Algebra & Number Theory

Volume 6
2012
No. 7

msp

Algebra & Number Theory

msp.berkeley.edu/ant

See inside back cover or www.jant.org for submission instructions.

The subscription price for 2012 is US $175/year for the electronic version, and $275/year (+$40 shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Algebra & Number Theory (ISSN 1937-0652) at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

http://msp.org/

A NON-PROFIT CORPORATION

Typeset in LaTeX

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1289

On the rank of the fibers of rational elliptic surfaces

Cecília Salgado

We consider an elliptic surface $\pi : \mathscr{E} \to \mathbb{P}^1$ defined over a number field k and study the problem of comparing the rank of the special fibers over k with that of the generic fiber over $k(\mathbb{P}^1)$. We prove, for a large class of rational elliptic surfaces, the existence of infinitely many fibers with rank at least equal to the generic rank plus two.

1. Introduction

Let k be a number field, B a projective curve and $\pi : \mathscr{E} \to B$ an elliptic surface over k, that is, a projective surface endowed with a morphism π such that almost all fibers are genus-one curves and such that there is a section σ of π defined over k that will be fixed as the zero section. The generic fiber $\mathscr{E}_\eta$ is an elliptic curve over the function field $k(B)$. Since $k(B)$ is finitely generated over $\mathbb{Q}$, the Mordell–Weil Theorem is still valid in this context (Lang–Néron), and hence the set of $k(B)$-rational points of $\mathscr{E}_\eta$ is a finitely generated abelian group. Since all but finitely many fibers of π are elliptic curves over the number field k, it is natural to be interested in comparing the rank r of $\mathscr{E}_\eta(k(B))$ and the rank r_t of the Mordell–Weil group of a fiber $E_t(k)$ for $t \in B(k)$.

A theorem on specializations by Néron [1956] or its refinement by Silverman [1994, Theorem III.11.4; 1983] in the case where the base is a curve tells us that if $\mathscr{E}$ is nonsplit, then for all but finitely many fibers we have $r_t \geq r$.

Billard [1998, Theorem C] showed that if we assume $\mathscr{E}$ to be $\mathbb{Q}$-rational (birational to $\mathbb{P}^2$ over $\mathbb{Q}$) and nonisotrivial, then

$$\#\{t \in B(k) \mid r_t \geq r+1\} = \infty.$$

Three natural questions arise:

1) Can we replace $\mathbb{Q}$ by an arbitrary number field k and the hypothesis that $\mathscr{E}$ is k-rational by the hypothesis that $\mathscr{E}$ is k-unirational?

MSC2010: primary 14J27; secondary 11G05, 14D99.
Keywords: elliptic surface, rational surface, Mordell–Weil group, elliptic curve.

2) Can we improve the bound, that is, have $r_t \geq r+2$ for infinitely many $t \in B(k)$?
3) Can we obtain a similar result for elliptic surfaces that are geometrically rational (that is, fixed an algebraic closure $\bar{k}$ of k, $\mathscr{E}$ is $\bar{k}$-rational) but not k-rational or for non(geometrically) rational elliptic surfaces, such as K3 surfaces?

We remind the reader that a geometrically irreducible algebraic variety V is said to be k-unirational if there is a rational map of finite degree $\mathbb{P}^n \dashrightarrow V$ defined over k.

We will give a positive answer to question 1) and a partial answer to questions 2) and 3) in this article. Question 3) for K3 surfaces is addressed in the author's Ph.D. thesis [Salgado 2009] and will be explored in another article.

Let X be a smooth projective rational surface. We denote by ω_X^2 the self-intersection number of the canonical sheaf on X. This will be called the degree of X and will be denoted by d_X or by d when the dependency on X is clear. The corollary to the following theorem answers question 1).

Theorem 1.1. *Let $\pi : \mathscr{E} \to B \simeq \mathbb{P}^1$ be a k-unirational elliptic surface defined over a number field k. There is a curve $C \to B$ such that $C \simeq_k \mathbb{P}^1$ and*

$$\operatorname{rank} \mathscr{E}_C(k(C)) \geq \operatorname{rank} \mathscr{E}(k(B)) + 1,$$

where $\mathscr{E}_C = \mathscr{E} \times_B C$.

Since the curve C has infinitely many k-rational points, an application of Néron's or Silverman's Specialization Theorem yields the following corollary.

Corollary 1.2. *Let $\pi : \mathscr{E} \to B$ be a k-unirational elliptic surface. Then*

$$\#\{t \in B(k) \mid r_t \geq r+1\} = \infty.$$

Remark 1. Since rational surfaces of degree $d_X \geq 3$ such that $X(k) \neq \varnothing$, and del Pezzo surfaces of degree 2 having a k-rational point outside a certain divisor are always k-unirational, we conclude, from the remark above, that the class of rational elliptic surfaces to which Theorem 1.1 applies is quite large. For example, it contains all rational elliptic surfaces defined over k with three distinct types of reducible fiber and/or a fiber with a double component not of type I_0^*. But surfaces with generic rank over k equal to zero and no reducible fibers always have as k-minimal models del Pezzo surfaces of degree one and are therefore excluded from the hypothesis of Theorem 1.1.

Remark 2. Showing a result such as the above corollary for a rational elliptic surface having a del Pezzo surface of degree one as a k-minimal model is equivalent to showing that the k-rational points on X are Zariski dense, a well known open problem. In [Ulas 2008] one can find partial results towards that direction.

The following theorems, or more precisely their corollary, answer question 2) after strengthening the hypothesis of Theorem 1.1. In order to state them we introduce the following terminology: if $f : X \to Y$ is a birational morphism of surfaces, passing to an algebraic closure, it is composed of monoidal transformations or blow ups of points; we will refer to the set of $\bar{k}$-points where f^{-1} is not defined as the *blow up locus of* f. We will denote it by $\mathrm{Bl}(f)$. Note that if X is a rational elliptic surface and Y a rational model obtained after contracting (-1)-curves, then $\mathrm{Bl}(f)$ contains d_Y not necessarily distinct points, thus $\#\,\mathrm{Bl}(f) \le d_Y$ where # denotes the number of distinct points. If f is defined over k, the set $\mathrm{Bl}(f)$ is composed of $\mathrm{Gal}(\bar{k}|k)$-orbits.

Theorem 1.3. *Let $\pi : \mathscr{E} \to B \simeq \mathbb{P}^1$ be a rational elliptic surface defined over a number field k such that there is a k-birational morphism $f : \mathscr{E} \to \mathbb{P}^2$ in which the zero section of $\mathscr{E}$ is contracted to a point $p_1 \in \mathbb{P}^2(k)$. Suppose that the blow up locus of f contains at least one orbit distinct from the one given by p_1 whose points are*, *together with p_1*, *in general position. Suppose also that $\mathscr{E}$ has at most one nonreduced fiber.*

Then there exists a finite covering $C \to B$ such that $C(k)$ is infinite and the surface $\mathscr{E}_C = \mathscr{E} \times_B C$ satisfies

$$\operatorname{rank} \mathscr{E}_C(k(C)) \ge \operatorname{rank} \mathscr{E}(k(B)) + 2.$$

Remark 3. It is simple to construct examples of rational elliptic surfaces satisfying the hypothesis of Theorem 1.3. Let f and g be two arbitrary cubics in $\mathbb{P}^2$ whose equations have coefficients in k. Suppose that they pass through two k-rational points p_1 and p_2. Then the elliptic surface given by the blow up of the intersection locus of f and g is certainly in this class.

Remark 4. The assumption that $\mathscr{E}$ has at most one nonreduced fiber excludes only one configuration of singular fibers, namely, (I_0^*, I_0^*). This case is left out because the surface might become trivial after a quadratic base change; see Lemma 2.1.

The result above also holds for some rational elliptic surfaces whose k-minimal models, after contracting the zero section, are not isomorphic to $\mathbb{P}^2$ but to other rational surfaces defined over k. These are the subject of the next theorem.

Theorem 1.4. *Let $\pi : \mathscr{E} \to B \simeq \mathbb{P}^1$ be a rational elliptic surface defined over a number field k. Suppose $\mathscr{E}$ does not have reducible fibers. Let X be a k-minimal model of $\mathscr{E}$ such that there exists a birational morphism $f : \mathscr{E} \to X$ in which the zero section of $\mathscr{E}$ is contracted. Suppose that X has degree d and satisfies one of the following*:

i) $d = 4, 5$ *or* 8.

ii) $d = 6$, $\#\,\mathrm{Bl}(f) = 6$ *and the largest* $\mathrm{Gal}(\bar{k}|k)$*-orbit in it has at most four points.*

Then there exists a finite covering $C \to B$ such that $C(k)$ is infinite and the surface $\mathscr{E}_C = \mathscr{E} \times_B C$ satisfies

$$\operatorname{rank} \mathscr{E}_C(k(C)) \geq \operatorname{rank} \mathscr{E}(k(B)) + 2.$$

Elliptic surfaces satisfying the hypothesis of the previous theorem are always the blow up of a del Pezzo surface Y. Indeed, let Y be the surface obtained after contracting the zero section of $\mathscr{E}$. Since Y is a rational surface, all we have to check is that $-K_Y$ is ample. In fact, the anticanonical divisor of Y satisfies $(-K_Y)^2 > 0$ and $-K_Y.D > 0$. The former because $K_{\mathscr{E}}^2 = 0$ and Y is obtained by contracting a curve in $\mathscr{E}$. The latter follows from the fact that $\mathscr{E}$ has no reducible fibers and thus $-K_{\mathscr{E}}.C \geq 0$ for all $C \in \operatorname{Div}(\mathscr{E})$ with equality if and only if $C \equiv -K_{\mathscr{E}}$. By the Nakai–Moishezon Theorem $-K_Y$ is ample.

Once again an application of Néron–Silverman's Specialization Theorem yields the following corollary to Theorems 1.3 and 1.4.

Corollary 1.5. *Let $\pi : \mathscr{E} \to B$ be an elliptic surface as in Theorem 1.3 or 1.4. For $t \in B(k)$, let r_t be the rank of the fiber above the point t and r the generic rank. Then*

$$\#\{t \in B(k) \mid r_t \geq r + 2\} = \infty.$$

Remark 5. Since geometrically, that is, over $\bar{k}$, a rational elliptic surface is isomorphic to the blow up of nine not necessarily distinct points in $\mathbb{P}^2$ (see Proposition 2.2), a k-minimal model X of a rational elliptic surface satisfies $1 \leq \omega_X^2 \leq 9$. As we suppose that the elliptic surface always has a section defined over the base field k which is contractible, X also verifies $X(k) \neq \varnothing$.

Remark 6. Theorems 1.3 and 1.4 are valid for a larger class of rational elliptic surfaces. The choice of the cases stated was made for the sake of simplicity. The reader is invited to consult the appendix for examples of cases to which the conclusion of Theorems 1.3 and 1.4 still applies.

This text is divided as follows: Sections 2 and 3 contain geometric and arithmetic preliminaries, respectively. Section 4 is dedicated to the proof of Theorem 1.1. It contains a key proposition that reduces the proof of this theorem to the construction of a linear pencil of genus zero curves defined over the base field k. The proofs of Theorems 1.3 and 1.4 are given in Section 5, where we give a case-by-case construction of two pencils of curves of genus zero satisfying certain geometric conditions. The last section sheds some light from analytic number theory into the problem. There, we combine the results obtained in this article with analytic conjectures to get better, but conditional, bounds for the ranks.

2. Geometric preliminaries

2A. ***Base change.*** Let $\pi : \mathscr{E} \to B$ be an elliptic surface endowed with a section and $\iota : C' \subset \mathscr{E}$ an irreducible curve. Let $\nu : C \to C'$ be its normalization. If C' is not contained in a fiber, the composition $\varphi = \pi \circ \iota \circ \nu$ is a finite covering $\varphi : C \to B$.

$$\begin{tikzcd} & & \mathscr{E} \arrow[d, "\pi"] \\ C \arrow[r, "\nu"] \arrow[rr, bend right, "\varphi"'] & C' \arrow[ur, "\iota"] \arrow[r] & B \end{tikzcd}$$

We obtain a new elliptic surface by taking the following fibered product:

$$\pi_C : \mathscr{E}_C = \mathscr{E} \times_B C \to C.$$

Each section σ of $\mathscr{E}$ naturally induces a section in $\mathscr{E}_C$:

$$(\sigma, \mathrm{id}) : C = B \times_B C \to \mathscr{E} \times_B C.$$

We call these sections *old sections*. The surface $\mathscr{E}_C$ has also a *new section* given by

$$\sigma_C^{\mathrm{new}} = (\iota \circ \nu, \mathrm{id}) : C \to \mathscr{E} \times_B C.$$

Remark 7. Given a smooth elliptic surface $\mathscr{E}$, the elliptic surface obtained after a base change of $\mathscr{E}$ is not necessarily smooth. When this is the case, we will replace the base-changed surface by its relatively minimal model without further notice.

Remark 8. If C is not contained in a fiber nor in a section, then the new section is different from the *old* ones, but it is not necessarily linearly independent in the Mordell–Weil group.

Remark 9. Since we want $B(k)$ to be infinite, we are naturally led to consider the base curves B such that $B \simeq \mathbb{P}^1$, or with geometric genus $g(B) = 1$ and rank $B(k) \geq 1$.

If $\mathscr{E}$ is a rational elliptic surface then, in general, after a quadratic base change we obtain an elliptic K3 surface $\mathscr{E}'$; nevertheless, if $\mathscr{E}$ has a nonreduced fiber, that is, a fiber of type $*$, and the base change is ramified above the place corresponding to this fiber and above the place corresponding to a reduced fiber, then the base-changed surface $\mathscr{E}'$ is still rational.

Lemma 2.1. *Let $\mathscr{E} \to B$ be a rational elliptic surface. Let $\varphi : C \to B$ be a degree-two morphism where C is rational. Then one of the following occurs*:

i) *$\mathscr{E}$ has a nonreduced fiber, that is, a fiber of type $*$, and the morphism φ is ramified above the place corresponding to it and above the place corresponding to a reduced fiber. In this case, $\mathscr{E}_C = \mathscr{E} \times_B C$ is a rational elliptic surface.*

ii) *$\mathscr{E}$ has two nonreduced fibers, which are necessarily of type I_0^*, and the morphism φ is ramified above both nonreduced fibers. In this case $\mathscr{E}_C \simeq E \times D$ where E is an elliptic curve and D is a curve of genus zero.*

iii) *$\mathscr{E}_C$ is a K3 surface.*

Proof. Let v be a place of B and w be a place of C above it. If w is not ramified, then the type of the fiber $(\mathscr{E}_C)_w$ is the same as that of $(\mathscr{E})_v$. In this case, since the degree of φ is two, there are two places w, w' above v. If w ramifies above v and $\mathscr{E}_v$ is a singular fiber, then fiber type changes, namely:

a) A fiber $(\mathscr{E})_v$ of type $*$ induces a fiber $(\mathscr{E}_C)_w$ whose contribution is $d_w = 2d_v - 12$, where d_v denotes the contribution of $(\mathscr{E})_v$, that is, the local Euler number of $\mathscr{E}_v$.

b) A reduced fiber, that is, a fiber $(\mathscr{E})_v$ of type I_n, II, III or IV transforms to a fiber $(\mathscr{E}_C)_w$ whose contribution to the Euler number of the surface is $d_w = 2d_v$.

Since $\mathscr{E}$ is rational, we have

$$\sum_{v \in B} d_v = 12.$$

Thus if $\mathscr{E}$ has a fiber, $(\mathscr{E})_v$, of type $*$, and φ is ramified above the place corresponding to it and above the place corresponding to a reduced fiber, then the Euler number of $\mathscr{E}_C$ is $\sum_{w \in C} d_w = 2d_v - 12 + 2(12 - d_v) = 12$, implying that $\mathscr{E}_C$ is rational.

If $\mathscr{E}$ has two nonreduced fibers $\mathscr{E}_{v_1}$ and $\mathscr{E}_{v_2}$, then since each contributes at least 6 to the Euler number, which is 12, we have that the contribution of each must be exactly 6. This implies that each nonreduced fiber is of type I_0^* and moreover, that these are the only singular fibers of $\mathscr{E}$. If φ ramifies above both $\mathscr{E}_{v_1}$ and $\mathscr{E}_{v_2}$ then by a) they both become nonsingular fibers. Since these were the unique singular fibers, the base-changed surface $\mathscr{E}_C$ has only smooth fibers. This entails $\mathscr{E}_C \simeq E \times D$ where E is an elliptic curve and d is a curve of genus zero.

Otherwise, that is, if φ is ramified only above reduced fibers, then the Euler number of $\mathscr{E}_C$ is $2d_v + 2(12 - d_v) = 24$, and hence $\mathscr{E}_C$ is a K3 surface. □

2B. ***Construction of rational elliptic surfaces.*** Let F and G be two distinct cubic curves in $\mathbb{P}^2$. We will also denote by F and G the two homogeneous cubic polynomials associated to these curves. Suppose F is smooth.

The pencil of cubics generated by F and G,

$$\Gamma := \{tF + uG \mid (t:u) \in \mathbb{P}^1\},$$

has nine base points (counted with multiplicities), namely the intersection points of the curves F and G. The blow up of these points in $\mathbb{P}^2$ defines a rational elliptic surface $\mathscr{E}_\Gamma$.

Conversely, over an algebraically closed field, we have the following proposition.

Proposition 2.2 [Miranda 1980]. *Over $\bar{k}$, every rational elliptic surface with a section is isomorphic to a surface $\mathscr{E}_\Gamma$ for a pencil of cubics Γ as above.*

This proposition motivates the following definition.

Definition 2.3. Let $\mathscr{E}$ be a rational elliptic surface. We say that a cubic pencil Γ in $\mathbb{P}^2$ induces $\mathscr{E}$ if $\mathscr{E}$ is $\bar{k}$-isomorphic to $\mathbb{P}^2$ blown up at the base locus of Γ.

The choice of the pencil Γ is noncanonical.

Suppose Γ induces $\mathscr{E}$ and $p_1, \dots, p_r$ are the distinct base points of Γ. The Picard group of $\mathscr{E}$ is generated by the strict transform of a cubic in Γ, the exceptional curves above each p_i and some of the (-2)-curves obtained in the process of blowing up the p_i in case the points have multiplicity strictly larger than one as base points or are in a nongeneral position (for example, three on a line or six on a conic). As the exceptional curves are the sections of the elliptic fibration, the above gives us the following information about the (geometric) Mordell–Weil group of $\mathscr{E}$.

Lemma 2.4. *Let Γ be a pencil of cubics in $\mathbb{P}^2$ and $\mathscr{E}$ the elliptic surface induced by Γ. Let s be the number of distinct base points of Γ and $\bar{r}$ be the geometric Mordell–Weil rank of $\mathscr{E}$, that is, the rank of the Mordell–Weil group over the algebraic closure $\bar{k}$. Then $\bar{r} \le s - 1$.*

Over a number field, a rational elliptic surface may have a minimal model other than $\mathbb{P}^2$. Hence we cannot assure the existence of a k-birational morphism between $\mathscr{E}$ and $\mathbb{P}^2$. We treat this situation in Section 3A.

2C. ***Néron–Tate height in elliptic surfaces.*** In [Shioda 1990], Shioda developed the theory of Mordell–Weil lattices. He remarked that the Néron–Severi and the Mordell–Weil groups modulo torsion have a lattice structure endowed with a pairing given essentially by the intersection pairing on the surface. In the Mordell–Weil group it coincides with the Néron–Tate height.

To define this pairing we need to introduce some notation:

Let Θ_v be a fiber with m_v components denoted by $\Theta_{v,i}$. Let $\Theta_{v,0}$ be the zero component (the one that intersects the zero section) and $A_v = ((\Theta_{v,i}.\Theta_{v,j}))_{1\le i,j\le m_v-1}$ a (negative definite) matrix. The pairing is given by the following formula:

$$\langle P, Q\rangle = \chi + (P.O) + (Q.O) - (P.Q) - \sum_{v\in R} \mathrm{contr}_v(P, Q),$$

where χ is the Euler characteristic of the surface, R is the set of reducible fibers, $(P.Q)$ the intersection of the sections given by P and Q, and $\mathrm{contr}_v(P, Q)$ gives the local contribution at v according to the intersection of P and Q with the fiber Θ_v: if P intersects $\Theta_{v,i}$ and Q intersects $\Theta_{v,j}$ then $\mathrm{contr}_v(P, Q) = -(A_v^{-1})_{i,j}$ if

$i, j \geq 1$ and $\mathrm{contr}_v(P, Q) = 0$ if one of the sections cuts the zero component. For a table of all possible values of contr_v according to the fiber type of Θ_v; see [Shioda 1990].

We will only use the fact that the Néron–Tate height of a section can be computed in terms of its numerical class.

3. Arithmetic preliminaries

3A. *Minimal models over perfect fields.* The theory in this subsection was developed by Enriques, Manin [1966; 1967] and Iskoviskikh [1979]. We state the main results. For proofs, we invite the reader to look at the bibliography cited above.

Theorem 3.1. *Let X be a smooth minimal rational surface defined over a perfect field k and let* $\mathrm{Pic}(X)$ *denote its Picard group over k. Then X is isomorphic to a surface in one of the following families*:

I. *A del Pezzo surface with* $\mathrm{Pic}(X) \simeq \mathbb{Z}$.

II. *A conic bundle such that* $\mathrm{Pic}(X) \simeq \mathbb{Z} \oplus \mathbb{Z}$.

Reciprocally, if X belongs to family I then it is (automatically) minimal. If X belongs to family II then it is not minimal if, and only if, $d = 3, 5, 6$ or $d = 8$ and X is isomorphic to the ruled surface $\mathbb{F}_1$. There are no minimal surfaces with $d = 7$.

Some surfaces endowed with a conic fibration are at the same time del Pezzo surfaces, namely, if $d = 3, 5, 6$ or $d = 1, 2, 4$ and X has two distinct conic fibrations, or $d = 8$ and $\bar{X} = X \times_k \bar{k} \simeq \mathbb{P}^1 \times \mathbb{P}^1$ or $\mathbb{P}^2$ blown up in one point.

Definition 3.2. We say that a surface X is k-birationally trivial or k-rational if there is a birational map $\mathbb{P}^2 \dashrightarrow X$ defined over k.

Theorem 3.3. *Every minimal rational surface such that $d \leq 4$ is k-birationally nontrivial.*

Theorem 3.4. *Every rational surface X of degree at least five such that $X(k) \neq \varnothing$ is k-rational and every rational surface X with $d \geq 3$ and $X(k) \neq \varnothing$ is k-unirational.*

Remark 10. A priori most k-minimal surfaces with $d \geq 1$ and $X(k) \neq \varnothing$ can be a k-minimal model of a rational elliptic surface. The condition $X(k) \neq \varnothing$ comes from the zero section that is defined over k and is contracted to a k-rational point. We will exclude the conic bundles such that $\bar{X} = X \times_k \bar{k}$ is isomorphic to $\mathbb{P}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(n))$ with $n \geq 3$ since these surfaces contain a curve with self intersection $-n$ and a rational elliptic surface contains no curves with self intersection $-n$ for $n \geq 3$.

Remark 11. If a surface X of degree $d = \omega_X^2$ is a k-minimal model of a rational elliptic surface $\mathscr{E}$, then $\mathscr{E}$ is isomorphic to the blow up of X in d points which form a Galois invariant set.

We finish this subsection giving a sufficient condition for $\bar{k}$-rational elliptic surfaces to be k-unirational.

Lemma 3.5. *Let $\mathscr{E}$ be a $\bar{k}$-rational elliptic surface defined over a perfect field k. If* $\operatorname{rank}\operatorname{Pic}(\mathscr{E}|k) \geq 5$ *then $\mathscr{E}$ is k-unirational.*

Proof. Since the Picard group of k-minimal rational surfaces has rank one or two, any k-minimal model of $\mathscr{E}$ will be obtained after contracting at least three (-1)-curves. Hence if X is a k-minimal model of $\mathscr{E}$ then $K_X^2 \geq 3$. Moreover, $X(k) \neq \varnothing$ since the image of the zero section will provide at least one k-rational point. It follows then from Theorem 3.4 that X is k-unirational. □

3B. ***Kummer theory.*** Let K be a number field or a function field and A an abelian variety. Let $P \in A(K)$ be a point of infinite order. We say that P is *indivisible by* n if for all $Q \in A(K)$ such that there exists a divisor d of n with $[d]Q = P$, we have $d = \pm 1$.

Remark 12. In [Hindry 1988], Hindry defines a point as being *indivisible* if it is indivisible by all natural numbers $m > 1$. One can replace this definition by the one introduced above, that is, P indivisible by m in the hypothesis of [Hindry 1988, Lemme 14]. In fact, let P be a point indivisible by m. We can write $P = [l]P_1$ with P_1 indivisible as in [Hindry 1988] and $(l, m) = 1$. There exist $u, v \in \mathbb{N}$ such that $ul + vm = 1$, which allows us to write

$$P_1 = [u]P + [vm]P_1,$$

and thus

$$K\big(\tfrac{1}{m}P\big) \subseteq K\big(\tfrac{1}{m}P_1\big) = K\big(\tfrac{u}{m}P\big) \subseteq K\big(\tfrac{1}{m}P\big).$$

Hence $K(\frac{1}{m}P_1) = K(\frac{1}{m}P)$. In the rest of this subsection we state some of the results that can be found in [Hindry 1988] about the degree of the extension $K(\frac{1}{m}P)$ taking this remark into account, that is, replacing the hypothesis P indivisible, by the hypothesis P indivisible by m.

Let $n \in \mathbb{N}$ and P be a point indivisible by m where m divides n. Denote by A_n the set of n-torsion points in $A(\bar{K})$ and $\frac{1}{m}P$ a point $Q \in A(\bar{k})$ such that $[m]Q = P$. The Galois group $G_{n,K,(1/m)P}$ of the extension

$$K\big(A_n, \tfrac{1}{m}P\big) \mid K(A_n)$$

can be viewed as a subgroup of A_m. Kummer theory for abelian varieties tells us that if K is a number field the group $G_{n,K,\frac{1}{m}P}$ is actually almost the whole group

A_m, that is, its image inside A_m by an injective map is of finite index which is bounded by a constant independent of P and m.

This theory was studied in full generality by Ribet, Bertrand, Bashmakov and others. We restrict ourselves to elliptic curves. For us k is a number field and $K = k(B)$ is the function field of a projective curve B.

We now state a typical result of Kummer theory for elliptic curves. This is valid in a more general context; see [Ribet 1979].

Theorem 3.6. *Let E be an elliptic curve defined over a number field k, $n \in \mathbb{N}$ and $P \in E(k)$ a point indivisible by m where m divides n. There exists a positive constant $f_0 = f_0(E, k)$ such that $|G_{n,k,(1/m)P}| \geq f_0.m$.*

Proof. See [Hindry 1988, Appendix 2] for when P is an indivisible point. □

Let $\pi : \mathscr{E} \to B$ be an elliptic surface and P a point indivisible by n in the generic fiber. Since Galois groups become smaller after specialization, after applying the previous theorem to a fiber $E = \pi^{-1}(t)$ defined over the number field k, we have

$$|G_{n,k(B),(1/m)P}| \geq |G_{n,k,(1/m)P}|,$$

which gives us a theorem as Theorem 3.6 for elliptic curves over function fields.

Theorem 3.7. *Let E be an elliptic curve defined over a function field K, $n \in \mathbb{N}$ and $P \in E(K)$ a point indivisible by m where m divides n. There exists a positive constant $f_0 = f_0(E, K)$ such that $|G_{n,K,(1/m)P}| \geq f_0.m$.*

Remark 13. If the surface $\mathscr{E}$ above is not isotrivial then the endomorphism ring of the generic fiber E is $\mathbb{Z}$ and the result above is stronger, namely, the group $G_{n,K,(1/m)P}$ is almost all the set of m-torsion points $\mathscr{E}_m$, that is, its image inside the m-torsion subgroup under an injective map is of finite index, bounded by a constant independent of P and m; see [Hindry 1988, Proposition 1].

We finish this subsection with a lemma about torsion points on elliptic curves that will be used in the next section during the proof of Proposition 4.2. See [Serre 1972] for a proof and more results on torsion points on elliptic curves.

Lemma 3.8. *Let E be an elliptic curve defined over a field k, which is either a number field or a function field, and $P \in E(\bar{k})[m] \setminus E(k)$ a point of order m. Then there exists an $\alpha > 0$ and a constant $c_{E,k}$ independent of the point P such that*

$$[k(P) : k] \geq c_E.m^{\alpha}.$$

4. Proof of Theorem 1.1

Let $\pi : \mathscr{E} \to B$ be as in Section 2. In the first subsection we show that given a nonconstant pencil of curves in $\mathscr{E}$ that are not contained in a fiber of $\pi : \mathscr{E} \to B$, then all but finitely many curves in it yield, after base change (see Section 2A), a *new section* that is independent of the *old sections*. The proof of Theorem 1.1 then depends only on a construction of a family of irreducible curves defined over k with infinitely many k-rational points, that is, $\mathbb{P}^1$ or genus-one curves with positive Mordell–Weil rank. This construction will be given in the second subsection.

4A. ***A key proposition.*** First we state the most useful criteria for us to determine when a *new section* is independent of the *old* ones.

Lemma 4.1. *An irreducible curve $C \subset \mathscr{E}$ that is not a component of a fiber induces a new section on $\mathscr{E} \times_B C$ independent of the old ones if and only if for every section $C_0 \subset \mathscr{E}$ and every $n \in \mathbb{N}^*$, the curve C is not a component of $[n]^{-1}(C_0)$.*

Let $\mathscr{C}$ be a family of curves in a projective surface X. We call $\mathscr{C}$ a *numerical family* if all its members belong to the same numerical class in the Néron–Severi group. We prove that in a numerical family it is enough to check the conditions of the lemma for a bounded n and a finite number of sections. Thus if the family is infinite, all but finitely many members induce new independent sections after base change.

Proposition 4.2. *Let $\mathscr{E} \to B$ be an elliptic surface defined over a number field k. Let $\mathscr{C}$ be a numerical family of curves inside $\mathscr{E}$. There exist an $n_0(\mathscr{C}) \in \mathbb{N}$ and a finite subset $\Sigma_0(\mathscr{C}) \subset \mathrm{Sec}(\mathscr{E})$ such that for C in the family $\mathscr{C}$, the new section induced by C is linearly dependent of the old ones if, and only if, $[n]C \in \Sigma_0$ for some $n \leq n_0$.*

Proof. Suppose $[n]C = C_0$ for some section C_0. We may assume such n to be minimal, that is, there does not exist $n' < n$ such that the curve $[n']C$ is a section. The proof is divided in two parts: bounding n from above using Kummer theory (see Section 3B), and then, for a fixed n such that $C_0 = [n]C$, showing that the set of sections in the same numerical class is finite by Néron–Tate height theory.

1) *Bounding n:* We define the degree of a curve in $\mathscr{E}$ by its intersection with a fiber:

$$\deg(C) = (C.F).$$

If C_0 is a section we have $\deg(C_0) = (C_0.F) = 1$. The degree of C, which will be denoted by h, is fixed within the family, since all curves belong to the same numerical class.

The map $[n]$ is not a morphism defined on the whole surface, but on an open set $U \subseteq \mathscr{E}$ which excludes the singular points in the fibers. Since sections do not intersect the fibers in singular points, they are contained in U. This allows us to write

$$\deg([n]^{-1}C_0) = (([n]^{-1}C_0).F) = n^2(C_0.F) = n^2.$$

Thus, $\lim_{n\to\infty} \deg[n]^{-1}(C_0) = \infty$.

Denote by K the field $k(B)$, by E the generic fiber of $\mathscr{E}$ and by P_0 the point in $E(K)$ corresponding to the section C_0. Let $P \in E(\bar{K})$ be such that $[n]P = P_0$

where n is minimal with respect to the expression above. We now show that n is bounded by a constant n_0 that depends only on E, K and h.

Note first that if P is a torsion point, we know by Lemma 3.8 that its order m is bounded by $c_E.m^\alpha \leq [K(P):K] = h$.

Now suppose P is of infinite order. Let m be the smallest positive integer such that there exist $P_1 \in E(K)$ and T a torsion point satisfying $[m]P = P_1 + T$. We claim that P_1 is indivisible by m. If l is a divisor of m such that $[l]Q = P_1$ with $Q \in E(K)$, then

$$[l][m/l]P = [l]Q + T$$

and hence there exists a torsion point T_1 such that $[m/l]P = Q + T_1$. By the minimality of m with respect to the equation above, we must have $l = \pm 1$.

Let $m' = mm_1$ where m_1 is the order of the torsion point T. Let T' be such that $[m]T' = T$ and put $P' = P + T'$; then $[m]P' = P_1$ and $T' \in E_{m'}$. Applying Theorem 3.7 we obtain

$$[K(P', E_{m'}) : K(E_{m'})] \geq m.f_1. \tag{1}$$

Now, note that $K(P, E_{m'}) = K(P', E_{m'})$ hence

$$h = [K(P):K] \geq [K(P, E_{m'}) : K(E_{m'})] = [K(P', E_{m'}) : K(E_{m'})] \geq f_1.m,$$

and thus m is bounded. Since T is defined over $K(P)$, its order m_1 is also bounded in terms of h, and thus $[mm_1]P = m_1 P_1 \in E(K)$ with $n \leq mm_1$ bounded as stated.

2) *The numerical class of a section:* Fix $C_1, \dots, C_r$ generators of the Mordell–Weil group.

For a fixed n, the intersection multiplicity $(([n]C).C_i)$ is also fixed, say equal to n_i, and it depends only on the numerical class of C. The same holds for the intersection of C with the zero section, say equal to m_0 and for the intersection with the fiber components Θ_v, say equal to l_{v,j_v}. The Néron–Tate height in an elliptic surface is uniquely determined by the intersection numbers above. The set

$$\Sigma_0 = \{ \text{ sections } C_0 \mid (C_0.C_i) = n_i, i = 1, \dots, r, (C_0.O) = m_0, (C_0.\Theta_{v,j_v}) = l_{v,j_v} \}$$

is finite since it is a set of points with bounded Néron–Tate height, thus there are only finitely many possible sections C_0 such that $C \subset [n]^{-1}C_0$. $\square$

Corollary 4.3. *Let $\mathscr{E}$ be an elliptic surface and $\mathscr{L}$ a nonconstant numerical family of curves on $\mathscr{E}$ whose members are not contained in the fibers of $\mathscr{E}$. Then for almost all member C of the pencil, the new section induced by C is independent of the old sections.*

Proof of Theorem 1.1. Let $\psi : \mathbb{P}^2 \dashrightarrow \mathscr{E}$ be a k-unirational map. Let $\mathscr{L}$ be given by the set of lines in $\mathbb{P}^2$. Then $\mathscr{L}^\psi = \{\psi(L) \mid L \in \mathscr{L}\}$ is an infinite family of curves in

$\mathscr{E}$ defined over k, whose general member is integral and of geometric genus zero. These curves cannot be all contained in a fiber of $\mathscr{E}$ as the family is infinite, with irreducible members, and there is only a finite number of reducible fibers for the elliptic fibration in $\mathscr{E}$.

The theorem follows from an application of Corollary 4.3 to the family $\mathscr{L}^{\psi}$. $\square$

5. Proof of Theorems 1.3 and 1.4

To prove Theorems 1.3 and 1.4 we need to produce two families of curves defined over k. These families must not only have infinitely many k-rational points, but also be such that the fibered product of two curves in different families is irreducible. The first subsection is devoted to the construction of such families, first done in the context of Theorem 1.3, that is, over $\mathbb{P}^2$, then in the settings of Theorem 1.4, that is, over other k-minimal surfaces. The proof of irreducibility of the generic member of the constructed families is given in the second subsection and is followed by the verification that such curves contain indeed infinitely many k-rational points. Finally, all is assembled in Section 5D to conclude the proofs of Theorems 1.3 and 1.4.

5A. ***Construction of linear pencils of rational curves.*** The results presented in this subsection are technical and may be skipped by the reader willing to accept the existence of two linear pencils of conics, that is, copies of $\mathbb{P}^1$ intersecting the fibers with multiplicity two in the surfaces satisfying the hypothesis of Theorem 1.3. Here we provide "case-by-case", depending on the configuration of blown up points, constructions of linear pencils of curves on $\mathscr{E}$ to which we apply Corollary 4.3. We construct two linear pencils of rational curves defined over k in a k-minimal model of the rational elliptic surface $\mathscr{E} \to B$. Since the base-changed surface fibers over the fibered product of these two curves over B, we fabricate those curves in a way that their fibered product has genus at most one. We state below sufficient conditions for this.

Lemma 5.1. *Let C_1 and C_2 be two smooth projective rational curves given with two distinct morphisms $\varphi_i : C_i \to B$ of degree 2 to a genus zero curve B. Then $g(C_1 \times_B C_2) \leq 1$.*

Proof. It is a simple application of the Hurwitz formula. $\square$

We can proceed to the constructions. They depend on the degree of the k-minimal model considered, as well as on the configuration of the blown up points under the action of the absolute Galois group $\mathrm{Gal}(\bar{k}|k)$. We start with the simplest case, namely, when $\mathscr{E}$ has a minimal model isomorphic to $\mathbb{P}^2$.

a) *A minimal model k-isomorphic to* $\mathbb{P}^2$. Let $p_1, \dots, p_9$ be the nine not necessarily distinct points in the blow up locus. Since the zero section is defined over k, at least one of the points above is k-rational, say p_1. If C is a curve of degree d in $\mathbb{P}^2$ with multiplicity m_i through p_i then its genus satisfies

$$g(C) \leq \frac{(d-1)(d-2)}{2} - \sum_i \frac{m_i(m_i-1)}{2}.$$

Denote by C' its strict transform under the blow up $\mathscr{E} \to \mathbb{P}^2$. Then the degree of the map given by the restriction of π to C' is given by

$$\deg(C' \to B) = (F.C') = 3d - \sum_i m_i$$

where F is a fiber of the elliptic fibration π.

Let $\mathscr{L}_1$ be the pencil of lines in $\mathbb{P}^2$ through p_1. Let $\mathscr{L}_1'$ be the pencil of curves in $\mathscr{E}$ given by the strict transforms of the curves in $\mathscr{L}_1$ and C' a curve in $\mathscr{L}_1'$ — from now on the superscript $'$ will denote the pencil or curve in the elliptic surface given by the strict transform of that in the minimal model. Then

$$\deg(C' \to B) = 3.1 - 1 = 2.$$

We now construct a second pencil of rational curves $\mathscr{L}_2$ such that the curves in the pencil of strict transforms induced in $\mathscr{E}$ satisfy Lemma 5.1.

Since the blow up is defined over k, the set formed by the other points is invariant under the action of $\mathrm{Gal}(\bar{k}|k)$. The construction depends on the size of the smallest orbit different from p_1 whose points are, together with p_1, in general position, that is, no three are collinear, no six lie in a conic and there is no cubic through eight of the points singular at one of them.

i) One other k-rational point p_2.

In this case we can construct $\mathscr{L}_2$ in a similar way as we did for $\mathscr{L}_1$: take

$$\mathscr{L}_2 = \{l \text{ a line in } \mathbb{P}^2 \text{ through } p_2\}.$$

Any curve in $\mathscr{L}_2'$ together with any curve in $\mathscr{L}_1'$ satisfies Lemma 5.1.

ii) Two conjugate (under $\mathrm{Gal}(\bar{k}|k)$) points p_2, p_3.

Let Λ be a pencil of cubics in $\mathbb{P}^2$ inducing $\mathscr{E}$ such that p_1, p_2, p_3 are base points of it. Since we suppose that points are in general position, p_1, p_2 and p_3 are not collinear. Let us first suppose that there are no other base points and thus that the multiplicities (m_1, m_2, m_3) of p_1, p_2, p_3 as base points of Λ are $(1, 4, 4)$, $(3, 3, 3)$, $(5, 2, 2)$ or $(7, 1, 1)$. In the first case, every cubic in Λ shares the same tangent line, say l_2, through p_2 as well as the same tangent line, say l_3, through p_3. We consider $\mathscr{L}_2$, the set of conics through p_2, p_3 with tangents l_i through p_i, for $i = 2, 3$. Let C be a conic in $\mathscr{L}_2$. Then C intersects the cubics of Λ in p_2 and p_3 with intersection multiplicity two and intersects in two other points. Hence, the morphism from the strict transform $\varphi_{C'} : C' \to B$, given by the restriction of the fibration to C', has degree two. In the remaining three cases all cubics of Λ share a tangent line, say l_1, through p_1. We take $\mathscr{L}_2$ to be the set of conics through p_1, p_2, p_3 with prescribed tangent l_1 through p_1. It is a linear pencil of conics, since the space of conics in $\mathbb{P}^2$ has dimension 5. A conic C in $\mathscr{L}_2$ also intersects the cubics in Λ at the point p_1 twice, at the points p_2, p_3 and at two other points. Thus if C is a conic in $\mathscr{L}_2$ then the morphism $\varphi_{C'}$ from C' to B also has degree two. As in i), any curve in $\mathscr{L}_2'$ together with any curve in $\mathscr{L}_1'$ satisfies the hypothesis of Lemma 5.1.

Now suppose there are other base points. If there is another orbit with two or four points one may apply construction iv) below independently of the configuration of these extra points. If there are three, five or six other conjugate base points then apply constructions iii), v), or vi), respectively, below.

iii) Three conjugate points p_2, p_3, p_4.

The construction here is simpler. The pencil of conics $\mathscr{L}_2$ through p_1, p_2, p_3 and p_4 is such that for every curve $C \in \mathscr{L}_2$ the morphism φ_C' has degree two.

iv) Four conjugate points p_2, p_3, p_4, p_5.

As in the previous case, but now we let $\mathscr{L}_2$ be the pencil of conics through p_2, p_3, p_4 and p_5.

v) Five conjugate points p_2, p_3, p_4, p_5, p_6.

Take $\mathscr{L}_2$ to be the pencil of cubics through $p_1, \ldots, p_6$ with a singularity at p_1. The degree of the morphism φ_C for $C \in \mathscr{L}_2$ is two $(9-5-2=2)$. So that together with a curve in $\mathscr{L}_1'$ the curve C satisfies the hypothesis of Lemma 5.1.

vi) Six conjugate points $p_2, \ldots, p_7$.

Consider $\mathscr{L}_2$ to be the pencil of quintics singular at each $p_2, \ldots, p_7$ that passes through p_1 as well. Since $\dim(H^0(\mathbb{P}^2, \mathcal{O}(5))) = 21$ and we have imposed 19 conditions, $\mathscr{L}_2$ forms at least a linear pencil. The curves in it are rational since $g \leq 6-6=0$.

The degree of φ_C' for $C \in \mathscr{L}_2$ is equal to $5.3-6.2-1=2$.

vii) Seven conjugate points $p_2, \ldots, p_8$.

Consider $\mathscr{L}_2$ the space of quartics through the seven points $p_2, \ldots, p_8$ with multiplicity at least three at p_1. That gives at most 13 conditions in a space of dimension 15. So $\mathscr{L}_2$ is at least a linear pencil. Its curves are rational and the degree of the induced morphism to B is equal to $4.3-7-3=2$.

viii) Eight conjugate points $p_2, \ldots, p_9$.

We consider highly singular curves. Take $\mathscr{L}_2$ to be the set of curves of degree 17 such that:

- The eight points $p_2, \ldots, p_9$ are singular with multiplicity at least six.

– It passes through p_1.

This gives us at most 169 conditions in a 171-dimensional space and thus at least a linear pencil. The degree of the morphism is also two and the curves have genus zero as in the previous cases.

Irreducibility of the curves constructed above: Since the points considered are in general position the curves constructed above are irreducible. This is trivially verified in cases i)–iv). In cases v)–viii) one can easily check that if the pencil is generically reducible then either its base points are in nongeneral position, that is, there are three collinear points, six lie on a conic or eight on a cubic singular at one of them, or the Galois orbits break into smaller orbits.

We now focus on the other possible minimal models over k. As observed after the statement of Theorem 1.4, we may suppose X is a del Pezzo surface. Let us first recall some geometric and arithmetic facts about those surfaces.

Since we are dealing with surfaces whose Picard group over k is small (isomorphic to $\mathbb{Z}.\omega_X$), the most natural place to look for curves defined over k is $H^0(X, \omega_X^{-n})$. We now recall the dimension of these spaces and the genus of the curves in them.

Lemma 5.2. *Let X be a k-minimal del Pezzo surface of degree $1 \leq (\omega_X.\omega_X) = d \leq 8$ defined over a number field k. Given points $p_1, \dots, p_j$ in $X(\bar{k})$ and nonnegative integers $n_1, \dots, n_j$, let $\mathscr{L} = \{s \in H^0(X, \omega_X^{-n}) \mid m_{p_i} \geq n_i\}$ where m_{p_i} denotes the multiplicity at the point p_i of the curve given by the divisor of zeros of the section s. The following hold*:

i) $\dim(\mathscr{L}) \geq \dfrac{d(n^2+n)}{2} + 1 - \sum_i \dfrac{n_i^2+n_i}{2}$.

ii) *If $L \in \mathscr{L}$ then* $g(L) \leq \dfrac{d(n^2-n)}{2} + 1 - \sum \dfrac{n_i^2-n_i}{2}$.

iii) *If $\pi : \mathscr{E} \to B$ is an elliptic surface obtained by blowing up $p_1, \dots, p_j$ and L' is the strict transform of L in $\mathscr{E}$ then* $\deg(\pi|_{L'} : L' \to B) = nd - \sum m_{p_i}$.

Proof. See [Kollár 1996, Chapter III Lemma 3.2.2]. □

We can now proceed to the construction of linear pencils on the k-minimal models. We recall that at least one point in the blow up locus of $f : \mathscr{E} \to X$ is k-rational, the one that comes from the contraction of the zero section. We suppose that $\mathscr{E}$ has no reducible fibers. It follows that all the points on the locus of f are distinct since the blow up of infinitely near points gives rise to (-2)-curves, and these are always components of reducible fibers. For (i) of Theorem 1.4 it is clearly sufficient to do the construction in the case where there are two orbits by the action of the Galois group: the one of the k-rational point and another one with the other $d-1$ points. As in the case where $\mathbb{P}^2$ was a k-minimal model, we will look for curves satisfying the hypothesis of Lemma 5.1.

b) *A minimal model isomorphic to a del Pezzo surface of degree eight:* Let $p_2, \dots, p_9$ be the points on the blow up locus of $f : \mathscr{E} \to X$. Let p_2 be the k-rational point. We consider the following pencils of curves in X:

$$\mathscr{L}_1 = \{s \in H^0(X, \omega_X^{-8}) \ \mid m_{p_2}(s) \geq 13,\ m_{p_i}(s) \geq 7 \text{ for } i = 3, \dots, 9\},$$
$$\mathscr{L}_2 = \{s \in H^0(X, \omega^{-22}) \mid m_{p_2}(s) \geq 13,\ m_{p_i}(s) \geq 23 \text{ for } i = 3, \dots, 9\}.$$

If $C_1' \in \mathscr{L}_1'$ then by Lemma 5.2

$$g(C_1') \leq 8\frac{(64-8)}{2} + 1 - \frac{(169-13)}{2} - 7\frac{(49-7)}{2} = 0$$

and $\deg(\varphi_{C_1}) = 64 - 13 - 49 = 2$. If $C_2' \in \mathscr{L}_2'$ then

$$g(C_2') \leq 8\frac{(22^2-22)}{2} + 1 - \frac{(169-13)}{2} - 7\frac{(23^2-23)}{2} = 0$$

and $\deg(\varphi_{C_2}) = 22.8 - 13 - 7.23 = 2$.

Thus C_1' and C_2' satisfy the hypothesis of Lemma 5.1. Since, by Lemma 5.2 (i), they belong to a linear pencil of curves, by Corollary 4.3 they can be chosen in a way such that the new sections induced by them in the base-changed surface are independent of the *old sections* and of each other.

Since there are no minimal rational surfaces of degree seven we now pass to surfaces of degree six.

c) *A minimal model isomorphic to a del Pezzo surface of degree six.* Let $p_4, \dots, p_9$ be the points on the blow up locus of f. We consider the possible orbits under the action of the absolute Galois group. We denote the cases by $(n_1, \dots, n_r)$ where r is the number of distinct orbits and n_i is the multiplicity of the points in the same orbit.

i) If the points lie in a $(1, 2, 3)$-configuration, then the blow up of the two points in the same orbit produces a surface of degree four to which we apply the constructions in e).

ii) If $(1, 1, n_3, n_4)$, let p_4 and p_5 be the k-rational points. The blow up of p_4 produces a surface of degree five to which we can apply the constructions in d).

d) *A minimal model isomorphic to a del Pezzo surface of degree five:* Here we consider the pencils

$$\mathscr{L}_1 = \{s \in H^0(X, \omega_X^{-2}) \ \mid m_{p_5}(s) \geq 4,\ m_{p_i}(s) \geq 1 \text{ for } i = 6, \dots, 9\}$$
$$\mathscr{L}_2 = \{s \in H^0(X, \omega_X^{-10}) \mid m_{p_5}(s) \geq 4,\ m_{p_i}(s) \geq 11 \text{ for } i = 6, \dots, 9\}.$$

We have

$$\dim(\mathscr{L}_1) \geq 16 - 10 - 4 = 2, \quad g(C_1) = 6 - 6 = 0,$$
$$\deg(f : C_1' \to D) = 10 - 4 - 4 = 2, \quad \dim(\mathscr{L}_2) \geq 5(110)/2 + 1 - 10 - 4(66) = 2.$$

e) *A minimal model isomorphic to a del Pezzo surface of degree four:* The pencils that we consider to prove Theorem 1.3 are

$$\mathcal{L}_1 = \{s \in H^0(X, \omega_X^{-1}) \mid m_{p_1}(s) \geq 2\},$$
$$\mathcal{L}_2 = \{s \in H^0(X, \omega_X^{-7}) \mid m_{p_1}(s) \geq 2, m_{p_i}(s) \geq 8, i = 2, 3, 4\}.$$

5B. *Irreducibility of the curves.* We prove that if $\mathscr{E}$ satisfies the hypothesis of Theorem 1.4 the curves constructed in the previous subsection are irreducible. First, we show that a series of Cremona transformations, that is, the blow up of three distinct and noncollinear points followed by the contraction of the three lines through them, reduces each pencil of curves produced above to a pencil of lines in $\mathbb{P}^2$ passing through a point. Cremona transformations are not in general automorphisms of the surface, but they are automorphisms of the Picard group. In particular a class is represented by a connected curve if and only if the transformed one under a Cremona transformation is represented by an connected curve. We then use the fact that $\mathscr{E}$ has no reducible fibers to show that the curves constructed are irreducible. (See [Testa 2009] for more on irreducibility of spaces of curves on del Pezzo surfaces.)

Lemma 5.3. *Let $\mathscr{E}$ be a rational elliptic surface with no reducible fibers, X a k-minimal model of $\mathscr{E}$ as in Theorem 1.4 and let $\mathscr{L}$ be one of the pencils of curves constructed in the previous subsection. Then the generic member of $\mathscr{L}$ is an irreducible curve with geometric genus zero.*

Proof. Let d be the degree of X. Let $f : \mathscr{E} \to X$ be the map corresponding to the blow up of a $\mathrm{Gal}(\bar{k}|k)$-invariant set of distinct points

$$P_1, \ldots, P_d \in X(\bar{k}).$$

Since $\mathscr{E}$ has no reducible fibers, the exceptional curves above P_i, for $i = 1, \ldots, d$, are all independent in the Mordell–Weil group of $\mathscr{E}$. Therefore, they provide a subset of a set of generators of the Picard group of $\mathscr{E}$. We can fix a basis for the geometric Picard group of $\mathscr{E}$ to be $\{L_0, \ldots, L_9\}$ where L_0 is the total transform of a line l in $\mathbb{P}^2$, $L_1, \ldots, L_d$ are the exceptional curves above $P_1, \ldots, P_d$ and $L_{d+1}, \ldots, L_9$ are also exceptional curves in $\mathscr{E} \times_k \bar{k}$.

Let $g : \mathscr{E} \to \mathbb{P}^2$ be a blow up presentation, defined over $\bar{k}$, factoring through f. We represent a curve C in $\mathscr{E}$ by its *numerical type*, that is, by the list of coordinates of its divisor class in the basis given by $\{L_0, \ldots, L_9\}$ of the Picard group:

$$(d, m_1, \ldots, m_9),$$

where d is the degree of the image of C in $\mathbb{P}^2$ with respect to $g : \mathscr{E} \to \mathbb{P}^2$, $m_i = m_{P_i}(C)$, the multiplicity of the curve C at the point P_i, $i = 1, \ldots, 9$.

Let C be a curve in X. Then the strict transform of C through f is a curve in $\mathscr{E}$ given by

$$f^{-1}(C) - \sum_{i=1,\dots,d} m_{p_i}(C) L_i.$$

We define the *numerical type* of C as the numerical type of its strict transform in $\mathscr{E}$.

For example if $X = \mathbb{P}^2$, lines in $\mathbb{P}^2$ through the point p_1 are represented by the class $(1, 1, 0, \dots, 0)$. If X is a del Pezzo surface of degree five, curves in $\mathscr{L}_1'$, where

$$\mathscr{L}_1 = \{s \in H^0(X, \omega_X^{-2}) \mid m_{P_5}(s) \geq 4, m_{P_i}(s) \geq 1, i = 6, \dots, 9\},$$

are represented by $(6, 2, 2, 2, 2, 4, 1, 1, 1, 1)$.

