v}$ converges to 0 in probability. Hochster's formula [Bruns and Herzog 1993, Theorem 5.5.1] implies that $\beta_{i,v}(S/I_\Delta)$ is nonzero if and only there is some subset $K \subseteq \Delta$ with $|K| = v$ and where $\widetilde{\mathrm{H}}_{v-i-1}(\Delta|_K) \neq 0$. By Lemma 2.4 it suffices to show that $P[\mathrm{Y}_v^s = 0] \to 1$ for $s = v - i - 1$. But by Lemma 4.3, we know $E[\mathrm{Y}_v^s] \to 0$, and since $\mathrm{Y}_v^s \geq 0$ and Y_v^s takes integer values, this implies that $P[\mathrm{Y}_v^s = 0] \to 1$. $\qquad\square$

5. Ein–Lazarsfeld asymptotic nonvanishing of syzygies

Whereas Theorem 1.6 provides the nonvanishing thresholds for individual Betti numbers, Question 1.2 asks about the simultaneous nonvanishing of more and more Betti numbers as $n \to \infty$. However, as we now illustrate, the proof of Theorem 1.6 is sufficiently strong to obtain simultaneous nonvanishing of the various Betti numbers.

Proof of Theorem 1.3. For each n, we partition the vertices into $r + 1$ sets $S_0, S_1, \ldots, S_r$ each of size approximately $n/(r + 1)$. Since $\Delta|_{S_s}$ is a random flag complex for any $0 \leq s \leq r$, the proof of Theorem 1.6 implies the existence of some $\Diamond_s$ in $\Delta|_{S_s}$ with probability $1 - o(1)$. Moreover, since r is fixed, we can assume that these all occur simultaneously. By construction, the vertices involved in $\Diamond_0, \Diamond_1, \ldots, \Diamond_r$ are all disjoint.

Fix some $0 < \epsilon < 1$. For each $0 \leq s \leq r$, fix some vertex $v \in \Diamond_s$. Since the complement of $\bigcup_{s=0}^{r} \Diamond_s$ consists of $n - O(1)$ vertices, the expected number of vertices $w \notin \bigcup_{s=0}^{r} \Diamond_s$ that are not connected with vertex v is $(n - O(1))(1 - p) \geq n - n^{1-\epsilon}$, at least for n sufficiently large. Since those conditions are independent, the weak law of large numbers implies that this happens with high probability. Call that set J and $J' \subseteq J$ be any subset. Since the only edges in $\Diamond_s \cup J'$ through the vertex v are the ones from

$\diamondsuit_s$, it follows $\widetilde{H}_s(\Delta|_{\diamondsuit_s \cup J'})$ is still nonzero. Since $|\diamondsuit_s \cup J'|$ ranges from $2s + 2$ to $n - n^{1-\epsilon} + 2s + 2$, it follows that $\beta_{i+1,i+s+2}(S/I_\Delta) \neq 0$ for all $s \leq i \leq n - n^{1-\epsilon} + s$ with high probability. In particular, with high probability we have

$$\lim_{n \to \infty} \rho_{s+1}(S/I_\Delta) \geq \lim_{n \to \infty} \frac{n - n^{1-\epsilon} + 1}{n} = 1.$$

Moreover, since the $\diamondsuit_s$ involve disjoint vertices, these nonvanishing conditions are independent in s, and we thus obtain the desired convergence of ρ_{s+1} for all s simultaneously. $\qquad\square$

The proof of Theorem 1.3 shows that if we cross the threshold for the appearance of subcomplexes of the form $\diamondsuit_s$, then we get nonvanishing across nearly the entire $(s + 1)$-th row of the Betti table. The appearance of $\diamondsuit_s$ subcomplexes thus accounts for why $\rho_{s+1}(S/I_\Delta)$ goes to 1.

Example 5.1. Here is the Betti table of S/I_Δ for a randomly chosen $\Delta \sim \Delta(18, 1/18^{0.6})$, as computed in Macaulay2:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
1	•	126	1203	5986	19491	45278	78385	103667	106356	85548	54408	27541	11118	3550	873	156	18	1
2	•	•	1	24	233	1282	4568	11261	19911	25743	24538	17229	8815	3204	786	117	8	•

As predicted by Theorem 1.3, the entries in rows 1 and 2 are almost all nonzero.

Though we do not compute a precise threshold for the Castelnuovo–Mumford regularity of S/I_Δ, we do obtain a linear bound.

Corollary 5.2. If $1/n^{1/r} \ll p \ll 1/n^{2/(2r+1)}$, then with high probability $r + 1 \leq \mathrm{reg}(S/I_\Delta) \leq 2r$.

Proof. Since $1/n^{1/r} \ll p$ we have that $\beta_{r+1,2r+2}(S/I_\Delta) \neq 0$ and thus $\mathrm{reg}(S/I_\Delta) \geq r$, with high probability. For the other direction, we let $s = 2r + 1$ so that $p \ll 1/n^{2/s}$. A simple computation shows that the expected number of $(s + 1)$-cliques in Δ is

$$\binom{n}{s+1} p^{\binom{s+1}{2}} \leq n^{s+1}(p^{s/2})^{s+1} \ll n^{s+1}(n^{-1})^{s+1} = 1.$$

Since the expected number of $(s + 1)$-cliques goes to zero, it follows that with high probability Δ has no subcomplex with $(s + 1)$-th homology and thus $\mathrm{reg}(S/I_\Delta) < s = 2r + 1$. $\qquad\square$

Question 5.3. Does $\mathrm{reg}(S/I_\Delta)$ converge in probability (with appropriate conditions on p)? More precisely, if $1/n^{1/r} \ll p \ll 1/n^{1/(r+1)}$ does $\mathrm{reg}(S/I_\Delta)$ converge to $r + 1$ in probability?

6. Normal distribution of quadratic strand

In this section, we prove Theorem 1.4 and Corollary 1.5.

Remark 6.1. For Δ as in Theorem 1.4, the second row of the Betti table of S/I_Δ is interesting as well, because $p = c/n$ is a boundary case for the nonvanishing in Theorem 1.3. In [Erdős and Rényi 1960, Theorem 5b], they prove that the 1-skeleton of Δ will contain a cycle with probability $1 -$

$\sqrt{1-c}e^{(c/2)+(c^2/4)}$. Among graphs containing at least one cycle, an argument similar to the proof of
Theorem 1.3 yields $n - n^{1-\epsilon}$ nonzero entries in the second row of the Betti table of S/I_Δ, and thus in
this case, S/I_Δ will have overlapping Betti numbers throughout two rows, similar to the case of a smooth
surface in Theorem 1.1.

Given a graph G, we define

$$\boldsymbol{H}_0(G, k) = \sum_{\alpha \in \binom{[n]}{k}} \widetilde{\mathrm{H}}_0(G|_\alpha)$$

as the sum of $\widetilde{\mathrm{H}}_0(G|_\alpha)$, where $\alpha \in \binom{[n]}{k}$ is a subset of the vertices of size k and where $G|_\alpha$ is the induced
subgraph. Hochster's formula [Bruns and Herzog 1993, Theorem 5.5.1] implies that if I_Δ is a Stanley–
Reisner ideal, then the Betti number $\beta_{k,k+1}(S/I_\Delta)$ equals $\boldsymbol{H}_0(G, k)$, where G is the one-skeleton of the
simplicial complex Δ. We can thus reduce Theorem 1.4 to the following computation about graphs:

Proposition 6.2. *Let $G \sim G(n, c/n)$ be a random graph with $0 < c < 1$. If $\{i_n\}$ is an integer sequence*
satisfying $i_n = n/2 + o(n)$, and if $C := (1-c)/2$, then

$$\frac{\boldsymbol{H}_0(G, i_n)}{Cn\binom{n}{i_n}} \to 1$$

in probability.

Proof. If we remove graphs from the distribution $G \in G(n, p)$ which arise with probability $o(1)$, then this
will not affect facts about convergence in probability. For instance, with probability $1 - o(1)$ a random
$G \sim G(n, c/n)$ with $c < 1$ will be the disjoint union of trees and components with a single cycle [Frieze
and Karoński 2016, p. 31]. Thus, we may restriction attention to graphs G which are the disjoint union of
trees and components with a single cycle. Moreover, since the expected number of cycles is constant
when $c < 1$, we conclude that with probability $1 - o(1)$, Δ has at most $n^{1-\epsilon}$ cycles for any fixed $0 < \epsilon < 1$.
We thus further restrict attention to the case where Δ is the disjoint union of trees and at most $n^{1-\epsilon}$
components each with a single cycle. We denote this restricted distribution of graphs by $\widetilde{G}(n, c/n)$ and
we henceforth choose $G \sim \widetilde{G}(n, c/n)$.

To prove the main result, we introduce several auxiliary random variables. For a graph G, we now set
$\mathrm{E}(G)$ to be the number of edges in G and we define $\mathrm{C}(G)$ to be the number of cycles in G. Finally, for a
pair of vertices $e \in \binom{[n]}{2}$, we define Z_e to be the indicator random variable of whether that pair of vertices
is an edge in G.

With this notation, and using our assumption that G is a disjoint union of trees and components
containing a single cycle, we have

$$\boldsymbol{H}_0(G, i_n) = \sum_{\alpha \in \binom{[n]}{i_n}} i_n - \mathrm{E}(G|_\alpha) + \mathrm{C}(G|_\alpha).$$

Ignoring the cycles, we get

$$\geq \sum_{\alpha \in \binom{[n]}{i_n}} i_n - \mathrm{E}(G|_\alpha) = \binom{n}{i_n}i_n - \sum_{\alpha \in \binom{[n]}{i_n}} \mathrm{E}(G|_\alpha).$$

We may rewrite the right-hand sum in terms of individual edges to obtain

$$= \binom{n}{i_n} i_n - \sum_{e \in \binom{[n]}{2}} \binom{n}{i_n - 2} Z_e.$$

But $E(G)$ is the sum of the Z_e, and thus we have

$$= \binom{n}{i_n} i_n - \binom{n}{i_n - 2} E(G).$$

By a similar argument, but where we do not ignore $C(G|_\alpha)$, we can use the fact that G has at most $n^{1-\epsilon}$ cycles to obtain an upper bound $H_0(G, i_n) \le \binom{n}{i_n} i_n - \binom{n}{i_n-2}(E(G) - n^{1-\epsilon})$:

$$\binom{n}{i_n} i_n - \binom{n}{i_n - 2} E(G) \le H_0(G, i_n) \le \binom{n}{i_n} i_n - \binom{n}{i_n - 2}(E(G) - n^{1-\epsilon}). \tag{6.3}$$

We have

$$\binom{n}{i_n - 2} = \binom{n}{i_n} \frac{i_n(i_n - 1)}{(n - i_n + 2)(n - i_n + 1)}$$

and since $i_n = n/2 + o(n)$ this yields that $\binom{n}{i_n-2} = \binom{n}{i_n}(1 + o(1))$. Applying this to (6.3) yields:

$$\binom{n}{i_n}\big(i_n - (1 + o(1))E(G)\big) \le H_0(G, i_n) \le \binom{n}{i_n}\big(i_n - (1 + o(1))(E(G) - n^{1-\epsilon})\big).$$

Recall that $C = (1 - c)/2$. We now divide through by $1/\big(Cn\binom{n}{i_n}\big)$. By rewriting $i_n = n/2 + o(n)$ and absorbing the $n^{1-\epsilon}$ term into the $o(n)$, the left-hand and right-hand bounds have the same form, and we obtain

$$\frac{H_0(G, i_n)}{Cn\binom{n}{i_n}} = \frac{(n/2) - E(G) + o(n) + o(1)E(G)}{Cn}.$$

Since $E(G)$ is a sum of independent random variables, one for each potential edge, this now essentially reduces to a weak law of large numbers argument. In particular, we have that the variance of $E(G)$ is $\binom{n}{2}p(1 - p)$ and the mean is $\binom{n}{2}p = c(n - 1)/2$. We apply Chebyshev's inequality to the random variable $E(G)/n$:

$$P\left[\left|\frac{c(n-1)}{2n} - \frac{E(G)}{n}\right| \ge \epsilon\right] \le \frac{\mathrm{Var}(E(G)/n)}{\epsilon^2} = \frac{\binom{n}{2}p(1 - p)}{n^2\epsilon^2}.$$

Since $p = c/n$ and $1 - p < 1$ this simplifies to $c(n - 1)/(2n^2\epsilon^2)$. For fixed ϵ we have

$$\lim_{n \to \infty} \frac{c(n - 1)}{2n^2\epsilon^2} = 0.$$

Since $\lim_{n \to \infty} c(n - 1)/(2n) = c/2$, we conclude that $E(G)/n$ converges to $c/2$ in probability. This implies that

$$\frac{H_0(G, i_n)}{Cn\binom{n}{i_n}} \to 1$$

in probability. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 1.4. Let G be the 1-skeleton of Δ. By Hochster's formula [Bruns and Herzog 1993, Theorem 5.5.1], $\beta_{i_n, i_n+1}(S/I_\Delta) = H_0(G, i_n)$. The statement is now an immediate corollary of Proposition 6.2.   □

Proof of Corollary 1.5. Let $C = (1-c)/2$. Using Theorem 1.4 and the normal approximation of the binomial distribution, e.g., [Boas 2006, (8.3), p. 762], we obtain that

$$\beta_{i_n, i_n+1}(S/I_\Delta) \sim Cn\binom{n}{i_n} \sim Cn\frac{2^{n+1}}{\sqrt{2\pi n}}e^{-a^2/2}.$$

Therefore we have

$$\frac{\sqrt{2\pi n}}{Cn2^{n+1}}\beta_{i_n, i_n+1}(S/I_\Delta) = \frac{\sqrt{2\pi}}{(1-c)2^n\sqrt{n}}\beta_{i_n, i_n+1}(S/I_\Delta) \sim e^{-a^2/2}.$$

Since the right-hand side is a constant, we have convergence in probability.   □

Conjecture 6.4. *In cases where Theorem 1.3 yields nonvanishing Betti numbers in row k, we conjecture that the k-th row of the Betti table will be normally distributed, in a manner similar to Corollary 1.5.*

7. Projective dimension estimates

We conclude with a corollary about Cohen–Macaulayness. For many values of p, we show that S/I_Δ will essentially never be Cohen–Macaulay. However, while the projective dimension almost never equals the codimension of S/I_Δ, with high probability the ratio of these quantities converges to 1 as $n \to \infty$.

Corollary 7.1. *For any $k \geq 1$, and any p satisfying $1/n^{2/3} \ll p \ll (\log n/n)^{2/(k+3)}$ we have that* $\text{codim}(S/I_\Delta)/\text{pdim}(S/I_\Delta) \to 1$ *in probability, yet the probability that S/I_Δ is Cohen–Macaulay goes to* 0.

First we prove a quick lemma bounding the dimension of Δ.

Lemma 7.2. *If $p \leq \epsilon$ for some $0 < \epsilon < 1$ then $P[\dim \Delta \geq \epsilon \cdot n] \to 0$ as $n \to \infty$.*

Proof. The dimension of Δ is the size of the largest k-clique in Δ. Let $N := \binom{n}{k}$. The expected number of k-cliques in Δ is $Np^N \leq N\epsilon^N$, which goes to zero as $n \to \infty$.   □

Note that [Bollobás and Erdős 1976, Theorem 1] provides a much sharper estimate of the dimension of Δ, though we will not need that.

Proof of Corollary 7.1. Lemma 7.2 shows that $\dim \Delta = o(n)$ with high probability. By Auslander–Buchsbaum, this implies that

$$n - o(n) \leq \text{codim}(S/I_\Delta) \leq \text{pdim}(S/I_\Delta) \leq n.$$

Thus the ratio between $\text{pdim}(S/I_\Delta)$ and $\text{codim}(S/I_\Delta)$ goes to 1 in probability.

For the statement on Cohen–Macaulayness, using Reisner's criterion [Bruns and Herzog 1993, Corollary 5.3.9] it suffices to show that there exists a vertex $v \in \Delta$ and an integer $i < \dim(\text{link}_\Delta(v))$ where $\widetilde{H}_i(\text{link}_\Delta(v)) \neq 0$. For $\Delta \sim \Delta(n, p)$ and a vertex v, the link of v is itself a random flag complex, namely $\text{link}_\Delta(v) \sim \Delta(np, p)$.

For convenience we write $m := np$. In terms of m we can rewrite the left-hand side of the original constraints on p as $1/m^2 \ll p$. For the right-hand side of the constraint, since $1/n \ll p$, we have $\log m \sim \log n$ so we get $p \ll (\log n/n)^{2/(k+3)} \sim (\log m/m)^{2/(k+1)}$. Thus the constraints in terms of m are

$$\frac{1}{m^2} \ll p \ll \left(\frac{\log m}{m}\right)^{2/(k+1)}.$$

For $1 \leq t \leq k$, we consider the interval $1/m^{2/t} \ll p \ll (\log m/m)^{2/(t+1)}$. Since $1/m^{2/(t+1)} \ll (\log m/m)^{2/(t+1)}$, the successive intervals overlap, and it suffices to show that for each of these intervals Δ is not Cohen–Macaulay with probability approaching 1.

First let us consider the case where $t \geq 2$. Setting $i := \lfloor t/2 \rfloor$ and applying [Kahle 2014a, Theorem 1.1] we have $\widetilde{H}_i(\text{link}_\Delta(v)) \neq 0$ with probability $1 - o(1)$. Since $1/m^{2/t} \ll p$, there exist $(t+1)$-cliques and thus $\dim(\text{link}_\Delta(v)) \geq t$ with probability $1 - o(1)$. Together these imply that Δ is not Cohen–Macaulay with probability $1 - o(1)$

We now consider the case $t = k = 1$, where we have $1/m^2 \ll p \ll \log m/m$. Thus we apply [Erdős and Rényi 1959, Theorem 1] to get $\widetilde{H}_0(\text{link}_\Delta(v)) \neq 0$ with probability $1 - o(1)$. On the other hand, since $1/m^2 \ll p$, we have 2-cliques and thus $\dim(\text{link}_\Delta(v)) \geq t$ with probability $1 - o(1)$ □

Acknowledgments

References

[Alon and Spencer 2016] N. Alon and J. H. Spencer, *The probabilistic method*, 4th ed., Wiley, Hoboken, NJ, 2016. MR Zbl

[Boas 2006] M. L. Boas, *Mathematical methods in the physical sciences*, Wiley, 2006. Zbl

[Bollobás and Erdős 1976] B. Bollobás and P. Erdős, "Cliques in random graphs", *Math. Proc. Cambridge Philos. Soc.* **80**:3 (1976), 419–427. MR Zbl

[Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. MR

[Conca et al. 2018] A. Conca, M. Juhnke-Kubitzke, and V. Welker, "Asymptotic syzygies of Stanley–Reisner rings of iterated subdivisions", *Trans. Amer. Math. Soc.* **370**:3 (2018), 1661–1691. MR Zbl

[Ein and Lazarsfeld 2012] L. Ein and R. Lazarsfeld, "Asymptotic syzygies of algebraic varieties", *Invent. Math.* **190**:3 (2012), 603–646. MR Zbl

[Ein et al. 2015] L. Ein, D. Erman, and R. Lazarsfeld, "Asymptotics of random Betti tables", *J. Reine Angew. Math.* **702** (2015), 55–75. MR Zbl

[Ein et al. 2016] L. Ein, D. Erman, and R. Lazarsfeld, "A quick proof of nonvanishing for asymptotic syzygies", *Algebr. Geom.* **3**:2 (2016), 211–222. MR Zbl

[Eisenbud 2005] D. Eisenbud, *The geometry of syzygies*, Graduate Texts in Mathematics **229**, Springer, New York, 2005. MR Zbl

[Eisenbud et al. 1994] D. Eisenbud, A. Reeves, and B. Totaro, "Initial ideals, Veronese subrings, and rates of algebras", *Adv. Math.* **109**:2 (1994), 168–187. MR Zbl

[Erdős and Rényi 1959] P. Erdős and A. Rényi, "On random graphs, I", *Publ. Math. Debrecen* **6** (1959), 290–297. MR

[Erdős and Rényi 1960] P. Erdős and A. Rényi, "On the evolution of random graphs", *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61. MR

[Frieze and Karoński 2016] A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, 2016. MR Zbl

[Kahle 2014a] M. Kahle, "Sharp vanishing thresholds for cohomology of random flag complexes", *Ann. of Math.* (2) **179**:3 (2014), 1085–1107. MR Zbl

[Kahle 2014b] M. Kahle, "Topology of random simplicial complexes: a survey", pp. 201–221 in *Algebraic topology: applications and new directions*, edited by U. Tillmann et al., Contemp. Math. **620**, Amer. Math. Soc., Providence, RI, 2014. MR Zbl

[Loera et al. 2017] J. A. D. Loera, S. Petrovic, L. Silverstein, D. Stasi, and D. Wilburne, "Random monomial ideals", 2017. arXiv

[Macaulay2] D. R. Grayson and M. E. Stillman, "Macaulay2, a software system for research in algebraic geometry", available at http://www.math.uiuc.edu/Macaulay2/.

[Peeva 2011] I. Peeva, *Graded syzygies*, Algebra and Applications **14**, Springer, London, 2011. MR Zbl

[Zhou 2014] X. Zhou, "Effective non-vanishing of asymptotic adjoint syzygies", *Proc. Amer. Math. Soc.* **142**:7 (2014), 2255–2264. MR Zbl

derman@math.wisc.edu   *Department of Mathematics, University of Wisconsin, Madison, WI, United States*

jkyang@umn.edu   *Department of Mathematics, University of Minnesota Twin Cities, Minneapolis 55455,*

Grothendieck rings for Lie superalgebras and the Duflo–Serganova functor

Crystal Hoyt and Shifra Reif

We show that the Duflo–Serganova functor on the category of finite-dimensional modules over a finite-dimensional contragredient Lie superalgebra induces a ring homomorphism on a natural quotient of the Grothendieck ring, which is isomorphic to the ring of supercharacters. We realize this homomorphism as a certain evaluation of functions related to the supersymmetry property. We use this realization to describe the kernel and image of the homomorphism induced by the Duflo–Serganova functor.

1. Introduction

The Duflo–Serganova functor was originally introduced in [Duflo and Serganova 2005] together with associated varieties of modules over Lie superalgebras. On the category of finite-dimensional modules, the Duflo–Serganova functor is a tensor functor which preserves the superdimension. This functor was used by Serganova [2011] to prove the conjecture of Kac and Wakimoto that the superdimension of a finite-dimensional module is zero if and only if the atypicality of the module is maximal. The Duflo–Serganova functor was also used to give an additional proof for the superdimension formula of $GL(m \mid n)$-modules in [Heidersdorf and Weissauer 2014], and has been applied to study Deligne categories in [Comes and Heidersdorf 2017; Entova-Aizenbud et al. 2015; Heidersdorf 2015; Heidersdorf and Weissauer 2015].

Given an odd element x in a Lie superalgebra $\mathfrak{g}$ satisfying $[x, x] = 0$, we have that $x^2 = 0$ in the universal enveloping algebra of $\mathfrak{g}$, and so for every $\mathfrak{g}$-module M, we can define the cohomology

$$M_x := \operatorname{Ker}_M x / x M.$$

In fact, M_x is a module for the Lie superalgebra

$$\mathfrak{g}_x := \operatorname{Ker} \operatorname{ad}_x / \operatorname{Im} \operatorname{ad}_x,$$

which is a Lie superalgebra of smaller rank than $\mathfrak{g}$. For example, if $\mathfrak{g} = \mathfrak{gl}(m \mid n)$ and x is a root vector, then $\mathfrak{g}_x = \mathfrak{gl}(m - 1 \mid n - 1)$. Duflo and Serganova [2005] defined the functor $DS_x : M \mapsto M_x$ from the category of $\mathfrak{g}$-modules to the category of $\mathfrak{g}_x$-modules, which we refer to as the Duflo–Serganova functor.

One of the difficulties that arises in using the Duflo–Serganova functor is that it is not exact. It is therefore surprising that it induces a ring homomorphism ds_x on a natural quotient of the Grothendieck ring

of the category of finite-dimensional $\mathfrak{g}$-modules. This quotient is defined by identifying the equivalence class of a module $[M]$ with $-[\Pi(M)]$, where Π is the shift of parity functor. We refer to this quotient as the *supercharacter ring* of $\mathfrak{g}$ and show that the homomorphism ds_x is indeed well defined.

Sergeev and Veselov [2011] described the supercharacter ring as a ring of functions admitting a certain supersymmetry condition. In this paper, we realize the homomorphism ds_x in terms of evaluation of functions related to the supersymmetry condition. For example, the supercharacter ring of the Lie supergroup $GL(m \mid n)$ corresponding to the Lie superalgebra $\mathfrak{gl}(m \mid n)$ is isomorphic to the ring of doubly symmetric Laurent polynomials in $x_1, \ldots, x_m, y_1, \ldots, y_n$ for which the evaluation $x_1 = y_1 = t$ is independent of t. If x is a root vector for the root $\varepsilon_i - \delta_j$ of $\mathfrak{gl}(m \mid n)$, then the homomorphism ds_x is given by the evaluation $x_i = y_j = t$, which is independent of the variable t after evaluation, by the supersymmetry property.

We use this realization to describe the kernel of the homomorphism ds_x when x is a root vector. In particular, we show that if $\mathfrak{g}$ is a Lie superalgebra of type I, the supercharacters of Kac modules form a basis for the kernel. When $\mathfrak{g}$ is a Lie superalgebra of type II, there are no Kac modules; however, we show that the kernel has a basis consisting of expressions similar to the supercharacters of Kac modules. These are the same expressions that were used by Gruson and Serganova [2010] to define Kazhdan–Lusztig polynomials for the orthosymplectic Lie superalgebras.

We also describe the image of ds_x. In particular, for $\mathfrak{g} = \mathfrak{sl}(m \mid n)$, $m \neq n$, and $\mathfrak{osp}(m \mid 2n)$, we show that the image is the supercharacter ring of G_x, where G_x is the Lie supergroup corresponding to the Lie superalgebra $\mathfrak{g}_x$. Moreover, we prove that the homomorphism induced by the Duflo–Serganova functor from the category of finite-dimensional G-modules to the category of finite-dimensional G_x-modules is surjective. For the exceptional Lie superalgebras, we explicitly describe the image using a set of generators.

2. Preliminaries

2A. *Lie superalgebras.* Lie superalgebras are a natural generalization of Lie algebras which first appeared in mathematical physics. In this paper, we study the finite-dimensional contragredient Lie superalgebras $\mathfrak{g} = \mathfrak{g}_{\bar{0}} \oplus \mathfrak{g}_{\bar{1}}$ with indecomposable Cartan matrix. These are the Lie superalgebras $\mathfrak{sl}(m \mid n)$, $m \neq n$, $\mathfrak{gl}(n \mid n)$, $\mathfrak{osp}(m \mid 2n)$, $D(2, 1, \alpha)$, $F(4)$, or $G(3)$. We also consider the case when $\mathfrak{g} = \mathfrak{gl}(m \mid n)$ is the general linear Lie superalgebra. These Lie superalgebras resemble reductive Lie algebras in their structure theory; in particular, they are defined by a Cartan matrix and they possess an even supersymmetric invariant bilinear form $(\cdot, \cdot)$ which has kernel equal to the center of $\mathfrak{g}$.

Fix a Cartan subalgebra $\mathfrak{h} \subset \mathfrak{g}_{\bar{0}} \subset \mathfrak{g}$, and consider the corresponding root space decomposition

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Delta} \mathfrak{g}_{\alpha}.$$

Then the set of roots $\Delta \subset \mathfrak{h}^*$ splits $\Delta = \Delta_{\bar{0}} \sqcup \Delta_{\bar{1}}$ into even roots $\Delta_{\bar{0}}$ and odd roots $\Delta_{\bar{1}}$. A choice of positive roots $\Delta^+ = \Delta_{\bar{0}}^+ \sqcup \Delta_{\bar{1}}^+$ determines a triangular decomposition of $\mathfrak{g}$ given by $\mathfrak{g} = \mathfrak{n}^+ \oplus \mathfrak{h} \oplus \mathfrak{n}^-$,

where $\mathfrak{n}^{\pm} = \bigoplus_{\alpha \in \Delta^{\pm}} \mathfrak{g}_{\alpha}$. Let $\rho_{\bar{0}} = \frac{1}{2} \sum_{\alpha \in \Delta_{\bar{0}}^+} \alpha$, $\rho_{\bar{1}} = \frac{1}{2} \sum_{\alpha \in \Delta_{\bar{1}}^+} \alpha$, and $\rho = \rho_{\bar{0}} - \rho_{\bar{1}}$. The Weyl group W of $\mathfrak{g}$ is by definition the Weyl group of $\mathfrak{g}_{\bar{0}}$. The sign map sgn : $W \to \{\pm 1\}$ is defined by $w \mapsto (-1)^{l(w)}$, where $l(w)$ denotes the length of w as a product of simple reflections with respect to a set of simple roots for $\mathfrak{g}_{\bar{0}}$.

The space $\mathfrak{h}^*$ inherits an even supersymmetric bilinear form $(\cdot, \cdot)$. A root $\beta \in \Delta_{\bar{1}}$ is called isotropic if $(\beta, \beta) = 0$. Two roots $\alpha, \beta \in \Delta$ are called orthogonal if $(\alpha, \beta) = 0$. The maximal number of linearly independent mutually orthogonal isotropic roots is called the defect of $\mathfrak{g}$. We denote by $\Delta_{\mathrm{iso}} := \{\beta \in \Delta_{\bar{1}} \mid (\beta, \beta) = 0\}$ the set of all isotropic roots and by $\Delta_{\mathrm{iso}}^+ = \Delta_{\mathrm{iso}} \cap \Delta^+$ the set of positive isotropic roots, and we let $\rho_{\mathrm{iso}} := \frac{1}{2} \sum_{\alpha \in \Delta_{\mathrm{iso}}^+} \alpha$. We define

$$\mathscr{S}_{\mathfrak{g}} = \{B \subset \Delta_{\mathrm{iso}} \mid B = \{\beta_1, \ldots, \beta_k\} \mid (\beta_i, \beta_j) = 0, \ \beta_i \neq \pm\beta_j\}\} \tag{2-1}$$

to be the set of subsets of linearly independent mutually orthogonal isotropic roots.

The space $\mathfrak{h}^*$ has a natural basis $\varepsilon_1, \ldots, \varepsilon_m, \delta_1, \ldots, \delta_n$, which for $\mathfrak{gl}(m \mid n)$ and $\mathfrak{osp}(m \mid 2n)$ satisfies $(\varepsilon_i, \varepsilon_j) = \delta_{ij} = -(\delta_i, \delta_j)$ and $(\varepsilon_i, \delta_j) = 0$. The roots of $\mathfrak{g}$ have a nice presentation in this basis (see [Cheng and Wang 2012; Musson 2012] for more details). Let $Q_{\mathfrak{g}} = \mathrm{span}_{\mathbb{Z}} \Delta$ be the root lattice of $\mathfrak{g}$, and let $Q_{\mathfrak{g}}^+ = \mathrm{span}_{\mathbb{Z}} \Delta^+$. The parity function $p : \Delta \to \mathbb{Z}_2$ extends uniquely to a linear function $p : Q_{\mathfrak{g}} \to \mathbb{Z}_2$. The root lattice $Q_{\mathfrak{g}}$ is contained in the integral weight lattice $P_{\bar{0}}$ for $\mathfrak{g}_{\bar{0}}$, where

$$P_{\bar{0}} = \left\{\lambda \in \mathfrak{h}^* \ \middle| \ \frac{2(\lambda, \alpha)}{(\alpha, \alpha)} \in \mathbb{Z} \text{ for all } \alpha \in \Delta_{\bar{0}}\right\}.$$

The set of dominant integral weights

$$P_{\bar{0}}^+ = \left\{\lambda \in P_{\bar{0}} \ \middle| \ \frac{2(\lambda, \alpha)}{(\alpha, \alpha)} \geq 0 \text{ for all } \alpha \in \Delta_{\bar{0}}\right\}$$

is the set of highest weights of finite-dimensional simple $\mathfrak{g}_{\bar{0}}$-modules.

The category of finite-dimensional modules $\mathscr{F}_{\mathfrak{g}}$ over a Lie superalgebra $\mathfrak{g}$ is not semisimple; that is, there exist indecomposable modules which are not irreducible. For example, a Lie superalgebra $\mathfrak{g}$ of type I has a decomposition $\mathfrak{g} = \mathfrak{g}_{-1} \oplus \mathfrak{g}_{\bar{0}} \oplus \mathfrak{g}_{+1}$, so one can define the Kac module of highest weight $\lambda \in P_{\bar{0}}^+$ as

$$K(\lambda) = \mathrm{Ind}_{\mathfrak{g}_{\bar{0}} \oplus \mathfrak{g}_{+1}}^{\mathfrak{g}} L_{\bar{0}}(\lambda),$$

where $L_{\bar{0}}(\lambda)$ is the finite-dimensional simple $\mathfrak{g}_{\bar{0}}$-module of highest weight λ and $\mathfrak{g}_{+1}$ acts trivially on $L_{\bar{0}}(\lambda)$. Then $K(\lambda)$ is a finite-dimensional, indecomposable $\mathfrak{g}$-module with a unique simple quotient $L(\lambda)$, where λ is the highest weight with respect to the distinguished choice of simple roots, and $K(\lambda)$ is simple (i.e., $K(\lambda) = L(\lambda)$) if and only if λ is a typical weight: $(\lambda + \rho, \beta) \neq 0$ for all $\beta \in \Delta_{\mathrm{iso}}$ (see, for example, [Cheng and Wang 2012, Chapter 2] for more details).

If $G_{\bar{0}}$ is a simply connected and connected Lie group corresponding to the Lie algebra $\mathfrak{g}_{\bar{0}}$ [Serganova 2014], and $\mathscr{F}_G$ is the full subcategory of $\mathscr{F}_{\mathfrak{g}}$ consisting of all finite-dimensional $G_{\bar{0}}$-integrable modules, then $\mathscr{F}_G$ is equivalent to the category of finite-dimensional modules over the corresponding algebraic supergroup G [Serganova 2014].

2B. *Supercharacter rings of Lie superalgebras.* The character theory of Lie superalgebras is a rich area of research which has led to interesting applications in number theory [Kac and Wakimoto 1994; 2014]. For a finite-dimensional $\mathfrak{g}$-module M, with weight decomposition $M = \bigoplus_{\mu \in \mathfrak{h}^*} M^\mu$ and weight spaces $M^\mu = M_{\bar{0}}^\mu \oplus M_{\bar{1}}^\mu$, the supercharacter of M is defined to be

$$\operatorname{sch} M = \sum_{\mu \in \mathfrak{h}^*} (\dim M_{\bar{0}}^\mu - \dim M_{\bar{1}}^\mu) e^\mu,$$

while the character of M is given by $\operatorname{ch} M = \sum (\dim M_{\bar{0}}^\mu + \dim M_{\bar{1}}^\mu) e^\mu$. A finite-dimensional simple $\mathfrak{g}$-module is determined by its supercharacter, as well as by its character [Sergeev and Veselov 2011, Proposition 4.2].

The supercharacter ring $\mathcal{J}_\mathfrak{g}$ of a Lie superalgebra $\mathfrak{g}$ is defined to be the image of the map

$$\operatorname{sch} : \mathcal{F}_\mathfrak{g} \to \mathbb{Z}[P_{\bar{0}}]^W,$$

where $\mathbb{Z}[P_{\bar{0}}] := \mathbb{Z}\{e^\mu \mid \mu \in P_{\bar{0}}\}$. For an element $f \in \mathcal{J}_\mathfrak{g}$, with $f = \sum_{\mu \in P_{\bar{0}}} c_\mu e^\mu$, we call the set $\operatorname{Supp} f = \{\mu \in P_{\bar{0}} \mid c_\mu \neq 0\}$ the support of f.