Since Cremona transformations are given by composing birational maps, to show that a curve is connected, it is sufficient to verify that curves constructed in the previous subsection can be (Cremona-)transformed into a curve whose numerical type is one of $(1, 1, 0, \dots, 0), \dots, (1, 0, \dots, 0, 1)$, that is, into a line through one of the p_i.

After applying Cremona transformations successively, one can check that all curves constructed in the previous subsection are in the same class as a line through a point and are thus connected curves.

We give below the result of Cremona transformations applied to curves in $\mathscr{L}_1'$ where $\mathscr{L}_1 = \{s \in H^0(X, \omega_X^{-2}) \mid m_{p_5}(s) \geq 4, m_{p_i}(s) \geq 1, i = 6, \dots, 9\}$ and X is a del Pezzo surface of degree five.

Curves in this family are encoded by $(6, 2, 2, 2, 2, 4, 1, 1, 1, 1)$. Applying Cremona transformations successively, these become $(4, 0, 0, 2, 2, 2, 1, 1, 1, 1)$, then $(2, 0, 0, 0, 0, 0, 1, 1, 1, 1)$, and finally, $(1, 0, 0, 0, 0, 0, 0, 0, 0, 1)$.

Now that we have verified that the curves constructed in the previous subsection are connected, we still have to show that the only connected component is irreducible.

Let C be a curve in one of the families constructed previously. Then $C.F = 2$ and C has geometric genus zero. Thus if C is a reducible curve then it satisfies one of the following:

i) $C = C_1 \cup C_2 \cup (F_1 + \cdots + F_m)$ where C_1 and C_2 are sections and F_i are components of reducible fibers.

ii) $C = D \cup (G_1 + \cdots + G_m)$ where D is an irreducible genus zero curve such that $D.F = 2$ and G_j are components of reducible fibers.

By Proposition 4.2, case i) can only occur for finitely many curves in a numerical family. Thus, we may suppose that the generic members of the families constructed satisfy case ii). As there are only finitely many reducible components of the fibers, the curve D in case ii) is such that $\dim |D| \geq 1$. Since $\mathscr{E}$ has no reducible fibers, both pencils $\mathscr{L}_1$ and $\mathscr{L}_2$ constructed have irreducible generic members. $\square$

5C. ***Infinitely many rational points on the new base.*** In order to prove the corollaries we must show that the base curve of the new elliptic surface (the base-changed one) has infinitely many k-rational points. To prove Corollary 1.2 one needs only one base change, thus one has to prove that infinitely many among the rational curves constructed in the previous section have a k-rational point. This is assured since the surface X where the pencil is constructed is k-unirational [Manin and Tsfasman 1986, Theorem 3.5.1] so, in particular, has a Zariski dense set of k-rational points.

Lemma 5.4. *Suppose $\mathscr{E}(k)$ is Zariski dense in $\mathscr{E}$. Let $\{D_t\}_{t\in\mathbb{P}^1}$ be a nonconstant pencil of genus zero curves defined over k in $\mathscr{E}$. Then infinitely many of its curves are k-rational.*

For Corollary 1.5 the base curve is in general an elliptic curve. Since our constructions give us families of possible new bases, we look at these families as elliptic surfaces and we show that these elliptic surfaces have a nontorsion section.

Theorem 5.5. *Let $\pi : \mathscr{E} \to B$ be a rational elliptic surface defined over a number field k. Let $\mathscr{L}_1 = \{C_t \mid t \in \mathbb{P}^1\}$ and $\mathscr{L}_2 = \{D_u \mid u \in \mathbb{P}^1\}$ be two base point free linear systems of k-rational curves in $\mathscr{E}$ defined over k with $C_t(k)$, $D_u(k) \neq \varnothing$ for infinitely many t, u, such that the morphism given by the restriction of the elliptic fibration to it has degree two. Then for infinitely many $t \in \mathbb{P}^1(k)$ and infinitely many $u \in \mathbb{P}^1(k)$ we have $\#(C_t \times_B D_u)(k) = \infty$.*

Proof. Let $C \in \mathscr{L}_1$ be such that the base-changed elliptic surface $\mathscr{E}_C = \mathscr{E} \times_B C$ has generic rank strictly larger than the generic rank of $\mathscr{E}$. The surface $\mathscr{E}_C$ admits a second elliptic fibration (since $\mathscr{L}_2$ is base point free it is not necessary to blow up points to have the fibration), namely, $\mathscr{E}_C \to \mathbb{P}^1(u)$ where $\mathbb{P}^1(u)$ is the index-set of the pencil of curves $\mathscr{L}_2$. The fibers of the latter are exactly the curves $D_u \times_B C$. We will show that infinitely many among them have positive rank by showing that this fibration is covered by an elliptic fibration of positive rank.

The natural morphism $C \to \mathbb{P}^1(u)$ gives us the surface $\mathscr{E}_C \times_{\mathbb{P}^1(u)} C \to C$. Fix $(\mathrm{id}, \mathrm{id}, \mathrm{id}) : C \to \mathscr{E}_C \times_{\mathbb{P}^1(u)} C$ as the zero section. The involution ι on C with respect to the double cover $\varphi_C : C \to B$ gives us another section for the fibration $\mathscr{E}_C \times_{\mathbb{P}^1(u)} C \to C$, namely, $(\iota, \mathrm{id}, \mathrm{id})$. It intersects the zero section on the points corresponding to the ramification points a and b of the morphism $C \to B$. The intersection (Q_i, Q_i, Q_i) where $Q_i = \varphi_C^{-1}(t_i)$, is a singular point on the fiber where it is located if and only if t_i is also a ramification point for $\varphi_{D_{t_i}} : D_{t_i} \to B$, see [Grothendieck 1960, Corollaire 3.2.7, pp.108]. We have two possibilities:

(i) The intersection point is not a ramification point for $\varphi_{D_{t_i}} : D_{t_i} \to B$.

Since torsion sections do not meet at a nonsingular point (see for example [Miranda and Persson 1989, Lemma 1.1]) the section given by $(\iota, \mathrm{id}, \mathrm{id})$ has infinite order.

The fibration $\mathscr{E}_C \times_{\mathbb{P}^1(u)} C \to C$ has positive generic rank and thus, by Néron–Silverman's specialization theorem, infinitely many fibers with positive rank. Hence the fibration $\mathscr{E}_C \to \mathbb{P}^1(u)$ has also infinitely many fibers with positive rank.

(ii) The intersection point is a ramification point for $\varphi_{D_{t_i}} : D_{t_i} \to B$.

In this case the curve $C \times_B D_{t_i}$ is rational. We must consider two cases:

(iia) The curve D_{t_i} induces an independent new section in $\mathscr{E} \times_B D_{t_i}$ and we are done.

(iib) The generic rank of $\mathscr{E} \times_B D_{t_i}$ is equal to the generic rank of $\mathscr{E}$.

Choose another curve $C \in \mathscr{L}_1$ in the beginning of the proof. Since only finitely many D_u do not contribute with an extra section after base change by Corollary 4.3, if (iib) holds for almost all curves $C_t \in \mathscr{L}_1$ then the set

$$R = \bigcup_{t \in \mathbb{P}^1} \{b \in B(k) \mid \phi_t : C_t \to B \text{ is ramified above } b\}$$

is finite. By Lemma 5.6 below we conclude n that all morphisms $\phi_t : C_t \to B$ ramify above the same points. In this case we start over by fixing a curve $D \in \mathscr{L}_2$ such that the surface $\mathscr{E}_D$ has generic rank strictly larger than the generic rank of $\mathscr{E}$. The curves $C_t \times_B D$ where C_t varies in $\mathscr{L}_1$ induce an elliptic fibration on $\mathscr{E}_D$. Since infinitely many of them contribute with a new independent section, we will be either in case (i) or case (iia). □

Lemma 5.6. *Let $\pi : X \to B$ be a fibration on curves from a smooth proper surface X defined over a number field k to a smooth proper curve B. Let $f : X \to \mathbb{P}^1$ be a genus zero fibration on X such that the fibers of f are not fibers of π. Let π_t be the restriction of π to the fiber $f^{-1}(t)$, for $t \in \mathbb{P}^1(k)$. Let*

$$R = \bigcup_{t \in \mathbb{P}^1} \{b \in B(k) \mid \pi_t \textit{ is ramified above } b\}.$$

Then R is either infinite or equal to $\{b \in B(k) \mid \pi_{t_0} \textit{ is ramified above } b\}$ for any $t_0 \in B(k)$.

5D. *Proof of Theorems 1.3 and 1.4.* Let $\mathscr{E}$ be as in the hypothesis of Theorem 1.3 or Theorem 1.4. Let $\mathscr{L}_1' = \{C_t\}_{\{t \in \mathbb{P}^1\}}$ and $\mathscr{L}_2' = \{D_u\}_{\{u \in \mathbb{P}^1\}}$ be the two pencils of rational curves constructed in the previous subsections according to the possible minimal models of $\mathscr{E}$. By Corollary 4.3 all but finitely many curves $C_t \in \mathscr{L}_1$ induce a new section in $\mathscr{E}_{C_t}$ independent of the old sections. For each $t \in \mathbb{P}^1_k$ the pencil $\mathscr{L}_{2,t}' = \{D_u \times_B C_t\}$ of curves in $\mathscr{E}_{C_t}$ also satisfies Corollary 4.3 and thus for all but finitely many $u \in \mathbb{P}^1_k$ the curve $D_u \times_B C_t$ induces a new section in $\mathscr{E}_{D_u \times_B C_t}$ independent of the old sections coming from $\mathscr{E}_{C_t}$. Thus, after excluding finitely many $t \in \mathbb{P}^1$ and finitely many $u \in \mathbb{P}^1$, the surface $\mathscr{E}_{D_u \times_B C_t}$ satisfies

$$\mathrm{rk}(\mathscr{E}_{D_u \times_B C_t}(k(D_u \times_B C_t))) \geq \mathrm{rk}(\mathscr{E}_{C_t}(k(C_t))) + 1 \geq \mathrm{rk}(\mathscr{E}(k(B))) + 2.$$

Moreover, since we excluded only finitely many of each t and u, by Theorem 5.5 we may choose the curve $D_u \times_B C_t$ such that it has infinitely many k-rational points.

Remark 14. The proof of Theorem 1.4 breaks down if the k-minimal model considered, X, is a surface of degree 6 such that the blow up locus of f contains a Galois orbit with 5 points, or if it has degree 3 or 2. Although we are still able to construct a (family of) rational curve(s) defined over the ground field, the generic member of such a family has two connected components, which cannot be used to finish the proof of the theorem. Moreover, X has no families of irreducible conics (that is, rational curves intersecting the anticanonical divisor of the rational elliptic surface with multiplicity two) defined over the ground field.

6. Corollaries from analytic number theory

To an elliptic curve E over a number field k, we can associate a sign $W(E|k)$ intrinsically via the product of local signs $W_v(E|k)$ (for a complete definition of the local sign, see for example [Rohrlich 1993]).

The parity conjecture may be stated in the following form:

Conjecture 6.1. Let E be an elliptic curve over a number field k. Let r be the rank of its Mordell Weil group. Then $W(E|k) = (-1)^r$.

Remark 15. The previous conjecture is a weak version of the Birch–Swinnerton-Dyer conjecture.

Over the field $\mathbb{Q}$ we know from the work of Wiles that E is modular and that $W(E|k)$ is the sign of the functional equation of the L-function $L(E, s)$.

Let $\pi : \mathscr{E} \to B$ be an elliptic surface over k and $U \subseteq B$ an affine open subset over which $\mathscr{E}$ is an abelian scheme. Note

$$U_{\pm}(k) = \{t \in U(k) \mid W(\mathscr{E}_t) = \pm 1\}.$$

Modulo the parity conjecture, we also have $U_+(k) = \{t \in U(k) \mid \operatorname{rank} \mathscr{E}_t(k) \text{ is even}\}$ and $U_-(k) = \{t \in U(k) \mid \operatorname{rank} \mathscr{E}_t(k) \text{ is odd}\}$.

There are examples for which $W(E_t)$ is constant, but they all correspond to isotrivial surfaces; see for example [Cassels and Schinzel 1982]. In the nonisotrivial case and for $B \simeq \mathbb{P}^1$, H. Helfgott [2003] has shown, under classical conjectures, that the sets $U_{\pm}(k)$ are infinite. This is established unconditionally in some interesting cases by Helfgott [2003; 2004] and Manduchi [1995]. Much less has been done in the case B is a genus one curve such that $B(k)$ is infinite. But, from previous work cited above, it seems reasonable to conjecture the following:

Conjecture 6.2. Let $\mathscr{E} \to B$ be a nonisotrivial elliptic surface defined over a number field k such that $g(B) = 1$ and $B(k)$ is infinite. Then $U_+(k)$ and $U_-(k)$ are infinite.

This allows us to state the following better but conditional result.

Theorem 6.3 (modulo Conjectures 6.2 and 6.1). *Let $\mathscr{E} \to B$ be a rational elliptic surface satisfying the hypothesis of Theorem 1.3. Then*

$$\#\{t \in B(k) \mid r_t \geq r+3\} = \infty,$$

where r is the generic rank and r_t the rank of the fiber $\pi^{-1}(t)$.

Appendix

In this appendix we deal with rational elliptic surfaces with one nonreduced fiber.

If the blown up points are in nongeneral position, then the constructions given in Section 5A may yield reducible curves. Nevertheless, we are still able to deal with some of these cases since some of the special Galois invariant configurations of the base points of a pencil of cubic curves in the plane yield elliptic surfaces with fiber types that are easier to treat, namely, nonreduced fibers.

If the surface has a unique nonreduced fiber, then, depending on the structure of a pencil inducing $\mathscr{E}$, we will be able to prove the rank jumps for infinitely many fibers by first base-changing by a curve in $\mathscr{L}_1'$ where $\mathscr{L}_1$ is the pencil of lines through p_1 constructed in i) of Section 5A. The proposition below tells us that the resulting base-changed elliptic surface is still rational and satisfies the hypothesis of Theorem 1.1, that is, it is a k-unirational elliptic surface.

Proposition 6.4. *Let $\mathscr{E} \to B$ be a rational elliptic surface defined over a number field k such that there is a k-birational morphism $\mathscr{E} \to \mathbb{P}^2$ contracting the zero section to a point $p_1 \in \mathbb{P}^2(k)$. Suppose $\mathscr{E}$ has a unique fiber of type $*$, induced by a cubic curve of the form $3m$ or $m \cup 2l$ where m is a line through p_1 and l is another line.*

Then for all but finitely many $L_1 \in \mathscr{L}_1$ the morphism $L_1' \to B$ is ramified over the place corresponding to the fiber of type $*$. Moreover, the surface $\mathscr{E} \times_B L_1'$ is k-unirational for all but finitely many $L_1' \in \mathscr{L}_1'$.

Proof. Let F be the nonreduced fiber of $\mathscr{E}$ given in the hypothesis.

Suppose first that F is induced by the triple line $3m$ where m is a line through p_1. Note that p_1 is a base point with multiplicity at least 3 and thus f factors through the blow up of p_1 and two infinitely near points to it. The first blow up of p_1 transforms $3m$ into $3m' + 2E_1$ where E_1 is the exceptional curve above p_1 and m' is the strict transform of m. The strict transform of L_1 intersects $2E_1$, but does not intersect the curve $3m'$. The second blow up is that of p_1', the intersection point of $3m'$ and $2E_1$. Since the strict transform of L_1 by the first blow up does not pass through p_1', this curve or its intersection with other divisors is unaffected by the remaining blow ups. Thus L_1', the strict transform of L_1 by f, intersects the multiplicity-two component of F in a single point. This assures that the map $\varphi_{L_1} : L_1' \to B$ is ramified above the place corresponding to F.

Now suppose that F is induced by $m \cup 2l$ where l is another line. We may suppose that L_1 does not pass through other base points and hence L_1 intersects $2l$ at a point that is not in the blow up locus of f. This implies that L_1' intersects F at the double component corresponding to the strict transform of $2l$, and thus φ_{L_1} ramifies above the place corresponding to F.

By Lemma 2.1, $\mathscr{E} \times_B L_1'$ is rational. The existence of a nonreduced fiber in $\mathscr{E}$ assures that $\operatorname{rank}(\operatorname{Pic}(\mathscr{E})|k) \geq 4$, since the components of the fiber that do not intersect the zero section contribute with at least two divisors to the Picard group over k. By Corollary 4.3, the rank of the surface $\mathscr{E} \times_B L_1'$ over k is strictly larger than that of $\mathscr{E}$, for all but finitely many $L_1 \in \mathscr{L}_1$, and thus $\operatorname{rank}(\operatorname{Pic}(\mathscr{E} \times_B L_1')|k) \geq 5$. By Lemma 3.5, $\mathscr{E} \times_B L_1'$ is k-unirational. □

We apply Theorem 1.1 to the surfaces satisfying the hypothesis of the previous proposition. This gives us the following theorem.

Theorem 6.5. *Let $\mathscr{E} \to B$ be a rational elliptic surface as in Proposition 6.4. Then there is a finite covering $C \to B$ such that $C \simeq_k \mathbb{P}^1$, and the surface $\mathscr{E}_C = \mathscr{E} \times_B C$ satisfies* $\operatorname{rank} \mathscr{E}_C(k(C)) \geq \operatorname{rank} \mathscr{E}(k(B)) + 2$.

As before, we get the following corollary.

Corollary 6.6. *Let $\pi : \mathscr{E} \to B$ be an elliptic surface as in Proposition 6.4. For $t \in B(k)$, let r_t be the rank of the fiber above the point t and r the generic rank. Then*

$$\#\{t \in B(k) \mid r_t \geq r + 2\} = \infty.$$

Acknowledgements

The author would like to thank Marc Hindry for suggesting this beautiful problem and for several fruitful discussions on the topic, Damiano Testa, Daniel Bertrand, Douglas Ulmer, Jean-Louis Colliot-Thélène and Ronald van Luijk for discussions and suggestions that improved the quality of the paper. Finally the author would like to thank the referee for the careful reading of the article as well as for important suggestions that improved the readability and the quality of the paper.

References

[Billard 1998] H. Billard, “Sur la répartition des points rationnels de surfaces elliptiques”, *J. Reine Angew. Math.* **505** (1998), 45–71. MR 99m:14039 Zbl 0948.14018

[Cassels and Schinzel 1982] J. W. S. Cassels and A. Schinzel, “Selmer’s conjecture and families of elliptic curves”, *Bull. London Math. Soc.* **14**:4 (1982), 345–348. MR 84d:14028 Zbl 0474.14010

[Grothendieck 1960] A. Grothendieck, “Éléments de géométrie algébrique, I: Le langage des schémas”, *Inst. Hautes Études Sci. Publ. Math.* **4** (1960), 5–228. MR 36 #177a Zbl 0118.36206

[Helfgott 2003] H. A. Helfgott, *Root numbers and the parity problem*, thesis, Princeton University, 2003, Available at http://tel.archives-ouvertes.fr/docs/00/04/84/09/PDF/tel-00010129.pdf.

[Helfgott 2004] H. A. Helfgott, "On the behaviour of root numbers in families of elliptic curves", preprint, 2004. arXiv math/0408141

[Hindry 1988] M. Hindry, "Autour d'une conjecture de Serge Lang", *Invent. Math.* **94**:3 (1988), 575–603. MR 89k:11046 Zbl 0638.14026

[Iskovskikh 1979] V. A. Iskovskikh, "Minimal models of rational surfaces over arbitrary fields", *Izv. Akad. Nauk SSSR Ser. Mat.* **43**:1 (1979), 19–43. In Russian; translated in *Math. USSR-Izv* **14**:1 (1980), 17–39. MR 80m:14021 Zbl 0412.14012

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergeb. Math. Grenzgeb. (3) **32**, Springer, Berlin, 1996. MR 98c:14001 Zbl 0877.14012

[Manduchi 1995] E. Manduchi, "Root numbers of fibers of elliptic surfaces", *Compositio Math.* **99**:1 (1995), 33–58. MR 96j:11076 Zbl 0878.14028

[Manin 1966] Y. I. Manin, "Rational surfaces over perfect fields", *Inst. Hautes Études Sci. Publ. Math.* **30** (1966), 55–113. In Russian. MR 37 #1373 Zbl 0171.41701

[Manin 1967] Y. I. Manin, "Rational surfaces over perfect fields, II", *Mat. Sb.* (*N.S.*) **72** (1967), 161–192. In Russian; translated in *Math. USSR, Sb.* **1**: (1967), 141–168. MR 37 #1374

[Manin and Tsfasman 1986] Y. I. Manin and M. A. Tsfasman, "Rational varieties: algebra, geometry, arithmetic", *Uspekhi Mat. Nauk* **41**:2(248) (1986), 43–94. In Russian; translated in *Russian Math. Surveys* **41**:2 (1986), 51–116. MR 87k:11065 Zbl 0621.14029

[Miranda 1980] R. Miranda, "On the stability of pencils of cubic curves", *Amer. J. Math.* **102**:6 (1980), 1177–1202. MR 82j:14009 Zbl 0468.14009

[Miranda and Persson 1989] R. Miranda and U. Persson, "Torsion groups of elliptic surfaces", *Compositio Math.* **72**:3 (1989), 249–267. MR 91e:14035 Zbl 0717.14019

[Néron 1956] A. Néron, "Propriétés arithmétiques de certaines familles de courbes algébriques", pp. 481–488 in *Proceedings of the International Congress of Mathematicians* (Amsterdam, 1954), vol. 3, edited by J. Gerretsen and J. de Groot, E. P. Noordhoff, Groningen, 1956. MR 19,321b Zbl 0074.15901

[Ribet 1979] K. A. Ribet, "Kummer theory on extensions of abelian varieties by tori", *Duke Math. J.* **46**:4 (1979), 745–761. MR 81g:14019 Zbl 0428.14018

[Rohrlich 1993] D. E. Rohrlich, "Variation of the root number in families of elliptic curves", *Compositio Math.* **87**:2 (1993), 119–151. MR 94d:11045 Zbl 0791.11026

[Salgado 2009] C. Salgado, *Rang des surfaces elliptiques: théorèmes de comparaison*, thesis, Université Denis Diderot, 2009, Available at www.math.leidenuniv.nl/~salgado/thesececi9.pdf.

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR 52 #8126 Zbl 0235.14012

[Shioda 1990] T. Shioda, "On the Mordell–Weil lattices", *Comment. Math. Univ. St. Paul.* **39**:2 (1990), 211–240. MR 91m:14056 Zbl 0725.14017

[Silverman 1983] J. H. Silverman, "Heights and the specialization map for families of abelian varieties", *J. Reine Angew. Math.* **342** (1983), 197–211. MR 84k:14033 Zbl 0505.14035

[Silverman 1994] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer, New York, 1994. MR 96b:11074 Zbl 0911.14015

[Testa 2009] D. Testa, "The irreducibility of the spaces of rational curves on del Pezzo surfaces", *J. Algebraic Geom.* **18**:1 (2009), 37–61. MR 2010a:14057 Zbl 1165.14024

[Ulas 2008] M. Ulas, "Rational points on certain del Pezzo surfaces of degree one", *Glasg. Math. J.* **50**:3 (2008), 557–564. MR 2009g:11034 Zbl 1223.14041

Communicated by Jean-Louis Colliot-Thélène
Received 2010-11-18 Revised 2011-12-19 Accepted 2012-01-24

salgado@im.ufrj.br *Mathematisch Instituut, Universiteit Leiden, Niels Bohrweg, 1, 2333 CA Leiden, The Netherlands*

Current address: *Instituto de Matemática, Universidade Federal do Rio de Janeiro, 21941-909 - Rio de Janeiro, RJ, Brazil*

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1315

Néron's pairing and relative algebraic equivalence

Cédric Pépin

Let R be a complete discrete valuation ring with algebraically closed residue field k and fraction field K. Let X_K be a proper smooth and geometrically connected scheme over K. Néron defined a canonical pairing on X_K between 0-cycles of degree zero and divisors which are algebraically equivalent to zero. When X_K is an abelian variety, and if one restricts to those 0-cycles supported on K-rational points, Néron gave an expression of his pairing involving intersection multiplicities on the Néron model A of A_K over R. When X_K is a curve, Gross and Hriljac gave independently an analogous description of Néron's pairing, but for arbitrary 0-cycles of degree zero, by means of intersection theory on a proper flat regular R-model X of X_K.

We show that these intersection computations are valid for an arbitrary scheme X_K as above and arbitrary 0-cycles of degree zero, by using a proper flat *normal and semifactorial* model X of X_K over R. When $X_K = A_K$ is an abelian variety, and $X = \overline{A}$ is a semifactorial compactification of its Néron model A, these computations can be used to study the relative algebraic equivalence on $\overline{A}/R$. We then obtain an interpretation of Grothendieck's duality for the Néron model A, in terms of the Picard functor of $\overline{A}$ over R. Finally, we give an explicit description of Grothendieck's duality pairing when A_K is the Jacobian of a curve of index one.

1. Introduction

Let R be a complete discrete valuation ring with algebraically closed residue field k and fraction field K. Let X_K be a proper smooth and geometrically connected

Partially supported by the FWO project ZKC1235-PS-011 G.0415.10 at KU Leuven.
MSC2010: primary 14K30; secondary 14G40, 14K15, 11G10.
Keywords: Néron's symbol, Picard functor, Néron models, duality, Grothendieck's pairing.

scheme over K. Denote by $Z_0^0(X_K)$ the group of 0-cycles of degree zero on X_K, and by $\mathrm{Div}^0(X_K)$ the group of divisors which are algebraically equivalent to zero on X_K. For each $c_K \in Z_0^0(X_K)$ and $D_K \in \mathrm{Div}^0(X_K)$ with disjoint supports, Néron attached a rational number

$$\langle c_K, D_K\rangle \in \mathbb{Q},$$

by using the unique (up to constant) Néron function associated to D_K. This defines a bilinear pairing $\langle\cdot,\cdot\rangle$; see [Néron 1965, II 9.3].

Suppose first that $X_K = A_K$ is an abelian variety, and denote by A its Néron model over R. By definition of A, any K-rational point of A_K extends to a section of A over R. Then, if c_K is supported on K-rational points, Néron showed that the pairing attached to A_K can be decomposed as follows:

$$\langle c_K, D_K\rangle = i(c_K, D_K) + j(c_K, D_K), \tag{1}$$

where $i(c_K, D_K)$ is the intersection multiplicity $(\bar{c}_K.\overline{D}_K) \in \mathbb{Z}$ of the schematic closures in A, and $j(c_K, D_K) \in \mathbb{Q}$ depends only on the specialization of c_K on the group Φ_A of connected components of the special fiber A_k; see [Néron 1965, III 4.1; Lang 1983, 11.5.1].

Suppose now that X_K is a curve, and denote by X a proper flat regular model of X_K over R. Let M be the intersection matrix of the special fiber X_k of X/R: if $\Gamma_1, \dots, \Gamma_\nu$ are the irreducible components of X_k equipped with their reduced scheme structure, the (i, j)-th entry of M is the intersection number $(\Gamma_i \cdot \Gamma_j)$. Let $D_K \in \mathrm{Div}^0(X_K)$ and let $\overline{D}_K$ be its closure in X. Computing the degree $(\overline{D}_K.\Gamma_i)$ of $\overline{D}_K$ along each Γ_i, we get a vector $\rho(\overline{D}_K) \in \mathbb{Z}^\nu$. Next, as a consequence of intersection theory on X, there exists a vector $V \in \mathbb{Q}^\nu$ such that $\rho(\overline{D}_K) = MV$. Denote again by V the $\mathbb{Q}$-linear combination of the Γ_i where the coefficient of Γ_i is the i-th entry of V. Then, for any $c_K \in Z_0^0(X_K)$ whose support is disjoint from that of D_K, the following formula holds:

$$\langle c_K, D_K\rangle = (\bar{c}_K.\overline{D}_K) + (\bar{c}_K.(-V)), \tag{2}$$

where the second intersection number is defined by $\mathbb{Q}$-linearity from the $(\bar{c}_K.\Gamma_i)$. See [Gross 1986; Hriljac 1985; Lang 1988, III 5.2]. Now let J_K be the Jacobian of X_K and let J be its Néron model over R. Following the point of view of Bosch and Lorenzini [2002, 4.3], it results from Raynaud's theory of the Picard functor $\mathrm{Pic}_{X/R}$ [Raynaud 1970, Section 8] that the term $(\bar{c}_K.(-V))$ depends only on the specialization of $(c_K) \in J_K(K)$ into the group of components Φ_J of J_k.

In Section 2, we provide a unified approach to these two descriptions of Néron's pairing. More precisely, for an arbitrary proper geometrically normal and geometrically connected scheme X_K, there always exists some proper flat normal semi-factorial model X of X_K over R [Pepin 2011, Theorem 2.6]. Recall that X/R is *semifactorial* if the restriction homomorphism on Picard groups $\mathrm{Pic}(X) \to \mathrm{Pic}(X_K)$ is surjective. Note that a regular model is semifactorial. Using the theory of the Picard functor of semifactorial models, we define a pairing $[\cdot,\cdot]$ on X_K involving intersection multiplicities on X (Definition 2.1.1). It turns out that this pairing depends only on X_K, and *coincides* with Néron's pairing when the latter is defined, that is, when X_K is smooth:

$$\langle\cdot,\cdot\rangle = [\cdot,\cdot] \tag{3}$$

(Theorem 2.2.1). If $X_K = A_K$ is an abelian variety and $X = \overline{A}$ is a semifactorial compactification of its Néron model A, then equality (3) provides decomposition (1) for 0-cycles supported on K-rational points. If X_K is a curve and X a proper flat regular model of X_K, then the intersection matrix of X_k is defined, and equality (3) is exactly formula (2).

In Section 3, we consider an abelian variety A_K, with dual A'_K. By definition, the abelian variety A'_K parametrizes the divisors on A_K which are algebraically equivalent to zero, that is, $A'_K = \mathrm{Pic}^0_{A_K/K}$. Now, let A'/R be the Néron model of A'_K, and denote by $(A')^0$ its identity component. By restricting to the generic fiber, the group of sections $(A')^0(R)$ can be viewed as a subgroup of $A'_K(K)$. On the other hand, let $\overline{A}$ be a normal semifactorial compactification of A, let $\mathrm{Pic}_{\overline{A}/R}$ be its relative Picard functor, and let $\mathrm{Pic}^0_{\overline{A}/R}$ be the component of the zero section. By restricting to the generic fiber, the group $\mathrm{Pic}^0_{\overline{A}/R}(R)$ can be viewed as a subgroup of $\mathrm{Pic}^0_{A_K/K}(K)$.

In Theorem 3.2.1, we investigate the relationship between the two groups

$$(A')^0(R) \quad \text{and} \quad \mathrm{Pic}^0_{\overline{A}/R}(R) \quad (\text{contained in } A'_K(K) = \mathrm{Pic}^0_{A_K/K}(K)).$$

We show that they are *equal* as soon as the duality conjecture of Grothendieck about A and A' is true [SGA 7 I 1972, IX 1.3]. More precisely, Grothendieck defined a pairing between the component groups of the special fibers of A and A', and he conjectured that this pairing is perfect. This duality statement has been proved in many situations (see the introduction of [Bosch and Lorenzini 2002] for a detailed list of the known cases, and also [Loerke 2009]), but it remains open in equal characteristic $p > 0$. Here, we give an equivalent formulation of Grothendieck's conjecture, in terms of Cartier divisors on $\overline{A}$. As a consequence, when the conjecture is true, we obtain the equality $(A')^0(R) = \mathrm{Pic}^0_{\overline{A}/R}(R)$. As a Cartier divisor on $\overline{A}$ is said to be algebraically equivalent to zero relative to R if its image into $\mathrm{Pic}_{\overline{A}/R}(R)$ is contained $\mathrm{Pic}^0_{\overline{A}/R}(R)$, the latter equality says that these divisors are parametrized by $(A')^0$. The main ingredients for the proof are a theorem of Bosch and Lorenzini about Néron's and Grothendieck's pairings [Bosch and Lorenzini 2002, 4.4], and the study of the pairing $[\cdot,\cdot]$ introduced above, especially for 0-cycles supported on *nonrational* points (Proposition 3.4.2).

In Section 4, we examine the relationship between Néron's and Grothendieck's pairing for the Jacobian of a curve, following Bosch and Lorenzini [2002, 4.6] and Lorenzini [2008, 3.4]. Here we take into account the *index of the curve* (Theorem 4.1.1). As a consequence, we obtain the perfectness of Grothendieck's pairing when this index is prime to the characteristic of the residue field k (Corollary 4.1.2).

2. Néron's pairing and intersection multiplicities

In this article, let us adopt the following terminology: a *divisor* on a scheme will always be a *Cartier divisor*.

2.1. ***A canonical pairing computed on semifactorial models.*** Let R be a discrete valuation ring with fraction field K and residue field k. *We assume R complete and k algebraically closed.* Let X_K be a proper geometrically normal and geometrically connected scheme over K. From [Pepin 2011, Theorem 2.6], there exists a model X/R of X_K, that is, an R-scheme with generic fiber X_K, which is proper, flat, normal and *semifactorial*: every invertible sheaf on X_K can be extended to an invertible sheaf on X. To each 0-cycle $c_K \in Z_0^0(X_K)$ and divisor $D_K \in \mathrm{Div}^0(X_K)$ with support disjoint from that of c_K, we will attach a number $[c_K, D_K]_X \in \mathbb{Q}$ using intersection multiplicities on X. For this purpose, let us first recall some definitions and one result.

Intersection multiplicities. Let X/R be a proper R-scheme. Let c_K be a 0-cycle on the generic fiber X_K. Denote by $\bar{c}_K$ its schematic closure in X: if $c_K = \sum_i n_i[x_{K,i}]$, then $\bar{c}_K = \sum_i n_i[\bar{x}_{K,i}]$, where $\bar{x}_{K,i}$ is the closure in X of the closed point $x_{K,i}$ of X_K. On the other hand, let Δ be a divisor on X whose support does not meet that of c_K. The *intersection multiplicity* $(\bar{c}_K.\Delta)$ of $\bar{c}_K$ and Δ on X is defined as follows. Let x_K be a point of the support of c_K. Let $Z := \bar{x}_K$ be its schematic closure in X. This is an integral scheme, finite and flat over R, which is local because R is henselian. Set $x_k := Z \cap X_k$. If $f \in K(X)$ is a local equation for Δ in a neighborhood of x_k, then $(\bar{c}_K.\Delta)_{x_k}$ is the order of $f|_Z$ at x_k: writing $f|_Z = a/b$ with regular $a, b \in \mathcal{O}(Z)$, then

$$(\bar{c}_K.\Delta)_{x_k} = \mathrm{length}_{\mathcal{O}(Z)}\big(\mathcal{O}(Z)/(a)\big) - \mathrm{length}_{\mathcal{O}(Z)}\big(\mathcal{O}(Z)/(b)\big)$$

[Fulton 1998, page 8]. The whole intersection multiplicity $(\bar{c}_K.\Delta)$ is defined by $\mathbb{Z}$-linearity.

Let us also give another description of $(\bar{c}_K.\Delta)_{x_k}$, which will be useful in the sequel. As R is excellent, the normalization $\widetilde{Z} \to Z$ is finite. Moreover, as k is algebraically closed,

$$\mathrm{length}_{\mathcal{O}(Z)}\big(\mathcal{O}(Z)/(a)\big) = \mathrm{length}_R\big(\mathcal{O}(Z)/(a)\big),$$

for any regular $a \in \mathcal{O}(Z)$, and the same formula holds with Z replaced by $\widetilde{Z}$ [loc. cit., Appendix A.1.3]. But

$$\operatorname{length}_R\big(\mathcal{O}(Z)/(a)\big) = \operatorname{length}_R\big(\mathcal{O}(\widetilde{Z})/(a)\big)$$

for any regular $a \in \mathcal{O}(Z)$; see [Bosch et al. 1990, end of page 237]. Thus, if $f \in K(X)$ is a local equation for Δ in a neighborhood of x_k, we have obtained that

$$(\bar{c}_K.\Delta)_{x_k} = \begin{cases} \operatorname{length}_{\mathcal{O}(\widetilde{Z})}\big(\mathcal{O}(\widetilde{Z})/(f)\big) & \text{if } f|_{\widetilde{Z}} \in \mathcal{O}(\widetilde{Z}), \\ -\operatorname{length}_{\mathcal{O}(\widetilde{Z})}\big(\mathcal{O}(\widetilde{Z})/(f^{-1})\big) & \text{otherwise.} \end{cases}$$

Relative algebraic equivalence and relative τ-equivalence [Raynaud 1970, 3.2d; SGA 6 1971, XIII 4]. If G is a commutative group scheme locally of finite type over a field, the *identity component* G^0 of G is the open subscheme of G whose underlying topological space is the connected component of the identity element of G. The *τ-component* of G is open subgroup scheme G^τ of G which is the inverse image of the torsion subgroup of G/G^0. When G is a commutative group functor over a scheme T, whose fibers are representable by schemes locally of finite type, the *identity component* and *τ-component* of G are the subfunctors G^τ of G whose fibers are the G_t^0, $t \in T$ and G_t^τ, $t \in T$, respectively. Note that $G^0 \subseteq G^\tau$.

Let $Z \to T$ be a proper morphism of schemes. Then the fibers of the Picard functor $\operatorname{Pic}_{Z/T}$ are representable by schemes locally of finite type [Murre 1964; Oort 1962]. Let $\mathcal{L}$ be an invertible $\mathcal{O}_Z$-module. The sheaf $\mathcal{L}$ is said to be *algebraically equivalent to zero relative to T* if its image into $\operatorname{Pic}_{Z/T}(T)$ belongs to the subgroup $\operatorname{Pic}^0_{Z/T}(T)$, that is $\mathcal{L}_t \in \operatorname{Pic}^0_{Z_t/t}(t)$ for all $t \in T$. When there is no ambiguity about the base scheme T, we will just say that $\mathcal{L}$ is *algebraically equivalent to zero*. Similarly, the sheaf $\mathcal{L}$ is said to be *τ-equivalent to zero relative to T* if its image into $\operatorname{Pic}_{Z/T}(T)$ belongs to the subgroup $\operatorname{Pic}^\tau_{Z/T}(T)$, that is $\mathcal{L}_t \in \operatorname{Pic}^\tau_{Z_t/t}(t)$ for all $t \in T$. If D is a divisor on Z, it is *algebraically equivalent to zero*, or *τ-equivalent to zero*, respectively, *relative to T* if the associated invertible sheaf $\mathcal{O}_Z(D)$ is. We denote by $\operatorname{Div}^0(Z/T)$ and $\operatorname{Div}^\tau(Z/T))$ the groups of divisors on Z which are algebraically equivalent to zero and τ-equivalent to zero, respectively, relative to T. Then $\operatorname{Div}^0(Z/T) \subseteq \operatorname{Div}^\tau(Z/T)$.

Relative algebraic equivalence and semifactoriality. Let X/R be a proper flat *semifactorial* R-scheme. Suppose that the generic fiber X_K is geometrically normal and geometrically connected. Its Picard variety $\operatorname{Pic}^0_{X_K/K,\mathrm{red}}$ is then an abelian variety [FGA VI 1966, 3.2]. Let A/R be its Néron model, and let n be the exponent of the component group of the special fiber of A. In this situation, [Pepin 2011, Corollary 3.14] can be read as follows: for any divisor D_K on X_K which is algebraically equivalent to zero, there exists a divisor Δ on X which is algebraically equivalent to zero relative to R and whose generic fiber Δ_K is equal to nD_K.

Definition 2.1.1. Let X_K be a proper, geometrically normal and geometrically connected scheme over K. Let X/R be a proper, flat, normal and semifactorial model of X_K over R.

Consider $c_K \in Z_0^0(X_K)$ and $D_K \in \mathrm{Div}^\tau(X_K)$ with disjoint supports. Let $\bar{c}_K$ be the schematic closure of c_K in X. Choose any $(n, \Delta) \in (\mathbb{Z} \setminus \{0\}) \times \mathrm{Div}^\tau(X/R)$ such that $\Delta_K = nD_K$. Then set

$$[c_K, D_K]_X := \frac{1}{n}(\bar{c}_K.\Delta) \in \mathbb{Q}.$$

This definition makes sense because $\frac{1}{n}\Delta \in \mathrm{Div}^\tau(X/R) \otimes_{\mathbb{Z}} \mathbb{Q}$ is uniquely determined by D_K, up to a rational multiple of the principal divisor X_k. Indeed, if (n', Δ') is another choice in Definition 2.1.1, then the divisor $n'\Delta - n\Delta'$ is τ-equivalent to zero on X and equal to zero on X_K. Thus, as X is normal, this difference is a rational multiple of X_k [Raynaud 1970, 6.4.1 3]. Now note that $(\bar{c}_K.X_k)$ is equal to the degree of c_K, which is zero, so that $\frac{1}{n}(\bar{c}_K.\Delta) = \frac{1}{n'}(\bar{c}_K.\Delta')$.

Next, one checks easily that the symbol $[\,\cdot\,,\cdot\,]_X$ is *bilinear* (in its definition domain). To prove that this pairing does not depend on the choice of X, we will use the following lemma.

We will denote by $(\cdot)_*$ and $(\cdot)^*$ the push-forward of cycles and the pull-back of divisors respectively; see [Fulton 1998, 20.1.3] and [Liu 2002, 7.1.29, 7.1.33, 7.1.34], respectively.

Lemma 2.1.2. *Let X and X' be integral schemes, proper over R. Let $\varphi : X \to X'$ be an R-morphism. Let $c_K \in Z_0^0(X_K)$ and let $\bar{c}_K$ be its schematic closure in X. Let Δ' be a divisor on X' whose support does not meet that of $(\varphi_K)_*c_K$. Then the following projection formula holds*:

$$\bar{c}_K.\varphi^*\Delta' = \varphi_*\bar{c}_K.\Delta'.$$

In particular, let X and X' be proper, flat, normal and semifactorial schemes over R, with geometrically normal and geometrically connected generic fibers, so that $[\,\cdot\,,\cdot\,]_X$ and $[\,\cdot\,,\cdot\,]_{X'}$ are defined. Let $\varphi : X \to X'$ be an R-morphism. Let $c_K \in Z_0^0(X_K)$, and let $D'_K \in \mathrm{Div}^\tau(X'_K)$ whose support does not meet that of $(\varphi_K)_*c_K$. Then the following equality holds:

$$[c_K, (\varphi_K)^*D'_K]_X = [(\varphi_K)_*c_K, D'_K]_{X'}.$$

Proof. Let us first note that the divisors $\varphi^*\Delta'$ (and $(\varphi_K)^*D'_K$) are well-defined. Indeed, as φ is proper, its image Y is a closed subset of X'. Endow Y with its reduced scheme structure. As X is reduced, φ factors through Y:

$$\begin{tikzcd} X \arrow[r, "\psi"] \arrow[dr, "\varphi"'] & Y \arrow[d, hook, "\iota"] \\ & X'. \end{tikzcd}$$

Now, by hypothesis, the support of Δ' is disjoint from that of $(\varphi_K)_* c_K$. In particular, Y is not contained in the support of Δ'. So the pullback $\iota^*\Delta'$ is well-defined. Next, X and Y being integral and ψ dominant, $\varphi^*\Delta' := \psi^*(\iota^*\Delta')$ is well-defined.

Let us now recall the proof of the projection formula $\bar{c}_K.\varphi^*\Delta' = \varphi_*\bar{c}_K.\Delta'$. Let x_K be a closed point of the support of c_K, let Z be its schematic closure in X, set $x_k := Z \cap X_k$ and let $\widetilde{Z}$ be the normalization of Z. The reduced scheme $V := \varphi(Z)$ is the schematic closure of $\varphi(x_K)$ and we have $\varphi(x_k) = V \cap X_k$. Denote by $\widetilde{V}$ the normalization of V. The morphism φ induces a finite surjective morphism $Z \to V$, which in turn induces a finite surjective morphism $\widetilde{Z} \to \widetilde{V}$ (R is excellent). Let f' be a local equation of Δ' at $\varphi(x_k)$. Suppose for example that $f'|_{\widetilde{V}} \in \mathbb{O}(\widetilde{V})$. The equality $(\varphi_*\bar{c}_K.\Delta')_{x_k} = (\bar{c}_K.\varphi^*\Delta')_{x_k}$ to be proved can be written as

$$[K(Z):K(V)] \cdot \operatorname{length}\big(\mathbb{O}(\widetilde{V})/(f')\big) = \operatorname{length}\big(\mathbb{O}(\widetilde{Z})/(\varphi^* f')\big).$$

But $[K(Z):K(V)]$ is equal to the ramification index of the discrete valuation rings extension $\mathbb{O}(\widetilde{V}) \to \mathbb{O}(\widetilde{Z})$. Consequently, the above formula is true.

Now, when the pairings $[\,\cdot\,,\cdot\,]_X$ and $[\,\cdot\,,\cdot\,]_{X'}$ are defined, the projection formula can be written as the equality $[c_K, (\varphi_K)^* D'_K]_X = [(\varphi_K)_* c_K, D'_K]_{X'}$. Indeed, let Δ' be a divisor which is τ-equivalent to zero on X' and let n' be a nonzero integer such that $(\Delta')_K = n' D'_K$. The direct image $\varphi_*\bar{c}_K$ of the schematic closure of c_K coincides with the schematic closure of $(\varphi_K)_* c_K$. Thus, by definition,

$$n'[(\varphi_K)_* c_K, D'_K]_{X'} = \varphi_*\bar{c}_K.\Delta'.$$

The divisor $\varphi^*\Delta'$ is τ-equivalent to zero on X, and satisfies $(\varphi^*\Delta')_K = n'(\varphi_K)^* D'_K$. Hence, by definition,

$$n'[c_K, (\varphi_K)^* D'_K]_X = \bar{c}_K.\varphi^*\Delta'. \qquad \square$$

In the situation of Definition 2.1.1, let X' be another proper flat normal semifactorial R-model of X_K. Consider the graph Γ of the rational map $X \dashrightarrow X'$ induced by the identity on the generic fibers. By definition, this is the schematic closure of the graph of the identity morphism $X_K \to X'_K$ in $X \times_R X'$. In particular, this is a closed subscheme of $X \times_R X'$, proper and flat over R, with generic fiber isomorphic to X_K. Applying [Pepin 2011, Theorem 2.6], we can find an R-scheme $\widetilde{X}$ which is proper flat normal and semifactorial, together with an R-morphism $\widetilde{X} \to \Gamma$ which is an isomorphism on the generic fibers. Composing with the two projections from $X \times_R X'$ to X and X', we get arrows

$$\begin{tikzcd} & \widetilde{X} \arrow[dl] \arrow[dr] & \\ X & & X' \end{tikzcd}$$

which are isomorphisms on the generic fibers. Now, Lemma 2.1.2 shows that the pairings $[\cdot,\cdot]_X$ and $[\cdot,\cdot]_{X'}$ both coincide with $[\cdot,\cdot]_{\widetilde{X}}$. In conclusion, the pairing $[\cdot,\cdot]_X$ depends only on X_K, and not on the choice of X.

Let us summarize the above considerations:

Proposition 2.1.3. *Let X_K be a proper, geometrically normal and geometrically connected scheme over K. There exists a pairing*

$$[\cdot,\cdot] : Z_0^0(X_K) \times \mathrm{Div}^\tau(X_K) \to \mathbb{Q},$$

defined for the pairs (c_K, D_K) such that the supports of c_K and D_K are disjoint, and which can be computed as follows.

Let X/R be any proper flat normal and semifactorial model of X_K over R. Let $\bar{c}_K$ be the schematic closure of c_K in X. Choose $(n, \Delta) \in (\mathbb{Z} \setminus \{0\}) \times \mathrm{Div}^\tau(X/R)$ such that $\Delta_K = nD_K$. Then we have

$$[c_K, D_K] = \frac{1}{n}(\bar{c}_K.\Delta) \in \mathbb{Q}.$$

2.2. ***Comparison with Néron's pairing.*** As before, let R be a complete discrete valuation ring with fraction field K and algebraically closed residue field k. Let X_K be a proper smooth and geometrically connected scheme over K. Let v be the *normalized* valuation on K, which maps any uniformizing element of R to $1 \in \mathbb{Z}$. We fix an algebraic closure $\overline{K}$ of K, and we still denote by v the unique valuation on $\overline{K}$ extending v. Néron attached to X_K a pairing $\langle \cdot,\cdot \rangle$ with respect to the valuation v [Néron 1965, II 9.3]. This is a pairing

$$\langle \cdot,\cdot \rangle : Z_0^0(X_K) \times \mathrm{Div}^\tau(X_K) \to \mathbb{R},$$

defined for (c_K, D_K) when the supports of c_K and D_K are disjoint (the definition of Néron's pairing is briefly reviewed at the beginning of the proof of Theorem 2.2.1). Actually, Néron considers the subgroup $\mathrm{Div}^0(X_K) \subseteq \mathrm{Div}^\tau(X_K)$ to consist of divisors which are algebraically equivalent to zero on X_K. However, the group $(\mathbb{R}, +)$ being divisible, the real number $\langle c_K, D_K \rangle$ is naturally defined when D_K is only τ-equivalent to zero. Néron shows in [loc. cit., III 4.2] that the pairing takes values in $\mathbb{Q}$. This fact will be recovered and made more precise below (Corollary 2.2.2).

Our goal in this subsection is to prove the following common generalization of Néron [1965, III 4.1], Gross [1986], Hriljac [1985], Lang [1988, III 5.2] and Bosch–Lorenzini [2002, 4.3], over a complete discrete valuation ring R with algebraically closed residue field k and fraction field K.

Theorem 2.2.1. *For every proper, smooth and geometrically connected scheme over K, the pairing $[\,\cdot\,,\cdot\,]$ defined in Proposition 2.1.3 coincides with Néron's pairing $\langle\,\cdot\,,\cdot\,\rangle$ defined in* [Néron 1965, II.9, Theorem 3].

In particular, the pairing $[\,\cdot\,,\cdot\,]$ generalizes Néron's pairing to K-schemes which are proper geometrically normal and geometrically connected, but not necessarily smooth.

Before proving the theorem, let us note a consequence of Proposition 2.1.3.

Corollary 2.2.2. *Let X_K be a proper, geometrically normal and geometrically connected scheme over K. Let n be the exponent of the component group of the special fiber of the Néron model of the Picard variety $A_K = \mathrm{Pic}^0_{X_K/K,\mathrm{red}}$. Then Néron's pairing on $Z_0^0(X_K) \times \mathrm{Div}^0(X_K)$ takes values in $(1/n)\,\mathbb{Z}$.*

Proof. As recalled before Definition 2.1.1, the exponent n has the following property: for any $D_K \in \mathrm{Div}^0(X_K)$ and any proper flat normal and semifactorial model X of X_K, there exists $\Delta \in \mathrm{Div}^0(X/R)$ such that $\Delta_K = nD_K$. In particular, for any $D_K \in \mathrm{Div}^0(X_K)$, we can choose this integer n, together with a divisor $\Delta \in \mathrm{Div}^0(X/R)$, to compute

$$[c_K, D_K] = \frac{1}{n}(\bar{c}_K.\Delta) \in \frac{1}{n}\,\mathbb{Z}\,.$$

Now Theorem 2.2.1 asserts that $\langle c_K, D_K\rangle = [c_K, D_K]$. □

Corollary 2.2.2 provides a refinement of [Néron 1965, III 4.2]. More precisely, Néron shows that the pairing

$$\langle\,\cdot\,,\cdot\,\rangle : Z_0^0(X_K) \times \mathrm{Div}^0(X_K) \to \mathbb{R}$$

takes values in $(1/2n'ab)\,\mathbb{Z}$, where n', a and b are defined as follows. The integer n' is the exponent of the component group of the special fiber of the Néron model of the Albanese variety A'_K of X_K. Conjecturally, n' is equal to n; see Section 3.1. Next, a is the smallest positive integer such that there exists a map $h : X_K \to A'_K$ from X_K to its Albanese variety, with the property that for any divisor $D_K \in \mathrm{Div}^0(X_K)$, there exists a divisor $W_K \in \mathrm{Div}^0(A'_K)$ such that h^*W_K is linearly equivalent to aD_K. We can have $a > 1$ if $X_K(K)$ is empty. Finally, b is the smallest degree of a polarization of the Albanese variety A'_K.

In [Mazur and Tate 1983, (1.5) and (2.3); Lang 1983, 11.5.1-11.5.2], it is proved that $\langle c_K, D_K\rangle$ belongs to $(1/n')\,\mathbb{Z}$ when X_K is an abelian variety and if c_K is supported on rational points. This statement is also a consequence of [Bosch and

Lorenzini 2002, 4.4]. Moreover, note that Néron's pairing can take the value $1/n$, for instance when X_K is an elliptic curve; see [loc. cit., Example 5.8].

Let us go back to Theorem 2.2.1. To prove the theorem, we will use the characterization of Néron's pairing given in [Lang 1983, 11.3.2] and that we recall now.