For a fixed choice of positive roots $\Delta^+ = \Delta_{\bar{0}}^+ \sqcup \Delta_{\bar{1}}^+$, we denote the super Weyl denominator by $R = R_{\bar{0}}/R_{\bar{1}}$ where $R_{\bar{0}} = \prod_{\alpha \in \Delta_{\bar{0}}^+} (1 - e^{-\alpha})$ and $R_{\bar{1}} = \prod_{\alpha \in \Delta_{\bar{1}}^+} (1 - e^{-\alpha})$. Note that the supercharacter of the Kac module equals

$$\operatorname{sch} K(\lambda) = e^{-\rho} R^{-1} \cdot \operatorname{ch} L_{\bar{0}}(\lambda),$$

where $\Delta^+ = \Delta_{\bar{0}}^+ \sqcup \Delta_{\bar{1}}^+$ is the distinguished choice of simple roots.

The Grothendieck group of the category $\mathcal{F}_\mathfrak{g}$ is defined by taking the free abelian group generated by the elements $[M]$ which represent each isomorphism class of finite-dimensional $\mathfrak{g}$-modules, and modding out by the relations $[M_1] - [M_2] + [M_3]$ for all exact sequences $0 \to M_1 \to M_2 \to M_3 \to$. Since $\mathcal{F}_\mathfrak{g}$ is closed under tensor products, the Grothendieck group inherits a natural ring structure.

The Grothendieck ring of $\mathcal{F}_\mathfrak{g}$ has a natural quotient described as follows. Let Π denote the parity reversing functor from $\mathcal{F}_\mathfrak{g}$ to itself, and let $\mathcal{K}_\mathfrak{g}$ denote the quotient of the Grothendieck ring of $\mathcal{F}_\mathfrak{g}$ by the ideal $\langle [\Pi(M)] + [M] \mid M$ is a $\mathfrak{g}$-module$\rangle$. The map $\operatorname{sch} : \mathcal{K}_\mathfrak{g} \to \mathbb{Z}[P_{\bar{0}}]^{W_{\bar{0}}}$ given on generators by $[M] \mapsto \operatorname{sch} M$ is injective [Sergeev and Veselov 2011, Proposition 4.4], and its image is the supercharacter ring $\mathcal{J}_\mathfrak{g}$ of $\mathfrak{g}$.

Remark 1. In this paper, we identify the rings $\mathcal{K}_\mathfrak{g}$ and $\mathcal{J}_\mathfrak{g}$ under this isomorphism, and use the notation $\mathcal{J}_\mathfrak{g}$ to denote this ring. Given a module $M \in \mathcal{F}_\mathfrak{g}$, we write $[M]$ for its image in $\mathcal{J}_\mathfrak{g}$.

Sergeev and Veselov [2011] gave an explicit description of supercharacter rings for basic classical Lie superalgebras as follows. The supercharacter ring of $\mathfrak{g}$ is isomorphic to the space of supersymmetric exponential functions

$$\mathcal{J}_\mathfrak{g} = \{f \in \mathbb{Z}[P_{\bar{0}}]^W \mid D_\beta f \text{ is in the ideal generated by } (e^\beta - 1) \text{ for any } \beta \in \Delta_{\mathrm{iso}}\} \tag{2-2}$$

where $D_\beta(e^\lambda) = (\lambda, \beta)e^\beta$. Sergeev and Veselov [2011, §7] also described the supercharacter ring $\mathscr{J}_G \subset \mathscr{J}_{\mathfrak{g}}$ for the Lie supergroup G corresponding to the Lie superalgebra $\mathfrak{g}$ as a ring of Laurent polynomials subject to some additional conditions. Recall the basis $\varepsilon_1, \ldots, \varepsilon_m, \delta_1, \ldots, \delta_n$ of $\mathfrak{h}^*$, and define $x_i := e^{\varepsilon_i}$, $y_j := e^{\delta_j}$, $u_i = x_i + x_i^{-1}$, and $v_j = y_j + y_j^{-1}$.

$GL(m \mid n)$: *The supercharacter ring of $GL(m \mid n)$ is*

$$\mathscr{J}_G = \left\{ f \in \mathbb{Z}[x_1^{\pm 1}, \ldots, x_m^{\pm 1}, y_1^{\pm 1}, \ldots, y_n^{\pm 1}]^{S_m \times S_n} \ \middle| \ y_j \frac{\partial f}{\partial y_j} + x_i \frac{\partial f}{\partial x_i} \in \langle y_j - x_i \rangle \right\}. \tag{2-3}$$

$SL(m \mid n)$, $m \neq n$: *The supercharacter ring of $SL(m \mid n)$ for $m \neq n$ is the quotient of (2-3) by the ideal*
$\langle x_1 \cdots x_m - y_1 \cdots y_n \rangle$.

$B(m \mid n)$: *The supercharacter ring of $OSP(2m + 1 \mid 2n)$ is*

$$\mathscr{J}_G = \left\{ f \in \mathbb{Z}[u_1, \ldots, u_m, v_1, \ldots, v_n]^{S_m \times S_n} \ \middle| \ u_i \frac{\partial f}{\partial u_i} + v_j \frac{\partial f}{\partial v_j} \in \langle u_i - v_j \rangle \right\}.$$

$C(n + 1)$: *The supercharacter ring of $OSP(2 \mid 2n)$ is*

$$\mathscr{J}_G = \left\{ f \in \mathbb{Z}[u_1, v_1, \ldots, v_n]^{S_m} \ \middle| \ u_1 \frac{\partial f}{\partial u_1} + v_j \frac{\partial f}{\partial v_j} \in \langle u_1 - v_j \rangle \right\}.$$

$D(m \mid n)$, $m \geq 2$: *The supercharacter ring of $OSP(2m \mid 2n)$ for $m \geq 2$ is*

$$\mathscr{J}_G = \left\{ f \in \mathbb{Z}[u_1, \ldots, u_m, v_1, \ldots, v_n]^{S_m \times S_n} \ \middle| \ u_i \frac{\partial f}{\partial u_i} + v_j \frac{\partial f}{\partial v_j} \in \langle u_i - v_j \rangle \right\}.$$

Remark 2. Note that $f \in \mathscr{J}_{GL(m|n)}$ if and only if it is supersymmetric in $x_1, \ldots, x_m, y_1, \ldots y_n$, that is, if it is invariant under permutation of $x_1, \ldots, x_m$ and of $y_1, \ldots, y_n$, and if the substitution $x_1 = y_1 = t$ made in f is independent of t (see for example [Musson 2012, §12]).

2C. *The Duflo–Serganova functor.* The idea behind the Duflo–Serganova functor is simple and natural. For any odd element $x \in \mathfrak{g}_{\bar{1}}$ of a finite-dimensional contragredient Lie superalgebra $\mathfrak{g}$ which satisfies $[x, x] = 0$, we have that $x^2 = 0$ in the universal enveloping algebra of $\mathfrak{g}$, and so for any finite-dimensional $\mathfrak{g}$-module M we can define the cohomology

$$M_x := \mathrm{Ker}_M x / x M. \tag{2-4}$$

Then M_x is in fact a module over the Lie superalgebra

$$\mathfrak{g}_x := \mathfrak{g}^x / [x, \mathfrak{g}],$$

where $\mathfrak{g}^x = \{a \in \mathfrak{g} \mid [x, a] = 0\}$ is the centralizer of x in $\mathfrak{g}$ [Duflo and Serganova 2005, Lemma 6.2]. The Duflo–Serganova functor $DS_x : \mathscr{F}_{\mathfrak{g}} \to \mathscr{F}_{\mathfrak{g}_x}$ is defined from the category of finite-dimensional $\mathfrak{g}$-modules to the category of finite-dimensional $\mathfrak{g}_x$-modules by sending $M \mapsto M_x$.

The Duflo–Serganova functor is a cohomology functor and hence is a symmetric monoidal tensor functor; that is, for $\mathfrak{g}$-modules M, N one has a natural isomorphism $M_x \otimes N_x \to (M \otimes N)_x$ [Serganova 2011]. Moreover, the Duflo–Serganova functor commutes with direct sums; however, it is not exact.

Let $X_{\mathfrak{g}} = \{x \in \mathfrak{g}_{\bar{1}} : [x, x] = 0\}$, and let $\mathscr{S}_{\mathfrak{g}}$ be the set of subsets of mutually orthogonal isotropic roots (see (2-1)). Then the $G_{\bar{0}}$-orbits of $X_{\mathfrak{g}}$ are in one-to-one correspondence with the W-orbits of $\mathscr{S}_{\mathfrak{g}}$ via the correspondence

$$B = \{\beta_1, \ldots, \beta_k\} \mapsto x = x_{\beta_1} + \cdots + x_{\beta_k} \in X_{\mathfrak{g}}, \tag{2-5}$$

where each $x_{\beta_i} \in \mathfrak{g}_{\beta_i}$ is chosen to be nonzero [Duflo and Serganova 2005, Theorem 4.2].

The Lie superalgebra $\mathfrak{g}_x$ can be naturally embedded into $\mathfrak{g}^x \subset \mathfrak{g}$, in such a way that $\mathfrak{h}_x = \mathfrak{h} \cap \mathfrak{g}_x$ is a Cartan subalgebra of $\mathfrak{g}_x$ and the root spaces of $\mathfrak{g}_x$ are root spaces of $\mathfrak{g}$ [Duflo and Serganova 2005, Lemma 6.3]. More explicitly, Duflo and Serganova proved the following:

If $B = \{\beta_1, \ldots, \beta_k\} \in \mathscr{S}$ and $x = x_{\beta_1} + \cdots + x_{\beta_k}$ for some nonzero $x_{\beta_i} \in \mathfrak{g}_{\beta_i}$, then $\mathfrak{g}^x \subset \mathfrak{g}$ can be decomposed into a semidirect sum $\mathfrak{g}^x = [x, \mathfrak{g}] \oplus \mathfrak{g}_x$, where $\mathfrak{g}_x = \mathfrak{h}_x \oplus \left(\bigoplus_{\alpha \in \Delta_x} \mathfrak{g}_\alpha\right)$, the subspace $\mathfrak{h}_x = \mathfrak{h} \cap \mathfrak{g}_x$ is a Cartan subalgebra of $\mathfrak{g}_x$, and

$$\Delta_x = \{\alpha \in \Delta \mid (\alpha, \beta) = 0 \text{ for all } \beta \in B \text{ and } \pm\alpha \notin B\} \tag{2-6}$$

is the root system of $\mathfrak{g}_x$.

For each finite-dimensional contragredient Lie superalgebra $\mathfrak{g}$ with irreducible Cartan matrix, we can explicitly describe the isomorphism type of $\mathfrak{g}_x$. If $\mathscr{B} = \{\beta_1, \ldots, \beta_k\} \in \mathscr{S}$ and $x = x_{\beta_1} + \cdots + x_{\beta_k}$ for some nonzero $x_{\beta_i} \in \mathfrak{g}_{\beta_i}$, then by [Duflo and Serganova 2005, Remark 6.4] we have the following description. In particular, the defect of $\mathfrak{g}_x$ equals the defect of $\mathfrak{g}$ minus k. Note that in the last three columns the defect of $\mathfrak{g}$ is 1 and $k = 1$:

$\mathfrak{g}$	$\mathfrak{gl}(m \mid n)$	$\mathfrak{sl}(m \mid n), m \neq n$	$\mathfrak{osp}(m \mid 2n)$	$D(2, 1, \alpha)$	F_4	G_3
$\mathfrak{g}_x$	$\mathfrak{gl}(m - k \mid n - k)$	$\mathfrak{sl}(m - k \mid n - k)$	$\mathfrak{osp}(m - 2k \mid 2n - 2k)$	$\mathbb{C}$	$\mathfrak{sl}(3)$	$\mathfrak{sl}(2)$

Remark 3. Note that when $\mathfrak{g}_x$ is simple, the embedding $\mathfrak{g}^x \subset \mathfrak{g}$ is determined by the condition that the root spaces of $\mathfrak{g}_x$ are mapped into the respective root spaces of $\mathfrak{g}$, since in this case $\mathfrak{h}_x \subset [\mathfrak{n}_x^+, \mathfrak{n}_x^-]$. For $\mathfrak{g} = \mathfrak{gl}(m, n)$, we take the matrix embedding of $\mathfrak{g}_x = \mathfrak{gl}(m - k \mid n - k)$ into $\mathfrak{gl}(m \mid n)$ which has $2k$ zero rows and $2k$ zero columns at the locations $r_i, n + s_i$, for $i = 1, \ldots, k$, when $B = \{\beta_i = \varepsilon_{r_i} - \delta_{s_i}\}_{i=1,\ldots,k}$ is the set of maximal isotropic roots defining x.

3. The Duflo–Serganova functor and the supercharacter ring

In this section, we prove that the Duflo–Serganova functor $DS_x : \mathscr{F}_{\mathfrak{g}} \to \mathscr{F}_{\mathfrak{g}_x}$ induces a ring homomorphism $ds_x : \mathscr{J}_{\mathfrak{g}} \to \mathscr{J}_{\mathfrak{g}_x}$, and we realize it as a certain evaluation of the functions $f \in \mathscr{J}_{\mathfrak{g}}$ related to the supersymmetry property defining $\mathscr{J}_{\mathfrak{g}}$.

3A. *The ring homomorphism induced by the Duflo–Serganova functor.* Let $\mathfrak{g}$ be a finite-dimensional contragredient Lie superalgebra with indecomposable Cartan matrix, or let $\mathfrak{g} = \mathfrak{gl}(m, n)$, and fix a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$. Let $B = \{\beta_1, \dots, \beta_k\} \in \mathscr{S}_{\mathfrak{g}}$, $x \in X_{\mathfrak{g}}$, and $x = x_{\beta_1} + \cdots + x_{\beta_k}$ for nonzero $x_{\beta_i} \in \mathfrak{g}_{\beta_i}$. Fix an embedding $\mathfrak{g}_x \subset \mathfrak{g}^x \subset \mathfrak{g}$ with Cartan subalgebra $\mathfrak{h}_x = \mathfrak{h} \cap \mathfrak{g}_x$ (see Section 2C).

Lemma 4. *For $\mathfrak{g}$-modules M and N we have*

(1) $\mathrm{sch}\, M_x(h) = \mathrm{sch}\, M(h)$ *for all $h \in \mathfrak{h}_x$ and*

(2) *if $\mathrm{sch}\, M = \mathrm{sch}\, N$, then $\mathrm{sch}\, M_x = \mathrm{sch}\, N_x$.*

Proof. We have an exact sequence $0 \to \ker_M x \to M \to xM \to 0$ of $\mathfrak{h}^x$-invariant spaces. Thus, $M / \ker_M x \cong \Pi(xM)$ as $\mathfrak{h}^x$-modules, where Π switches the parity of a superspace, and so $\mathrm{sch}(M / \ker_M x)(h) = \mathrm{sch}\, \Pi(xM)(h)$ for all $h \in \mathfrak{h}^x$. Hence, for all $h \in \mathfrak{h}_x \subset \mathfrak{h}^x$ we have that

$$\mathrm{sch}\, M(h) = \mathrm{sch}\, \ker x(h) + \mathrm{sch}\, \Pi(xM)(h) = \mathrm{sch}\, \ker x(h) - \mathrm{sch}\, M(h) = \mathrm{sch}(\ker_M x / xM)(h)$$
$$= \mathrm{sch}(M_x)(h). \qquad \square$$

Remark 5. The following example shows that Lemma 4 does not hold if we replace supercharacter by character. It also shows that the Duflo–Serganova functor is not exact.

Example 6. Let $\mathfrak{g} = \mathfrak{gl}(2 \mid 1)$ with the standard choice of simple roots $\{\alpha = \varepsilon_1 - \varepsilon_2,\ \beta = \varepsilon_2 - \delta_1\}$. Let $K(0)$ be the Kac module with highest weight zero, and denote the highest weight vector by v_0. Then $K(0) = \mathrm{span}\{v_0, f_\beta v_0, f_{\alpha+\beta} v_0, f_\beta f_{\alpha+\beta} v_0\}$, where $f_\beta \in \mathfrak{g}_{-\beta}$ and $f_{\alpha+\beta} \in \mathfrak{g}_{-\alpha-\beta}$ are nonzero. The maximal submodule of $K(0)$ is $\overline{K}(0) := \mathrm{span}\{f_\beta v_0, f_{\alpha+\beta} v_0, f_\beta f_{\alpha+\beta} v_0\}$, and the simple quotient of $K(0)$ is isomorphic to the trivial $\mathfrak{g}$-module $L(0)$. Clearly, the $\mathfrak{g}$-modules $K(0)$ and $L(0) \oplus \overline{K}(0)$ have the same character and supercharacter.

Let us show that for $x = f_\beta$, the $\mathfrak{g}_x$-modules $K(0)_x$ and $(L(0) \oplus \overline{K}(0))_x$ have the same supercharacter but not the same character. In this case, $\mathfrak{g}_x = \mathfrak{gl}(1 \mid 0)$. By a direct computation using (2-4) and the basis given above, one can check that $K(0)_x = \{0\}$, $L(0)_x \cong \mathbb{C}_{1|0}$, and $\overline{K}(0)_x \cong \mathbb{C}_{0|1}$, where $\mathbb{C}_{1|0}$ and $\mathbb{C}_{0|1}$ are the even and odd trivial $\mathfrak{g}_x$-modules, respectively. Thus, $\mathrm{ch}\, K(0)_x = \mathrm{sch}\, K(0)_x = 0$ and $\mathrm{sch}(L(0) \oplus \overline{K}(0))_x = 0$, while $\mathrm{ch}(L(0) \oplus \overline{K}(0))_x = 2$.

Definition 7. We define $ds_x : \mathscr{J}_{\mathfrak{g}} \to \mathscr{J}_{\mathfrak{g}_x}$ on the generators $[M] \in \mathscr{J}_{\mathfrak{g}}$, where $M \in \mathscr{F}_{\mathfrak{g}}$, by

$$ds_x([M]) = [DS_x(M)],$$

and we extend linearly to $\mathscr{J}_{\mathfrak{g}}$.

It is not difficult to show that ds_x is a well defined linear map using Lemma 4. The fact that ds_x is a ring homomorphism then follows from the fact that DS_x is a tensor functor. Hence, we have:

Proposition 8. *Let $\mathfrak{g}$ be a finite-dimensional contragredient Lie superalgebra, and let $x \in \mathfrak{g}_{\bar{1}}$ nonzero such that $[x, x] = 0$. The functor $DS_x : \mathscr{F}_{\mathfrak{g}} \to \mathscr{F}_{\mathfrak{g}_x}$ induces a ring homomorphism on the corresponding supercharacter rings $ds_x : \mathscr{J}_{\mathfrak{g}} \to \mathscr{J}_{\mathfrak{g}_x}$.*

Remark 9. The proofs in Section 3A also work for modules in the BGG category $\mathbb{O}$, and so the Duflo–Serganova functor induces a group homomorphism on the quotient of the Grothendieck group by the parity. However, it is not a ring homomorphism since category $\mathbb{O}$ is not closed under tensor products.

3B. *Realization of the ring homomorphism.* Given $f \in \mathcal{J}_{\mathfrak{g}}$ we can realize $f : \mathfrak{h} \to \mathbb{C}$ as a supersymmetric function in the variables $x_1, \ldots, x_m, y_1, \ldots, y_n$ with $x_i = e^{\varepsilon_i}$ and $y_j = e^{\delta_j}$, using the supercharacter ring description of Sergeev and Veselov (see (2-2)). (Note that for $F(4)$ we take $x_i = e^{(1/2)\varepsilon_i}$ and $y_j = e^{(1/2)\delta_j}$.)

Theorem 10. *Suppose $ds_x : \mathcal{J}_{\mathfrak{g}} \to \mathcal{J}_{\mathfrak{g}_x}$ is defined by $x = x_{\beta_1} + \cdots + x_{\beta_k}$ for nonzero $x_{\beta_i} \in \mathfrak{g}_{\beta_i}$, where $B = \{\beta_1, \ldots, \beta_k\} \in \mathcal{S}_{\mathfrak{g}}$.*

(1) Then for any $f \in \mathcal{J}_{\mathfrak{g}}$,

$$ds_x(f) = f|_{\mathfrak{h}_x}.$$

(2) If $B = \{\varepsilon_1 - \delta_1\}$, then $ds_x(f)$ is given by substituting $x_1 = y_1$ into f, that is,

$$ds_x f = f|_{x_1=y_1}.$$

If $\mathfrak{g} = F(4)$ or $D(2, 1, \alpha)$ and $B = \{\frac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 - \delta_1)\}$ or $B = \{\varepsilon_1 - \varepsilon_2 - \varepsilon_3\}$, then $ds_x f$ is given by substituting $y_1 = x_1 x_2 x_3$ or $x_1 = x_2 x_3$ into f, respectively.

(3) If $B = \{\beta_i = a_i \varepsilon_{r_i} - b_i \delta_{s_i}\}_{i=1,\ldots,k}$ for some $a_i, b_i \in \{\pm 1\}$, then $ds_x f$ is given by substituting $x_{r_i}^{a_i} = y_{s_i}^{b_i}$ into f, that is,

$$ds_x f = f|_{x_{r_i}^{a_i} = y_{s_i}^{b_i}, \, i=1,\ldots,k}.$$

(4) For any $f \in \mathcal{J}_{\mathfrak{g}}$,

$$ds_x(f) = f|_{\beta_1 = \cdots = \beta_k = 0}.$$

Proof. It suffices to prove (1) for a spanning set of $\mathcal{J}_{\mathfrak{g}}$. Suppose $[M] \in \mathcal{K}_{\mathfrak{g}}$ corresponds to a module $M \in \mathcal{F}_{\mathfrak{g}}$. By Lemma 4, we have

$$ds_x([M]) = [DS_x(M)] = \operatorname{sch} M_x = (\operatorname{sch} M)|_{\mathfrak{h}_x} = [M]|_{\mathfrak{h}_x} \in \mathcal{J}_{\mathfrak{g}_x}.$$

Hence, $ds_x(f) = f|_{\mathfrak{h}_x}$ for any $f \in \mathcal{J}_{\mathfrak{g}}$.

To prove (2), fix $f \in \mathcal{J}_{\mathfrak{g}}$, and suppose that $x \in \mathfrak{g}_\beta$. If $\beta = \varepsilon_1 - \delta_1$, then the evaluation $f_{x_1=y_1=t}$ is well defined and independent of t due to the supersymmetry property of $f \in \mathcal{J}_{\mathfrak{g}}$. Thus,

$$f|_{x_1=y_1=t} := f(t, x_2, \ldots, x_m \mid t, y_2, \ldots, y_n)$$

is equal to the restriction of f to the hyperplane $x_1 - y_1 = 0$. Since $\mathfrak{h}_x \subset \mathfrak{h}^x = \{h \in \mathfrak{h} \mid \beta(h) = 0\}$ belongs to the hyperplane $x_1 - y_1 = 0$, we have proven that $ds_x f = f|_{\mathfrak{h}_x} = f|_{x_1=y_1}$. The cases $\beta = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 - \delta_1$ and $\beta = \varepsilon_1 - \varepsilon_2 - \varepsilon_3$ are similar.

Now (3) can be proven using arguments similar to that of (2) and the fact that

$$\mathfrak{h}^x = \{h \in \mathfrak{h} \mid \beta(h) = 0 \text{ for all } \beta \in B\}.$$

Finally, (4) follows from (3) since if $\beta_i = a_i \varepsilon_{r_i} - b_i \delta_{r_i}$, then $\beta_i = 0$ if and only if $x_{r_i}^{a_i} y_{r_i}^{-b_i} = e^{\beta_i} = 1$ if and only if $x_{r_i}^{a_i} = y_{r_i}^{b_i}$. $\qquad\qquad\square$

Corollary 11. *If $x = x_{\beta_1} + \cdots + x_{\beta_k}$ where $x_{\beta_i} \in \mathfrak{g}_{\beta_i}$ and $B = \{\beta_1, \ldots, \beta_k\} \in \mathscr{S}$, then for all $f \in \mathcal{J}_{\mathfrak{g}}$*

$$ds_x(f) = ds_{x_{\beta_1}} \circ \cdots \circ ds_{x_{\beta_k}}(f).$$

4. The kernel of the ring homomorphism

In this section, we give a $\mathbb{Z}$-basis for the kernel of ds_x when $x \in \mathfrak{g}_\beta$ is a root vector of an isotropic root β for the Lie superalgebra $\mathfrak{g}$. Our basis is given by elements of the following form.

Definition 12. For each $\lambda \in P_{\bar{0}}$, we define

$$k(\lambda) := R^{-1} \cdot \sum_{w \in W} (-1)^{l(w)+p(w(\rho)-\rho)} e^{w(\lambda+\rho)-\rho}.$$

Here $p(w(\rho) - \rho)$ denotes the parity of $w(\rho) - \rho$, which is well defined since $w(\rho) - \rho \in Q$. Note that the element $w(\rho) - \rho$ may be odd, e.g., in $\mathfrak{osp}(1 \mid 2)$.

For each $\lambda \in P_{\bar{0}}^+$, the expression $k(\lambda)$ is in $\mathbb{Z}[P_{\bar{0}}]^W$ since it is the product of the W-invariant polynomial $e^{\rho_{\bar{1}}} R_{\bar{1}}$ and the character of a finite-dimensional $\mathfrak{g}_{\bar{0}}$-module given by the Weyl character formula. Moreover, since the evaluation $k(\lambda)|_{\beta=0}$ equals zero for any $\beta \in \Delta_{\text{iso}}$, we have that $k(\lambda) \in \mathcal{J}_{\mathfrak{g}}$. It is clear that $k(\lambda)$ is in the kernel of ds_x for any $x \in \mathfrak{g}_\beta$, since $ds_x(R_{\bar{1}}) = 0$.

For Lie superalgebras of type I with the distinguished choice of simple roots, $k(\lambda)$ is the supercharacter of a Kac module when $\lambda \in P_{\bar{0}}^+$, whereas in type II, $k(\lambda)$ is a virtual supercharacter. Similar virtual characters were used by Gruson and Serganova [2010] to study the character formula of simple modules over orthosymplectic Lie superalgebras.

We need the following definition to prove the main result in this section for Lie superalgebras of type II.

Definition 13. Given a finite-dimensional Lie superalgebra $\mathfrak{g}$ with root system Δ, we define a Lie algebra $\tilde{\mathfrak{g}}$ as follows. We let $\tilde{\Delta}$ be the root system with positive even roots given by

$$\tilde{\Delta}^+ := \left\{ \alpha \in \Delta_{\bar{0}}^+ \;\middle|\; \frac{\alpha}{2} \notin \Delta_{\bar{1}} \right\} \cup \{\alpha \in \Delta_{\bar{1}}^+ \mid \alpha \notin \Delta_{\text{iso}}\},$$

and we let $\tilde{\mathfrak{g}}$ be the semisimple Lie algebra with root system $\tilde{\Delta}$. If $\Delta_{\bar{1}} = \Delta_{\text{iso}}$, then $\tilde{\mathfrak{g}} = \mathfrak{g}_{\bar{0}}$. If $\mathfrak{g} = B(m \mid n)$, $G(3)$, then $\tilde{\mathfrak{g}} \cong B_m \times B_n$, $G_2 \times A_1$, respectively. We set $\tilde{\rho} := \frac{1}{2} \sum_{\alpha \in \tilde{\Delta}^+} \alpha$. Note that $\rho = \tilde{\rho} - \rho_{\text{iso}}$, since $\beta \in \Delta_{\text{iso}}$ if and only if $\beta \in \Delta_{\bar{1}}$ but $2\beta \notin \Delta_{\bar{0}}$. Let $P_{\tilde{\mathfrak{g}}}^+$ denote the set of dominant integral weights of $\tilde{\mathfrak{g}}$. Then $P_{\bar{0}}^+ \subset P_{\tilde{\mathfrak{g}}}^+$ and the Weyl group of $\tilde{\mathfrak{g}}$ is isomorphic to the Weyl group of $\mathfrak{g}_{\bar{0}}$. We extend the definition of $k(\lambda)$ to $\lambda \in P_{\tilde{\mathfrak{g}}}$ by letting

$$k(\lambda) := R^{-1} \cdot \sum_{w \in W} (-1)^{l(w)+p(w(\lambda+\rho)-\rho)} e^{w(\lambda+\rho)-\rho}.$$

We have the following lemma.

Lemma 14. *The set $\{k(\mu) \mid \mu \in P_{\tilde{\mathfrak{g}}}^+ + \rho_{\mathrm{iso}}\}$ is linearly independent.*

Proof. To prove linear independence we consider a completion of $\mathbb{Z}[P_{\tilde{\mathfrak{g}}}]$, where we allow expansions in the domain $|e^{-\alpha}| < 1$ for $\alpha \in \tilde{\Delta}^+$. Note that in this completion, $R^{-1} = \sum_{\nu \in -Q_{\tilde{\mathfrak{g}}}^+} b_\nu e^\nu$ for some $b_\nu \in \mathbb{Z}$. For each $\mu \in P_{\tilde{\mathfrak{g}}}^+ + \rho_{\mathrm{iso}}$, we will show that $\mu + \rho$ is a strictly dominant element of $P_{\tilde{\mathfrak{g}}}$, that is, $w(\mu + \rho) < \mu + \rho$ for $w \in W$, $w \neq 1$. Indeed, if $\mu \in P_{\tilde{\mathfrak{g}}}^+ + \rho_{\mathrm{iso}}$, then $\mu + \rho = \lambda + \tilde{\rho}$ for some $\lambda \in P_{\tilde{\mathfrak{g}}}^+$. Since $\lambda + \tilde{\rho}$ is strictly dominant with respect to $\tilde{\mathfrak{g}}$, it is also strictly dominant with respect to $\mathfrak{g}_{\bar{0}}$ and the claim follows. Thus,

$$k(\mu) = e^\mu + \sum_{\nu \in \mu - Q_{\tilde{\mathfrak{g}}}^+} a_\nu e^\nu$$

and linear independence follows. $\qquad\square$

Remark 15. Note that if one takes the distinguished choice of simple roots for $\mathfrak{gl}(m, n)$, then $P^+ = P^+ + \rho_{\mathrm{iso}}$, since in this case $(\rho_{\mathrm{iso}}, \alpha) = 0$ for every even root α.

The following lemma is used in the proof of the main theorem of this section.

Lemma 16. *For each $\mu \in P_{\tilde{\mathfrak{g}}}^+$, we have $k(\mu + \rho_{\mathrm{iso}}) = e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 - e^{-\alpha}) \cdot \mathrm{ch}\, L_{\tilde{\mathfrak{g}}}(\mu)$.*

Proof. For any element $g \in \mathbb{Z}[P_{\tilde{\mathfrak{g}}}]$ with $\mathrm{Supp}\, g \subset \mu + Q_{\tilde{\mathfrak{g}}}$, we write $g = \sum_{\lambda \in Q_{\tilde{\mathfrak{g}}}} c_{\mu+\lambda} e^{\mu+\lambda}$, and we define $\bar{g} = \sum_{\lambda \in Q_{\tilde{\mathfrak{g}}}} (-1)^{p(\lambda)} c_{\mu+\lambda} e^{\mu+\lambda}$, where $p : Q_{\tilde{\mathfrak{g}}} \to \mathbb{Z}_2$ is the parity function. Clearly, this operation is an involution. So we have that

$$\overline{e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 - e^{-\alpha}) \cdot \mathrm{ch}\, L_{\tilde{\mathfrak{g}}}(\mu)} = (-1)^{p(\rho_{\mathrm{iso}})} e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 + e^{-\alpha}) \cdot \mathrm{sch}\, L_{\tilde{\mathfrak{g}}}(\mu)$$

$$= (-1)^{p(\rho_{\mathrm{iso}})} e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 + e^{-\alpha}) \frac{\sum_{w \in W} (-1)^{l(w) + p(w(\mu+\tilde{\rho}) - \tilde{\rho})} e^{w(\mu+\tilde{\rho}) - \tilde{\rho}}}{\prod_{\alpha \in \tilde{\Delta}_{\bar{0}}^+} (1 - e^{-\alpha})}$$

$$= \prod_{\alpha \in \Delta_{\bar{1}}^+} (1 + e^{-\alpha}) \cdot \frac{\sum_{w \in W} (-1)^{l(w) + p(w(\mu+\rho_{\mathrm{iso}}+\rho) - \rho)} e^{w((\mu+\rho_{\mathrm{iso}})+\tilde{\rho} - \rho_{\mathrm{iso}}) - \tilde{\rho} + \rho_{\mathrm{iso}}}}{\prod_{\alpha \in \tilde{\Delta}_{\bar{0}}^+} (1 - e^{-\alpha}) \prod_{\alpha \in \Delta_{\bar{1}}^+ \setminus \Delta_{\mathrm{iso}}^+} (1 + e^{-\alpha})}$$

$$= \prod_{\alpha \in \Delta_{\bar{1}}^+} (1 + e^{-\alpha}) \frac{\sum_{w \in W} (-1)^{l(w) + p(w((\mu+\rho_{\mathrm{iso}})+\rho) - \rho)} e^{w((\mu+\rho_{\mathrm{iso}})+\rho) - \rho}}{\prod_{\alpha \in \Delta_{\bar{0}}^+} (1 - e^{-\alpha})}$$

$$= \overline{k(\mu + \rho_{\mathrm{iso}})},$$

and hence, the claim follows. $\qquad\square$

We have the following theorem.

Theorem 17. *If β is an odd isotropic root and $x \in \mathfrak{g}_\beta$, then the set*

$$\{k(\lambda) \mid \lambda \in P_{\bar{0}}^+ + \rho_{\mathrm{iso}}\} \tag{4-1}$$

is a $\mathbb{Z}$-basis for the kernel of $ds_x : \mathcal{J}_{\mathfrak{g}} \to \mathcal{J}_{\mathfrak{g}_x}$.

Proof. Linear independence of the set (4-1) follows from Lemma 14 since $P_{\bar{0}}^{+} \subset P_{\tilde{\mathfrak{g}}}^{+}$. So it only remains to show that the set (4-1) spans the kernel of $ds_x : \mathcal{J}_{\mathfrak{g}} \to \mathcal{J}_{\mathfrak{g}_x}$.

Let $f \in \mathcal{J}_{\mathfrak{g}}$ such that $ds_x(f) = 0$. According to Theorem 10, this means that the restriction of f to the hyperplane $\beta = 0$ is zero, or equivalently, substituting $e^{-\beta} = 1$ yields zero. Hence, f is divisible by $(1 - e^{-\beta})$. Since f is W-invariant and $W\beta = \Delta_{\text{iso}}$, it follows that f is divisible by $e^{\rho_{\text{iso}}} \prod_{\alpha \in \Delta_{\text{iso}}^{+}} (1 - e^{-\alpha})$.

Write

$$f = e^{\rho_{\text{iso}}} \prod_{\alpha \in \Delta_{\text{iso}}^{+}} (1 - e^{-\alpha}) \cdot g.$$

Then g is a W-invariant element of $\mathbb{Z}[P_{\bar{0}}]$, since both f and $e^{\rho_{\text{iso}}} \prod_{\alpha \in \Delta_{\text{iso}}^{+}} (1 - e^{-\alpha})$ are W-invariant.

Case 1. First suppose that $\mathfrak{g}$ does not have nonisotropic roots; then $\Delta_{\text{iso}}^{+} = \Delta_{\bar{1}}^{+}$ and $\rho_{\text{iso}} = \rho_{\bar{1}}$. By the theory of symmetric functions,

$$g = \sum_{\mu \in P_{\bar{0}}^{+}}^{\text{finite}} a_\mu \operatorname{ch} L_{\mathfrak{g}_{\bar{0}}}(\mu),$$

for some $a_\mu \in \mathbb{Z}$, where $P_{\bar{0}}^{+}$ is the set of highest weights of finite-dimensional $\mathfrak{g}_{\bar{0}}$-modules (see for example [Macdonald 1995]).