An element c_K of $Z_0^0(X_K)$ can be written uniquely as a difference of two positive 0-cycles with disjoint supports: $c_K = c_K^+ - c_K^-$. Denoting by deg the degree of a 0-cycle, let us set

$$\deg^+ c_K := \deg(c_K^+) = \deg(c_K^-) \geq 0.$$

Lemma 2.2.3 [Lang 1983, 11.3.2]. *Suppose that for each projective smooth and geometrically connected scheme X_K over K, we are given a bilinear pairing*

$$\begin{aligned} Z_0^0(X_K) \times \mathrm{Div}^0(X_K) &\to \mathbb{R} \\ (c_K, D_K) &\mapsto \delta(c_K, D_K) \end{aligned}$$

such that the following properties are true:

(1) *If D_K is a principal divisor on X_K, then $\delta(c_K\,, D_K) = 0$.*

(2) *Let $\varphi_K : X_K \to X'_K$ be a K-morphism. For all $c_K \in Z_0^0(X_K)$, and for all $D'_K \in \mathrm{Div}^0(X'_K)$ whose support does not meet that of the* 0*-cycle $(\varphi_K)_* c_K$, the following equality holds*:

$$\delta(c_K, (\varphi_K)^* D'_K) = \delta((\varphi_K)_* c_K, D'_K).$$

(3) *For $D_K \in \mathrm{Div}^0(X_K)$ fixed and* $\deg^+ c_K$ *bounded, the values $\delta(c_K, D_K)$ are bounded.*

Then $\delta(c_K, D_K) = 0$ for all c_K, D_K and X_K.

Remark 2.2.4. In the statement of [Lang 1983, 11.3.2], one reads "projective variety V over K" instead of "projective smooth and geometrically connected scheme X_K over K". According to the general conventions of [loc. cit., page 21], a "variety over K" is a "geometrically integral scheme of finite type over K". However, the given proof of [loc. cit., 11.3.2] works if and only if the Albanese variety of each V is an abelian variety. The latter is true, for example, if each V is geometrically normal, or if each V is smooth. For our purposes, namely the proof of Theorem 2.2.1, we need the version of the lemma where all the V are smooth.

Proof of Theorem 2.2.1. Starting from the existence of *Néron functions* on a proper smooth and geometrically connected K-scheme X_K [Néron 1965, II 8.2], let us recall the definition of Néron's pairing. Let

$$c_K = \sum_i n_i [x_{K,i}] \in Z_0^0(X_K)$$

and take $D_K \in \mathrm{Div}^0(X_K)$ whose support $\mathrm{Supp}(D_K)$ does not contain any of the $x_{K,i}$. Let $\lambda_{D_K} : (X_K - \mathrm{Supp}(D_K))(\overline{K}) \to \mathbb{R}$ be a Néron function associated to D_K. For each i, the scheme $x_{K,i} \otimes_K \overline{K}$ is supported on some $\overline{K}$-points $x_{\overline{K},j_i}$, $j_i = 1, \ldots, s_i$, where s_i is the separable degree of $K(x_{K,i})/K$. Denoting by l_i the inseparable degree of $K(x_{K,i})/K$, then

$$\lambda_{D_K}(x_{K,i}) := \sum_{j_i=1}^{s_i} l_i \lambda_{D_K}(x_{\overline{K},j_i}) \quad \text{and} \quad \langle c_K, D_K \rangle := \sum_i n_i \lambda_{D_K}(x_{K,i}).$$

The real number $\langle c_K, D_K \rangle$ is well-defined because λ_{D_K} is unique up to constant and c_K has degree zero.

Comparison of the pairings for a principal divisor D_K. Let us keep the previous notation, and suppose that $D_K = \mathrm{div}_{X_K} f$ for a nonzero $f \in K(X_K)$. Let $z \in (X_K - \mathrm{Supp}(\mathrm{div}_{X_K} f))(\overline{K})$, mapping to a closed point $x_K \in X_K$. The evaluation of f at z is defined by the pull-back $z^* : \mathcal{O}_{X_K,x_K} \to \overline{K}$, that is, $f(z) := z^* f$. The formula $\lambda_f(z) = v(f(z))$ then defines a Néron function for the divisor $\mathrm{div}_{X_K} f$.

Fix an i. There is a 1-1 correspondence between the $x_{\overline{K},j_i}$ and the K-embeddings of the residue field extension $K(x_{K,i})/K$ into $\overline{K}/K$. By pulling back the valuation v, each of these embeddings induces a valuation on $K(x_{K,i})$. However, as R is complete, these valuations are equal to the unique valuation on $K(x_{K,i})$ which extends the normalized valuation on K, and that we can also denote by v. Consequently,

$$\lambda_f(x_{K,i}) = \sum_{j_i=1}^{s_i} l_i v(f(x_{K,i})) = [K(x_{K,i}) : K] v(f(x_{K,i}))$$

where $f(x_{K,i})$ is the image of f by the canonical surjection $\mathcal{O}_{X_K,x_{K,i}} \to K(x_{K,i})$.

Now, take the schematic closure Z_i of $x_{K,i}$ in X, denote by $\widetilde{Z}_i$ its normalization and set $x_{k,i} := X_k \cap Z_i$. The ring $\mathcal{O}(\widetilde{Z}_i)$ is a discrete valuation ring with fraction field $K(x_{K,i})$. So it is precisely the valuation ring of v in $K(x_{K,i})$. As k is algebraically closed, its ramification index over R is equal to $[K(x_{K,i}) : K]$. From this observation, we get

$$v(f(x_{K,i})) = \begin{cases} 1/[K(x_{K,i}) : K] \,\mathrm{length}_{\mathcal{O}(\widetilde{Z}_i)}\big(\mathcal{O}(\widetilde{Z}_i)/(f)\big) & \text{if } f|_{\widetilde{Z}_i} \in \mathcal{O}(\widetilde{Z}_i), \\ -1/[K(x_{K,i}) : K] \,\mathrm{length}_{\mathcal{O}(\widetilde{Z}_i)}\big(\mathcal{O}(\widetilde{Z}_i)/(f^{-1})\big) & \text{otherwise.} \end{cases}$$

We have thus obtained $[K(x_{K,i}) : K] v(f(x_{K,i})) = (\bar{c}_K . \mathrm{div}_X f)_{x_{k,i}}$ (recall the beginning of Section 2.1). But $\mathrm{div}_X f$ is a divisor on X which is τ-equivalent to zero and extends $\mathrm{div}_{X_K} f$. The desired equality $\langle c_K, \mathrm{div}_{X_K} f \rangle = [c_K, \mathrm{div}_{X_K} f]$ follows.

Functoriality of the pairing $[\,\cdot\,,\cdot\,]$. Let $\varphi_K : X_K \to X'_K$ be a K-morphism of proper smooth and geometrically connected schemes over K. Let us show that for all $c_K \in Z_0^0(X_K)$, and for all $D'_K \in \mathrm{Div}^\tau(X'_K)$ whose support does not meet that of the

0-cycle $(\varphi_K)_* c_K$, the following equality holds

$$[c_K, (\varphi_K)^* D'_K] = [(\varphi_K)_* c_K, D'_K].$$

Let X/R (resp. X'/R) be a proper flat normal semifactorial model of X_K (resp. X'_K). Consider the graph Γ of the rational map $X \dashrightarrow X'$ defined by φ_K. Applying Theorem 2.6 of [Pepin 2011] to Γ, we obtain a proper flat normal semifactorial $\widetilde{X}/R$ and R-morphisms

$$\begin{tikzcd} & \widetilde{X} \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\ X & & X' \end{tikzcd}$$

such that on the generic fibers, α is an isomorphism and β coincides with φ_K. In particular, the pairing $[\,\cdot\,,\cdot\,]$ for X_K can be computed on $\widetilde{X}$, and the desired functoriality follows from Lemma 2.1.2 applied to β.

The pairing $\delta(\,\cdot\,,\cdot\,)$. At this point, we recall that for any proper smooth and geometrically connected scheme X_K over K, there exists a nonzero integer a and a map $X_K \to A'_K$ from X_K to its Albanese variety, with the property that for any divisor $D_K \in \mathrm{Div}^\tau(X_K)$, there exists a divisor $W_K \in \mathrm{Div}^0(A'_K)$ such that $h^* W_K$ is well-defined and linearly equivalent to $a D_K$ [Néron 1965, II 2.1]. Let $c_K \in Z_0^0(X_K)$ and $D_K \in \mathrm{Div}^\tau(X_K)$ with disjoint supports. Keep the previous notation. After *moving* W_K *on the projective smooth scheme* A'_K if necessary (see [Liu 2002, 9.1.11], for example), we can assume that the support of $h^* W_K$ does not meet that of c_K. Then, using the functoriality of $[\,\cdot\,,\cdot\,]$, we can write

$$a[c_K, D_K] = [c_K, h^* W_K] + [c_K, \mathrm{div}_{X_K} f] = [h_* c_K, W_K] + [c_K, \mathrm{div}_{X_K} f]$$

for some nonzero $f \in K(X_K)$. By definition, Néron's pairing has the same functoriality property as $[\,\cdot\,,\cdot\,]$. And we have seen that both pairings coincide for principal divisors. Consequently, as A'_K is projective smooth geometrically connected over K, Theorem 2.2.1 is proved if we know that both pairings coincide on such schemes. So, until the end of the proof, we will only consider the pairings for projective smooth geometrically connected schemes. Furthermore, by $\mathbb{Z}$-linearity, we can only consider divisors which are algebraically equivalent to zero.

Now, both $[\,\cdot\,,\cdot\,]$ and $\langle\,\cdot\,,\cdot\,\rangle$ are bilinear in their definition domain, and they coincide for principal divisors. Using a *moving lemma on the projective smooth scheme* X_K, we see that

$$\delta(c_K, D_K) := \langle c_K, D_K\rangle - [c_K, D_K]$$

is well-defined on the *whole* product $Z_0^0(X_K) \times \mathrm{Div}^0(X_K)$. And conditions (1) and (2) of Lemma 2.2.3 are satisfied by δ.

Condition (3) *of Lemma 2.2.3 is satisfied by* $\delta(\cdot\,,\cdot)$. Denote by $\overline{R}$ the valuation ring of v in $\overline{K}$.

Fix $D_K \in \mathrm{Div}^0(X_K)$. Let $(n, \Delta) \in (\mathbb{Z}\setminus\{0\}) \times \mathrm{Div}^\tau(X_K)$ satisfying $\Delta_K = nD_K$. Represent the divisor Δ by a family $(U_t, g_t)_{t=1,\dots,m}$, where the U_t are affine open subsets of X and the g_t are rational functions on X. Let E_t be the set of $\overline{K}$-points of X_K which extend to $\overline{R}$-points of U_t. As X is proper over R, we see that $X(\overline{K}) = \cup_{t=1}^m E_t$. The family $(U_{t,K}, g_t)_{t=1,\dots,m}$ represents the divisor nD_K on X_K. Let us choose a Néron function λ_{nD_K} on X_K. By definition, we can find some v-continuous locally bounded functions $\alpha_t : U_{t,K}(\overline{K}) \to \mathbb{R}$ such that

$$\lambda_{nD_K}(z) = v(g_t(z)) + \alpha_t(z)$$

for all $z \in (U_{t,K} - \mathrm{Supp}(D_K))(\overline{K})$. As E_t is bounded in $U_t(\overline{K})$ (by construction), the function α_t is bounded on E_t.

Let $c_K = \sum_i n_i[x_{K,i}] \in Z_0^0(X_K)$ whose support does not meet that of D_K. Fix an i, let Z_i be the schematic closure of $x_{K,i}$ in X, set $x_{k,i} := X_k \cap Z_i$ and let t_i be such that $Z_i \subset U_{t_i}$. The same local computation as in the case of a principal divisor shows that

$$(\bar{c}_K.\Delta)_{x_{k,i}} = [K(x_{K,i}) : K]v(g_{t_i}(x_{K,i})) = \sum_{j_i=1}^{s_i} l_i v(g_{t_i}(x_{K,i})).$$

On the other hand, keeping the same notation as in the beginning of the proof,

$$\langle c_K, nD_K\rangle = \sum_i n_i \sum_{j_i=1}^{s_i} l_i \lambda_{nD_K}(x_{\overline{K},j_i}).$$

Consequently,

$$n\delta(c_K, D_K) = \sum_i n_i \sum_{j_i=1}^{s_i} l_i \alpha_{t_i}(x_{\overline{K},j_i}).$$

By construction, the $\overline{K}$-point $x_{\overline{K},j_i}$ of X_K belongs to E_{t_i}. Denoting by $|\cdot|$ the usual absolute value on $\mathbb{R}$, and setting

$$B := \max_{t=1,\dots,m} (\sup_{E_t} |\alpha_t|) \in \mathbb{R},$$

we obtain

$$|\delta(c_K, D_K)| \le \frac{1}{|n|} \sum_i |n_i|[K(x_{K,i}) : K]B = \frac{2B}{|n|} \deg^+ c_K.$$

As the divisor D_K is fixed, the numbers n and B are fixed, and so the right-hand side of the above inequality is bounded if $\deg^+ c_K$ is. □

Let us note the following properties of the pairing $[\cdot\,,\cdot]$, and consequently of Néron's pairing.

Proposition 2.2.5. *Let X_K be a proper, geometrically normal and geometrically connected scheme over K. Let $c_K \in Z_0^0(X_K)$ and let $D_K \in \mathrm{Div}^\tau(X_K)$ with disjoint supports. If c_K or D_K is rationally equivalent to zero, then $[c_K, D_K] \in \mathbb{Z}$.*

Proof. The case where D_K is rationally equivalent to zero follows directly from the definition of $[\,\cdot\,,\cdot\,]$: if $D_K = \mathrm{div}_K f$ with $f \in K(X_K)\setminus\{0\}$, then $[c_K, \mathrm{div}_K f] = (c_K.\mathrm{div} f) \in \mathbb{Z}$.

Let us now suppose that c_K is rationally equivalent to zero. As $[, D_K]$ is $\mathbb{Z}$-linear, we have to show that if $c_K = (\varphi_K)_*\mathrm{div}_{C_K} f$ for some K-morphism

$$\varphi_K : C_K \to X_K$$

from a proper normal connected curve C_K to X_K, and some nonzero $f \in K(C_K)$, then

$$[c_K, D_K] \in \mathbb{Z}\,.$$

As R is excellent, there exists a proper flat *regular* model C/R of C_K. On the other hand, let us consider a proper flat normal semifactorial model X/R of X_K. After replacing C by a desingularization of the graph of the rational map $C \dashrightarrow X$ induced by φ_K, we can suppose that φ_K extends to an R-morphism $\varphi : C \to X$. If Δ is a divisor on X which is τ-equivalent to zero and such that $\Delta_K = nD_K$ for some integer $n \neq 0$, then

$$[c_K, D_K] := \frac{1}{n}\big(\overline{(\varphi_K)_*\mathrm{div}_{C_K} f}.\Delta\big) = \frac{1}{n}\big(\overline{\mathrm{div}_{C_K} f}.\varphi^*\Delta\big)$$

by the projection formula (Lemma 2.1.2). Let us write

$$\mathrm{div}_C f = \overline{\mathrm{div}_{C_K} f} - V \quad \text{and} \quad \varphi^*\Delta = \overline{(\varphi_K)^*\Delta_K} - W$$

for some vertical divisors V and W on C/R. Denote by $\Gamma_1, \dots, \Gamma_\nu$ the reduced irreducible components of C_k, by M the intersection matrix associated to C_k (as defined in the introduction), and by $\rho : \mathrm{Pic}(C) \to \mathbb{Z}^\nu$ the degree homomorphism $(E) \mapsto (E \cdot \Gamma_i)_{i=1,\dots,\nu}$. Following [Bosch et al. 1990, 9.2/13], the divisor E on the *R-curve* C is algebraically equivalent to zero if and only if (E) belongs to the kernel of ρ. Therefore the τ-equivalence relation and the algebraic equivalence relation on C/R are the same, and the linear equivalence classes of $\varphi^*\Delta$ and $\mathrm{div}_C f$ belongs to the kernel of ρ. Thus we get:

$$\rho\big(\overline{\mathrm{div}_{C_K} f}\big) = \rho(V) = MV \quad \text{and} \quad \rho\big(\overline{(\varphi_K)^*\Delta_K}\big) = \rho(W) = MW,$$

where we have identified a vertical divisor on C/R with an element of $\mathbb{Z}^\nu$. Next, we use that the matrix M is *symmetric* to obtain

$$\big(\overline{\mathrm{div}_{C_K} f}.W\big) = {}^tW\rho\big(\overline{\mathrm{div}_{C_K} f}\big) = {}^tWMV = {}^tVMW = \big(\overline{(\varphi_K)^*\Delta_K}.V\big).$$

Then it follows that

$$[c_K, D_K] = \frac{1}{n}\big(\overline{(\varphi_K)^*\Delta_K}.\mathrm{div}_C f\big) = \big(\overline{(\varphi_K)^*D_K}.\mathrm{div}_C f\big) \in \mathbb{Z}. \qquad \square$$

Remark 2.2.6. Let us keep the notation of the proof of 2.2.5. If the curve C_K is *geometrically* normal and *geometrically* connected, the pairing $[\,\cdot\,,\cdot\,]$ is defined on C_K and

$$\big(\overline{(\varphi_K)^*D_K}.\mathrm{div}_C f\big) = [(\varphi_K)^*D_K, \mathrm{div}_{C_K} f].$$

In other words, in this case, the proof consists in using the functoriality of the pairing $[\,\cdot\,,\cdot\,]$, then showing that it is *symmetric* for curves, and finally applying the definition of the pairing for a principal divisor. The symmetry property of Néron's pairing $\langle\,\cdot\,,\cdot\,\rangle$ for such a curve is well-known: for example see [Lang 1983, 11.3.6 and 11.3.7]. But here, there is no reason for the curve C_K coming from the rational equivalence relation to satisfy the above *geometric* hypotheses. So we could not use directly the properties of the pairing $\langle\,\cdot\,,\cdot\,\rangle$. However, over an excellent discrete valuation ring, there is no need of these geometric hypotheses on C_K for the existence of the regular model C/R. So we have been able to prove the proposition for the pairing $[\,\cdot\,,\cdot\,]$, and thus also for Néron's pairing $\langle\,\cdot\,,\cdot\,\rangle$ thanks to Theorem 2.2.1.

3. Duality and algebraic equivalence for models of abelian varieties

3.1. ***Grothendieck's duality for Néron models.*** Let us recall here Grothendieck's duality theory for Néron models of abelian varieties, as developed in [SGA 7 I 1972, VII, VIII, IX].

Let R be a discrete valuation ring with perfect residue field k and fraction field K. Let A_K be an abelian variety over K, with dual A'_K. Let A/R, A'/R be the Néron models of A_K, A'_K, and Φ_A, $\Phi_{A'}$ be the étale k-group schemes of connected components of the special fibers A_k, A'_k.

By definition, the abelian variety A'_K represents the identity component $\mathrm{Pic}^0_{A_K/K}$ of the Picard functor of A_K, and the canonical isomorphism $A'_K = \mathrm{Pic}^0_{A_K/K}$ is given by the Poincaré sheaf $\mathcal{P}_K$ on $A_K \times_K A'_K$ birigidified along the unit sections of A_K and A'_K. Now, this sheaf is canonically endowed with the structure of a *biextension of* (A_K, A'_K) *by* $\mathbb{G}_{m,K}$ [loc. cit., VII 2.9.5]. Then the duality theory for Néron models is to understand how this biextension *extends* at the level of Néron models. For this, Grothendieck attached to $\mathcal{P}_K$ a canonical pairing

$$\langle\,\cdot\,,\cdot\,\rangle : \Phi_A \times_k \Phi_{A'} \to \mathbb{Q}/\mathbb{Z},$$

which measures the obstruction to extending $\mathcal{P}_K$ as a biextension of (A, A') by $\mathbb{G}_{m,R}$. The duality statement is: *this pairing is a perfect duality* [loc. cit., IX 1.3]. As mentioned in the introduction, it has been proved in various situations, including the semistable case [SGA 7 I 1972, IX 11.4; Werner 1997] and the mixed characteristic case [Bégueri 1980]. In general, the duality statement remains a conjecture.

3.2. ***Duality and Picard functor.*** Keep the notation of the previous subsection. By [Pepin 2011, Corollary 2.23], it is always possible to find an *R-compactification of A*, that is, an open R-immersion of A into a proper R-scheme $\overline{A}$ with dense image, such that $\overline{A}/R$ is flat, $\overline{A}$ is normal and the canonical map $\mathrm{Pic}(\overline{A}) \to \mathrm{Pic}(A)$ is surjective. Note that, in particular, $\overline{A}/R$ is semifactorial: the map $\mathrm{Pic}(A) \to \mathrm{Pic}(A_K)$ is surjective because A is regular, so that $\mathrm{Pic}(\overline{A}) \to \mathrm{Pic}(A_K)$ is surjective by composition. As $\overline{A}/R$ is proper, it makes sense to consider the notion of algebraic equivalence on $\overline{A}$ relative to R using the identity component of the Picard functor $\mathrm{Pic}_{\overline{A}/R}$, as defined in Section 2.1. Our goal in this section is to understand the duality from the point of view of algebraic equivalence, starting from the canonical isomorphism $A'_K = \mathrm{Pic}^0_{A_K/K}$. To do this, we need the following notions.

$\mathbb{Q}$-divisors and relative τ-equivalence. Let Z be a *normal* locally noetherian scheme, so that the canonical homomorphism from the group of divisors on Z into that of 1-codimensional cycles is *injective* [EGA IV$_4$ 1967, 21.6.9(i)]. A 1-codimensional cycle C on Z is said to be a *$\mathbb{Q}$-divisor* if there exists $n \in \mathbb{Z}\setminus\{0\}$ such that nC is a divisor.

Let $Z \to T$ be a proper morphism of schemes, with Z locally noetherian and normal. A $\mathbb{Q}$-divisor C on Z is said to be *τ-equivalent to zero relative to T* (or *τ-equivalent to zero* if there is no ambiguity on the base scheme T) if there exists $n \in \mathbb{Z}\setminus\{0\}$ such that nC is a divisor on Z which is τ-equivalent to zero relative to T (see Section 2.1). The group of classes of $\mathbb{Q}$-divisors on Z which are τ-equivalent to zero relative to T, modulo the principal divisors, will be denoted by $\mathrm{Pic}^{\mathbb{Q},\tau}(Z/T)$.

When $Z = \overline{A}$, the restriction to the generic fiber induces an injective morphism

$$\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R) \hookrightarrow \mathrm{Pic}^{\tau}_{A_K/K}(K) = \mathrm{Pic}^{0}_{A_K/K}(K) = A'_K(K).$$

The fact that $\mathrm{Pic}^{\tau}_{A_K/K}(K) = \mathrm{Pic}^{0}_{A_K/K}(K)$ can be found in [Mumford 1974, (v) p. 75]. To see that the above morphism is injective, let (C) be in its kernel. After modifying C by a principal divisor if necessary, we can assume that $C_K = 0$, that is, the support of C is contained in the special fiber $\overline{A}_k$ of $\overline{A}/R$. Let n be a nonzero integer such that nC is a divisor on $\overline{A}$ which is τ-equivalent to zero relative to R. As $\overline{A}_k$ admits at least one irreducible component Γ with multiplicity 1 (the component containing the unit element of A_k), the vertical divisor nC is principal [Raynaud 1970, 6.4.1 3]. In other words, there exists an integer m such that $nC = m\,\mathrm{div}(\pi)$, where π is a uniformizing element of R. Taking the associated cycles, and comparing the coefficients of Γ, we obtain that n divides m. Consequently, the $\mathbb{Q}$-divisor C is a principal divisor, whence the injectivity.

By definition, the group $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ contains the group $\mathrm{Pic}^0(\overline{A}/R)$ of classes of divisors on $\overline{A}$ which are algebraically equivalent to zero relative to R, modulo principal divisors. Now, when R is complete with algebraically closed residue field, we know from [Pepin 2011, Corollary 3.14] that the image of the composition

$$\mathrm{Pic}^0(\overline{A}/R) \hookrightarrow \mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R) \hookrightarrow A'_K(K)$$

contains the subgroup $(A')^0(R)$ of $A'_K(K)$.

Conversely, we will show that Grothendieck's duality statement for A and A' is equivalent to the following assertion: *the image of* $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R) \hookrightarrow A'_K(K)$ *is contained in the subgroup* $(A')^0(R)$.

Theorem 3.2.1. *Let R be a complete discrete valuation ring with algebraically closed residue field k and fraction field K. Let A_K be an abelian variety over K, with dual A'_K. Let A and A' be the Néron models of A_K and A'_K, respectively, over R. Let $\overline{A}$ be a proper flat normal model of A_K over R, equipped with a dense open R-immersion $A \to \overline{A}$, such that the induced map* $\mathrm{Pic}(\overline{A}) \to \mathrm{Pic}(A)$ *is surjective. Let* $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ *be the group of $\mathbb{Q}$-divisors on $\overline{A}$ which are τ-equivalent to zero relative to R, modulo the principal divisors. Then, the duality statement recalled in 3.1 is equivalent to the following*:

The image of the restriction map $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R) \hookrightarrow A'_K(K)$ *is contained in the subgroup* $(A')^0(R)$.

Let $\mathrm{Pic}^0(\overline{A}/R)$ *be the group of divisors on $\overline{A}$ which are algebraically equivalent to zero relative to R, modulo the principal divisors. Then, when the duality statement is true, the inclusion* $\mathrm{Pic}^0(\overline{A}/R) \hookrightarrow \mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ *is an equality, and there is a canonical commutative diagram*

$$\begin{tikzcd}
\mathrm{Pic}^0(A_K) \arrow[r, "\sim"] & A'_K(K) \\
\mathrm{Pic}^0(\overline{A}/R) \arrow[r, "\sim"] \arrow[u, hook] & (A')^0(R) \arrow[u, hook]
\end{tikzcd}$$

where the vertical maps are injective, and the horizontal maps are bijective.

See the end of Section 3.4 for the proof.

Remark 3.2.2. With the notation of Theorem 3.2.1, the canonical morphisms of abstract groups

$$\mathrm{Pic}^0(A_K) \to \mathrm{Pic}^0_{A_K/K}(K), \quad \mathrm{Pic}^0(\overline{A}/R) \to \mathrm{Pic}^0_{\overline{A}/R}(R)$$

are isomorphisms. For the second one, note that $\mathrm{Pic}_{\overline{A}/R}$ can be defined using the étale topology, and that R is strictly henselian. Note also that, when $\overline{A}$ is locally factorial (e.g., regular), the group $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ coincides with the group $\mathrm{Pic}^\tau(\overline{A}/R)$

of divisors on $\overline{A}$ which are τ-equivalent to zero relative to R, modulo the principal divisors, which in turn can be identified with the group $\mathrm{Pic}^{\tau}_{\overline{A}/R}(R)$.

The last assertion of Theorem 3.2.1 provides a refinement of [Pepin 2011, Corollary 3.14] in the present case $X = \overline{A}$. Here, when Grothendieck's duality holds, we obtain a *necessary and sufficient condition* for an invertible sheaf which is algebraically equivalent to zero on A_K to extend into an invertible sheaf on $\overline{A}$ which is algebraically equivalent to zero relative to R: the corresponding point $a'_K \in A'_K(K)$ must extend in the identity component of A'. Thus, conjecturally, *the group* $(A')^0(R)$ *parametrizes the invertible sheaves on* $\overline{A}$ *which are algebraically equivalent to zero relative to* R.

To make the link between Grothendieck's duality for A and A', and algebraic equivalence on $\overline{A}$, we need some preparation about *nonrational* 0-cycles on A_K, especially those which are supported on *inseparable* points over K.

3.3. ***About nonrational* 0*-cycles on abelian varieties.*** Let K be a field, and denote by $\overline{K}$ its algebraic closure. Let A_K be an abelian variety over K. Let d be a positive integer and let $\mathrm{Hilb}^d_{A_K/K}$ be the Hilbert scheme of points of degree d on A_K. The Grothendieck–Deligne norm map

$$\sigma_d : \mathrm{Hilb}^d_{A_K/K} \to A_K^{(d)}$$

defined in [SGA 4 III 1973, (6.3.4.1) on p. 435 = XVII-184] (see also [Bosch et al. 1990, pages 252–254]) maps $\mathrm{Hilb}^d_{A_K/K}$ to the d-fold *symmetric* product $A_K^{(d)}$. On the other hand, the map

$$A_K^d \to A_K, \quad (x_1, \dots, x_d) \mapsto x_1 + \cdots + x_d,$$

induces a map

$$m_d : A_K^{(d)} \to A_K.$$

Let us set

$$\mathcal{S}_d := m_d \circ \sigma_d : \mathrm{Hilb}^d_{A_K/K} \to A_K.$$

Let $a_K \in A_K$ be a closed point of degree d, that is to say, the residue field extension $K(a_K)/K$ has degree d. It corresponds to a rational point

$$h(a_K) \in \mathrm{Hilb}^d_{A_K/K}(K).$$

We will need an explicit description of its image $\mathcal{S}_d(h(a_K)) \in A_K(K)$, when considered as an element of $A_{\overline{K}}(\overline{K})$.

Let us consider the artinian $\overline{K}$-scheme $a_K \otimes_K \overline{K}$. It is supported on some $a_j \in A_{\overline{K}}(\overline{K})$, $j = 1, \dots, s$, where s is the separable degree of $K(a_K)/K$. The length of each local component of $a_K \otimes_K \overline{K}$ is equal to the inseparable degree of $K(a_K)/K$, and will be denoted by l. So the effective 0-cycle associated to $a_K \otimes_K \overline{K}$ is

$$\sum_{j=1}^{s} l[a_j] \in Z_0(A_{\overline{K}}).$$

We are going to show that

$$\mathcal{S}_d(h(a_K)) = \sum_{j=1}^{s} l a_j \in A_{\overline{K}}(\overline{K}).$$

Note that, in particular, this will show that the right-hand-side of the equality belongs to $A_K(K)$.

Lemma 3.3.1. *Let C be an artinian algebra over an algebraically closed field $\overline{K}$. Let $C_1, \ldots, C_s$ be the local components of C, with respective lengths $l_1, \ldots, l_s$, and let $u_j : C_j \to \overline{K}$ be the canonical surjection from C_j to its residue field. Then, for all*

$$c = (c_1, \ldots, c_s) \in C = C_1 \times \cdots \times C_s,$$

the following formula holds for the norm of c over $\overline{K}$:

$$N_{C/\overline{K}}(c) = \prod_{i=1}^{s} (u_j(c_j))^{l_j}.$$

Proof. We can assume that C is local, with length l. Let $\mathfrak{m}$ be the maximal ideal of C. Let n be the smallest integer such that $\mathfrak{m}^n = 0$. Choose a basis $\mathscr{E} = \mathscr{E}_0 \coprod \cdots \coprod \mathscr{E}_{n-1}$ of C over $\overline{K}$ which is adapted to the filtration

$$0 = \mathfrak{m}^n \subset \mathfrak{m}^{n-1} \subset \cdots \subset \mathfrak{m} \subset C,$$

i.e., $\mathscr{E}_i$ is contained in $\mathfrak{m}^i \setminus \mathfrak{m}^{i+1}$ and induces a basis of the $\overline{K}$-vector space $\mathfrak{m}^i / \mathfrak{m}^{i+1}$.

Fix $c \in C$ and let M be the matrix of multiplication-by-c in the basis $\mathscr{E}$. Write $c = \lambda + \epsilon$ with $\lambda \in \overline{K}$ and $\epsilon \in \mathfrak{m}$. Then M is a $l \times l$ lower triangular matrix, with all diagonal entries equal to λ. Hence $N_{C/\overline{K}}(c) = \lambda^l$, as required. □

We use the lemma to compute $\sigma_d(h(a_K))$, considered as an element of $A^{(d)}_{\overline{K}}(\overline{K})$. Let C be the $\overline{K}$-algebra of global sections of the scheme $a_K \otimes_K \overline{K}$. Set

$$\mathrm{TS}^d_{\overline{K}}(C) := (C^{\otimes d})^{\mathfrak{S}_d} \subseteq C^{\otimes d}$$

where $\mathfrak{S}_d$ is the symmetric group acting on $C^{\otimes d}$ by permuting factors. By definition, the point $\sigma_d(h(a_K)) \in (a_K \otimes_K \overline{K})^{(d)}(\overline{K}) \subset A^{(d)}_{\overline{K}}(\overline{K})$ corresponds to the unique $\overline{K}$-algebra homomorphism

$$\mathrm{TS}^d_{\overline{K}}(C) \to \overline{K}, \quad c^{\otimes d} \mapsto N_{C/\overline{K}}(c).$$

Now, from Lemma 3.3.1, this homomorphism is induced by the point

$$\big(a_1, \dots, a_1, a_2, \dots, a_2, \dots, a_s, \dots, a_s\big) \in A^d_{\overline{K}}(\overline{K}),$$

where a_j is repeated l times.

Next, the element $\mathcal{S}_d(h(a_K)) \in A_{\overline{K}}(\overline{K})$ is just the sum

$$m_d(\sigma_d(h(a_K))) = \sum_{j=1}^{s} la_j \in A_{\overline{K}}(\overline{K}),$$

as claimed.

Notation 3.3.2. The above K-morphisms $\mathcal{S}_d$ induce a homomorphism

$$\mathcal{S} : Z_0(A_K) \to A_K(K)$$

from the group of 0-cycles on A_K to that of K-rational points: if $a_K \in A_K$ is a closed point of degree d, defining $h(a_K) \in \mathrm{Hilb}^d_{A_K/K}(K)$, then $\mathcal{S}([a_K]) := \mathcal{S}_d(h(a_K))$.

We will also need to "translate divisors on A_K by nonrational points".

Let $\mathrm{Div}_{A_K/K}$ be the scheme of relative effective divisors on A_K [FGA VI 1966, 4.1]. Fix a positive integer d and consider the map

$$A^d_K \times_K \mathrm{Div}_{A_K/K} \to \mathrm{Div}_{A_K/K}$$

which is given by the functorial formula

$$((a_1, \dots, a_d), D) \mapsto D_{a_1} + \cdots + D_{a_d},$$

where D_a is obtained from D by translation by the section a. By symmetry, it induces a map

$$A^{(d)}_K \times_K \mathrm{Div}_{A_K/K} \to \mathrm{Div}_{A_K/K}\,.$$

By composing with the norm map σ_d, the latter gives rise to a map

$$\mathrm{Hilb}^d_{A_K/K} \times_K \mathrm{Div}_{A_K/K} \to \mathrm{Div}_{A_K/K}\,.$$

Let $a_K \in A_K$ be a closed point of degree d and let D_K be an effective divisor on A_K. Denote by $(D_K)_{a_K} \in \mathrm{Div}_{A_K/K}(K)$ the image of $(h(a_K), D_K)$ by the previous arrow. As above, write

$$\sum_{r=1}^{d} [a_{\overline{K},r}]$$

for the 0-cycle associated to $a_K \otimes_K \overline{K}$. In this expression, repetitions are allowed. Then, using the above computation of $\sigma_d(h(a_K))$, we see that $(D_K)_{a_K}$, as an element of the group $\mathrm{Div}_{A_{\overline{K}}/\overline{K}}(\overline{K})$, is equal to

$$\sum_{r=1}^{d}(D_{\overline{K}})_{a_{\overline{K},r}},$$

where $D_{\overline{K}}$ denotes the pull-back of D_K on $A_{\overline{K}}$. When a_K is étale over K, it is easy to see that the latter divisor descends on A_K. But this turns out to be true in general because of the above construction. Moreover, this description shows that the formation of $(D_K)_{a_K}$ is additive in D_K. We can thus associate a divisor $(D_K)_{a_K}$ on A_K to *any* divisor D_K in the following way: identifying divisors on A_K with 1-codimensional cycles, first use the above to define $(D_K)_{a_K}$ when D_K is a prime cycle, and then extend by $\mathbb{Z}$-linearity.

Notation 3.3.3. If c_K is a 0-cycle on A_K and D_K a divisor on A_K, define the divisor $(D_K)_{c_K}$ on A_K by $\mathbb{Z}$-linearity from the above situation where c_K is a closed point.

3.4. ***Relative algebraic equivalence on semifactorial compactifications.*** Our goal in this subsection is to prove Theorem 3.2.1. So, until the end of the subsection, *we fix a complete discrete valuation ring R with algebraically closed residue field k and fraction field K.*

The starting point is the link between Grothendieck's pairing and Néron's pairing, which has been established by Bosch and Lorenzini: Grothendieck's pairing is the *specialization* of Néron's pairing.

Theorem 3.4.1 [Bosch and Lorenzini 2002, 4.4]. *Keep the notation of Theorem 3.2.1. Moreover, let Φ_A and $\Phi_{A'}$ be the groups of connected components of A_k and A'_k, respectively. On the one hand, consider Grothendieck's pairing [SGA 7 I 1972, IX 1.3]*

$$\langle\cdot,\cdot\rangle : \Phi_A \times \Phi_{A'} \to \mathbb{Q}/\mathbb{Z},$$

and on the other hand, consider Néron's pairing [Néron 1965, II 9.3]

$$\langle\cdot,\cdot\rangle : Z_0^0(A_K) \times \mathrm{Div}^0(A_K) \to \mathbb{Q}$$

(*defined for (c_K, D_K) when the supports of c_K and D_K are disjoint*).

Let $(a, a') \in \Phi_A \times \Phi_{A'}$. Fix a point $a_K \in A_K(K)$ specializing to a, and a divisor $D'_K \in \mathrm{Div}^0(A_K)$ whose image in $A'_K(K)$ specializes to a'. Assume that a_K and 0_K do not belong to the support of D'_K. Then

$$\langle a, a'\rangle = -\langle\, [a_K] - [0_K]\,,\, D'_K\rangle \quad \mathrm{mod}\,\mathbb{Z}\,.$$

The following is a key result about the pairing $[\cdot,\cdot]$ defined in Proposition 2.1.3.

Proposition 3.4.2. *Let X_K be a proper geometrically normal and geometrically connected scheme over K. Let X be a proper flat normal semifactorial model of X_K over R. Let ν be the number of irreducible components of the special fiber X_k. There exist some 0-cycles of degree zero $c_{K,1}, \dots, c_{K,\nu}$ on X_K, with the following property*:

If D_K is a divisor on X_K which is τ-equivalent to zero, whose support is disjoint from those of the $c_{K,i}$, and if $[c_{K,i}, D_K]$ is an integer for all $i = 1, \dots, \nu$, then there exists a $\mathbb{Q}$-divisor on X which is τ-equivalent to zero relative R, with generic fiber D_K.

Proof. Let U be the open subset of X consisting of the regular points. As X is normal, for any irreducible closed subset C of codimension 1 in X, the intersection $C \cap U$ is a dense open subset of C. Furthermore, for any 1-codimensional cycle C on X, the restriction $C|_U$ is a *divisor* on U.

Next, let $\Gamma_1, \dots, \Gamma_\nu$ be the reduced irreducible components of X_k. Let $\xi_1, \dots, \xi_\nu$ be the generic points of $\Gamma_1, \dots, \Gamma_\nu$. Set $d_i := \mathrm{length}(\mathcal{O}_{X_k,\xi_i})$. From [Raynaud 1970, 7.1.2], there exists, for all $i = 1, \dots, \nu$, an R-immersion $u_i : Z_i \to U$, with Z_i finite and flat over R, with rank d_i, such that $u_{i,k}(Z_{i,k})$ is a point $x_{i,k}$ of Γ_i. Then the intersection multiplicity of Z_i and $\Gamma_j \cap U$ is equal to 1 if $i = j$, and 0 otherwise. In particular, the generic fiber of Z_i is a closed point $x_{K,i} \in U_K$ of degree d_i. Moreover, as Z_i is proper over R, the immersion $Z_i \to X$ is closed. Finally, setting $d := \gcd(d_i, i = 1, \dots, \nu)$, an appropriate $\mathbb{Z}$-linear combination of the $x_{K,i}$ provides a 0-cycle c_K on X_K of degree d. We set

$$c_{K,i} := [x_{K,i}] - \frac{d_i}{d} c_K \in Z_0^0(X_K).$$

Let $D_K \in \mathrm{Div}^\tau(X_K)$ whose support is disjoint from those of the $c_{K,i}$. Choose $\Delta \in \mathrm{Div}^\tau(X/R)$ with a nonzero integer n such that $\Delta_K = n D_K$. Denoting by $\overline{D}_K$ the schematic closure of D_K in X, we can view Δ as a 1-codimensional cycle on X, and write

$$\Delta = n\overline{D}_K + \sum_{i=1}^{\nu} n_i \Gamma_i$$

for some integers $n_1, \dots, n_\nu$. Set $V := \sum_{i=1}^{\nu} n_i \Gamma_i$. As the schematic closures $\bar{c}_{K,i}$ of the $c_{K,i}$ in X are contained in U (by construction), the following computation is valid:

$$\bar{c}_{K,i}.\Delta = n(\bar{c}_{K,i}.\overline{D}_K) + (\bar{x}_{K,i}.V) - \frac{d_i}{d}(\bar{c}_K.V) = n(\bar{c}_{K,i}.\overline{D}_K) + n_i - \frac{d_i}{d}(\bar{c}_K.V).$$

Assume that $[c_{K,i}, D_K]$ belongs to $\mathbb{Z}$. Then, the left-hand side of the above equality belongs to $n\,\mathbb{Z}$. Consequently, there exists $r_i \in \mathbb{Z}$ such that

$$nr_i = n_i - \frac{d_i}{d}(\bar{c}_K.V).$$

Now, consider the vertical cycle (with integral coefficients)

$$W := (\bar{c}_K.V)\frac{1}{d}[X_k].$$

By definition,

$$V - W = n\sum_{i=1}^{\nu} r_i\Gamma_i, \quad \text{that is,} \quad \Delta - W = n(\overline{D}_K - \sum_{i=1}^{\nu} r_i\Gamma_i).$$

The cycle $D := \overline{D}_K - \sum_{i=1}^{\nu} r_i\Gamma_i$ is equal to D_K on the generic fiber. This is a $\mathbb{Q}$-divisor on X which is τ-equivalent to zero because dnD is a divisor on X which is τ-equivalent to zero. □

Keep the notation of Proposition 3.4.2. Even if X/R admits a section, so that d is equal to 1, the closed point $x_{K,i}$ is *not* rational as soon as the special fiber X_k is not reduced at the generic point of the irreducible component Γ_i. Therefore, if we want to combine Theorem 3.4.1 and Proposition 3.4.2 when $X = \overline{A}$ (notation of Theorem 3.2.1), we need to compare the values of Néron's pairing on the *abelian variety* A_K for 0-cycles which are supported on nonrational points, with its values for 0-cycles of the form $[a_K] - [0_K]$, with $a_K \in A_K(K)$. Here we will use the constructions of Section 3.3, together with some biduality argument. To take care of the conditions on supports involved in the computations of Néron's pairings, let us first note the following lemma.

Lemma 3.4.3. *Let A_K be an abelian variety over K with dual A'_K. Let $a'_K \in A'_K(K)$ and let $\mathscr{E}$ be a finite set of closed points of A_K. Then there exists a Poincaré divisor on $A_K \times_K A'_K$, that is, a divisor such that the invertible sheaf $\mathcal{O}_{A_K \times_K A'_K}(P)$ is a Poincaré sheaf which is birigidified along $0_K \in A_K(K)$ and $0'_K \in A'_K(K)$, satisfying the following conditions*:

(1) *$P_{0_K} := P|_{0_K \times_K A'_K}$ and $P_{0'_K} := P|_{A_K \times_K 0'_K}$ are well-defined and equal to zero.*

(2) *$P_{a'_K} := P|_{A_K \times_K a'_K}$ is well-defined, and its support does not meet $\mathscr{E}$.*

(3) *For all $a_K \in \mathscr{E}$, $P_{a_K} := P|_{a_K \times_K A'_K}$ is well-defined, and its support does not meet $\{0'_K, a'_K\}$.*

Proof. Consider the finite set $\mathscr{F}$ whose elements are the following closed points of the product $A_K \times_K A'_K$:

$$a_K \times_K 0'_K \text{ or } a_K \times_K a'_K, \text{ with } a_K \in (\{0_K\} \coprod \mathscr{E}).$$

Let $\mathscr{P}$ be a Poincaré sheaf on $A_K \times_K A'_K$, birigidified along $0_K \in A_K(K)$ and $0'_K \in A'_K(K)$. Choose an arbitrary divisor Q such that $\mathscr{O}_{A_K \times_K A'_K}(Q) \simeq \mathscr{P}$. Using a moving lemma on the product $A_K \times_K A'_K$ if necessary [Liu 2002, 9.1.11], one can assume that the support of Q is disjoint from the finite set $\mathscr{F}$. As $0_K \times_K 0'_K \in \mathscr{F}$, the divisors $Q|_{0_K \times_K A'_K}$ and $Q|_{A_K \times_K 0'_K}$ are well-defined, and are principal. Then

$$P := Q - p_2^*(Q|_{0_K \times_K A'_K}) - p_1^*(Q|_{A_K \times_K 0'_K})$$

(where $p_1 : A_K \times_K A'_K \to A_K$ and $p_2 : A_K \times_K A'_K \to A'_K$ are the projections) is a Poincaré divisor again.

Now, let $a_K \in (\{0_K\} \coprod \mathscr{E})$. Then $a_K \times_K a'_K$ does not belong to the support $\mathrm{Supp}(Q)$ of Q because $a_K \times_K a'_K \in \mathscr{F}$. Next, $a_K \times_K a'_K \notin \mathrm{Supp}(p_2^*(Q|_{0_K \times_K A'_K}))$: indeed, $0_K \times_K a'_K \in \mathscr{F}$ by definition, hence $0_K \times_K a'_K \notin \mathrm{Supp}(Q)$, and consequently $a'_K \notin \mathrm{Supp}(Q|_{0_K \times_K A'_K})$. Finally $a_K \times_K a'_K \notin \mathrm{Supp}(p_1^*(Q|_{A_K \times_K 0'_K}))$, because otherwise $a_K \in \mathrm{Supp}(Q|_{A_K \times_K 0'_K})$ and $a_K \times_K 0'_K \in \mathrm{Supp}(Q)$, which is not the case because $a_K \times_K 0'_K \in \mathscr{F}$. We have thus shown that the point $a_K \times_K a'_K$ does not belong to the support of P. Similarly, the point $a_K \times_K 0'_K$ does not belong to the support of P. In conclusion:

(1) $P|_{0_K \times_K A'_K}$ and $P|_{A_K \times_K 0'_K}$ are well-defined, and are equal to zero, by definition of P.

(2) $P|_{A_K \times_K a'_K}$ is well-defined, and its support does not meet $\mathscr{E}$, because $a_K \times_K a'_K \notin \mathrm{Supp}(P)$ for all $a_K \in \mathscr{E}$.

(3) $P|_{a_K \times_K A'_K}$ is well-defined for all $a_K \in \mathscr{E}$, and its support does not meet $\{0'_K, a'_K\}$, because $a_K \times_K a'_K \notin \mathrm{Supp}(P)$ and $a_K \times_K 0'_K \notin \mathrm{Supp}(P)$ for all $a_K \in \mathscr{E}$. $\square$

We can now proceed to the announced comparison of some values of Néron's pairing.

Proposition 3.4.4. *Let A_K be an abelian variety with dual A'_K. Let $c_K \in Z_0^0(A_K)$ and $D'_K \in \mathrm{Div}^0(A_K)$. Assume that the support of D'_K is disjoint from that of c_K and that of $[\mathscr{S}(c_K)] - [0_K]$ (Notation 3.3.2). Then the following relation between values of Néron's pairing on A_K is true*:

$$\langle c_K, D'_K \rangle \equiv \langle\, [\mathscr{S}(c_K)] - [0_K]\,, D'_K \rangle \quad \mod \mathbb{Z}\,.$$

Proof. Let $a'_K \in A'_K(K)$ corresponding to D'_K. Let $\mathscr{E}$ be a finite set of closed points of A_K, containing the supports of c_K and $[\mathscr{S}(c_K)] - [0_K]$. From Lemma 3.4.3, there exists a Poincaré divisor P satisfying the following conditions:

(1) $P_{0_K} := P|_{0_K \times_K A'_K}$ and $P_{0'_K} := P|_{A_K \times_K 0'_K}$ are well-defined and equal to zero.

(2) $P_{a'_K} := P|_{A_K \times_K a'_K}$ is well-defined, and its support does not meet $\mathscr{E}$.

(3) $P_{a_K} := P|_{a_K \times_K A'_K}$ is well-defined for all $a_K \in \mathscr{E}$, and its support does not meet $\{0'_K, a'_K\}$.

Then, the divisors D'_K and $P_{a'_K}$ are linearly equivalent. Consequently, we can assume $D'_K = P_{a'_K}$ (Proposition 2.2.5).

Write $c_K = c_K^+ - c_K^-$ where c_K^+ and c_K^- are positive 0-cycles with disjoint supports. Let L/K be a finite field extension such that

$$c_K^+ \otimes_K L = \sum_{r=1}^{d} [a_{r,+}] \quad \text{and} \quad c_K^- \otimes_K L = \sum_{r=1}^{d} [a_{r,-}]$$

where $d := \deg c_K^+ = \deg c_K^-$ and with $a_{r,+}$, $a_{r,-}$ in $A_L(L)$ (repetitions allowed). Computing Néron's pairings over K and over L with normalized valuations, we get

$$\langle c_K, P_{a'_K}\rangle_{A_K} = \frac{1}{e_L}\left\langle \sum_{r=1}^{d} [a_{r,+}] - \sum_{r=1}^{d} [a_{r,-}]\,, (P_L)_{a'_L}\right\rangle_{A_L},$$

where P_L is the pull-back of P over L, the point $a'_L \in A'_L(L)$ is the image of $a'_K \in A'_K(K)$ by the inclusion $A'_K(K) \subseteq A'_L(L)$, and e_L is the ramification index of L/K. As $(P_L)_{0'_L} = 0$, the *reciprocity law* for Néron's pairing [Lang 1983, 11.4.2][1] asserts that the right-hand side of the equality is equal to the (well-defined) quantity

$$\frac{1}{e_L}\left\langle [a'_L] - [0'_L]\,, \sum_{r=1}^{d}(P_L)_{a_{r,+}} - \sum_{r=1}^{d}(P_L)_{a_{r,-}}\right\rangle_{A'_L}.$$

Now, with Notation 3.3.3, the divisor $\sum_{r=1}^{d}(P_L)_{a_{r,+}} - \sum_{r=1}^{d}(P_L)_{a_{r,-}}$ is precisely the pull-back over L of the divisor P_{c_K} on A'_K. Furthermore, as the Poincaré map

$$A_L(L) \to \operatorname{Pic}^0_{A'_L/L}(L)$$

is a group homomorphism, the divisors P_{c_K} and $P_{S(c_K)}$ are linearly equivalent on A'_L, and thus on A'_K (because $\operatorname{Pic}^0_{A'_K/K}(K)$ is contained in $\operatorname{Pic}^0_{A'_K/K}(L)$). Let $f \in K(A'_K)$ be such that $P_{c_K} - P_{S(c_K)} = \operatorname{div}(f)$. As the normalized valuation on K takes values in $\mathbb{Z}$, the (well-defined) pairing

$$\frac{1}{e_L}\langle\, [a'_L] - [0'_L]\,, (\operatorname{div}(f))_L\rangle_{A'_L} = \langle\, [a'_K] - [0'_K]\,, \operatorname{div}(f)\rangle_{A'_K}$$

is an *integer*. Consequently,

$$\langle c_K\,, P_{a'_K}\rangle_{A_K} \equiv \langle\, [a'_K] - [0'_K]\,, P_{\mathcal{S}(c_K)}\rangle_{A'_K} \mod \mathbb{Z}\,.$$

As $P_{0_K} = 0$ and $P_{0'_K} = 0$, we conclude by using once again the reciprocity law. $\square$

[1] Here we use the reciprocity law in the case where the divisorial correspondence is the Poincaré divisor P_L. By using a definition of Néron's pairing relying on the Poincaré biextension (see [Zahrin 1972, §5; Mazur and Tate 1983, §2]), the reciprocity law for P_L is a direct consequence of the *biduality* of abelian varieties.

We can now interpret Grothendieck's obstruction (Section 3.1) in terms of relative algebraic equivalence.

Theorem 3.4.5. *Keep the notation of Theorem 3.2.1. Moreover, let Φ_A and $\Phi_{A'}$ be the group of connected components of A_k and A'_k, respectively.*

Let $a' \in \Phi_{A'}$. Lift a' to a point $a'_K \in A'_K(K)$, representing the linear equivalence class of a divisor D'_K on A_K. Then the homomorphism

$$\langle \cdot\,, a' \rangle : \Phi_A \to \mathbb{Q}/\mathbb{Z}$$

induced by Grothendieck's pairing is identically zero if and only if D'_K can be extended to a $\mathbb{Q}$-divisor on $\overline{A}$ which is τ-equivalent to zero relative to R.

Proof. Suppose that the obstruction $\langle \cdot\,, a' \rangle$ vanishes. Choose 0-cycles of degree zero $c_{K,1}, \dots, c_{K,\nu}$ on A_K satisfying the conclusion of Proposition 3.4.2 when applied to the model $\overline{A}/R$ of A_K. To prove that D'_K extends to a $\mathbb{Q}$-divisor on $\overline{A}$ which is τ-equivalent to zero, we can replace D'_K by any divisor on A_K which is linearly equivalent to D'_K. In particular, using moving lemma [Liu 2002, 9.1.11], we can assume that the support of D'_K does not meet the finite set

$$\{0_K, S(c_{K,1}), \dots, S(c_{K,\nu})\} \coprod_{i=1}^{\nu} \mathrm{Supp}(c_{K,i}).$$

Then, as $\langle \cdot\,, a' \rangle = 0$, we get from Bosch–Lorenzini's Theorem 3.4.1 that

$$\langle\, [\mathcal{S}(c_{K,i})] - [0_K]\,, D'_K \rangle \in \mathbb{Z}$$

for all $i = 1, \dots, \nu$. Proposition 3.4.4 and Theorem 2.2.1 then imply that

$$[c_{K,i}\,, D'_K] \in \mathbb{Z}$$

for all $i = 1, \dots, \nu$. Due to the choice of the $c_{K,i}$, the divisor D'_K can then be extended to a $\mathbb{Q}$-divisor on $\overline{A}$ which is τ-equivalent to zero.