By the Weyl character formula for semisimple Lie algebras, we have that

$$\begin{aligned}
f &= e^{\rho_{\bar{1}}} R_{\bar{1}} \cdot g \\
&= e^{\rho_{\bar{1}}} R_{\bar{1}} \sum_{\mu \in P_{\bar{0}}^{+}} a_\mu \operatorname{ch} L_{\mathfrak{g}_{\bar{0}}}(\mu) \\
&= e^{\rho_{\bar{1}}} R_{\bar{1}} \sum_{\mu \in P_{\bar{0}}^{+}} a_\mu \frac{\sum_{w \in W} (-1)^{l(w)} e^{w(\mu + \rho_0)}}{e^{\rho_{\bar{0}}} R_{\bar{0}}} \\
&= e^{\rho_{\bar{1}}} R_{\bar{1}} \sum_{\lambda \in P_{\bar{0}}^{+} + \rho_{\bar{1}}} b_\lambda \frac{\sum_{w \in W} (-1)^{l(w)} e^{w(\lambda + \rho_0 - \rho_1)}}{e^{\rho_{\bar{0}}} R_{\bar{0}}} \\
&= \sum_{\lambda \in P_{\bar{0}}^{+} + \rho_{\bar{1}}} b_\lambda k(\lambda),
\end{aligned}$$

where $b_\lambda := a_{\lambda - \rho_{\bar{1}}}$. For each $w \in W$, the parity of $w(\rho)$ equals the parity of ρ, since $\rho \in P_{\bar{0}}$. Hence, the last equality follows.

Case 2. Suppose that $\mathfrak{g}$ has nonisotropic roots. Since $P_{\bar{0}} \subset P_{\tilde{\mathfrak{g}}}$, by the theory of characters of Lie algebras

$$g = \sum_{\mu \in P_{\tilde{\mathfrak{g}}}^{+}}^{\text{finite}} a_\mu \operatorname{ch} L_{\tilde{\mathfrak{g}}}(\mu)$$

for some $a_\mu \in \mathbb{Z}$. By Lemma 16, we have that

$$
\begin{aligned}
f &= e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 - e^{-\alpha}) \cdot g \\
&= e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 - e^{-\alpha}) \sum_{\mu \in P_{\tilde{\mathfrak{g}}}^+} a_\mu \operatorname{ch} L_{\tilde{\mathfrak{g}}}(\mu) \\
&= \sum_{\mu \in P_{\tilde{\mathfrak{g}}}^+} a_\mu \cdot e^{\rho_{\mathrm{iso}}} \prod_{\alpha \in \Delta_{\mathrm{iso}}^+} (1 - e^{-\alpha}) \cdot \operatorname{ch} L_{\tilde{\mathfrak{g}}}(\mu) \\
&= \sum_{\mu \in P_{\tilde{\mathfrak{g}}}^+} a_\mu \cdot k(\mu + \rho_{\mathrm{iso}}) \\
&= \sum_{\lambda \in P_{\tilde{\mathfrak{g}}}^+ + \rho_{\mathrm{iso}}} b_\lambda k(\lambda) \quad (4\text{-}2)
\end{aligned}
$$

where $b_\lambda := a_{\lambda - \rho_{\mathrm{iso}}}$. We are left to show that $b_\lambda = 0$ for $\lambda \notin P_{\bar{0}}^+ + \rho_{\mathrm{iso}}$. Since $\operatorname{Supp} f \subset P_{\bar{0}}$, $\operatorname{Supp} k(\lambda) \subset P_{\bar{0}}$, the elements $k(\lambda)$ for $\mu \in P_{\tilde{\mathfrak{g}}}^+ + \rho_{\mathrm{iso}}$ are linearly independent, and the sum in (4-2) is finite, we conclude that

$$
f = \sum_{\lambda \in P_{\bar{0}}^+ + \rho_{\mathrm{iso}}} b_\lambda k(\lambda). \qquad \square
$$

Corollary 18. *Let G be one of the Lie supergroups $SL(m \mid n)$, $m \neq n$, $GL(m \mid n)$, or $SOSP(m \mid 2n)$, and let $\mathfrak{g}$ be the corresponding Lie superalgebra. Let β be an odd isotropic root and $x \in \mathfrak{g}_\beta$, and let $DS_x : \mathscr{F}_G \to \mathscr{F}_{G_x}$ be the Duflo–Serganova functor from the category $\mathscr{F}_G$ of finite-dimensional G-modules to the category $\mathscr{F}_{G_x}$ of finite-dimensional G_x-modules, where G_x denotes the Lie supergroup corresponding to the Lie superalgebra $\mathfrak{g}_x$. Then the kernel of the induced ring homomorphism $ds_x : \mathscr{J}_G \to \mathscr{J}_{G_x}$ has a $\mathbb{Z}$-basis*

$$
\{k(\lambda) \mid \lambda \in P_G^+ + \rho_{\mathrm{iso}}\},
$$

where P_G^+ is the set of highest weights for finite-dimensional G-modules.

Proof. Let $P_G \subset P_{\bar{0}}$ be the sublattice of integral weights of finite-dimensional $G_{\bar{0}}$-modules. Then for $G = GL(m \mid n)$ or $SOSP(m \mid 2n)$

$$
P_G = \left\{ \sum_{i=1}^m \lambda_i \varepsilon_i + \sum_{j=1}^n \mu_j \delta_j \;\middle|\; \lambda_i, \mu_j \in \mathbb{Z} \right\},
$$

and the supercharacter ring for the category of finite-dimensional G-modules $\mathscr{F}_G$ is

$$
\mathscr{J}_G = \left\{ f \in \mathbb{Z}[x_1^{\pm 1}, \ldots, x_m^{\pm 1}, y_1^{\pm 1}, \ldots, y_n^{\pm 1}]^W \;\middle|\; y_j \frac{\partial f}{\partial y_j} + x_i \frac{\partial f}{\partial x_i} \in \langle y_j - x_i \rangle \right\}
$$

as shown in [Sergeev and Veselov 2011, §7] (note that this ring is therein denoted by $J(\mathfrak{g})_0$). If $G = SL(m \mid n)$, $m \neq n$, then

$$P_G = \left\{ \sum_{i=1}^{m} \lambda_i \varepsilon_i + \sum_{j=1}^{n} \mu_j \delta_j \ \Big| \ \lambda_i, \mu_j \in \mathbb{Z}, \ \sum_{i=1}^{m} \lambda_i - \sum_{j=1}^{n} \mu_j = 0 \right\},$$

and the supercharacter ring for the category of finite-dimensional G-modules $\mathscr{F}_G$ is

$$\mathscr{J}_G = \left\{ f \in \mathbb{Z}[x_1^{\pm 1}, \ldots, x_m^{\pm 1}, y_1^{\pm 1}, \ldots, y_n^{\pm 1}]^W / \langle x_1 \cdots x_m - y_1 \cdots y_n \rangle \ \Big| \ y_j \frac{\partial f}{\partial y_j} + x_i \frac{\partial f}{\partial x_i} \in \langle y_j - x_i \rangle \right\}$$

as shown in [Sergeev and Veselov 2011, §7]. Since in both cases $\mathscr{J}_G = \mathscr{J}_\mathfrak{g} \cap \mathbb{Z}[P_G]$, the kernel of the homomorphism $ds_x : \mathscr{J}_G \to \mathscr{J}_{G_x}$ equals $\mathrm{Ker}_G \, ds_x = \mathrm{Ker}_\mathfrak{g} \, ds_x \cap \mathbb{Z}[P_G]$, where $\mathrm{Ker}_\mathfrak{g} \, ds_x$ is the kernel of the corresponding homomorphism $ds_x : \mathscr{J}_\mathfrak{g} \to \mathscr{J}_{\mathfrak{g}_x}$. It follows from the linear independence of the elements $k(\lambda)$ and the fact that $\lambda \in P_G$ if and only if $\mathrm{Supp} \, k(\lambda) \in P_G$ that $\mathrm{Ker}_G \, ds_x = \mathrm{span}\{k(\lambda) \mid \lambda \in P_G + \rho_{\mathrm{iso}}\}$. Since $P_G^+ = P_\mathfrak{g}^+ \cap P_G$, the claim follows. $\square$

Remark 19. On the level of categories, it was shown in [Boe et al. 2012] that a module M over a type-I finite-dimensional contragredient Lie superalgebra has a filtration of Kac modules or dual Kac modules if and only if $DS_x(M) = 0$ for all $x \in X_\mathfrak{g}^-$ or $x \in X_\mathfrak{g}^+$, respectively, where $X_\mathfrak{g}^\pm = X_\mathfrak{g} \cap \mathfrak{n}^\pm$ and $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$ is the triangular decomposition with respect to the distinguished choice of simple roots.

5. The image of the ring homomorphism

5A. *Image of ds_x for classical Lie superalgebras.* Let $\mathfrak{g}$ be one of the Lie superalgebras: $\mathfrak{sl}(m \mid n)$, $m \neq n$, $\mathfrak{gl}(m \mid n)$, and $\mathfrak{osp}(m \mid 2n)$. In this section, we describe the image of ds_x for every $x \in X_\mathfrak{g}$. We use the realization of ds_x given in Theorem 10 and the explicit description of the supercharacter rings given by Sergeev and Veselov [2011, §7].

Theorem 20. *Let G be one of the Lie supergroups $SL(m \mid n)$, $m \neq n$, $GL(m \mid n)$, or $OSP(m, 2n)$ and $\mathfrak{g}$ be the corresponding Lie superalgebra. For any $x \in X_\mathfrak{g}$, the Duflo–Serganova functor $DS_x : \mathscr{F}_G \to \mathscr{F}_{G_x}$ from the category $\mathscr{F}_G$ of finite-dimensional G-modules to the category $\mathscr{F}_{G_x}$ of finite-dimensional G_x-modules induces a surjective ring homomorphism on the corresponding supercharacter rings $ds_x : \mathscr{J}_G \to \mathscr{J}_{G_x}$.*

Proof. We will use Corollary 11 to reduce to the case that $x \in \mathfrak{g}_\beta$ for some isotropic root β. Using the realization of ds_x given in Theorem 10, we will show that ds_x transfers a certain set of generators of the supercharacter ring $\mathscr{K}_G$ to a set of generators of the supercharacter ring $\mathscr{J}_{G_x}$. We use the same set of generators of $\mathscr{J}_G$ that Sergeev and Veselov [2011, §7] used to give explicit descriptions of supercharacter rings over basic Lie superalgebras and their corresponding Lie supergroups.

$GL(m, n)$: The supercharacter ring of $GL(m, n)$ is generated by $(x_1 \cdots x_m)/(y_1 \cdots y_n)$, $(y_1 \cdots y_n)/(x_1 \cdots x_m)$, $h_k(x_1, \ldots, x_m, y_1, \ldots, y_n)$, and $h_k(x_1^{-1}, \ldots, x_m^{-1}, y_1^{-1}, \ldots, y_n^{-1})$, $k \in \mathbb{Z}_{>0}$, where

$$\chi_G(t) = \frac{\prod_{i=1}^{m}(1 - x_i t)}{\prod_{j=1}^{n}(1 - y_j t)} = \sum_{k=0}^{\infty} h_k(x_1, \ldots, x_m, y_1, \ldots, y_n) t^k. \tag{5-1}$$

$SL(m, n)$, $m \neq n$: The supercharacter ring of $SL(m, n)$, $m \neq n$, is generated by $h_k(x_1, \ldots, x_m, y_1, \ldots, y_n)$ and $h_k(x_1^{-1}, \ldots, x_m^{-1}, y_1^{-1}, \ldots, y_n^{-1})$, $k \in \mathbb{Z}_{>0}$, where h_k is given by (5-1).

$OSP(2m+1, 2n)$: The supercharacter ring of $OSP(2m+1, 2n)$ is generated by $h_k(x_1, \ldots, x_m, y_1, \ldots, y_n)$, $k \in \mathbb{Z}_{>0}$, where

$$\chi_G(t) = \frac{\prod_{j=1}^{n}(1 - y_j t)(1 - y_j^{-1} t)}{(1 - t)\prod_{i=1}^{m}(1 - x_i t)(1 - x_i^{-1} t)} = \sum_{k=0}^{\infty} h_k(x_1, \ldots, x_m, y_1, \ldots, y_n)t^k.$$

$OSP(2, 2n)$: The supercharacter ring of $OSP(2, 2n)$ is generated by $h_k(x_1, y_1, \ldots, y_n)$, $k \in \mathbb{Z}_{>0}$, where

$$\chi_G(t) = \frac{\prod_{i=1}^{m}(1 - y_j t)(1 - y_j^{-1} t)}{(1 - x_1 t)(1 - x_1^{-1} t)} = \sum_{k=0}^{\infty} h_k(x_1, y_1, \ldots, y_n)t^k.$$

$OSP(2m, 2n)$, $m \geq 2$: The supercharacter ring of $OSP(2m, 2n)$ is generated by $h_k(x_1, \ldots, x_m, y_1, \ldots, y_n)$, $k \in \mathbb{Z}_{>0}$, where

$$\chi_G(t) = \frac{\prod_{p=1}^{n}(1 - y_j t)(1 - y_j^{-1} t)}{\prod_{i=1}^{m}(1 - x_i t)(1 - x_i^{-1} t)} = \sum_{k=0}^{\infty} h_k(x_1, \ldots, x_m, y_1, \ldots, y_n)t^k.$$

By Theorem 10, $ds_x(h_k^{\mathfrak{g}}) = (h_k^{\mathfrak{g}})|_{\beta=0}$. Since χ_G is W-invariant and $W\beta = \Delta_{\text{iso}}$ for any $\beta \in \Delta_{\text{iso}}$, it suffices to consider the case that $\beta = \varepsilon_1 - \delta_1$. In this case, $\beta = 0$ if and only if $x_1 = y_1$. It is not difficult to check that $\chi_G(t)|_{x_1=y_1} = \chi_{G_x}$, and hence, $ds_x(h_k^{\mathfrak{g}}) = h_k^{\mathfrak{g}_x}$. Thus, all the generators of $\mathcal{J}_{G_x}$ are in the image of ds_x.

The general case for arbitrary $x \in X_{\mathfrak{g}}$ now follows from Corollary 11, since the composition of surjective maps is surjective. □

Proposition 21. *Let* $\mathfrak{g} = \mathfrak{sl}(m \mid n)$, $m \neq n$, *or* $\mathfrak{g} = \mathfrak{osp}(m \mid 2n)$. *Then for any* $x \in X_{\mathfrak{g}}$, *the image of* $ds_x : \mathcal{J}_{\mathfrak{g}} \to \mathcal{J}_{\mathfrak{g}_x}$ *is the supercharacter ring* $\mathcal{J}_{G_x}$ *of the Lie supergroup* G_x.

Proof. We use Theorems 20 and 10, together with the description of the rings $\mathcal{J}_{\mathfrak{g}}$ given by Sergeev and Veselov [2011] to prove that the image of the map $ds_x : \mathcal{J}_{\mathfrak{g}} \to \mathcal{J}_{\mathfrak{g}_x}$ equals $\mathcal{J}_{G_x}$ in the case that $x \in \mathfrak{g}_\beta$ is an isotropic root β. The claim for any element $x \in X_{\mathfrak{g}}$ then follows Corollary 11.

The supercharacter ring of $\mathfrak{g} = \mathfrak{sl}(m \mid n)$, $m \neq n$, is $\mathcal{J}_{\mathfrak{g}} = \mathcal{J}_G \oplus \bigoplus_{a \in \mathbb{C}/\mathbb{Z}} J(\mathfrak{g})_a$, where

$$J(\mathfrak{g})_a = (x_1 \cdots x_n)^a \prod_{i,j}(1 - x_i y_j^{-1})\mathbb{Z}[x^{\pm 1}, y^{\pm 1}]_0^{S_m \times S_n},$$

and $\mathbb{Z}[x^{\pm 1}, y^{\pm 1}]_0^{S_m \times S_n}$ is the quotient of the ring $\mathbb{Z}[x_1^{\pm 1}, \ldots, x_m^{\pm 1}, y_1^{\pm 1}, \ldots, y_n^{\pm 1}]^{S_m \times S_n}$ by ideal $\langle x_1 \cdots x_m - y_1 \cdots y_n \rangle$. Clearly, $f|_{\beta=0} = f|_{x_i=y_j} = 0$ for any $f \in J(\mathfrak{g})_a$. Hence, $ds_x(f) = 0$ for any $x \in X_{\mathfrak{g}}$ and $f \in J(\mathfrak{g})_a$.

If $\mathfrak{g} = B(m \mid n)$, $C(n + 1)$, or $D(m \mid n)$, then $\mathcal{J}_{\mathfrak{g}} = \mathcal{J}_G \oplus \tilde{\mathcal{J}}$ and $ds_x(f) = 0$ for all $f \in \tilde{\mathcal{J}}$. Indeed, for $\beta = \pm\varepsilon_i \pm \delta_j$ it is not difficult to check that $f|_{\beta=0} = f|_{x_i^{\pm 1} = y_j^{\pm 1}} = f|_{u_i = v_j} = 0$.

The supercharacter ring of $\mathfrak{g} = B(m \mid n)$ is $\mathcal{J}_{\mathfrak{g}} = \mathcal{J}_G \oplus J_{\mathfrak{g},1/2}$, where

$$J_{\mathfrak{g},1/2} = \left\{ \prod_{i=1}^{m}(x_i^{1/2} + x_i^{-1/2}) \prod_{i,j}(u_i - v_j)g \mid g \in \mathbb{Z}[u_1, \ldots, u_m, v_1, \ldots, v_n]^{S_m \times S_n} \right\}.$$

The supercharacter ring of $\mathfrak{g} = C(n+1)$ is $\mathcal{J}_{\mathfrak{g}} = \mathcal{J}_G \oplus \left(J(\mathfrak{g})_0^- \oplus \bigoplus_{a \in \mathbb{C}/\mathbb{Z}} J(\mathfrak{g})_a \right)$, where

$$J(\mathfrak{g})_0^- = \left\{ x_1 \prod_{j=1}^{n}(u_1 - v_j)g \mid g \in \mathbb{Z}[u_1, v_1, \ldots, v_n]^{S_n} \right\},$$

$$J(\mathfrak{g})_a = x_1^a \prod_{j=1}^{n}(1 - x_1 y_j)(1 - x_1 y_j^{-1})\mathbb{Z}[x_1^{\pm 1}, y_1^{\pm 1}, \ldots, y_n^{\pm 1}]^W.$$

The supercharacter ring of $\mathfrak{g} = D(m \mid n)$ is $\mathcal{J}_{\mathfrak{g}} = \mathcal{J}_G \oplus \left(J(\mathfrak{g})_0^- \oplus J_{\mathfrak{g},1/2} \right)$, where

$$J(\mathfrak{g})_0^- = \left\{ \omega \prod_{i,j}(u_i - v_j)g \mid g \in \mathbb{Z}[u_1, \ldots, u_m, v_1, \ldots, v_n]^{S_m \times S_n} \right\},$$

$$J_{\mathfrak{g},1/2} = \prod_{i,j}(u_i - v_j)((x_1 \ldots x_m)^{1/2}\mathbb{Z}[u_1, \ldots, u_m, v_1, \ldots, v_n])^W. \qquad \square$$

Proposition 22. *Let $\mathfrak{g} = \mathfrak{gl}(m \mid n)$ and $x \in X_{\mathfrak{g}}$. The image of $ds_x : \mathcal{J}_{\mathfrak{g}} \to \mathcal{J}_{\mathfrak{g}_x}$ is*

$$\bigoplus_{a \in \mathbb{C}/\mathbb{Z}} (x_1 \cdots x_{m-k})^a (y_1 \cdots y_{n-k})^{-a} \mathcal{J}_{G_x},$$

where k is the size of $\psi(x) \in S_{\mathfrak{g}}$ under the bijection $\psi : X_{\mathfrak{g}}/G_{\bar{0}} \to S_{\mathfrak{g}}/W$, and $\mathcal{J}_{G_x}$ is the supercharacter ring of the Lie supergroup G_x.

Proof. By Sergeev and Veselov [2011], the supercharacter ring of $\mathfrak{gl}(m \mid n)$ is $\mathcal{J}_{\mathfrak{g}} = \bigoplus_{a,b \in \mathbb{C}/\mathbb{Z}} J(\mathfrak{g})_{a,b}$ where $J(\mathfrak{g})_{0,0} = \mathcal{J}_G$,

$$J(\mathfrak{g})_{a,b} = (x_1 \cdots x_m)^a (y_1 \cdots y_n)^{-a} J(\mathfrak{g})_{0,0}$$

when $a + b \in \mathbb{Z}$, but $a \notin \mathbb{Z}$, and

$$J(\mathfrak{g})_{a,b} = (x_1 \cdots x_m)^a (y_1 \cdots y_n)^b \prod_{i,j}(1 - x_i y_j^{-1})\mathbb{Z}[x_1^{\pm 1}, \ldots, x_m^{\pm 1}, y_1^{\pm 1}, \ldots, y_n^{\pm 1}]^{S_m \times S_n}$$

when $a + b \notin \mathbb{Z}$.

Then we have that $f|_{x_i = y_j} = 0$ for any $f \in J(\mathfrak{g})_{a,b}$ with $a + b \notin \mathbb{Z}$. By Theorem 20, $ds_x(J(\mathfrak{g})_{0,0}) = J(\mathfrak{g}_x)_{0,0}$. Since $ds_x(f) = f|_{x_{r_i} = y_{s_i}, \, i=1,\ldots,k}$ by Theorem 10, we have that

$$ds_x(J(\mathfrak{g})_{a,b}) = (x_1 \cdots x_{m-k})^a (y_1 \cdots y_{n-k})^{-a} J(\mathfrak{g}_x)_{0,0} = J(\mathfrak{g}_x)_{a,b}$$

when $a + b \in \mathbb{Z}$, but $a \notin \mathbb{Z}$. $\qquad \square$

5B. *The image of ds_x for the exceptional Lie superalgebras.* In this section, we describe the image of ds_x for the Lie superalgebras $G(3)$, $F(4)$, and $D(2, 1, \alpha)$, using the explicit description of the super-character rings given by Sergeev and Veselov [2011, §7].

Since $G(3)$, $F(4)$, and $D(2, 1, \alpha)$ have defect 1, we may assume that $x \in \mathfrak{g}_\beta$ for some isotropic root β. Moreover, since $W\beta = \Delta_{\bar{1}}$, it suffices to describe the image for a fixed choice of β.

5B1. $G(3)$. Let $\beta = \varepsilon_3 + \delta_1$. Then $\mathfrak{g}_x \cong \mathfrak{sl}(2)$ with $\Delta_x = \{\pm(\varepsilon_1 - \varepsilon_2)\}$. The supercharacter ring of $G(3)$ is

$$\mathcal{J}_\mathfrak{g} = \{g(w) + (v_1 - u_1)(v_1 - u_2)(v_1 - u_3)h \mid h \in \mathbb{Z}[u_1, u_2, u_3, v_1]^{S_3}, \ g \in \mathbb{Z}[w]\},$$

where $y_1 = e^{\delta_1}$, $v_1 = y_1 + y_1^{-1}$, $x_i = e^{\varepsilon_i}$, $u_i = x_i + x_i^{-1}$ for $i = 1, 2, 3$, and

$$w = v_1^2 - v_1(u_1 + u_2 + u_3 + 1) + u_1u_2 + u_1u_3 + u_2u_3.$$

Note that $x_1x_2x_3 = 1$, so $u_3 = x_1x_2 + x_1^{-1}x_2^{-1}$.

Theorem 10 implies that $ds_x(f) = f|_{y_1 = x_3^{-1} = x_1x_2}$ for every $f \in \mathcal{J}_\mathfrak{g}$. Hence, $ds_x(f) = ds_x(g(w))$ since $(v_1 - u_3)|_{y_1 = x_3^{-1} = x_1x_2} = 0$. Thus, the image of ds_x is the polynomial ring $\mathbb{Z}[w_x]$ generated by the element

$$w_x := w|_{y_1 = x_3^{-1} = x_1x_2} = \frac{x_1}{x_2} + \frac{x_2}{x_1} \in \mathcal{J}_{\mathfrak{g}_x}.$$

Note that $w_x + 1$ is the supercharacter of the adjoint representation of $\mathfrak{sl}(2)$, and that $x_1/x_2 + x_2/x_1$ equals $x_1^2 + x_2^2$ in $\mathcal{J}_{\mathfrak{g}_x}$ due to the relation $x_1x_2 = 1$. Finally, we obtain that

$$\mathrm{Im}\, ds_x = \mathbb{Z}[x_1^2 + x_2^{-2}] \subsetneq \mathcal{J}_{G_x} = \mathcal{J}_{SL(2)} = \mathbb{Z}[x_1^{\pm 1}, x_2^{\pm 1}]^{S_2}/\langle x_1x_2 - 1\rangle \cong \mathbb{Z}[x_1 + x_1^{-1}].$$

5B2. $F(4)$. Let $\beta = \frac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 - \delta_1)$. Then $\mathfrak{g}_x \cong \mathfrak{sl}(3)$ with $\Delta_x = \{\varepsilon_i - \varepsilon_j \mid 1 \leq i, j \leq 3\}$. The supercharacter ring of $F(4)$ is

$$\mathcal{J}_\mathfrak{g} = \{g(w_1, w_2) + Qh \mid h \in \mathbb{Z}[x_1^{\pm 2}, x_2^{\pm 2}, x_3^{\pm 2}, (x_1x_2x_3)^{\pm 1}, y_1^{\pm 1}]^{W_0}, \ g \in \mathbb{Z}[w_1, w_2]\},$$

where $y_1 = e^{(1/2)\delta_1}$, $x_i = e^{(1/2)\varepsilon_i}$ for $i = 1, 2, 3$, and

$$Q = (y_1 + y_1^{-1} - x_1x_2x_3 - x_1^{-1}x_2^{-1}x_3^{-1})\prod_{i=1}^{3}\left(y_1 + y_1^{-1} - \frac{x_1x_2x_3}{x_i^2} - \frac{x_i^2}{x_1x_2x_3}\right),$$

$$w_k = \sum_{i \neq j}\frac{x_i^{2k}}{x_j^{2k}} + \sum_{i=1}^{3}(x_i^{2k} + x_i^{-2k}) + y_1^{2k} + y_1^{-2k} - (y_1^k + y_1^{-k})\prod_{i=1}^{3}(x_i^k + x_i^{-k}), \quad k = 1, 2.$$

Theorem 10 implies that $ds_x(f) = f|_{x_1x_2x_3 = y_1}$ for every $f \in \mathcal{J}_\mathfrak{g}$. Hence, $ds_x(f) = ds_x(g(w_1, w_2))$ since $Q|_{x_1x_2x_3 = y_1} = 0$. Thus, the image of ds_x is generated by the elements

$$w_x^1 := w_1|_{x_1x_2x_3 = y_1} = \sum_{i \neq j}\frac{x_i^2}{x_j^2},$$

$$w_x^2 := w_2|_{x_1x_2x_3 = y_1} = \sum_{i \neq j}\frac{x_i^4}{x_j^4},$$

and is a proper subring of $\mathcal{J}_{G_x} = \mathcal{J}_{SL(3)} = \mathbb{Z}[x_1^{\pm 1}, x_2^{\pm 1}, x_3^{\pm 1}]^{S_3}/\langle x_1x_2x_3 - 1\rangle$.

5B3. $D(2, 1, \alpha)$. Let $\beta = \varepsilon_1 - \varepsilon_2 - \varepsilon_3$. Then $\mathfrak{g}_x \cong \mathbb{C}$.

If $\alpha \notin \mathbb{Q}$, then the supercharacter ring of $D(2, 1, \alpha)$ is

$$\mathcal{J}_{\mathfrak{g}} = \{c + Qh \mid c \in \mathbb{Z}, \ h \in \mathbb{Z}[u_1, u_2, u_3]\},$$

where $x_i := e^{\varepsilon_i}$, $u_i = x_i + x_i^{-1}$ for $i = 1, 2, 3$, and

$$Q = (x_1 - x_2 x_3)(x_2 - x_1 x_3)(x_3 - x_1 x_2)(1 - x_1 x_2 x_3)x_1^{-2} x_2^{-2} x_3^{-2}$$
$$= u_1^2 + u_2^2 + u_3^2 - u_1 u_2 u_3 - 4.$$

If $\alpha = p/q \in \mathbb{Q}$, then the supercharacter ring of $D(2, 1, \alpha)$ is

$$\mathcal{J}_{\mathfrak{g}} = \{g(w_\alpha) + Qh \mid g \in \mathbb{Z}[w_\alpha], \ h \in \mathbb{Z}[u_1, u_2, u_3]\},$$

where

$$w_\alpha = (x_1 + x_1^{-1} - x_2 x_3 - x_2^{-1} x_3^{-1})\frac{(x_2^p - x_2^{-p})(x_3^q - x_3^{-q})}{(x_2 - x_2^{-1})(x_3 - x_3^{-1})} + x_2^p x_3^{-q} + x_2^{-p} x_3^q.$$

By Theorem 10, $ds_x(f) = f|_{x_1 = x_2 x_3}$ for every $f \in \mathcal{J}_{\mathfrak{g}}$. Since $Q|_{x_1 = x_2 x_3} = 0$, $ds_x(f) = c$ for some $c \in \mathbb{Z}$ when $\alpha \notin \mathbb{Q}$, while $ds_x(f) = ds_x(g(w_\alpha))$ when $\alpha \in \mathbb{Q}$. Thus, the image of ds_x is $\mathbb{Z} \subset \mathcal{J}_{\mathbb{C}}$ when $\alpha \notin \mathbb{Q}$ and the image is $\mathbb{Z}[w_\alpha] \subset \mathcal{J}_{\mathbb{C}}$ when $\alpha \in \mathbb{Q}$.

Acknowledgments

References

[Boe et al. 2012] B. D. Boe, J. R. Kujawa, and D. K. Nakano, "Complexity for modules over the classical Lie superalgebra $\mathfrak{gl}(m \mid n)$", *Compos. Math.* **148**:5 (2012), 1561–1592. MR Zbl

[Cheng and Wang 2012] S.-J. Cheng and W. Wang, *Dualities and representations of Lie superalgebras*, Graduate Studies in Math. **144**, American Mathematical Society, Providence, RI, 2012. MR Zbl

[Comes and Heidersdorf 2017] J. Comes and T. Heidersdorf, "Thick ideals in Deligne's category $\mathrm{Rep}(O_\delta)$", *J. Algebra* **480** (2017), 237–265. MR Zbl

[Duflo and Serganova 2005] M. Duflo and V. Serganova, "On associated variety for Lie superalgebras", preprint, 2005. arXiv

[Entova-Aizenbud et al. 2015] I. Entova-Aizenbud, V. Hinich, and V. Serganova, "Deligne categories and the limit of categories Rep($GL(m \mid n)$)", preprint, 2015. arXiv

[Gruson and Serganova 2010] C. Gruson and V. Serganova, "Cohomology of generalized supergrassmannians and character formulae for basic classical Lie superalgebras", *Proc. Lond. Math. Soc.* (3) **101**:3 (2010), 852–892. MR Zbl

[Heidersdorf 2015] T. Heidersdorf, "On supergroups and their semisimplified representation categories", preprint, 2015. arXiv

[Heidersdorf and Weissauer 2014] T. Heidersdorf and R. Weissauer, "Cohomological tensor functors on representations of the general linear supergroup", preprint, 2014. arXiv

[Heidersdorf and Weissauer 2015] T. Heidersdorf and R. Weissauer, "Pieri type rules and $GL(2 \mid 2)$ tensor products", preprint, 2015. arXiv

[Kac and Wakimoto 1994] V. G. Kac and M. Wakimoto, "Integrable highest weight modules over affine superalgebras and number theory", pp. 415–456 in *Lie theory and geometry*, edited by J.-L. Brylinski et al., Progr. Math. **123**, Birkhäuser, Boston, 1994. MR Zbl

[Kac and Wakimoto 2014] V. G. Kac and M. Wakimoto, "Representations of affine superalgebras and mock theta functions", *Transform. Groups* **19**:2 (2014), 383–455. MR Zbl

[Macdonald 1995] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford University, 1995. MR Zbl

[Musson 2012] I. M. Musson, *Lie superalgebras and enveloping algebras*, Graduate Studies in Math. **131**, American Mathematical Society, Providence, RI, 2012. MR Zbl

[Serganova 2011] V. Serganova, "On the superdimension of an irreducible representation of a basic classical Lie superalgebra", pp. 253–273 in *Supersymmetry in mathematics and physics*, edited by S. Ferrara et al., Lecture Notes in Math. **2027**, Springer, 2011. MR Zbl

[Serganova 2014] V. Serganova, "Finite dimensional representations of algebraic supergroups", pp. 603–632 in *Proceedings of the International Congress of Mathematicians* (Seoul, 2014), vol. 1, edited by S. Y. Jang et al., Kyung Moon Sa, Seoul, 2014. MR Zbl

[Sergeev and Veselov 2011] A. N. Sergeev and A. P. Veselov, "Grothendieck rings of basic classical Lie superalgebras", *Ann. of Math.* (2) **173**:2 (2011), 663–703. MR Zbl

crystal.hoyt@weizmann.ac.il   *Department of Mathematics, Weizmann Institute of Science, ORT Braude College, Rehovot, Israel*

shifra.reif@biu.ac.il   *Department of Mathematics, Bar-Ilan University, Ramat-Gan, Israel*

Dynamics on abelian varieties in positive characteristic

Jakub Byszewski and Gunther Cornelissen
Appendix by Robert Royals and Thomas Ward

We study periodic points and orbit length distribution for endomorphisms of abelian varieties in characteristic $p > 0$. We study rationality, algebraicity and the natural boundary property for the dynamical zeta function (the latter using a general result on power series proven by Royals and Ward in the appendix), as well as analogues of the prime number theorem, also for tame dynamics, ignoring orbits whose order is divisible by p. The behavior is governed by whether or not the action on the local p-torsion group scheme is nilpotent.

Introduction

The study of the orbit structure of a dynamical system starts by considering periodic points, which, as advocated by Smale [1967, §1.4] and Artin and Mazur [1965], can be approached by considering *dynamical zeta functions*. More precisely, let S denote a set (typically, a topological space, differentiable manifold, or an algebraic variety), let $f \colon S \to S$ be a map on a set S (typically, a homeomorphism, a diffeomorphism, or a regular map), and denote by f_n the number of fixed points of the n-th iterate

$f^n = f \circ f \circ \cdots \circ f$ (n times), i.e., the number of *distinct* solutions in S of the equation $f^n(x) = x$. Let us say that f is *confined* if f_n is finite for all n, and use the notation $f \circlearrowright S$ to indicate that f satisfies this assumption. For such f, the basic question is to find patterns in the sequence $(f_n)_{n \geqslant 1}$: Does it grow in some controlled way? Does it satisfy a recurrence relation, so that finitely many f_n suffice to determine all? These questions are recast in terms of the (full) dynamical zeta function, defined as $\zeta_f(z) := \exp\left(\sum f_n z^n / n\right)$. Typical questions are:

(Q1) Is ζ_f (generically) a rational function? [1967, Problem 4.5]

(Q2) Is ζ_f algebraic as soon as it has a nonzero radius of convergence? [Artin and Mazur 1965, Question 2 on p.84]

Answers to these questions vary widely depending on the situation considered; we quote some results that provide context for our study. The dynamical zeta function $\zeta_f(z)$ is rational when f is an endomorphism of a real torus [Baake et al. 2010, Theorem 1]; f is a rational function of degree $\geqslant 2$ on $\mathbf{P}^1(\mathbf{C})$ [Hinkkanen 1994, Theorem 1]; or f is the Frobenius map on a variety X defined over a finite field $\mathbf{F}_q$, so that f_n is the number of $\mathbf{F}_{q^n}$-rational points on X and $\zeta_f(z)$ is the Weil zeta function of X [Dwork 1960; Grothendieck 1965, Corollary 5.2]. Our original starting point for this work was Andrew Bridy's automaton-theoretic proof that $\zeta_f(z)$ is transcendental for separable dynamically affine maps on $\mathbf{P}^1(\overline{\mathbf{F}}_p)$, e.g., for the power map $x \mapsto x^m$ where m is coprime to p ([Bridy 2012, Theorem 1] and [Bridy 2016, Theorems 1.2 and 1.3]). Finally, we mention that $\zeta_f(z)$ has natural boundary (namely, it does not extend analytically beyond the disk of convergence) for some explicit automorphisms of solenoids, e.g., the map dual to doubling on $\mathbf{Z}[1/6]$ (see Bell, Miles, and Ward [2014]).