Conversely, suppose that there is a $\mathbb{Q}$-divisor D' on $\overline{A}$ which is τ-equivalent to zero, with generic fiber D'_K. To prove that $\langle \cdot\,, a' \rangle = 0$, we can assume that 0_K does not belong to the support of D'_K, by adding to D' the divisor of a rational function on $\overline{A}$ if needed. Let n' be a nonzero integer such that $\Delta' := n'D'$ is a divisor on $\overline{A}$ which is τ-equivalent to zero. For each $a_K \in A_K(K)$ which is not in the support of D'_K, we get:

$$[\,[a_K] - [0_K]\,, D'_K] = \frac{1}{n'}\big([\overline{a_K}] - [\overline{0_K}].\Delta'\big) = \big([\overline{a_K}] - [\overline{0_K}].D'\big) \in \mathbb{Z}\,.$$

The first equality holds by definition of the pairing $[\,\cdot\,,\cdot\,]$, and the second one is true because $[\overline{a_K}] - [\overline{0_K}]$ is contained in the regular locus of $\overline{A}$. Now observe that an element $a \in \Phi_A$ can always be lifted to a point $a_K \in A_K(K)$ which is not in the support of D'_K. Thus, it follows from Theorem 2.2.1 and Bosch–Lorenzini's Theorem 3.4.1 that the obstruction $\langle\,\cdot\,, a'\rangle$ vanishes. □

Proof of Theorem 3.2.1. By biduality of abelian varieties, Grothendieck's duality statement is equivalent to the following: the obstruction $\langle\,\cdot\,, a'\rangle$ vanishes if and only if $a' = 0$.

Suppose that this assertion is true. Let $(C) \in \mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ and let a'_K be its canonical image in $A'_K(K)$. By Theorem 3.4.5, the obstruction $\langle\,\cdot\,, a'\rangle$ vanishes. Hence $a' = 0$, that is, $a_{K'} \in (A')^0(R)$.

Conversely, suppose that the canonical image of $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ in $A'_K(K)$ is contained in $(A')^0(R)$. Let $a' \in \Phi_{A'}$, and assume that the corresponding obstruction $\langle\,\cdot\,, a'\rangle$ vanishes. Choose a lifting $a'_K \in A'_K(K)$ of a'. Then, by Theorem 3.4.5, the point a'_K belongs to the image of $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$. In particular, it belongs to $(A')^0(R)$, and $a' = 0$.

Thus, we have proved that Grothendieck's conjecture is equivalent to the fact that the image of $\mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R)$ in $A'_K(K)$ is contained in $(A')^0(R)$. Now suppose that the conjecture is true. Then, from [Pepin 2011, Corollary 3.14], we obtain isomorphisms

$$\mathrm{Pic}^0(\overline{A}/R) \xrightarrow{\sim} \mathrm{Pic}^{\mathbb{Q},\tau}(\overline{A}/R) \xrightarrow{\sim} (A')^0(R).$$

The last assertion of Theorem 3.2.1 follows. □

4. Grothendieck's pairing for Jacobians

4.1. ***Statement of the results.*** Let R be a complete discrete valuation ring with algebraically closed residue field k and fraction field K. Let X_K be a proper smooth geometrically connected curve over K, and let $J_K := \mathrm{Pic}^0_{X_K/K}$ be its Jacobian. Denote by J and J' the Néron models of J_K and J'_K over R, respectively, and Φ_J and $\Phi_{J'}$ the groups of connected components of the special fiber of J/R and J'/R, respectively. Theorems 3.4.1 and 2.2.1 describe Grothendieck's pairing associated to J_K in terms of intersection multiplicities on some compactification $\overline{J}$ of J. It is natural to wonder if these computations can be replaced by intersection computations *on a proper flat regular model X of X_K*.

Assume that $X_K(K)$ is nonempty. In this case, the curve X_K can be *embedded into J_K*, and can be used to define a classical theta divisor on J_K. Then, using Theorem 3.4.1, Bosch and Lorenzini described Grothendieck's pairing associated to J_K in terms of the Néron pairing *on X_K*, and so in terms of intersection multiplicities on X, thanks to Gross's and Hriljac's Theorems [Gross 1986; Hriljac 1985]. Their precise result is as follows. Let M be the intersection matrix of the special fiber of X/R: if $\Gamma_1, \dots, \Gamma_\nu$ are the irreducible components of X_k equipped with their reduced scheme structure, the (i, j)-th entry of M is the intersection number $(\Gamma_i \cdot \Gamma_j)$. Denote by Φ_M the torsion part of the cokernel of $M : \mathbb{Z}^\nu \to \mathbb{Z}^\nu$. According to

Raynaud's work on the sheaf $\mathrm{Pic}_{X/S}$, there is a canonical isomorphism $\Phi_J = \Phi_M$; see [Bosch et al. 1990, 9.6/1]. Now, on the product $\Phi_M \times \Phi_M$, there is the canonical pairing

$$\langle\,\cdot\,,\,\cdot\,\rangle_M : \Phi_M \times \Phi_M \to \mathbb{Q}/\mathbb{Z}$$
$$(\overline{T}, \overline{T}') \mapsto ({}^tS/n)M(S'/n') \mod \mathbb{Z}$$

for any $n, n' \in \mathbb{Z}\setminus\{0\}$ and $S, S' \in \mathbb{Z}^\nu$ such that $MS = nT$, $MS' = n'T'$. Now let $(a, a') \in \Phi_J \times \Phi_{J'}$. By identifying J_K and J'_K with the help of the opposite of the canonical principal polarization defined by a theta divisor, Grothendieck's pairing of a and a' can be computed by the formula

$$\langle a, a'\rangle = \langle a, a'\rangle_M$$

[Bosch and Lorenzini 2002, Theorem 4.6].

Now assume that $X_K(K)$ is empty. Choosing a field extension L/K such that $X_K(L)$ is nonempty, one can consider a theta divisor on J_L, and it is a classical fact that the associated canonical principal polarization is defined over K. Using its opposite, one can still identify Φ_J with $\Phi_{J'}$, and thus $\Phi_{J'}$ with Φ_M (as k is algebraically closed, the identification $\Phi_J = \Phi_M$ holds without assuming that $X_K(K)$ is nonempty). Then the authors of [Bosch and Lorenzini 2002] ask if both pairings $\langle\,\cdot\,,\,\cdot\,\rangle$ and $\langle\,\cdot\,,\,\cdot\,\rangle_M$ still coincide in this situation [loc. cit., Remark 4.9]. In [Lorenzini 2008, Theorem 3.4], Lorenzini gives a positive answer to this question when the special fiber of X/R admits two irreducible components C_i and C_j with multiplicities d_i and d_j such that $(C_i \cdot C_j) > 0$ and $\gcd(d_i, d_j) = 1$. Here we show that this result still holds if we only assume that *the global* gcd *of the multiplicities of the irreducible components of X_k is equal to* 1. Note that, due to the hypotheses on R and on X, this global gcd coincides with the *index* of the curve X_K, that is, the smallest positive degree of a divisor on X_K [Raynaud 1970, 7.1.6 1].

Theorem 4.1.1. *Let R be a complete discrete valuation ring with algebraically closed residue field k and fraction field K. Let X_K be a proper smooth geometrically connected curve over K, with index d. Let J_K be the Jacobian of X_K, identified with its dual using the opposite of its canonical principal polarization. Let X/R be a proper flat regular model of X_K. The following relation between Grothendieck's pairing for J_K and the above pairing defined by the intersection matrix M of X_k is true*:

$$d\langle a, a'\rangle = d\langle a, a'\rangle_M.$$

In particular, we get the following partial answers to Grothendieck's conjecture [SGA 7 I 1972, IX 1.3] in this case:

Corollary 4.1.2. *Keep the notation of Theorem 4.1.1. Then*:

- *The kernel of Grothendieck's pairing for J_K is killed by d.*

- *If d is prime to the characteristic of k, then Grothendieck's pairing for J_K is perfect.*

Proof. From [Bosch and Lorenzini 2002, Theorem 1.3], the pairing $\langle\cdot,\cdot\rangle_M$ is a perfect duality. So the first point follows directly from Theorem 4.1.1. For the second point, denote by p the characteristic of k. Then Grothendieck's pairing is perfect when restricted to the prime-to-p parts of the component groups: [SGA 7 I 1972, IX 11.3; Bertapelle 2001, Theorem 1]. Consequently, the second point follows again from the perfectness of $\langle\cdot,\cdot\rangle_M$ and Theorem 4.1.1. □

4.2. ***Proof of Theorem 4.1.1.*** Here are two lemmas to prepare the proof of the theorem.

Recall that, as R is complete with algebraically closed residue field, a classical result of Lang asserts that the Brauer group of K is zero, whence $\mathrm{Pic}^0(X_K) = J_K(K)$.

Lemma 4.2.1. *Let $a, a' \in \Phi_J = \Phi_M$, and choose divisors D_K, D'_K on X_K with disjoint supports, such that $a_K := (D_K)$, $a'_K := (D'_K) \in J_K(K) = \mathrm{Pic}^0(X_K)$ specialize to a, a'. The relationship between the pairing $\langle\cdot,\cdot\rangle_M$ and Néron's pairing on X_K is given by*:

$$\langle a,\ a'\rangle_M = -\langle D_K,\ D'_K\rangle \mod \mathbb{Z}.$$

Proof. This is an immediate consequence of the definitions, and of the description of Néron's pairing for the curve X_K in terms of intersection multiplicities on X. Indeed, let $\rho : \mathrm{Pic}(X) \to \mathbb{Z}^\nu$ be the degree morphism $(Z) \mapsto (Z\cdot\Gamma_i)_{i=1,\dots,\nu}$. Denote by $\overline{D_K}$ the schematic closure of D_K in X. By definition of Raynaud's isomorphism $\Phi_J = \Phi_M$, the image of $\rho(\overline{D}_K) \in \mathbb{Z}^\nu$ in $\mathbb{Z}^\nu/\operatorname{Im} M$ is contained in the torsion part Φ_M, and the resulting element is precisely the image of $a \in \Phi_J$ under the isomorphism. In particular, there are $n, n' \in \mathbb{Z}\setminus\{0\}$ and $S, S' \in \mathbb{Z}^\nu$ such that $MS = n\rho(\overline{D}_K)$, $MS' = n'\rho(\overline{D}'_K)$, and by definition of the symmetric pairing $\langle\cdot,\cdot\rangle_M$, we get

$$\langle a,\ a'\rangle_M = ({}^tS'/n')\rho(\overline{D}_K) \mod \mathbb{Z}.$$

Under the identification $\bigoplus_{i=1}^{\nu} \mathbb{Z}\,\Gamma_i \simeq \mathbb{Z}^\nu$, the right-hand side can also be written as an intersection multiplicity:

$$\langle a,\ a'\rangle_M = \frac{1}{n'}(\overline{D}_K.S') = -\frac{1}{n'}\big(\overline{D}_K.(n'\overline{D}'_K - S')\big) \quad \in \mathbb{Q}/\mathbb{Z}.$$

Now, the equality $MS' = n'\rho(\overline{D}'_K)$ means that the divisor $n'\overline{D}'_K - S'$ on X is algebraically equivalent to zero relative to R ([Bosch et al. 1990] 9.2/13). Applying Theorem 2.2.1 to the curve X_K, we conclude that

$$\langle a,\ a'\rangle_M = -[D_K, D'_K] = -\langle D_K, D'_K\rangle \in \mathbb{Q}/\mathbb{Z}. \qquad \square$$

Next, the index d of X_K divides $g-1$ where g is the genus of X_K [Raynaud 1970, 9.5.1]. Let us fix a divisor E of degree d on X_K, and consider the linear equivalence class of divisors of degree $g-1$ given by

$$t_K := (g-1)d^{-1}(E) \in \mathrm{Pic}^{g-1}_{X_K/K}(K).$$

The canonical image of the $(g-1)$-fold symmetric product $X_K^{(g-1)}$ in $\mathrm{Pic}^{g-1}_{X_K/K}$ can be translated by t_K to a divisor on J_K, which we will denote by Θ. Then, by extending K and reducing to the case where $X_K(K)$ is nonempty, one sees that the canonical principal polarization φ of J_K can be written explicitly here as $\varphi(z) = -(\Theta_z - \Theta)$, where Θ_z is obtained from Θ by translation by the point z. On the other hand, denoting by Δ the diagonal of $X_K \times_K X_K$, the divisor $d\Delta - E \times_K X_K$ on $X_K \times_K X_K$ defines an element of $\mathrm{Pic}^0_{X_K/K}(X_K)$, hence a K-morphism $h : X_K \to \mathrm{Pic}^0_{X_K/K} = J_K$.

Lemma 4.2.2. *The following diagram of K-morphisms is commutative*:

$$\begin{tikzcd} & J'_K \arrow[dr, "h^*"] & \\ J_K \arrow[ur, "-\varphi"] \arrow[rr, "d"] & & J_K \end{tikzcd}$$

The commutativity can be stated as follows. Let $z \in J_K(\overline{K})$. Let Z be any divisor of degree 0 *on $X_{\overline{K}}$, whose linear equivalence class (Z) corresponds to z via the canonical isomorphism* $\mathrm{Pic}^0(X_{\overline{K}}) = J_K(\overline{K})$. *Then the following relation holds*:

$$h^*(\Theta_z - \Theta) = d(Z) \in \mathrm{Pic}^0(X_{\overline{K}}) = J_K(\overline{K}).$$

In particular, there is a nonempty open subset U_K of J_K such that $h^*\Theta_z$ is a well-defined divisor on X_K for all $z \in U_K(K)$, and whose degree does not depend on the point z.

Proof. To check that the diagram is commutative, one can replace K by its algebraic closure, and so we can assume that K is algebraically closed. As the pull-back by the multiplication-by-d on J_K acts as multiplication-by-d on the group $\mathrm{Pic}^0(J_K)$, the lemma then follows from the classical situation where X_K can be embedded into J_K using a rational point of X_K. □

Proof of Theorem 4.1.1. Let $(a, a') \in \Phi_J \times \Phi_J$. Choose a point $a_K \in J_K(K)$ which specializes to $a \in \Phi_J$. The point a_K corresponds, under the equality $J_K(K) = \mathrm{Pic}^0(X_K)$, to the linear equivalence class of a divisor $D(a)_K$ of degree 0 on X_K. Write $D(a)_K = D(a)_K^+ - D(a)_K^-$ with $D(a)_K^+$ and $D(a)_K^-$ positive with disjoint supports. Let L/K be a finite field extension such that

$$D(a)_K^+ \otimes_K L = \sum_{r=1}^{\alpha} [a_{r,+}] \quad \text{and} \quad D(a)_K^- \otimes_K L = \sum_{r=1}^{\alpha} [a_{r,-}],$$

where $\alpha := \deg D(a)_K^+ = \deg D(a)_K^-$ and with $a_{r,+}$, $a_{r,-}$ in $X_L(L)$ (repetitions allowed).

Next, still denoting by U_K the open subset of J_K provided by Lemma 4.2.2, one can find a'_K and z_K in $U_K(K)$ specializing respectively to a' and 0 in Φ_J, and such that

$$\begin{aligned} &da_K,\ 0_K \notin \operatorname{Supp}(\Theta_{a'_K} - \Theta_{z_K}) \\ &\bar{a}_{r,+},\ \bar{a}_{r,-} \notin \operatorname{Supp}((\Theta_{a'_K} - \Theta_{z_K})_L) \qquad \text{for all } r = 1, \dots, \alpha, \end{aligned}$$

where $\bar{a}_{r,+} := h(a_{r,+})$ and $\bar{a}_{r,-} := h(a_{r,-})$. The points a'_K and z_K correspond to the classes of some divisors $D(a')_K$ and $D(0)_K$ on X_K, under the identification $J_K(K) = \operatorname{Pic}^0(X_K)$. From Lemma 4.2.2, we get:

$$h^*(\Theta_{a'_K} - \Theta_{z_K}) = d(D(a')_K - D(0)_K) = d(a'_K - z_K)$$

in $\operatorname{Pic}^0(X_K) = J_K(K)$. And by construction, the K-point $d(a'_K - z_K)$ of J_K specializes to $da' \in \Phi_J$. As a consequence, Lemma 4.2.1 provides the formula:

$$\langle a, da' \rangle_M = -\langle D(a)_K, h^*(\Theta_{a'_K} - \Theta_{z_K}) \rangle_{X_K} \mod \mathbb{Z}$$

(note that $h^*(\Theta_{a'_K} - \Theta_{z_K})$ is a well-defined divisor, and not only a class, because $a'_K, z_K \in U_K(K)$).

Still working with normalized valuations to compute Néron's pairing, and using functoriality, we obtain:

$$\langle a, da' \rangle_M = -\frac{1}{e_L} \left\langle \sum_{r=1}^{\alpha} [\bar{a}_{r,+}] - [\bar{a}_{r,-}], (\Theta_{a'_K} - \Theta_{z_K})_L \right\rangle_{J_L} \mod \mathbb{Z},$$

where e_L is the ramification index of L/K. Then we apply the reciprocity law for Néron's pairing with the divisorial correspondence $(\delta^*\Theta - p_1^*\Theta - p_2^*\Theta)_L$, where δ, p_1 and p_2: $J_K \times_K J_K \to J_K$ are the difference map and the two projections, to get:

$$\langle a, da' \rangle_M = -\frac{1}{e_L} \left\langle [a'_L] - [z_L], \sum_{r=1}^{\alpha} (\Theta_L)^-_{\bar{a}_{r,+}} - (\Theta_L)^-_{\bar{a}_{r,-}} \right\rangle_{J_L} \mod \mathbb{Z}.$$

Here $(\Theta_L)^-$ stands for $[-1]^*(\Theta_L)$.

Now, with Notation 3.3.3, the divisor $\sum_{r=1}^{\alpha} (\Theta_L)^-_{\bar{a}_{r,+}} - (\Theta_L)^-_{\bar{a}_{r,-}}$ is the pull-back on J_L of the divisor $(\Theta^-)_{h_* D(a)}$ defined on J_K. On the other hand,

$$\begin{aligned} \sum_{r=1}^{\alpha} \bar{a}_{r,+} - \bar{a}_{r,-} &= \sum_{r=1}^{\alpha} (d[a_{r,+}] - E_L) - (d[a_{r,-}] - E_L) \\ &= d(D(a)_L) \in J_K(L) \\ &= da_K \in J_K(K). \end{aligned}$$

Therefore the theorem of the square on J_L shows that the two divisors $(\Theta^-)_{h_*D(a)}$ and $\Theta^-_{da_K} - \Theta^-$ on J_K are linearly equivalent over L, hence also over K ($J'_K(K)$ injects into $J'_L(L)$). From this observation, and the fact that the normalized valuation on K takes values in $\mathbb{Z}$, we deduce that

$$\langle a, da' \rangle_M = -\langle\, [a'_K] - [z_K]\, , \Theta^-_{da_K} - \Theta^- \rangle_{J_K} \quad \bmod \mathbb{Z}\,.$$

Applying once more the reciprocity law, we find

$$\langle a, da' \rangle_M = -\langle\, [da_K] - [0_K]\, , \Theta_{a'_K} - \Theta_{z_K} \rangle \quad \bmod \mathbb{Z}\,.$$

Finally, note that $(\Theta_{a'_K} - \Theta_{z_K}) = -\varphi(a'_K - z_K) \in J'(K)$ and $a'_K - z_K$ specializes to $a' \in \Phi_J$. Consequently, if we use $-\varphi$ to identify J_K with its dual, Theorem 3.4.1 tells us that

$$-\langle\, [da_K] - [0_K]\, , \Theta_{a'_K} - \Theta_{z_K} \rangle = \langle da, a' \rangle \quad \bmod \mathbb{Z}\,.$$

Whence $\langle a, da' \rangle_M = \langle da, a' \rangle$, as claimed. $\square$

Acknowledgments

I would like to thank Michel Raynaud for many enlightening discussions about Néron's and Grothendieck's pairings. I am indebted to Qing Liu for his valuable teaching of intersection theory on relative schemes. I am very grateful to Pascal Autissier, Siegfried Bosch, Dino Lorenzini and Laurent Moret-Bailly for their careful reading and for many helpful comments. I thank the referee for his meticulous reading and for many useful remarks.

References

[Bégueri 1980] L. Bégueri, *Dualité sur un corps local à corps résiduel algébriquement clos*, Mém. Soc. Math. France (N.S.) **4**, 1980. MR 82k:12019 Zbl 0502.14016

[Bertapelle 2001] A. Bertapelle, "On perfectness of Grothendieck's pairing for the l-parts of component groups", *J. Reine Angew. Math.* **538** (2001), 223–236. MR 2002k:14072 Zbl 1028.14017

[Bosch and Lorenzini 2002] S. Bosch and D. Lorenzini, "Grothendieck's pairing on component groups of Jacobians", *Invent. Math.* **148**:2 (2002), 353–396. MR 2003b:14053 Zbl 1061.14042

[Bosch et al. 1990] S. Bosch, W. Lütkebohmert, and M. Raynaud, *Néron models*, Ergeb. Math. Grenzgeb. (3) **21**, Springer, Berlin, 1990. MR 91i:14034 Zbl 0705.14001

[EGA IV_4 1967] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR 39 #220 Zbl 0153.22301

[FGA VI 1966] A. Grothendieck, "Technique de descente et théorèmes d'existence en géométrie algébrique, VI: Les schémas de Picard, propriétés générales", exposé 236 in *Séminaire Bourbaki,* 1961/1962, W. A. Benjamin, New York, 1966. Reprinted as pp. 221–243 in *Séminaire Bourbaki* **7**, Soc. Math. France, Paris, 1995. MR 1611207 Zbl 0238.14015

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergeb. Math. Grenzgeb. (3) **2**, Springer, Berlin, 1998. MR 99d:14003 Zbl 0885.14002

[Gross 1986] B. H. Gross, "Local heights on curves", pp. 327–339 in *Arithmetic geometry* (Storrs, CT, 1984), edited by G. Cornell and J. H. Silverman, Springer, New York, 1986. MR 861983 Zbl 0605.14027

[Hriljac 1985] P. Hriljac, "Heights and Arakelov's intersection theory", *Amer. J. Math.* **107**:1 (1985), 23–38. MR 86c:14024 Zbl 0593.14004

[Lang 1983] S. Lang, *Fundamentals of Diophantine geometry*, Springer, New York, 1983. MR 85j:11005 Zbl 0528.14013

[Lang 1988] S. Lang, *Introduction to Arakelov theory*, Springer, New York, 1988. MR 89m:11059 Zbl 0667.14001

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, 2002. MR 2003g:14001 Zbl 0996.14005

[Loerke 2009] K. Loerke, "Reduction of Abelian varieties and Grothendieck's pairing", preprint, 2009. arXiv 0909.4425v1

[Lorenzini 2008] D. Lorenzini, "Grothendieck's pairing for Jacobians and base change", *J. Number Theory* **128**:6 (2008), 1448–1457. MR 2009b:14088 Zbl 1198.14042

[Mazur and Tate 1983] B. Mazur and J. Tate, "Canonical height pairings via biextensions", pp. 195–237 in *Arithmetic and geometry*, vol. 1, edited by M. Artin and J. Tate, Progress in Mathematics **35**, Birkhäuser, Boston, MA, 1983. MR 85j:14081 Zbl 0574.14036

[Mumford 1974] D. Mumford, *Abelian varieties*, 2nd ed., Tata Inst. Fund. Res. Studies in Math. **5**, Oxford University Press, London, 1974. Zbl 0326.14012

[Murre 1964] J. P. Murre, "On contravariant functors from the category of pre-schemes over a field into the category of Abelian groups (with an application to the Picard functor)", *Inst. Hautes Études Sci. Publ. Math.* **23** (1964), 5–43. MR 34 #5836 Zbl 0142.18402

[Néron 1965] A. Néron, "Quasi-fonctions et hauteurs sur les variétés abéliennes", *Ann. of Math.* (2) **82** (1965), 249–331. MR 31 #3424 Zbl 0163.15205

[Oort 1962] F. Oort, "Sur le schéma de Picard", *Bull. Soc. Math. France* **90** (1962), 1–14. MR 25 #2070 Zbl 0123.13901

[Pepin 2011] C. Pepin, "Modèles semi-factoriels et modèles de Néron", preprint, 2011, Available at https://perswww.kuleuven.be/~u0079577/semi-factoriel-final.pdf. To appear in *Math. Ann.*

[Raynaud 1970] M. Raynaud, "Spécialisation du foncteur de Picard", *Inst. Hautes Études Sci. Publ. Math.* **38** (1970), 27–76. MR 44 #227 Zbl 0207.51602

[SGA 4 III 1973] M. Artin, A. Grothendieck, and J. L. Verdier (editors), *Théorie des topos et cohomologie étale des schémas. Tome 3. Exposes IX à XIX.* (Séminaire de Géométrie Algébrique du Bois-Marie 1963–1964 = SGA 4), Lecture Notes in Math. **305**, Springer, Berlin, 1973. MR 50 #7132 Zbl 0245.00002

[SGA 6 1971] A. Grothendieck, P. Berthelot, and L. Illusie, *Séminaire de Géométrie Algébrique du Bois Marie* 1966/1967*: Théorie des intersections et théorème de Riemann–Roch* (SGA 6), Lecture Notes in Mathematics **225**, Springer, Berlin, 1971. MR 50 #7133 Zbl 218.14001

[SGA 7 I 1972] A. Grothendieck, *Séminaire de Géométrie Algébrique du Bois Marie* 1967/1969*: Groupes de monodromie en géométrie algébrique, I* (SGA 7), Lecture Notes in Mathematics **288**, Springer, Berlin, 1972. MR 50 #7134 Zbl 0237.00013

[Werner 1997] A. Werner, "On Grothendieck's pairing of component groups in the semistable reduction case", *J. Reine Angew. Math.* **486** (1997), 205–215. MR 98j:14058 Zbl 0872.14037

[Zahrin 1972] J. G. Zahrin, “Néron pairing and quasicharacters”, *Math. USSR Izv.* **6**:3 (1972), 491–503. Zbl 0254.14012

Communicated by Barry Mazur
Received 2011-02-19 Revised 2011-12-21 Accepted 2012-01-18

cedric.pepin@wis.kuleuven.be *KU Leuven, Departement Wiskunde, Celestijnenlaan 200B, 3001 Heverlee, Belgium*
https://perswww.kuleuven.be/~u0079577/

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1349

Free subalgebras of quotient rings of Ore extensions

Jason P. Bell and Daniel Rogalski

Let K be a field extension of an uncountable base field k, let σ be a k-automorphism of K, and let δ be a k-derivation of K. We show that if D is one of $K(x;\sigma)$ or $K(x;\delta)$, then D either contains a free algebra over k on two generators, or every finitely generated subalgebra of D satisfies a polynomial identity. As a corollary, we show that the quotient division ring of any iterated Ore extension of an affine PI domain over k is either again PI, or else it contains a free algebra over its center on two variables.

1. Introduction

Many authors have noted that it is often the case that noncommutative division algebras have free subobjects. For example, the existence of nonabelian free groups inside the multiplicative group $D^\times$ of a division algebra D has been studied in several papers (see [Reichstein and Vonessen 1995; Chiba 1996] and the references therein). It is now known that if D is noncommutative and has uncountable center, then $D^\times$ contains a free subgroup on two generators [Chiba 1996].

The question of when a division k-algebra D contains a free k-subalgebra on two generators has also attracted much attention. The first result in this direction was obtained by Makar-Limanov [1983], who showed that if

$$A_1(k) = k\{x, y\}/(xy - yx - 1)$$

is the Weyl algebra over a field k of characteristic 0, then its quotient division algebra $D_1(k)$ does indeed contain such a free subalgebra. This result is perhaps surprising to those only familiar with localization in the commutative setting, and is in fact a good demonstration of how noncommutative localization is less well-behaved. In particular, the Weyl algebra A is an algebra of quadratic growth; that is, if we let V denote the k-vector subspace of A spanned by 1 and the images of x and y

Bell was supported by NSERC grant 31-611456. Rogalski was supported by NSF grant DMS-0900981.
MSC2010: primary 16K40; secondary 16S10, 16S36, 16S85.
Keywords: free algebra, division algebra, Ore extension, skew polynomial ring.

in A, then the dimension of V^n is a quadratic function of n. On the other hand, a free algebra on two generators has exponential growth. This is a good example of the principle that there is no nice relationship, in general, between the growth of a finitely generated algebra and the growth of other subalgebras of its quotient division algebra.

We note that by Lemma 1 of [Makar-Limanov and Malcolmson 1991], if a division k-algebra D contains a free k-subalgebra on two generators, then it contains a free F-subalgebra on two generators for any central subfield F. Thus the choice of base field is not an important consideration when considering the existence of free subalgebras, and need not even be mentioned. Now there are certain division algebras which cannot contain copies of free algebras on more than one generator for trivial reasons, for example, division algebras which are algebraic over their centers. Note also that a free algebra on two generators does not satisfy a polynomial identity. We say that a k-algebra R is *locally PI* if every finitely generated k-subalgebra of R is a polynomial identity ring (this is also easily seen to be independent of the choice of central base field k). An obvious necessary condition for a division algebra D to contain a noncommutative free algebra is that D not be locally PI. On the other hand, there are no known examples of division algebras which do not contain a free algebra on two generators, except locally PI ones.

In light of the discussion above, we say that a division algebra D satisfies the *free subalgebra conjecture* if D contains a free subalgebra on two generators if and only if D is not locally PI. Makar-Limanov [1984a] annunciated the FOFS (full of free subobjects) conjecture, one part of which was the statement that every division algebra D which is finitely generated (as a division algebra) and infinite-dimensional over its center contains a free subalgebra on two generators. It is easy to see, using a bit of PI theory, that this statement is equivalent to what we have called the free subalgebra conjecture here. Toby Stafford independently formulated a similar conjecture, as we have learned from Lance Small. As Makar-Limanov also notes, the conjecture is a bit provocative as stated, because it implies the resolution of the Kurosh problem for division rings. However, we will study the conjecture here only for special types of division rings in any case.

Since Makar-Limanov's original breakthrough, many authors have used his ideas to demonstrate the existence of free subalgebras on two generators in the quotient division algebras of many special classes of rings, especially certain Ore extensions, group algebras, and enveloping algebras of Lie algebras [Figueiredo et al. 1996; Lichtman 1999; Lorenz 1986; Makar-Limanov 1983; 1984b; 1984c; Makar-Limanov and Malcolmson 1991; Shirvani and Gonçalves 1998; 1999]. Our main aim here is to further develop the Ore extension case. Suppose that D is a division ring with automorphism $\sigma : D \to D$ and σ-derivation δ, and let $D(x; \sigma, \delta)$ be the quotient division ring of the Ore extension $D[x; \sigma, \delta]$. Lorenz [1986] showed that $k(t)(x;\sigma)$ contains a free subalgebra on two generators when σ has infinite order. Shirvani and J. Z. Gonçalves [1999] showed that if R is a k-algebra which is a UFD with field of fractions K and the property that $R^\times = k^\times$, and $\sigma : R \to R$ is a k-automorphism such that the σ-fixed subring of R is k, then $K(x;\sigma)$ contains a free subalgebra on two generators (in fact, even a free group algebra of rank $|k|$).

In this paper, we first give in Section 2 some new criteria for the existence of a free k-subalgebra on two generators in a division ring $D(x;\sigma,\delta)$, following the main idea of Makar-Limanov's original method. We then use these criteria to completely settle the free subalgebra conjecture for the case of Ore extensions of fields, assuming an uncountable base field. Our main results are the following.

Theorem 1.1. *Let K/k be a field extension and $\sigma : K \to K$ a k-automorphism.*

(1) *If k is uncountable, then the following are equivalent*:
   (i) *$K(x;\sigma)$ contains a free k-subalgebra on two generators.*
   (ii) *$K(x;\sigma)$ is not locally PI.*
   (iii) *K has an element lying on an infinite σ-orbit.*

(2) *If k is countable, the same conclusion as in* (1) *holds if either K/k is infinitely generated as a field extension, or if σ is induced by a regular k-automorphism of a quasiprojective k-variety with function field K.*

We expect that the free subalgebra conjecture for $K(x;\sigma)$ is always true, with no restrictions on k; in any case, the theorem above certainly covers the cases one is most likely to encounter.

We also study the derivation case, which is in fact easier and requires no assumption on the base field.

Theorem 1.2. *Let K be a field extension of a field k. If $\delta : K \to K$ is a k-derivation, then $K(x;\delta)$ contains a free k-subalgebra on two generators if and only if it is not locally PI.*

See Theorem 4.1 for a characterization of when $K(x;\delta)$ is locally PI.

In fact, a general Ore extension $K[x;\sigma,\delta]$ of a field K is isomorphic to one with either $\sigma = 1$ or with $\delta = 0$. So as a rather quick consequence of the theorems above, we obtain the following result.

Theorem 1.3. *The quotient division algebra of any iterated Ore extension of a PI domain which is affine over an uncountable field satisfies the free subalgebra conjecture.*

We note that our proofs are largely independent of past work in this subject, except that we assume Makar-Limanov's original result. Some authors have considered the more general question of the existence of k-free group algebras in a division ring D, and have also studied the cardinality of the rank of the largest such free group algebra. For simplicity, we stick to the context of the free subalgebra conjecture here. We mention that Shirvani and Gonçalves [1996] have shown that if the center k of D is uncountable, then the existence of a free group k-algebra of rank 2 in D is implied by the existence of a free k-algebra on two generators.

In this paper, we have tried to make as few assumptions as possible on the ground field k. In a follow-up paper [Bell and Rogalski 2011], we give stronger criteria for existence of free subalgebras, and thus verify the free subalgebra conjecture for some additional classes of algebras, in case k is uncountable.

2. Criteria for existence of free subalgebras of a division algebra

In this section we use ideas of Makar-Limanov to give a simple criterion that guarantees that a division algebra D contains a copy of a free algebra on two generators. We work over an arbitrary field k. As we noted in the introduction, the question of whether D contains a free k-subalgebra on two generators is independent of the choice of central subfield k.

Notation 2.1. We use the following notation:

(1) We let D denote a division algebra over a field k.

(2) We let $\sigma : D \to D$ denote a k-algebra automorphism of D.

(3) We let $\delta : D \to D$ denote any σ-derivation of D over k, that is, a k-linear map satisfying $\delta(ab) = \sigma(a)\delta(b) + \delta(a)b$ for all $a, b \in D$.

(4) We let $\psi = (\sigma - 1) + \delta : D \to D$ (this is also a σ-derivation) and set

$$E = \{u \in D \;:\; \psi(u) = 0\},$$

which is a division subring of D.

(5) We let $D[x; \sigma, \delta]$ be the Ore extension generated by D and the indeterminate x with relations $xa = \sigma(a)x + \delta(a)$ for $a \in D$, and let $D(x; \sigma, \delta)$ denote its quotient division algebra. As usual, if $\sigma = 1$ we omit σ from the notation, and if $\delta = 0$ we omit δ from the notation.

We now prove a sufficient condition for the ring $D(x; \sigma, \delta)$ to contain a free subalgebra. Compared to the original method of Makar-Limanov's, we choose a slightly different pair of elements, and we avoid the use of power series. We note that the characteristic of the base field has no effect in the following criterion.

Theorem 2.2. *Assume Notation 2.1 and let $b \in D$. If*

(1) $b \notin \sigma(E)$, *and*

(2) *for all $u \in D$, $\psi(u) \in \sigma(E) + \sigma(E)b$ implies $u \in E$,*

then the k-algebra generated by $b(1 - x)^{-1}$ and $(1 - x)^{-1}$ is a free subalgebra of $D(x; \sigma, \delta)$.

Proof. Let

$$\mathcal{S} = \{(i_1, \dots, i_r) \ : \ r \geq 1, i_1, \dots, i_r \in \{0, 1\}\} \cup \{\varnothing\}, \tag{2.1}$$

and for nonempty $I = (i_1, \dots, i_r) \in \mathcal{S}$, define length$(I) := r$ and

$$W_I := b^{i_1}(1-x)^{-1}b^{i_2}(1-x)^{-1} \cdots b^{i_r}(1-x)^{-1}. \tag{2.2}$$

If $I = \varnothing$, we define length$(\varnothing) := 0$ and $W_\varnothing := 1$. Note that the W_I are exactly the words in the generators $b(1-x)^{-1}$, $(1-x)^{-1}$, and so our task is to show that $\{W_I \mid I \in \mathcal{S}\}$ is linearly independent over k. It is also useful to define

$$V_I := (1-x)^{-1}b^{i_1}(1-x)^{-1} \cdots b^{i_r}(1-x)^{-1}, \tag{2.3}$$

for nonempty $I \in \mathcal{S}$, and to set $V_\varnothing := (1-x)^{-1}$.

For nonempty $I = (i_1, i_2, \dots, i_r)$, we define its truncation as $I' = (i_2, \dots, i_r)$, with the convention that if I has length 1, then $I' = \varnothing$. Note then that trivially from the definitions we have

$$(1-x)V_I = W_I = b^{i_1} V_{I'}, \qquad \text{for } I \neq \varnothing. \tag{2.4}$$

We claim that to prove that the W_I are k-independent. It is enough to prove that $\{V_I \mid I \in S\}$ is left D-independent. To see this, suppose the V_I are D-independent and that we have a nontrivial relation $\sum_{I \in S} c_I W_I = 0$, with $c_I \in k$ not all 0. We can assume that $c_\varnothing = 0$ by multiplying our relation through on the right by $(1-x)^{-1}$. Then

$$0 = \sum_{I \in S} c_I W_I = \sum_{I \neq \varnothing} c_I b^{i_1} V_{I'} = 0.$$

This forces, for each nonempty $I \in S$, the equation $c_I b^{i_1} + c_H b^{1-i_1} = 0$, where H is the other element of S which has truncation I'. But then $c_I = c_H = 0$, since $\{1, b\}$ is certainly k-independent, given that $b \notin \sigma(E)$. This contradicts the nontriviality of our chosen relation and the claim is proved.

The strategy is to prove by contradiction that $\{V_I \mid I \in S\}$ is left D-independent. In fact, it is more convenient to prove the seemingly stronger statement that this set is D-independent in the left factor D-space $D(x; \sigma, \delta)/D[x; \sigma, \delta]$. In other words, we work modulo polynomials. Equivalently, we suppose that we have a relation $\sum_{I \in S} \alpha_I V_I = p(x) \in D[x; \sigma, \delta]$, with $\alpha_I \in D$ not all 0. Among all such relations, we pick one with a minimal value of $d = \min(\text{length}(I) \mid \alpha_I \neq 0)$. Moreover, among these, we select one with the smallest number of nonzero α_I with length$(I) = d$. Note that certainly $d \geq 1$. By multiplying our relation by a nonzero element of D, we may also assume that $\alpha_J = 1$, for some J of length d.

Now for nonempty I, (2.4) can be rewritten as $xV_I = V_I - b^{i_1} V_{I'}$, and for $I = \varnothing$, we have $xV_I = V_I - 1$. Multiplying our relation on the left by x and applying these formulas, we obtain

$$\begin{aligned} x\sum_{I\in S}\alpha_I V_I &= \sum_{I\in S}\bigl[\sigma(\alpha_I)x+\delta(\alpha_I)\bigr]V_I \\ &= \sum_{I\in S}\bigl[\sigma(\alpha_I)+\delta(\alpha_I)\bigr]V_I-\sum_{I\neq\varnothing}\sigma(\alpha_I)b^{i_1}V_{I'}-\sigma(\alpha_\varnothing) \\ &= xp(x)\in D[x;\sigma,\delta]. \end{aligned}$$

Subtracting the original relation $\sum\alpha_I V_I = p(x)$, we get

$$\sum_{I\in S}\bigl[\sigma(\alpha_I)-\alpha_I+\delta(\alpha_I)\bigr]V_I-\sum_{I\neq\varnothing}\sigma(\alpha_I)b^{i_1}V_{I'}=(x-1)p(x)+\sigma(\alpha_\varnothing)\in D[x;\sigma,\delta]. \tag{2.5}$$

But notice that since $\alpha_J = 1$, the coefficient of V_J in this relation is now 0, while no new nonzero coefficients associated to V_I with I of length d have appeared. Thus by our assumption that we originally picked a minimal relation, all coefficients of the V_I on the left-hand side of (2.5) are 0. In particular,

$$\psi(\alpha_I)=\sigma(\alpha_I)-\alpha_I+\delta(\alpha_I)=0,$$

so $\alpha_I\in E$, for all I of length d. Also, if H is the other element of S with truncation J', then the coefficient of $V_{J'}$ in (2.5) is

$$\sigma(\alpha_{J'})-\alpha_{J'}+\delta(\alpha_{J'})-\sigma(\alpha_J)b^{j_1}-\sigma(\alpha_H)b^{1-j_1}=0.$$

Since H and J have length d, for $u=\alpha_{J'}$ we obtain

$$\psi(u)=\sigma(u)-u+\delta(u)\in\sigma(E)b+\sigma(E).$$

Note that also $\psi(u)\neq 0$, as the assumption $b\notin\sigma(E)$ implies that $\sigma(E)b+\sigma(E)$ is direct, and $\sigma(\alpha_J)=1$. The existence of such a u violates the hypothesis, so we have achieved a contradiction. Thus $b(1-x)^{-1}$ and $(1-x)^{-1}$ generate a free subalgebra of $D(x;\sigma,\delta)$, as claimed. □

The interaction between δ and σ in the criterion of the preceding theorem seems to make it hard to analyze in general. In practice, we will only use the theorem later in the special cases where $\delta=0$ or $\sigma=1$. In the rest of this section, we examine the criterion for the special case of $D(x;\sigma)$ more closely. As mentioned in the introduction, Makar-Limanov [1983] proved that the Ore quotient ring of the first Weyl algebra, $D_1(k)$, contains a free k-subalgebra on two generators when k has characteristic 0 (see also [Krause and Lenagan 2000, Theorem 8.17]). It is standard that $D_1(k)\cong k(u)(x;\sigma)$, where $\sigma(u)=u+1$, but we note that Theorem 2.2, as stated, does not recover Makar-Limanov's result. More specifically, taking $D=k(u)$, $\sigma(u)=u+1$, and $\delta=0$ in Notation 2.1, it is easy to see that $E=k$, but we have $\sigma(u)-u\in k$ with $u\notin k$; thus the criterion in Theorem 2.2 cannot be satisfied regardless of b. In fact, our criterion seems to be most useful when we combine it with Makar-Limanov's known result to give the following stronger criterion.

Theorem 2.3. *Assume the notation from Notation 2.1, with $\delta = 0$. Suppose that either k has characteristic 0, or else k has characteristic $p > 0$ and we have the additional condition that*

$$\{a \in D \mid \sigma^p(a) = a\} = E.$$

If there is $b \in D \setminus E$ such that the equation

$$\sigma(u) - u \ \in \ b + E \tag{2.6}$$

has no solutions for $u \in D$, then the k-algebra generated by $b(1-x)^{-1}$ and $(1-x)^{-1}$ is a free subalgebra of $D(x; \sigma)$.

Proof. Choose b as in the hypothesis. If $\sigma(u) - u \in E + Eb$ has no solutions except for $u \in E$, then we are done by Theorem 2.2. So we may assume there is a solution of the form $\sigma(u) - u = \alpha + \beta b$ with $\alpha, \beta \in E$ not both zero. As long as $\beta \neq 0$, we may replace u by $\beta^{-1}u$ and thus assume that $\beta = 1$, and so (2.6) has a solution, contradicting the hypothesis. Thus $\beta = 0$. Then $\alpha \neq 0$ and $y = u\alpha^{-1}$ satisfies $\sigma(y) = y + 1$. Then the elements $z = yx^{-1}$ and x satisfy the relation $xz - zx = 1$. If $\operatorname{char} k = 0$, then we see that the k-subalgebra R of $D(x; \sigma)$ generated by x and z is isomorphic to a factor of the Weyl algebra $A_1(k)$. Since the Weyl algebra is simple, $R \cong A_1(k)$, and so $D(x; \sigma)$ must contain a copy of $D_1(k)$, and hence a free k-algebra on two generators [Makar-Limanov 1983]. If instead $\operatorname{char} k = p > 0$, then we have $\sigma^p(y) = y$. It follows by the hypothesis that $y \in E$, but this contradicts $\sigma(y) = y + 1$. □

We end this section with a valuation-theoretic criterion that will be especially useful later when D is a field. Recall that a *discrete valuation* of a division ring D is a function $\nu : D^\times \to \mathbb{Z}$ such that $\nu(xy) = \nu(x) + \nu(y)$ and $\nu(x+y) \geq \min(\nu(x), \nu(y))$ for all $x, y \in D^\times$. It is easy to see that $\nu(x+y) = \min(\nu(x), \nu(y))$ if $\nu(x) \neq \nu(y)$. The valuation ν is *trivial* if $\nu(x) = 0$ for all $x \in D^\times$.

Lemma 2.4. *Assume the notation from Notation 2.1, with $\delta = 0$. Suppose that D has a nontrivial discrete valuation $\nu : D^\times \to \mathbb{Z}$, such that (i) $\nu(a) = 0$ for all $a \in E$; and (ii) for all $a \in D^\times$, $\nu(\sigma^n(a)) = 0$ for all $n \gg 0$ and all $n \ll 0$. If* $\operatorname{char} k = 0$, *or if* $\operatorname{char} k = p > 0$ *and $\{y \in D \mid \sigma^p(y) = y\} = E$, then $D(x; \sigma)$ contains a free subalgebra on two generators.*

Proof. For any given $u \in D$, by hypothesis $X_u = \{n \in \mathbb{Z} \mid \nu(\sigma^n(u)) < 0\}$ is a finite set, and if $X_u \neq \varnothing$, we call $\ell(u) = \max X_u - \min X_u$ the *length* of u. If $X_u \neq \varnothing$, then it is easy to see that $\ell(u - \sigma(u)) = \ell(u) + 1$.

By nontriviality, we can pick $b \in D$ such that $X_b \neq \varnothing$. Among all such b, choose one of minimal length, say $\ell(b) = d$. We claim that there are no solutions to the equation $u - \sigma(u) = b + e$ with $u \in D$ and $e \in E$. Suppose (u, e) does give such a solution. It is easy to see that $\ell(b + e) = \ell(b) = d$, since $\nu(\sigma^n(e)) = 0$ for all n by hypothesis (i). Now $X_u = \varnothing$ is clearly impossible, so u has a length. By minimality, $\ell(u) \geq d$, and so $\ell(u - \sigma(u)) \geq d + 1$, a contradiction. The result now easily follows from Theorem 2.3. □

3. The automorphism case

In this section, the goal is to use the criteria developed in the previous section to study when $K(x; \sigma)$ contains a free subalgebra, where K is a field, and thus to prove Theorem 1.1. In terms of Notation 2.1, we now write $D = K$ for a field K containing the base field k with k-automorphism $\sigma : K \to K$, and assume that $\delta = 0$. Then $K(x; \sigma)$ is also an algebra over the fixed subfield $E = \{a \in K \mid \sigma(a) = a\}$, and as already mentioned, we may change the base field to any central subfield without affecting the question of the existence of free subalgebras. Thus, it does no harm to replace k by E, and we assume that the base field k is the σ-fixed field for the rest of this section. We will frequently use in this section the exponent notation $b^\sigma := \sigma(b)$ for the action of an automorphism on an element.

The difficult direction of Theorem 1.1 is to prove that $K(x; \sigma)$ contains a free subalgebra on two generators if K contains an element a lying on an infinite σ-orbit. In this case, letting $K' = k(\dots, a^{\sigma^{-2}}, a^{\sigma^{-1}}, a, a^\sigma, \dots)$ be the subfield of K generated over k by the σ-orbit of a, it suffices to prove that $K'(x; \sigma)$ contains a noncommutative free subalgebra. Thus, in this section, we will often assume the following hypothesis.

Hypothesis 3.1. *Let K be a field with automorphism $\sigma : K \to K$, and let k be the fixed field of σ. Assume that there is an element $a \in K$ on an infinite σ-orbit such that $K = k(a^{\sigma^n} \mid n \in \mathbb{Z})$.*

The proof that $K(x; \sigma)$ satisfying Hypothesis 3.1 contains a free subalgebra naturally breaks up into two cases, depending on whether or not K/k is finitely generated as a field extension. The infinitely generated case is rather easily dispatched. We thank the referee very much for suggesting the elegant proof of the following proposition, which gives a simpler and more direct method for handling the infinitely generated case than our original.

Proposition 3.2. *Assume Hypothesis 3.1, and suppose that K is infinitely generated as a field extension of k. Then $K(x; \sigma)$ contains a free k-subalgebra on two generators.*

Proof. Write $a_j = a^{\sigma^j}$ for all $j \in \mathbb{Z}$. Suppose first that $k(a_i \mid i \geq 0)$ is still an infinitely generated field extension of k. In this case, we will show that in fact the countable system of elements $\{ax^j \mid j \geq 1\}$ generates a free k-subalgebra of $K(x;\sigma)$.

Let $y_j = ax^j$ for all $j \geq 1$. Then an arbitrary monomial in the y_j looks like

$$y_{j_1} y_{j_2} \dots y_{j_n} = a_0 a_{j_1} a_{j_1+j_2} \dots a_{j_1+j_2+\dots+j_{n-1}} x^{j_1+j_2+\dots+j_n},$$

for some $j_1, \dots, j_n \geq 1$. We claim that the set

$$S = \{a_0 a_{i_1} a_{i_2} \dots a_{i_m} \mid m \geq 1, 0 < i_1 < i_2 < \dots < i_m\} \cup \{a_0\}$$

is linearly independent over k. Suppose not, and pick a linear dependency relation over k in which the maximum n such that a_n appears in this relation is as small as possible. Clearly $n \geq 1$. Since every element of S is a product of distinct a_i, the dependency relation has the form $p\,a_n + q = 0$, where p, q are linear combinations of elements of S involving only a_i with $i < n$. If $p = 0$, we contradict the choice of n. Thus $a_n = q/p \in k(a_0, a_1, \dots, a_{n-1})$. Applying σ, this easily implies by induction that $k(a_0, a_1, \dots, a_{n-1}) = k(a_i \mid i \geq 0)$, contradicting the assumption that the latter field is infinitely generated as an extension of k. This establishes the claim that S is linearly independent over k. Together with the K-independence of the powers of x, this implies that the distinct monomials $y_{j_1} y_{j_2} \dots y_{j_n}$ are linearly independent over k. In other words, the y_i generate a free subalgebra of $K(x;\sigma)$ as required.

Suppose instead that $k(a_i \mid i \leq 0)$ is an infinitely generated field extension of k. A symmetric argument to the above shows that $\{ax^j \mid j \leq -1\}$ generates a free subalgebra of $K(x;\sigma)$. Finally, if both $k(a_i \mid i \leq 0)$ and $k(a_i \mid i \geq 0)$ are finitely generated field extensions of k, then $K = k(a_i \mid i \in \mathbb{Z})$ is also a finitely generated extension of k, contradicting the hypothesis. $\square$

Now we begin to tackle the case where K/k is a finitely generated field extension. The idea in this case is to construct an appropriate valuation satisfying the hypothesis of Lemma 2.4, using algebraic geometry, which we do in the next proposition. All geometric notions we use will be standard ones as defined in [Hartshorne 1977]. Though sometimes we cite passages in that reference that belong to sections with a blanket hypothesis that the base field k is algebraically closed, one can check in each case that this restriction is unnecessary.

Before proving the proposition, we review some basic geometric facts that we will need in particular, and prove a lemma. For convenience, we work only with quasiprojective varieties over the field k, as defined in [Hartshorne 1977, Section I.2], and for brevity we simply call these k-varieties. While for the most part we do not need schemes, it will be convenient occasionally to think of a variety as a scheme, as in [Hartshorne 1977, Proposition II.2.6], so that we may use the notion of the generic point of an irreducible subvariety.

Let X be a k-variety. Recall that there is a bijective correspondence between birational maps $\phi : X \dashrightarrow X$ and automorphisms of the field $k(X)$ of rational functions on X [Hartshorne 1977, Theorem I.4.4]. More explicitly, given a birational map $\phi : X \dashrightarrow X$, for $f \in k(X)$ we write $\phi(f) = f^{\phi} = f \circ \phi \in k(X)$; then $f \mapsto f^{\phi}$ is the corresponding automorphism of $k(X)$, and we call this automorphism $\phi : k(X) \to k(X)$ as well. Conversely, if we begin with an automorphism $\tau : K \to K$ of a finitely generated field extension K/k, we may be able to choose a k-variety X with function field $k(X) = K$ such that the corresponding birational map $\tau : X \dashrightarrow X$ is not just birational, but in fact a regular k-automorphism of X. In this case we say that $\tau : K \to K$ is induced by a regular k-automorphism of X. In general, however, it could be that no such choice of X exists; see, for example, [Rogalski 2009, Section 3] for more discussion of this issue.