   In this paper, we deal with these questions in a rather "rigid" algebraic situation, when $S = A(K)$ is the set of K-points on an abelian variety over an algebraically closed field of characteristic $p > 0$, and $f = \sigma$ is a confined endomorphism $\sigma \in \mathrm{End}(A)$ (reserving the notation f for the general case). It is plain that ζ_σ has nonzero radius of convergence (Proposition 5.2). We provide an *exact dichotomy* for rationality of zeta functions in terms of an arithmetical property of $\sigma \circlearrowright A$. Call σ *very inseparable* if $\sigma^n - 1$ is a separable isogeny for all $n \geqslant 1$. The terminology at first may appear confusing, but notice that the multiplication-by-m map for an integer m is very inseparable precisely when $p \mid m$, i.e., when it is an inseparable isogeny or zero. For another example, if A is defined over a finite field, the corresponding (inseparable) Frobenius is very inseparable.

Theorem A (Theorems 4.3 and 6.3). *Suppose that $\sigma \colon A \to A$ is a confined endomorphism of an abelian variety A over an algebraically closed field K of characteristic $p > 0$. Then σ is very inseparable if and only if it acts nilpotently on the local p-torsion subgroup scheme $A[p]^0$. Furthermore, the following dichotomy holds*:

 (i) *If σ is very inseparable, then (σ_n) is linear recurrent, and $\zeta_\sigma(z)$ is rational.*

 (ii) *If σ is not very inseparable, then (σ_n) is nonholonomic (see Definition 1.1 below), and $\zeta_\sigma(z)$ is transcendental.*

Since the local p-torsion group scheme has trivial group of K-points, in the given characterization of very inseparability it is essential to use the scheme structure of $A[p]^0$. When A is ordinary — which happens along a Zariski dense subspace in the moduli space of abelian varieties — very inseparable endomorphisms form a proper ideal in the endomorphism ring. Thus, in relation to question (Q1) above, in our case rationality is *not* generic at all.

The proofs proceed as follows: The number σ_n is the quotient of the degree of $\sigma^n - 1$ by its inseparability degree. We use arithmetical properties of the endomorphism ring of A and the action of its elements on the p-divisible subgroup to study the structure of these degrees as a function of n, showing that their ℓ-valuations are of the form "(periodic sequence) $\times$ (periodic power of $|n|_\ell$)" (Propositions 2.3 and 2.7). The emerging picture is that the degree is a very regular function of n essentially controlled by linear algebra/cohomology, but to study the inseparability degree, one needs to use geometry. The crucial tool is a general commutative algebra lemma (Lemma 2.1). We find that for some positive integers q, ϖ,

$$d_n := \deg(\sigma^n - 1) = \sum_{i=1}^{r} m_i \lambda_i^n \quad \text{for some } m_i \in \mathbf{Z} \text{ and distinct } \lambda_i \in \mathbf{C}^*, \quad \text{and}$$

$$\deg_i(\sigma^n - 1) = r_n |n|_p^{s_n} \quad \text{for } \varpi\text{-periodic sequences } r_n \in \mathbf{Q}^*, s_n \in \mathbf{Z}_{\leqslant 0}. \tag{1}$$

Note in particular that this implies that the *degree zeta function*

$$D_\sigma(z) := \exp\left(\sum d_n z^n / n\right) = \prod_{i=1}^{r}(1 - \lambda_i z)^{-m_i},$$

(called the "false zeta function" by Smale [1967, p.768]) is rational. In Proposition 3.1, we then prove an adaptation of the Hadamard quotient theorem in which one of the series displays such periodic behavior, but the other is merely assumed holonomic. From this, we can already deduce the rationality or transcendence of ζ_σ. In contrast to Bridy's result, we make no reference to the theory of automata.

Example B. We present as a warm up example the case where E is an ordinary elliptic curve over $\mathbf{F}_3$ and let $\sigma = [2]$ be the doubling map and $\tau = [3]$ the tripling map, where everything can be computed explicitly. Although the example lacks some of the features of the general case, we hope this will help the reader to grasp the basic ideas. For this example, some facts follow from the general theory in [Bridy 2016]; and, since $\zeta_\sigma(z)$ equals the dynamical zeta function induced by doubling on the direct product of the circle and the solenoid dual to $\mathbf{Z}[1/6]$ [Bell et al. 2014], some properties could be deduced from the existing literature, which we will not do.

First of all,

$$\deg(\sigma^n - 1) = (2^n - 1)^2 = 4^n - 2 \cdot 2^n + 1 \quad \text{and} \quad \deg(\tau^n - 1) = (3^n - 1)^2 = 9^n - 2 \cdot 3^n + 1.$$

The corresponding degree zeta functions are:

$$D_\sigma(z) = \frac{(1 - 2z)^2}{(1 - 4z)(1 - z)} \quad \text{and} \quad D_\tau(z) = \frac{(1 - 3z)^2}{(1 - 9z)(1 - z)}.$$

From the definition, σ is not very inseparable but τ is. In fact, $\tau_n = \deg(3^n - 1)$ and $\zeta_\tau = D_\tau$ but, since we are on an ordinary elliptic curve (where $E[p^m]$ is of order p^m), we find

$$\sigma_n = (2^n - 1)^2 |2^n - 1|_3 = (2^n - 1)^2 r_n^{-1} |n|_3^{-s_n}, \quad \text{with } \varpi = 2; \, r_{2k} = 3, \, s_{2k} = -1; \, r_{2k+1} = 1, \, s_{2k+1} = 0.$$

In our first proof of the transcendence of $\zeta_\sigma(z)$, we use the fact that σ_{2n} differs from a linear recurrence by a factor $|n|_3$ to argue that it is not holonomic.

Since we are on an ordinary curve, the local 3-torsion group scheme is $E[3]^0 = \mu_3$, which has $\text{End}(E[3]^0) = \mathbf{F}_3$ in which the only nilpotent element is the zero element. Thus, we can detect very inseparability of σ or τ by their image under $\text{End}(E) \to \text{End}(E[3]^0) = \mathbf{F}_3$ being zero, and indeed, $\tau = [3]$ maps to zero, but $\sigma = [2]$ does not. $\diamond$

In some cases, we prove a stronger result. Let Λ denote a dominant root of the linear recurrence (1) satisfied by $\deg(\sigma^n - 1)$, i.e., $\Lambda \in \{\lambda_i\}$ has $|\Lambda| = \max|\lambda_i|$. In Proposition 5.1, we prove some properties of Λ, e.g., that $\Lambda > 1$ is real and $1/\Lambda$ is a pole of ζ_σ.

Theorem C (Theorem 5.5). *If $\sigma : A \to A$ is a confined, not very inseparable endomorphism of an abelian variety A over an algebraically closed field K of characteristic $p > 0$ such that Λ is the unique dominant root, then the dynamical zeta function $\zeta_\sigma(z)$ has a natural boundary along $|z| = 1/\Lambda$.*

This result implies nonholonomicity and hence transcendence for such functions; our proof of Theorem C is independent of that of Theorem A. The existence of a natural boundary follows from the fact that the logarithmic derivative of ζ_σ can be expressed through certain "adelically perturbed" series that satisfy Mahler-type functional equations in the sense of [Bell et al. 2013], and hence have accumulating poles (proven in the Appendix by Royals and Ward). From the theorem we see, in connection with question (Q2) above, that a "generic" ζ_σ is far from algebraic (not even holonomic), despite having a positive radius of convergence.

Example B (continued). The dominant roots are $\Lambda_\sigma = 4$ and $\Lambda_\tau = 9$, which are simple. Since ζ_τ is rational, it extends meromorphically to $\mathbf{C}$. We prove that $\zeta_\sigma(z)$ has a natural boundary at $|z| = \frac{1}{4}$, as follows. It suffices to prove this for the function $Z(z) = z\zeta_\sigma'(z)/\zeta_\sigma(z) = \sum \sigma_n z^n$, which we can expand as

$$Z(z) = \sum_{2 \nmid n} (2^n - 1)^2 z^n + \tfrac{1}{3} \sum_{2 \mid n} |n|_3 (2^n - 1)^2 z^n;$$

if we write $f(t) = \sum |n|_3 t^n$, then

$$Z(z) = \frac{z(1 + 28z^2 + 16z^4)}{(1 - 16z^2)(1 - 4z^2)(1 - z^2)} + \tfrac{1}{3}(f(16z^2) - 2f(4z^2) + f(z^2)).$$

It suffices to prove that $f(t)$ has a natural boundary at $|t| = 1$, and this follows from the fact that f satisfies the functional equation

$$f(z) = \frac{z^2 + z}{1 - z^3} + \tfrac{1}{3}f(z^3),$$

and hence acquires singularities at the dense set in the unit circle consisting of all third power roots of unity. ◇

[Section 6](#) constitutes a purely arithmetic geometric study of the notion of very inseparability. We prove that very inseparable isogenies are inseparable and that an isogeny $\sigma : E \to E$ of an elliptic curve E is very inseparable if and only if it is inseparable. We give examples where very inseparability is not the same as inseparability even for simple abelian varieties. We study very inseparability using the description of $A[p]^0$ through Dieudonné modules, from which it follows that very inseparable endomorphisms are precisely those of which a power factors through the Frobenius morphism.

Example D. Let E denote an ordinary elliptic curve over a field of characteristic 3 and set $A = E \times E$; then the map $[2] \times [3]$ is inseparable but not very inseparable, since there exist n for which $2^n - 1$ is divisible by 3. In this case, $\text{End}(A[3]^0)$ is the two-by-two matrix algebra over $\mathbf{F}_3$, which contains noninvertible nonnilpotent elements, and under $\text{End}(A) \to \text{End}(A[3]^0) = M_2(\mathbf{F}_3)$, $[2] \times [3]$ is mapped to the matrix $\text{diag}(2, 0)$, which is such an element. ◇

We then introduce the *tame zeta function* ζ_σ^*, defined as

$$\zeta_\sigma^*(z) := \exp\left(\sum_{p \nmid n} \sigma_n \frac{z^n}{n} \right), \tag{2}$$

summing only over n that are not divisible by p. The full zeta function ζ_σ is an infinite product of tame zeta functions of p-power iterates of σ ([Proposition 7.2](#)). Thus, one "understands" the full zeta function by understanding those tame zeta functions. Our main result in this direction says that the tame zeta function belongs to a cyclic extension of the field of rational functions:

Theorem E ([Theorem 7.3](#)). *For any* (*very inseparable or not*) $\sigma \circlearrowleft A$, *a positive integer power of the tame zeta function* ζ_σ^* *is rational.*

The minimal such integral power $t_\sigma > 0$ seems to be an interesting arithmetical invariant of $\sigma \circlearrowleft A$; for example, on an ordinary elliptic curve E, one can choose t_σ to be a p-th power for $\sigma \circlearrowleft E$, but for a certain endomorphism of a supersingular elliptic curve, $t_\sigma = p^2(p + 1)$ (cf. [Proposition 7.4](#)).

Example B (continued). The tame zeta function for σ is, by direct computation,

$$\zeta_\sigma^*(z) = \exp\left(\frac{1}{3} \sum_{3 \nmid n, \, 2 \mid n} (2^n - 1)^2 \frac{z^n}{n} + \sum_{3 \nmid n, \, 2 \nmid n} (2^n - 1)^2 \frac{z^n}{n} \right)$$

$$= \sqrt[9]{\frac{F_2(z)^9 F_{64}(z^6)}{F_8(z^3)^3 F_4(z^2)^3}}, \quad \text{where } F_a(z) := \frac{(1 - az)^2}{(1 - a^2 z)(1 - z)},$$

and hence $t_\sigma = 9$. Note that even for the very inseparable τ, $\zeta_\tau^*(z) = D_\tau(z)/\sqrt[3]{D_{\tau^3}(z^3)}$ is not rational, and $t_\tau = 3$. ◇

In Section 8, we investigate functional equations for ζ_σ and ζ_σ^* under $z \mapsto 1/(\deg(\sigma)z)$. For very inseparable σ, there is such a functional equation (which can also be understood cohomologically), but not for ζ_σ having a natural boundary. On the other hand, we show that all tame zeta functions satisfy a functional equation when continued to their Riemann surface (see Theorem 8.3).

In Section 9, we study the distribution of prime orbits for $\sigma \circlearrowleft A$. Let P_ℓ denote the number of prime orbits of length ℓ for σ. In case of a unique dominant root, we deduce sharp asymptotics for P_ℓ of the form

$$P_\ell = \frac{\Lambda^\ell}{\ell r_\ell |\ell|_p^{s_\ell}} + O(\Lambda^{\Theta\ell}), \quad \text{where } \Theta := \max\{\mathrm{Re}(s) : D_\sigma(\Lambda^{-s}) = 0\}. \tag{3}$$

We average further, as in the prime number theorem (PNT). Define the *prime orbit counting function* $\pi_\sigma(X)$ and the *tame prime orbit counting function* $\pi_\sigma^*(X)$ by

$$\pi_\sigma(X) := \sum_{\ell \leqslant X} P_\ell \quad \text{and} \quad \pi_\sigma^*(X) := \sum_{\substack{\ell \leqslant X \\ p \nmid \ell}} P_\ell.$$

Again, whether or not σ is very inseparable is related to the limit behavior of these functions.

Theorem F (Theorems 9.5 and 9.9). *If $\sigma \circlearrowleft A$ has a unique dominant root $\Lambda > 1$, then, with ϖ as in (1) and for X taking integer values, we have:*

(i) *If σ is very inseparable, $\lim_{X \to +\infty} X\pi_\sigma(X)/\Lambda^X$ exists and equals $\Lambda/(\Lambda - 1)$.*

(ii) *If σ is not very inseparable, then $X\pi_\sigma(X)/\Lambda^X$ is bounded away from zero and infinity, its set of accumulation points is a union of a Cantor set and finitely many points (in particular, it is uncountable), and every accumulation point is a limit along a sequence of integers X for which (X, X) converges in the topological group*

$$\{(a, x) \in \mathbf{Z}/\varpi\mathbf{Z} \times \mathbf{Z}_p : a \equiv x \mod |\varpi|_p^{-1}\}.$$

(iii) *For any $k \in \{0, \ldots, p\varpi - 1\}$, the limit $\displaystyle\lim_{\substack{X \to +\infty \\ X \equiv k \mod p\varpi}} X\pi_\sigma^*(X)/\Lambda^X =: \rho_k$ exists.*

An expression for ρ_k in terms of arithmetic invariants can be found in (39). We also present an analogue of Mertens' second theorem (Proposition 9.10) on the asymptotics of

$$\mathrm{Mer}(\sigma) := \sum_{\ell \leqslant X} P_\ell/\Lambda^\ell$$

in X. It turns out that, in contrast to the PNT analogue, this type of averaged asymptotics is insensitive to the endomorphism being very inseparable or not.

Example B (continued). Including a subscript for σ or τ in the notation, Möbius inversion relates $P_{\sigma,\ell}$ to the values of σ_ℓ, and hence of λ_i, r_n, s_n; we find for the very inseparable τ that $P_{\tau,\ell} = 9^\ell/\ell + O(3^\ell)$, which we can sum to the analogue of the prime number theorem $\pi_\tau(X) \sim 9/8 \cdot 9^X/X$. The situation is
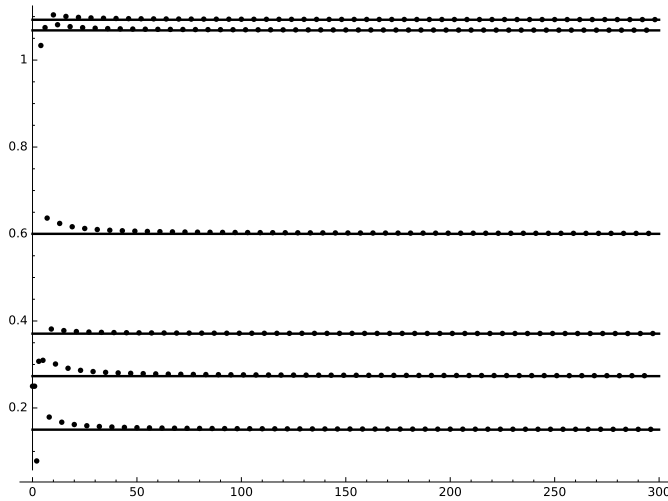
Figure 1. Plot of $X \mapsto X\pi_\sigma^*(X)/4^X$, where σ is doubling on an ordinary elliptic curve in characteristic 3 (dots) and the six limit values as computed from (39) (horizontal solid lines).

different for the not very inseparable σ, where

$$P_{\sigma,\ell} = \frac{4^\ell}{\ell} \begin{cases} |3\ell|_3 & \text{if } \ell \text{ is even,} \\ 1 & \text{if } \ell \text{ is odd} \end{cases} + O(2^\ell), \tag{4}$$

and $\pi_\sigma(X)X/4^X$ has uncountably many limit points in the interval $[1/12, 4/3]$ (following the line of thought set out in [Everest et al. 2007]).

We find as main term in $\text{Mer}(\tau)$ the X-th harmonic number $\sum_{\ell \leqslant X} 1/\ell$, and, taking into account the constant term from summing error terms in (3), we get $\text{Mer}(\tau) \sim \log X + c$ for some $c \in \mathbf{R}$. On the other hand, a more tedious computation gives $\text{Mer}(\sigma) \sim 5/8 \log X + c'$ for some $c' \in \mathbf{R}$.

Concerning the tame case, Figure 1 shows a graph (computed in SageMath [SageMath 2016]) of the function $\pi_\sigma^*(X)X/4^X$, in which one sees six different accumulation points. The values ρ_k can be computed in closed form as rational numbers by noticing that if we sum (4) only over ℓ not divisible by 3, we can split it into a finite sum over different values of ℓ modulo 6. We show the computed values in Table 1, which match the asymptotics in the graph.[1] $\diamond$

We briefly discuss convergence rates in the above theorem (compare, e.g., [Pollicott and Sharp 1998]) in relation to analogues of the Riemann hypothesis (see Proposition 9.11): there is a function $M(X)$

---

[1]An amusing observation is the similarity between Figure 1 and the final image in the notorious paper by Fermi, Pasta, Ulam and Tsingou (see the very suggestive Figures 4.3 and 4.5 in the modern account [Benettin et al. 2008]): the time averaged fraction of the energy per Fourier mode in the eponymous particle system seems to converge to distinct values, whereas mixing would imply convergence to a unique value; by work of Izrailev and Chirikov the latter seems to happen at higher energy densities. This suggests an analogy (not in any way mathematically precise) between "very inseparable" and "ergodic/mixing/high energy density".

$k \bmod 6$	$\rho_k \cdot 2^{-2} \cdot 3^3 \cdot 5 \cdot 7 \cdot 13$	ρ_k (numerical)
0	839	0.27317867317867
1	$17 \cdot 193$	1.06829466829467
2	$2^2 \cdot 461$	0.60040700040700
3	461	0.15010175010175
4	$17 \cdot 67$	0.37085877085877
5	$2^2 \cdot 839$	1.09271469271469

Table 1. Exact and numerical values of the six limit values in Figure 1.

determined by the combinatorial information $(p, \Lambda, \varpi, (r_n), (s_n))$ associated to $\sigma \circlearrowright A$ as in (1), such that for integer values X, we have

$$\pi_\sigma(X) = M(X) + O(\Lambda^{\Theta X})$$

where the "power saving" Θ is determined by the real part of zeros of the degree zeta function $D_\sigma(\Lambda^{-s})$. Said more colloquially, the main term reflects the growth rate (analogue of entropy) and inseparability, whereas the error term is insensitive to inseparability and determined purely by the action of σ on the total cohomology.

Example B (continued). If we collect the main terms using the function, for $k \in \{0, 1\}$,

$$F_k(\Lambda, X) = \sum_{\substack{\ell \leqslant X \\ \ell \equiv k \bmod 2}} \Lambda^\ell / \ell$$

we arrive at the following analogue of the Riemann hypothesis for σ:

$$\pi_\sigma(X) = M(X) + O(2^X), \quad \text{with } M(X) := \tfrac{1}{3} F_0(4, X) + F_1(4, X) - \sum_{i=1}^{\lfloor \log_3(X) \rfloor} \frac{2}{9^i} F_0\left(4^{3^i}, \left\lfloor \frac{X}{3^i} \right\rfloor\right).$$

See Figure 2 (computed in SageMath [SageMath 2016]) for an illustration. ◇

Example G. All our results apply to the situation where A is an abelian variety defined over a finite field $\mathbf{F}_q$ and σ is the Frobenius of $\mathbf{F}_q$, which is very inseparable. This implies known results about curves $C/\mathbf{F}_q$ when applied to the Jacobian $A = \mathrm{Jac}(C)$ of C, such as rationality of the zeta function and analogues of PNT (compare [Rosen 2002, Theorem 5.12]).

We finish this introduction by discussing some open problems and possible future research directions. In the near future, we hope to treat the case of linear algebraic groups, which will require different techniques. Our methods in this paper rest on the presence of a group structure preserved by the map. What happens in absence of a group structure is momentarily unclear to us, but we believe that the study of the tame zeta function in such a more general setup merits consideration. We will consider this for dynamically affine maps on $\mathbf{P}^1$ in the sense of [Bridy 2016] (not equal to, but still "close to" a group) in future work. It would be interesting to study direct relations between our results and that of compact group endomorphisms and S-integer dynamical systems — we briefly touch upon this at the end of Section 5.
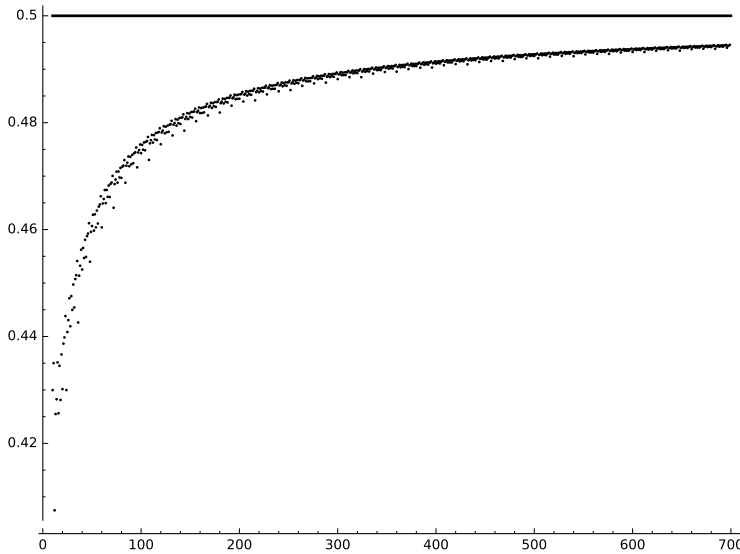
Figure 2. Plot of $X \mapsto \log_4 |\pi_\sigma(X) - M(X)|/X$ (dots) for integer $X \in [10, 700]$ and the solid line $\Theta = 1/2$, where σ is doubling on an ordinary elliptic curve in characteristic 3.

1. Generalities

Rationality and holonomicity. We start by recalling some basic facts about recurrence sequences.

Definition 1.1. A power series $f = \sum_{n \geqslant} a_n z^n \in \mathbf{C}[\![z]\!]$ is *holonomic* (or *D-finite*) if it satisfies a linear differential equation over $\mathbf{C}(z)$, i.e., if there exist polynomials $q_0, \ldots, q_d \in \mathbf{C}[z]$, not all zero, such that

$$q_0(z) f(z) + q_1(z) f'(z) + \ldots + q_d(z) f^{(d)}(z) = 0. \tag{5}$$

A sequence $(a_n)_{n \geqslant 1}$ is called *holonomic* if its associated generating function $f = \sum_{n \geqslant 1} a_n z^n \in \mathbf{C}[\![z]\!]$ is holonomic.

In the following lemma, we collect some well-known equivalences between properties of a sequence and its generating series:

Lemma 1.2. *Let $(a_n)_{n \geqslant 1}$ be a sequence of complex numbers.*

(i) *The following conditions are equivalent*:

    (a) *The sequence $(a_n)_{n \geqslant 1}$ satisfies a linear recurrence.*

    (b) *The power series $\sum_{n \geqslant 1} a_n z^n$ is in $\mathbf{C}(z)$.*

    (c) *There exist complex numbers λ_i and polynomials $q_i \in \mathbf{C}[z]$, $1 \leqslant i \leqslant s$, such that we have $a_n = \sum_{i=1}^{s} q_i(n) \lambda_i^n$ for n large enough.*

(ii) *The following conditions are equivalent*:

    (a) *The power series $f(z) = \exp\left(\sum_{n \geqslant 1} \frac{a_n}{n} z^n\right)$ is in $\mathbf{C}(z)$.*

(b) *There exist integers m_i and complex numbers λ_i, $1 \leqslant i \leqslant s$, such that the sequence a_n can be written as $a_n = \sum_{i=1}^{s} m_i \lambda_i^n$ for all $n \geqslant 1$.*

Furthermore, if all a_n are in $\mathbf{Q}$, then $f(z)$ is in $\mathbf{Q}(z)$.

(iii) *The following conditions are equivalent*:

(a) *The sequence $(a_n)_{n \geqslant 1}$ is holonomic.*

(b) *There exist polynomials $q_0, \ldots, q_d \in \mathbf{C}[z]$, not all zero, such that for all $n \geqslant 1$ we have*
   $$q_0(n)a_n + \ldots + q_d(n)a_{n+d} = 0.$$

Furthermore, if a power series $f(z) \in \mathbf{C}[\![z]\!]$ is algebraic over $\mathbf{C}(z)$, then it is holonomic.

Proof. Statement (i) follows from [Stanley 2012, Theorem 4.1.1 and Proposition 4.2.2]. Statement (ii) is [Stanley 2012, Example 4.8]; the final claim holds since $\mathbf{C}(z) \cap \mathbf{Q}((z)) = \mathbf{Q}(z)$ (see, e.g., [Milne 2013, Lemma 27.9]). Statement (iii) is [Stanley 1980, Theorems 1.5 and 2.1]. $\qquad \square$

Initial reduction from rational maps to confined endomorphisms. Let A denote an abelian variety over an algebraically closed field K. Rational maps on abelian varieties are automatically regular [Milne 2008, I.3.2], and are always compositions of an endomorphism and a translation [Milne 2008, I.3.7]. We say that a regular map $\sigma \colon A \to A$ is *confined* if the set of fixed points of σ^n is finite for all n, which we assume from now on. We use the notations from the introduction: σ_n is the number of fixed points of σ^n and ζ_σ is the Artin–Mazur dynamical zeta function of σ.

If σ is an endomorphism of A, confinedness is equivalent to the finiteness of the kernel $\ker(\sigma^n - 1)$ for all n, or the fact that all $\sigma^n - 1$ are isogenies [Milne 2008, I.7.1]. For arbitrary maps, the following allows us to restrict ourselves to the study of zeta functions of confined endomorphisms (where case (i) can effectively occur, for example, when σ is a translation by a nontorsion point):

Proposition 1.3. *Let $\sigma \colon A \to A$ be a confined regular map and write $\sigma = \tau_b \psi$, where τ_b is a translation by $b \in A(K)$ and ψ is an endomorphism of A. Then either*

(i) *$\sigma_n = 0$ for all n and hence $\zeta_\sigma(z) = 1$; or else*

(ii) *ψ is confined and $\zeta_\sigma(z) = \zeta_\psi(z)$.*

Proof. Iterates of σ are of the form

$$\sigma^n = \tau_{b^{(n)}} \psi^n, \quad \text{where } b^{(n)} = \sum_{i=0}^{n-1} \psi^i(b).$$

Thus, $\sigma_n = \psi_n$ if $b^{(n)} \in \mathrm{im}(\psi^n - 1)$ and $\sigma_n = 0$ otherwise. If $\sigma_n = 0$ for all n, then $\zeta_\sigma(z) = 1$. Otherwise, for some $m \geqslant 1$ we have $\sigma_m > 0$ and thus $b^{(m)} \in \mathrm{im}(\psi^m - 1)$, $\sigma_m = \psi_m$, and $\psi^m - 1$ is an isogeny. It follows that for all $k \geqslant 1$ we have $b^{(km)} = \sum_{i=0}^{k-1} \psi^{im}(b^{(m)})$ and hence $b^{(km)} \in \mathrm{im}(\psi^{km} - 1)$, $\sigma_{km} = \psi_{km}$, and $\psi^{km} - 1$ is an isogeny. Since $\psi^k - 1$ is a factor of $\psi^{km} - 1$, we conclude that ψ is a confined endomorphism, and hence $\psi^k - 1$ is surjective. In particular, $b^{(k)} \in \mathrm{im}(\psi^k - 1)$, so $\sigma_n = \psi_n$ for all n, and hence $\zeta_\sigma(z) = \zeta_\psi(z)$. $\qquad \square$

We make the following standing assumptions from now on, that we will not repeat in formulations of results. Only in Section 6 shall we temporarily drop the assumption of confinedness, since this will make exposition smoother (this will be clearly indicated).

> *Standing assumptions*: K is an algebraically closed field of characteristic $p > 0$. A is an abelian variety over K of dimension g. The endomorphism $\sigma \colon A \to A$ is confined.

2. Periodic patterns in (in)separability degrees

For now, we will consider ζ_σ as a *formal power series*

$$\zeta_\sigma(z) := \exp\left(\sum_{n \geqslant 1} \sigma_n \frac{z^n}{n}\right),$$

and postpone the discussion of complex analytic aspects to Section 5. Let $\deg_i(\tau)$ denote the inseparability degree of an isogeny $\tau \in \operatorname{End}(A)$ (a pure p-th power). We then have the basic equation

$$\sigma_n = \frac{\deg(\sigma^n - 1)}{\deg_i(\sigma^n - 1)}. \tag{6}$$

The strategy is to first consider the "false" (in the terminology of Smale [1967]) zeta function with σ_n replaced by the degree of $\sigma^n - 1$. This turns out to be a rational function. We then turn to study the inseparability degree, which is determined by the p-valuations of the other two sequences.

We start with a general lemma in commutative algebra that is our crucial tool for controlling the valuations of certain elements of sequences:

Lemma 2.1. *Let S denote a local ring with maximal ideal $\mathfrak{m}$ and residue field k of characteristic $p > 0$ such that the ring S/pS is artinian. For $\sigma \in S$ and a positive integer n, let $I_n := (\sigma^n - 1)S$. Let $\bar{\sigma}$ denote the image of σ in k.*

- (i) *If $\sigma \in \mathfrak{m}$, then $I_n = S$ for all n.*
- (ii) *If $\sigma \in S^*$, let e be the order of $\bar{\sigma}$ in k^*. Then:*
  - (a) *If $e \nmid n$, then $I_n = S$ (this happens in particular if $e = \infty$).*
  - (b) *If $e \mid n$ and $p \nmid m$, then $I_{mn} = I_n$.*
  - (c) *There exists an integer n_0 such that for all n with $e \mid n$ and $\operatorname{ord}_p(n) > n_0$, we have $I_{pn} = pI_n$.*

Proof. Part (i) is clear, so assume $\sigma \in S^*$. If $e \nmid n$, then $\sigma^n - 1$ is invertible in S, since $\bar{\sigma}^n - 1 \neq 0$ in k and hence $I_n = S$.

If $e \mid n$, we can assume without loss of generality that $e = 1$ (replacing σ by σ^e). Write $\sigma^n = 1 + \varepsilon$ for $\varepsilon \in \mathfrak{m}$. Then for m coprime to p, we immediately find

$$\sigma^{mn} - 1 = \varepsilon u$$

for a unit $u \in S^*$, and hence $I_{mn} = I_n$, which proves (b). On the other hand,

$$\sigma^{pn} - 1 = p\varepsilon v + \varepsilon^p \tag{7}$$

for some unit $v \in S^*$. This shows that $\sigma^{pn} - 1 = \varepsilon(pv + \varepsilon^{p-1}) \subseteq \varepsilon\mathfrak{m}$, which already implies that we get

$$I_{pn} \subseteq I_n \mathfrak{m}, \quad \text{for all } n. \tag{8}$$

Since S/pS is artinian, there exists an integer n_0 such that $\mathfrak{m}^{n_0} \subseteq pS$. By iterating (8) $n_0 + 1$ times, we have

$$I_n \subseteq p\mathfrak{m}, \quad \text{for all } n \text{ with } \mathrm{ord}_p(n) > n_0.$$

Assuming now that $\mathrm{ord}_p(n) > n_0$, we have $\varepsilon \in p\mathfrak{m}$, so $\varepsilon^p \in p\varepsilon\mathfrak{m}$. Hence we conclude from (7) that $\sigma^{pn} - 1 = p\varepsilon w$ for some unit $w \in S^*$, and hence $I_{pn} = pI_n$. $\qquad\square$

The degree zeta function. We start by considering the following zeta function with σ_n replaced by the degree of $\sigma^n - 1$.

Definition 2.2. The *degree zeta function* is defined as the formal power series

$$D_\sigma(z) := \exp\left(\sum_{n \geqslant 1} \frac{\deg(\sigma^n - 1)}{n} z^n \right).$$

Proposition 2.3.   (i) $D_\sigma(z) \in \mathbf{Q}(z)$.

(ii) *Let ℓ be a prime* (*which might or might not be equal to p*). *Then the sequence of ℓ-adic valuations* $(|\deg(\sigma^n - 1)|_\ell)_{n \geqslant 1}$ *is of the form*

$$|\deg(\sigma^n - 1)|_\ell = r_n \cdot |n|_\ell^{s_n}$$

for some periodic sequences (r_n) and (s_n) with $r_n \in \mathbf{Q}^*$ and $s_n \in \mathbf{N}$. Furthermore, there is an integer ω such that we have

$$r_n = r_{\gcd(n,\omega)} \quad \text{for } \ell \nmid n.$$

Proof. By [Grieve 2017, Cororllary 3.6], the degree of σ and the sequence $\deg(\sigma^n - 1)$ can be computed as

$$\deg \sigma = \prod_{i=1}^{k} \mathrm{Nrd}_{R_i/\mathbf{Q}}(\alpha_i)^{\nu_i}, \quad \deg(\sigma^n - 1) = \prod_{i=1}^{k} \mathrm{Nrd}_{R_i/\mathbf{Q}}(\alpha_i^n - 1)^{\nu_i},$$

where the R_i are finite-dimensional simple algebras over $\mathbf{Q}$, the α_i are elements of R_i, $\mathrm{Nrd}_{R_i/\mathbf{Q}}$ is the reduced norm, and the ν_i are positive integers. These formulæ come from replacing the variety A by an isogenous one that is a finite product of simple abelian varieties and applying the well-known results on the structure of endomorphism algebras of simple abelian varieties.