Suppose that X is a normal k-variety. Then for any prime divisor (that is, codimension-1 irreducible subvariety) C of X, there is a corresponding discrete valuation v_C of $K = k(X)$ such that $v_C(f)$ is the order of vanishing of f along C [Hartshorne 1977, Section II.6]. Recall that $\operatorname{Div} X$ is the free abelian group with basis the distinct prime divisors C on X. Given a rational function $f \in k(X)$, we define its corresponding *principal divisor* in $\operatorname{Div} X$ as

$$(f) = \sum_C v_C(f)C.$$

Now suppose that $\sigma : X \dashrightarrow X$ is a birational map of a normal projective k-variety. Then by Zariski's main theorem, the closed set where σ is undefined is at least codimension 2 in X. In other words, given any prime divisor C on X, σ is defined at the generic point η of C [Hartshorne 1977, Lemma V.5.1], or equivalently, σ is defined on an open subset of the points of C. We write $\sigma(C)$ to mean the closure of $\sigma(\eta)$. Since σ is merely birational, $\sigma(C)$ may be a closed subset of codimension greater than 1 in X, in which case we say that σ *contracts* C. On the other hand, if $\sigma(C)$ is again of codimension 1, then the birational map σ^{-1} must be defined at the generic point $\sigma(\eta)$ of $\sigma(C)$, and we conclude that σ is a local isomorphism from an open neighborhood of η to an open neighborhood of $\sigma(\eta)$.

Lemma 3.3. *Let $\sigma : X \dashrightarrow X$ be a birational map of a normal projective k-variety. If C is a prime divisor not contracted by σ, then $v_{\sigma(C)} = v_C \circ \sigma$. Moreover, if $f \in K = k(X)$ has the property that σ^{-1} contracts no prime divisor appearing with nonzero coefficient in $D = (f)$, and σ contracts no prime divisor appearing with nonzero coefficient in $E = (f^{\sigma})$, then $E = \sigma^{-1}(D)$.*

Proof. Let C be a prime divisor not contracted by σ, so that σ is a local isomorphism from a neighborhood of the generic point η of C to a neighborhood of the generic point $\sigma(\eta)$ of $\sigma(C)$. The formula $v_{\sigma(C)} = v_C \circ \sigma$ follows, since by definition,

$v_{\sigma(C)}(f)$ depends only on the image of f in the local ring $\mathbb{O}_{X,\sigma(\eta)}$ of functions defined in a neighborhood of $\sigma(\eta)$.

For the second statement, suppose that C is a prime divisor such that $v_C(f) \neq 0$. Then by hypothesis, σ^{-1} does not contract C, and so $B = \sigma^{-1}(C)$ is another prime divisor which is not contracted by σ, and using the previous paragraph, we have that $v_B(f^\sigma) = v_{\sigma(B)}(f) = v_C(f) \neq 0$. Conversely, if B is a prime divisor with $v_B(f^\sigma) \neq 0$, then by hypothesis, $C = \sigma(B)$ is another prime divisor and again $v_C(f) = v_{\sigma(B)}(f) = v_B(f^\sigma) \neq 0$. It follows that if $B_1, \ldots, B_m$ are the distinct prime divisors occurring with nonzero coefficient in (f^σ), then $C_1 = \sigma(B_1), \ldots, C_m = \sigma(B_m)$ are the distinct prime divisors occurring with nonzero coefficient in (f). Thus

$$\begin{aligned} E = (f^\sigma) &= \sum_{i=1}^{m} v_{B_i}(f^\sigma) B_i = \sum_{i=1}^{m} v_{\sigma(B_i)}(f) B_i \\ &= \sum_{i=1}^{m} v_{C_i}(f) \sigma^{-1}(C_i) = \sigma^{-1}\Big(\sum_{i=1}^{m} v_{C_i}(f) C_i\Big) = \sigma^{-1}(D), \end{aligned}$$

as required. □

We remark that a similar (but easier) argument shows that if instead $\sigma : X \to X$ is a regular k-automorphism of any normal k-variety, then the formulas $v_{\sigma(C)} = v_C \circ \sigma$ and $(f^\sigma) = \sigma^{-1}[(f)]$ as in the previous result always hold for any prime divisor C and any $f \in k(X)$.

The assumption in the next result that K/k is totally transcendental (that is, that every element $b \in K \setminus k$ is transcendental over k) will be easily removed in the proof of Theorem 1.1.

Proposition 3.4. *Assume Hypothesis 3.1, and that K/k is a totally transcendental finitely generated field extension of k. Suppose that either k is uncountable, or that there is a quasiprojective k-variety X with function field K such that $\sigma : K \to K$ is induced by a regular k-automorphism of X. Then $K(x; \sigma)$ contains a free subalgebra on two generators.*

Proof. Note that if an element b in K has finite order under σ, it is algebraic over k, and thus in k by the assumption that K/k is totally transcendental. Thus all elements in K/k have infinite order under σ. In particular, K/k must have transcendence degree at least 1.

Assume first that there is a quasiprojective k-variety X with function field K such that $\sigma : K \to K$ is induced by a regular k-automorphism of X, which we give the same name $\sigma : X \to X$. An automorphism of a variety induces an automorphism of its normalization, by the universal property of the normalization [Hartshorne 1977, Exercise II.3.8]. Thus by replacing X with its normalization (which is again quasiprojective by a standard result), we can assume that X is normal. Suppose that X has an prime divisor C lying on an infinite σ-orbit of divisors. Then we take the valuation v_C of K, which satisfies the hypotheses of Lemma 2.4 since $v_{\sigma^i(C)} = v_C \circ \sigma^i$ by the remark following Lemma 3.3, and since a rational function has a pole or zero along at most finitely many of the divisors $\sigma^i(C)$. We conclude by that lemma that $K(x; \sigma)$ contains a free k-algebra on two generators. (Note that the extra hypothesis of Lemma 2.4 in characteristic p holds since K/k is totally transcendental.)

Otherwise, all prime divisors C of X lie on finite σ-orbits. We now apply a similar argument as in [Bell et al. 2010, Theorem 5.7] to show that this implies the existence of a σ^n-eigenvector in $K \setminus k$ for some n. Note that X certainly has infinitely many distinct prime divisors, since $\dim X \geq 1$. Thus we can pick a sequence of rational functions $f_1, f_2, \dots \in K \setminus k$ such that for each i, f_i has a zero or pole along some prime divisor not appearing in (f_j) for all $1 \leq j < i$. Note that for any i, if n is a multiple of the order under σ of all of the divisors appearing in (f_i), then $(f_i^{\sigma^n}) = \sigma^{-n}[(f_i)] = (f_i)$ (using the remark following Lemma 3.3), and thus $u_i = f_i^{\sigma^n}/f_i$ has principal divisor $(u_i) = 0$. In other words, u_i is in $G = \Gamma(X, \mathbb{O}_X)^*$, the units group of the ring of global regular functions on X. Now $H = G/(\overline{k}^* \cap G)$ is a finitely generated abelian group [Bell et al. 2010, Lemma 5.6(2)] which is easily seen to be torsion-free, where $\overline{k}$ is the algebraic closure of k. (We note that in order to apply this result of Bell et al., we need X to be quasiprojective.) In fact, since K/k is totally transcendental, we have that $H = G/(k^* \cap G)$; say this group has rank d. Then we can choose $n > 0$ such that $(f_i^{\sigma^n}) = (f_i)$ for all $1 \leq i \leq d+1$, and so $u_i = f_i^{\sigma^n}/f_i$ is in G for all $1 \leq i \leq d+1$. This forces $\lambda = u_1^{a_1} u_2^{a_2} \dots u_{d+1}^{a_{d+1}} \in k$ for some integers a_i, not all 0. Then $g = f_1^{a_1} f_2^{a_2} \dots f_{d+1}^{a_{d+1}}$ satisfies $\sigma^n(g) = \lambda g$. Moreover, $g \notin k$, because otherwise (f_{d+1}) would involve only prime divisors occurring among the (f_i) with $1 \leq i < d+1$. Now if λ is a root of 1, then $\sigma^m(g) = g$ for some $m > 0$, which implies that g is algebraic over k, contradicting that K/k is totally transcendental. So λ has infinite multiplicative order. Then $L = k(g)$ is a rational function field over k to which the automorphism σ^n restricts as an infinite order automorphism, and it suffices to show that $L(x; \sigma^n)$ contains a free subalgebra on two generators. This follows from another application of Lemma 2.4 to L and its automorphism $\sigma^n|_L$, choosing the valuation associated to the maximal ideal $(g-1)$ of $k[g]$, which lies on an infinite $\sigma^n|_L$-orbit.

Next, we assume instead that k is uncountable. In this case, we will have to work with a birational map of a variety only, but will be able to perform a similar argument to the above by choosing f_i such that (f_i) avoids the places where the birational map is not an isomorphism. Since K/k is finitely generated, it is well-known that we can choose a normal projective k-variety X such that $k(X) = K$. The automorphism $\sigma : K \to K$ corresponds to a birational map $\sigma : X \dashrightarrow X$. Since σ is an isomorphism on some open subset of X, σ must contract at most finitely many prime divisors. Thus the set S of prime divisors on X which are contracted by some σ^n with $n \in \mathbb{Z}$ is countable.

We now show that we can find a plentiful supply of rational functions whose σ-iterates have divisors entirely avoiding the bad set S. Pick any element $h \in K \setminus k$. Consider $h + \lambda$ as $\lambda \in k$ varies. The divisors along which $h + \lambda_1$ and $h + \lambda_2$ have a zero are disjoint if $\lambda_1 \neq \lambda_2$. For a given $i \in \mathbb{Z}$, there are countably many λ such that $h^{\sigma^i} + \lambda$ has a zero along a divisor in S. Since k is uncountable, there are uncountably many λ such that $h^{\sigma^i} + \lambda$ has zeroes only along divisors not in S, for all i. Fix such a λ and put $g = 1/(h + \lambda)$; thus g^{σ^i} has poles only along divisors not in S, for all i. By the same argument, for uncountably many μ, the rational functions $g^{\sigma^i} + \mu$ have no zeroes in S, for all i. Let $f_\mu = g + \mu$. By construction, there are uncountably many $\mu \in k$ such that $f_\mu^{\sigma^i}$ has no zeroes or poles in S, for all i; and moreover, f_{μ_1} and f_{μ_2} have disjoint zeroes if $\mu_1 \neq \mu_2$.

Now we may essentially repeat the argument of the first half of the proof. If there is a prime divisor C not in S which lies on an infinite σ-orbit, then since the equation $v_{\sigma^i(C)} = v_C \circ \sigma^i$ holds for all $i \in \mathbb{Z}$, by Lemma 3.3, this allows us to apply Lemma 2.4 to the valuation v_C to conclude that $K(x;\sigma)$ contains a free subalgebra on two generators. Otherwise, by the construction of the f_μ, we may choose a sequence of rational functions $f_1, f_2, \dots$ from among the uncountably many f_μ's whose σ-iterates all avoid the set S, such that each (f_i) has a zero along some prime divisor not appearing in (f_j), for all $1 \leq j < i$. Lemma 3.3 implies that the equation $(f_i^{\sigma^n}) = \sigma^{-n}[(f_i)] = (f_i)$, which was needed in the third paragraph of the proof, still holds for any n which is a multiple of the (necessarily finite) order under σ of the prime divisors occurring in (f_i). Thus, the same argument as in the third paragraph of the proof goes through to construct a free subalgebra of $K(x;\sigma)$ in this case also. $\square$

Proof of Theorem 1.1. Let $\sigma : K \to K$ be an automorphism of K/k. If every element of K lies on a finite σ-orbit, then setting $K_n = \{x \in K \mid \sigma^n(x) = x\}$, we will have $K = \bigcup_{n\geq 1} K_n$, and thus $K(x;\sigma) = \bigcup_{n\geq 1} K_n(x;\sigma)$ is a directed union of PI algebras. Thus it is locally PI, and cannot possibly contain a free subalgebra on two generators.

To complete the proof, we assume that there is an element $a \in K$ lying on an infinite σ-orbit, and need to prove that $K(x;\sigma)$ contains a free subalgebra. We have seen that we may assume the conditions in Hypothesis 3.1, so that k is the fixed field of σ and $K = k(a^{\sigma^n} \mid n \in \mathbb{Z})$. If K/k is infinitely generated as a field extension, then we are done by Proposition 3.2, with no assumptions on the base field k necessary. Suppose instead that K/k is finitely generated, and that we have either that (i) k is uncountable, or (ii) there is a k-automorphism σ of a quasiprojective k-variety X with $k(X) = K$ inducing $\sigma : K \to K$ (we may assume that X is normal, just as in the proof of Proposition 3.2). To apply Proposition 3.4, we need to reduce to the totally transcendental case. If $L \subseteq K$ is the subfield of elements algebraic over k, then L/k is also finitely generated, and thus $[L : k] < \infty$. The elements in L have finite order under σ, and thus there is a single power σ^d such that $\sigma^d(b) = b$ for all $b \in L$. Let $K' := k(a^{\sigma^{nd}} \mid n \in \mathbb{Z})$. We now replace (K, σ) by (K', σ^d). By construction, K' is still generated by the σ^d-iterates of a single element a on an infinite orbit. The field $k' = \{b \in K \mid \sigma^d(b) = b\}$ certainly contains L; in fact, k' is algebraic over the field k of σ-fixed elements, and so $k' = L$. Thus the fixed field of $\sigma^d : K' \to K'$ is now $K' \cap L$, and the field extension $K'/(K' \cap L)$ is now totally transcendental. If we have hypothesis (ii), then X is an L-variety since rational functions which are algebraic over k must be global regular functions (X is normal), and $\sigma^d : X \to X$ is now an L-automorphism. Thus in either case (i) or (ii), replacing the triple $(K/k, \sigma, a)$ with $(K'/(K' \cap L), \sigma^d, a)$ preserves the hypothesis, and now $K'/(K' \cap L)$ is totally transcendental. By Proposition 3.4, $K'(x; \sigma^d)$ contains a free k-subalgebra on two generators, and thus so does the larger division algebra $K(x; \sigma)$. □

We close this section with a few remarks related to the main theorem.

Remark 3.5. Consider an arbitrary (possibly countable) base field k, and k-automorphism $\sigma : K \to K$, where K contains an element a lying on an infinite σ-orbit. Although our methods do not in full generality allow us to conclude that $K(x; \sigma)$ contains a free subalgebra on two generators, we may always conclude that $K(t_1, \ldots, t_n)(x; \sigma)$ contains a free subalgebra on two generators for some n, where the t_i are commuting indeterminates and we extend σ to $K(t_1, \ldots, t_n)$ by setting $\sigma(t_i) = t_i$ for all i. To see this, note that adjoining an uncountable set of indeterminates $\{t_\alpha\}$ to K, we can consider $K(\{t_\alpha\})(x; \sigma)$ as an algebra over the uncountable field $k(\{t_\alpha\})$. Then Theorem 1.1 applies and shows that $K(\{t_\alpha\})(x; \sigma)$ contains a free subalgebra on two generators; but note that these generators live in $K(t_1, \ldots, t_n)(x; \sigma)$ for some finite subset $\{t_1, \ldots, t_n\}$ of $\{t_\alpha\}$.

In characteristic 0, we can do even better and conclude that $K(t)(x; \sigma)$ contains a free subalgebra. This easily follows from an application of Lemma 2.4 to the discrete valuation v of $K(t)$ corresponding to the maximal ideal $(t - a)$ of $K[t]$. In [Bell and Rogalski 2011, Theorem 2.6], we show how this idea provides an alternate proof of Theorem 1.1 when k is uncountable of characteristic 0, since when k is uncountable, we prove that a division k-algebra D contains a free subalgebra on two generators if and only if $D(t)$ does [Bell and Rogalski 2011, Proposition 2.1].

Remark 3.6. The division rings $K(x; \sigma)$ really can be only locally PI rather than PI. For example, let $K = \mathbb{C}(y_1, y_2, \ldots)$ be a function field in infinitely many indeterminates and define an automorphism $\sigma : K \to K$ which fixes $\mathbb{C}$ and has

$\sigma(y_n) = \zeta_n y_n$ for a primitive n-th root of unity ζ_n. If $K_n = \mathbb{C}(y_n)$, then it is easy to check that $K_n(x; \sigma)$ is a subdivision ring of $K(x; \sigma)$ with PI degree exactly n.

4. The derivation case

Assume Notation 2.1, with $K = D$ a field extension of k, $\sigma = 1$, and $\delta : K \to K$ a k-derivation, so $E = \{a \in K \mid \delta(a) = 0\}$. In this section, we show that $K(x; \delta)$ satisfies the free subalgebra conjecture. Since $K(x; \delta)$ is an E-algebra, as usual we can and do replace the base field k by the central subfield E. In fact, the analysis of $K(x; \delta)$ is much easier than the automorphism case. In characteristic 0, this reduces rather trivially to the case of the Weyl algebra (as other authors have also observed). So our main contribution here is to consider the characteristic p case.

Theorem 4.1. *Let $\delta : K \to K$ be a derivation of a field, where $k = \{a \in K \mid \delta(a) = 0\}$.*

(1) *If* $\operatorname{char} k = 0$, *then $K(x; \delta)$ contains a free subalgebra if and only if $\delta \neq 0$.*

(2) *If* $\operatorname{char} k = p > 0$, *then $K(x; \delta)$ contains a free subalgebra if and only if there is an element $a \in K$ such that setting $F_i = k\big(a, \delta(a), \dots, \delta^{i-1}(a)\big)$, one has $F_i \subsetneq F_{i+1}$ for all $i \geq 0$.*

Proof. (1) If $\delta \neq 0$, say $\delta(a) \neq 0$, take $y = a$ and $z = x(\delta(a))^{-1}$. Then $yz - zy = 1$, so D contains a copy of the Weyl algebra $A_1(k)$, and hence a free algebra on two generators by Makar-Limanov's original result [1983]. On the other hand, if $\delta = 0$, then $K(x; \delta) \cong K(x)$ is commutative and cannot contain a noncommutative free subalgebra.

(2) Since $\operatorname{char} k = p$, note that $\delta(b^p) = 0$ for all $b \in K$, so k contains all p-th powers. Now fix $a \in K$ and consider the fields $F_i = k\big(a, \delta(a), \dots, \delta^{i-1}(a)\big)$. For each $i \geq 1$, if $F_{i-1} \subsetneq F_i$, then since $[\delta^{i-1}(a)]^p \in k \subseteq F_{i-1}$, we must have $[F_i : F_{i-1}] = p$ and $F_{i-1} \subseteq F_i$ is a purely inseparable simple extension.

Suppose that $F_{i-1} \subsetneq F_i$ for all $i \geq 1$, let $F = \bigcup_{i \geq 0} F_i$, and note that F is closed under δ; so it is enough to prove that $F(x; \delta)$ contains a free subalgebra. Write $b_i = \delta^i(a)$. By the analysis of the previous paragraph, it easily follows that F has a k-basis consisting of all words in the b_i of the form

$$\{b_0^{e_1} b_1^{e_2} \dots b_m^{e_m} \mid 0 \leq e_j \leq p - 1\}.$$

Now we apply the criterion of Theorem 2.2, with the choice $b = b_0 = a$. Thus it is sufficient to prove the claim that if $u \in F$ satisfies $\delta(u) \in k + kb$, then $u \in k$. To obtain this claim, suppose that u satisfies $\delta(u) \in k + kb$ with $u \notin k$, so there exists some $d \geq 0$ such that $u \in F_{d+1} \setminus F_d$. We can write u as

$$u = \sum_{i=0}^{p-1} u_i b_d^i,$$

where each $u_i \in F_d$ and $u_i \neq 0$ for some $i > 0$. Thus we see that

$$\delta(u) = \sum_{i=0}^{p-1} i u_i b_d^{i-1} b_{d+1} + \sum_{i=0}^{p-1} \delta(u_i) b_d^i,$$

where the second sum is contained in F_{d+1}. Since some $u_i \neq 0$ with $i \neq 0$, we have $\delta(u) \notin F_{d+1}$, contradicting the assumption $\delta(u) \in k + kb$. This proves the claim, and so $K(x; \delta)$ contains a free algebra on two generators.

On the other hand, suppose that for some $a \in K$, the sequence of fields

$$F_i = k\big(a, \delta(a), \ldots, \delta^{i-1}(a)\big)$$

has $F_i = F_{i+1}$ for some i. Then it is easy to see that $F_n = F_i$ for all $n \geq i$, and so F_i is a δ-invariant subfield, of finite degree over k. If this happens for every $a \in K$, then every finite subset of K is contained in a δ-invariant subfield F of finite degree over k, and so is contained in the PI division ring $F(x; \delta)$. Thus $K(x; \delta)$ is locally PI and does not contain a noncommutative free subalgebra. □

Proof of Theorem 1.2. Examining the proofs of parts (1) and (2) of Theorem 4.1, we see that $K(x; \delta)$ contains a free subalgebra if and only if it is not locally PI. □

An interesting example of part (2) of Theorem 4.1 is obtained by taking $K = \mathbb{F}_p(x_0, x_1, \ldots)$ to be a rational function field in infinitely many indeterminates over the field of p elements, and defining $\delta(x_i) = x_{i+1}$ for all $i \geq 0$. The ring $K(x; \delta)$ then contains a free algebra in two generators over $\mathbb{F}_p$. This ring has appeared before in the literature and has other interesting properties. In particular, Resco and Small [1993] studied this ring as an example of a noetherian affine algebra which becomes non-noetherian after base field extension.

5. Summary theorems

In this section, we apply our results to show that the free subalgebra conjecture holds for a large class of algebras formed from iterated Ore extensions. We state our summary theorems over an uncountable field for convenience, though they hold over an arbitrary field whenever the iterated Ore extension is built out of extensions satisfying Theorem 1.1(2).

Before proving our main theorem, we make an easy observation. The reason that we have not yet considered Ore extensions with both an automorphism and derivation is the following fact.

Lemma 5.1. *Let D be a PI division algebra with automorphism σ and σ-derivation δ. Then $D[x; \sigma, \delta]$ is isomorphic either to $D[x'; \sigma']$ for some other automorphism σ', or else to $D[x'; \delta']$ for some derivation δ'.*

Proof. This is presumably well-known, but we sketch the proof since it is elementary. Let $Z = Z(D)$. For any $a \in D$, $b \in Z$, we have $\delta(ab) = \delta(ba)$, and so

$$\sigma(a)\delta(b) + \delta(a)b = \sigma(b)\delta(a) + \delta(b)a. \tag{5.7}$$

Since $\sigma(b) \in Z$ also, we have $\delta(a)[b - \sigma(b)] = \delta(b)a - \sigma(a)\delta(b)$. Then if there is any $b \in Z$ such that $\sigma(b) \neq b$, we must have $\delta(a) = (b - \sigma(b))^{-1}[\delta(b)a - \sigma(a)\delta(b)]$ for all $a \in D$. In this case, making the change of variable $x' = x + (b - \sigma(b))^{-1}\delta(b)$, one easily checks that $D[x; \sigma, \delta] \cong D[x'; \sigma, 0]$.

Otherwise, $\sigma(b) = b$ for all $b \in Z$; in other words, σ is trivial on the center. By the Skolem–Noether theorem, σ is an inner automorphism of D, say $\sigma(a) = d^{-1}ad$ for all a and some $d \in D^{\times}$. Then the change of variable $x' = dx$ gives

$$D[x; \sigma, \delta] \cong D[x'; 1, d\delta]. \qquad \square$$

Theorem 5.2. *Let k be an uncountable field. The following results hold*:

(1) *Let A be any PI domain which is a k-algebra with automorphism σ and σ-derivation δ (over k). Then the quotient division algebra of $A[x; \sigma, \delta]$ satisfies the free subalgebra conjecture.*

(2) *If A is any affine k-algebra which is an Ore domain such that $Q(A)$ satisfies the free subalgebra conjecture, then $Q(A[x; \sigma, \delta])$ also satisfies the conjecture.*

Proof. (1) Let D be the quotient division algebra of A, so that R has quotient ring $Q(R) = D(x; \sigma, \delta)$. By Lemma 5.1, it is enough to consider the two special cases $D(x; \sigma)$ and $D(x; \delta)$. If $K = Z(D)$, then σ restricts to K and D is finite over K since D is PI. It is easy to see that it is enough to prove the free subalgebra conjecture for $K(x; \sigma)$. Now Theorem 1.1 gives the result.

Similarly, considering $D(x; \delta)$, we have $\delta(K) \subseteq K$ (use (5.7)), and so we easily reduce to the case of $K(x; \delta)$. We are done by Theorem 1.2.

(2) If A is not locally PI, then by assumption, $Q(A)$ contains a free subalgebra on two generators. Then there is an embedding $Q(A) \subseteq Q(A[x; \sigma, \delta])$, and of course $Q(A[x; \sigma, \delta])$ is also not locally PI, so we are done in this case. If instead A is locally PI, then it is actually PI by the assumption that A is affine. Now part (1) applies. $\square$

Proof of Theorem 1.3. An easy induction using parts (1) and (2) of Theorem 5.2 shows that any iterated Ore extension of an affine PI domain over an uncountable field has a quotient division algebra satisfying the free subalgebra conjecture. $\square$

Acknowledgments

We thank James Zhang, Sue Sierra, Agata Smoktunowicz, Toby Stafford, and Lance Small for many valuable discussions. We thank Jairo Gonçalves for pointing out an error in a previous version of the paper. Finally, we are grateful to the referee for providing the proof of Proposition 3.2, for suggesting Remark 3.5, and for many helpful comments which improved the exposition of the paper.

References

[Bell and Rogalski 2011] J. P. Bell and D. Rogalski, "Free subalgebras of division algebras over uncountable fields", preprint, 2011. arXiv 1112.0041

[Bell et al. 2010] J. Bell, D. Rogalski, and S. J. Sierra, "The Dixmier–Moeglin equivalence for twisted homogeneous coordinate rings", *Israel J. Math.* **180** (2010), 461–507. MR 2011m:14005 Zbl 1219.16030

[Chiba 1996] K. Chiba, "Free subgroups and free subsemigroups of division rings", *J. Algebra* **184**:2 (1996), 570–574. MR 97g:16030 Zbl 0865.16014

[Figueiredo et al. 1996] L. M. V. Figueiredo, J. Z. Gonçalves, and M. Shirvani, "Free group algebras in certain division rings", *J. Algebra* **185**:2 (1996), 298–313. MR 97j:16030 Zbl 0866.16011

[Goncalves and Shirvani 1996] J. Goncalves and M. Shirvani, "On free group algebras in division rings with uncountable center", *Proc. Amer. Math. Soc.* **124**:3 (1996), 685–687. MR 96f:16020 Zbl 0844.16012

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Krause and Lenagan 2000] G. R. Krause and T. H. Lenagan, *Growth of algebras and Gelfand–Kirillov dimension*, 2nd ed., Graduate Studies in Mathematics **22**, American Mathematical Society, Providence, RI, 2000. MR 2000j:16035 Zbl 0957.16001

[Lichtman 1999] A. I. Lichtman, "Free subalgebras in division rings generated by universal enveloping algebras", *Algebra Colloq.* **6**:2 (1999), 145–153. MR 2001k:16030 Zbl 0971.16010

[Lorenz 1986] M. Lorenz, "On free subalgebras of certain division algebras", *Proc. Amer. Math. Soc.* **98**:3 (1986), 401–405. MR 87m:16040 Zbl 0615.16002

[Makar-Limanov 1983] L. Makar-Limanov, "The skew field of fractions of the Weyl algebra contains a free noncommutative subalgebra", *Comm. Algebra* **11**:17 (1983), 2003–2006. MR 84j:16012 Zbl 0521.16015

[Makar-Limanov 1984a] L. Makar-Limanov, "On free subobjects of skew fields", pp. 281–285 in *Methods in ring theory* (Antwerp, 1983), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **129**, Reidel, Dordrecht, 1984. MR 770598 Zbl 0539.16015

[Makar-Limanov 1984b] L. Makar-Limanov, "On free subsemigroups of skew fields", *Proc. Amer. Math. Soc.* **91**:2 (1984), 189–191. MR 85j:16022 Zbl 0512.16018

[Makar-Limanov 1984c] L. Makar-Limanov, "On group rings of nilpotent groups", *Israel J. Math.* **48**:2-3 (1984), 244–248. MR 86c:16006 Zbl 0556.16006

[Makar-Limanov and Malcolmson 1991] L. Makar-Limanov and P. Malcolmson, "Free subalgebras of enveloping fields", *Proc. Amer. Math. Soc.* **111**:2 (1991), 315–322. MR 91f:16023 Zbl 0719.16008

[Reichstein and Vonessen 1995] Z. Reichstein and N. Vonessen, "Free subgroups of division algebras", *Comm. Algebra* **23**:6 (1995), 2181–2185. MR 96a:16032 Zbl 0833.16015

[Resco and Small 1993] R. Resco and L. W. Small, "Affine Noetherian algebras and extensions of the base field", *Bull. London Math. Soc.* **25**:6 (1993), 549–552. MR 94m:16029 Zbl 0802.16014

[Rogalski 2009] D. Rogalski, "GK-dimension of birationally commutative surfaces", *Trans. Amer. Math. Soc.* **361**:11 (2009), 5921–5945. MR 2010j:14008 Zbl 1181.14005

[Shirvani and Gonçalves 1998] M. Shirvani and J. Z. Gonçalves, "Free group algebras in the field of fractions of differential polynomial rings and enveloping algebras", *J. Algebra* **204**:2 (1998), 372–385. MR 99g:16033 Zbl 0932.16029

[Shirvani and Gonçalves 1999] M. Shirvani and J. Z. Gonçalves, "Large free algebras in the ring of fractions of skew polynomial rings", *J. London Math. Soc.* (2) **60**:2 (1999), 481–489. MR 2000m:16041 Zbl 0983.16023

Communicated by Susan Montgomery
Received 2011-03-10 Revised 2012-01-06 Accepted 2012-02-07

jpb@math.sfu.ca *Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada*
http://www.math.sfu.ca/~jpb/

drogalsk@math.ucsd.edu *Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, United States*
http://www.math.ucsd.edu/~drogalsk/

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1369

Classes de cycles motiviques étales

Bruno Kahn

Résumé. Soit X une variété lisse sur un corps k et soit l un nombre premier. On construit une suite exacte

$$0 \to H^0(X, \mathcal{H}^3_{\text{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \to C_{\text{tors}} \to 0$$

où $\mathcal{H}^i_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))$ et $\mathcal{H}^i_{\text{cont}}(\mathbb{Z}_l(2))$ sont les faisceaux Zariski associés aux cohomologies étale et continue et C est le conoyau de la classe de cycle l-adique définie par Jannsen sur $CH^2(X) \otimes \mathbb{Z}_l$ si $l \neq \operatorname{car} k$ et une variante de celle-ci si $l = \operatorname{car} k$. Si $k = \mathbb{C}$, cela fournit une autre démonstration d'un théorème de Colliot-Thélène–Voisin, qui évite l'utilisation de la conjecture de Bloch–Kato en degré 3. Si k est séparablement clos et $l \neq \operatorname{car} k$, on obtient, toujours dans l'esprit de Colliot-Thélène et Voisin, une suite exacte

$$\begin{aligned} 0 \to \operatorname{Griff}^2(X, \mathbb{Z}_l)_{\text{tors}} &\to H^3_{\text{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \\ &\to H^0(X, \mathcal{H}^3_{\text{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to \operatorname{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \to 0 \end{aligned}$$

où $H^3_{\text{tr}}(X, \mathbb{Z}_l(2))$ est le quotient de la cohomologie l-adique par le premier cran de la filtration par le coniveau et $\operatorname{Griff}^2(X, \mathbb{Z}_l)$ est un groupe de Griffiths l-adique.

Si k est la clôture algébrique d'un corps fini et si X est « de type abélien » et vérifie la conjecture de Tate, alors $\operatorname{Griff}^2(X, \mathbb{Z}_l)$ est de torsion et $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est fini si $H^3_{\text{tr}}(X, \mathbb{Q}_l(2)) = 0$. D'autre part, un théorème de Schoen donne un exemple où $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est fini mais $H^3_{\text{tr}}(X, \mathbb{Q}_l(2)) \neq 0$.

An English abstract appears on the next page.

MSC2010 : primary 14C25; secondary 14E22, 14F20, 14F42, 14GXX.
Keywords : cycle class map, unramified cohomology, continuous étale cohomology, motivic cohomology.

Abstract. Let X be a smooth variety over a field k, and l be a prime number. We construct an exact sequence

$$0 \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \to C_{\mathrm{tors}} \to 0$$

where $\mathcal{H}^i_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))$ and $\mathcal{H}^i_{\mathrm{cont}}(\mathbb{Z}_l(2))$ are the Zariski sheaves associated to étale and continuous étale cohomology and C is the cokernel of Jannsen's l-adic cycle class on $CH^2(X) \otimes \mathbb{Z}_l$ if $l \neq \operatorname{char} k$ or a variant of it if $l = \operatorname{char} k$. If $k = \mathbb{C}$, this gives another proof of a theorem of Colliot-Thélène and Voisin, avoiding a recourse to the Bloch–Kato conjecture in degree 3. If k is separably closed and $l \neq \operatorname{char} k$, still in the spirit of Colliot-Thélène and Voisin we get an exact sequence

$$0 \to \operatorname{Griff}^2(X, \mathbb{Z}_l)_{\mathrm{tors}} \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \\ \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to \operatorname{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \to 0$$

where $H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2))$ is the quotient of l-adic cohomology by the first step of the coniveau filtration and $\operatorname{Griff}^2(X, \mathbb{Z}_l)$ is an l-adic Griffiths group.

If k is the algebraic closure of a finite field and X is "of abelian type" and satisfies the Tate conjecture, $\operatorname{Griff}^2(X, \mathbb{Z}_l)$ is torsion and $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ is finite provided $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l(2)) = 0$. On the other hand, a theorem of Schoen gives an example where $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ is finite but $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l(2)) \neq 0$.

1. Introduction

Soient k un corps, X une k-variété lisse et l un nombre premier différent de $\operatorname{car} k$. Uwe Jannsen [1988, Lemma 6.14] a défini une classe de cycle l-adique

$$CH^2(X) \otimes \mathbb{Z}_l \xrightarrow{\mathrm{cl}^2} H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) \tag{1-1}$$

à valeurs dans sa cohomologie étale continue. En imitant sa construction à partir d'un théorème de Geisser et Levine [2000], on obtient une variante p-adique de (1-1) si k est de caractéristique $p > 0$, où le second membre est une version continue de la cohomologie de Hodge–Witt logarithmique. Notons $\mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))$ (resp. $\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))$) le faisceau Zariski associé au préfaisceau $U \mapsto H^3_{\text{ét}}(U, \mathbb{Q}_l/\mathbb{Z}_l(2))$ (resp. $U \mapsto H^3_{\mathrm{cont}}(U, \mathbb{Z}_l(2))$). Le but de cet article est de démontrer l'analogue l-adique d'un théorème de Jean-Louis Colliot-Thélène et Claire Voisin [2010, théorème 3.6] :

Théorème 1.1. *Soit l un nombre premier quelconque et soit C le conoyau de* (1-1)*. On a une suite exacte*

$$0 \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \xrightarrow{f} H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \xrightarrow{g} C_{\mathrm{tors}} \to 0.$$

Noter que C_{tors} est fini si $H^4_{\text{ét}}(X, \mathbb{Z}_l(2))$ est un $\mathbb{Z}_l$-module de type fini : ceci se produit pour $l \neq p$ si k est fini, ou plus généralement si les groupes de cohomologie galoisienne de k à coefficients finis sont finis, par exemple (comme me l'a fait remarquer J.-L. Colliot-Thélène) si k est un corps local [supérieur] ou un corps séparablement clos. Dans ces cas, le théorème 1.1 implique donc que $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est extension d'un groupe fini par un groupe divisible. (Pour $l = p$, C_{tors} est fini si k est fini et X projective d'après [Gros et Suwa 1988a, p. 589, proposition 4.18].)

Lorsque $k = \mathbb{C}$, ceci donne une autre démonstration du théorème de Colliot-Thélène–Voisin en utilisant l'isomorphisme de comparaison entre cohomologies de Betti et l-adique et sa compatibilité aux classes de cycles respectives.

La démonstration de Colliot-Thélène et Voisin utilise l'exactitude du complexe de faisceaux Zariski de cohomologie de Betti

$$0 \to \mathcal{H}^3_{\text{cont}}(\mathbb{Z}(2)) \to \mathcal{H}^3_{\text{cont}}(\mathbb{Q}(2)) \to \mathcal{H}^3_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(2)) \to 0.$$

Son exactitude à gauche découle du théorème de Merkurjev–Suslin, c'est-à-dire la conjecture de Bloch–Kato en degré 2 ; celle à droite découle de la conjecture de Bloch–Kato en degré 3, dont la démonstration a été conclue récemment par Voevodsky et un certain nombre d'auteurs.

La démonstration donnée ici évite le recours à cette dernière conjecture : elle ne repose que sur le théorème de Merkurjev–Suslin plus un formalisme triangulé un peu sophistiqué, mais dont, je pense, la sophistication est bien inférieure aux ingrédients de la preuve du théorème de Voevodsky et al.

Son principe est le suivant. La classe de cycle (1-1) se prolonge en une classe « étale »

$$H^4_{\text{ét}}(X, \mathbb{Z}(2)) \otimes \mathbb{Z}_l \to H^4_{\text{ét}}(X, \mathbb{Z}_l(2)) \tag{1-2}$$

où le terme de gauche est un groupe de cohomologie motivique étale de X. Le théorème de comparaison de la cohomologie motivique étale à coefficients finis avec la cohomologie étale des racines de l'unité tordues ou de Hodge–Witt logarithmique (théorème 2.6 a) et b)) implique que (1-2) est de noyau divisible et de conoyau sans torsion. On en déduit une surjection g de noyau divisible dans le théorème 1.1 à l'aide de la suite exacte

$$0 \to CH^2(X) \to H^4_{\text{ét}}(X, \mathbb{Z}(2)) \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(2))) \to 0,$$

qui est rappelée/établie dans la proposition 2.9. Ceci est fait au §3B. La détermination du noyau est plus technique et je renvoie au §3F pour les détails.

Pour justifier la structure du noyau et du conoyau de (1-2), il faut considérer le « cône » de l'application classe de cycle : ceci est expliqué au §3A.

On obtient de plus des renseignements supplémentaires sur le groupe

$$H^0(X, \mathcal{H}^3_{\text{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} :$$

1. Si k est fini et X projective lisse, dans la classe $B_{\mathrm{Tate}}(k)$ de [Kahn 2003, définition 1 b)] ce groupe est nul (§5A).

2. Si k est séparablement clos et $l \neq \operatorname{car} k$, toujours dans l'esprit de [Colliot-Thélène et Voisin 2010] on a une suite exacte (corollaire 4.7) :

$$0 \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \to 0$$

   où $\mathrm{Griff}^2(X, \mathbb{Z}_l)$ est le groupe des cycles de codimension 2 à coefficients l-adiques, modulo l'équivalence algébrique, dont la classe de cohomologie l-adique est nulle, et $H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2))$ est le quotient de $H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2))$ par le premier cran de la filtration par le coniveau.

   Comme le groupe $H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2)))$ est sans torsion (lemme 3.12), on en déduit une suite exacte (proposition 4.12) :

$$\begin{aligned} 0 \to \mathrm{Griff}^2(X, \mathbb{Z}_l)_{\mathrm{tors}} \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \\ \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \to 0. \end{aligned}$$

3. Si k est la clôture algébrique d'un corps fini k_0 et que X provient de la classe $B_{\mathrm{Tate}}(k_0)$ de [Kahn 2003], le groupe $\mathrm{Griff}^2(X, \mathbb{Z}_l)$ est de torsion et la suite exacte ci-dessus se simplifie en :

$$0 \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to 0$$

   (théorème 5.2).

   En particulier, $H^0(X, \mathcal{H}^3_{\mathrm{\acute{e}t}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est fini dès que $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l(2)) = 0$ (ceci est conjecturalement vrai sur tout corps séparablement clos, cf. remarque 4.13), mais un exemple de Schoen montre que cette condition n'est pas nécessaire (proposition 5.5 et théorème 5.6).

2. Groupes de Chow supérieurs

Cette section comporte presque exclusivement des rappels sur les groupes de Chow supérieurs : le lecteur au courant peut la parcourir rapidement. Elle a pour but principal de fournir une preuve complète de la proposition 2.9, évitant le complexe $\Gamma(2)$ de Lichtenbaum.

2A. ***Groupes de Chow supérieurs.*** Soit k un corps. Bloch [1986] associe à tout k-schéma algébrique X des complexes de groupes abéliens $z^n(X, \cdot)$ ($n \geq 0$), concentrés en degrés (homologiques) ≥ 0 : rappelons qu'on pose

$$\Delta^p = \operatorname{Spec} k[t_0, \dots, t_p]/\left(\sum t_i - 1\right),$$

que $z^n(X, p)$ est le groupe abélien libre sur les fermés intègres de codimension n de $X \times_k \Delta^p$ qui rencontrent les faces proprement, et que la différentielle d_p est obtenue comme somme alternée des intersections avec les faces de dimension $p-1$. Les groupes d'homologie $CH^n(X, p)$ de $z^n(X, \bullet)$ sont les *groupes de Chow supérieurs* de X : on a $CH^n(X, 0) = CH^n(X)$ par construction.

Les $z^n(X, \bullet)$ sont contravariants pour les morphismes plats, en particulier étales ; ils définissent en fait des complexes de faisceaux sur le petit site étale d'un schéma lisse X donné. Ils sont aussi covariants pour les morphismes propres, en particulier pour les immersions fermées.

Si Y est un fermé de X, on notera ici

$$\begin{aligned} z^n_Y(X, \bullet) &= \mathrm{Fib}\big(z^n(X, \bullet) \xrightarrow{j^*} z^n(X-Y, \bullet)\big), \\ CH^n_Y(X, p) &= H_p(z^n_Y(X, \bullet)), \end{aligned}$$

où $j : X - Y \to X$ est l'immersion ouverte complémentaire et Fib désigne la fibre homotopique (décalé du *mapping cone*). Si on tensorise par un groupe abélien A, on écrit $CH^n_Y(X, p, A)$.

On a le théorème fondamental suivant, qui est une vaste généralisation du lemme de déplacement de Chow ([Bloch 1986, théorèmes 3.1 et 4.1], preuves corrigées dans [Bloch 1994]) :

Théorème 2.1. a) *Les groupes de Chow supérieurs sont contravariants pour les morphismes quelconques de but lisse entre variétés quasi-projectives. Ils commutent aux limites projectives filtrantes à morphismes de transition affines.*

b) *Soient X un k-schéma quasi-projectif équidimensionnel, $i : Y \to X$ un fermé équidimensionnel et $j : U \to X$ l'ouvert complémentaire. Soit d la codimension de Z dans X. Alors le morphisme naturel*

$$z^{n-d}(Y, \bullet) \xrightarrow{i_*} z^n_Y(X, \bullet)$$

est un quasi-isomorphisme.

c) *On dispose de produits d'intersection*

$$CH^m(X, p) \times CH^n(X, q) \to CH^{m+n}(X, p+q) \tag{2-1}$$

pour X quasi-projectif lisse.

De la partie b) de ce théorème, on déduit que pour tout groupe abélien A, la théorie cohomologique à supports

$$(X, Z) \mapsto CH^n_Z(X, \bullet, A)$$

définie sur la catégorie des k-schémas quasi-projectifs lisses [Colliot-Thélène et al. 1997, Definition 5.1.1 a)] vérifie l'axiome COH1 de loc. cit., p. 53. D'après loc. cit., théorème 7.5.2, on a donc des isomorphismes pour tout (n, p)

$$CH^n(X, p, A) \xrightarrow{\sim} H^{-p}_{\mathrm{Zar}}(X, z^n(-, \cdot) \otimes A) \xrightarrow{\sim} H^{-p}_{\mathrm{Nis}}(X, z^n(-, \cdot) \otimes A)$$

pour X quasi-projectif lisse.

Si X est seulement lisse, le second isomorphisme persiste.

Définition 2.2. Soit X un k-schéma lisse (essentiellement de type fini), et soit τ une topologie de Grothendieck moins fine que la topologie étale sur la catégorie des k-schémas lisses de type fini : en pratique $\tau \in \{\mathrm{Zar}, \mathrm{Nis}, \text{ét}\}$. On note $A_X(n)_\tau$ le complexe de faisceaux $z^n(-, \cdot) \otimes A[-2n]$ sur X_τ et $H^*_\tau(X, A(n))$ l'hypercohomologie de X à coefficients dans le complexe $A_X(n)$ (pour la topologie τ). C'est la *cohomologie motivique de poids n à coefficients dans* A pour la topologie concernée.

Pour simplifier, on supprime τ de la notation si $\tau = \mathrm{Zar}$ ou Nis (voir ci-dessus).

On a donc un isomorphisme, pour X quasi-projectif lisse :

$$CH^n(X, 2n - i) \xrightarrow{\sim} H^i(X, A(n)). \tag{2-2}$$

On a

$$\mathbb{Z}_X(0) = \mathbb{Z}, \tag{2-3}$$

$$\mathbb{Z}_X(1) \simeq \mathbb{O}^*_X[-1]. \tag{2-4}$$

([Bloch 1986, corollaire 6.4] pour le second quasi-isomorphisme).

L'isomorphisme (2-4) se généralise ainsi :

Théorème 2.3 [Nesterenko et Suslin 1989; Totaro 1992]. *Supposons que* $X =$ Spec K, *où* K *est un corps. L'isomorphisme* (2-4) *et les produits* (2-1) *induisent des isomorphismes*

$$K^M_n(K) \xrightarrow{\sim} H^n(K, \mathbb{Z}(n)),$$

$$K^M_n(K)/m \xrightarrow{\sim} H^n(K, \mathbb{Z}/m(n)),$$

pour $m > 0$, *où* K^M_n *désigne la* K*-théorie de Milnor.*

Démonstration. Voir les travaux cités pour le premier énoncé ; le second s'en déduit puisque $H^{n+1}(K, \mathbb{Z}(n)) = 0$. □

Remarque 2.4. L'isomorphisme (2-2) vaut pour $i \geq 2n$, même si X n'est pas quasi-projectif. En effet, le terme de droite est l'aboutissement d'une suite spectrale de coniveau [Colliot-Thélène et al. 1997, Remark 5.1.3 (3)] qui, grâce au théorème 2.1, prend la forme suivante :

$$E_1^{p,q} = \bigoplus_{x \in X^{(p)}} H^{q-p}(k(x), A(n-p)) \Rightarrow H^{p+q}(X, A(n)).$$

On a $H^i(F, A(r)) = 0$ pour $i > r$ et tout corps F, car $A_F(r)$ est un complexe concentré en degrés $\leq r$. On en déduit déjà que $H^i(X, A(n)) = 0$ pour $i > 2n$, et on a évidemment $CH^n(X, p, A) = 0$ pour $p < 0$. Quant à $H^{2n}(X, A(n))$, il s'insère dans une suite exacte

$$E_1^{n-1,n} \xrightarrow{d_1} E_1^{n,n} \to H^{2n}(X, A(n)) \to 0$$

qui s'identifie à la suite exacte

$$\bigoplus_{x \in X_{(1)}} k(x)^* \otimes A \xrightarrow{\mathrm{Div}} \bigoplus_{x \in X_{(0)}} A \to CH^n(X) \otimes A \to 0$$

via (2-3) et (2-4) (l'identification de la différentielle d_1 à l'application diviseurs est facile à partir du théorème 2.1 appliqué pour $n = 1$). D'autre part, on calcule aisément que $CH^n(X, 0, A) = CH^n(X) \otimes A$ sans supposer X quasi-projectif.

Le lemme suivant raffine une partie de la remarque 2.4 : sa démonstration est moins élémentaire.

Lemme 2.5. *Le complexe $\mathbb{Z}_X(n)$ est concentré en degrés $\leq n$: autrement dit, $\mathcal{H}^i(\mathbb{Z}_X(n)) = 0$ pour $i > n$.* □

Démonstration. La théorie cohomologique à supports

$$(X, Y) \mapsto H^*_Y(X, \mathbb{Z}(n))$$

vérifie les axiomes COH1 et COH3 de [Colliot-Thélène et al. 1997] : le premier, « excision étale », résulte facilement du théorème 2.1 b) et le second, invariance par homotopie, est démontré dans [Bloch 1986, théorème 2.1]. Il résulte alors de [Colliot-Thélène et al. 1997, Corollary 5.1.11] qu'elle vérifie la conjecture de Gersten (c'est déjà démontré dans [Bloch 1986, théorème 10.1]). En particulier, les faisceaux $\mathcal{H}^i(\mathbb{Z}(n))$ s'injectent dans leur fibre générique, et on est réduit au cas évident d'un corps de base. □

2B. *Comparaisons.* À partir de maintenant, X désigne un k-schéma lisse.

Théorème 2.6. a) *Si m est inversible dans k, il existe un quasi-isomorphisme canonique*

$$(\mathbb{Z}/m)_X(n)_{\text{ét}} \xrightarrow{\sim} \mu_m^{\otimes n}.$$

b) *Si k est de caractéristique $p > 0$ et $r \geq 1$, il existe un quasi-isomorphisme canonique*

$$(\mathbb{Z}/p^r)_X(n)_{\text{ét}} \xrightarrow{\sim} \nu_r(n)[-n]$$

où $\nu_r(n)$ est le n-ème faisceau de Hodge–Witt logarithmique.

c) *Soit α la projection de $X_{\text{ét}}$ sur X_{Zar}. Alors la flèche d'adjonction*

$$\mathbb{Q}_X(n) \to R\alpha_*\mathbb{Q}_X(n)_{\text{ét}}$$

est un isomorphisme.

Démonstration. a) et b) sont dus à Geisser–Levine : a) est [2001, Theorem 1.5] et b) est [2000, Theorem 8.5] (si k est parfait : voir le corollaire A.7 en général).[1] Quant à c), c'est un fait général pour un complexe de faisceaux Zariski C de $\mathbb{Q}$-espaces vectoriels sur un schéma normal S : on se ramène au cas où S est local et où $C = A[0]$ est un faisceau concentré en degré 0. Alors A est un faisceau constant de $\mathbb{Q}$-espaces vectoriels et cela résulte de [Deninger 1988, Theorem 2.1]. □

Définition 2.7. Soit $n \geq 0$. Pour l premier différent de $\operatorname{car} k$, on note

$$\mathbb{Q}_l/\mathbb{Z}_l(n) = \varinjlim_r \mu_{l^r}^{\otimes n}.$$

Pour $l = \operatorname{car} k$, on note

$$\mathbb{Q}_l/\mathbb{Z}_l(n) = \varinjlim_r \nu_r(n)[-n].$$

Enfin, on note

$$\mathbb{Q}/\mathbb{Z}(n) = \bigoplus_l \mathbb{Q}_l/\mathbb{Z}_l(n).$$

C'est un objet de la catégorie dérivée des groupes abéliens sur le gros site étale de $\operatorname{Spec} k$.

Le théorème 2.6 montre qu'on a un isomorphisme, pour tout X lisse sur k :

$$(\mathbb{Q}/\mathbb{Z})_X(n)_{\text{ét}} \xrightarrow{\sim} (\mathbb{Q}/\mathbb{Z})(n)_{|X}. \tag{2-5}$$

Nous utiliserons cette identification dans la suite sans mention ultérieure.

On a alors :

Corollaire 2.8. *Pour tout $i > n+1$, l'homomorphisme de faisceaux Zariski*

$$\mathcal{H}^{i-1}(R\alpha_*\mathbb{Q}/\mathbb{Z}(n)) \to \mathcal{H}^i(R\alpha_*\mathbb{Z}_X(n)_{\text{ét}})$$

émanant du théorème 2.6 a) et b) est un isomorphisme.

1. Cette dernière référence indique que l'hypothèse « X lisse sur k » devrait être remplacée par « X régulier de type fini sur k » dans une grande partie de ce texte.

Démonstration. Dans la suite exacte de faisceaux Zariski

$$\mathcal{H}^{i-1}(R\alpha_*\mathbb{Q}_X(n)_{\text{ét}}) \to \mathcal{H}^{i-1}(R\alpha_*\mathbb{Q}/\mathbb{Z}(n)) \to \mathcal{H}^{i}(R\alpha_*\mathbb{Z}_X(n)_{\text{ét}}) \to \mathcal{H}^{i}(R\alpha_*\mathbb{Q}_X(n)_{\text{ét}})$$

les deux termes extrêmes sont nuls d'après le théorème 2.6 c) et le lemme 2.5. □

2C. ***Cohomologie étale de complexes non bornés.*** Si X est un schéma de dimension cohomologique étale *a priori* non finie et si C est un complexe de faisceaux étales sur X, non borné inférieurement, la considération de l'hypercohomologie $H^*_{\text{ét}}(X, C)$ soulève au moins trois difficultés : 1) une définition en forme ; 2) la commutation aux limites ; 3) la convergence de la suite spectrale d'hypercohomologie.

Le premier problème est maintenant bien compris : il faut prendre une résolution K-injective, ou fibrante, de C ; voir par exemple [Spaltenstein 1988, Theorem 4.5 et Remark 4.6].

Le second et le troisième problèmes sont plus délicats. Dans le cas de $\mathbb{Z}(n)$, le second et implicitement le troisième est résolu dans [Kahn 1997, §B.3 p. 1114] (pour la cohomologie motivique de Suslin–Voevodsky). Rappelons l'argument : en utilisant le théorème 2.6, on peut insérer $R\alpha_*\mathbb{Z}_X(n)_{\text{ét}}$ dans un triangle exact

$$R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \to \mathbb{Q}_X(n) \to R\alpha_*(\mathbb{Q}/\mathbb{Z})_X(n)_{\text{ét}} \xrightarrow{+1} .$$

Si X est de dimension de Krull finie, l'hypercohomologie Zariski du second terme se comporte bien, et celle du troisième terme aussi puisque c'est l'hypercohomologie étale d'un complexe borné.