After tensoring with $\overline{\mathbf{Q}}$, the algebras R_i become isomorphic to a finite product of matrix algebras over $\overline{\mathbf{Q}}$. For matrix algebras the notion of reduced norm coincides with the notion of determinant, and since the determinant of a matrix is equal to the product of its eigenvalues, we obtain formulæ of the form

$$\deg(\sigma) = \prod_{i=1}^{q} \xi_i, \quad \deg(\sigma^n - 1) = \prod_{i=1}^{q} (\xi_i^n - 1), \tag{9}$$

with $\xi_i \in \overline{\mathbf{Q}}$ (with possible repetitions to take care of multiplicities) and $q = 2g$ (since deg is a polynomial function of degree $2g$). Multiplying out the terms in this expression, we finally obtain a formula of the form

$$\deg(\sigma^n - 1) = \sum_{i=1}^{r} m_i \lambda_i^n, \tag{10}$$

for some $m_i \in \mathbf{Z}$ and $\lambda_i \in \overline{\mathbf{Q}}$. Now (i) follows from 1.2(ii).

In order to prove (ii), we will use (9). Consider a finite extension L of the field of ℓ-adic numbers $\mathbf{Q}_\ell$ obtained by adjoining all ξ_i with $1 \leqslant i \leqslant q$. There is a unique extension of the valuation $|\cdot|_\ell$ to L that we continue to denote by the same symbol. Then we have

$$|\deg(\sigma^n - 1)|_\ell = \prod_{i=1}^{q} |\xi_i^n - 1|_\ell.$$

We now claim that for $\xi \in L$, we have

$$|\xi^n - 1|_\ell = \begin{cases} |\xi|_\ell^n & \text{if } |\xi|_\ell > 1, \\ r_n^\xi |n|_\ell^{s_n^\xi} & \text{if } |\xi|_\ell = 1, \\ 1 & \text{if } |\xi|_\ell < 1, \end{cases} \tag{11}$$

where $(r_n^\xi)_n$ and $(s_n^\xi)_n$ are certain periodic sequences, $r_n^\xi \in \mathbf{R}^*$, $s_n^\xi \in \{0, 1\}$. The first and the last line of the claim are immediate, and the second one follows from applying Lemma 2.1 to the ring of integers $S = \mathcal{O}_L$ with $\sigma = \xi$, as follows: set $a_n = |\xi^n - 1|_\ell^{-1}$ and let e_ξ be the order of ξ in the residue field of S (note that e_ξ is not divisible by ℓ). Then by Lemma 2.1 there exists an integer N such that $a_n = 1$ if $e_\xi \nmid n$; $a_{mn} = a_n$ if $e_\xi \mid n$ and $\ell \nmid m$; and $a_{\ell n} = \ell a_n$ if $e_\xi \mid n$ and $\mathrm{ord}_\ell(n) \geqslant N$. Therefore, it suffices to set $(r_n^\xi, s_n^\xi) = (1, 0)$ for $e_\xi \nmid n$; $(r_n^\xi, s_n^\xi) = (a_{e_\xi \ell^\nu}^{-1}, 0)$ for $e_\xi \mid n$ and $\nu := \mathrm{ord}_\ell(n) < N$; and $(r_n^\xi, s_n^\xi) = (a_{e_\xi \ell^N}^{-1} \ell^N, 1)$ for $e_\xi \mid n$ and $\mathrm{ord}_\ell(n) \geqslant N$. Note that for $\ell \nmid n$ we have

$$r_n^\xi = \begin{cases} 1 & \text{if } e_\xi \nmid n, \\ a_{e_\xi}^{-1} & \text{if } e_\xi \mid n. \end{cases}$$

Multiplying together formulæ (11) for $\xi = \xi_1, \ldots, \xi_q$, we obtain

$$|\deg(\sigma^n - 1)|_\ell = \rho^n r_n |n|_\ell^{s_n},$$

where

$$\rho = \prod_{i=1}^{q} \max(|\xi_i|_\ell, 1) \geqslant 1$$

and (r_n) and (s_n) are periodic sequences, $r_n \in \mathbf{R}^*$, $s_n \in \mathbf{N}$. We claim that $\rho = 1$ (that is, there is no i such that $|\xi_i|_\ell > 1$). Indeed, we know that $\deg(\sigma^n - 1)$ is an integer, and hence $\rho^n r_n |n|_\ell^{s_n} \leqslant 1$ for all n. Thus, taking $n \to \infty$, $\ell \nmid n$, we get $\rho = 1$ and $r_n \in \mathbf{Q}^*$. This finishes the proof of the formula for $|\deg(\sigma^n - 1)|_\ell$.

Furthermore, we have

$$r_n = \prod_{e_{\xi_i} \mid n} a_{e_{\xi_i}}^{-1}, \quad \text{for } \ell \nmid n,$$

and hence the final formula holds with $\omega = \text{lcm}(e_{\xi_1}, \ldots, e_{\xi_q})$. $\qquad\square$

Remark 2.4. We present an alternative, cohomological description of the degree zeta function $D_\sigma(z)$. Fix a prime $\ell \neq p$ and let $\mathrm{H}^i := \mathrm{H}^i_{\text{ét}}(A, \mathbf{Q}_\ell) = \bigwedge^i (V_\ell A)^\vee$ denote the i-th ℓ-adic cohomology group of A, ($V_\ell A = T_\ell A \otimes_{\mathbf{Z}_\ell} \mathbf{Q}_\ell$, $T_\ell A$ is the Tate module and $^\vee$ denotes the dual); then

$$D_\sigma(z) = \prod_{i=1}^{2g} \det(1 - \sigma^* z | \mathrm{H}^i)^{(-1)^{i+1}}. \tag{12}$$

This follows in the same way as for the Weil zeta function: let $\Gamma_{\sigma^n} \subseteq A \times A$ denote the graph of σ^n and $\Delta \subseteq A \times A$ is the diagonal [Milne 2013, 25.6]. The Lefschetz fixed point theorem [Milne 2013, 25.1] implies that

$$(\Gamma_{\sigma^n} \cdot \Delta) = \sum_{i=0}^{2g} (-1)^i \, \text{tr}(\sigma^n | \mathrm{H}^i).$$

Now Γ_{σ^n} intersects Δ precisely along the (finite flat) group torsion group scheme $A[\sigma^n - 1]$, and hence the intersection number $(\Gamma_{\sigma^n} \cdot \Delta)$ is the order of this group scheme, which is $\deg(\sigma^n - 1)$. Then the standard determinant-trace identity [Milne 2013, 27.5] implies the result (12).

The characteristic polynomial of σ_* acting on H^1 has integer coefficients independent of the choice of ℓ and its set of roots is precisely the set of algebraic numbers ξ_i from the proof of Proposition 2.3 (with multiplicities), see, e.g., [Mumford 2008, IV.19, Theorems 3 and 4].

Example 2.5. Suppose A is an abelian variety over a finite field $\mathbf{F}_q$ and σ is the q-Frobenius. Then $\sigma^n - 1$ is separable for all n, so $\sigma_n = \deg(\sigma^n - 1)$ for all n, and $\zeta_\sigma(z) = D_\sigma(z)$ is exactly the Weil zeta function of $A/\mathbf{F}_q$. Thus, we recover the rationality of that function for abelian varieties; note that this is an "easy" case: by cutting A with suitable hyperplanes, we are reduced to the case of (Jacobians of) curves, hence essentially to the Riemann–Roch theorem for global function fields proven by F. K. Schmidt in 1927.

The inseparability degree. As in Proposition 2.3, we can control the regularity in the sequence of inseparability degrees, with some more (geometric) work; this is relevant in the light of (6). We start with a decomposition lemma in commutative algebra:

Lemma 2.6. *Let R be a (commutative) ring and let M be an R-module such that for every $m \in M$ the ring $R/\text{ann}(m)$ is artinian. Let $\mathfrak{m}$ be a maximal ideal of R. Then the localization $M_{\mathfrak{m}}$ is equal to*

$$M_{\mathfrak{m}} = M[\mathfrak{m}^\infty] := \{m \in M : \mathfrak{m}^k m = 0 \text{ for some } k \geqslant 1\}$$

and

$$M = \bigoplus_{\mathfrak{m}} M_{\mathfrak{m}},$$

the direct sum being taken over all maximal ideals $\mathfrak{m}$ of R.

Proof. Assume first that the module M is finitely generated, say, with generators $m_1, \ldots, m_s$. Set $I = \mathrm{ann}(M)$. Then M is of finite length as a surjective image of the module $\bigoplus_{i=1}^{s} R/\mathrm{ann}(m_i)$ and hence the ring R/I is artinian, since it can be regarded as a submodule of M^s via the embedding $r \mapsto (rm_1, \ldots, rm_s)$. Therefore, the ideal I is contained in only finitely many maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_s$ of R, and for the remaining maximal ideals $\mathfrak{m}$ of R we have $M_{\mathfrak{m}} = 0$. The artinian ring R/I decomposes as the product

$$R/I \simeq \prod_{i=1}^{s} R_{\mathfrak{m}_i}/I R_{\mathfrak{m}_i}. \tag{13}$$

Since $I = \mathrm{ann}(M)$, we have $M \otimes_R R/I \simeq M$ and $M \otimes_R R_{\mathfrak{m}_i}/I R_{\mathfrak{m}_i} \simeq M_{\mathfrak{m}_i}$. Thus, tensoring (13) with M, we obtain an isomorphism

$$M \to M_{\mathfrak{m}_1} \oplus \cdots \oplus M_{\mathfrak{m}_s}.$$

Since the modules $M_{\mathfrak{m}_i}$ are also of finite length, we see that each $M_{\mathfrak{m}_i}$ is annihilated by some power of the maximal ideal $\mathfrak{m}_i$.

We now turn to the case of an arbitrary module M. Consider the canonical map

$$\Phi \colon M \to \prod_{\mathfrak{m}} M_{\mathfrak{m}},$$

the product being taken over all maximal ideals $\mathfrak{m}$ of R. Restricting Φ to finitely generated submodules $N \subseteq M$, and using the (already established) claim for finitely generated modules, we conclude that the image of Φ is in fact contained in $\bigoplus_{\mathfrak{m}} M_{\mathfrak{m}}$ and that the induced map

$$\Phi \colon M \to \bigoplus_{\mathfrak{m}} M_{\mathfrak{m}}$$

(that we continue to denote by the same letter) is an isomorphism. For a maximal ideal $\mathfrak{n}$ of R, multiplication by elements outside of $\mathfrak{n}$ is bijective on $M_{\mathfrak{n}}$. Therefore, restricting Φ to $M[\mathfrak{m}^\infty]$ shows that $M[\mathfrak{m}^\infty] = M_{\mathfrak{m}}[\mathfrak{m}^\infty]$. Finally, we conclude from the case of finitely generated modules that every element in $M_{\mathfrak{m}}$ is annihilated by some power of the maximal ideal $\mathfrak{m}$. Thus, $M[\mathfrak{m}^\infty] = M_{\mathfrak{m}}$. $\square$

Proposition 2.7. *The inseparability degree of $\sigma^n - 1$ satisfies*

$$\deg_i(\sigma^n - 1) = r_n \cdot |n|_p^{s_n} \tag{14}$$

for periodic sequences (r_n) and (s_n) with $r_n \in \mathbf{Q}^*$ and $s_n \in \mathbf{Z}$, $s_n \leqslant 0$. Furthermore, there is an integer ω such that we have

$$r_n = r_{\gcd(n,\omega)} \quad \text{for } p \nmid n.$$

Proof. The strategy of the proof is as follows: since $\deg_i(\sigma^n - 1)$ is a power of p, it is sufficient to compute $|\deg(\sigma^n - 1)|_p$ and $|\sigma_n|_p$. The former number has been already computed in Proposition 2.3(ii); for the latter, we study the p-primary torsion of A as an R-module, where, not to have to worry about

noncommutative arithmetic, we work with the ring $R = \mathbf{Z}[\sigma] \subseteq \mathrm{End}(A)$. Note that R need not be a Dedekind domain. Let $X := A(K)_{\mathrm{tor}}$ denote the subgroup of torsion points of $A(K)$. It has a natural structure of an R-module, and as an abelian group is divisible; in fact,

$$X \simeq \left(\mathbf{Z}\left[\frac{1}{p^\infty}\right]/\mathbf{Z}\right)^f \oplus \bigoplus_{q \neq p}\left(\mathbf{Z}\left[\frac{1}{q^\infty}\right]/\mathbf{Z}\right)^{2g},$$

where f is the p-rank of A, and

$$\mathbf{Z}\left[\frac{1}{q^\infty}\right] = \bigcup_{k \geqslant 1}\mathbf{Z}\left[\frac{1}{q^k}\right].$$

As R acts on X, the localization $R_{\mathfrak{m}}$ acts on $X_{\mathfrak{m}}$ for each maximal ideal $\mathfrak{m}$ of R. Since X is torsion as an abelian group, the conditions of Lemma 2.6 are satisfied, and hence we have $X_{\mathfrak{m}} = X[\mathfrak{m}^\infty]$ and

$$X = \bigoplus_{\mathfrak{m}} X_{\mathfrak{m}},$$

the sum being taken over all maximal ideals $\mathfrak{m}$ of R. For an element $\tau \in R$, we have

$$X[\tau] = \bigoplus_{\mathfrak{m}} X_{\mathfrak{m}}[\tau].$$

Since $X_{\mathfrak{m}} = X[\mathfrak{m}^\infty]$, for any prime number q we have $X_{\mathfrak{m}}[q^\infty] = 0$ if $q \notin \mathfrak{m}$ and $X_{\mathfrak{m}}[q^\infty] = X_{\mathfrak{m}}$ if $q \in \mathfrak{m}$, and hence we get

$$X[q^\infty] = \bigoplus_{q \in \mathfrak{m}} X_{\mathfrak{m}}.$$

Thus the groups $X_{\mathfrak{m}}$ for $q \in \mathfrak{m}$ are q-power torsion. It follows that for $\tau \in R$, $\tau \neq 0$, we can compute

$$|X[\tau]|_q = \prod_{q \in \mathfrak{m}} |X_{\mathfrak{m}}[\tau]|_q. \tag{15}$$

Since X is a divisible abelian group, the groups $X_{\mathfrak{m}}$, being quotients of X, are also divisible. Thus, the surjectivity of $p \colon X_{\mathfrak{m}} \to X_{\mathfrak{m}}$ implies that there is a short exact sequence

$$0 \to X_{\mathfrak{m}}[p] \to X_{\mathfrak{m}}[p\tau] \xrightarrow{p} X_{\mathfrak{m}}[\tau] \to 0. \tag{16}$$

Let σ be an element of R, let $e_{\mathfrak{m}}$ denote the order of $\bar{\sigma}$ in $(R_{\mathfrak{m}}/\mathfrak{m}R_{\mathfrak{m}})^*$ for maximal ideals $\mathfrak{m}$ of R with $p \in \mathfrak{m}$ and $\sigma \notin \mathfrak{m}$. Note that $e_{\mathfrak{m}}$ is then coprime with p. Applying (16) to $\tau = \sigma^n - 1$ and using Lemma 2.1, we get

$$|X_{\mathfrak{m}}[\sigma^{mn} - 1]|_p = \begin{cases} 1 & \text{for } \sigma \in \mathfrak{m}, \\ 1 & \text{for } \sigma \notin \mathfrak{m} \text{ and } e_{\mathfrak{m}} \nmid mn, \\ |X_{\mathfrak{m}}[\sigma^n - 1]|_p & \text{for } \sigma \notin \mathfrak{m}, \ p \nmid m \text{ and } e_{\mathfrak{m}} \mid n, \\ |X_{\mathfrak{m}}[\sigma^n - 1]|_p \cdot |X_{\mathfrak{m}}[p]|_p & \text{for } \sigma \notin \mathfrak{m}, \ m = p, e_{\mathfrak{m}} \mid n, \text{ and } \mathrm{ord}_p(n) \gg 0. \end{cases}$$

Arguing in the same way as in the proof of Proposition 2.3, we conclude that there exist periodic sequences $(r_n^{\mathfrak{m}})_n$ and $(s_n^{\mathfrak{m}})_n$ with $r_n^{\mathfrak{m}} \in \mathbf{Q}^*$ and $s_n^{\mathfrak{m}} \in \mathbf{N}$ such that

$$|X_{\mathfrak{m}}[\sigma^n - 1]|_p = r_n^{\mathfrak{m}} |n|_p^{s_n^{\mathfrak{m}}} \quad \text{for } n \geqslant 1. \tag{17}$$

Furthermore, $r_n^{\mathfrak{m}} = 1$ and $s_n^{\mathfrak{m}} = 0$ for all n if $\sigma \in \mathfrak{m}$, and

$$r_n^{\mathfrak{m}} = r_{\gcd(n, e_{\mathfrak{m}})}^{\mathfrak{m}} \quad \text{for } \sigma \notin \mathfrak{m} \text{ and } p \nmid n.$$

Applying (15) to $\tau = \sigma^n - 1$ and $q = p$, we get the equality

$$|\sigma_n|_p = \prod_{p \in \mathfrak{m}} |X_{\mathfrak{m}}[\sigma^n - 1]|_p.$$

Taking the product of the formulæ (17) over all maximal ideals $\mathfrak{m}$ of R with $p \in \mathfrak{m}$, we obtain periodic sequences $(r_n')_n$ and $(s_n')_n$ with $r_n' \in \mathbf{Q}^*$ and $s_n' \in \mathbf{N}$ such that

$$|\sigma_n|_p = r_n' |n|_p^{s_n'}$$

and

$$r_n' = r_{\gcd(n, \omega')}' \quad \text{for } p \nmid n,$$

where

$$\omega' = \operatorname{lcm}\{e_{\mathfrak{m}} \mid \sigma \notin \mathfrak{m}\}.$$

Writing

$$\deg_{\mathrm{i}}(\sigma^n - 1) = \frac{\deg(\sigma^n - 1)}{\sigma_n} = \frac{|\sigma_n|_p}{|\deg(\sigma^n - 1)|_p}$$

and using Proposition 2.3(ii), we get sequences (r_n) and (s_n) satisfying all stated properties except that it might be that $s_n > 0$ for some n. However, since $\deg_{\mathrm{i}}(\sigma^n - 1)$ is an integer, letting ϖ be the common period of (r_n) and (s_n), we automatically get $s_n \leqslant 0$ for all n such that the arithmetic sequence $n + \varpi \mathbf{N}$ contains terms divisible by arbitrarily high powers of p. For all the remaining n we have $\operatorname{ord}_p(n) < \operatorname{ord}_p(\varpi)$, and thus whenever $s_n > 0$, we replace s_n by 0 and r_n by $r_n |n|_p^{s_n}$, obtaining the claim. $\qquad\square$

3. A holonomic version of the Hadamard quotient theorem

The next proposition is our basic tool from the theory of recurrent sequences. It bears some resemblance to the Hadamard quotient theorem (which is used in its proof), and to conjectural generalizations of it as proposed by Bellagh and Bézivin [2011, "Question" in §1] (using holonomicity instead of linear recurrence) and Dimitrov [2013, Conjecture in 1.1] (using algebraicity instead of linear recurrence). In our special case, the proof relies on the quotient sequence having a specific form.

Proposition 3.1. *Let* $(a_n)_{n \geqslant 1}$, $(b_n)_{n \geqslant 1}$, $(c_n)_{n \geqslant 1}$ *be sequences of nonzero complex numbers such that*

$$a_n = b_n c_n$$

for all n. *Assume that*:

(i) $(a_n)_{n \geqslant 1}$ *satisfies a linear recurrence.*

(ii) $(b_n)_{n \geqslant 1}$ *is holonomic.*

(iii) $(c_n)_{n \geqslant 1}$ *is of the form* $c_n = r_n |n|_p^{s_n}$ *for a prime* p *and periodic sequences* $(r_n)_{n \geqslant 1}$, $(s_n)_{n \geqslant 1}$ *with* $r_n \in \mathbf{Q}^*$, $s_n \in \mathbf{Z}$.

Then the sequence $(c_n)_{n \geqslant 1}$ *is bounded.*

Proof. Note that $c_n \neq 0$ for all n. Since the sequence $(b_n)_{n \geqslant 1}$ given by $b_n = a_n/c_n$ is holonomic, by Lemma 1.2(iii) there exist polynomials $q_0, \ldots, q_d \in \mathbf{C}[z]$ such that

$$q_0(n)\frac{a_n}{c_n} = -\sum_{i=1}^{d} q_i(n+i)\frac{a_{n+i}}{c_{n+i}}, \quad \text{for } n \geqslant 1. \tag{18}$$

We may further assume that $q_0 \neq 0$ (otherwise, replace for $i = 1, \ldots, d$ the polynomials q_i by $(z-1)q_i$ and shift the relation by one). Suppose $c_n = r_n |n|_p^{s_n}$ is not bounded and let ϖ be the common period of both (r_n) and (s_n). The unboundedness of $(c_n)_{n \geqslant 1}$ means that there exists an integer $j \geqslant 1$ with $s_j < 0$ such that there are elements in the arithmetic sequence $\{j + \varpi n \mid n \geqslant 0\}$ which are divisible by an arbitrarily high power of p. Fix such j and write $s := s_j$. Let ν be an integer such that $p^\nu > \max(d, \varpi)$ and let $\Pi = \mathrm{lcm}(\varpi, p^\nu)$. Note that $\mathrm{ord}_p \Pi = \nu$. By the assumption on $\{j + \varpi n \mid n \geqslant 0\}$, there exists an integer J such that $J \equiv j \pmod{\varpi}$ and $J \equiv 0 \pmod{p^\nu}$. By the definition of the sequence $(c_n)_{n \geqslant 1}$, for $n \equiv J \pmod{\Pi}$ the values $c_{n+1}, \ldots, c_{n+d}$ are uniquely determined (i.e., do not depend on n). Substituting such n into (18), we obtain a formula of the form

$$\frac{a'_n}{|n|_p^s} = b'_n \quad \text{for } n \equiv J \pmod{\Pi},$$

where

$$a'_n = q_0(n)\frac{a_n}{r_j} \quad \text{and} \quad b'_n = -\sum_{i=1}^{d} q_i(n+i)\frac{a_{n+i}}{c_{n+i}}$$

are linear recurrence sequences along the arithmetic sequence $n \equiv J \pmod{\Pi}$ (here we use the fact that the values $c_{n+1}, \ldots, c_{n+d}$ do not depend on n, and that linear recurrence sequences form an algebra). Note that the values of $(a'_n)_{n \geqslant 1}$ are nonzero for sufficiently large n, and hence so are $(b'_n)_{n \geqslant 1}$. By Lemma 1.2(i), a subsequence of a linear recurrence sequence along an arithmetic sequence is a linear recurrence sequence. Since the sequence

$$|n|_p^s = \frac{a'_n}{b'_n}$$

takes values in a finitely generated ring (namely $\mathbf{Z}[1/p]$), we conclude from the Hadamard quotient theorem [Rumely 1988; van der Poorten 1988, Théorème] that the sequence $(|J + \Pi n|_p^s)_{n \geqslant 0}$ satisfies a linear recurrence, say

$$\gamma_0|J + \Pi n|_p^s + \gamma_1|J + \Pi(n + 1)|_p^s + \cdots + \gamma_e|J + \Pi(n + e)|_p^s = 0, \quad \text{for } n \text{ large enough,} \qquad (19)$$

where $\gamma_0, \ldots, \gamma_e \in \mathbf{C}$, $\gamma_0 \neq 0$. Let μ be an integer such that $p^\mu > \Pi d$. Since $\nu = \mathrm{ord}_p(\Pi) \leqslant \mathrm{ord}_p(J)$, we can find an integer $\Pi' > 0$ such that $\Pi\Pi' \equiv -J \pmod{p^\mu}$. Then for $n \equiv \Pi' \pmod{p^{\mu-\nu}}$ the values of

$$|J + \Pi(n + 1)|_p^s, \ldots, |J + \Pi(n + e)|_p^s$$

are independent of n (actually, $|J + \Pi(n + j)|_p^s = p^{-\nu s}|j|_p^s$ for $j = 1, \ldots, e$), and hence by (19) so is the value of $\gamma_0|J + \Pi n|_p^s$ for n sufficiently large. Substituting $n = \Pi' + ip^{\mu-\nu}$ with $i = 0, \ldots, p-1$, we get a contradiction, since there is exactly one value of i for which $|J + \Pi(\Pi' + ip^{\mu-\nu})|_p^s < p^{-\mu s}$. $\qquad \square$

4. Rationality properties of dynamical zeta functions

We prove a general rational/transcendental dichotomy in terms of the following arithmetical property:

Definition 4.1. An endomorphism $\sigma \in \mathrm{End}(A)$ is called *very inseparable* if $\sigma^n - 1$ is a separable isogeny for all n.

Note that the zero map is very inseparable. The notion "very inseparable" makes sense for arbitrary (not necessarily confined) endomorphisms, but such very inseparable endomorphisms are then automatically confined. We will study the geometric meaning of very inseparability in greater detail in Section 6; here we content ourselves with discussing the case of elliptic curves.

Example 4.2. If $A = E$ is an elliptic curve, things simplify greatly (compare [Bridy 2016, §5]): there exists a (nonarchimedean) absolute value $|\cdot|$ on the ring $\mathrm{End}(E)$ such that $\deg_i(\tau) = |\tau|^{-1}$ for $\tau \in \mathrm{End}(E)$. It is immediate that inseparable isogenies together with the zero map form an ideal in $\mathrm{End}(E)$ and that an inseparable isogeny σ (i.e., $|\sigma| < 1$) is very inseparable (i.e., $|\sigma^n - 1| = 1$ for all n). Neither of these statements is true in general for higher-dimensional abelian varieties.

Theorem 4.3. (i) *If σ is very inseparable, then $\zeta_\sigma(z) \in \mathbf{Q}(z)$ is rational.*

(ii) *If σ is not very inseparable, then the sequence (σ_n) is not holonomic and $\zeta_\sigma(z)$ is transcendental over $\mathbf{C}(z)$.*

Proof. Suppose we are in case (i), so $\sigma^n - 1$ is separable for all n. Since $\sigma_n = \deg(\sigma^n - 1)$, Proposition 2.3(i) implies that $\zeta_\sigma(z)$ is a rational function of z.

In case (ii), set $a_n = \deg(\sigma^n - 1)$, $b_n = \sigma_n$, and $c_n = \deg_i(\sigma^n - 1)$. By Proposition 2.3(i), (a_n) satisfies a linear recurrence. By Proposition 2.7, $c_n = r_n|n|_p^{s_n}$ for periodic $r_n \in \mathbf{Q}^*$ and $s_n \in \mathbf{Z}$. Assume, by contradiction, that b_n is holonomic, i.e., that the sequence (b_n) is holonomic. The sequences (a_n), (b_n), and (c_n) then satisfy all the conditions of Proposition 2.7, and we conclude that the sequence (c_n) is bounded. However, the following proves that (c_n) is unbounded:

Lemma 4.4. *If σ is not very inseparable, then the sequence $\deg_i(\sigma^n - 1)$ is unbounded.*

Proof. By assumption, there exists n_0 for which $\sigma^{n_0} - 1$ is inseparable. Write $\sigma^{n_0} = 1 + \psi$ with ψ inseparable; then

$$\sigma^{n_0 p} - 1 = (1 + \psi)^p - 1 = \psi(\psi^{p-1} + p\chi),$$

for some endomorphism $\chi : A \to A$. Since p has identically zero differential, the map $\psi^{p-1} + p\chi$ is inseparable, and hence

$$\deg_i(\sigma^{n_0 p} - 1) \geqslant 1 + \deg_i(\psi) = 1 + \deg_i(\sigma^{n_0} - 1),$$

and the result follows by iteration. □

To show the transcendence of $\zeta_\sigma(z)$ over $\mathbf{C}(z)$, suppose it is algebraic. Then so would be

$$z\frac{\zeta'_\sigma(z)}{\zeta_\sigma(z)} = z(\log(\zeta_\sigma(z)))' = \sum \sigma_n z^n.$$

This contradicts the fact that σ_n is not holonomic. □

Corollary 4.5. *At most one of the functions*

$$\zeta_\sigma(z) = \exp\left(\sum_{n \geqslant 1} \sigma_n \frac{z^n}{n}\right) \quad and \quad \frac{1}{\zeta_\sigma(z)} = \exp\left(\sum_{n \geqslant 1} -\sigma_n \frac{z^n}{n}\right)$$

is holonomic.

Proof. Assume that both these functions are holonomic. Since the class of holonomic functions is closed under taking the derivative and the product [Stanley 1980, Theorem 2.3], we conclude that $z\zeta'_\sigma(z)/\zeta_\sigma(z)$ is holonomic, contradicting Theorem 4.3(ii). □

Remark 4.6. It is not true that the multiplicative inverse of a holonomic function is necessarily holonomic. Harris and Shibuya [1985] proved that this happens precisely if the logarithmic derivative of the function is algebraic. We do not know whether $\zeta_\sigma(z)$ is holonomic for not very inseparable σ, but Theorem 5.5 will show that $\zeta_\sigma(z)$ is not holonomic for a large class of maps.

Remark 4.7. If σ is not assumed to be confined, we could change the definition of σ_n by considering σ_n to be the number of fixed points of σ^n whenever it is finite, and 0 otherwise. This is in the spirit of [Artin and Mazur 1965], where only isolated fixed points of diffeomorphisms of manifolds were considered. In this case, we could still prove a variant of Theorem 4.3 saying that if σ is a (not necessarily confined) endomorphism of A such that there exist n such that $\sigma^n - 1$ is an isogeny of arbitrarily high inseparability degree, then (σ_n) is not holonomic; one needs to use the fact that (the proof of) Proposition 3.1 holds even if we do not insist that a_n and b_n be nonzero and instead demand that $c_n = 1$ if $a_n = 0$. Note, however, that without the assumption that σ is confined, $\zeta_\sigma(z)$ could be an algebraic but not rational function. For example, let E be a supersingular elliptic curve over a field of characteristic 2, let $A = E \times E$, and $\sigma = [2] \times [-1]$. Then

$$\zeta_\sigma(z) = \frac{1 - 2z}{1 + 2z}\sqrt{\frac{(1+z)(1+4z)}{(1-z)(1-4z)}}.$$

5. Complex analytic aspects

We now turn to questions of convergence and analytic continuation.

Radius of convergence. From the proof of Proposition 2.3, we pick up the formula

$$\deg(\sigma^n - 1) = \prod_{i=1}^{q}(\xi_i^n - 1) = \sum_{i=1}^{r} m_i \lambda_i^n, \tag{20}$$

where we note for future use that $q = 2g$, $\prod_{i=1}^{q} \xi_i = \deg(\sigma)$, and λ_i are of the form $\lambda_i = \prod_{j \in I} \xi_j$ for some $I \subseteq \{1, \ldots, q\}$, each occurring with sign $(-1)^{|I|}$. Recall that $\{\lambda_i\}$ are called the *roots* of the linear recurrence, and λ_i is called a *dominant root* if it is of maximal absolute value amongst the roots. The roots $\{\lambda_i\}$ of the recurrence should not be confused with the roots $\{\xi_i\}$ of the characteristic polynomial of σ on H^1 (the dual of the ℓ-adic Tate module for any choice of $\ell \neq p$).

The following proposition follows from (20) and the fact that $\deg(\sigma^n - 1)$ takes only positive values.

Proposition 5.1.   (i) *The ξ_i are not roots of unity.*

(ii) *The linear recurrent sequence $\deg(\sigma^n - 1)$ has a dominant positive real root, denoted Λ.*

(iii) $\Lambda = \prod_{i=1}^{q} \max\{|\xi_i|, 1\} \geqslant 1$ *is the Mahler measure of the characteristic polynomial of σ acting on H^1.*

(iv) $\Lambda = 1$ *if and only if σ is nilpotent.*

(v) $\deg(\sigma^n - 1)$ *has a unique dominant root if and only if there is no ξ_i with $|\xi_i| = 1$.*

(vi) *If $\deg(\sigma^n - 1)$ has a unique dominant root Λ, then Λ has multiplicity 1.*

Proof. (i) This is clear since σ is confined.

(ii) If not, $\deg(\sigma^n - 1)$ would be negative infinitely often by a result of Bell and Gerhold [2007, Theorem 2].

(iii) Denote temporarily $\tilde{\Lambda} = \prod_{i=1}^{q} \max\{|\xi_i|, 1\}$. We will prove shortly that $\tilde{\Lambda} = \Lambda$. Formula (20) implies that $\Lambda \leqslant \tilde{\Lambda}$ and

$$a_1(n) := \sum_{|\lambda_j| = \tilde{\Lambda}} m_j \lambda_j^n$$

equals

$$a_1(n) = (-1)^t P^n \prod_{j \in J}(\xi_j^n - 1), \tag{21}$$

where t is the number of indices i such that $|\xi_i| < 1$, $P := \prod_{|\xi_i| > 1} \xi_i$, and $J \subseteq \{1, \ldots, q\}$ denotes the set of indices i such that $|\xi_i| = 1$. Since the right hand side of (21) is nonzero, we conclude that $\tilde{\Lambda} = \Lambda$. Finally, by Remark 2.4, ξ_i are the roots of the indicated characteristic polynomial.

(iv) Since none of the ξ_i is a root of unity, and since the set $\{\xi_i\}$ is closed under Galois conjugation, Kronecker's theorem implies that either some ξ_i has absolute value $|\xi_i| > 1$, in which case $\Lambda > 1$, or else all ξ_i are 0. The latter is equivalent to σ acting nilpotently on H^1, and hence σ is nilpotent since $\mathrm{End}(A)$ embeds into (the opposite ring of) $\mathrm{End}(H^1)$.

(v) From (21) we immediately get that if $J = \varnothing$, then $\deg(\sigma^n - 1)$ has a unique dominant root. Conversely, if $J \neq \varnothing$, then substituting $n = 0$ into (21) gives $\sum m_j = 0$, and hence in the formula there are at least two distinct values of λ_j occurring, and the dominant root is not unique.

(vi) We have already proved that if there is a unique dominant root, then $J = \varnothing$. Thus we read from (21) that the multiplicity of Λ is ± 1. Since $\deg(\sigma^n - 1)$ takes only positive values, the multiplicity is in fact 1. $\square$

Proposition 5.2. *The radius of convergence of the power series defining $\zeta_\sigma(z)$ is $1/\Lambda > 0$.*

Proof. Note first that we have a trivial bound $\sigma_n = O(\Lambda^n)$, which implies that the power series $\zeta_\sigma(z)$ is majorized by $\exp(\sum_{n \geqslant 1} C\Lambda^n z^n / n) = (1 - \Lambda z)^{-C}$ for some constant $C > 0$. Thus the radius of convergence of $\zeta_\sigma(z)$ is at least $1/\Lambda$. If σ is nilpotent, the maps $\sigma^n - 1$ are all invertible, and hence $\sigma_n = 1$ and $\zeta_\sigma(z) = 1/(1 - z)$. Assume thus that σ is not nilpotent, and hence by Proposition 5.1(iv), $\Lambda > 1$.

For the other inequality, we write the linear recurrence sequence $\deg(\sigma^n - 1) = \sum_{i=1}^r m_i \lambda_i^n$ as the sum of two linear recurrence sequences $a_1(n)$ and $a_2(n)$, $a_1(n)$ as in (21) containing the terms with λ_i of absolute value $\tilde{\Lambda} = \Lambda$, and $a_2(n)$ containing the terms where λ_i is of strictly smaller absolute value.