2D. ***Conjecture de Beilinson–Lichtenbaum.*** Cette conjecture concerne la comparaison entre $H^*(X, A(n))$ et $H^*_{\text{ét}}(X, A(n))$, pour $A = \mathbb{Z}/m$, cf. [Geisser et Levine 2001, Theorem 1.6]. Si m est une puissance d'un nombre premier $l \neq \operatorname{car} k$, elle est équivalente d'après [Geisser et Levine 2001] à la conjecture de Bloch–Kato en poids n (pour le nombre premier l) ; donc en poids 2, au théorème de Merkurjev–Suslin. Sur un corps de caractéristique zéro, ceci avait été antérieurement démontré dans [Suslin et Voevodsky 2000].

De plus, pour $l = \operatorname{car} k$, une version de cette conjecture est démontrée par Geisser et Levine [2000] ; voir théorème A.5. En ajoutant à tout ceci le théorème 2.6 c), la conjecture de Beilinson–Lichtenbaum en poids n se retraduit en un triangle exact [Voevodsky 2003, Theorem 6.6]

$$\mathbb{Z}_X(n) \to R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \to \tau_{\geq n+2}R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \to \mathbb{Z}_X(n)[1]. \tag{2-6}$$

Ce triangle exact contient l'énoncé (« Hilbert 90 en poids n ») :

$$\mathcal{H}^{n+1}(R\alpha_*\mathbb{Z}_X(n)_{\text{ét}}) = 0.$$

2E. ***Une suite exacte.***

Proposition 2.9. *Notons $\mathcal{H}^3_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(2))$ le faisceau Zariski associé au préfaisceau $U \mapsto H^3_{\text{ét}}(U, \mathbb{Q}/\mathbb{Z}(2))$. Pour toute k-variété lisse X, on a une suite exacte courte :*

$$0 \to CH^2(X) \to H^4_{\text{ét}}(X, \mathbb{Z}(2)) \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(2))) \to 0. \tag{2-7}$$

Démonstration. En prenant l'hypercohomologie de Zariski de X à valeurs dans le triangle (2-6) pour $n = 2$, on trouve une suite exacte

$$0 \to H^4(X, \mathbb{Z}(2)) \to H^4_{\text{ét}}(X, \mathbb{Z}(2)) \to H^0(X, R^4\alpha_*\mathbb{Z}(2)_{\text{ét}}) \to H^5(X, \mathbb{Z}(2)).$$

D'après la remarque 2.4, on a $H^4(X, \mathbb{Z}(2)) = CH^2(X)$ et $H^5(X, \mathbb{Z}(2)) = 0$. D'autre part, le triangle exact

$$\mathbb{Z}_X(2)_{\text{ét}} \to \mathbb{Q}_X(2)_{\text{ét}} \to (\mathbb{Q}/\mathbb{Z})_X(2)_{\text{ét}}$$

provenant du théorème 2.6 donne une longue suite exacte de faisceaux

$$R^3\alpha_*\mathbb{Q}_X(2)_{\text{ét}} \to R^3\alpha_*(\mathbb{Q}/\mathbb{Z})_X(2)_{\text{ét}} \to R^4\alpha_*\mathbb{Z}_X(2)_{\text{ét}} \to R^4\alpha_*\mathbb{Q}_X(2)_{\text{ét}}.$$

On a $R^3\alpha_*\mathbb{Q}_X(2)_{\text{ét}} = R^4\alpha_*\mathbb{Q}_X(2)_{\text{ét}} = 0$ puisque $\mathbb{Z}_X(2)$ est concentré en degrés ≤ 2, d'après le théorème 2.6 c). Ce qui conclut, via l'isomorphisme (2-5). □

Remarques 2.10. 1) La suite exacte (2-7) apparaît dans [Kahn 1996, théorème 1.1, équation (9)], avec $\mathbb{Z}(2)$ remplacé par le complexe de Lichtenbaum $\Gamma(2)$; à la 2-torsion près, elle est déjà chez Lichtenbaum [1990, Theorem 2.13 et Remark 2.14]. Il est probable qu'on a un isomorphisme

$$\Gamma(2, X) \simeq \tau_{\geq 1}\left(z^2(X, \cdot)[-4]\right) \tag{2-8}$$

dans $D(X_{\text{Zar}})$ pour tout k-schéma lisse X.[2] Une fonctorialité suffisante de cet isomorphisme impliquerait qu'il peut s'étalifier. Dans [Block 1995, Theorem 7.2], un isomorphisme (2-8) est construit pour $X = \operatorname{Spec} k$. Mais (2-8) ne semble pas apparaître dans la littérature en général.

2) En se reposant sur la conjecture de Bloch–Kato en poids 3, on obtient de la même manière une suite exacte

$$0 \to H^2(X, \mathcal{K}_3) \to H^5_{\text{ét}}(X, \mathbb{Z}(3)) \to H^0(X, \mathcal{H}^4_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(3)) \\ \to CH^3(X) \to H^6_{\text{ét}}(X, \mathbb{Z}(3)).$$

Cette suite apparaît dans [Kahn 2003, remarque 4.10], sauf que le premier terme est $H^5(X, \mathbb{Z}(3))$; son identification avec $H^2(X, \mathcal{K}_3)$ se fait à l'aide de la suite spectrale de coniveau de la remarque 2.4.

2. Par ailleurs, la conjecture de Beilinson–Soulé prédit que $\mathbb{Z}(2) \to \tau_{\geq 1}\mathbb{Z}(2)$ est un quasi-isomorphisme, mais elle n'a pas d'importance pour ce travail.

2F. ***D'autres suites exactes.***

Proposition 2.11. *On a des suites exactes*

$$0 \to H^3(X, \mathbb{Z}/m(2)) \to H^3_{\text{ét}}(X, \mathbb{Z}/m(2)) \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Z}/m(2))) \\ \to CH^2(X) \otimes \mathbb{Z}/m \to H^4_{\text{ét}}(X, \mathbb{Z}/m(2))$$

$(m > 0)$,

$$0 \to H^3(X, \mathbb{Q}/\mathbb{Z}(2)) \to H^3_{\text{ét}}(X, \mathbb{Q}/\mathbb{Z}(2)) \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(2))) \\ \to CH^2(X) \otimes \mathbb{Q}/\mathbb{Z} \to H^4_{\text{ét}}(X, \mathbb{Q}/\mathbb{Z}(2)).$$

Démonstration. Elles s'obtiennent comme dans la preuve de la proposition 2.9 en prenant la cohomologie des triangles exacts

$$\mathbb{Z}/m(2) \to R\alpha_*(\mathbb{Z}/m)_{\text{ét}}(2) \to \tau_{\geq 3} R\alpha_*(\mathbb{Z}/m)_{\text{ét}}(2) \xrightarrow{+1}$$

$$\mathbb{Q}/\mathbb{Z}(2) \to R\alpha_*(\mathbb{Q}/\mathbb{Z})_{\text{ét}}(2) \to \tau_{\geq 3} R\alpha_*(\mathbb{Q}/\mathbb{Z})_{\text{ét}}(2) \xrightarrow{+1}$$

obtenus en tensorisant (2-6) par $\mathbb{Z}/m$ ou $\mathbb{Q}/\mathbb{Z}$ au sens dérivé. □

On reconnaît donc dans $H^3(X, \mathbb{Z}/m(2))$ le groupe $NH^3_{\text{ét}}(X, \mathbb{Z}/m(2))$ de Suslin [1987, §4]. On peut sans doute montrer que la seconde suite exacte coïncide avec celle de [Colliot-Thélène et al. 1983, p. 790, Remark 2].

3. Cohomologie l-adique et p-adique

Dans cette section, k est un corps quelconque, de caractéristique $p \geq 0$.

3A. ***Classe de cycle l-adique et p-adique.*** Soit l un nombre premier quelconque. Pour toute k-variété lisse X, on a des applications « classe de cycle l-adique »

$$H^i_{\text{ét}}(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l \to H^i_{\text{cont}}(X, \mathbb{Z}_l(n)). \tag{3-1}$$

Ces homomorphismes proviennent d'un morphisme de complexes (dans la catégorie dérivée de la catégorie des complexes de faisceaux sur $X_{\text{ét}}$)

$$\mathbb{Z}_X(n)_{\text{ét}} \overset{L}{\otimes} \mathbb{Z}_l \to \mathbb{Z}_l(n)^c_X \tag{3-2}$$

où

$$\mathbb{Z}_l(n)^c_X = \begin{cases} R \varprojlim \mu_{l^r}^{\otimes n} & \text{si } l \neq p \\ R \varprojlim \nu_r(n)[-n] & \text{si } l = p. \end{cases}$$

Cette construction est décrite dans [Kahn 2002, §1.4, en particulier (1.8)] pour $l \neq p$ et dans [Kahn 2003, §3.5] pour $l = p$. Elle repose sur celles de Geisser–Levine aux crans finis ([2001] pour $l \neq p$ et [2000, démonstration du théorème 8.3] pour $l = p$).

Remarque 3.1. Pour $l \neq p$ et $i = 2n$, la composée de (3-1) avec l'homomorphisme $CH^n(X) \otimes \mathbb{Z}_l \to H^{2n}_{\text{ét}}(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l$ (2-7) n'est autre que la classe de cycle de Jannsen [1988, Lemma 6.14] : cela résulte de la construction même dans [Geisser et Levine 2001] de l'isomorphisme du théorème 2.6 a). Pour $l = p$, il est moins clair que (3-1) soit compatible avec la classe de cycle de Gros [1985, p. 50, définition 4.1.7 et p. 55, proposition 4.2.33]. Cela doit pouvoir se vérifier directement ; comme je n'en aurai pas besoin, je laisse cet « exercice »aux lecteurs intéressés.

Notons les isomorphismes évidents :

$$\mathbb{Z}_l(n)^c_X \overset{L}{\otimes} \mathbb{Z}/l^r \xrightarrow{\sim} \begin{cases} \mu_{l^r}^{\otimes n} & \text{si } l \neq p \\ \nu_r(n)[-n] & \text{si } l = p. \end{cases} \tag{3-3}$$

Définition 3.2. On note respectivement $K_X(n)_{\text{ét}}$ et $K_X(n)$ le choix d'un cône du morphisme (3-2) et du morphisme composé

$$\mathbb{Z}_X(n) \overset{L}{\otimes} \mathbb{Z}_l \to R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \overset{L}{\otimes} \mathbb{Z}_l \to R\alpha_*\mathbb{Z}_l(n)^c_X$$

de sorte qu'on a un morphisme

$$K_X(n) \to R\alpha_* K_X(n)_{\text{ét}}$$

compatible avec le morphisme $\mathbb{Z}_X(n) \overset{L}{\otimes} \mathbb{Z}_l \to R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \overset{L}{\otimes} \mathbb{Z}_l$.

Remarque 3.3. Rappelons que $K_X(n)$ et $K_X(n)_{\text{ét}}$ ne sont uniques qu'à isomorphisme non unique près ; le morphisme $K_X(n) \to R\alpha_* K_X(n)_{\text{ét}}$ n'a pas non plus d'unicité particulière. En particulier, ces choix ne sont fonctoriels en X que pour les immersions ouvertes : cela suffira pour nos besoins ici. Toutefois, on pourrait faire des choix plus rigides (fonctoriels pour les morphismes quelconques entre schémas lisses), quitte à travailler dans des catégories de modèles convenables.

En vertu du théorème 2.6, (3-3) implique immédiatement :

Proposition 3.4. *Le morphisme* (3-2)$\otimes\mathbb{Z}/l^r$ *est un isomorphisme pour tout entier* $r \geq 1$. *Autrement dit, les faisceaux de cohomologie de* $K_X(n)_{\text{ét}}$ *sont uniquement* l*-divisibles.* □

Corollaire 3.5. *Pour tout* (X, i, n), *le noyau de* (3-1) *est* l*-divisible et son conoyau est sans* l*-torsion.*

Démonstration. On a une suite exacte

$$H^{i-1}_{\text{ét}}(X, K(n)) \to H^i_{\text{ét}}(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l \to H^i_{\text{cont}}(X, \mathbb{Z}_l(n)) \to H^i_{\text{ét}}(X, K(n))$$

(où $H^*_{\text{ét}}(X, K(n)) := \mathbb{H}^*_{\text{ét}}(X, K_X(n)_{\text{ét}})$), dont les termes extrêmes sont uniquement divisibles. □

3B. ***Démonstration du théorème 1.1 : première partie.*** On va démontrer :

Proposition 3.6. *Soit C le conoyau de* (1-1). *On a une surjection*

$$H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \twoheadrightarrow C_{\text{tors}} \tag{3-4}$$

de noyau divisible.

Démonstration. Utilisons la suite exacte (2-7) : en chassant dans le diagramme commutatif de suites exactes

$$\begin{array}{ccccccccc}
0 \to K & \longrightarrow & CH^2(X) \otimes \mathbb{Z}_l & \longrightarrow & H^4_{\text{cont}}(X, \mathbb{Z}_l(2)) & \longrightarrow & C \to 0 \\
\downarrow & & \downarrow & & {\scriptstyle =}\downarrow & & \downarrow \\
0 \to K_{\text{ét}} & \longrightarrow & H^4_{\text{ét}}(X, \mathbb{Z}(2)) \otimes \mathbb{Z}_l & \longrightarrow & H^4_{\text{cont}}(X, \mathbb{Z}_l(2)) & \longrightarrow & C_{\text{ét}} \to 0
\end{array}$$

(définissant K, $K_{\text{ét}}$, C et $C_{\text{ét}}$), on en déduit une suite exacte

$$0 \to K \to K_{\text{ét}} \to H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \to C \to C_{\text{ét}} \to 0. \tag{3-5}$$

On conclut en utilisant le corollaire 3.5. □

3C. ***Suites spectrales de coniveau.*** Soit C un complexe de faisceaux Zariski sur X. Par une technique bien connue remontant à Grothendieck et Hartshorne (cf. [Colliot-Thélène et al. 1997, 1.1]) on obtient une suite spectrale

$$E_1^{p,q} = \bigoplus_{x \in X^{(p)}} H_x^{p+q}(X, C) \Rightarrow H^{p+q}(X, C).$$

Cette suite spectrale est clairement naturelle en $C \in D(X_{\text{Zar}})$. On notera de manière suggestive :

$$E_2^{p,q} = A^p(X, H^q(C))$$

de sorte qu'on a des morphismes « edge »

$$H^n(X, C) \to A^0(X, H^n(C)). \tag{3-6}$$

Lorsque C vérifie la conjecture de Gersten, on a des isomorphismes canoniques

$$A^p(X, H^q(C)) \simeq H^p(X, \mathcal{H}^q(C)).$$

D'après [Colliot-Thélène et al. 1997, Corollary 5.1.11] c'est le cas pour $C = R\alpha_* \mathbb{Z}_l(n)^c_X$. En effet, pour $l \neq p$, la théorie cohomologique à supports correspondante vérifie les axiomes COH1 (excision étale, on dit maintenant Nisnevich) et COH3 (invariance par homotopie) de [Colliot-Thélène et al. 1997] ; pour $l = p$, elle vérifie COH1 et COH5. Ce dernier axiome est la « formule du fibré projectif » : il résulte de [Gros 1985]. Si k est un corps fini, il faut adjoindre à ces axiomes l'axiome COH6 de [Colliot-Thélène et al. 1997, p. 64] (existence de transferts pour les extensions finies) : il est standard.

C'est également le cas pour $C = \mathbb{Z}_X(n)$, cf. preuve du lemme 2.5. Par contre ce n'est pas clair pour $C = K_X(n)$: en effet, la règle $(X, Y) \mapsto H^*_Y(X, K(n))$ ne définit pas une théorie cohomologique à supports sans un choix cohérent des cônes $K_X(n)$. Plus précisément, cette règle n'est pas a priori fonctorielle en (X, Y) pour les morphismes quelconques de paires. On ne peut donc pas lui appliquer la théorie de Bloch–Ogus–Gabber développée dans [Colliot-Thélène et al. 1997]. La considération des suites spectrales de coniveau nous permettra de contourner ce problème.

3D. ***Un encadrement de la cohomologie non ramifiée.*** L'identification du noyau de (3-4) est plus délicate. À titre préparatoire, on va pousser l'analyse du §3A un peu plus loin en faisant intervenir la conjecture de Bloch–Kato en degré n.

Par l'axiome de l'octaèdre (et la conjecture de Bloch–Kato, cf. (2-6)), on a un diagramme commutatif de triangles distingués dans $D(X_{\mathrm{Zar}})$, où $K_X(n)$ et $K_X(n)_{\text{ét}}$ ont été introduits dans la définition 3.2 :

$$\begin{tikzcd}
\mathbb{Z}_X(n) \overset{L}{\otimes} \mathbb{Z}_l \arrow[r] \arrow[d] & R\alpha_*\mathbb{Z}_l(n)^c_X \arrow[r] \arrow[d, "="'] & K_X(n) \arrow[d, "f"'] \\
R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \overset{L}{\otimes} \mathbb{Z}_l \arrow[r] \arrow[d] & R\alpha_*\mathbb{Z}_l(n)^c_X \arrow[r] \arrow[d] & R\alpha_*K_X(n)_{\text{ét}} \arrow[d] \\
\tau_{\geq n+2}R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \overset{L}{\otimes} \mathbb{Z}_l \arrow[r] & 0 \arrow[r] & C
\end{tikzcd}$$

et où C est par définition « le » cône de f. On a donc un zig-zag d'isomorphismes

$$C \xrightarrow{\sim} \tau_{\geq n+2}R\alpha_*\mathbb{Z}_X(n)_{\text{ét}} \overset{L}{\otimes} \mathbb{Z}_l[1] \xleftarrow{\sim} \tau_{\geq n+1}R\alpha_*(\mathbb{Q}_l/\mathbb{Z}_l)(n)_{\text{ét}}$$

où l'isomorphisme de gauche provient du diagramme ci-dessus, et celui de droite provient du corollaire 2.8. D'où un triangle exact

$$K_X(n) \to R\alpha_*K_X(n)_{\text{ét}} \to \tau_{\geq n+1}R\alpha_*(\mathbb{Q}_l/\mathbb{Z}_l)(n)_{\text{ét}} \xrightarrow{+1} .$$

Comme le deuxième terme est uniquement divisible (proposition 3.4) et que le troisième est de torsion, cela montre que

$$K_X(n) \otimes \mathbb{Q} \xrightarrow{\sim} R\alpha_*K_X(n)_{\text{ét}}. \tag{3-7}$$

On en déduit :

Lemme 3.7. *Soit $l \neq p$. Sous la conjecture de Bloch–Kato en degré n, le groupe $H^i(X, K(n))$ est uniquement divisible pour $i \leq n$ et on a une suite exacte courte*

$$0 \to H^{n+1}(X, K(n)) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathcal{H}^{n+1}(\mathbb{Q}_l/\mathbb{Z}_l(n))) \to H^{n+2}(X, K(n))_{\text{tors}} \to 0.$$

Le même énoncé vaut pour $l = p$, en utilisant le théorème A.5. □

Le point est maintenant d'identifier les termes extrêmes de la suite du lemme 3.7 à des groupes plus concrets : nous n'y parvenons que pour $n = 2$ au §3F. Mais pour n quelconque, notons la suite exacte

$$0 \to \mathrm{Coker}\big(H^{n+2}(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l \to H^{n+2}_{\mathrm{cont}}(X, \mathbb{Z}_l(n))\big) \\ \to H^{n+2}(X, K(n)) \to H^{n+3}(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l \quad (3\text{-}8)$$

et les homomorphismes évidents :

$$H^{n+1}(X, K(n)) \xrightarrow{\alpha} A^0(X, H^{n+1}(K(n))) \xleftarrow{\beta} A^0(X, H^{n+1}_{\mathrm{cont}}(\mathbb{Z}_l(n))) \qquad (3\text{-}9)$$

où α est l'homomorphisme (3-6). La suite exacte (3-8) en induit une sur les sous-groupes de torsion. Pour $n = 2$, le dernier terme est nul : on retrouve ainsi la proposition 3.6. Pour $n > 2$, la première flèche de (3-8) n'a plus de raison d'être surjective sur la torsion. Notons tout de même que pour $n = 3$, le dernier terme de (3-8) n'est autre que $CH^3(X) \otimes \mathbb{Z}_l$ (comparer aux remarques 2.10).

D'autre part :

Lemme 3.8. *L'homomorphisme β de* (3-9) *est un isomorphisme pour tout $n \geq 0$.*

Démonstration. Dans le diagramme commutatif

$$\begin{array}{ccc} \bigoplus_{x \in X^{(0)}} H^{n+1}_x(X, \mathbb{Z}_l(n)) & \longrightarrow & \bigoplus_{x \in X^{(1)}} H^{n+2}_x(X, \mathbb{Z}_l(n)) \\ \downarrow & & \downarrow \\ \bigoplus_{x \in X^{(0)}} H^{n+1}_x(X, K(n)) & \longrightarrow & \bigoplus_{x \in X^{(1)}} H^{n+2}_x(X, K(n)) \end{array}$$

les deux flèches verticales sont des isomorphismes. En effet, elle s'insèrent dans des suites exactes du type

$$H^{n+1}_x(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l \to H^{n+1}_x(X, \mathbb{Z}_l(n)) \to H^{n+1}_x(X, K(n)) \to H^{n+2}_x(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l$$

$$H^{n+2}_x(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l \to H^{n+2}_x(X, \mathbb{Z}_l(n)) \to H^{n+2}_x(X, K(n)) \to H^{n+3}_x(X, \mathbb{Z}(n)) \otimes \mathbb{Z}_l$$

où dans la première suite, x est de codimension 0 et dans la seconde suite, x est de codimension 1. Sans perte de généralité, on peut supposer X connexe et alors, pour son point générique η (cf. théorème 2.1 a)) :

$$H^i_\eta(X, \mathbb{Z}(n)) := \varinjlim_U H^i(U, \mathbb{Z}(n)) \xrightarrow{\sim} H^i(k(X), \mathbb{Z}(n)) = 0 \text{ pour } i > n.$$

Pour x de codimension 1, on a

$$H^i_x(X, \mathbb{Z}(n)) := \varinjlim_{U \ni x} H^i_{Z_U}(U, \mathbb{Z}(n))$$

où $Z_U = \overline{\{x\}} \cap U$. Grâce au théorème 2.1 b) et a), cette limite devient

$$\varinjlim_{U \ni x} H^{i-2}(Z_U, \mathbb{Z}(n-1)) = H^{i-2}(k(x), \mathbb{Z}(n-1)) = 0 \text{ pour } i-2 > n-1.$$

Le lemme en découle. $\square$

3E. *Cohomologie à supports de $K(n)$.* On aura aussi besoin des deux lemmes suivants au prochain numéro :

Lemme 3.9. *Soit $l \neq \operatorname{car} k$.*
Soit $Y \subset X$ un couple lisse de codimension d. Alors il existe des isomorphismes

$$H^{i-2d}_{\mathrm{Zar}}(Y, K(n-d)) \xrightarrow{\sim} H^i_Y(X, K(n)) \quad (n \geq 0, i \in \mathbb{Z})$$

contravariants pour les immersions ouvertes $U \hookrightarrow X$.

Attention : ce lemme affirme l'existence d'isomorphismes de pureté, mais ne dit rien sur leur caractère canonique ou fonctoriel au-delà de la contravariance énoncée. (On peut faire en sorte qu'ils soient contravariants pour les morphismes quelconques entre variétés lisses, mais c'est plus technique et inutile ici.)

Démonstration. Notons i l'immersion fermée $Y \to X$. On remarque que le diagramme de $D(Y_{\mathrm{Zar}})$

$$\begin{tikzcd}
\mathbb{Z}(n-d)_Y \otimes \mathbb{Z}_l[-2d] \arrow[r, "\mathrm{cl}^{n-d}_Y"] \arrow[d, "f"'] & R\alpha_*\mathbb{Z}_l(n-d)^c_Y[-2d] \arrow[d, "g"'] \\
Ri^!_{\mathrm{Zar}}\mathbb{Z}(n)_X \otimes \mathbb{Z}_l \arrow[r, "\mathrm{cl}^n_X"] & R\alpha_* Ri^!_{\text{ét}}\mathbb{Z}_l(n)^c_X
\end{tikzcd}$$

où f est induit par le théorème 2.1 b) et g est donné par le théorème de pureté en cohomologie étale, est commutatif : cela résulte tautologiquement de la construction des classes de cycles motiviques dans [Geisser et Levine 2001]. Par conséquent, ce diagramme s'étend en un diagramme commutatif de triangles exacts

$$\begin{tikzcd}
\mathbb{Z}(n-d)_Y \otimes \mathbb{Z}_l[-2d] \arrow[r, "\mathrm{cl}^{n-d}_Y"] \arrow[d, "f"'] & R\alpha_*\mathbb{Z}_l(n-d)^c_Y[-2d] \arrow[r] \arrow[d, "g"'] & K(n-d)_Y[-2d] \arrow[r, "+1"] \arrow[d, "h"'] & {} \\
Ri^!_{\mathrm{Zar}}\mathbb{Z}(n)_X \otimes \mathbb{Z}_l \arrow[r, "\mathrm{cl}^n_X"] & R\alpha_* Ri^!_{\text{ét}}\mathbb{Z}_l(n)^c_X \arrow[r] & Ri^!_{\mathrm{Zar}}K(n)_X \arrow[r, "+1"] & {}
\end{tikzcd}$$

(Rien n'est dit sur un choix privilégié de h.) Comme f et g sont des quasi-isomorphismes, h en est un aussi, d'où l'énoncé.

Comme h est un morphisme dans la catégorie dérivée des faisceaux Zariski sur Y, la contravariance annoncée est tautologique pour $U \hookrightarrow X$ tel que $U \cap Y \neq \varnothing$, et elle est sans contenu lorsque $U \cap Y = \varnothing$. $\square$

Lemme 3.10. *Soit $l = p = \operatorname{car} k$. Soit $Y \subset X$ un couple lisse de codimension d. Alors il existe des homomorphismes*

$$H^{i-2d}(Y, K(n-d)) \xrightarrow{h^i} H^i_Y(X, K(n)) \quad (n \geq 0, i \in \mathbb{Z})$$

contravariants pour les immersions ouvertes $U \hookrightarrow X$. Ce sont des isomorphismes pour $i \leq n+d$.

Démonstration. On raisonne comme dans la démonstration du lemme 3.9, en utilisant cette fois le théorème 2.6 b) et les résultats de Gros [1985]. D'après [Gros 1985, (3.5.3) et théorème 3.5.8], on a

$$R^q i^! \nu_r(n) = \begin{cases} 0 & \text{si } q \neq d, d+1 \\ \nu_r(n-d) & \text{si } q = d. \end{cases} \tag{3-10}$$

La formule (3-10) et sa compatibilité aux classes de cycles motiviques (elle est à la base de leur construction) fournit un diagramme commutatif dans $D(Y_{\mathrm{Zar}})$

$$\begin{array}{ccc} \mathbb{Z}(n-d)_Y \otimes \mathbb{Z}_p & \longrightarrow & R\alpha_*\mathbb{Z}_p(n-d)^c[-2d] \\ {\scriptstyle f}\downarrow & & {\scriptstyle g}\downarrow \\ Ri^!_{\mathrm{Zar}}\mathbb{Z}(n)_X \otimes \mathbb{Z}_p & \longrightarrow & R\alpha_* Ri^!_{\text{ét}}\mathbb{Z}_p(n)^c. \end{array} \tag{3-11}$$

Il en résulte un morphisme

$$K(n-d)_Y \xrightarrow{h} Ri^!_{\mathrm{Zar}} K(n)_X$$

complétant le carré ci-dessus en un diagramme commutatif de triangles exacts. Ceci fournit les homomorphismes h^i du lemme.

Dans le diagramme (3-11), f est un isomorphisme (théorème 2.1 b)). Par (3-10), le cône de g est acyclique en degrés $\leq n+d$. Par conséquent, le cône de h est acyclique en degrés $\leq n+d$, ce qui donne la bijectivité de h^i pour $i \leq n+d$. □

3F. *Fin de la démonstration du théorème 1.1.* On prend maintenant $n = 2$. Le résultat principal est :

Proposition 3.11. *Pour $n = 2$, l'homomorphisme α de* (3-9) *est surjectif de noyau* $A^1(X, H^2(K(2)))$, *uniquement divisible.*

Démonstration. Notons $E_2^{a,b} = A^a(X, H^b(K(n))$: c'est la cohomologie d'un certain complexe de Cousin.

En utilisant les lemmes 3.9 et 3.10, on trouve que $E_1^{a,b} = 0$ pour

$l \neq p$: $a > 2$ et $a+b < 2a$; $a = 2$ et $a+b \leq 4$.

$l = p$: $a > 2$ et $a+b < 2+a$; $a = 2$ et $a+b \leq 4$.

(En particulier, $E_2^{2,2}=0$ puisque $K(0)=0$!) On en déduit une suite exacte

$$0\to A^1(X,H^2(K(2)))\to H^3(X,K(2))\to A^0(X,H^3(K(2)))\to 0. \qquad (3\text{-}12)$$

Mais $A^1(X,H^2(K(2)))$ est l'homologie du complexe

$$E_1^{0,2}\to E_1^{1,2}\to E_1^{2,2}$$

dont tous les termes sont encore dans le domaine d'application des lemmes 3.9 et 3.10 (isomorphismes de pureté). D'après le lemme 3.7, ils sont uniquement divisibles, ainsi donc que $E_2^{1,2}$. □

Le théorème 1.1 résulte maintenant de la proposition 3.6, du lemme 3.7, du lemme 3.8 et de la proposition 3.11. □

3G. *Un complément.* Notons pour conclure :

Lemme 3.12. *Le $\mathbb{Z}_l$-module $H^0(X,\mathscr{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2)))$ est sans torsion.*

Démonstration. On fait comme dans [Colliot-Thélène et Voisin 2010, théorème 3.1] (cet argument remonte à [Bloch et Srinivas 1983] pour la cohomologie de Betti) : le théorème de Merkurjev–Suslin implique que le faisceau $\mathscr{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))$ est sans torsion. (Pour $l=p$, utiliser [Geisser et Levine 2000].) □

4. Cas d'un corps de base séparablement clos

Soient k un corps séparablement clos et X une k-variété lisse. On veut préciser le théorème 1.1 dans ce cas, toujours dans l'esprit de [Colliot-Thélène et Voisin 2010].

4A. *Lien avec les cycles de Tate.* Le lemme suivant est démontré dans [Colliot-Thélène et Kahn 2011]. Il montre que les cycles de Tate entiers fournissent un bon analogue des cycles de Hodge entiers considérés dans [Colliot-Thélène et Voisin 2010] :

Lemme 4.1. *Soient G un groupe profini et M un $\mathbb{Z}_l$-module de type fini muni d'une action continue de G. Soit*

$$M^{(1)}=\bigcup_U M^U$$

où U décrit les sous-groupes ouverts de G. Alors $M/M^{(1)}$ est sans torsion.

On en déduit :

Lemme 4.2. *Supposons que k soit la clôture séparable d'un corps de type fini et que $l\neq \operatorname{car} k$. Alors le groupe fini C_{tors} du théorème 1.1 est aussi le sous-groupe de torsion de*

$$\operatorname{Coker}\left(CH^2(X)\otimes\mathbb{Z}_l\to H^4_{\mathrm{cont}}(X,\mathbb{Z}_l(2))^{(1)}\right).$$

Sous la conjecture de Tate, ce conoyau est entièrement de torsion (*pour X non nécessairement propre*, *cf.* [Jannsen 1990, p. 114, Theorem 7.10 a)]). □

On peut d'ailleurs supprimer l'hypothèse que k soit la clôture séparable d'un corps de type fini. En général, écrivons $k=\bigcup_\alpha k_\alpha$, où k_α décrit l'ensemble (ordonné filtrant) des clôtures séparables des sous-corps de type fini de k sur lesquels X est définie. Pour tout α, notons X_α le k_α-modèle de X correspondant. Si $k_\alpha \subset k_\beta$, on a des isomorphismes

$$H^4_{\mathrm{cont}}(X_\alpha, \mathbb{Z}_l(2)) \xrightarrow{\sim} H^4_{\mathrm{cont}}(X_\beta, \mathbb{Z}_l(2)) \xrightarrow{\sim} H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))$$

et on peut définir

$$H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))^{(1)} := \varinjlim_\alpha H^4_{\mathrm{cont}}(X_\alpha, \mathbb{Z}_l(2))^{(1)}.$$

Il s'agit en fait d'une limite inductive d'isomorphismes puisque, si X est définie sur $k^0_\alpha \subset k_\alpha$ de type fini et de clôture séparable k_α et que $k_\beta \supset k_\alpha$, l'homomorphisme $\mathrm{Gal}(k_\beta/k_\beta k^0_\alpha) \to \mathrm{Gal}(k_\alpha/k^0_\alpha)$ est surjectif.

4B. ***Lien avec le groupe de Griffiths.*** Si $k=\mathbb{C}$ et X est projective, Colliot-Thélène et Voisin ont établi un lien entre $H^0(X, \mathcal{H}^3(\mathbb{Q}/\mathbb{Z}(2)))$ et le groupe de Griffiths dans [Colliot-Thélène et Voisin 2010, §4.2]. Reprenons cette idée en l'amplifiant un peu.

Voici d'abord une définition de groupes de Griffiths et de groupes d'équivalence homologique dans le contexte l-adique. Supposons k séparablement clos si $l \neq \operatorname{car} k$, et k algébriquement clos si $l = \operatorname{car} k$. Par un argument bien connu [Bloch et Ogus 1974, Lemma 7.10 ; Bloch 2010, Lecture 1, Lemma 1.3]), pour toute k-variété lisse X, le sous-groupe de $CH^n(X)$ formé des cycles algébriquement équivalents à zéro est l-divisible ; les diagrammes commutatifs

$$\begin{array}{ccc} CH^n(X)\otimes\mathbb{Z}_l & \xrightarrow{\mathrm{cl}^n} & H^{2n}_{\mathrm{cont}}(X, \mathbb{Z}_l(n)) \\ \downarrow & & \downarrow \\ CH^n(X)\otimes\mathbb{Z}/l^s & \xrightarrow{\mathrm{cl}^n_s} & H^{2n}_{\mathrm{\acute{e}t}}(X, \mathbb{Z}/l^s(n)) \end{array}$$

et l'isomorphisme

$$H^{2n}_{\mathrm{cont}}(X, \mathbb{Z}_l(n)) \xrightarrow{\sim} \varprojlim_s H^{2n}_{\mathrm{\acute{e}t}}(X, \mathbb{Z}/l^s(n)) \tag{4-1}$$

montrent donc que cl^n_s et cl^n se factorisent à travers l'équivalence algébrique.

(Précisons. L'isomorphisme (4-1) est valable pourvu que le système projectif $(H^{2n-1}_{\mathrm{\acute{e}t}}(X, \mathbb{Z}/l^s(n)))_{s\geq 1}$ soit de Mittag-Leffler. Pour $l \neq \operatorname{car} k$ c'est vrai parce que les termes sont finis ; pour $l = \operatorname{car} k$ et X projective c'est expliqué dans [Colliot-Thélène et al. 1983, p. 783], donc il faut a priori supposer X projective dans ce cas.)

Ceci donne un sens à :

Définition 4.3. Soit X une k-variété lisse où k est séparablement clos si $l \neq \operatorname{car} k$ et algébriquement clos si $l = \operatorname{car} k$; dans ce dernier cas, on suppose aussi X projective. Soit $A \in \{\mathbb{Z}_l, \mathbb{Q}_l, \mathbb{Z}/l^n, \mathbb{Q}_l/\mathbb{Z}_l\}$. On note

$$\operatorname{Griff}^n(X, A) = \operatorname{Ker}\big(A^n_{\mathrm{alg}}(X) \otimes A \xrightarrow{\mathrm{cl}^n} H^{2n}_{\mathrm{cont}}(X, A(n))\big)$$
$$A^n_{\mathrm{hom}}(X, A) = \operatorname{Im}\big(A^n_{\mathrm{alg}}(X) \otimes A \xrightarrow{\mathrm{cl}^n} H^{2n}_{\mathrm{cont}}(X, A(n))\big).$$

Remarque 4.4. Si $k = \mathbb{C}$, on a $\operatorname{Griff}^n(X, \mathbb{Z}_l) = \operatorname{Griff}^n(X) \otimes \mathbb{Z}_l$ par l'isomorphisme de comparaison entre cohomologies de Betti et l-adique, où $\operatorname{Griff}^n(X)$ est défini à l'aide de la cohomologie de Betti.

On a la version l-adique de [Bloch et Ogus 1974, Theorem 7.3] :

Proposition 4.5. *Supposons $l \neq \operatorname{car} k$. Notons $A^n_{\mathrm{alg}}(X)$ le groupe des cycles de codimension n sur X modulo l'équivalence algébrique. Dans la suite spectrale de coniveau*

$$E_r^{p,q} \Rightarrow H^{p+q}(X, \mathbb{Z}_l(n))$$

pour la cohomologie l-adique de X, on a un isomorphisme

$$A^n_{\mathrm{alg}}(X) \otimes \mathbb{Z}_l \xrightarrow{\sim} E_2^{n,n}$$

induit par l'isomorphisme

$$Z^n(X) \otimes \mathbb{Z}_l \xrightarrow{\sim} E_1^{n,n}$$

donné par les isomorphismes de pureté.

Démonstration. C'est la même que celle de [Bloch et Ogus 1974, preuve du théorème 7.3], *mutatis mutandis*. Plus précisément, la première étape est identique. Dans la deuxième étape, on remplace la désingularisation à la Hironaka des cycles de codimension n de X par une désingularisation à la de Jong [1996, Theorem 4.1] ; pour obtenir des résultats entiers, on utilise le théorème de Ofer Gabber (travail en cours) disant qu'on peut trouver une telle désingularisation de degré premier à l. Enfin, l'argument transcendant de Bloch–Ogus pour prouver que équivalences algébrique et homologique coïncident pour les diviseurs sur une variété lisse Y est remplacé par le suivant : par [Bloch et Ogus 1974, Lemma 7.10], le noyau de $CH^n(Y) \to A^n_{\mathrm{alg}}(Y)$ est l-divisible, donc les suites exactes de Kummer

$$\operatorname{Pic}(X) \xrightarrow{l^n} \operatorname{Pic}(X) \to H^2(X, \mathbb{Z}/l^n(1))$$

définissent des injections

$$0 \to A^1_{\mathrm{alg}}(X)/l^n \to H^2(X, \mathbb{Z}/l^n(1))$$

d'où à la limite

$$0 \to A^1_{\mathrm{alg}}(X) \otimes \mathbb{Z}_l \to H^2_{\mathrm{cont}}(X, \mathbb{Z}_l(1))$$

puisque $A^1_{\mathrm{alg}}(X)$ est un $\mathbb{Z}$-module de type fini. □

Notons que ces arguments ne nécessitent pas que X soit projective (pour le dernier, voir [Kahn 2006, théorème 3]). Si le théorème de Gabber n'évitait pas $l = p$, la démonstration s'étendrait à ce nombre premier.

Convention 4.6. *À partir de maintenant, l est supposé différent de* car k *sauf mention expresse du contraire.* La raison essentielle pour cela est que cette restriction apparaît dans la proposition 4.5 (cf. le commentaire ci-dessus).

Corollaire 4.7 [Colliot-Thélène et Voisin 2010, Theorem 2.7]. *On a une suite exacte*

$$H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) \xrightarrow{\alpha} H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \to 0.$$

Démonstration. Cela résulte de la suite exacte

$$H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) \to E_2^{0,3} \to E_2^{2,2} \xrightarrow{c} H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))$$

provenant de la suite spectrale de Bloch–Ogus en poids 2, de l'identification de $E_2^{0,3}$ à $H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2)))$, de celle de $E_2^{2,2}$ à $A^2_{\mathrm{alg}}(X) \otimes \mathbb{Z}_l$ (proposition 4.5) et de celle de c à l'application classe de cycle. □

L'analogue complexe du corollaire suivant devrait figurer dans [Colliot-Thélène et Voisin 2010] :

Corollaire 4.8. a) *On a une suite exacte, modulo des groupes finis* :

$$H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathcal{H}^3(\mathbb{Q}_l/\mathbb{Z}_l(2))) \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \to 0$$

(*cf. définition 4.3*).

b) *Le groupe $H^0(X, \mathcal{H}^3(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est dénombrable.*

c) *Si* car $k = 0$, *il existe X/k projective lisse telle que son corang soit infini pour l convenable.*

(Précisons : « modulo des groupes finis » signifie « dans la localisation de la catégorie des groupes abéliens relative à la sous-catégorie épaisse des groupes abéliens finis ».)

Démonstration. a) résulte du théorème 1.1 et du corollaire 4.7. Pour b), le terme de gauche dans la suite de a) est de cotype fini, donc dénombrable, et le terme de droite l'est aussi (propriété classique des cycles modulo l'équivalence algébrique). Enfin, d'après [Schoen 2002], on a des exemples de X et de nombres premiers l (même sur $\bar{\mathbb{Q}}$) où $\mathrm{Griff}^2(X)/l$ est infini ; en utilisant [Colliot-Thélène et Voisin

2010, Proposition 4.1] (voir aussi corollaire 4.15 du présent article), on en déduit que $\mathrm{Griff}^2(X) \otimes \mathbb{Q}_l/\mathbb{Z}_l$ est de corang infini. $\square$

4C. ***Quelques calculs de groupes de torsion.*** On veut maintenant préciser le corollaire 4.8 a) en décrivant explicitement le noyau de l'application

$$H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z}.$$

Définition 4.9. Soit A un $\mathbb{Z}_l$-module de la forme $\mathbb{Z}_l, \mathbb{Q}_l, \mathbb{Z}/l^n, \mathbb{Q}_l/\mathbb{Z}_l$. On note $NH^3_{\mathrm{cont}}(X, A)$ le premier cran de la filtration par le coniveau sur $H^3_{\mathrm{cont}}(X, A)$ et

$$H^3_{\mathrm{tr}}(X, A) = \frac{H^3_{\mathrm{cont}}(X, A)}{NH^3_{\mathrm{cont}}(X, A)}.$$

Si on a un twist à la Tate, on note $NH^3_{\mathrm{cont}}(X, A(n)) := NH^3_{\mathrm{cont}}(X, A)(n)$.

Notons que $NH^3_{\mathrm{cont}}(X, A(2)) = H^3(X, A(2))$ (cohomologie motivique de Nisnevich) pour $A = \mathbb{Z}/l^n$ ou $\mathbb{Q}_l/\mathbb{Z}_l$, d'après la proposition 2.11.

Remarque 4.10. Si X est propre, les $\mathbb{Z}_l$-modules $H^3_{\mathrm{tr}}(X, A)$ sont des invariants birationnels de X, avec l'action de $\mathrm{Gal}(k/k_0)$ si X est défini sur un sous-corps k_0 de clôture séparable k. C'est dû à Grothendieck [1968, 9.4].

Le lemme 3.12 implique :

Lemme 4.11. *Le $\mathbb{Z}_l$-module de type fini $H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2))$ est sans torsion.* $\square$

Par définition de $NH^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2))$, la suite exacte du corollaire 4.7 se raffine en une suite exacte :

$$0 \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \to 0$$

qui montre incidemment que $\mathrm{Griff}^2(X, \mathbb{Z}_l)$ est un invariant birationnel pour X propre et lisse (cf. remarque 4.10). En réutilisant le lemme 3.12, on en déduit :

Proposition 4.12. *On a une suite exacte*

$$\begin{aligned} 0 \to \mathrm{Griff}^2(X, \mathbb{Z}_l)_{\mathrm{tors}} &\to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \\ &\to H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \to 0. \square \end{aligned}$$

Remarque 4.13. Dans cette remarque, nous adoptons la convention contravariante pour les motifs purs sur un corps. Supposons que X, de dimension d, vérifie la conjecture standard C et la conjecture de nilpotence suivante : l'idéal

$$\mathrm{Ker}\big(CH^d(X \times X) \otimes \mathbb{Q} \to A^d_{\mathrm{num}}(X \times X) \otimes \mathbb{Q}\big)$$

de l'anneau des correspondances de Chow est nilpotent. Ces propriétés sont vérifiées par exemple si X est une variété abélienne : voir [Kleiman 1968] pour la première (résultat de Lieberman et Kleiman) et [Kimura 2005] pour la seconde. Alors le motif numérique de X admet une décomposition de Künneth, qui se relève en une décomposition de Chow–Künneth de son motif de Chow :

$$h(X) = \bigoplus_{i=0}^{2d} h^i(X).$$

Mais le théorème de semi-simplicité de Jannsen [1992] implique que chaque facteur numérique $h^i_{\mathrm{num}}(X)$ admet une décomposition plus fine, provenant de sa décomposition isotypique :

$$h^i_{\mathrm{num}}(X) = \bigoplus_{j=0}^{i/2} h^{i,j}_{\mathrm{num}}(X)(-j)$$

où $h^{i,j}_{\mathrm{num}}(X)$ est effectif mais aucun facteur simple de $h^{i,j}_{\mathrm{num}}(X)(1)$ n'est effectif. Cette décomposition se relève de nouveau pour donner une décomposition de Chow–Künneth raffinée (voir [Kahn et al. 2007, théorème 7.7.3]) :

$$h(X) = \bigoplus_{i=0}^{2d} \bigoplus_{j=0}^{i/2} h^{i,j}(X)(-j).$$

Notons Ab la catégorie des groupes abéliens, $\mathscr{A}$ le quotient de Ab par la sous-catégorie épaisse des groupes abéliens d'exposant fini et, pour tout anneau commutatif R, $\mathrm{Chow}(k, R)$ la catégorie des motifs de Chow à coefficients dans R. On observe que le foncteur

$$\mathrm{Hom}(-,-) : \mathrm{Chow}(k, \mathbb{Z})^{\mathrm{op}} \times \mathrm{Chow}(k, \mathbb{Z}) \to \mathrm{Ab}$$

s'étend en un foncteur

$$\mathrm{Hom}(-,-) : \mathrm{Chow}(k, \mathbb{Q})^{\mathrm{op}} \times \mathrm{Chow}(k, \mathbb{Q}) \to \mathscr{A}.$$

Par construction, $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l) = H^*_{\mathrm{cont}}(h^{3,0}(X), \mathbb{Q}_l)$. Au moins si $d = 3$, on peut montrer que, d'autre part,

$$\mathrm{Griff}^2(X, \mathbb{Z}_l) \simeq \mathrm{Hom}(h^{3,0}(X), \mathbb{L}) \otimes \mathbb{Z}_l$$

où $\mathbb{L}$ est le motif de Lefschetz et l'isomorphisme est dans $\mathscr{A}$. Ainsi, la proposition 4.12 et le théorème 1.1 montrent que (si $d = 3$) *la nullité de $h^{3,0}(X)$ entraîne la finitude de $H^0(X, \mathscr{H}^3(\mathbb{Q}_l/\mathbb{Z}_l(2)))$.*

D'autre part, la conjecture de Bloch–Beilinson–Murre [Jannsen 1994] implique que la nullité de $h^{3,0}(X)$ découle de celle de $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l)$: conjecturalement, celle-ci est donc suffisante pour impliquer la finitude du groupe $H^0(X, \mathscr{H}^3(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ (au moins si $\dim X = 3$).

Ceci est une variante de la conjecture 4.5 de [Colliot-Thélène et Voisin 2010]. On verra au théorème 5.2 c) qu'elle est vraie si k est la clôture algébrique d'un corps fini k_0 et que X provient de la classe $B_{\mathrm{Tate}}(k_0)$ de [Kahn 2003].

On peut se demander si la réciproque est vraie. Elle est fausse (théorème 5.6).

Le lemme 4.11 et la proposition 2.11 donnent un diagramme commutatif de suites exactes

$$\begin{array}{ccccccccc} 0\to & NH^3_{\mathrm{cont}}(X,\mathbb{Z}_l(2))\otimes\mathbb{Q}/\mathbb{Z} & \to & H^3_{\mathrm{cont}}(X,\mathbb{Z}_l(2))\otimes\mathbb{Q}/\mathbb{Z} & \to & H^3_{\mathrm{tr}}(X,\mathbb{Z}_l(2))\otimes\mathbb{Q}/\mathbb{Z} & \to 0 \\ & {\scriptstyle a}\downarrow & & \downarrow & & {\scriptstyle b}\downarrow & \\ 0\to & H^3_{\mathrm{Nis}}(X,\mathbb{Q}_l/\mathbb{Z}_l(2)) & \to & H^3_{\mathrm{\acute{e}t}}(X,\mathbb{Q}_l/\mathbb{Z}_l(2)) & \to & H^3_{\mathrm{tr}}(X,\mathbb{Q}_l/\mathbb{Z}_l(2)) & \to 0 \end{array} \tag{4-2}$$

dans lequel la flèche verticale centrale est injective de conoyau fini, isomorphe à $H^4_{\mathrm{cont}}(X,\mathbb{Z}_l(2))_{\mathrm{tors}}$. Par le lemme du serpent, on en déduit :

Proposition 4.14. *Avec les notations de* (4-2), *a est injective et on a une suite exacte*

$$0\to \operatorname{Ker} b\to \operatorname{Coker} a\to H^4_{\mathrm{cont}}(X,\mathbb{Z}_l(2))_{\mathrm{tors}}\to \operatorname{Coker} b\to 0.$$

Voici une application de la proposition 4.14.

Corollaire 4.15. *Soit $H^3(X,\mathbb{Q}_l/\mathbb{Z}_l(2))^0$ le noyau de la composition*

$$H^3(X,\mathbb{Q}_l/\mathbb{Z}_l(2))\to H^3_{\mathrm{\acute{e}t}}(X,\mathbb{Q}_l/\mathbb{Z}_l(2))\to H^4_{\mathrm{cont}}(X,\mathbb{Z}_l(2))_{\mathrm{tors}}.$$

Alors $\operatorname{Im} a\subset H^3(X,\mathbb{Q}_l/\mathbb{Z}_l(2))^0$ (*notations de* (4-2)) *et on a un isomorphisme*

$$\operatorname{Griff}^2(X,\mathbb{Z}_l)_{\mathrm{tors}}\xrightarrow{\sim}\operatorname{Ker} b\xrightarrow{\sim}\operatorname{Coker} a^0$$

où $a^0 : NH^3_{\mathrm{cont}}(X,\mathbb{Z}_l(2))\otimes\mathbb{Q}/\mathbb{Z}\xrightarrow{a} H^3(X,\mathbb{Q}_l/\mathbb{Z}_l(2))^0$ est l'application induite par a.

Démonstration. La première assertion est évidente. Notons a^0 l'application induite : la suite exacte de la proposition 4.14 induit donc un isomorphisme

$$\operatorname{Ker} b\xrightarrow{\sim}\operatorname{Coker} a^0.$$

D'autre part, la suite exacte de la proposition 4.12 s'insère dans un diagramme commutatif de suites exactes

$$\begin{array}{ccccc} 0\to\operatorname{Griff}^2(X,\mathbb{Z}_l)_{\mathrm{tors}} & \to & H^3_{\mathrm{tr}}(X,\mathbb{Z}_l(2))\otimes\mathbb{Q}/\mathbb{Z} & \to & H^0(X,\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2)))\otimes\mathbb{Q}/\mathbb{Z} \\ & & {\scriptstyle b}\downarrow & & {\scriptstyle c}\downarrow \\ & & H^3_{\mathrm{tr}}(X,\mathbb{Q}_l/\mathbb{Z}_l(2)) & \to & H^0(X,\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \end{array}$$

Comme le faisceau $\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))$ est sans torsion, la suite

$$0\to H^0(X,\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2)))\to H^0(X,\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Q}_l(2)))\to H^0(X,\mathcal{H}^3_{\mathrm{cont}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$$

est exacte, ce qui signifie que c est injective dans le diagramme ci-dessus. On en déduit un isomorphisme

$$\mathrm{Griff}^2(X, \mathbb{Z}_l)_{\mathrm{tors}} \xrightarrow{\sim} \operatorname{Ker} b,$$

d'où le corollaire. □

4D. ***Les homomorphismes de groupes*** $\boldsymbol{A^2_{\mathrm{hom}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}_l/\mathbb{Z}_l \to A^2_{\mathrm{hom}}(X, \mathbb{Q}_l/\mathbb{Z}_l)}$ ***et*** $\boldsymbol{\mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}_l/\mathbb{Z}_l \to \mathrm{Griff}^2(X, \mathbb{Q}_l/\mathbb{Z}_l)}$. On garde les notations de la définition 4.3.

Proposition 4.16. *On a des suites exactes*

$$A^2_{\mathrm{hom}}(X, \mathbb{Z}_l)_{\mathrm{tors}} \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}_l/\mathbb{Z}_l \to \mathrm{Griff}^2(X, \mathbb{Q}_l/\mathbb{Z}_l) \\ \to A^2_{\mathrm{hom}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}_l/\mathbb{Z}_l \to A^2_{\mathrm{hom}}(X, \mathbb{Q}_l/\mathbb{Z}_l) \to 0$$

et

$$C_{\mathrm{tors}} \to A^2_{\mathrm{hom}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}_l/\mathbb{Z}_l \to A^2_{\mathrm{hom}}(X, \mathbb{Q}_l/\mathbb{Z}_l),$$

où C est comme dans (1-1) (*voir théorème 1.1*). *En particulier*, *l'application*

$$\mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}_l/\mathbb{Z}_l \to \mathrm{Griff}^2(X, \mathbb{Q}_l/\mathbb{Z}_l)$$

est de noyau et de conoyau finis.