Since all ξ_j with $j \in J$ are algebraic numbers on the unit circle but not roots of unity, a theorem of Gel'fond [1960, Theorem 3] implies that for any $\varepsilon > 0$ and $n = n(\varepsilon)$ sufficiently large,

$$\prod_{j \in J} |\xi_j^n - 1| > \Lambda^{-n\varepsilon}$$

and hence $|a_1(n)| > \Lambda^{n(1-\varepsilon)}$ for sufficiently large n. The formula in Proposition 2.7 implies that $\deg_i(\sigma^n - 1) = O(n^s)$ for some integer s, and hence it follows from (6) that $\sigma_n > \Lambda^{n(1-2\varepsilon)}$ for sufficiently large n. For the lower bound, analogous reasoning proves that the radius of convergence of $\zeta_\sigma(z)$ is at most $1/\Lambda^{1-2\varepsilon}$, implying the claim. $\square$

Remark 5.3. The value $\log \Lambda$ describes the growth rate of the number of periodic points and plays the role of entropy as defined in the presence of a topology or a measure. It is the logarithm of the spectral radius of σ acting on the total (ℓ-adic) cohomology of A — even in the not very inseparable case — as in a result of Friedland's [1991] in the context of complex dynamics.

The degree zeta function. The degree zeta function $D_\sigma(z)$ is a rational function, and hence admits a meromorphic continuation to the entire complex plane. Actually,

$$D_\sigma(z) = \prod_{i=1}^r (1 - \lambda_i z)^{-m_i},$$

written in terms of the parameters in (20), immediately provides the extension. Poles (with multiplicity m_i) occur at $1/\lambda_i$ with $m_i > 0$; zeros (with multiplicity m_i) occur at $1/\lambda_i$ with $m_i < 0$. We may describe the behavior of zeros and poles more precisely.

Proposition 5.4. *Assume that σ is not nilpotent. Let $\Lambda' := \max\{|\lambda_i| : |\lambda_i| < \Lambda\} < \Lambda$.*

   (i) *The function $D_\sigma(z)$ has a pole at $1/\Lambda$.*

(ii) *The function $D_\sigma(z)$ has a zero z_0 with $|z_0| = 1/\Lambda'$ and is holomorphic in the annulus $1/\Lambda < |z| < 1/\Lambda'$.*

(iii) $\Lambda' \geqslant \sqrt{\Lambda}$.

Proof. In order to prove (i), we need to show that the multiplicity m of Λ is positive. If Λ is a dominant root, this follows from Proposition 5.1(vi). If Λ is not a dominant root and $m < 0$, the sequence $\deg(\sigma^n - 1) - m\Lambda^n$ is a linear recurrent sequence with positive values and no dominant positive real root, contradicting [Bell and Gerhold 2007, Theorem 2].

Let us now prove (ii). Let ρ denote the minimal value of $|\xi_i|$ and $|\xi_i|^{-1}$ that is strictly larger than 1, i.e.,

$$\rho = \min(\min\{|\xi_i| : |\xi_i| > 1\}, \min\{|\xi_i|^{-1} : 0 < |\xi_i| < 1\});$$

it exists since by Proposition 5.1(iv), $\Lambda > 1$. Write the set of indices $\{1, \ldots, q\} = J_<^- \cup J^- \cup J \cup J^+ \cup J_>^+$, where membership $i \in J_*^*$ is defined by the corresponding condition in the second row of the following table:

$J_<^-$	J^-	J	J^+	$J_>^+$
$|\xi_i| < \rho^{-1}$	$|\xi_i| = \rho^{-1}$	$|\xi_i| = 1$	$|\xi_i| = \rho$	$|\xi_i| > \rho$

From (20) we see that there is no λ_j with $\Lambda/\rho < |\lambda_j| < \Lambda$ and that the terms λ_j with $|\lambda_j| = \Lambda/\rho$ arise as products $\prod_{i \in I} \xi_i$ where I contains $J_>^+$, I is disjoint from $J_<^-$, $I \cap J$ can be anything and *either* I contains all except one $i \in J^+$ *or* I contains all $i \in J^+$ and exactly one $i \in J^-$.

Setting as before $P := \prod_{i \in J^+ \cup J_>^+} \xi_i$ and $t = \#(J_<^- \cup J^-)$, we get

$$\sum_{|\lambda_j| = \Lambda/\rho} m_j \lambda_j^n = (-1)^{t-1} P^n \prod_{j \in J}(\xi_j^n - 1)\left(\sum_{i \in J^+} \xi_i^{-n} + \sum_{i \in J^-} \xi_i^n\right). \tag{22}$$

Since the right-hand side is not identically zero as a function of n, we conclude that $\Lambda' = \Lambda/\rho$. We consider two cases.

Case 1: $J = \varnothing$. Then by Proposition 5.1(vi), $P = \Lambda$ has multiplicity 1 and hence from (21) we conclude that t is even. Therefore by (22) all λ_i with $|\lambda_i| = \Lambda'$ have multiplicity $m_i < 0$, and hence correspond to zeros of $D_\sigma(z)$.

Case 1: $J \neq \varnothing$. Substituting $n = 0$ into (21) shows that the sum of multiplicities m_i of λ_i with $|\lambda_i| = \Lambda$ is 0. By (22), the same is true for multiplicities m_j of λ_j with $|\lambda_j| = \Lambda'$. Thus there is some λ_i with $|\lambda_i| = \Lambda'$ and $m_i < 0$.

For the proof of (iii), note that since $\Lambda' = \Lambda/\rho$, the stated inequality is equivalent to $\Lambda \geqslant \rho^2$. Since $\Lambda = \prod \max\{|\xi_i|, 1\}$, it is enough to prove that there are at least two elements in the (nonempty) set $J^+ \cup J_>^+$. Since $q = 2g$ is even, it suffices to prove that both $\#J$ and $t = \#(J^- \cup J_<^-)$ are even. Since ξ_i with $|\xi_i| = 1$ occur in complex conjugate pairs, $\#J$ is even, and the corresponding term in (21) is real positive. In the course of proof of Proposition 5.2 we have shown that the sum $a_1(n)$ dominates the remaining terms, and hence is positive for large n. Hence we find from (21) that $P > 1$ and t is even. $\square$

Analytic continuation/natural boundary. When σ is very inseparable, $\zeta_\sigma(z)$ coincides with the degree zeta function $D_\sigma(z)$ and hence is a rational function. One may wonder whether a Pólya–Carlson dichotomy holds for the functions $\zeta_\sigma(z)$, meaning that, when they are not rational as above, they admit a natural boundary as complex function (and hence they are nonholonomic; in this context also called "transcendentally transcendental").

We confirm this for a large class of such maps, providing at the same time another proof of their transcendence (and even nonholonomicity). The crucial tool is Theorem A.1 that Royals and Ward prove in the Appendix of this paper.

Theorem 5.5. *Suppose that σ is not very inseparable and that Λ is the unique dominant root. Then the function $\zeta_\sigma(z)$ has the circle $|z| = 1/\Lambda$ as its natural boundary. In particular, $\zeta_\sigma(z)$ is not holonomic.*

Proof. We start by the observation that $\zeta_\sigma(z)$ has the same natural boundary as $Z_\sigma(z) := \sum \sigma_n z^n$ if the latter function has natural boundary [Bell et al. 2014, Lemma 1]. Next, we find an expression

$$Z_\sigma(z) = \sum_{i=1}^{r} m_i \sum_{n \geqslant 1} r_n^{-1} |n|_p^{-s_n} (\lambda_i z)^n,$$

where m_i and λ_i are as in (10) and r_n and s_n are as in Proposition 2.7. We now apply Theorem A.1: in the notation of that theorem, we choose S to be the set of primes containing p and all primes ℓ for which $|r_n|_\ell \neq 1$ for some n. By periodicity of (r_n), the set S is finite. Let $a_n := \deg_i(\sigma^n - 1) = r_n |n|_p^{s_n}$. Suppose ϖ is a common period for (r_n) and (s_n). For $\ell \in S$, set $n_\ell = \varpi$, $c_{\ell,k} = |r_k|_\ell$; for $\ell \neq p$, set $e_{\ell,k} = 0$, and set $e_{p,k} = -s_k$. Then $|a_n|_S = a_n^{-1}$, and hence we can write

$$Z_\sigma(z) = \sum_{i=1}^{r} m_i f(\lambda_i z),$$

where f is the function associated to (a_n) as in Theorem A.1. Since σ is not very inseparable, by Lemma 4.4 the sequence (a_n) takes infinitely many values. We find that the term $f(\lambda_i z)$ has a natural boundary along $|z| = 1/|\lambda_i|$. If Λ is the *unique* λ_i of maximal absolute value, then the dense singularities along this circle cannot be canceled by other terms, and we conclude that $Z_\sigma(z)$ has a natural boundary along $|z| = 1/\Lambda$, and the same holds for $\zeta_\sigma(z)$. Since a holonomic function has only finitely many singularities (corresponding to the zeros of $q_0(z)$ if the series function satisfies (5), compare to [Flajolet et al. 2004/06, Theorem 1]), $\zeta_\sigma(z)$ cannot be holonomic. □

Question 5.6. Is $|z| = 1/\Lambda$ a natural boundary for $\zeta_\sigma(z)$ for any not very inseparable σ (even without the assumption of a unique dominant root)?

Metrizable group endomorphisms with the same zeta function. Given the analogy between our results and some properties of metrizable group endomorphisms, one may ask for the following more formal relationship:

Question 5.7. Can one associate to an action of $\sigma \circlearrowright A$ an endomorphism of a compact metrizable abelian group $\tau \circlearrowright G$ with the same Artin–Mazur zeta function, i.e., $\zeta_\sigma = \zeta_\tau$?

The analogue of this question over the complex numbers is trivial, as one may take $G = A(\mathbf{C})$. The degree zeta function $D_\sigma(z)$ artificially equals the Artin–Mazur zeta function of an endomorphism τ of a $2g$-dimensional real torus whose matrix has the same characteristic polynomial as that of σ acting on $T_\ell(A)$ for any $\ell \neq p$ (e.g., the companion matrix). This implies that for a very inseparable $\sigma \circlearrowright A$, indeed, $\zeta_\sigma(z) = \zeta_\tau(z)$.

Even in the not very inseparable case, it is sometimes possible to construct such $\tau \circlearrowright G$, like we did for the example in the introduction.

In general, it would be natural to consider the induced action of σ on the torsion subgroup $A(K)_{\mathrm{tor}}$ (dual of the total Tate module $\prod T_\ell(A)$). This provides the correct contribution $|\sigma_n|_\ell$ at all primes $\ell \neq p$; for such ℓ, the size of the cokernel of $\sigma^n - 1$ acting on $T_\ell(A)$ is precisely $|\sigma_n|_\ell^{-1}$. However, at $\ell = p$, we found no such natural group in general, and it seems that $|\sigma_n|_p$ is genuinely determined by the geometry of the p-torsion subgroup scheme.

6. Geometric characterization of very inseparable endomorphisms

In this section, we analyze the condition of very inseparability from a geometric point of view as well as its relation to inseparability. For this, it is advantageous to *temporarily drop the assumption of confinedness* and consider a general $\sigma \in \mathrm{End}(A)$.

Elementary properties. We start by listing properties of very inseparability that follow more or less directly from the definition. For this, we first write out a very basic property:

Lemma 6.1. *Whether $\sigma \in \mathrm{End}(A)$ is a separable isogeny or not is determined by its action on the finite commutative group scheme $A[p]$, i.e., by its image under the map $\mathrm{End}(A) \to \mathrm{End}(A[p])$.*

Proof. If two endomorphisms $\sigma, \tau \colon A \to A$ induce the same map on $A[p]$, then $\sigma - \tau$ vanishes on the group scheme $A[p]$, and hence it factors through the map $[p] \colon A \to A$. Thus $\sigma - \tau = p\nu$ for some $\nu \colon A \to A$, and hence the map $\mathrm{End}(A)/p\,\mathrm{End}(A) \hookrightarrow \mathrm{End}(A[p])$ is injective. Since an endomorphism $A \to A$ is a separable isogeny if and only if it induces an isomorphism on the tangent space, and since every map of the form $p\nu$ induces the zero map on the tangent space, we conclude that σ is a separable isogeny if and only if τ is a separable isogeny. $\qquad\square$

Proposition 6.2. *Let $\sigma \in \mathrm{End}(A)$.*

(i) *The endomorphism σ is very inseparable if and only if $\sigma^n - 1$ is a separable isogeny for all $n \leqslant p^{4g^2}$.*

(ii) *If $A = A_1 \times A_2$ with A_1 and A_2 abelian varieties and $\sigma = \sigma_1 \times \sigma_2$ with $\sigma_i \in \mathrm{End}(A_i)$, then σ is very inseparable if and only if σ_1 and σ_2 are both very inseparable.*

(iii) *Multiplication $[m] \colon A \to A$ by an integer m is very inseparable if and only if m is divisible by p.*

(iv) *An endomorphism of an elliptic curve is very inseparable if and only if it is either an inseparable isogeny or zero.*

(v) *If E is an elliptic curve over a field of characteristic* 3, *then the isogeny* $\sigma := [2] \times [3]$ *on* $A := E \times E$ *is inseparable but not very inseparable.*

Proof. To prove (i), observe that by Lemma 6.1, it suffices to look at the images of $\sigma^n - 1$ in the ring $\operatorname{End}(A)/p\operatorname{End}(A)$. Since $\operatorname{End} A$ is finite free of rank at most $4g^2$, this ring is finite of cardinality $\leqslant p^{4g^2}$, and hence the sequence of images of $\sigma^n - 1$ is ultimately periodic (i.e., periodic except for a finite number of n) with all possible values already occurring for $n \leqslant p^{4g^2}$.

Property (ii) is immediate from the definition.

Since an endomorphism of an abelian variety is a separable isogeny if and only if its differential is surjective, to prove (iii), observe that the differential of the multiplication by $m^n - 1$ map is still given by multiplication by $m^n - 1$ and hence is surjective if and only if it is nonzero, i.e., when p does not divide $m^n - 1$. The latter happens for all $n \geqslant 1$ if and only if $p \mid m$.

Statement (iv) was already discussed in Example 4.2.

Property (v) follows immediately from (ii) and (iii). □

Using the local group scheme $A[p]^0$. The category of finite commutative group schemes over K is abelian and decomposes as the product of the category of finite étale and the category of finite local group schemes (see, e.g., [Goren 2002, A§4]). The group scheme $A[p]$ decomposes canonically as the product of the étale part $A[p]_{\text{ét}}$ and the local part $A[p]^0$. We now provide a geometric characterization of (very) inseparability using the local p-torsion subgroup scheme, as in Theorem A in the introduction.

Theorem 6.3. *Let $\sigma \in \operatorname{End}(A)$.*

(i) *σ is a separable isogeny if and only if it induces an isomorphism on $A[p]^0$.*

(ii) *σ is very inseparable if and only if it induces a nilpotent map on $A[p]^0$.*

Proof. Under the splitting $A[p] = A[p]_{\text{ét}} \times A[p]^0$, the morphism $\sigma[p]$ induced by σ on $A[p]$ splits as a product morphism $\sigma[p] = \sigma[p]_{\text{ét}} \times \sigma[p]^0$. Therefore, we have

$$\ker \sigma[p] = \ker \sigma[p]_{\text{ét}} \times \ker \sigma[p]^0. \tag{23}$$

An isogeny σ is separable if and only if $\ker \sigma$ is étale.

We turn to the proof of (i). In one direction, first assume that σ is a separable isogeny. Then $\ker \sigma$ is étale, and hence so is its subgroup scheme $\ker \sigma[p]$. From the decomposition (23), we conclude that $\ker \sigma[p]^0$ is both étale and local, hence trivial. Since $A[p]^0$ is a finite group scheme, the map $\sigma[p]^0$ is an isomorphism.

For the other direction, assume first that σ is *not an isogeny*. Let B be the reduced connected component of 0 of $\ker \sigma$. Then B is an abelian subvariety, $B[p]^0$ is a nontrivial group scheme (because multiplication by p on B is not étale) and is contained in the kernel of $\sigma[p]^0$ and hence $\sigma[p]^0$ is not an isomorphism.

Secondly, assume that σ is an *inseparable isogeny*. Then $\ker \sigma$ is not étale. We have $\ker \sigma \subseteq A[n]$ for $n = \deg \sigma$. Writing $n = p^t u$ with u coprime with p, we get a decomposition $\ker \sigma = \ker \sigma[p^t] \times \ker \sigma[u]$.

The group scheme $\ker \sigma[u]$ is étale (as a subgroup scheme of $A[u]$), and hence $\ker \sigma[p^t]$ cannot be étale, which means that $\ker \sigma[p^t]^0$ is nontrivial. For each integer r, we have an exact sequence

$$0 \to \ker \sigma[p^{r-1}]^0 \to \ker \sigma[p^r]^0 \xrightarrow{p^{r-1}} \ker \sigma[p]^0.$$

Applying this inductively for $r = t, t-1, \ldots, 2$, we conclude that $\ker \sigma[p]^0$ is nontrivial, and hence the morphism $\sigma[p]^0$ is not an isomorphism. This proves (i).

For the proof of (ii), consider the natural homomorphism $\varphi \colon \operatorname{End}(A) \to \operatorname{End}(A[p]^0)$. Since $\operatorname{End}(A)$ is a finite $\mathbf{Z}$-algebra, and since $p \in \ker \varphi$, the ring $R := \operatorname{im}(\varphi)$ is a finite $\mathbf{F}_p$-algebra. By part (i), the map $\sigma^n - 1$ is a separable isogeny if and only if its image $\varphi(\sigma^n - 1)$ is a unit in $\operatorname{End}(A[p]^0)$. We claim that $\varphi(\sigma^n - 1)$ is then a unit in R; in fact, the ring R is a finite $\mathbf{F}_p$-algebra, and hence there exists a monic polynomial $f \in \mathbf{F}_p[t]$, $f = t^d + a_{d-1}t^{d-1} + \cdots + a_0$, of lowest degree such that $f(\sigma^n - 1) = 0$. If the constant term a_0 of f is different than zero, then we easily see that $\sigma^n - 1$ is invertible in R, its inverse being $-a_0^{-1} \sum_{i=0}^{d-1} (\sigma^n - 1)^i$. If on the other hand $a_0 = 0$, then $\sigma^n - 1$ is a two-sided zero-divisor in R, hence in $\operatorname{End}(A[p]^0)$, and therefore cannot be a unit in $\operatorname{End}(A[p]^0)$. Thus, our claim is now reduced to the proof of the following lemma. $\qquad \square$

Lemma 6.4. *Let R be a finite* (*not necessarily commutative*) $\mathbf{F}_p$*-algebra and let $r \in R$. Then the following conditions are equivalent*:

(i) *For all positive integers $r^n - 1$ is invertible.*

(ii) *The element r is nilpotent.*

Proof. Let J denote the Jacobson radical of R. The ring R is artinian and hence the ring $\overline{R} = R/J$ is semisimple [Lam 1991, 4.14]. For an element $s \in R$, denote the image of s in $\overline{R}$ by $\bar{s}$. Then s is invertible in R if and only if $\bar{s}$ is invertible in $\overline{R}$ [Lam 1991, 4.18] and s is nilpotent if and only if $\bar{s}$ is nilpotent (this follows from the fact that the Jacobson radical of an artinian ring is nilpotent, see [Lam 1991, 4.12]). Thus we have reduced the claim to the case of a semisimple ring $\overline{R}$.

By the Wedderburn–Artin theorem [Lam 1991, 3.5], a semisimple ring is a product of matrix rings over division rings which in our case need to be finite, and hence by another theorem of Wedderburn [Lam 1991, 13.1] are commutative. Thus we can decompose the ring $\overline{R}$ as a product of matrix rings over finite fields

$$\overline{R} \simeq \prod_{i=1}^{s} \mathrm{M}_{n_i}(\mathbf{F}_{q_i}).$$

Clearly, each of the properties in the statement of the lemma can be considered separately for each term in this product, and we are reduced to proving that a matrix N over a finite field has the property that $N^n - 1$ is invertible for all $n \geqslant 1$ if and only if N is nilpotent.

If N is nilpotent, then all the matrices $N^n - 1$ are invertible, since in any ring the sum of a unit and a nilpotent that commute with each other is a unit. Conversely, if N is not nilpotent, then N has some eigenvalue $\lambda \neq 0$, perhaps in a larger (but still finite) field. Let $n \geqslant 1$ be such that $\lambda^n = 1$ (such n always exists in a *finite* field). Then the matrix $N^n - 1$ is not invertible. $\qquad \square$

We have some immediate corollaries (where Corollary 6.5(i) refines Lemma 6.1):

Corollary 6.5. *Let $\sigma \in \mathrm{End}(A)$.*

(i) *Whether σ is a separable isogeny or not, or very inseparable or not, is determined by its action on $A[p]^0$, i.e., on its image under the map*

$$\mathrm{End}(A) \to \mathrm{End}(A[p]^0).$$

(ii) *Very inseparable isogenies are inseparable.*

(iii) *There exists a simple abelian surface with a confined isogeny that is inseparable but not very inseparable and for which inseparable isogenies together with the zero map do not form an ideal.*

Proof. Statement (i) is immediate from Theorem 6.3. Statement (ii) follows from Theorem 6.3, since nilpotents are not invertible. Concerning (iii), the following is an example of a simple abelian variety A and an inseparable but not very inseparable isogeny σ (all computational data used can be found at [LMFDB Collaboration 2013]). Consider the isogeny class of supersingular abelian surfaces over $\mathbf{F}_5$ of p-rank 0 with characteristic polynomial of the Frobenius π equal to $x^4 + 25 = 0$. The splitting field $L := \mathbf{Q}(\pi) = \mathbf{Q}(i, \sqrt{10})$ has no real embeddings, hence by [Waterhouse 1969, Theorem 6.1] there exists a simple abelian surface A with endomorphism ring $\mathbb{O}_L = \mathbf{Z}[i, \pi]$ (the ring of integers in L, containing both π and $5/\pi = -i\pi$). Consider $\sigma = i - 2 = \pi^2/5 - 2$, with characteristic polynomial $\sigma^2 + 4\sigma + 5 = 0$. The endomorphism σ is a confined isogeny since on a simple abelian variety these are exactly the endomorphisms that are neither zero nor roots of unity. Denoting the reduction of σ modulo 5 by $\bar{\sigma}$, we find that

$$\bar{\sigma}^2 = \bar{\sigma}. \tag{24}$$

Note that $A[p] = A[p]^0$ and hence there is an injective map $\mathbb{O}_L/5\mathbb{O}_L \hookrightarrow \mathrm{End}(A[p]^0)$. Now σ is separable if and only if $\bar{\sigma}$ is an isomorphism on $A[p]^0$, which, by (24), happens exactly if $\bar{\sigma} = 1$. But then $\sigma = 5\psi + 1$ for some $\psi \in \mathbb{O}_L$, which does not hold. Hence σ is inseparable. On the other hand, σ is very inseparable if and only if $\bar{\sigma}$ is nilpotent on $A[p]^0$, which, by (24), happens exactly if $\bar{\sigma} = 0$. This means that $\sigma = 5\psi$ for some $\psi \in \mathbb{O}_L$, which does not hold either. Hence σ is not very inseparable.

Let $\sigma' = -i - 2$. We similarly prove that σ' is inseparable, and yet the map $\sigma + \sigma' = -4$ is a separable isogeny. Hence the set of inseparable isogenies together with the zero map is not closed under addition. $\square$

Using Dieudonné modules. The structure of the endomorphism ring of the local group scheme $A[p]^0$ can be computed explicitly using the theory of Dieudonné modules, and we will use this to deduce some more results on very inseparability.

The group schemes $A[p]$ and $A[p]^0$ are objects in the category $\mathscr{C}_K$ of finite commutative group schemes over K annihilated by p. By covariant Dieudonné theory [Goren 2002, A§5] there is an equivalence of categories

$$D \colon \mathscr{C}_K \to \text{finite length left } \mathbf{E}\text{-modules,}$$

where $\mathbf{E} = K[F, V]$ denotes the noncommutative ring of polynomials with relations

$$FV = VF = 0, F\lambda = \lambda^p F \quad \text{and} \quad V\lambda^p = \lambda V \text{ for } \lambda \in K.$$

We may consider being a very inseparable endomorphism or a separable isogeny as a property of the image of an endomorphism under the map $\text{End}(A) \to \text{End}_{\mathbf{E}}(D(A[p]^0))$.

Example 6.6. If A is an ordinary elliptic curve, then $A[p]^0 \cong \mu_p$, so $\text{End}(A[p]^0) = \mathbf{F}_p$. If A is a supersingular elliptic curve, the local group scheme $A[p]^0$ is the unique nonsplit self-dual extension of α_p by α_p. The Dieudonné module is $D(A[p]^0) = \mathbf{E}/\mathbf{E}(V + F)$ [Goren 2002, A.5.4] and a computation [Goren 2002, A.5.8] gives a ring isomorphism

$$\text{End}(A[p]^0) \cong \text{End}_{\mathbf{E}}(\mathbf{E}/\mathbf{E}(V + F)) \cong \left\{ \begin{pmatrix} a^p & b \\ 0 & a \end{pmatrix} : a \in \mathbf{F}_{p^2}, b \in K \right\}.$$

From these computations, one also sees directly that noninvertible elements are nilpotent in $\text{End}(A[p]^0)$ in both the ordinary and the supersingular case, giving an alternative proof of 6.2(iv).

Proposition 6.7. *Let $\sigma \in \text{End}(A)$ and set $\mathcal{D} := D(A[p])^0$.*

(i) *σ is a separable isogeny (respectively, very inseparable endomorphism) if and only if its image in $\text{End}_{K[F]}(\mathcal{D}/V\mathcal{D})$ is invertible (respectively, nilpotent).*

(ii) *σ is very inseparable if and only if a power of σ factors through the p-Frobenius map $\text{Fr}: A \mapsto A^{(p)}$.*

(iii) *If $\text{End}(A)$ is commutative, the set of very inseparable endomorphisms forms an ideal in $\text{End}(A)$.*

(iv) *There exists an abelian variety for which the set of very inseparable endomorphisms is not closed under either addition or multiplication (in particular, it is not an ideal).*

(v) *Let A denote a simple ordinary abelian variety defined over a finite field $\mathbf{F}_q \subseteq K$ with (commutative) endomorphism ring $\mathbb{O} := \text{End}(A)$ and Frobenius endomorphism π. Set $R := \mathbf{Z}[\pi, q/\pi]$. Then $R \subseteq \mathbb{O}$ and if $p \nmid [\mathbb{O}:R]$, then any isogeny of A is very inseparable if and only if it is inseparable. This is in particular true if $q = p \geqslant 5$.*

Proof. We first prove (i). The relations in $\mathbf{E}$ imply that $V\mathbf{E}$ is a two-sided ideal in $\mathbf{E}$. In this way, σ, as an $\mathbf{E}$-endomorphism of $\mathcal{D}$, gives rise to an endomorphism $\tilde{\sigma}$ of the $\mathbf{E}/V\mathbf{E} = k[F]$-module $\mathcal{D}/V\mathcal{D}$. The first claim is that σ is nilpotent if and only if $\tilde{\sigma}$ is. The interesting direction is where $\tilde{\sigma}$ is nilpotent, meaning that $\sigma^n(\mathcal{D}) \subseteq V\mathcal{D}$ for some n. Since V is nilpotent on $\mathcal{D}$ [Goren 2002, A.5], say $V^d\mathcal{D} = 0$, we can iterate the equation to get $\sigma^{nd}(\mathcal{D}) \subseteq V^d\mathcal{D} = 0$. Secondly, we claim that σ is invertible if and only if $\tilde{\sigma}$ is so. Again, the interesting direction is when $\tilde{\sigma}$ is invertible. If we let $\mathcal{D}'$ denote the image of $\sigma: \mathcal{D} \to \mathcal{D}$, then $\mathcal{D}'$ is an $\mathbf{E}$-submodule of $\mathcal{D}$ and $\mathcal{D} = \mathcal{D}' + V\mathcal{D}$. Iterating this sufficiently many times, we find that

$$\mathcal{D} = \mathcal{D}' + V\mathcal{D} = \mathcal{D}' + V\mathcal{D}' + V^2\mathcal{D} = \cdots = \mathcal{D}' + V\mathcal{D}' + \cdots + V^{d-1}\mathcal{D}' \subseteq \mathcal{D}'.$$

This shows that σ is surjective, and, since it is an endomorphism of the underlying finite-dimensional vector space, it is then automatically injective.

In order to prove (ii), note that the Dieudonné module $D(A^{(p)}[p]^0)$ can be identified with $\mathcal{D} = D(A[p]^0)$ with the **E**-action twisted by the geometric Frobenius map $\psi \colon K \to K$, $\psi(\lambda) = \lambda^{1/p}$. Under this identification, the map induced by the p-Frobenius Fr$\colon A \to A^{(p)}$ on the Dieudonné modules is the ψ-semilinear map $V \colon \mathcal{D} \to \mathcal{D}$ [Goren 2002, A.5]. Moreover, the map V is nilpotent.

If σ is very inseparable, there exists n with $\sigma^n|_{A[p]^0} = 0$. Since $A[\text{Fr}] \subseteq A[p]^0$, we have $\sigma^n|_{A[\text{Fr}]} = 0$ and hence σ^n factors through Fr. Conversely, suppose that $\sigma^n = \tau \circ \text{Fr}$ for some $\tau \colon A^{(p)} \to A$. Passing to the Dieudonné modules, and using the fact that the map $D(\tau)$ is ψ^{-1}-semilinear (and hence commutes with V), we see that $D(\sigma^n)\mathcal{D} \subseteq V\mathcal{D}$, so $D(\sigma)$ is nilpotent modulo V. By part (i), we find that σ is very inseparable.

For the proof of (iii), note that, without any assumptions on the ring End(A), the set I of maps in End(A) that factor through the p-Frobenius Fr is a left ideal in End(A). Therefore by (ii), if the ring End(A) is commutative, the set of very inseparable maps in End(A) coincides with the radical of I, and hence is an ideal.

For (iv), consider $A = E \times E$ for an ordinary elliptic curve E. Then End$(A) = \text{M}_2(\text{End}(E))$ surjects onto End$(A[p]^0) = \text{M}_2(\mathbf{F}_p)$ (see Example 6.6). The set of very inseparable endomorphisms corresponds under this map to matrices whose image in $\text{M}_2(\mathbf{F}_p)$ is nilpotent, and it suffices to remark that the set of nilpotent elements in $\text{M}_2(\mathbf{F}_p)$ is not closed under neither addition nor multiplication.

For (v), we indeed have $R \subseteq \mathbb{O}$ by [Waterhouse 1969, 7.4]. Let $\sigma \in \mathbb{O}$ and observe that the coprimality of $[\mathbb{O}:R]$ to p implies that there exists an integer N coprime to p with $N\sigma \in R$. Therefore, it suffices to prove the equivalence of inseparability and very inseparability for elements of R. Represent such an element $\sigma \in R$ by

$$\sum_{i \geqslant 1} a_i \pi^i + \sum_{j \geqslant 0} b_j (\pi')^j,$$

with $\pi' = q/\pi$ and $a_i, b_i \in \mathbf{Z}$ (the terms containing both π and π' may be omitted since they do not change the image of σ in End$(\mathcal{D})$). Since A is defined over $\mathbf{F}_q$ with $q = p^r$, we have $\pi = \text{Fr}^r$ and $\pi' = \text{Ver}^r$, where Ver$\colon A^{(p)} \to A$ is the Verschiebung. On the level of Dieudonné modules, Fr maps to V and Ver maps to F [Goren 2002, A.5], so σ maps to the endomorphism

$$\tilde{\sigma} := \sum b_j F^{rj} \in \text{End}_{K[F]}(\mathcal{D}/V\mathcal{D}).$$

In the ordinary case, the Dieudonné modules of $A[p]$ and $A[p]^0$ are

$$D(A[p]) = (\mathbf{E}/(V, 1-F) \oplus \mathbf{E}/(F, 1-V))^g \quad \text{and} \quad \mathcal{D} = D(A[p]^0) = (\mathbf{E}/(V, 1-F))^g$$

(since this is the subgroup scheme of $D(A[p])$ on which V is nilpotent [Goren 2002, A.5]). Hence $F = 1$ in End$(\mathcal{D}/V\mathcal{D}) = \text{M}_g(\mathbf{F}_p)$, and $\tilde{\sigma} := \sum b_j$ is a scalar multiplication; therefore, it is nilpotent if and only if it is zero (i.e., noninvertible).

The final claim follows from a result of Freeman and Lauter [2008, Proposition 3.7]. $\square$

We were unable to answer the following natural questions:

Question 6.8. (i) Can one construct a *simple* abelian variety for which very inseparable endomorphisms do not form an ideal?

(ii) Consider the subset of the moduli space of abelian varieties of given dimension and given degree of polarization consisting of those abelian varieties A for which inseparable isogenies are very inseparable. Is this locus dense in the moduli space? Recall that, by a result of Norman and Oort [1980, Theorem 3.1], the ordinary locus is dense.

7. The tame zeta function

We revert to our standard assumptions and define the following general "tame" version of the Artin–Mazur zeta function for varieties over fields of positive characteristic (the construction is somewhat reminiscent of that of the Artin–Hasse exponential):

Definition 7.1. Let K denote an algebraically closed field of positive characteristic $p > 0$, X/K an algebraic variety, and let $f : X \to X$ denote a confined morphism. The *tame zeta function* ζ_f^* is defined as the formal power series

$$\zeta_f^*(z) := \exp\left(\sum_{p \nmid n} f_n \frac{z^n}{n}\right), \tag{25}$$

summing only over n that are not divisible by p.

A basic observation is:

Proposition 7.2. *We have identities of formal power series*

$$\zeta_{X,f}(z) = \prod_{i \geq 0} \sqrt[p^i]{\zeta_{X,f^{p^i}}^*(z^{p^i})} \tag{26}$$

and

$$\zeta_{X,f}^*(z) = \zeta_{X,f}(z) / \sqrt[p]{\zeta_{X,f^p}(z^p)}. \tag{27}$$

Proof. For the first identity (26), we do a formal computation, splitting the sum over n into parts where n is exactly divisible by a given power p^i of p (denoted $p^i \parallel n$):

$$\begin{aligned}
\zeta_{X,f}(z) &= \exp\left(\sum_{i \geq 0} \sum_{p^i \parallel n} \frac{f_n}{n} z^n\right) \\
&= \exp\left(\sum_{i \geq 0} \sum_{p \nmid m} \frac{f_{p^i m}}{p^i m} z^{p^i m}\right) \\
&= \exp\left(\sum_{i \geq 0} \frac{1}{p^i} \sum_{p \nmid m} \frac{(f^{p^i})_m}{m} (z^{p^i})^m\right) \\
&= \prod_{i \geq 0} \exp\left(\frac{1}{p^i} \log(\zeta_{f^{p^i}}^*(z^{p^i}))\right).
\end{aligned}$$

For the second identity (27), we compute as follows:

$$\zeta^*_{X,f}(z) = \exp\left(\sum_{n\geqslant 1}\frac{f_n}{n}z^n - \sum_{k\geqslant 1}\frac{f_{pk}}{pk}z^{pk}\right) = \exp\left(\sum_{n\geqslant 1}\frac{f_n}{n}z^n\right)\bigg/\exp\left(\frac{1}{p}\sum_{k\geqslant 1}\frac{(f^p)_k}{k}z^{pk}\right). \qquad \square$$

Theorem 7.3. *For $\sigma \circlearrowright A$, there exists an integer $t > 0$ (depending on σ) such that $(\zeta^*_\sigma)^t$ is a rational function. In particular, ζ^*_σ is algebraic.*

Proof. Proposition 2.7 implies that for $p \nmid n$ the inseparability degree $\deg_i(\sigma^n - 1) = r_n$ is periodic of period ω with $r_n = r_{\gcd(n,\omega)}$. Let μ denote the Möbius function. For $n \mid \omega$, define rational numbers α_n by

$$\alpha_n = \frac{1}{n}\sum_{e\mid n}\frac{\mu(n/e)}{r_e}. \tag{28}$$

By Möbius inversion and the equality $r_n = r_{\gcd(n,\omega)}$, we get

$$\frac{1}{r_n} = \sum_{d\mid\gcd(n,\omega)}d\alpha_d \quad \text{for all } n\geqslant 1.$$

Therefore,

$$\begin{aligned}
\zeta^*_\sigma(z) &= \exp\left(\sum_{p\nmid n}\frac{\deg(\sigma^n - 1)}{nr_n}z^n\right)\\
&= \exp\left(\sum_{d\mid\omega}\alpha_d\sum_{p\nmid m}\frac{\deg(\sigma^{dm} - 1)}{m}z^{dm}\right)\\
&= \prod_{d\mid\omega}\left(\exp\left(\sum_{p\nmid m}\frac{\deg(\sigma^{dm} - 1)}{m}z^{dm}\right)\right)^{\alpha_d}.
\end{aligned}$$

Using the notation of Proposition 2.3(i), we can rewrite this as

$$\zeta^*_\sigma(z) = \prod_{d\mid\omega}\left(D_{\sigma^d}(z^d)\big/\sqrt[p]{D_{\sigma^{pd}}(z^{pd})}\right)^{\alpha_d} \tag{29}$$

and hence the result follows from the rationality of the degree zeta functions. $\qquad \square$

The minimal exponent $t_\sigma > 0$ for which $\zeta^*_\sigma(z) \in \mathbf{Q}(z)$ is an invariant of the dynamical system $\sigma \circlearrowright A$. We briefly discuss the arithmetic significance of such t_σ, by considering both ordinary and supersingular elliptic curves.

Proposition 7.4. *Let E denote an elliptic curve, $\sigma \in \mathrm{End}(E)$, and let t_σ be the minimal positive integer for which $\zeta^*_\sigma(z)^{t_\sigma} \in \mathbf{Q}(z)$.*

(i) *If E is ordinary, t_σ is a pure p-th power.*

(ii) *There exists a (supersingular) E and $\sigma \circlearrowright E$ for which t_σ is not a pure p-th power.*

Proof. If σ is an endomorphism of an ordinary elliptic curve, then there is a valuation $|\cdot|$ on the quotient field L of the endomorphism ring that extends the p-valuation and such that $\deg_i \sigma = |\sigma|$ (cf. Example 4.2). If σ is very inseparable, $\zeta_\sigma^*(z)$ is rational, and the claim is clear. Otherwise, let s be the minimal positive integer for which $M := |\sigma^s - 1| < 1$. We find that for integers n not divisible by p,

$$r_n = \deg_i(\sigma^n - 1) = \begin{cases} 1 & \text{if } s \nmid n, \\ M & \text{if } s \mid n. \end{cases} \tag{30}$$

Substituting this into (28), we get $\omega = s$. If $s = 1$, we have $\alpha_1 = 1/M$, and if $s > 1$, we find

$$\alpha_n = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \mid s, \ 1 < n < s, \\ (1 - M)/(Ms) & \text{if } n = s. \end{cases} \tag{31}$$

Since p splits in L [Deuring 1941, §2.10], the valuation $|\cdot|$ has residue field $\mathbf{F}_p$, and hence $s \mid (p - 1)$. From (29), it follows that $\zeta_\sigma^*(z)$ is a product of rational functions to powers $1/p$ and $(1 - M)/(Mps)$ (and $1/(Mp)$ if $s = 1$). Now with $M = p^{-r}$ for some $r \geqslant 1$, we find that $(1 - M)/(Mps) = (p^r - 1)/p^{r+1}s$, which has denominator a power of p, since s divides $p - 1$. This proves (i).

For (ii) consider a supersingular elliptic curve $A = E$. We have already seen in Example 4.2 that the inseparability degree of an isogeny is detected by a valuation on the quaternion algebra $\mathrm{End}(E) \otimes \mathbf{Q}$, on which we now briefly elaborate. The ring $\mathbb{O} = \mathrm{End}(E)$ is a maximal order in a quaternion algebra, and its completion $\mathbb{O}_p = \mathrm{End}(E) \otimes_{\mathbf{Z}} \mathbf{Z}_p$ is an order in the unique quaternion division algebra D over $\mathbf{Q}_p$ [Deuring 1941]. There exists a valuation $v \colon D \to \mathbf{Z}$ on D with the property that $\mathbb{O}_p = \{x \in D : v(x) \geqslant 0\}$. Let $\mathfrak{p} = \{x \in \mathbb{O} : v(x) \geqslant 1\}$. Then $\mathfrak{p}$ is a two-sided maximal ideal in $\mathbb{O}$ with $p\mathbb{O}_p = \mathfrak{p}^2 \mathbb{O}_p$ and we have an isomorphism $\mathbb{O}/\mathfrak{p} \simeq \mathbf{F}_{p^2}$. The inseparable degree of an isogeny $\sigma \in \mathbb{O}$ is given by the formula $\deg_i(\sigma) = p^{v(\sigma)}$, cf. [Bridy 2016, Proposition 5.5].

Let $\sigma \in \mathbb{O}$ be an endomorphism such that its image in $\mathbb{O}/\mathfrak{p} \simeq \mathbf{F}_{p^2}$ generates the multiplicative group of the field and such that $v(\sigma^{p^2-1} - 1) = 1$. Then for integers n not divisible by p we have

$$\deg_i(\sigma^n - 1) = \begin{cases} 1 & \text{if } (p^2 - 1) \nmid n, \\ p & \text{if } (p^2 - 1) \mid n. \end{cases} \tag{32}$$

Let us prove that such σ exists: choose elements $\sigma_0, \tau \in \mathbb{O}$ such that the image of σ_0 in $\mathbb{O}/\mathfrak{p} \simeq \mathbf{F}_{p^2}$ generates the multiplicative group of the field and $v(\tau) = 1$. Then one of the elements $\sigma_0, \sigma_0 + \tau$ satisfies the desired conditions.

Furthermore, the degree is of the form $\deg(\sigma^n - 1) = m^n - \lambda^n - (\lambda')^n + 1$ for $\lambda, \lambda' \in \overline{\mathbf{Q}}$ and $m := \lambda\lambda' \in \mathbf{Z}$. Using the convenient notation

$$\mathscr{X}(z) := \frac{\sqrt[p]{1 - z^p}}{1 - z},$$

a somewhat tedious computation, splitting the terms in $\log \zeta_\sigma^*(z)$ to take into account the cases in (32), gives that

$$\zeta_\sigma^*(z) = \frac{g_1(z)}{\sqrt[p(p+1)]{g_{p^2-1}(z)}}, \quad \text{where } g_i(z) := \frac{\mathscr{X}(z^i)\,\mathscr{X}((mz)^i)}{\mathscr{X}((\lambda z)^i)\,\mathscr{X}((\lambda' z)^i)}.$$

Note that $\mathscr{L}(z)$ is itself a p-th root of a rational function. We conclude that $t = p^2(p+1)$ suffices to have $\zeta_\sigma^*(z)^t \in \mathbf{Q}(z)$ but $\zeta_\sigma^*(z)^t$ is not rational for any choice of t as a pure p-th power. $\qquad\square$

8. Functional equations

In this section, we study the existence of functional equations for full and tame zeta functions on abelian varieties. Assume throughout the section that σ is an isogeny. Under the transformation $z \mapsto 1/\deg(\sigma)z$, we will find a functional equation for zeta functions of very inseparable endomorphisms, and a "Riemann surface" version of a functional equation for the tame zeta function. Since this transformation does not make sense for ζ_σ as a formal power series, D_σ, ζ_σ, and ζ_σ^* are therefore considered as genuine functions of a complex variable, and the symbols are understood to refer to their (maximal) analytic continuations.

Proposition 8.1. *The degree zeta function $D_\sigma(z)$ (cf. Definition 2.2) satisfies a functional equation of the form*

$$D_\sigma\left(\frac{1}{\deg(\sigma)z}\right) = D_\sigma(z).$$

Proof. We use the notations from (20). It is clear that the multiset of λ_i is stable under the involution $\lambda \mapsto \deg(\sigma)/\lambda$. From this symmetry, we obtain a functional equation for the exponential generating function $D_\sigma(z) = \prod_{i=1}^r (1 - \lambda_i z)^{-m_i}$ of the form

$$D_\sigma\left(\frac{1}{\deg(\sigma)z}\right) = (-z)^{\sum_{i=1}^r m_i} \prod_{i=1}^r \lambda_i^{m_i} D_\sigma(z).$$

Substituting $n = 0$ into (20) gives $\sum_{i=1}^r m_i = 0$ and a direct computation using the form of λ_i and the fact that q is even shows that $\prod_{i=1}^r \lambda_i^{m_i} = 1$, which gives the claim. $\qquad\square$

Remark 8.2. The functional equation for $D_\sigma(z)$ can be placed in the cohomological framework from Remark 2.4: consider the Poincaré duality pairing $\langle \cdot, \cdot \rangle \colon \mathrm{H}^i \times \mathrm{H}^{2g-i} \otimes \mathbf{Q}_\ell(g) \to \mathbf{Q}_\ell$, under which $\langle \sigma_* x, y \rangle = \langle x, \sigma^* y \rangle$, with $\sigma_* \sigma^* = [\deg \sigma]$. Hence if σ^* has eigenvalues α_i on H^i, then σ_* has eigenvalues $\deg(\sigma)/\alpha_i$ on H^{2g-i}, but these sets are the same by duality. In this way the functional equation picks up a factor $z^{\chi(A)}$, where $\chi(A)$ is the ℓ-adic Euler characteristic of A. But here, $\chi(A) = 0$ (since the i-th ℓ-adic Betti number of an abelian variety of dimension g is the binomial coefficient $\binom{2g}{i}$).

Theorem 8.3.   (i) *If σ is very inseparable, then $\zeta_\sigma(z)$ extends to a meromorphic function on the entire complex plane and satisfies a functional equation of the form*

$$\zeta_\sigma\left(\frac{1}{\deg(\sigma)z}\right) = \zeta_\sigma(z).$$

(ii) *If σ is not very inseparable and Λ is the unique dominant root, then $\zeta_\sigma(z)$ cannot satisfy a functional equation under $z \mapsto 1/\deg(\sigma)z$; actually, the intersection of the domains of $\zeta_\sigma(z)$ and $\zeta_\sigma(1/\deg(\sigma)z)$ is empty.*

(iii) *For any confined σ, let X_σ denote the concrete Riemann surface of the algebraic function $\zeta_\sigma^*(z)$ (a finite covering of the Riemann sphere). Then there exists an involution $\tau \in \mathrm{Aut}(X_\sigma)$ such that the meromorphic extension $\zeta_\sigma^*\colon X_\sigma \to \hat{\mathbf{C}}$ fits into a commutative diagram of the form*

$$
\begin{array}{ccc}
X_\sigma & \xrightarrow{\ \tau\ } & X_\sigma \\
{\scriptstyle\zeta_\sigma^*}\downarrow & & \downarrow{\scriptstyle\zeta_\sigma^*} \\
\hat{\mathbf{C}} & \xrightarrow{\ \mathrm{id}\ } & \hat{\mathbf{C}}.
\end{array}
\tag{33}
$$

Proof. If σ is very inseparable, then $\zeta_\sigma = D_\sigma$, and the result follows from Proposition 8.1.

If σ is not very inseparable and Λ is the unique dominant root, then by Theorem 5.5 the function $\zeta_\sigma(z)$ has a natural boundary on $|z| = 1/\Lambda$. Thus $\zeta_\sigma(z)$ and $\zeta_\sigma(1/\deg(\sigma)z)$ are commonly defined only on $\Lambda/\deg(\sigma) < |z| < 1/\Lambda$ which is empty when $\Lambda^2 \geqslant \deg(\sigma)$. By Proposition 5.1(iii), we have $\Lambda^2 \geqslant \Lambda \geqslant \prod |\xi_i| = \deg \sigma$, so this always holds.

For the third part of the theorem, consider (29) that expresses the function ζ_σ^* in terms of degree zeta functions. Write $\alpha_d/p = A_d/B_d$ for coprime integers A_d, B_d, let N denote the least common multiple of B_d over all $d \mid \omega$ and set $\beta_d := N\alpha_d/p \in \mathbf{Z}$. Then ζ_σ^* extends to a function on the Riemann surface X_σ corresponding to the projective curve defined by the affine equation

$$
y^N = \prod_{d \mid \omega} \left(\frac{D_{\sigma^d}(x^d)^p}{D_{\sigma^{pd}}(x^{pd})} \right)^{\beta_d}
$$

given by $\zeta_\sigma^*(x, y) = y$. By the fact that all D_σ satisfy the functional equation as in Proposition 8.1, the map $\tau\colon X_\sigma \to X_\sigma$, $\tau(x, y) = (1/(\deg(\sigma)x), y)$ is an involution of X_σ (we use that $\deg(\sigma^r) = \deg(\sigma)^r$ for any integer r). The same functional equations then prove that the diagram (33) commutes. $\qquad\square$

9. Prime orbit growth

In this section, we consider the prime orbit growth for a confined endomorphism $\sigma\colon A \to A$. We are interested in possible analogues of the prime number theorem (PNT), much like Parry and Pollicott [1983] proved for axiom A flows. In our case, it follows almost immediately from the rationality of their zeta functions that such an analogue holds for very inseparable σ. In general, however, as we will see, the prime orbit counting function displays infinitely many forms of limiting behavior. Nevertheless, the (weaker) analogue of Chebyshev's bounds and Mertens' second theorem hold. In accordance with our philosophy, we also consider counting only "tame" prime orbits (i.e, of length coprime to p), and in this case we see finitely many forms of limiting behavior, detectable from properties of the p-divisible group. Finally, we briefly discuss good main and error terms reflecting analogues of the Riemann hypothesis.

Notations/Definitions 9.1. A *prime orbit* O of length $\ell =: \ell(O)$ of $\sigma\colon A \to A$ is a set

$$
O = \{x, \sigma x, \sigma^2 x, \ldots, \sigma^\ell x = x\} \subseteq A(K)
$$

of exact cardinality ℓ. Letting P_ℓ denote the number of prime orbits of length ℓ for σ, the *prime orbit counting function* is $\pi_\sigma(X) := \sum_{\ell \leqslant X} P_\ell$.

As formal power series, the zeta function of σ admits a product expansion

$$\zeta_\sigma(z) = \prod_O \frac{1}{1 - z^{\ell(O)}},$$

where the product runs over all prime orbits. Since $\sigma_n = \sum_{\ell \mid n} \ell P_\ell$, Möbius inversion implies that $P_\ell = \frac{1}{\ell} \sum_{n \mid \ell} \mu\left(\frac{\ell}{n}\right)\sigma_n$. Our proofs will exploit the fact that the numbers σ_n differ from the linear recurrent sequence $\deg(\sigma^n - 1)$ only by a multiplicative factor, the inseparable degree, that grows quite slowly.

To avoid complications, we make the following assumption:

> *Standing assumption/notations*:
> The dominant root $\Lambda > 1$ is unique.
> The ϖ-periodic sequences (r_n) and (s_n), $s_n \leqslant 0$, are as in (14).
> All asymptotic formulæ in this section hold for integer values of the parameter.

By Proposition 5.1(vi), this implies that $\Lambda > 1$ is of multiplicity one. We start with a basic proposition describing the asymptotics of P_ℓ. Interestingly, the error terms are determined by the zeros of the degree zeta function. This appears to be a rather strong result with a very easy proof, dependent on the exponential growth.

Proposition 9.2. $P_\ell = \Lambda^\ell / (\ell r_\ell |\ell|_p^{s_\ell}) + O(\Lambda^{\Theta\ell})$, *where* $\Theta := \max\{\mathrm{Re}(s) : D_\sigma(\Lambda^{-s}) = 0\} \in \left[\frac{1}{2}, 1\right)$.

Proof. From (10), we get $\deg(\sigma^n - 1) = \Lambda^n + O(\Lambda^{\Theta\ell})$ for

$$\Theta := \max_{|\lambda_i| \neq \Lambda} \frac{\log|\lambda_i|}{\log(\Lambda)}.$$

By Proposition 5.4, this equals the largest real part of a zero of $D_\sigma(\Lambda^{-s})$, and $1/2 \leqslant \Theta < 1$. Hence

$$\sigma_\ell = \frac{\deg(\sigma^\ell - 1)}{\deg_i(\sigma^\ell - 1)} = \frac{\Lambda^\ell}{r_\ell |\ell|_p^{s_\ell}} + O(\Lambda^{\Theta\ell}).$$

Expressing the number of prime orbits in terms of the number of fixed points, we get

$$P_\ell = \frac{1}{\ell} \sum_{n \mid \ell} \mu\left(\frac{\ell}{n}\right)\sigma_n = \frac{\sigma_\ell}{\ell} + \frac{1}{\ell} \sum_{\substack{n \mid \ell \\ n < \ell}} \mu\left(\frac{\ell}{n}\right)\sigma_n.$$

Since $|\mu(\ell/n)\sigma_n| \leqslant \deg(\sigma^n - 1) \leqslant M\Lambda^n$ for some constant M depending only on σ, we get

$$\left| \sum_{\substack{n \mid \ell \\ n < \ell}} \mu\left(\frac{\ell}{n}\right)\sigma_n \right| \leqslant \ell M \Lambda^{\ell/2},$$

and since $\Theta \geqslant \frac{1}{2}$, the claim follows. $\qquad\square$

The remainder of this section is dedicated to a study of what happens to the asymptotics if we further average in ℓ, like in the prime number theorem or Mertens' theorem. We will see that between PNT and Mertens' theorem, information about σ being very inseparable or not gets lost.

The next lemma is formulated in a general way and will be applied several times in order to asymptotically replace factors "$1/\ell$" for $\ell \leqslant X$ by "$1/X$". This leads to simplified main terms at the cost of worse error terms (we will discuss another approach leading to a "complicated main term with good error term" at the end of the section).

Lemma 9.3. *Let (a_ℓ) be a bounded sequence and let $\Lambda > 1$ be a real number. Then*

$$\sum_{\ell \leqslant X} \frac{a_\ell}{\ell} \Lambda^{\ell - X} = \frac{1}{X} \sum_{\ell \leqslant X} a_\ell \Lambda^{\ell - X} + O(1/X^2).$$

Proof. Write

$$\sum_{\ell \leqslant X} \frac{a_\ell}{\ell} \Lambda^{\ell - X} - \frac{1}{X} \sum_{\ell \leqslant X} a_\ell \Lambda^{\ell - X} = \sum_{\ell \leqslant X} \frac{a_\ell (X - \ell)}{X \ell} \Lambda^{\ell - X}.$$

With $M := \sup|a_\ell| < +\infty$, the "top half" of this sum can be bounded as follows:

$$\left| \sum_{X/2 \leqslant \ell \leqslant X} \frac{a_\ell (X - \ell)}{X \ell} \Lambda^{\ell - X} \right| \leqslant \frac{2M}{X^2} \sum_{i \geqslant 0} i \Lambda^{-i} = O(1/X^2)$$

while the "bottom half" is easily seen to be $O(X \Lambda^{-X/2})$, whence the claim. $\square$

(Non)analogues of PNT and analogues of Chebyshev's estimates. The first application is to the following "fluctuating" asymptotics for the prime orbit counting function:

Proposition 9.4.
$$\frac{X \pi_\sigma(X)}{\Lambda^X} = \sum_{\ell \leqslant X} \frac{1}{r_\ell |\ell|_p^{s_\ell}} \Lambda^{\ell - X} + O(1/X).$$

Proof. By Proposition 9.2 we see that

$$\frac{X \pi_\sigma(X)}{\Lambda^X} = X \sum_{\ell \leqslant X} P_\ell \Lambda^{-X} = X \sum_{\ell \leqslant X} \left(\frac{1}{\ell r_\ell |\ell|_p^{s_\ell}} \Lambda^{\ell - X} + \Lambda^{-X} O(\Lambda^{\Theta \ell}) \right).$$

The error terms in this sum form a geometric series and hence decrease exponentially. Applying Lemma 9.3 to the main term, we find the stated result. $\square$

The next theorem discusses the analogue of the PNT in our setting; an analogue of Chebyshev's 1852 determination of the order of magnitude of the prime counting function holds in general, but the analogue of the PNT holds only for very inseparable endomorphisms. The result for general endomorphisms is similar in spirit to that for the 3-adic doubling map considered in [Everest et al. 2005, Theorem 3], S-integer dynamical systems in [Everest et al. 2007] (from which we take the terminology "detector group"), or to Knieper's theorem [1997, Theorem B] on the asymptotics of closed geodesics on rank one manifolds of nonpositive curvature.

Theorem 9.5. (i) *The order of magnitude of $\pi_\sigma(X)$ is $\pi_\sigma(X) \asymp \Lambda^X/X$, in the sense that the function $X\pi_\sigma(X)/\Lambda^X$ is bounded away from 0 and ∞.*

(ii) *Consider the "detector" group*

$$G_\sigma := \{(a, x) \in \mathbf{Z}/\varpi\mathbf{Z} \times \mathbf{Z}_p : a \equiv x \bmod |\varpi|_p^{-1}\}.$$

If (X_n) is a sequence of integers such that $X_n \to +\infty$ and (X_n, X_n) has a limit in the group G_σ, then the sequence $X_n\pi_\sigma(X_n)/\Lambda^{X_n}$ converges, and every accumulation point of $X\pi_\sigma(X)/\Lambda^X$ arises in this way.

(iii) (a) *If σ is very inseparable, $\lim_{X\to+\infty} X\pi_\sigma(X)/\Lambda^X$ exists and equals $\Lambda/(\Lambda - 1)$.*

(b) *If σ is not very inseparable, then the set of accumulation points of $X\pi_\sigma(X)/\Lambda^X$ is a union of a Cantor set and finitely many points. In particular, it is uncountable.*

Proof. For (i), we estimate the value of $X\pi_\sigma(X)/\Lambda^X$ in terms of the sum in Proposition 9.4. The bound from above is trivial; for the bound from below we consider the terms with $\ell = X - 1$ and $\ell = X$ and note that for at least one of these indices we have $|\ell|_p = 1$. We thus obtain the bounds

$$\frac{1}{\Lambda \max(r_\ell)} \leqslant \liminf_{X\to+\infty} \frac{X\pi_\sigma(X)}{\Lambda^X} \leqslant \limsup_{X\to+\infty} \frac{X\pi_\sigma(X)}{\Lambda^X} \leqslant \frac{\Lambda}{\Lambda - 1}. \tag{34}$$

To prove (ii), the formula in Proposition 9.4 may be rewritten as

$$\frac{X\pi_\sigma(X)}{\Lambda^X} = \sum_{\ell=0}^{X-1} \frac{1}{r_{X-\ell}|X - \ell|_p^{s_{X-\ell}}} \Lambda^{-\ell} + O(1/X). \tag{35}$$

If (X_n) is as indicated, i.e., if $X_n \bmod \varpi$ stabilizes (say at the value $\varpi_0 \bmod \varpi$) and X_n converges to some x in $\mathbf{Z}_p$, then individual summands in (35) have a well-defined limit while the whole sum is bounded uniformly in n by the convergent series $\sum_{t=0}^\infty \Lambda^{-t}$. Thus

$$\lim_{n\to+\infty} \frac{X_n\pi_\sigma(X_n)}{\Lambda^{X_n}} = \sum_{\ell=0}^\infty \frac{1}{r_{\varpi_0-\ell}|x - \ell|_p^{s_{\varpi_0-\ell}}} \Lambda^{-\ell}, \tag{36}$$

where (r_n) and (s_n) are prolonged to periodic sequences for $n \in \mathbf{Z}$ in an obvious manner; if x is a positive integer, then the term corresponding to $\ell = x$ should be construed as $\Lambda^{-\ell}/r_{\varpi_0-\ell}$ if $s_{\varpi_0-\ell} = 0$, and 0 otherwise.

We now prove (iii). When σ is very inseparable, $\varpi = 1$, $r_n = 1$, $s_n = 0$, and Proposition 9.4 implies the result by summing the geometric series $\sum_{k\geqslant 0} \Lambda^{-k} = 1/(1 - 1/\Lambda)$ in (36). Note that the result also follows by Tauberian methods applied to the rational zeta function $\zeta_\sigma = D_\sigma$.

In the case of general σ, we consider the map $\varphi \colon G_\sigma \to \mathbf{R}$ which associates to an element $(\varpi_0, x) \in G_\sigma$ the limit

$$\varphi(\varpi_0, x) = \lim_{n\to+\infty} \frac{X_n\pi_\sigma(X_n)}{\Lambda^{X_n}}$$

for a sequence (X_n) of integers such that $X_n \to +\infty$ and X_n has the limit (ϖ_0, x) in G_σ. By (36), this map is continuous. We will show that in some neighborhood of each point the map φ is either constant or a homeomorphism. Note that since G_σ is compact, the set of accumulation points of $X\pi_\sigma(X)/\Lambda^X$ is equal to the image of φ.

Choose $\varpi_0 \mod \varpi$, two distinct elements $x, y \in \mathbf{Z}_p$ and two sequences of integers (X_n) and (Y_n) which tend to infinity and such that $X_n \mod \varpi = Y_n \mod \varpi = \varpi_0$ and $X_n \to x$ and $Y_n \to y$ in $\mathbf{Z}_p$. Then by (36) we have

$$\varphi(\varpi_0, x) - \varphi(\varpi_0, y) = \sum_{\ell=0}^{\infty} a_\ell, \tag{37}$$

where

$$a_\ell = \frac{1}{r_{\varpi_0 - \ell}} \left(\frac{1}{|x - \ell|_p^{s_{\varpi_0 - \ell}}} - \frac{1}{|y - \ell|_p^{s_{\varpi_0 - \ell}}} \right) \Lambda^{-\ell}.$$

Let $k \geqslant 0$ be such that $|x - y|_p = p^{-k}$. The terms a_ℓ are nonzero if and only if $\ell \equiv x \pmod{p^{k+1}}$ or $\ell \equiv y \pmod{p^{k+1}}$ and furthermore $s_{\varpi_0 - \ell} \neq 0$. Note that whether such ℓ exists depends only on the values of $x - \varpi_0$ and $y - \varpi_0$ modulo $\gcd(p^{k+1}, \varpi)$. For ℓ with $a_\ell \neq 0$, the terms a_ℓ can be bounded from below:

$$|a_\ell| \geqslant \frac{1}{r_{\varpi_0 - \ell}} (p^{k s_{\varpi_0 - \ell}} - p^{(k+1)s_{\varpi_0 - \ell}}) \Lambda^{-\ell} \geqslant \frac{1}{2r_{\varpi_0 - \ell}} p^{k s_{\varpi_0 - \ell}} \Lambda^{-\ell}$$

while clearly $|a_\ell| \leqslant \Lambda^{-\ell}$ for any ℓ.

We now consider two cases depending on whether or not there exists ℓ such that $a_\ell \neq 0$.

Case 1: Assume first that there exists ℓ such that $a_\ell \neq 0$ and let ℓ_0 be the smallest such ℓ. Since any other such ℓ differs from ℓ_0 by a multiple of p^k, we get

$$\left| \sum_{\ell=0}^{\infty} a_\ell \right| \geqslant \left(\frac{1}{2r_{\varpi_0 - \ell_0}} p^{k s_{\varpi_0 - \ell_0}} - \frac{\Lambda^{-p^k}}{1 - \Lambda^{-p^k}} \right) \Lambda^{-\ell_0}.$$

Since the sequences (r_ℓ) and (s_ℓ) take only finitely many values, the expression on the right is positive for k larger than a constant K_0 which depends only on σ but not on x, y, or ϖ_0. Therefore from (37) we conclude that if $|x - y|_p \leqslant p^{-K_0}$, then $\varphi(\varpi_0, x) \neq \varphi(\varpi_0, y)$.

Case 2: If $a_\ell = 0$ for all ℓ, then by (37) we have $\varphi(\varpi_0, x) = \varphi(\varpi_0, y)$. Let p^v be the largest power of p dividing ϖ. Recall that whether $a_\ell = 0$ for all ℓ depends only on the values of $x - \varpi_0$ and $y - \varpi_0$ modulo $\gcd(p^{k+1}, \varpi)$. Therefore if $a_\ell = 0$ for all ℓ for $|x - y| = p^{-k}$ with $k \geqslant v$, then the map φ is locally constant in a neighborhood of (ϖ_0, x).

Replacing K_0 with $\max(K_0, v)$ if necessary, we see that the map $\varphi \colon G_\sigma \to \mathbf{R}$ restricted to open compact subsets

$$B(\varpi_0, x) = \{(\varpi_0, y) \in G_\sigma : |x - y|_p \leqslant p^{-K_0}\} \subseteq G_\sigma$$

is either injective (corresponding to Case 1) or constant (corresponding to Case 2). Since G_σ is a disjoint union of finitely many subsets $B(\varpi_0, x)$, and since each $B(\varpi_0, x)$ is topologically a Cantor set, we conclude that the image of φ is a union of finitely many (possibly no) Cantor sets and finitely many points.

In order to finish the proof, it is enough to note that if σ is very inseparable, then there exists $(\varpi_0, x) \in G_\sigma$ for which Case 1 holds, so the image of φ contains a Cantor set. Indeed, by Lemma 4.4 there exists an integer ϖ_0 such that $s_{\varpi_0} < 0$. It is then easy to see that Case 1 holds for this choice of ϖ_0 and $x = 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Example 9.6. If σ is the (very inseparable) Frobenius (relative to $\mathbf{F}_q$) on an abelian variety $A/\mathbf{F}_q$ of dimension g, then $\Lambda = q^g$ and we find that $\sum_{\ell \leqslant X} P_\ell \sim q^{g(X+1)}/(X(q^g - 1))$, where P_ℓ is the number of closed points of A with residue field $\mathbf{F}_{q^\ell}$.

Our warm up example from the introduction illustrates what happens in the not very inseparable case.

Tame prime orbit counting. Now consider the analogous question in the tame case.

Definition 9.7. The *tame prime orbit counting function* is $\pi_\sigma^*(X) := \sum_{\substack{\ell \leqslant X \\ p \nmid \ell}} P_\ell$.

Remark 9.8. The tame zeta function $\zeta_\sigma^*(z)$ is not exactly equal to the formal Euler product over orbits of length coprime to p, but rather (notice the difference with (26)):

$$\prod_{p \nmid \ell(O)} \frac{1}{1 - z^{\ell(O)}} = \prod_{i \geqslant 0} \sqrt[p^i]{\zeta_\sigma^*(z^{p^i})}.$$

We find only finitely many possible kinds of limiting behavior, governed by the values of the periodic sequence (r_n) (the warm up example from the introduction illustrates this).

Theorem 9.9. *For any $k \in \{0, \ldots, p\varpi - 1\}$ the limit*

$$\lim_{\substack{X \to +\infty \\ X \equiv k \bmod p\varpi}} \frac{X\pi_\sigma^*(X)}{\Lambda^X} = \rho_k \tag{38}$$

exists (*so there is convergence along sequences of values of X that converge in the "tame detector group" $G_\sigma^* := \mathbf{Z}/p\varpi$) and is given by*

$$\rho_k = \frac{1}{\Lambda^{p\varpi} - 1} \sum_{\substack{1 \leqslant n \leqslant p\varpi \\ p \nmid n}} \frac{\Lambda^{\langle n-k \rangle}}{r_n}, \tag{39}$$

where $\langle x \rangle$ denotes the representative for $x \bmod p\varpi$ in $\{1, \ldots, p\varpi\}$.

Proof. By Proposition 9.2 we have

$$\pi_\sigma^*(X) = \sum_{\ell \leqslant X,\, p \nmid \ell} \left(\frac{\Lambda^\ell}{\ell r_\ell} + O(\Lambda^{\Theta\ell}) \right).$$

The error terms in this formula form a geometric progression and hence are $O(\Lambda^{\Theta X})$. Multiplying by Λ^{-X} and applying Lemma 9.3, we get

$$\frac{\pi_\sigma^*(X)}{\Lambda^X} = \frac{1}{X\Lambda^X} \sum_{\ell \leqslant X,\, p \nmid \ell} \Lambda^\ell \frac{1}{r_\ell} + O(1/X^2).$$

We split the sum by values of r_n, as follows:

$$\lim_{X \to +\infty} \frac{X\pi_\sigma^*(X)}{\Lambda^X} = \lim_{X \to +\infty} \frac{1}{\Lambda^X} \left(\sum_{\substack{1 \leqslant n \leqslant p\varpi \\ p \nmid n}} \frac{1}{r_n} \sum_{s=0}^{\left\lfloor \frac{X-n}{p\varpi} \right\rfloor} \Lambda^{n+sp\varpi} \right) = \lim_{X \to +\infty} \left(\sum_{\substack{1 \leqslant n \leqslant p\varpi \\ p \nmid n}} \frac{\Lambda^{p\varpi \left\lfloor \frac{X-n}{p\varpi} \right\rfloor + p\varpi + n - X}}{r_n(\Lambda^{p\varpi} - 1)} \right).$$

The limit does not converge in general, but if we put $X = Yp\varpi + k$ for fixed k and $Y \to +\infty$, we find the indicated result, since $p\varpi \left\lfloor \frac{k-n}{p\varpi} \right\rfloor + p\varpi + n - k = \langle n - k \rangle$. $\qquad\square$

We refer to the example in the introduction for some explicit computations and graphs.

Analogue of Mertens' theorem. The PNT is equivalent to the statement that the reciprocals of the primes up to X sum, up to a constant, to $\log \log X + o(1/\log X)$. Mertens' second theorem is the same statement but with the weaker error term $O(1/\log X)$. It turns out that the analogue of this last theorem in our setting does hold, and very inseparable and not very inseparable endomorphisms behave in the same way.

Proposition 9.10. *For some $c \in \mathbf{Q}$ and $c' \in \mathbf{R}$ we have $\sum_{\ell \leqslant X} P_\ell / \Lambda^\ell = c \log X + c' + O(1/X)$.*

Proof. From Proposition 9.2 we find

$$\sum_{\ell \leqslant X} P_\ell / \Lambda^\ell = \sum_{\ell \leqslant X} \left(\frac{1}{\ell r_\ell |\ell|_p^{s_\ell}} + O(\Lambda^{(\Theta - 1)\ell}) \right).$$

The error terms in this formula sum to $c'' + O(\Lambda^{(\Theta-1)X})$ for some $c'' \in \mathbf{R}$ and the main terms sum to

$$\sum_{j=1}^{\varpi} \frac{1}{r_j} B_{-s_j, j}(X),$$

where for integers $s \geqslant 0$, $\varpi > 0$, and j, we set

$$B_{s,j}(X) := \sum_{\substack{n \leqslant X \\ n \equiv j \bmod \varpi}} \frac{|n|_p^s}{n}.$$

The proposition follows from

$$B_{s,j}(X) = c_{s,j} \log X + c'_{s,j} + O(1/X), \tag{40}$$

for constants $c_{s,j} \in \mathbf{Q}$ and $c'_{s,j} \in \mathbf{R}$. The case $s = 0$ is well-known and we will thus limit ourselves to the case $s > 0$. To prove (40), we first consider the related sum

$$A_{s,j}(X) = \sum_{\substack{n \leqslant X \\ n \equiv j \bmod \varpi}} |n|_p^s$$

and we claim that

$$A_{s,j}(X) = c_{s,j} X + O(1) \quad \text{with } c_{s,j} \in \mathbf{Q}. \tag{41}$$

Then Abel summation gives

$$B_{s,j}(X) = \frac{A_{s,j}(X)}{X} + \int_1^X \frac{A_{s,j}(t)}{t^2} \, dt,$$

so (40) follows, setting $c'_{s,j} = c_{s,j} + \int_1^\infty (A_{s,j}(t) - c_{s,j}t) \, dt/t^2 \in \mathbf{R}$. To prove (41), observe that the arithmetic sequence $j + \varpi \mathbf{N}$ might or might not contain terms divisible by arbitrarily high power of p depending on whether $|j|_p \leqslant |\varpi|_p$ or $|j|_p > |\varpi|_p$. In the latter case the sequence $|n|_p$ for $n \equiv j \pmod{\varpi}$ is constant, and the asymptotic formula for $A_{s,j}$ is clear. In the former case we write k for the power of p dividing ϖ. In the formula defining $A_{s,j}$, we isolate terms with a given value of $|n|_p$. For each integer $q \geqslant k$ the number of terms $n \equiv j \pmod{\varpi}$ with $n \leqslant X$ and $|n|_p = p^{-q}$ is $p - 1/(p^{q-k+1}\varpi)X + O(1)$, the implicit constant being independent of q. We thus get the asymptotic formula

$$A_{s,j} = \sum_{q \geqslant k} p^{-sq} \left(\frac{p-1}{p^{q-k+1}\varpi} X + O(1) \right) = c_{s,j} X + O(1),$$

with $c_{s,j} = (p-1)p^{s(1-k)}/((p^{s+1}-1)\varpi)$. □

Error terms in the PNT. We now briefly discuss how to identify good main terms and error terms in the asymptotics for the number of prime orbits. From Proposition 9.2, it is immediate that

$$\pi_\sigma(X) = M(X) + O(\Lambda^{\Theta X})$$

with "main term"

$$M(X) := \sum_{\ell \leqslant X} \frac{\Lambda^\ell}{\ell r_\ell |\ell|_p^{s_\ell}}$$

depending only on the data $(p, \Lambda, \varpi, (r_n), (s_n))$ and the power saving in the error term is dictated by the zeros of the degree zeta function D_σ.