Démonstration. On a un diagramme commutatif de suites exactes

$$\begin{tikzcd}
A^2_{\mathrm{hom}}(X, \mathbb{Z}_l)_{\mathrm{tors}} \arrow[r] & \mathrm{Griff}^2(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \arrow[r] \arrow[d] & A^2_{\mathrm{alg}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \arrow[r] \arrow[d, "\wr"'] & A^2_{\mathrm{hom}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathrm{Griff}^2(X, \mathbb{Q}_l/\mathbb{Z}_l) \arrow[r] & A^2_{\mathrm{alg}}(X, \mathbb{Q}_l/\mathbb{Z}_l) \arrow[r] & A^2_{\mathrm{hom}}(X, \mathbb{Q}_l/\mathbb{Z}_l) \arrow[r] & 0
\end{tikzcd}$$

qui donne la première suite de la proposition, par application du lemme du serpent. Pour la seconde, on utilise le diagramme commutatif de suites exactes

$$\begin{tikzcd}
\left(\dfrac{H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))}{N^2 H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))}\right)_{\mathrm{tors}} \arrow[r] & A^2_{\mathrm{hom}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \arrow[r] \arrow[d] & H^4_{\mathrm{cont}}(X, \mathbb{Z}_l) \otimes \mathbb{Q}/\mathbb{Z} \arrow[d] \\
0 \arrow[r] & A^2_{\mathrm{hom}}(X, \mathbb{Q}_l/\mathbb{Z}_l) \arrow[r] & H^4_{\text{ét}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2))
\end{tikzcd}$$

en remarquant que la flèche verticale de droite est injective. □

4E. ***Le sous-groupe de torsion de*** $\boldsymbol{CH^2(X, \mathbb{Z}_l)_{\mathrm{alg}}}$**.** Terminons cette analyse de la torsion en déterminant celle de $CH^2(X, \mathbb{Z}_l)_{\mathrm{alg}}$, sous-groupe de $CH^2(X) \otimes \mathbb{Z}_l$ formé des classes de cycles algébriquement équivalentes à zéro, lorsque X est propre : voir corollaire 4.21. Pour cela nous avons besoin de la proposition suivante :

Proposition 4.17. *Supposons k séparablement clos, X/k propre et lisse et $i < 2n$. Soit l un nombre premier ; si $l = \operatorname{car} k$, on suppose k algébriquement clos. Alors l'image de l'application cycle* (3-1) *est égale à $H^i_{\mathrm{cont}}(X, \mathbb{Z}_l(n))_{\mathrm{tors}}$. En particulier, $H^i_{\text{ét}}(X, \mathbb{Z}(n))$ est extension d'un groupe de torsion (fini pour $l \neq p$) par un groupe divisible, et*

$$H^i_{\text{ét}}(X, \mathbb{Z}(n)) \otimes \mathbb{Q}_l/\mathbb{Z}_l = 0.$$

Démonstration. Étant donné le corollaire 3.5, il suffit de montrer que (3-1) a une image de torsion. On reprend les arguments de Colliot-Thélène et Raskind [1985] : d'après le théorème 2.1 a) et le §2C, on a

$$H^i_{\text{ét}}(X, \mathbb{Z}(n)) = \varinjlim_{\alpha} H^i_{\text{ét}}(X_\alpha, \mathbb{Z}(n))$$

où X_α parcourt un ensemble ordonné filtrant de modèles de X sur des sous-corps F_α de type fini sur le corps premier. Il suffit donc de savoir que $H^i(X, \mathbb{Z}_l(n))^G$ est de torsion, où G est le groupe de Galois absolu de $F_\alpha^{p^{-\infty}}$. On le voit en se ramenant au cas d'un corps de base fini par changement de base propre et lisse (SGA4 pour $l \neq p$, [Gros et Suwa 1988a, p. 590, théorème 2.1] pour $l = p$), où cela résulte de la démonstration par Deligne de la conjecture de Weil [Deligne 1974] pour $l \neq p$ et du complément de Katz et Messing [1974] pour $l = p$. □

Remarque 4.18. Supposons $n = 2$. En utilisant la suite spectrale de coniveau de la remarque 2.4, on obtient un isomorphisme $H^i(X, \mathbb{Z}(2)) \simeq H^{i-2}(X, \mathcal{K}_2)$: alors l'énoncé n'est autre que celui de [Colliot-Thélène et Raskind 1985, Theorem 1.8 et 2.2] pour $l \neq p$, et de [Gros et Suwa 1988a, p. 604, corollaire 2.2 et p. 605, théorème 3.1] pour $l = p$.

Corollaire 4.19. *Sous ces hypothèses, les homomorphismes*

$$H^3(X, \mathbb{Q}_l/\mathbb{Z}_l(2)) \to CH^2(X)\{l\},$$
$$H^3_{\text{ét}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2)) \to H^4_{\text{ét}}(X, \mathbb{Z}(2))\{l\}$$

sont bijectifs.

Démonstration. Pour le second, cela résulte de la suite exacte des coefficients universels et de la proposition 4.17 appliquée pour $(i, n) = (3, 2)$. Pour le premier, même raisonnement en utilisant le fait que l'homomorphisme

$$H^i(X, \mathbb{Z}(2)) \to H^i_{\text{ét}}(X, \mathbb{Z}(2))$$

est bijectif pour $i \leq 3$ par (2-6) (qui résulte en poids 2 du théorème de Merkurjev–Suslin). □

En particulier, on obtient une injection

$$H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}_l/\mathbb{Z}_l \hookrightarrow H^4_{\text{ét}}(X, \mathbb{Z}(2)). \tag{4-3}$$

Corollaire 4.20. *Sous les mêmes hypothèses, soit N le noyau de l'homomorphisme $H^4_{\text{ét}}(X, \mathbb{Z}(2)) \otimes \mathbb{Z}_l \to H^4_{\text{cont}}(X, \mathbb{Z}_l(2))$. Alors* (4-3) *induit un isomorphisme*

$$H^3_{\text{cont}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}_l/\mathbb{Z}_l \xrightarrow{\sim} N_{\text{tors}}.$$

Corollaire 4.21. *Sous les mêmes hypothèses, on a un isomorphisme canonique* :

$$CH^2(X, \mathbb{Z}_l)_{\text{alg}}\{l\} \xrightarrow{\sim} NH^3_{\text{cont}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z}.$$

Démonstration. On va réutiliser le complexe $K(2)$ de la définition 3.2. Considérons le diagramme commutatif de suites exactes :

$$\begin{tikzcd}
 & & 0 \arrow[d] & & \\
 & & H^3_{\text{cont}}(X, \mathbb{Z}_l(2))_{\text{tors}} \arrow[d] & & \\
0 \arrow[r] & NH^3_{\text{cont}}(X, \mathbb{Z}_l(2)) \arrow[r] \arrow[d, "\theta"'] & H^3_{\text{cont}}(X, \mathbb{Z}_l(2)) \arrow[r] \arrow[d] & H^0(X, \mathcal{H}^3_{\text{cont}}(\mathbb{Z}_l(2))) \arrow[r] \arrow[d, "\wr"'] & \mathrm{Griff}^2(X, \mathbb{Z}_l) \arrow[r] & 0 \\
0 \arrow[r] & H^1(X, \mathcal{H}^2(K(2))) \arrow[r] & H^3(X, K(2)) \arrow[r] \arrow[d] & H^0(X, \mathcal{H}^3_{\text{cont}}(K(2)) \arrow[r] & 0 & \\
 & & CH^2(X, \mathbb{Z}_l)_{\text{hom}} \arrow[d] & & \\
 & & 0 & &
\end{tikzcd}$$

où $CH^2(X, \mathbb{Z}_l)_{\text{hom}}$ est le noyau de la classe de cycle sur $CH^2(X) \otimes \mathbb{Z}_l$. L'exactitude à gauche de la suite verticale découle de la proposition 4.17, celle de la première suite horizontale du corollaire 4.7, la seconde suite horizontale est (3-12), enfin l'isomorphisme vertical est le lemme 3.8. La flèche θ est induite par le diagramme.

Tout d'abord, le lemme 3.12 implique via ce diagramme que

$$NH^3_{\text{cont}}(X, \mathbb{Z}_l(2))_{\text{tors}} \xrightarrow{\sim} H^3_{\text{cont}}(X, \mathbb{Z}_l(2))_{\text{tors}}.$$

Appliquons maintenant le lemme du serpent aux deux suites exactes horizontales : on obtient un isomorphisme

$$\operatorname{Ker} \theta \xrightarrow{\sim} H^3_{\text{cont}}(X, \mathbb{Z}_l(2))_{\text{tors}}$$

et une suite exacte

$$0 \to \operatorname{Coker} \theta \to CH^2(X, \mathbb{Z}_l)_{\text{hom}} \xrightarrow{\phi} \mathrm{Griff}^2(X, \mathbb{Z}_l) \to 0$$

et on calcule que ϕ est la projection naturelle. Finalement on obtient une suite exacte

$$0 \to NH^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2))/\mathrm{tors} \to H^1(X, \mathcal{H}^2(K(2))) \to CH^2(X, \mathbb{Z}_l)_{\mathrm{alg}} \to 0.$$

et l'isomorphisme du corollaire découle maintenant de la proposition 3.11 et de la suite exacte des Tor à coefficients $\mathbb{Q}/\mathbb{Z}$. □

Remarque 4.22. Si k est la clôture algébrique d'un corps fini, alors $CH^n(X, \mathbb{Z}_l)_{\mathrm{alg}}$ est un groupe de torsion pour toute k-variété projective lisse X et tout $n \geq 0$ (réduction au cas d'une courbe par l'argument de correspondances de Bloch, cf. [Schoen 1995, preuve de la proposition 2.7]). En particulier, le corollaire 4.21 décrit le groupe $CH^2(X, \mathbb{Z}_l)_{\mathrm{alg}}$ tout entier. (Voir aussi §5C.)

5. Cas d'un corps de base fini et de sa clôture algébrique

5A. ***Cas d'un corps fini.*** Soit k un corps fini. Rappelons d'abord la classe $B_{\mathrm{Tate}}(k)$ de [Kahn 2003, définition 1 b)]

Définition 5.1. Une k-variété projective lisse X est dans $B_{\mathrm{Tate}}(k)$ si

(i) Il existe une k-variété abélienne A et une extension finie k'/k telles que le motif de Chow de $X_{k'}$ à coefficients rationnels soit facteur direct de celui de $A_{k'}$.

(ii) X vérifie la conjecture de Tate (sur l'ordre des pôles de $\zeta(X, s)$ aux entiers ≥ 0).

On sait montrer qu'étant donné (i), (ii) est conséquence de (donc équivalent à) la conjecture de Tate cohomologique pour la cohomologie l-adique, pour un nombre premier l donné pouvant être égal à la caractéristique de k (cela résulte de [Kahn 2003, lemme 1.9], cf. [Colliot-Thélène et Kahn 2011, remarque 3.10]).

Considérons les notations de la preuve de la proposition 3.6. Si k est fini et si $X \in B_{\mathrm{Tate}}(k)$, $K_{\text{ét}}$ et $C_{\text{ét}}$ sont finis (ibid., théorème 3.6 et lemme 3.7), donc $K = K_{\text{ét}} = C_{\text{ét}} = 0$ et (3-5) devient un isomorphisme

$$H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2))) \xrightarrow{\sim} C. \tag{5-1}$$

En particulier, $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est fini et $H^0(X, \mathcal{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} = 0$ (théorème 1.1). En réalité, même le groupe $H^0(X, \mathcal{H}^3_{\text{ét}}(\mathbb{Q}/\mathbb{Z}(2)))$ est fini : cela résulte de la proposition 2.9 et de la génération finie de $H^4_{\text{ét}}(X, \mathbb{Z}(2))$ [Kahn 2003, corollaire 3.8 c) et e)].

Conjecturalement, toute variété projective lisse est dans $B_{\mathrm{Tate}}(k)$.

5B. ***Cas de la clôture algébrique d'un corps fini.*** Le but de ce numéro est de démontrer :

Théorème 5.2. *Soient k la clôture algébrique d'un corps fini k_0, $X_0 \in B_{\mathrm{Tate}}(k_0)$, $X = X_0 \otimes_{k_0} k$ et $l \neq \operatorname{car} k$.*

a) $\mathrm{Griff}^2(X, \mathbb{Z}_l)$ *est de torsion.*

b) *On a une suite exacte*

$$0 \to \mathrm{Griff}^2(X, \mathbb{Z}_l) \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}/\mathbb{Z} \to H^0(X, \mathscr{H}^3_{\mathrm{cont}}(\mathbb{Z}_l(2))) \otimes \mathbb{Q}/\mathbb{Z} \to 0.$$

c) *Si $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l(2)) = 0$, le groupe $H^0(X, \mathscr{H}^3(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ est fini ; dans ce cas, il est isomorphe à C_{tors}.*

Remarque 5.3. Le corollaire 4.15 donne une autre description de $\mathrm{Griff}^2(X, \mathbb{Z}_l)$.

Démonstration. a) $\Rightarrow$ b) par la proposition 4.12 et b) $\Rightarrow$ c) par le théorème 1.1. Montrons a). Soit k_1 une extension finie de k_0, et $X_1 = X_0 \otimes_{k_0} k_1$. D'après [Kahn 2003, théorème 3.6 et corollaire 3.8 e)], l'homomorphisme

$$H^4_{\text{ét}}(X_1, \mathbb{Z}(2)) \otimes \mathbb{Z}_l \to H^4_{\mathrm{cont}}(X_1, \mathbb{Z}_l(2))$$

est bijectif. D'après (2-7), l'homomorphisme

$$CH^2(X_1) \otimes \mathbb{Z}_l \to H^4_{\mathrm{cont}}(X_1, \mathbb{Z}_l(2))$$

est donc injectif. Or dans la suite exacte

$$0 \to H^1(G_1, H^3_{\mathrm{cont}}(X, \mathbb{Z}_l(2))) \to H^4_{\mathrm{cont}}(X_1, \mathbb{Z}_l(2)) \to H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))^{G_1} \to 0 \quad (5\text{-}2)$$

(où $G_1 = \mathrm{Gal}(k/k_1)$), le groupe de gauche est fini d'après Weil I [Deligne 1974]. Il en résulte que le noyau de

$$CH^2(X_1) \otimes \mathbb{Z}_l \to H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))$$

est fini pour tout k_1, d'où la conclusion en passant à la limite. □

5C. ***Un exemple de Schoen.*** J'avais initialement pensé que la réciproque du théorème 5.2 c) est vraie. En réalité elle est fausse : cela résulte d'un calcul de C. Schoen [1995]. Dans cet article, Schoen considère $X = E^3$ sur $k = \bar{\mathbb{F}}_p$, où E est la courbe elliptique d'équation $x^3 + y^3 + z^3 = 0$, et montre que, si $p \equiv 1 \pmod 3$:

$$\mathrm{Griff}^2(X)\{l\} \simeq (\mathbb{Q}_l/\mathbb{Z}_l)^2$$

pour $l \equiv -1 \pmod 3$ [Schoen 1995, Theorem 0.1]. Le groupe $\mathrm{Griff}^2(X)$ est défini comme le quotient du groupe des cycles à coefficients entiers qui sont homologiquement équivalents à zéro par le sous-groupe de ceux qui sont algébriquement équivalents à zéro. Commençons par clarifier le lien entre ce groupe et $\mathrm{Griff}^2(X, \mathbb{Z}_l)$:

Proposition 5.4. *Soient k_0 un corps fini de clôture algébrique k, $X_0 \in B_{\mathrm{Tate}}(k_0)$ et $X = X_0 \otimes_{k_0} k$. Alors, pour tout $n \geq 0$, l'homomorphisme évident*

$$\mathrm{Griff}^n(X) \otimes \mathbb{Z}_l \to \mathrm{Griff}^n(X, \mathbb{Z}_l)$$

est bijectif.

Démonstration. Soit A un groupe abélien quelconque. Pour une variété lisse X sur un corps quelconque, on peut définir les cycles à coefficients dans A modulo l'équivalence rationnelle, ou algébrique. Notons ces groupes $CH^*(X, A)$ et $A^*_{\mathrm{alg}}(X, A)$. Je dis que les homomorphismes

$$CH^*(X) \otimes A \to CH^*(X, A),$$
$$A^*_{\mathrm{alg}}(X) \otimes A \to A^*_{\mathrm{alg}}(X, A)$$

sont bijectifs : par exemple on peut décrire $A^n_{\mathrm{alg}}(X, A)$ comme le conoyau d'un homomorphisme

$$\bigoplus_{(V,v_0,v_1)} \bigoplus_{x\in((X\times V)^{(n)})'} A \xrightarrow{s_0^*-s_1^*} \bigoplus_{x\in X^{(n)}} A$$

où (V, v_0, v_1) décrit l'ensemble des classes d'isomorphismes de k-variétés lisses V munies de deux points rationnels v_0 et v_1.

Plaçons-nous maintenant dans la situation de la proposition. Notons $CH^n(X)_{\mathrm{hom}}$ le noyau de $\mathrm{cl}^n : CH^n(X) \to H^{2n}_{\mathrm{cont}}(X, \mathbb{Z}_l(n))$. Je dis que l'isomorphisme

$$CH^n(X) \otimes \mathbb{Z}_l \xrightarrow{\sim} CH^n(X, \mathbb{Z}_l)$$

envoie $CH^n(X)_{\mathrm{hom}} \otimes \mathbb{Z}_l$ sur $CH^n(X, \mathbb{Z}_l)_{\mathrm{hom}}$. En effet, soit $x \in CH^n(X, \mathbb{Z}_l)_{\mathrm{hom}}$. Écrivant $x = \sum \alpha_i x_i$ avec $\alpha_i \in \mathbb{Z}_l$, $x_i \in CH^n(X)$, on peut (quitte à augmenter k_0) supposer que x provient de $x_0 \in CH^n(X_0, \mathbb{Z}_l)$. On a évidemment $x_0 \in CH^n(X_0, \mathbb{Z}_l)_{\mathrm{hom}}$; le même raisonnement que dans la preuve du théorème 5.2 (utilisant le fait que $X_0 \in B_{\mathrm{Tate}}(k_0)$) montre alors que x_0 est de torsion. Mais, pour tout groupe abélien M, on a des isomorphismes

$$M\{l\} \xrightarrow{\sim} M\{l\} \otimes \mathbb{Z}_l \xrightarrow{\sim} (M \otimes \mathbb{Z}_l)\{l\}$$

puisque $(M/M\{l\}) \otimes \mathbb{Z}_l$ est sans l-torsion. Donc $x_0 \in CH^n(X_0)_{\mathrm{hom}}$ et $x \in CH^n(X)_{\mathrm{hom}}$.

Il résulte de ceci que l'homomorphisme induit

$$\mathrm{Griff}^n(X) \otimes \mathbb{Z}_l = A^n_{\mathrm{alg}}(X)_{\mathrm{hom}} \otimes \mathbb{Z}_l \to A^n_{\mathrm{alg}}(X, \mathbb{Z}_l)_{\mathrm{hom}} = \mathrm{Griff}^n(X, \mathbb{Z}_l)$$

est surjectif, donc bijectif, d'où l'énoncé. □

Proposition 5.5. *Sous les hypothèses de la proposition 5.4, les conditions suivantes sont équivalentes :*

(i) $H^0(X, \mathcal{H}^3(\mathbb{Q}_l/\mathbb{Z}_l(2)))$ *est fini.*

(ii) *Le monomorphisme* $\mathrm{Griff}^2(X, \mathbb{Z}_l) \to H^3_{\mathrm{tr}}(X, \mathbb{Z}_l(2)) \otimes \mathbb{Q}_l/\mathbb{Z}_l$ *du théorème 5.2 b) est surjectif.*

(iii) $\operatorname{corang} \mathrm{Griff}^2(X, \mathbb{Z}_l) \geq \dim H^3_{\mathrm{tr}}(X, \mathbb{Q}_l(2))$.

(iv) *L'application b de* (4-2) *est nulle.*

(v) $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2))$ *est fini, quotient de* $H^4_{\mathrm{cont}}(X, \mathbb{Z}_l)_{\mathrm{tors}}$.

Si X est une variété abélienne (*il suffit que* $H^3(X, \mathbb{Z}_l)$ *et* $H^4(X, \mathbb{Z}_l)$ *soient sans torsion*), *ces conditions sont encore équivalentes à* :

(vi) *Pour tout* $n \geq 1$, *l'homomorphisme* $H^3_{\text{ét}}(X, \mathbb{Z}/l^n) \to H^3_{\text{ét}}(k(X), \mathbb{Z}/l^n)$ *est* nul.

Démonstration. Les équivalences (i) $\iff$ (ii) $\iff$ (iii) résultent du théorème 5.2 b) et du théorème 1.1. Les équivalences (ii) $\iff$ (iv) $\iff$ (v) résultent du théorème 5.2 a), du corollaire 4.15 et de la proposition 4.14 (ou plus directement du diagramme (4-2)).

Si (vi) est vrai, il est vrai stablement (c'est-à-dire à coefficients $\mathbb{Q}_l/\mathbb{Z}_l$), ce qui est équivalent à la nullité de $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l/\mathbb{Z}_l)$, d'où (iv). Réciproquement, montrons que (v) $\Longrightarrow$ (vi) si $H^3_{\mathrm{cont}}(X, \mathbb{Z}_l)$ et $H^4_{\mathrm{cont}}(X, \mathbb{Z}_l)$ sont sans torsion. De (v) on déduit que $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2)) = 0$, ce qui donne (vi) stablement. Pour l'obtenir à coefficients finis, considérons le diagramme commutatif aux lignes exactes :

$$\begin{array}{ccccccccc}
0 \to & H^2_{\mathrm{Nis}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2))/l^n & \to & H^3_{\mathrm{Nis}}(X, \mathbb{Z}/l^n(2)) & \to & {}_{l^n}H^3_{\mathrm{Nis}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2)) & \to 0 \\
& \downarrow & & \downarrow & & \wr\downarrow & \\
0 \to & H^2_{\text{ét}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2))/l^n & \to & H^3_{\text{ét}}(X, \mathbb{Z}/l^n(2)) & \to & {}_{l^n}H^3_{\text{ét}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2)) & \to 0.
\end{array}$$

Comme $H^3_{\mathrm{cont}}(X, \mathbb{Z}_l)$ est sans torsion, $H^2_{\text{ét}}(X, \mathbb{Q}_l/\mathbb{Z}_l(2))$ est divisible et le terme en bas à gauche est nul. La flèche verticale centrale est donc surjective, ce qui donne l'énoncé pour $H^3_{\text{ét}}(X, \mathbb{Z}/l^n(2))$. □

Théorème 5.6. *Soient p un nombre premier* $\equiv 1 \pmod 3$, $k = \bar{\mathbb{F}}_p$, *et E la courbe elliptique sur k d'équation* $x^3 + y^3 + z^3 = 0$. *Posons* $X = E^3$. *Si* $l \equiv -1 \pmod 3$, *les conditions de la proposition 5.5 sont vérifiées.*

Démonstration. Pour commencer, observons que $X \in B_{\mathrm{Tate}}(\mathbb{F}_p)$. Cela résulte du théorème de Spieß [1999], ou simplement de [Soulé 1984, théorème 3] puisque $\dim X = 3$.

Montrons (iii). D'après [Schoen 1995, Theorem 0.1] et la proposition 5.4, on a $\mathrm{Griff}^2(X, \mathbb{Z}_l) \simeq (\mathbb{Q}_l/\mathbb{Z}_l)^2$; il faut donc montrer que $H^3_{\mathrm{tr}}(X, \mathbb{Z}_l)$ est de rang ≤ 2. Comme X est une variété abélienne, on a un isomorphisme

$$\Lambda^3 H^1_{\mathrm{cont}}(X, \mathbb{Q}_l) \xrightarrow{\sim} H^3_{\mathrm{cont}}(X, \mathbb{Q}_l).$$

L'hypothèse sur p assure que E est ordinaire (cf. [Schoen 1995, p. 46]). Soient α, β les nombres de Weil de E sur $\mathbb{F}_p$: on a $\alpha\beta = p$, et $K := \mathbb{Q}(\alpha) = \mathbb{Q}(\mu_3)$ (ibid.).

L'espace vectoriel $H^1_{\mathrm{cont}}(X, \mathbb{Q}_l)$ est somme de trois exemplaires de $H^1_{\mathrm{cont}}(E, \mathbb{Q}_l)$: il est donc de rang 6. Soit $(v_1, v_2, v_3, w_1, w_2, w_3)$ une base du $\mathbb{Q}_l \otimes K$-module libre $H^1_{\mathrm{cont}}(X, \mathbb{Q}_l) \otimes K$ formée de vecteurs propres pour l'action du Frobenius ϕ, avec $\phi v_i = \alpha v_i$, $\phi w_i = \beta w_i$. Le $\mathbb{Q}_l \otimes K$-module $H^3_{\mathrm{cont}}(X, \mathbb{Q}_l) \otimes K$ est libre de rang 20, de base les tenseurs purs de degré 3 construits sur les v_i, w_j. Par construction, cette base B est formée de vecteurs propres pour l'action de Frobenius.

Soit $b \in B$. Si $b \notin \{v_1 \wedge v_2 \wedge v_3, w_1 \wedge w_2 \wedge w_3\}$, b est divisible par $c = v_i \wedge w_j$ pour un couple (i, j). La valeur propre de $c \in H^2_{\mathrm{cont}}(X, \mathbb{Q}_l) \otimes K$ est égale à p ; en particulier, $c \in H^2_{\mathrm{cont}}(X, \mathbb{Q}_l)$. Par le théorème de Tate (dû dans ce cas particulier à Deuring), $c \otimes \mathbb{Q}_l(1) \in H^2_{\mathrm{cont}}(X, \mathbb{Q}_l(1))$ est de la forme $\mathrm{cl}(\gamma)$ pour un diviseur $\gamma \in \mathrm{Pic}(X) \otimes \mathbb{Z}_l$: il en résulte que $b \in NH^3_{\mathrm{cont}}(X, \mathbb{Q}_l)$.

Ceci montre que $H^3_{\mathrm{tr}}(X, \mathbb{Q}_l) \otimes K$ est engendré par $b = v_1 \wedge v_2 \wedge v_3$ et $b' = w_1 \wedge w_2 \wedge w_3$, et donc que $\dim H^3_{\mathrm{tr}}(X, \mathbb{Q}_l) \leq 2$. □

Remarque 5.7. Comme $H^*_{\mathrm{cont}}(X, \mathbb{Z}_l) \to H^*_{\text{ét}}(X, \mathbb{Z}/l^n)$ est surjectif, le calcul fait dans la preuve du théorème 5.6 montre a priori que l'image de $H^3_{\text{ét}}(X, \mathbb{Z}/l^n)$ dans $H^3(F, \mathbb{Z}/l^n)$ est de rang ≤ 2, où $F = k(X)$. On aimerait bien démontrer sa nullité (l'énoncé (vi) de la proposition 5.5) directement : il s'agit de voir que, si $x_1, x_2, x_2 \in H^1_{\text{ét}}(X, \mathbb{Z}/l^n)$, le cup-produit $x_1 \cdot x_2 \cdot x_3$ est nul dans $H^3_{\text{ét}}(F, \mathbb{Z}/l^n)$.

On peut se limiter aux triplets (x_1, x_2, x_3) tels que

$$x_i \in \mathrm{Im}\left(H^1_{\text{ét}}(E, \mathbb{Z}/l^n) \xrightarrow{\pi_i^*} H^1_{\text{ét}}(X, \mathbb{Z}/l^n)\right)$$

pour une valeur de i, où π_i est la i-ème projection, et sans perte de généralité, supposer $i = 1$. Alors x_1 définit une isogénie $f : E' \to E$ de degré l^n. Soit $F' = k(X')$, où $X' = E' \times E \times E$: d'après Merkurjev–Suslin, la nullité de $x_1 \cdot x_2 \cdot x_3$ dans $H^3_{\text{ét}}(F, \mathbb{Z}/l^n)$ équivaut au fait que $x_2 \cdot x_3 \in H^2(F, \mathbb{Z}/l^n) \simeq {}_{l^n}Br(F)$ est une norme dans l'extension F'/F. Peut-on montrer ceci directement ?

5D. ***Autres corps.*** Si $cd(k) \leq 1$, la suite exacte (5-2) persiste [Jannsen 1988, Theorem 3.3]. Malheureusement, elle ne semble pas apporter d'informations supplémentaires très utiles, sauf peut-être dans le cas d'un corps quasi-fini que je n'ai pas exploré.

Considérons les notations de la preuve de la proposition 3.6. Si k est de type fini mais n'est pas fini, je ne sais pas s'il faut espérer que K est de torsion, même sous toutes les conjectures habituelles (Jannsen le suggère dans [Jannsen 1994, Lemma 2.7]). On peut remplacer $H^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2))$ par le groupe plus fin

$$\tilde{H}^4_{\mathrm{cont}}(X, \mathbb{Z}_l(2)) = \varinjlim H^4_{\mathrm{cont}}(\mathscr{X}, \mathbb{Z}_l(2))$$

où $\mathscr{X}$ décrit les modèles réguliers de X, de type fini sur $\mathrm{Spec}\,\mathbb{Z}$ (cf. [Jannsen 1990, (11.6.1)]). En caractéristique p, par passage à la limite, la conjecture de Tate–Beilinson implique alors que $K_{\text{ét}}$ est de torsion [Kahn 2005, théorème 60, (iii)]. De plus, cette conjecture implique que $CH^2(X)$ est de type fini (comme quotient de $CH^2(\mathscr{X})$ pour un modèle $\mathscr{X}$ lisse de type fini), donc que K est fini. Par contre, elle n'implique pas a priori que $H^4_{\text{ét}}(X, \mathbb{Z}(2))$ est de type fini (dans les suites exactes de Gysin pour un diviseur, le terme suivant est de la forme $H^3_{\text{ét}}(Z, \mathbb{Z}(1)) = Br(Z)\dots$) donc il se pourrait fort bien que $K_{\text{ét}}$ ait une partie divisible non triviale.

Le bon objet avec lequel travailler pour des variétés ouvertes sur un corps fini est $H^4_W(\mathscr{X}, \mathbb{Z}(2))$ (cohomologie Weil-étale) : c'est celui qui permet d'attraper tout $\tilde{H}^4_{\text{cont}}(X, \mathbb{Z}_l(2))$ pour $l \neq p$ [Kahn 2005, théorème 64]. Mais cela a l'air compliqué, cf. [Kahn 2003, (3.2)] ou [Kahn 2005, théorème 62 (ii)].

Annexe : Cohomologie de Hodge–Witt logarithmique sur des corps imparfaits

Dans [Geisser et Levine 2000], Geisser et Levine comparent la cohomologie motivique modulo p d'un corps de caractéristique p quelconque avec sa cohomologie de Hodge(-Witt) logarithmique, mais n'en déduisent une comparaison globale que pour des variétés lisses sur un corps parfait. Le but de ce numéro est de rappeler les bases de cette comparaison et de se débarrasser de manière « triviale » de l'hypothèse de perfection, à l'aide d'une observation classique de Quillen [1973, p. 133, démonstration du théorème 5.11].

Cohomologie de Hodge–Witt logarithmique. Soit X un schéma de caractéristique p. On lui associe son pro-complexe de de Rham–Witt [Illusie 1979, p. 548, 1.12]

$$(W_r\Omega^\bullet_X)_{r\geq 1}$$

qui est un système projectif de faisceaux d'algèbres différentielles graduées sur $X_{\text{ét}}$, prolongeant (pour $\cdot = 0$) le pro-faisceau des vecteurs de Witt et (pour $r = 1$) le complexe des différentielles de Kähler. Il est muni d'un opérateur $F : W_r\Omega^n_X \to W_{r-1}\Omega^n_X$ [Illusie 1979, p. 562, théorème 2.17].

Si X est défini sur un corps parfait k, on a évidemment

$$W_r\Omega^\bullet_X = W_r\Omega^\bullet_{X/k}.$$

On a des applications « de Teichmüller » (multiplicatives)

$$\mathbb{O}_X \to W_r\mathbb{O}_X, \quad x \mapsto \underline{x} = (x, 0, \dots, 0, \dots)$$

[Illusie 1979, p. 505, (1.1.7)], qu'on utilise pour définir les homomorphismes

$$d\log : \mathbb{O}^*_X/(\mathbb{O}^*_X)^{p^r} \to W_r\Omega^1_X, \quad x \mapsto d\underline{x}/\underline{x} \tag{A-1}$$

[Illusie 1979, p. 580, (3.23.1)]. On définit alors $W_r\Omega^n_{X,\log}$ comme le sous-faisceau de $W_r\Omega^n_X$ engendré localement pour la topologie étale par les sections de la forme

$d\log x_1 \wedge \cdots \wedge d\log x_n$ [Illusie 1979, p. 596, (5.7.1)] ; comme dans [Geisser et Levine 2000], nous noterons simplement ce faisceau $\nu_r(n)_X$.

Lemme A.1. *Pour tout $x \in \Gamma(X, \mathbb{O}_X)$ et pour tout $r \geq 1$, on a*

$$\underline{x} \wedge \underline{1-x} = 0 \in \Gamma(X, W_r\Omega_X^2).$$

Démonstration. (Illusie) Le morphisme $X \to \mathbb{A}^1_{\mathbb{F}_p}$ défini par x nous ramène au cas universel $X = \operatorname{Spec} \mathbb{F}_p[t]$, $x = t$. Mais alors $W_r\Omega_X^2 = 0$ puisque $\dim X = 1$. □

Le symbole logarithmique. Supposons $X = \operatorname{Spec} k$, où k est un corps. Le lemme A.1 implique que l'homomorphisme $d\log$ de (A-1) induit un *symbole logarithmique*

$$d\log : K_n^M(k)/p^r \to \nu_r(n)_k, \qquad \text{(A-2)}$$
$$\{x_1, \dots, x_n\} \mapsto d\log(x_1) \wedge \cdots \wedge d\log(x_n).$$

Soit K le corps des fonctions d'un k-schéma lisse X, où k est parfait de caractéristique p. Un point x de codimension 1 de X définit une valuation discrète v sur K, de corps résiduel $E = k(x)$. Le théorème de pureté de Gros [1985, p. 46, théorème 3.5.8] et la longue suite exacte de cohomologie à supports définissent des homomorphismes résidus

$$\nu_r(n)_K \xrightarrow{\partial_v} \nu_r(n-1)_E. \qquad \text{(A-3)}$$

Lemme A.2. *Le diagramme*

$$\begin{array}{ccc} K_n^M(K)/p^r & \xrightarrow{\partial_v} & K_{n-1}^M(E)/p^r \\ {\scriptstyle d\log}\downarrow & & {\scriptstyle d\log}\downarrow \\ \nu_r(n)_K & \xrightarrow{\partial_v} & \nu_r(n-1)_E \end{array}$$

où la flèche horizontale du haut est le résidu en K-théorie de Milnor, est commutatif au signe près.

Démonstration. Pour $n = 1, 2$ c'est fait dans [Gros et Suwa 1988b, p. 625, lemme 4.11]. La démonstration ne se propage pas tout à fait à $n > 2$ car elle utilise la formule explicite donnant $\partial(\{x, y\})$ pour $x, y \in K^*$. Pour la propager, il suffit toutefois de remarquer que $K_n^M(K)$ est engendré par les symboles de la forme $\{u_1, \dots, u_{n-1}, x\}$ avec $u_i \in O_v^*$ et $x \in K^*$. □

Le morphisme de comparaison. Supposons X lisse sur un corps parfait k. D'après [Illusie 1979, p. 597, théorème 5.7.2], on a une suite exacte de pro-faisceaux étales

$$0 \to \nu_\cdot(n)_X \to W_\cdot\Omega_X^n \xrightarrow{1-F} W_\cdot\Omega_X^n \to 0$$

qui en fait n'interviendra pas ici. De plus, on a le théorème suivant, dû à Gros et Suwa :

Théorème A.3. *On a une suite exacte de faisceaux zariskiens*

$$0 \to \alpha_* \nu_r(n)_X \to \bigoplus_{x \in X^{(0)}} \big(\nu_r(n)_{k(x)}\big)_{\overline{\{x\}}} \xrightarrow{\partial} \bigoplus_{x \in X^{(1)}} \big(\nu_r(n-1)_{k(x)}\big)_{\overline{\{x\}}} \xrightarrow{\partial} \cdots$$

où α désigne la projection $X_{\text{ét}} \to X_{\text{Zar}}$ et les différentielles ∂ sont construites à partir des résidus (A-3).

Démonstration. Voir [Gros et Suwa 1988b, corollaire 1.6] ou [Colliot-Thélène et al. 1997, p. 70, Example 7.4 (3)]. □

Supposons maintenant X régulier de type fini sur un corps k (de caractéristique p). Supposons d'abord k de type fini sur $\mathbb{F}_p$: alors X admet un modèle $\mathscr{X}$ régulier, donc lisse, de type fini sur $\mathbb{F}_p$. Soit $j : X \to \mathscr{X}$ la pro-immersion ouverte correspondante : on a évidemment

$$\nu_r(n)_X = j^* \nu_r(n)_{\mathscr{X}}$$

puisque les anneaux semi-locaux de X sont certains anneaux semi-locaux de $\mathscr{X}$.

Interprétons maintenant $K_n^M(K)/p^n$ comme $H^n(K, \mathbb{Z}/p(n))$, cf. théorème 2.3. Vu la remarque 2.4 et le théorème A.3, les homomorphismes (A-2) et le lemme A.2 induisent des homomorphismes de faisceaux

$$\mathscr{H}^n(\mathbb{Z}/p^r(n)_{\mathscr{X}}) \to \alpha_* \nu_r(n)_{\mathscr{X}}$$

et donc des morphismes dans $D^-(\mathscr{X}_{\text{Zar}})$

$$\mathbb{Z}/p^r(n)_{\mathscr{X}} \to \alpha_* \nu_r(n)_{\mathscr{X}}[-n]$$

puisque $\mathscr{H}^i(\mathbb{Z}/p^r(n)_{\mathscr{X}}) = 0$ pour $i > n$ (lemme 2.5).

D'où, en appliquant j^*, des morphismes dans $D^-(X_{\text{Zar}})$

$$\mathbb{Z}/p^r(n)_X \to \alpha_* \nu_r(n)_X[-n]. \tag{A-4}$$

Si k est quelconque, écrivons

$$k = \varinjlim_{\alpha} k_\alpha \quad \text{et} \quad X = \varprojlim_{\alpha} X_\alpha,$$

où les k_α sont de type fini sur $\mathbb{F}_p$ et X_α est un k_α-schéma régulier de type fini, de sorte que $k_\alpha \subset k_\beta$ induise un isomorphisme $X_\beta \xrightarrow{\sim} X_\alpha \otimes_{k_\alpha} k_\beta$. On a évidemment :

$$\nu_r(n)_X = \varinjlim_{\alpha} \pi_\alpha^* \nu_r(n)_{X_\alpha},$$
$$\mathscr{H}^n(\mathbb{Z}/p^r(n)_X) = \varinjlim_{\alpha} \pi_\alpha^* \mathscr{H}^n(\mathbb{Z}/p^r(n)_{X_\alpha}),$$

où $\pi_\alpha : X \to X_\alpha$ est le morphisme canonique. Ceci étend la définition de (A-4) au cas où le corps de base est quelconque. On voit de même :

Proposition A.4 (cf. [Quillen 1973, démonstration du théorème 5.11]). *La suite exacte du théorème A.3 s'étend à tout X régulier de type fini sur un corps.* □

Le théorème de Geisser–Levine.

Théorème A.5. *Soit X un schéma régulier de type fini sur un corps k de caractéristique p. Alors le morphisme* (A-4) *est un isomorphisme.*

Démonstration. Il s'agit de voir que

$$\mathcal{H}^i(\mathbb{Z}/p^r(n)_X) \simeq \begin{cases} 0 & \text{si } i \neq n, \\ \alpha_* \nu_r(n)_X & \text{si } i = n, \end{cases}$$

le dernier isomorphisme étant induit par (A-4). L'énoncé est clair pour $i > n$, cf. lemme 2.5.

1) $X = \operatorname{Spec} k$: c'est le théorème de Bloch–Gabber–Kato pour $i = n$ [Bloch et Kato 1986, p. 117, corollaire 2.8] et celui de Geisser–Levine [2000, théorème 1.1] pour $i < n$.

2) X lisse sur k parfait : on se réduit à 1) en utilisant le théorème A.3, le lemme A.2 et la conjecture de Gersten pour la cohomologie motivique, cf. preuve du lemme 2.5.

3) k de type fini sur $\mathbb{F}_p$: on se ramène à 2) par la technique du numéro précédent.

4) k quelconque : on se ramène à 3) par passage à la limite. □

Remarque A.6. On pourrait court-circuiter les étapes 2) et 3), dans l'esprit de la proposition A.4.

Corollaire A.7. *Soit X un schéma régulier de type fini sur un corps k de caractéristique p. Alors le morphisme α^** (A-4) *est un isomorphisme, où α est la projection $X_{\text{ét}} \to X_{\text{Zar}}$.* □

Remerciements

Cet article est l'élaboration d'un texte rédigé en janvier 2010, lui-même directement inspiré de plusieurs discussions avec Jean-Louis Colliot-Thélène autour de son travail avec Claire Voisin [Colliot-Thélène et Voisin 2010], alors en cours d'achèvement. Il a ensuite bénéficié de la lecture de [Colliot-Thélène et Voisin 2010], ainsi que de discussions avec Colliot-Thélène autour d'un article commun en projet [Colliot-Thélène et Kahn 2011] : je le remercie de sa lecture critique de versions préliminaires de ce texte, et notamment de m'avoir fait remarquer qu'il n'est pas nécessaire de supposer k de type fini sur son sous-corps premier dans le théorème 1.1. Je remercie l'IMPA de Rio de Janeiro pour son hospitalité pendant la fin de sa rédaction, ainsi que la coopération franco-brésilienne pour son soutien. Pour finir, je remercie Luc Illusie de m'avoir indiqué une démonstration (triviale !) du lemme A.1, lemme qui à ma connaissance ne figure nulle part dans la littérature.

Bibliographie

[Bloch 1986] S. Bloch, “Algebraic cycles and higher K-theory”, *Adv. in Math.* **61**:3 (1986), 267–304. MR 88f:18010 Zbl 0608.14004

[Bloch 1994] S. Bloch, “The moving lemma for higher Chow groups”, *J. Algebraic Geom.* **3**:3 (1994), 537–568. MR 96c:14007 Zbl 0830.14003

[Bloch 2010] S. Bloch, *Lectures on algebraic cycles*, 2nd éd., New Mathematical Monographs **16**, Cambridge University, 2010. MR 2011h:14009 Zbl 1201.14006

[Bloch et Kato 1986] S. Bloch et K. Kato, “p-adic étale cohomology”, *Inst. Hautes Études Sci. Publ. Math.* **63** (1986), 107–152. MR 87k:14018 Zbl 0613.14017

[Bloch et Ogus 1974] S. Bloch et A. Ogus, “Gersten’s conjecture and the homology of schemes”, *Ann. Sci. École Norm. Sup.* (4) **7** (1974), 181–201. MR 54 #318 Zbl 0307.14008

[Bloch et Srinivas 1983] S. Bloch et V. Srinivas, “Remarks on correspondences and algebraic cycles”, *Amer. J. Math.* **105**:5 (1983), 1235–1253. MR 85i:14002 Zbl 0525.14003

[Block 1995] S. Block, “A spectral sequence for motivic cohomology”, preprint, 1995, Voir http :// www.math.uiuc.edu/K-theory/0062/.

[Colliot-Thélène et Kahn 2011] J. Colliot-Thélène et B. Kahn, “Cycles de codimension 2 et H^3 non ramifié pour les variétés sur les corps finis”, preprint, 2011. arXiv 1104.3350v1

[Colliot-Thélène et Raskind 1985] J.-L. Colliot-Thélène et W. Raskind, “$\mathcal{K}_2$-cohomology and the second Chow group”, *Math. Ann.* **270**:2 (1985), 165–199. MR 86m:14005 Zbl 0536.14004

[Colliot-Thélène et Voisin 2010] J. Colliot-Thélène et C. Voisin, “Cohomologie non ramifiée et conjecture de Hodge entière”, preprint, 2010. arXiv 1005.2778v1

[Colliot-Thélène et al. 1983] J.-L. Colliot-Thélène, J.-J. Sansuc et C. Soulé, “Torsion dans le groupe de Chow de codimension deux”, *Duke Math. J.* **50**:3 (1983), 763–801. MR 85d:14010 Zbl 0574.14004

[Colliot-Thélène et al. 1997] J.-L. Colliot-Thélène, R. T. Hoobler et B. Kahn, “The Bloch-Ogus-Gabber theorem”, pp. 31–94 dans *Algebraic K-theory* (Toronto, 1996), édité par V. P. Snaith, Fields Inst. Commun. **16**, Amer. Math. Soc., Providence, RI, 1997. MR 98j:14021 Zbl 0911.14004

[Deligne 1974] P. Deligne, “La conjecture de Weil, I”, *Inst. Hautes Études Sci. Publ. Math.* **43** (1974), 273–307. MR 49 #5013 Zbl 0287.14001

[Deninger 1988] C. Deninger, “A proper base change theorem for nontorsion sheaves in étale cohomology”, *J. Pure Appl. Algebra* **50**:3 (1988), 231–235. MR 89f:14016 Zbl 0672.14010

[Geisser et Levine 2000] T. Geisser et M. Levine, “The K-theory of fields in characteristic p”, *Invent. Math.* **139**:3 (2000), 459–493. MR 2001f:19002 Zbl 0957.19003

[Geisser et Levine 2001] T. Geisser et M. Levine, “The Bloch–Kato conjecture and a theorem of Suslin-Voevodsky”, *J. Reine Angew. Math.* **530** (2001), 55–103. MR 2003a:14031 Zbl 1023.14003

[Gros 1985] M. Gros, *Classes de Chern et classes de cycles en cohomologie de Hodge–Witt logarithmique*, Mém. Soc. Math. France (N.S.) **21**, 1985. MR 87m:14021 Zbl 0615.14011

[Gros et Suwa 1988a] M. Gros et N. Suwa, “Application d’Abel–Jacobi p-adique et cycles algébriques”, *Duke Math. J.* **57**:2 (1988), 579–613. MR 89h:14006a Zbl 0697.14005

[Gros et Suwa 1988b] M. Gros et N. Suwa, “La conjecture de Gersten pour les faisceaux de Hodge–Witt logarithmique”, *Duke Math. J.* **57**:2 (1988), 615–628. MR 89h:14006b Zbl 0715.14011

[Grothendieck 1968] A. Grothendieck, “Le groupe de Brauer, III : Exemples et compléments”, pp. 88–188 dans *Dix exposés sur la cohomologie des schémas*, Advanced Studies Pure Math. **3**, North-Holland, Amsterdam, 1968. MR 39 #5586c Zbl 0198.25901

[Illusie 1979] L. Illusie, “Complexe de de Rham–Witt et cohomologie cristalline”, *Ann. Sci. École Norm. Sup.* (4) **12**:4 (1979), 501–661. MR 82d:14013

[Jannsen 1988] U. Jannsen, “Continuous étale cohomology”, *Math. Ann.* **280**:2 (1988), 207–245. MR 89a:14022 Zbl 0649.14011

[Jannsen 1990] U. Jannsen, *Mixed motives and algebraic K-theory*, Lecture Notes in Mathematics **1400**, Springer, Berlin, 1990. MR 91g:14008 Zbl 0691.14001

[Jannsen 1992] U. Jannsen, “Motives, numerical equivalence, and semi-simplicity”, *Invent. Math.* **107**:3 (1992), 447–452. MR 93g:14009 Zbl 0762.14003

[Jannsen 1994] U. Jannsen, “Motivic sheaves and filtrations on Chow groups”, pp. 245–302 dans *Motives* (Seattle, WA, 1991), édité par U. Jannsen et al., Proc. Sympos. Pure Math. **55**, Amer. Math. Soc., Providence, RI, 1994. MR 95c:14006 Zbl 0811.14004

[de Jong 1996] A. J. de Jong, “Smoothness, semi-stability and alterations”, *Inst. Hautes Études Sci. Publ. Math.* **83** (1996), 51–93. MR 98e:14011 Zbl 0916.14005

[Kahn 1996] B. Kahn, “Applications of weight-two motivic cohomology”, *Doc. Math.* **1** (1996), 395–416. MR 98b:14007 Zbl 0883.19002

[Kahn 1997] B. Kahn, “La conjecture de Milnor (d’après V. Voevodsky)”, pp. 379–418 dans *Séminaire Bourbaki*, Astérisque **245**, Société Mathématique de France, Paris, 1997. Reprinted as pp. 1105–1149 in *Handbook of K-theory*, Springer, Berlin, 2005. MR 1627119 Zbl 0916.19001

[Kahn 2002] B. Kahn, “The Geisser–Levine method revisited and algebraic cycles over a finite field”, *Math. Ann.* **324**:3 (2002), 581–617. MR 2003h:14034 Zbl 1014.14004

[Kahn 2003] B. Kahn, “Équivalences rationnelle et numérique sur certaines variétés de type Abélien sur un corps fini”, *Ann. Sci. École Norm. Sup.* (4) **36**:6 (2003), 977–1002. MR 2005e:14037 Zbl 1073.14034

[Kahn 2005] B. Kahn, “Algebraic K-theory, algebraic cycles and arithmetic geometry”, pp. 351–428 dans *Handbook of K-theory*, vol. 1, édité par E. M. Friedlander et D. R. Grayson, Springer, Berlin, 2005. MR 2007b:14016 Zbl 1115.19003

[Kahn 2006] B. Kahn, “Sur le groupe des classes d’un schéma arithmétique”, *Bull. Soc. Math. France* **134**:3 (2006), 395–415. MR 2007d:14052 Zbl 1222.14048

[Kahn et al. 2007] B. Kahn, J. P. Murre et C. Pedrini, “On the transcendental part of the motive of a surface”, pp. 143–202 dans *Algebraic cycles and motives*, vol. 2, édité par J. Nagel et C. Peters, London Math. Soc. Lecture Note Ser. **344**, Cambridge Univ., 2007. MR 2009c:14010 Zbl 1130.14008

[Katz et Messing 1974] N. M. Katz et W. Messing, “Some consequences of the Riemann hypothesis for varieties over finite fields”, *Invent. Math.* **23** (1974), 73–77. MR 48 #11117 Zbl 0275.14011

[Kimura 2005] S.-I. Kimura, “Chow groups are finite dimensional, in some sense”, *Math. Ann.* **331**:1 (2005), 173–201. MR 2005j:14007 Zbl 1067.14006

[Kleiman 1968] S. L. Kleiman, “Algebraic cycles and the Weil conjectures”, pp. 359–386 dans *Dix exposés sur la cohomologie des schémas*, Advanced Studies Pure Math. **3**, North-Holland, Amsterdam, 1968. MR 45 #1920 Zbl 0198.25902

[Lichtenbaum 1990] S. Lichtenbaum, "New results on weight-two motivic cohomology", pp. 35–55 dans *The Grothendieck Festschrift*, vol. 3, Progr. Math. **88**, Birkhäuser, Boston, MA, 1990. MR 92m:14030 Zbl 0809.14004

[Nesterenko et Suslin 1989] Y. P. Nesterenko et A. A. Suslin, "Homology of the general linear group over a local ring, and Milnor's K-theory", *Izv. Akad. Nauk SSSR Ser. Mat.* **53**:1 (1989), 121–146. In Russian ; translated in *Math. USSR-Izv.* **34** :1 (1990), 121–145. MR 90a:20092 Zbl 0668.18011

[Quillen 1973] D. Quillen, "Higher algebraic K-theory, I", pp. 85–147 dans *Algebraic K-theory, I : Higher K-theories* (Seattle, WA, 1972), édité par H. Bass, Lecture Notes in Math. **341**, Springer, Berlin, 1973. MR 49 #2895 Zbl 0292.18004

[Schoen 1995] C. Schoen, "On the computation of the cycle class map for nullhomologous cycles over the algebraic closure of a finite field", *Ann. Sci. École Norm. Sup.* (4) **28**:1 (1995), 1–50. MR 95j:14010 Zbl 0839.14004

[Schoen 2002] C. Schoen, "Complex varieties for which the Chow group mod n is not finite", *J. Algebraic Geom.* **11**:1 (2002), 41–100. MR 2002h:14004 Zbl 1059.14010

[Soulé 1984] C. Soulé, "Groupes de Chow et K-théorie de variétés sur un corps fini", *Math. Ann.* **268**:3 (1984), 317–345. MR 86k:14017 Zbl 0573.14001

[Spaltenstein 1988] N. Spaltenstein, "Resolutions of unbounded complexes", *Compositio Math.* **65**:2 (1988), 121–154. MR 89m:18013 Zbl 0636.18006

[Spieß 1999] M. Spieß, "Proof of the Tate conjecture for products of elliptic curves over finite fields", *Math. Ann.* **314**:2 (1999), 285–290. MR 2000d:14012 Zbl 0941.11026

[Suslin 1987] A. A. Suslin, "Torsion in K_2 of fields", *K-Theory* **1**:1 (1987), 5–29. MR 89a:11123 Zbl 0635.12015

[Suslin et Voevodsky 2000] A. Suslin et V. Voevodsky, "Bloch–Kato conjecture and motivic cohomology with finite coefficients", pp. 117–189 dans *The arithmetic and geometry of algebraic cycles* (Banff, 1998), édité par B. B. Gordon et al., NATO Sci. Ser. C Math. Phys. Sci. **548**, Kluwer, Dordrecht, 2000. MR 2001g:14031 Zbl 1005.19001

[Totaro 1992] B. Totaro, "Milnor K-theory is the simplest part of algebraic K-theory", *K-Theory* **6**:2 (1992), 177–189. MR 94d:19009 Zbl 0776.19003

[Voevodsky 2003] V. Voevodsky, "Motivic cohomology with $\mathbf{Z}/2$-coefficients", *Publ. Math. Inst. Hautes Études Sci.* **98** (2003), 59–104. MR 2005b:14038b Zbl 1057.14028

Communicated by Raman Parimala
Received 2011-04-14 Revised 2011-06-12 Accepted 2011-07-17

kahn@math.jussieu.fr *Institut de Mathématiques de Jussieu, CNRS, 4 place Jussieu, Case 247, 75252 Cedex 05 Paris, France*

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1409

Higher-order Maass forms

Roelof Bruggeman and Nikolaos Diamantis

The spaces of Maass forms of even weight and of arbitrary order are studied. It is shown that, if we allow exponential growth at the cusps, these spaces are as large as algebraic restrictions allow. These results also apply to higher-order holomorphic forms of even weight.

1. Introduction

Occasionally, invariants of classical holomorphic modular forms can be studied effectively by means of generating functions that are nonanalytic. An example is the Eisenstein series modified with modular symbols. It is defined by

$$E^*(z,s) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_0(N)} \langle f,\gamma\rangle\,\mathrm{Im}(\gamma z)^s, \tag{1-1}$$

where Γ_∞ is the subgroup of translations of the congruence group $\Gamma_0(N)$, f is a weight-2 newform, and $\langle f,\gamma\rangle$ denotes its modular symbol $-2\pi i\int_\infty^{\gamma\infty} f(w)\,dw$. The function $E^*(-,s)$ is not analytic, but rather an eigenfunction of the Laplace operator $-y^2\partial_y^2 - y^2\partial_x^2$ with eigenvalue $s-s^2$. Its study has led to important results about modular symbols, such as the proof that the suitably normalized modular symbols follow the normal distribution [Petridis and Risager 2004]. A crucial feature of $E^*(-,s)$ is that it is not invariant under the action of $\Gamma_0(N)$, but instead it is $\Gamma_0(N)$-invariant of order 2.

This function was one of the motivating examples for the systematic study of invariants of order q for a Γ-module V (over $\mathbb{C}$), that is, $v\in V$ satisfying

$$v\,|\,(1-\gamma_1)(1-\gamma_2)\dots(1-\gamma_q)=0 \tag{1-2}$$

for all $\gamma_1,\dots,\gamma_q\in\Gamma$. The usual space of invariants V^Γ consists of the invariants of order 1. Also, $E^*(-,s)$ is invariant of order 2 in terms of the right regular representation of $\Gamma_0(N)$ on the space of functions on $\mathfrak{H}$ that are eigenfunctions of the Laplacian.

MSC2010: primary 11F12; secondary 11F37, 11F99, 30F35.
Keywords: higher-order automorphic forms, Maass forms.

Although higher-order invariants have been classified in several cases and from various perspectives [Chinta et al. 2002; Diamantis and O'Sullivan 2008; Diamantis and Sim 2008; Deitmar 2008; 2009], the real-analytic case to which the important function $E^*(-, s)$ belongs has not been fully addressed up to now. This is perhaps not surprising, given that such functions can contain very rich and complex information, as the example of $E^*(-, s)$ shows. The resolution of the problem of classification of higher-order Maass forms is the subject of the present paper. This task includes various aspects that are often automatic in the classification of higher-order invariants of other spaces, and it requires new techniques from the theory of families of automorphic forms and perturbation theory. The most important of these aspects, all of independent interest, are:

Firstly, while in the setting of [Diamantis and O'Sullivan 2008] the "size" of the space of higher-order invariants is expressed by its (finite) dimension, in the present case a different concept is required because the relevant spaces are, in general, infinitely dimensional. This concept is *maximal perturbability*. Specifically, let $V^{\Gamma,q}$ be the space of invariants of order q of a general group Γ and a Γ-module V. If Γ is finitely generated, there is an exact sequence

$$0 \longrightarrow V^{\Gamma,q} \longrightarrow V^{\Gamma,q+1} \longrightarrow (V^{\Gamma})^{n(\Gamma,q)}, \tag{1-3}$$

where the natural number $n(\Gamma, q)$ is determined by the structure of Γ. If for every $q \geq 1$ the map in (1-3) is surjective, we call the Γ-module V *maximally perturbable*. Our choice of the word "perturbable" stems from the fact that derivatives of families of automorphic form can lead to higher-order automorphic forms; see Section 4C3. A derivative of order q leads to an invariant v of order $q+1$ with the special property that there is an invariant $w \in V^{\Gamma}$ such that for all choices of $\gamma_1, \dots, \gamma_q \in \Gamma$, there is $\mu(\gamma_1, \dots, \gamma_q) \in \mathbb{C}$ with

$$v \,|\, (\gamma_1 - 1) \dots (\gamma_q - 1) = \mu(\gamma_1, \dots, \gamma_q) w. \tag{1-4}$$

We call v a *perturbation* of w. On the other hand, the existence of perturbations in which $\mu(\gamma_1, \dots, \gamma_q)$ is not symmetric in the γ_j implies that not all perturbations come from differentiation of families.

Secondly, the choice of the space V is a nontrivial matter. The obvious choice of functions with polynomial growth at the cusps leads to modules that are not maximally perturbable; we have to allow exponential growth. See Theorems 4.3 and 4.2.

Finally, Fourier expansions are not straightforward extensions of their classical counterparts either, because of the lack of usual invariance. Indeed, Section 7 is devoted to the development of a higher-order version of the theory of Fourier terms and expansions.

Our results imply that if V is maximally perturbable, an invariant in V^{Γ} has "many" perturbations. At the same time, the complexity of the constructions partly explains why it is surprisingly difficult, even in simple situations, to explicitly construct perturbations of a given invariant. In Section 4C we investigate this for the simplest example of an automorphic form that one can think of, the constant function 1. We succeed in giving a basis for all holomorphic perturbations and for all harmonic perturbations of 1 up to order 3.

As in the classical case, the general Maass setting discussed so far includes holomorphic higher-order invariants. Important examples of the latter emerge from problems in the theory of classical modular forms: in [Goldfeld 1995] and [Diamantis 1999], certain "period integrals" are associated to derivatives of L-functions of weight-2 cusp forms in a way analogous to the link between values of L-functions and modular integrals [Manin 1972]. Specifically, let f be a newform of weight 2 for $\Gamma_0(N)$, and let $L_f(s)$ be its L-function. If $L_f(1) = 0$, then $L'_f(1)$ can be written as a linear combination of integrals of the form

$$\int_0^{\gamma(0)} f(z)u(z)\,dz, \qquad \gamma \in \Gamma_0(N), \tag{1-5}$$

plus some "lower-order terms". Here $u(z) := \log \eta(z) + \log \eta(Nz)$, where η is the Dedekind eta function. The differential $f(z)u(z)\,dz$ is not $\Gamma_0(N)$-invariant. It does satisfy a transformation law which is reminiscent of (1-2) with $q = 2$, but is not quite $\Gamma_0(N)$-invariant, as it has an additional term. If it were invariant, the value of the derivative at 1 would be expressed as the value of the actual L-function of second-order $\Gamma_0(N)$ at 1. That could be advantageous for the study of $L'_f(1)$ in terms of the outstanding conjectures (Beilinson, Birch–Swinnerton-Dyer, etc.), especially since there is now evidence that a motivic structure underlies higher-order forms (see [Diamantis and Sreekantan 2006; Sreekantan 2009]).

Here we show that it is indeed possible to obtain a second-order $\Gamma_0(N)$-invariant function from $u(z)$, provided we move to a different domain. This domain is the universal covering group $\tilde{G}$, defined in detail in Section 5A. It was convenient and more general to carry out the entire study (that is, of both Maass and holomorphic higher-order invariants) on $\tilde{G}$.

The main theorems of the paper (Theorems 6.5 and 6.8) classify the spaces into which we incorporate the above two important examples ($E^*(-, s)$ and $u(z)$), in the sense that we show that these spaces are as large as they can a priori be.

2. Structure of the paper

In Section 3, we first discuss higher-order invariants for general groups and modules. Here we define the property of maximal perturbability.

In Section 4, Maass forms on $\mathfrak{H}$ (both general and holomorphic) are defined, and the first two main theorems of the paper (4.2 and 4.3) are stated. Since

$\hom(\Gamma_{\mathrm{mod}}, \mathbb{C}) = \{0\}$, there are no higher-order invariants for the full modular group $\Gamma_{\mathrm{mod}} = \mathrm{PSL}_2(\mathbb{Z})$. For examples, we go to a subgroup of order 6 and show that some well-known functions lead to higher-order invariants. To get all perturbations of 1 up to order 3, we have to look also at less known functions.

In Section 5, the universal covering group $\tilde{G}$ is introduced, and basic facts about $\tilde{G}$ are given. Section 6 starts with the interpretation of $\log \eta$ as a second-order form on the universal covering group $\tilde{G}$ for the inverse image $\tilde{\Gamma}_{\mathrm{mod}}$ of the modular group in $\tilde{G}$. We define Maass forms on the universal covering group in Section 6, and, in Theorems 6.5 and 6.8, we state the counterparts of Theorems 4.2 and 4.3 for forms on the universal covering group. The section concludes with concrete examples of low-order forms for the discrete subgroup $\tilde{\Gamma}_{\mathrm{mod}}$ of $\tilde{G}$.

Section 7 is of independent interest. A theory of Fourier expansions for higher-order forms is developed. Working on the universal covering group, we have to handle invariants for the commutative group generated by a parabolic element and the center of $\tilde{G}$. So all perturbations we meet are commutative.

The proof of Theorems 6.5 and 6.8 is the content of Section 8. We start with the maximal perturbability of the space of all functions on $\tilde{G}$. Step by step, we impose more and more analytic restrictions, such as smoothness, growth behavior at cusps, and the behavior under certain differential operators. For each step, we show that maximal perturbability is preserved. For several steps, this involves induction with respect to the order. At the end, we complete the proof of our main results by an application of spectral theory.

3. Higher-order invariants

In this section, we discuss higher-order invariants in general and then specialize their study to discrete cofinite subgroups $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$. We introduce the concept of a maximally perturbable Γ-module to make precise the statement that there are as many higher-order invariants of a given type as one can expect.

3A. ***Higher-order invariants on general groups.*** The concept of higher-order invariant functions on the upper half-plane is a special case of the concept of higher-order invariants for any group Γ and any Γ-module V. We work with *right* Γ-modules, and write the action as $v \mapsto v|\gamma$. It should be clear from the context when we refer to this general meaning of $|$ and when to the more narrow meaning given in the Introduction. We define the *higher-order invariants* inductively:

$$\begin{aligned} V^{\Gamma,1} &= V^{\Gamma} = \{v \in V : v|\gamma = v \text{ for all } \gamma \in \Gamma\}, \\ V^{\Gamma,q+1} &= \{v \in V : v|(\gamma - 1) \in V^{\Gamma,q} \text{ for all } \gamma \in \Gamma\}. \end{aligned} \tag{3-1}$$

We set $V^{\Gamma,0} = \{0\}$.

Now let Γ be finitely generated and let I be the augmentation ideal in the group ring $\mathbb{C}[\Gamma]$, generated by $\gamma - 1$ with $\gamma \in \Gamma \setminus \{1\}$. A fundamental role in this paper will be played by the map

$$\mathsf{m}_q : V^{\Gamma,q+1} \to \hom_{\mathbb{C}[\Gamma]}(I^{q+1}\backslash I^q, V^\Gamma).$$

To define it, we first quote from [Deitmar 2009] (before Proposition 1.2):

$$V^{\Gamma,q} \cong \hom_{\mathbb{C}[\Gamma]}(I^q\backslash\mathbb{C}[\Gamma], V). \tag{3-2}$$

The isomorphism is induced by the map $\varphi \mapsto \varphi(1)$ from $\hom_{\mathbb{C}[\Gamma]}(\mathbb{C}[\Gamma], V)$ to V. Next, we note that $I^{q+1}\backslash I^q$ is generated by

$$I^{q+1} + (\gamma_1 - 1)\dots(\gamma_q - 1),$$

with $\gamma_i \in \Gamma$. To each $v \in V^{\Gamma,q+1}$ we associate the map on $I^{q+1}\backslash I^q$ sending this element to $v\,|\,(\gamma_1 - 1)\dots(\gamma_q - 1)$. This map is well-defined because

$$v\,|\,(\gamma_1 - 1)\dots(\gamma_{q+1} - 1) = 0.$$

In this way, we obtain a map m_q from $V^{\Gamma,q+1}$ to

$$\hom_{\mathbb{C}[\Gamma]}(I^{q+1}\backslash I^q, V) \cong \hom_{\mathbb{C}[\Gamma]}(I^{q+1}\backslash I^q, V^\Gamma)$$

(since the action induced on $I^{q+1}\backslash I^q$ by the operation of Γ is trivial). It is easy to see that the kernel of m_q is $V^{\Gamma,q}$, and thus we obtain the exact sequence

$$0 \longrightarrow V^{\Gamma,q} \longrightarrow V^{\Gamma,q+1} \xrightarrow{\mathsf{m}_q} \hom_{\mathbb{C}[\Gamma]}(I^{q+1}\backslash I^q, V). \tag{3-3}$$

The map m_q may or may not be surjective, and we will interpret the phrase "as large as possible" as surjectivity of m_q for all $q \in \mathbb{N}$.

Definition 3.1. Let Γ be a finitely generated group. We will call a Γ-module V *maximally perturbable* if the linear map $\mathsf{m}_q : V^{\Gamma,q+1} \to \hom_{\mathbb{C}[\Gamma]}(I^{q+1}\backslash I^q, V^\Gamma)$ is surjective for all $q \geq 1$.

A reformulation of this definition, which is occasionally easier to use, uses the finite dimension

$$n(\Gamma, q) := \dim_{\mathbb{C}}(I^{q+1}\backslash I^q). \tag{3-4}$$

V is maximally perturbable if and only if $V^{\Gamma,q+1}/V^{\Gamma,q} \cong (V^\Gamma)^{n(\Gamma,q)}$ for all $q \in \mathbb{N}$.

The numbers $n(\Gamma, q)$ are determined by the algebraic structure of the group Γ, and increase quickly with q for many discrete $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$. For a maximally perturbable Γ-module with a nonzero space V^Γ of invariants, the sizes of the spaces of higher-order invariants $V^{\Gamma,q}$ also increase quickly, restricted only by the exact sequence in (3-3).

In [Diamantis and Sim 2008], higher-order cusp forms of weight k for a discrete group Γ are considered in the space of holomorphic functions on $\mathfrak{H}$ with exponential decay at the cusps that moreover are invariant under the parabolic transformations. The dimensions of these spaces are computed and generally turn out to be strictly smaller than $n(\Gamma, q)$ times the dimension of the spaces of invariants. So the corresponding Γ-module is not maximally perturbable.

The map $\varphi \mapsto \big((\gamma_1, \dots, \gamma_1) \mapsto \varphi \,\big|\, (\gamma_1 - 1) \dots (\gamma_q - 1)\big)$ induces an isomorphism $\hom_{\mathbb{C}[\Gamma]}(I^{q+1} \backslash I^q, V^\Gamma) \cong \mathrm{Mult}^q(\Gamma, V^\Gamma)$, where $\mathrm{Mult}^q(\Gamma, V^\Gamma)$ is the space of maps $\Gamma^q \to V^\Gamma$ inducing group homomorphisms $\Gamma \to \mathbb{C}$ on each of their coordinates. For a finitely generated group Γ, $\mathrm{Mult}^q(\Gamma, V^\Gamma) \cong \mathrm{Mult}^q(\Gamma, \mathbb{C}) \otimes_{\mathbb{C}} V^\Gamma$, where $\mathrm{Mult}^q(\Gamma, \mathbb{C})$ is the q-th tensor power of $\hom(\Gamma, \mathbb{C})$. This description suggests that it may be useful to consider the following special higher-order invariants:

Definition 3.2. Let $q \in \mathbb{N}$. For any group Γ and any Γ-module V, we call $f \in V^{\Gamma,q}$ a *perturbation* of $\varphi \in V^\Gamma$ if there exists $\mu_f \in \mathrm{Mult}^q(\Gamma, \mathbb{C})$ such that for all $\gamma_1, \dots, \gamma_q \in \Gamma$,

$$f \,|\, (\gamma_1 - 1) \dots (\gamma_q - 1) = \mu_f(\gamma_1, \dots, \gamma_q)\varphi. \tag{3-5}$$

We call a perturbation *commutative* if μ_f is invariant under all permutations of its arguments. If not, we call it noncommutative.

3B. *Canonical generators.* In this section, we recall the "canonical generators" of cofinite discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$, and use them to show that certain modules are maximally perturbable.

Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be a cofinite discrete group of motions in the upper half-plane $\mathfrak{H}$. A system of *canonical generators* for Γ consists of:

- Parabolic generators $P_1, \dots, P_{n_{\mathrm{par}}}$, each conjugate in $\mathrm{PSL}_2(\mathbb{R})$ to $\pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. We shall assume that Γ has cusps: $n_{\mathrm{par}} \geq 1$.
- Elliptic generators $E_1, \dots, E_{n_{\mathrm{ell}}}$, with $n_{\mathrm{ell}} \geq 0$. Each E_j is conjugate to $\pm\left(\begin{smallmatrix} \cos(\pi/v_j) & \sin(\pi/v_j) \\ -\sin(\pi/v_j) & \cos(\pi/v_j) \end{smallmatrix}\right)$ in $\mathrm{PSL}_2(\mathbb{R})$ for some $v_j \geq 2$.
- Hyperbolic generators $H_1, \dots, H_{2g}$, with $g \geq 0$, each conjugate in $\mathrm{PSL}_2(\mathbb{R})$ to the image $\pm\left(\begin{smallmatrix} t & 0 \\ 0 & t^{-1} \end{smallmatrix}\right)$, $t > 1$, of a diagonal matrix.

See, for example, [Lehner 1964, Chapter VII.4, p. 241] or [Petersson 1948, §3]. The relations are given by the condition that each $E_j^{v_j}$ equals Id for $j = 1, \dots, n_{\mathrm{ell}}$, and one large relation

$$P_1 \dots P_{n_{\mathrm{par}}} E_1 \dots E_{n_{\mathrm{ell}}} [H_1, H_2] \dots [H_{2g-1} H_{2g}] = \mathrm{Id}. \tag{3-6}$$

The choice of canonical generators is not unique, but the numbers n_{par}, n_{ell} and g, and the elliptic orders $v_1, \dots, v_{n_{\mathrm{ell}}}$, are uniquely determined by Γ.

Each group homomorphism $\Gamma \to \mathbb{C}$ vanishes on the E_j, and is determined by its values on $H_1, \dots, H_{2g}$, $P_1, \dots, P_{n_{\mathrm{par}}-1}$; hence, since Γ has cusps,

$$\dim \hom(\Gamma, \mathbb{C}) = n_{\mathrm{par}} - 1 + 2g. \tag{3-7}$$

We put $t(\Gamma) = n_{\text{par}} + 2g$ and denote P_1 by $A_1, \ldots, P_{n_{\text{par}}-1}$ by $A_{n_{\text{par}}-1}$, H_1 by $A_{n_{\text{par}}}, \ldots, H_{2g}$ by $A_{t(\Gamma)-1}$. The group Γ is generated by $E_1, \ldots, E_{n_{\text{ell}}}$ and $A_1, \ldots, A_{t(\Gamma)-1}$.

For the modular group, we have $n_{\text{par}} = 1$, $P_1 = \pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $n_{\text{ell}} = 2$, $E_1 = \pm\left(\begin{smallmatrix} 1 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, $E_2 = \pm S := \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $g = 0$, and hence $\hom(\Gamma_{\text{mod}}, \mathbb{C}) = \{0\}$ and $t(\Gamma_{\text{mod}}) = 1$.

In the sequel, we will need a basis for $I^{q+1} \backslash I^q$. Arguing as in Lemma 2.1 in [Deitmar 2009], we can deduce that the elements

$$\boldsymbol{b}(\boldsymbol{i}) = (A_{\boldsymbol{i}(1)} - 1) \ldots (A_{\boldsymbol{i}(q)} - 1), \tag{3-8}$$

where $\boldsymbol{i}$ runs over all $(t(\Gamma) - 1)$-tuples of elements of $\{1, \ldots, t(\Gamma) - 1\}$, form a basis of $I^{q+1} \backslash I^q$. We do not give a proof here, since it follows from the more general result in Proposition 5.1.

4. Maass forms

We turn to spaces of functions on the upper half-plane that contain the classical holomorphic automorphic forms and the more general Maass forms. The first main results of this paper are stated in Theorems 4.2 and 4.3. In Section 4C, we give some explicit examples of higher-order Maass forms.

4A. ***General Maass forms.*** Let Γ be a cofinite discrete subgroup Γ of the group $G = \mathrm{PSL}_2(\mathbb{R})$. For each cusp κ, we choose $g_\kappa \in \mathrm{PSL}_2(\mathbb{R})$ such that

$$\kappa = g_\kappa \infty \qquad \text{and} \qquad g_\kappa^{-1} \Gamma_\kappa g_\kappa = \left\{ \pm\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z} \right\}. \tag{4-1}$$

Here, Γ_κ is the set of elements of Γ fixing κ. The elements g_κ are determined up to right multiplication by elements $\pm\left(\begin{smallmatrix} a & b \\ 0 & a^{-1} \end{smallmatrix}\right) \in G$. We choose the g_κ for cusps in the same Γ-orbit so that $g_{\gamma\kappa} \in \gamma g_\kappa \Gamma_\infty$.

We further consider a generalization of the action $|$ considered in the last section. For a fixed k and for a $f : \mathfrak{H} \to \mathbb{C}$, we set

$$f\big|_k \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(z) = (cz + d)^{-k} f\left(\frac{az+b}{cz+d}\right). \tag{4-2}$$

We finally set

$$L_k = -y^2 \partial_x^2 - y^2 \partial_y^2 + iky\partial_x - ky\partial_y + \frac{k}{2}\left(1 - \frac{k}{2}\right). \tag{4-3}$$

With this notation, we have:

Definition 4.1. Let $k \in 2\mathbb{Z}$ and $\lambda \in \mathbb{C}$.

i) $\mathcal{M}_k(\Gamma, \lambda)$ denotes the space of smooth functions $f : \mathfrak{H} \to \mathbb{C}$ such that $L_k f = \lambda f$ and for which there is some $a \in \mathbb{R}$ such that

$$f(g_\kappa(x + iy)) = \mathrm{O}(y^a), \qquad (y \to \infty) \tag{4-4}$$

uniformly for x in compact sets in $\mathbb{R}$, for *all* cusps κ of Γ.

ii) $\mathscr{E}_k(\Gamma, \lambda)$ denotes the space of smooth functions f such that $L_k f = \lambda f$ and for which there is some $a \in \mathbb{R}$ such that

$$f(g_\kappa(x+iy)) = \mathrm{O}(e^{ay}), \qquad (y \to \infty) \tag{4-5}$$

uniformly for x in compact sets in $\mathbb{R}$, for *all* cusps κ of Γ.

iii) We denote the invariants in these spaces by

$$E_k(\Gamma, \lambda) := \mathscr{E}_k(\Gamma, \lambda)^\Gamma \qquad \text{and} \qquad M_k(\Gamma, \lambda) := \mathscr{M}_k(\Gamma, \lambda)^\Gamma. \tag{4-6}$$

We call the elements of $E_k(\Gamma, \lambda)$ (resp. $M_k(\Gamma, \lambda)$) *Maass forms of polynomial (resp. exponential) growth of weight k and eigenvalue $\lambda \in \mathbb{C}$ for Γ.*

Remarks. i) Since L_k is elliptic, all its eigenfunctions are automatically real-analytic. (See, for example, [Lang 1975, §5 of Appendix A4], and the references therein.) If f is holomorphic, then it is an eigenfunction of L_k with eigenvalue $(k/2)(1-k/2)$.

ii) The space $M_k(\Gamma, \lambda)$ is known to have finite dimension. The space $E_k(\Gamma, \lambda)$ has, for groups Γ with cusps, infinite dimension. The subspace of $E_k(\Gamma, \lambda)$ corresponding to a fixed value of a in the bound $\mathrm{O}(e^{ay})$ has finite dimension. (The first statement is due to Maass. See [Maass 1983, p. 190, Theorem 28] for the case that $\Gamma \subset \mathrm{PSL}_2(\mathbb{Z})$. The three statements hold for all Γ. See, for instance, the discussion in Section 9.5 of [Bruggeman 1994].)

iii) In an alternative definition, suitable for functions not necessarily holomorphic, one replaces the Maass forms f as defined above by $h(z) = y^{k/2} f(z)$. Then invariance under (4-2) becomes invariance under the action

$$f\big|_k \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(z) = e^{-ik \arg(cz+d)} f\left(\frac{az+b}{cz+d}\right), \tag{4-7}$$

and the eigenproperty in terms of the Laplacian

$$\left(-y^2\partial_x^2 - y^2\partial_y^2 + iky\partial_x\right)h = \lambda h. \tag{4-8}$$

The formulation of the growth conditions remains unchanged. Now antiholomorphic automorphic forms $a(z)$ of weight k give Maass forms $h(z) = y^{k/2} a(z)$ of weight $-k$.

Our main result for general Maass forms on $\mathfrak{H}$ is:

Theorem 4.2. *Let Γ be a cofinite discrete group of motions in $\mathfrak{H}$ with cusps. Then the Γ-module $\mathscr{E}_k(\Gamma, \lambda)$ is maximally perturbable for each $k \in 2\mathbb{Z}$ and each $\lambda \in \mathbb{C}$.*

In the course of the proof in Section 8, we will see that even if we start with Maass forms with polynomial growth, the construction of higher-order invariants will lead us to functions that have exponential growth.

4B. ***Holomorphic automorphic forms.*** For even k, the space $\mathscr{E}_k(\Gamma, \lambda_k)$, with $\lambda_k = (k/2)(1-k/2)$, contains the subspace $\mathscr{E}_k^{\mathrm{hol}}(\Gamma, \lambda_k)$, where the condition $L_k f = \lambda_k f$ is replaced by the stronger condition that f is holomorphic. In the alternative definition, condition (4-8) is replaced by the condition that $z \mapsto y^{-k/2} f(z)$ is holomorphic. The space $\mathscr{E}_k^{\mathrm{hol}}(\Gamma, \lambda_k)$ is a Γ-submodule of $\mathscr{E}_k(\Gamma, \lambda_k)$. We also have the Γ-submodule $\mathscr{M}_k^{\mathrm{hol}}(\Gamma, \lambda_k) = \mathscr{M}_k(\Gamma, \lambda_k) \cap \mathscr{E}_k^{\mathrm{hol}}(\Gamma, \lambda_k)$ of $\mathscr{M}_k(\Gamma, \lambda_k)$.

The space $\mathscr{M}_k^{\mathrm{hol}}(\Gamma, \lambda_k)^\Gamma$ is the usual space of entire weight-k automorphic forms for Γ, and $\mathscr{E}_k^{\mathrm{hol}}(\Gamma, \lambda_k)^\Gamma$ is the space of meromorphic automorphic forms with singularities only at cusps. Sometimes, as in [Bruinier et al. 2008], the elements of $\mathscr{E}_k^{\mathrm{hol}}(\Gamma, \lambda_k)^\Gamma$ are called *weakly holomorphic*. There the elements of $\mathscr{E}_k(\Gamma, \lambda_k)^\Gamma$ are called *harmonic weak Maass forms*. We prefer to use the term *harmonic* for Maass forms in $\mathscr{E}_k(\Gamma, 0)^\Gamma$. (Note that $\lambda_k \neq 0$ for $k \neq 0, 2$.)

Our main result for holomorphic automorphic forms on $\mathfrak{H}$ is:

Theorem 4.3. *Let Γ be a cofinite discrete group of motions in $\mathfrak{H}$ with cusps. Then $\mathscr{E}_k^{\mathrm{hol}}(\Gamma, k/2 - k^2/4)$ is maximally perturbable for each $k \in 2\mathbb{Z}$.*

4C. ***Examples of harmonic and holomorphic forms of orders 2 and 3.*** According to Theorems 4.2 and 4.3, there are plenty of examples of higher-order Maass forms for cofinite groups with cusps for which $\dim_{\mathbb{C}} \hom(\Gamma, \mathbb{C}) \geq 1$. (See the discussion following Definition 3.1.) It is, however, not very easy to exhibit explicit examples.

For the modular group $\Gamma_{\mathrm{mod}} = \mathrm{PSL}_2(\mathbb{Z})$, the space $\hom(\Gamma_{\mathrm{mod}}, \mathbb{C})$ is zero. Hence, it does not accept higher-order invariants. For the commutator subgroup $\Gamma_{\mathrm{com}} = [\Gamma_{\mathrm{mod}}, \Gamma_{\mathrm{mod}}]$, we will employ three different approaches to exhibit full sets of perturbations of 1 (as defined in Definition 3.2) of orders 2 and 3. A reader only interested in the existence of higher-order forms may prefer to skip this subsection.

4C1. *Holomorphic perturbation of* 1. In [Lehner 1964, Chapter XI, §3E, p. 362], one finds various facts concerning Γ_{com}. It is freely generated by $D = \pm\left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and $C = \pm\left(\begin{smallmatrix} 2 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. It has no elliptic elements, and one cuspidal orbit $\Gamma_{\mathrm{com}}\infty = \mathbb{P}^1_{\mathbb{Q}}$. The group $(\Gamma_{\mathrm{com}})_\infty$ fixing ∞ is generated by $\pm\left(\begin{smallmatrix} 1 & 6 \\ 0 & 1 \end{smallmatrix}\right)$. We have $t(\Gamma_{\mathrm{com}}) = 3$.

The space of holomorphic cusp forms of weight 2 has dimension $g = 1$. We use the basis element η^4 (power of the Dedekind eta function). The map

$$H(z) = -2\pi i \int_\infty^z \eta(\tau)^4 \, d\tau = -6e^{\pi i z/3} + \mathrm{O}\big(e^{7\pi i z/3}\big) \qquad \text{(4-9)}$$

induces an embedding of $\Gamma_{\mathrm{com}} \backslash \mathfrak{H}$ into an elliptic curve, which can be described as $\mathbb{C}/\Lambda$, with

$$\Lambda = \varpi\, \mathbb{Z}[\rho], \qquad \varpi = \pi^{1/2}\Gamma\big(\tfrac{1}{6}\big)\big/\big(6\sqrt{3}\,\Gamma\big(\tfrac{2}{3}\big)\big), \qquad \rho = e^{\pi i/3}. \qquad \text{(4-10)}$$

(See computations in [Bruggeman 1994, §15.2–3].) The map H maps $\mathfrak{H}$ onto $\mathbb{C}\smallsetminus\Lambda$, and satisfies for $\gamma\in\Gamma_{\mathrm{com}}$

$$H(\gamma z)=H(z)+\lambda(\gamma),\qquad \lambda(\gamma)=-2\pi i\int_{\infty}^{\gamma\infty}\eta(\tau)^4\,d\tau, \tag{4-11}$$

where $\lambda(C)=\rho\varpi$ and $\lambda(D)=\bar\rho\varpi$. So the lattice Λ is the image of $\lambda:\Gamma_{\mathrm{com}}\to\mathbb{C}$, and $\hom(\Gamma_{\mathrm{com}},\mathbb{C})=\mathrm{Mult}^1(\Gamma_{\mathrm{com}},\mathbb{C})$ has λ, $\bar\lambda$ as a basis. We note that the kernel $\ker(\lambda)$ is a subgroup with infinite index in Γ_{com}; it is in fact the commutator subgroup of Γ_{com}. The element $\pm\left(\begin{smallmatrix}1&6\\0&1\end{smallmatrix}\right)$ generating the subgroup of Γ_{com} fixing ∞ is in $\ker(\lambda)$. Since $\ker(\lambda)$ has no elliptic elements, composition with H gives a bijection from the holomorphic functions on $\mathbb{C}\smallsetminus\Lambda$ to the holomorphic $\ker(\lambda)$-invariant functions on $\mathfrak{H}$.

Clearly, H is a holomorphic second-order perturbation of 1 with linear form λ. It is also a *harmonic perturbation* of 1, that is, a perturbation which is harmonic as a function. By conjugation, we obtain the antiholomorphic harmonic perturbation of 1 with linear form $\bar\lambda$.

According to Theorem 4.3, there should also be a holomorphic second-order perturbation of 1 with a linear form that is linearly independent of λ. Here we can use the Weierstrass zeta function

$$\zeta(u;\,\Lambda)=\frac{1}{u}+\sum_{\omega\in\Lambda}{}'\Big(\frac{1}{u-\omega}+\frac{1}{\omega}+\frac{u}{\omega^2}\Big). \tag{4-12}$$

See, for example, [Koecher and Krieg 1998, Chapter I, §6]. It is holomorphic on $\mathbb{C}\smallsetminus\Lambda$ and satisfies $\zeta(u+\omega;\,\Lambda)=\zeta(u;\,\Lambda)+\mathsf{h}(\omega)$ for all $\omega\in\Lambda$, where $\mathsf{h}\in\hom(\Lambda,\mathbb{C})$ is linearly independent of $\omega\mapsto\omega$. (The classical notation for h is η. We write h to avoid confusion with the Dedekind eta function.) Pulling back this zeta function to $\mathfrak{H}$, we get a second-order holomorphic perturbation of 1:

$$W(z)=\zeta(H(z);\,\Lambda), \tag{4-13}$$

with the linear form $\gamma\mapsto\mathsf{h}(\lambda(\gamma))$. The Laurent expansion of the Weierstrass zeta function at 0 starts with $\zeta(u;\,\Lambda)=u^{-1}+\mathrm{O}(u^3)$. Hence, W has a Fourier expansion at ∞ starting with

$$W(z)=-\tfrac{1}{6}e^{-\pi iz/3}+\mathrm{O}(e^{\pi iz}). \tag{4-14}$$

This shows that W has exponential growth at the cusps.

We may carry this out also for holomorphic forms of order 3, to obtain the following commutative perturbations of 1 of order 3:

(4-15)

f	$H(z)^2$	$H(z)\,W(z)$	$W(z)^2$
μ_f	$2\,\lambda\otimes\lambda$	$\lambda\otimes(\mathsf{h}\circ\lambda)+(\mathsf{h}\circ\lambda)\otimes\lambda$	$2(\mathsf{h}\circ\lambda)\otimes(\mathsf{h}\circ\lambda)$

We know that there also exist noncommutative holomorphic perturbations of order 3. To find an explicit example, we have to work on $\mathfrak{H}$, since the group Λ acting on $\mathbb{C}$ is abelian.

The closed holomorphic 1-forms

$$\omega = -2\pi i \eta(\tau)^4 \, d\tau \quad \text{and} \quad \omega_1 = -2\pi i W(\tau)\, \eta(\tau)^4 \, d\tau$$

on $\mathfrak{H}$ transform as follows under Γ_{com}:

$$\omega | \gamma = \omega, \quad \omega_1 | \gamma = \omega_1 + \mathsf{h}(\lambda(\gamma))\omega. \tag{4-16}$$

For an arbitrary base point $z_0 \in \mathfrak{H}$, we put

$$K(z) = \int_{z_0}^{z} \omega_1. \tag{4-17}$$

This defines a holomorphic function on $\mathfrak{H}$ that satisfies, for $\gamma \in \Gamma_{\mathrm{com}}$,

$$K | (\gamma - 1)(z) = \int_{z}^{\gamma z} \omega_1,$$

and hence for $\gamma, \delta \in \Gamma_{\mathrm{com}}$,

$$\begin{aligned} K | (\gamma - 1)(\delta - 1)(z) &= \left(\int_{\gamma z}^{\gamma \delta z} - \int_{z}^{\delta z} \right) \omega_1 = \int_{z}^{\delta z} \omega_1 | \gamma - \int_{z}^{\delta z} \omega_1 \\ &= \mathsf{h}(\lambda(\gamma)) \int_{z}^{\delta z} \omega = \mathsf{h}(\lambda(\gamma))\lambda(\delta). \end{aligned}$$

Thus, we have a holomorphic third-order noncommutative perturbation K of 1 with nonsymmetric multilinear form $(\mathsf{h} \circ \lambda) \otimes \lambda$. Since holomorphic forms are harmonic in weight zero, these perturbations are also harmonic perturbations of 1.

4C2. *Iterated integrals.* The construction of the third-order form K in (4-17) is closely related to the iterated integrals used in [Diamantis and Sreekantan 2006] to prove maximal perturbability of spaces of smooth functions.

The idea is that we have two closed Γ_{com}-invariant differential forms on $\mathfrak{H}$, $dH(z) = \omega = -2\pi i \eta(z)^4 \, dz$ and

$$\omega_0 = dW(z) = -\wp(H(z))\, d(H(z)),$$

where $\wp(u; \Lambda) = -(d/du)\zeta(u; \Lambda)$ is the Weierstrass $\wp$-function. If $t \mapsto z(t)$, $0 \le t \le 1$ is a path in $\mathfrak{H}$ from z_0 to z_1, then

$$\begin{aligned} \int_{t_2=0}^{1} \int_{t_1=0}^{t_2} \omega_0(z(t_1))\omega(z(t_2)) &= \int_{t_2=0}^{1} \bigl(W(z(t_2)) - W(z_0) \bigr)\, dH(z(t_2)) \\ &= -2\pi i \int_{t=0}^{1} W(z(t))\eta(z(t))^4 z'(t)\, dt \\ &\qquad\qquad - W(z_0)\bigl(H(z_1) - H(z_0)\bigr) \\ &= K(z_1) - W(z_0)\bigl(H(z_1) - H(z_0)\bigr) \end{aligned}$$

depends only on z_0 and z_1, not on the actual path. For a fixed base point z_0, the holomorphic function $z_1 \mapsto W(z_0)\big(H(z_1) - H(z_0)\big)$ is invariant of order 2. So up to lower-order terms, the invariant K is given by an iterated integral, as in (3) of [Diamantis and Sreekantan 2006]; see also [Chen 1971].

4C3. *Differentiation of families.* We start by considering a general finitely generated group Γ acting on a space X. We will use the notation $f\,|\,\gamma(x) = f(\gamma x)$ for the action induced on functions defined on X. We consider a family of characters of Γ of the form $\chi_r(\gamma) = e^{ir\cdot\alpha(\gamma)}$, where $r \cdot \alpha(\gamma) = r_1\alpha_1(\gamma) + \cdots + r_n\alpha_n(\gamma)$ for $\alpha_1, \ldots, \alpha_n \in \hom(\Gamma, \mathbb{R})$ and r varying over an open set U in $\mathbb{R}^n$. In this way, χ_r is a family of unitary characters.

We consider a C^∞ family $r \mapsto f_r$ on a neighborhood $U \subset \mathbb{R}^n$ of 0 of functions $X \to \mathbb{C}$ that satisfy

$$f_r(\gamma x) = \chi_r(\gamma) f_r(x), \qquad (\gamma \in \Gamma). \tag{4-18}$$

We assume that χ_0 is the trivial character and that f_0 is a Γ-invariant function f.

We now set $h(x) = \partial_{r_j} f_r(x)\big|_{r=0}$, for one of the coordinates of r. The transformation behavior gives $h(\gamma x) = i\alpha_j(\gamma) f(x) + h(x)$, or, rewritten,

$$h\,|\,\gamma - h = i\alpha_j(\gamma) f.$$

The function h is a second-order perturbation of f, with $i\alpha_j$ as the corresponding element of $\hom(\Gamma, \mathbb{C})$. This can be generalized using a routine inductive argument:

Proposition 4.4. *For all multi-indices $a \in \mathbb{N}^n$, the derivative*

$$f^{(a)}(x) := \partial_r^a f_r(x)\big|_{r=0}$$

is a commutative perturbation of f with order $1 + |a|$. Here $\partial_r^a = \partial_{r_1}^{a_1} \ldots \partial_{r_n}^{a_n}$ and $|a| = a_1 + a_2 + \cdots + a_n$.

Remark. Proposition 4.4 shows that commutative perturbations can arise as infinitesimal perturbations of a family of automorphic forms. That is our motivation to use the word *perturbation* in Definition 3.2.

Application to harmonic perturbations of 1. We use the method of differentiation of families to produce explicit harmonic higher-order forms for Γ_{com} of order 3. We employ families studied in [Bruggeman 1994].

Since Γ_{com} is free on the generators $C = \pm\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ and $D = \pm\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, the character group of Γ_{com} is isomorphic to $\mathbb{C}^* \times \mathbb{C}^*$. We can parametrize the characters by

$$\chi_{v,w}(\gamma) = e^{iv\lambda(\gamma) + iw\overline{\lambda(\gamma)}}, \tag{4-19}$$

where (v, w) runs through $\mathbb{C}^2$, and where $\lambda \in \hom(\Gamma_{\mathrm{com}}, \mathbb{C})$ is as defined in (4-11). We are interested only in (v, w) in a neighborhood of $0 \in \mathbb{C}^2$.

In [Bruggeman 1994, §15.5], it is shown that there is a meromorphic Eisenstein family $E(v, w, s)$ of automorphic forms for Γ_{com}, with the character $\chi_{v,w}$ and eigenvalue $\frac{1}{4}-s^2$ for $\omega_0=-y^2(\partial_x^2+\partial_y^2)$. (In [Bruggeman 1994], the discussion of the family E is made in the context of families of automorphic forms of varying weight which are thus defined on the covering group $\tilde{\Gamma}_{\mathrm{com}}$. However, in §15.5 the weight is zero, and the automorphic forms are, in effect, on the discrete group Γ_{com}.) The restriction to $s=\frac{1}{2}$ exists [Bruggeman 1994, §15.6] and forms a meromorphic family $(v, w)\mapsto f(v, w; z)$ on $\mathbb{C}^2$ such that $f(v, w; \gamma z)=\chi_{v,w}(\gamma)f(v, w; z)$, and $L_0 f(v, w; z)=0$ for the dense set of (v, w) at which f is holomorphic. There is a meromorphic family $(v, w)\mapsto h(v, w; \cdot)$ on $\mathbb{C}$ such that

$$f(v, w; z)=h(v, w; H(z)),$$

satisfying $h(v, w; u+\lambda)=e^{iv\lambda+iw\bar{\lambda}}h(v, w; u)$ [Bruggeman 1994, §15.1–6]. Chapter 15 of [Bruggeman 1994] gives a complicated but explicit construction (obtained with the help of D. Zagier) of such a family h with Jacobi theta functions.

Specifically, in §15.6.11, the function h is expressed as a sum

$$h(v, w; u)=G_{(v+w)\varpi/2\pi}(u, w)+G_{-(v+w)\varpi/2\pi}(-\bar{u}, -v), \tag{4-20}$$

where the function $G_\mu(u, w)$, for $\mu\notin\mathbb{Z}$ and $0<\operatorname{Im} u<\frac{1}{2}\varpi\sqrt{3}$, is given by

$$G_\mu(u, w)=\sum_{m=-\infty}^{\infty}\frac{1}{\mu+m}\,\frac{\xi^{\mu+m}}{\eta q^m-1}, \tag{4-21}$$

with $q=-e^{-\pi\sqrt{3}}$, $\xi=e^{2\pi i u/\varpi}$, and $\eta=e^{-w\varpi\sqrt{3}}$. We compute a part of the expansion in powers of v and w at $v=w=0$. With the substitution $u=H(z)$, the coefficients provide us with higher-order harmonic modular forms for Γ_{com}. Some of these we have seen above. Denoting $f=-2\pi/\varpi^2\sqrt{3}$, we find:

term of	on $\mathbb{C}$	on $\mathfrak{H}$
1	f	f (constant function)
v	ifu	$ifH(z)$
w	$if\bar{u}$	$if\overline{H(z)}$
v^2	$(-f/2)u^2$	$(-f/2)H(z)^2$
w^2	$(-f/2)\bar{u}^2$	$(-f/2)\overline{H(z)}^2$

(4-22)

The coefficient of vw gives a third-order form

$$\begin{aligned} b_{1,1}(u) :=& \frac{\pi}{\sqrt{3}}\Big(\Big(\frac{u}{\varpi}-\frac{i\sqrt{3}}{2}\Big)^2+\Big(\frac{\bar{u}}{\varpi}+\frac{i\sqrt{3}}{2}\Big)^2+1\Big)\\ &+S(u)+S(\varpi\rho-u)+S(-\bar{u})+S(\varpi\rho+\bar{u}), \end{aligned} \tag{4-23}$$

with
$$S(u) := \sum_{m=1}^{\infty} \frac{e^{2\pi i m u/\varpi}}{m\,(q^m-1)}, \qquad \rho = \frac{1}{2} + \frac{i}{2}\sqrt{3}.$$

By $B_{1,1}(z) = b_{1,1}(H(z))$ we denote the corresponding harmonic third-order perturbation of 1 on $\mathfrak{H}$. The way $B_{1,1}$ has been derived, together with the proof of Proposition 4.4, ensures that it is a perturbation of 1 with a multilinear form that is a multiple of $\lambda \otimes \bar{\lambda} + \bar{\lambda} \otimes \lambda$.

However, $b_{1,1}(u)$ is represented by (4-23) only on the region $0 < \operatorname{Im} u < \frac{1}{2}\varpi\sqrt{3}$. By further computations, we arrive at expressions for it on larger regions, and can determine the associated bilinear form. We then see that the pull-back $-f^{-1}B_{1,1} = -f^{-1}b_{1,1} \circ H$ is a harmonic commutative perturbation of 1 for the multilinear form μ determined by the following values at the generators C and CD of Γ_{com}:

$$\mu(g,h) = \begin{cases} 2\varpi^2 & \text{if } g = h = C \text{ or } CD, \\ \varpi^2 & \text{if } g = C,\ h = CD, \text{ or if } g = CD,\ h = C. \end{cases} \tag{4-24}$$

We have used the values of λ given below (4-11). With these values at the generators, μ coincides with $\lambda \otimes \bar{\lambda} + \bar{\lambda} \otimes \lambda$ as predicted above by the way $B_{1,1}$ was constructed.

Proposition 4.4 shows that differentiation of families produces only commutative perturbations. However, by Theorem 4.2, there are noncommutative third-order harmonic perturbations of 1. We can obtain such perturbations from $B_{1,1}$ upon decomposing it as $B_{1,1} = A + B$ for a holomorphic function A and an antiholomorphic function B. Specifically, in view of (4-23), for those $z \in \mathfrak{H}$ for which $H(z)$ is in the upper half of the fundamental hexagon for $\mathbb{C}/\Lambda$, we can set

$$\begin{aligned} A(z) &= \frac{\pi}{2\sqrt{3}} + \frac{\pi}{\sqrt{3}}\left(\frac{H(z)}{\varpi} - \frac{i\sqrt{3}}{2}\right)^2 + S(H(z)) + S\big(\varpi\rho - H(z)\big), \\ B(z) &= \frac{\pi}{2\sqrt{3}} + \frac{\pi}{\sqrt{3}}\left(\frac{\overline{H(z)}}{\varpi} + \frac{i\sqrt{3}}{2}\right)^2 + S\big(-\overline{H(z)}\big) + S\big(\varpi\rho + \overline{H(z)}\big). \end{aligned} \tag{4-25}$$

More computations lead to the conclusion that $-f^{-1}A$ is a noncommutative holomorphic third-order holomorphic perturbation of 1 with multilinear form $\bar{\lambda} \otimes \lambda$, and that the multilinear form of the anticommutative third-order perturbation of 1 given by $-f^{-1}B = -f^{-1}(B_{1,1} - A)$ is $\big(\lambda \otimes \bar{\lambda} + \bar{\lambda} \otimes \lambda\big) - \bar{\lambda} \otimes \lambda = \lambda \otimes \bar{\lambda}$.

5. Universal covering group

5A. ***Universal covering group of*** $\mathbf{SL_2}(\mathbb{R})$**.** For our purpose, it suffices to describe the universal covering group $\tilde{G}$ of $\mathrm{SL}_2(\mathbb{R})$ as the Lie group with underlying analytic space the product $\mathfrak{H} \times \mathbb{R}$ with the group operations uniquely defined by the requirements that $(i, 0)$ be the unit element and that

$$\operatorname{pr}_2(z,\vartheta)=\begin{pmatrix}\sqrt{y} & x/\sqrt{y}\\ 0 & 1/\sqrt{y}\end{pmatrix}\begin{pmatrix}\cos\vartheta & \sin\vartheta\\ -\sin\vartheta & \cos\vartheta\end{pmatrix} \tag{5-1}$$

be a surjective group homomorphism $\operatorname{pr}_2:\tilde{G}\to \mathrm{SL}_2(\mathbb{R})$. We will often use the lift $g\mapsto\tilde{g}:\mathrm{SL}_2(\mathbb{R})\to\tilde{G}$ given by

$$\widetilde{\begin{pmatrix}a & b\\ c & d\end{pmatrix}}=\Bigl(\frac{ai+b}{ci+d},\,-\arg(ci+d)\Bigr),$$

with the convention that the argument takes values in $(-\pi,\pi]$. It satisfies, for all $(z,\vartheta)\in\tilde{G}$,

$$\widetilde{\begin{pmatrix}a & b\\ c & d\end{pmatrix}}(z,\vartheta)=\Bigl(\frac{az+b}{cz+d},\,\vartheta-\arg(cz+d)\Bigr). \tag{5-2}$$

By $\operatorname{pr}:\tilde{G}\to\mathrm{PSL}_2(\mathbb{R})$ we denote the composition of pr_2 and the natural map $\mathrm{SL}_2(\mathbb{R})\to\mathrm{PSL}_2(\mathbb{R})$.

We will use the following homomorphisms of Lie groups:

$$\begin{aligned}
n:\mathbb{R}\to\tilde{G},\quad & n(x)=(x+i,0), & \operatorname{pr}n(x)&=\begin{bmatrix}1 & x\\ 0 & 1\end{bmatrix},\\
a:\mathbb{R}^*_{>0}\to\tilde{G},\quad & a(y)=(iy,0), & \operatorname{pr}a(y)&=\begin{bmatrix}\sqrt{y} & 0\\ 0 & 1/\sqrt{y}\end{bmatrix},\\
k:\mathbb{R}\to\tilde{G},\quad & k(\vartheta)=(0,\vartheta), & \operatorname{pr}k(\vartheta)&=\begin{bmatrix}\cos\vartheta & \sin\vartheta\\ -\sin\vartheta & \cos\vartheta\end{bmatrix}.
\end{aligned} \tag{5-3}$$

The Lie algebra of $\tilde{G}$ is isomorphic to the Lie algebra of $\mathrm{SL}_2(\mathbb{R})$ and of $\mathrm{PSL}_2(\mathbb{R})$. A basis of the complex Lie algebra is $\boldsymbol{W}$, $\boldsymbol{E}^+$, $\boldsymbol{E}^-$, with $\boldsymbol{W}$ corresponding to $\left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$ in the Lie algebra of $\mathrm{SL}_2(\mathbb{R})$, and $\boldsymbol{E}^\pm$ corresponding to $\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)\pm i\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$. The corresponding left-invariant differential operators are, in the coordinates $(x+iy,\vartheta)$ on $\tilde{G}$:

$$\boldsymbol{W}=\partial_\vartheta,\qquad \boldsymbol{E}^\pm=e^{\pm 2i\vartheta}(\pm 2iy\partial_x+2y\partial_y\mp i\partial_\vartheta). \tag{5-4}$$

The *Casimir operator*

$$\omega=-\tfrac{1}{4}\boldsymbol{E}^\pm\boldsymbol{E}^\mp+\tfrac{1}{4}\boldsymbol{W}^2\mp\frac{i}{2}\boldsymbol{W}=-y^2\partial_y^2-y^2\partial_x^2+y\partial_x\partial_\vartheta \tag{5-5}$$

generates the center of the enveloping algebra of the Lie algebra, and determines a differential operator that commutes with left and with right translation.

5B. ***Cofinite discrete subgroups.*** To a cofinite discrete subgroup Γ of $\mathrm{PSL}_2(\mathbb{R})$ we associate its inverse image $\tilde{\Gamma}:=\operatorname{pr}^{-1}\Gamma$ in $\tilde{G}$. This gives a bijective correspondence between cofinite discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$ and cofinite discrete subgroups of $\tilde{G}$ that contain the center $\tilde{Z} = \langle\zeta\rangle$, where $\zeta := k(\pi)$. The projection pr induces an isomorphism $\Gamma \cong \tilde{\Gamma}/\tilde{Z}$.

As an example, we consider the *modular group* $\Gamma_{\mathrm{mod}} = \mathrm{PSL}_2(\mathbb{Z})$, with corresponding group $\tilde{\Gamma}_{\mathrm{mod}} \subset \tilde{G}$. It is known that $\mathrm{PSL}_2(\mathbb{Z})$ is presented by the generators $S = \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and relations $S^2 = (TS)^2 = I$.

Set

$$s := k(-\pi/2) = \widetilde{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}} \qquad \text{and} \qquad t := n(1