Finding Θ geometrically: Finding Θ can sometimes be approached geometrically, as follows. Recall that ξ_i are roots of the characteristic polynomial of σ acting on H^1 and all λ_i are products of such roots (corresponding to the characteristic polynomial of σ acting on $\mathrm{H}^i = \wedge^i \mathrm{H}^1$ for various i). Suppose that

$$|\xi_i|^2 = a \tag{42}$$

for all i and a fixed integer a. Then $\Lambda = a^g$ and $\Theta = 1 - 1/(2g)$, so we get an error term of the form $O(a^{g-1/2})$. By [Mumford 2008, Chapter 4, Application 2], condition (42) happens if for some polarization on A with Rosati involution $'$, we have $\sigma\sigma' = a$ in $\text{End}(A)$. In Weil's proof of the analogue of the Riemann hypothesis for abelian varieties $A/\mathbf{F}_q$, it is shown that this holds for σ the q-Frobenius with $a = q^g$.

Another expression for the main term: One may express the main term $M(X)$ as follows. For $k \in \{0, \ldots, \varpi - 1\}$, define

$$F_k(\Lambda, X) = \sum_{\substack{\ell \leqslant X \\ \ell \equiv k \bmod \varpi}} \Lambda^\ell / \ell; \tag{43}$$

then

$$M(X) = \sum_{k=0}^{\varpi-1} r_k^{-1}\left(F_k(\Lambda, X) + \sum_{i \geqslant 1} p^{(s_k-1)i}(1 - p^{-s_k}) \sum_{\substack{0 \leqslant k' < \varpi \\ p^i k' \equiv k \bmod \varpi}} F_{k'}\left(\Lambda^{p^i}, \left\lfloor \frac{X}{p^i} \right\rfloor\right)\right). \tag{44}$$

We collect the information in the following proposition.

Proposition 9.11. *With $M(X)$ the function defined in* (44) *using* (43), *depending only on the data* $(p, \Lambda, \varpi, (r_n), (s_n))$ *(i.e., the growth rate Λ and the inseparability degree pattern), we have for integer values of X,*

$$\pi_\sigma(X) = M(X) + O(\Lambda^{\Theta X})$$

where

$$\Theta = \{\text{Re}(s) : s \text{ is a zero of } D_\sigma(\Lambda^{-s})\}. \qquad \square$$

A worked example is in the introduction.

The tame case: In the tame setting, one similarly finds $\pi_\sigma^*(X) = M^*(X) + O(\Lambda^{\Theta X})$ with

$$M^*(X) = \sum_{k=0}^{\varpi-1} r_k^{-1}\left(F_k(\Lambda, X) - \frac{1}{p} \sum_{\substack{0 \leqslant k' < \varpi \\ pk' \equiv k \bmod \varpi}} F_{k'}\left(\Lambda^p, \left\lfloor \frac{X}{p} \right\rfloor\right)\right).$$

Remark 9.12. Due to its exponential growth as a function of a real variable X, it is not possible to approximate $M(\lfloor X \rfloor)$ by a continuous function with error $O(\Lambda^{\vartheta X})$ for any $\vartheta < 1$. Note that $F_k(\Lambda, X)$ can be evaluated using the Lerch transcendent.

Appendix: Adelic perturbation of power series
by Robert Royals and Thomas Ward

The result in this appendix comes from the thesis of Royals [2015], the first author, and arose there in connection with the following question about "adelic perturbation" of linear recurrence sequences. Write $|m|_S = \prod_{\ell \in S} |m|_\ell$ for $m \in \mathbf{Q}$ and S a set of primes, and for an integer sequence $a = (a_n)$ define a

function $f_{a,S}$ by $f_{a,S}(z) = \sum_{n=1}^{\infty} |a_n|_S a_n z^n$. If a is an integer linear recurrence sequence, does $f_{a,S}$ satisfy a Pólya–Carlson dichotomy? That is, does $f_{a,S}$ admit a natural boundary whenever it does not define a rational function? This remains open, but for certain classes of linear recurrence and for $|S| < \infty$, the following theorem is the key step in the argument.

Theorem A.1. *Let $a = (a_n)$ be an integer sequence with the property that for every prime ℓ there exist constants n_ℓ in $\mathbf{Z}_{>0}$, $(c_{\ell,i})_{i=0}^{n_\ell-1}$ in $\mathbf{Q}^{n_\ell}$, and $(e_{\ell,i})_{i=0}^{n_\ell-1}$ in $\mathbf{Z}_{\geqslant 0}^{n_\ell}$ such that $|a_n|_\ell = c_{\ell,k}|n|_\ell^{e_{\ell,k}}$ if $n \equiv k \mod n_\ell$. Let S be a finite set of primes and write $f(z) = \sum_{n \geqslant 1} |a_n|_S z^n$. If the sequence $(|a_n|_S)$ takes infinitely many values, then f admits the unit circle as a natural boundary. Otherwise, f is a rational function.*

The method of proof is reminiscent of Mahler's, in which functional equations allow one to conclude that certain functions have singularities along a dense set of roots of unity (compare [Bell et al. 2013]).

For the proof, it is necessary to consider a slightly more general setup. Assume that S is a finite set of primes and for each $\ell \in S$ there is an associated positive integer e_ℓ, write e for the collection $(e_\ell)_{\ell \in S}$, and write $F_{S,e,r}(z) = \sum_{n \geqslant 0} |n - r|_{S,e} z^n$ for some $r \in \mathbf{Q}$, where $|n|_{S,e} = \prod_{\ell \in S} |n|_\ell^{e_\ell}$. Notice that there is always a bound of the shape

$$\frac{A}{n^B} \ll |n - r|_\ell \leqslant \max\{1, |r|_\ell\},$$

for constants $A, B > 0$, so the radius of convergence of $F_{S,e,r}$ is 1. If $|r|_\ell > 1$ for some $\ell \in S$ then $|n - r|_\ell = |r|_\ell$ for all $n \in \mathbf{N}$, and so

$$F_{S,e,r}(z) = |r|_\ell^{e_\ell} \sum_{n \geqslant 0} |n - r|_{S-\{\ell\},e} z^n = |r|_\ell^{e_\ell} F_{S-\{\ell\},e,r}(z)$$

wherever these series are defined. Thus as far as the question of a natural boundary is concerned, we may safely assume that $|r|_\ell \leqslant 1$ for all $\ell \in S$.

Now let $\ell \in S$ be fixed. Since $|r|_\ell \leqslant 1$, we can write

$$r = r_0 + r_1 \ell + r_2 \ell^2 + \cdots$$

with $r_i \in \{0, 1, \ldots, \ell - 1\}$ for all $i \geqslant 0$. For $r \in \mathbf{Q}$ let the positive integer $r_0 + r_1 \ell + \cdots + r_{e-1} \ell^{e-1}$ be written as $r \mod \ell^e$. In particular, $r \mod \ell^e$ is the smallest nonnegative integer with

$$|r - (r \mod \ell^e)|_\ell \leqslant \ell^{-e}.$$

If $n = p_1^{e_1} \cdots p_j^{e_j}$ for distinct primes p_i, then write $r \mod n$ for the smallest nonnegative integer satisfying

$$|r - (r \mod n)|_{p_i} \leqslant p_i^{-e_i}$$

for $i = 1, \ldots, j$ (which exists by the Chinese remainder theorem).

Next we will obtain some functional equations for $F_{S,e,r}$. For $m \geqslant 0$, we write $t_m = (r - (r \mod \ell^m))/\ell^m$. Note that $|t_m|_p \leqslant 1$ for all $p \in S$ and $m \geqslant 0$. We claim that for any $m \geqslant 1$ we have the equality

$$F_{S,e,t_{m-1}}(z) = F_{S-\{\ell\},e,t_{m-1}}(z) + \ell^{-e_\ell} z^{r_{m-1}} F_{S,e,t_m}(z^\ell) - z^{r_{m-1}} F_{S-\{\ell\},e,t_m}(z^\ell). \tag{45}$$

Indeed, we compare directly the coefficients of z^n on both sides of this equation. The coefficient on the left is $|n - t_{m-1}|_{S,e}$. The coefficient on the right is $|n - t_{m-1}|_{S-\{\ell\},e}$ if $\ell \nmid (n - t_{m-1})$ and

$$|n - t_{m-1}|_{S-\{\ell\},e} + \ell^{-e_\ell} \left| \frac{n - r_{m-1}}{\ell} - t_m \right|_{S,e} - \left| \frac{n - r_{m-1}}{\ell} - t_m \right|_{S-\{\ell\},e}$$

otherwise. Since $(n - r_{m-1})/\ell - t_m = (n - t_{m-1})/\ell$ and $|\ell|_{S-\{\ell\},e} = 1$, after an easy manipulation we see that both these coefficients are equal and hence we get (45).

Combining formulæ(45) for $m = 1, \dots, s$, we obtain the equality:

$$F_{S,e,r}(z) = F_{S-\{\ell\},e,r}(z) - (\ell^{e_\ell} - 1) \sum_{k=1}^{s-1} \frac{1}{\ell^{ke_\ell}} z^{r \bmod \ell^k} F_{S-\{\ell\},e,t_k}(z^{\ell^k})$$
$$- \ell^{-(s-1)e_\ell} z^{r \bmod \ell^s} F_{S-\{\ell\},e,t_s}(z^{\ell^s}) + \ell^{-se_\ell} z^{r \bmod \ell^s} F_{S,e,t_s}(z^{\ell^s}). \quad (46)$$

Since we have $|t_s|_p \leqslant 1$ for all $p \in S$ and $s \geqslant 0$, the coefficients in the power series $F_{S-\{\ell\},e,t_s}(z^{\ell^s})$ and $F_{S,e,t_s}(z^{\ell^s})$ are bounded by 1, and hence for $|z| < 1$ we can bound the two latter terms in (46) by

$$|-\ell^{-(s-1)e_\ell} z^{r \bmod \ell^s} F_{S-\{\ell\},e,t_s}(z^{\ell^s}) + \ell^{-se_\ell} z^{r \bmod \ell^s} F_{S,e,t_s}(z^{\ell^s})| \leqslant (\ell^{-(s-1)e_\ell} + \ell^{-se_\ell}) \sum_{n \geqslant 0} |z|^{n\ell^s}.$$

Thus by passing in (46) with s to infinity, we obtain:

$$F_{S,e,r}(z) = F_{S-\{\ell\},e,r}(z) - (\ell^{e_\ell} - 1) \sum_{k \geqslant 1} \frac{1}{\ell^{ke_\ell}} z^{r \bmod \ell^k} F_{S-\{\ell\},e,t_k}(z^{\ell^k}). \quad (47)$$

Lemma A.2. *Let S be a finite set of primes, $e = \{e_\ell \mid \ell \in S\}$ the associated exponents, and $n > 1$ an integer divisible by some prime $q \notin S$. Then there is a constant $c_{n,e,S} > 0$ such that for any primitive n-th root of unity μ and for all $\lambda \in [0, 1)$ we have $|F_{S,e,r}(\lambda\mu)| < c_{n,e,S}$.*

The constant $c_{n,e,S}$ does not depend on r under the assumption that $|r|_\ell \leqslant 1$ for all $\ell \in S$.

Proof. We proceed by induction on the cardinality of S. For $S = \varnothing$ we have

$$F_{S,e,r}(z) = \sum_{m \geqslant 0} |m - r|_{\varnothing,e} z^m = \frac{1}{1 - z},$$

and the existence of the claimed constant is clear. Now suppose that $|S| \geqslant 1$, let $p \in S$ and write

$$F_{S,e,r}(z) = F_{S-\{p\},e,r}(z) - (p^{e_p} - 1) \sum_{k \geqslant 1} \frac{1}{p^{ke_p}} z^{r \bmod p^k} F_{S-\{p\},e,t_k}(z^{p^k}).$$

So,

$$|F_{S,e,r}(z)| \leqslant |F_{S-\{p\},e,r}(z)| + (p^{e_p} - 1) \sum_{k \geqslant 1} \frac{1}{p^{ke_p}} |z^{r \bmod p^k}| |F_{S-\{p\},e,t_k}(z^{p^k})|$$

$$\leqslant (p^{e_p} - 1) \sum_{k \geqslant 0} \frac{1}{p^{ke_p}} |F_{S-\{p\},e,t_k}(z^{p^k})|$$

for $|z| \leqslant 1$. If $z = \lambda\mu$ for some $\lambda \in [0, 1)$ and μ is a primitive n-th root of unity with $q \mid n$, then $z^{p^k} = \lambda'\mu'$ where $\lambda' \in [0, 1)$ and μ' is a primitive n'-th root of unity with $q \mid n'$, and n' is one of finitely many possible values. Thus by the inductive hypothesis there is a constant c with $|F_{S-\{p\},e,t_k}(z^{p^k})| < c$ for all k, and hence $|F_{S,e,r}(z)| < (p^{e_p} - 1)cp^{e_p}/(p^{e_p} - 1)$. Taking this as $c_{n,e,S}$ gives the lemma. $\qquad\square$

Lemma A.3. *Let S be a finite set of primes and let $r \in \mathbf{Q}$ be such that $|r|_p \leqslant 1$ for all $p \in S$. Suppose that $n \geqslant 1$ is an integer divisible only by primes in S, and that μ is a primitive n-th root of unity. Writing $n = p_1^{f_1} \cdots p_j^{f_j}$ where $p_1, \ldots, p_j$ are distinct primes in S and $f_i \geqslant 1$ for all $i = 1, \ldots, j$, we have*

$$|F_{S,e,r}(\lambda\mu)| \to \infty$$

as $\lambda \to 1^-$. More precisely,

$$\mathrm{Re}((-1)^j \mu^{-(r \bmod n)} F_{S,e,r}(\lambda\mu)) \to \infty$$

as $\lambda \to 1^-$ and there exists a constant $c'_{n,e,S}$ (which does not depend on r and λ) such that

$$|\mathrm{Im}((-1)^j \mu^{-(r \bmod n)} F_{S,e,r}(\lambda\mu))| < c'_{n,e,S} \quad and \quad \mathrm{Re}((-1)^j \mu^{-(r \bmod n)} F_{S,e,r}(\lambda\mu)) > -c'_{n,e,S}.$$

Proof. We again write $z = \lambda\mu$ and define the function $\varphi_{S,e,r,\mu}(\lambda)$ by the formula

$$\varphi_{S,e,r,\mu}(\lambda) = (-1)^j \mu^{-(r \bmod n)} F_{S,e,r}(\lambda\mu),$$

where j is the number of prime factors of n.

We proceed by induction on the number of distinct prime factors in n starting with $n = 1$. In this case $\varphi_{S,e,r,\mu}(\lambda) = \sum_{m \geqslant 0} |m - r|_{S,e} \lambda^m$ for each m, $\lambda^m \to 1^-$ as $\lambda \to 1^-$, and $|m - r|_{S,e} = 1$ infinitely often. This shows that the real part tends to infinity as $\lambda \to 1^-$ and is bounded from below by 0. The imaginary part is bounded as $F_{S,e,r}(\lambda)$ is real for all $\lambda \in [0, 1)$.

Now let $p_1, \ldots, p_j \in S$ be distinct, and let $n = \prod_{i=1}^j p_i^{f_i}$ with $f_i \geqslant 1$ for all i. Let $p = p_1$ and use the variables $r_0, r_1, \ldots$ to indicate the p-adic coefficients of r and $t_0, t_1, \ldots$ to indicate the values $t_k = (r - r \bmod p^k)/p^k$ for all k. Assume first that $f_1 = 1$. We will apply the functional equation (47). For all $k \geqslant 1$, μ^{p^k} is a primitive (n/p)-th root of unity and the formula $t_k = (r - r \bmod p^k)/p^k$ implies that

$$r \bmod n \equiv r \bmod p^k + p^k(t_k \bmod (n/p)) \pmod{n}.$$

Thus (47) after some manipulation gives

$$\varphi_{S,e,r,\mu}(\lambda) = \varphi_{S-\{p\},e,r,\mu}(\lambda) + (p^{e_p} - 1) \sum_{k=1}^{\infty} \frac{\lambda^{r \bmod p^k}}{p^{ke_p}} \varphi_{S-\{p\},e,t_k,\mu^{p^k}}(\lambda^{p^k}).$$

The leading term in this expression is bounded by Lemma A.2, and the inductive hypothesis applied to the terms $\varphi_{S-\{p\},e,r,\mu^{p^k}}(\lambda^{p^k})$ shows that their real part tends to $+\infty$ as $\lambda \to 1^-$ and is bounded away from $-\infty$ independently of r and λ. Since these terms appear within the geometric progression $\sum_{k=1}^{\infty} p^{-ke_p}$, we obtain that

$$\varphi_{S,e,r,\mu}(\lambda) \to \infty$$

as $\lambda \to 1^-$ and the same argument proves the latter claim. This proves the inductive step for the case $f_1 = 1$.

We will use this as the base case for a second inductive proof for $f_1 > 1$. The argument in this case is similar except that we will use the functional equation (45) instead of (47). As before, μ^p is a primitive (n/p)-th root of unity and

$$r \bmod n \equiv r \bmod p + p(t_1 \bmod (n/p)) \pmod{n}.$$

Thus (45) after some manipulation gives

$$\varphi_{S,e,r,\mu}(\lambda) = \varphi_{S-\{p\},e,r,\mu}(\lambda) + p^{-e_p}\lambda^{r \bmod p}\varphi_{S,e,t_1,\mu^p}(\lambda^p) - \lambda^{r \bmod p}\varphi_{S-\{p\},e,t_1,\mu^p}(\lambda^p).$$

The first and the third terms in this expression are bounded by Lemma A.2, and hence the claim follows immediately from the inductive hypothesis applied to the term $\varphi_{S,e,t_1,\mu^p}(\lambda^p)$. This concludes the induction. $\square$

Proof of Theorem A.1. If $c_{\ell,k} = 0$ for some $\ell \in S$ and k we will automatically take $e_{\ell,k} = 0$ as the power of $|n|_\ell$ plays no role. Another case we wish to avoid is if for some ℓ and $k \in \{0, 1, \ldots, n_\ell - 1\}$, the value $|n|_\ell$ is constant for all $n \equiv k \bmod n_\ell$. Writing v_ℓ for the ℓ-adic order, this happens exactly when $v_\ell(n_\ell) > v_\ell(k)$, and in this case $|n|_\ell = |k|_\ell$. If this is the case and $e_{\ell,k} \neq 0$, then we will set $e_{\ell,k} = 0$ and substitute $c_{\ell,k}|k|_\ell^{e_{\ell,k}}$ for $c_{\ell,k}$. Let $N = \mathrm{lcm}\{n_p \mid p \in S\}$. For each $j \in \{0, 1, \ldots, N-1\}$ consider the value of $|a_n|_S$ when $n \equiv j \bmod N$. For each p, $n \equiv j \bmod N$ and thus $n \equiv j \bmod n_p$ as $n_p \mid N$. Let $k_{p,j}$ be the unique element of $\{0, 1, \ldots, n_p - 1\}$ such that $k_{p,j} \equiv j \bmod n_p$. So

$$|a_n|_S = \prod_{p \in S}|a_n|_p = \prod_{p \in S} c_{p,k_{p,j}}|n|_p^{e_{p,k_{p,j}}}$$

as $n \equiv j \equiv k_{p,j} \bmod n_p$ for all $p \in S$. If for any nonzero n with $n \equiv j \bmod N$ we have $|a_n|_S = 0$, or equivalently $a_n = 0$, we define $S_j = \varnothing$ and $d_j = 0$. If this is the case, then it follows that for this value n

$$0 = \prod_{p \in S} c_{p,k_{p,j}}|n|_p^{e_{p,k_{p,j}}}$$

and $|n|_p^{e_{p,k_{p,j}}} \neq 0$ implies that $c_{p,k_{p,j}} = 0$ for some $p \in S$. This in turn implies that $|a_m|_S = 0$ and hence $a_m = 0$ for any $m \equiv j \bmod N$. If, on the other hand, for some $n \equiv j \bmod N$ we have $|a_n|_S \neq 0$ then for all $m \equiv j \bmod N$ we have $|a_m|_S \neq 0$ and hence $c_{p,k_{p,j}} \neq 0$ for all $p \in S$. If for a prime $p \in S$ we have $v_p(N) > v_p(j)$, then for all $n \equiv j \bmod N$ we have $|n|_p = |j|_p$. We will split S into the disjoint union $S_j \sqcup S_j' \sqcup S_j''$, where

$$S_j = \{p \in S \mid v_p(N) \leqslant v_p(j) \text{ and } e_{p,k_{p,j}} \neq 0\},$$
$$S_j' = \{p \in S \mid v_p(N) > v_p(j) \text{ and } e_{p,k_{p,j}} \neq 0\},$$
$$S_j'' = \{p \in S \mid v_p(N) > v_p(j) \text{ and } e_{p,k_{p,j}} = 0\}.$$

Thus for all $n \equiv j \bmod N$ we have

$$|a_n|_S = \prod_{p \in S} c_{p,k_{p,j}} \cdot \prod_{p \in S_j'} |j|_p^{e_{p,k_{p,j}}} \cdot |n|_{S_j, e^{(j)}},$$

where $e^{(j)}$ denotes the collection of exponents $\{e_{p,k_j} \mid p \in S_j\}$. Set

$$d_j = \prod_{p \in S} c_{p,k_{p,j}} \cdot \prod_{p \in S'_j} |j|_p^{e_{p,k_{p,j}}}$$

and $|a_n|_S = d_j |n|_{S_j, e^{(j)}}$ for all $n \equiv j \bmod N$.

Assume that the sequence $(|a_n|_S)$ takes infinitely many values. This implies that there exists some j for which S_j is nonempty. By our assumption, for such j we have $d_j \neq 0$. Consider the family of sets $\{S_j \mid 0 \leqslant j < N\}$, partially ordered by inclusion. Since it is finite and the S_j are not all empty, there is a nonempty maximal element S_{j_0}. Write

$$f(z) = \sum_{n=1}^{\infty} |a_n|_S z^n = \sum_{j=0}^{N-1} \sum_{n \equiv j\,(N)} |a_n|_S z^n = \sum_{j=0}^{N-1} f_j(z)$$

where

$$
\begin{aligned}
f_j(z) &= \sum_{n \equiv j\,(N)} |a_n|_S z^n \\
&= \sum_{n \equiv j\,(N)} d_j |n|_{S_j, e^{(j)}} z^n \\
&= \sum_{k=0}^{\infty} d_j |kN + j|_{S_j, e^{(j)}} z^{kN+j} \\
&= d_j |N|_{S_j, e^{(j)}} \sum_{k=0}^{\infty} |k + j/N|_{S_j, e^{(j)}} z^{kN+j} \\
&= d_j |N|_{S_j, e^{(j)}} z^j g_j(z^N)
\end{aligned}
$$

with $g_j(z) = F_{S_j, e^{(j)}, -j/N}(z)$. Thus $f = h_1 + h_2$, where h_1 is the sum of the f_j with $S_j = S_{j_0}$ and h_2 is the sum of the f_j with $S_j \neq S_{j_0}$. Let $n = \prod_{q \in S_{j_0}} q^{f_q}$ be an integer divisible by every prime in S_{j_0} and by no other primes such that for each $q \in S_{j_0}$ we have $f_q > v_q(N)$ and let μ be a primitive n-th root of unity. If j with $0 \leqslant j < N$ has $S_j \neq S_{j_0}$ then $f_j(\lambda \mu) = d_j |N|_{S_j, e^{(j)}} (\lambda \mu)^j g_j(\lambda^N \mu^N)$ is bounded as $\lambda \to 1^-$ by Lemma A.2 as μ^N is an n/N-th root of unity and n/N is divisible by every prime in S_{j_0} and hence by some prime not in S_j by maximality of S_{j_0}. Thus $|h_2(\lambda \mu)|$ is bounded as $\lambda \to 1^-$. Suppose instead that $S_j = S_{j_0}$. By Lemma A.3 we have that

$$\operatorname{Re}((-1)^m (\mu^N)^{-(-j/N \bmod n/N)} g_j(z^N)) \to \infty$$

as $\lambda \to 1^-$ where $m = |S_{j_0}|$. Equivalently,

$$\operatorname{Re}((-1)^m \mu^{(j \bmod n)} g_j(z^N)) \to \infty,$$

and thus

$$\operatorname{Re}((-1)^m z^j g_j(z^N)) \to \infty$$

as $\lambda \to 1^-$. As the real part of every term in $h_1(z)$ goes to ∞, this means that

$$\text{Re}((-1)^m f(\lambda\mu)) \to \infty$$

as $\lambda \to 1^-$. Since this is true for any μ that is a $\left(\prod_{q \in S_{j_0}} q^{f_q}\right)$-th root of unity with each $f_q > v_q(N)$, these singularities form a dense set on the unit circle. It follows that f admits a natural boundary on the unit circle.

For the second part of the theorem, assume that the sequence $(|a_n|_S)$ takes only finitely many values. Then $(|a_n|_S)$ is periodic modulo N, and thus

$$f(z) = \sum_{j=1}^{N} \sum_{n \equiv j\,(N)} |a_j|_S z^n = \sum_{j=1}^{N} |a_j|_S \sum_{m=0}^{\infty} z^{mN+j} = \sum_{j=1}^{N} |a_j|_S \frac{z^j}{1-z^N},$$

completing the proof. $\qquad\square$

References

[Artin and Mazur 1965] M. Artin and B. Mazur, "On periodic points", *Ann. of Math.* (2) **81** (1965), 82–99. MR Zbl

[Baake et al. 2010] M. Baake, E. Lau, and V. Paskunas, "A note on the dynamical zeta function of general toral endomorphisms", *Monatsh. Math.* **161**:1 (2010), 33–42. MR Zbl

[Bell and Gerhold 2007] J. P. Bell and S. Gerhold, "On the positivity set of a linear recurrence sequence", *Israel J. Math.* **157** (2007), 333–345. MR Zbl

[Bell et al. 2013] J. P. Bell, M. Coons, and E. Rowland, "The rational-transcendental dichotomy of Mahler functions", *J. Integer Seq.* **16**:2 (2013), Article 13.2.10, 11. MR Zbl

[Bell et al. 2014] J. Bell, R. Miles, and T. Ward, "Towards a Pólya–Carlson dichotomy for algebraic dynamics", *Indag. Math.* (*N.S.*) **25**:4 (2014), 652–668. MR Zbl

[Bellagh and Bézivin 2011] A. Bellagh and J.-P. Bézivin, "Quotients de suites holonomes", *Ann. Fac. Sci. Toulouse Math.* (6) **20**:1 (2011), 135–166. MR Zbl

[Benettin et al. 2008] G. Benettin, A. Carati, L. Galgani, and A. Giorgilli, "The Fermi–Pasta–Ulam problem and the metastability perspective", pp. 152–189 in *The Fermi–Pasta–Ulam problem*, edited by G. Gallavotti, Lecture Notes in Phys. **728**, Springer, 2008. MR Zbl

[Bridy 2012] A. Bridy, "Transcendence of the Artin–Mazur zeta function for polynomial maps of $\mathbb{A}^1(\overline{\mathbb{F}}_p)$", *Acta Arith.* **156**:3 (2012), 293–300. MR Zbl

[Bridy 2016] A. Bridy, "The Artin–Mazur zeta function of a dynamically affine rational map in positive characteristic", *J. Théor. Nombres Bordeaux* **28**:2 (2016), 301–324. MR Zbl

[Deuring 1941] M. Deuring, "Die Typen der Multiplikatorenringe elliptischer Funktionenkörper", *Abh. Math. Sem. Hansischen Univ.* **14** (1941), 197–272. MR Zbl

[Dimitrov 2013] V. Dimitrov, "A note on a generalization of the Hadamard quotient theorem", 2013. arXiv

[Dwork 1960] B. Dwork, "On the rationality of the zeta function of an algebraic variety", *Amer. J. Math.* **82** (1960), 631–648. MR Zbl

[Everest et al. 2005] G. Everest, V. Stangoe, and T. Ward, "Orbit counting with an isometric direction", pp. 293–302 in *Algebraic and topological dynamics*, edited by S. Kolyada et al., Contemp. Math. **385**, Amer. Math. Soc., Providence, RI, 2005. MR Zbl

[Everest et al. 2007] G. Everest, R. Miles, S. Stevens, and T. Ward, "Orbit-counting in non-hyperbolic dynamical systems", *J. Reine Angew. Math.* **608** (2007), 155–182. MR Zbl

[Flajolet et al. 2004/06] P. Flajolet, S. Gerhold, and B. Salvy, "On the non-holonomic character of logarithms, powers, and the nth prime function", *Electron. J. Combin.* **11**:2 (2004/06), Article 2, 16. MR Zbl

[Freeman and Lauter 2008] D. Freeman and K. Lauter, "Computing endomorphism rings of Jacobians of genus 2 curves over finite fields", pp. 29–66 in *Algebraic geometry and its applications*, edited by J. Chaumine et al., Ser. Number Theory Appl. **5**, World Sci. Publ., Hackensack, NJ, 2008. MR Zbl

[Friedland 1991] S. Friedland, "Entropy of polynomial and rational maps", *Ann. of Math.* (2) **133**:2 (1991), 359–368. MR Zbl

[Gel' fond 1960] A. O. Gel' fond, *Transcendental and algebraic numbers*, Dover Publications, New York, 1960. MR Zbl

[Goren 2002] E. Z. Goren, *Lectures on Hilbert modular varieties and modular forms*, CRM Monograph Series **14**, American Mathematical Society, Providence, RI, 2002. MR Zbl

[Grieve 2017] N. Grieve, "Reduced norms and the Riemann–Roch theorem for Abelian varieties", *New York J. Math.* **23** (2017), 1087–1110. MR Zbl

[Grothendieck 1965] A. Grothendieck, "Formule de Lefschetz et rationalité des fonctions *L*", in *Séminaire Bourbaki* 1964/1965 (Exposé 279), W. A. Benjamin, Amsterdam, 1965. Reprinted as pp. 41–55 in *Séminaire Bourbaki* **9**, Soc. Math. France, Paris, 1995. MR Zbl

[Harris and Sibuya 1985] W. A. Harris, Jr. and Y. Sibuya, "The reciprocals of solutions of linear ordinary differential equations", *Adv. in Math.* **58**:2 (1985), 119–132. MR Zbl

[Hinkkanen 1994] A. Hinkkanen, "Zeta functions of rational functions are rational", *Ann. Acad. Sci. Fenn. Ser. A I Math.* **19**:1 (1994), 3–10. MR Zbl

[Knieper 1997] G. Knieper, "On the asymptotic geometry of nonpositively curved manifolds", *Geom. Funct. Anal.* **7**:4 (1997), 755–782. MR Zbl

[Lam 1991] T. Y. Lam, *A first course in noncommutative rings*, Graduate Texts in Mathematics **131**, Springer, 1991. MR Zbl

[LMFDB Collaboration 2013] LMFDB Collaboration, "The *L*-functions and modular forms database", electronic reference, 2013, Available at http://www.lmfdb.org. Home page of the abelian variety isogeny class 2.5.a$_a$ over $\mathbb{F}_5$.

[Milne 2008] J. S. Milne, "Abelian Varieties (v2.00)", 2008, Available at http://www.jmilne.org/math/.

[Milne 2013] J. S. Milne, "Lectures on etale cohomology (v2.21)", 2013, Available at www.jmilne.org/math.

[Mumford 2008] D. Mumford, *Abelian varieties*, 2nd ed., Tata Institute of Fundamental Research Studies in Mathematics **5**, Tata Institute of Fundamental Research, Bombay, 2008. MR Zbl

[Norman and Oort 1980] P. Norman and F. Oort, "Moduli of abelian varieties", *Ann. of Math.* (2) **112**:3 (1980), 413–439. MR Zbl

[Parry and Pollicott 1983] W. Parry and M. Pollicott, "An analogue of the prime number theorem for closed orbits of Axiom A flows", *Ann. of Math.* (2) **118**:3 (1983), 573–591. MR Zbl

[Pollicott and Sharp 1998] M. Pollicott and R. Sharp, "Exponential error terms for growth functions on negatively curved surfaces", *Amer. J. Math.* **120**:5 (1998), 1019–1042. MR Zbl

[van der Poorten 1988] A. J. van der Poorten, "Solution de la conjecture de Pisot sur le quotient de Hadamard de deux fractions rationnelles", *C. R. Acad. Sci. Paris Sér. I Math.* **306**:3 (1988), 97–102. MR Zbl

[Rosen 2002] M. Rosen, *Number theory in function fields*, Graduate Texts in Mathematics **210**, Springer, 2002. MR Zbl

[Royals 2015] R. Royals, *Arithmetic and dynamical systems*, Ph.D. thesis, University of East Anglia, 2015, Available at https://ueaeprints.uea.ac.uk/57191.

[Rumely 1988] R. Rumely, "Notes on van der Poorten's proof of the Hadamard quotient theorem, I, II", pp. 349–382, 383–409 in *Séminaire de Théorie des Nombres, Paris 1986–87*, edited by C. Goldstein, Progr. Math. **75**, Birkhäuser Boston, Boston, 1988. MR Zbl

[SageMath 2016] The Sage Developers, *SageMath, the Sage Mathematics Software System (Version 8.0)*, Sage Development Team, 2016, Available at http://www.sagemath.org.

[Smale 1967] S. Smale, "Differentiable dynamical systems", *Bull. Amer. Math. Soc.* **73** (1967), 747–817. MR Zbl

[Stanley 1980] R. P. Stanley, "Differentiably finite power series", *European J. Combin.* **1**:2 (1980), 175–188. MR Zbl

[Stanley 2012] R. P. Stanley, *Enumerative combinatorics, Volume 1*, 2nd ed., Cambridge Studies in Advanced Mathematics **49**, Cambridge University Press, 2012. MR Zbl

[Waterhouse 1969] W. C. Waterhouse, "Abelian varieties over finite fields", *Ann. Sci. École Norm. Sup.* (4) **2** (1969), 521–560. MR Zbl

jakub.byszewski@uj.edu.pl   *Wydział Matematyki i Informatyki, Uniwersytet Jagielloński, Kraków, Poland*

g.cornelissen@uu.nl   *Mathematisch Instituut, University of Utrecht, The Netherlands*

aradesh@gmail.com   *School of Mathematics, University of East Anglia, Norwich, United Kingdom*

t.b.ward@leeds.ac.uk   *Ziff Building, University of Leeds, United Kingdom*

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory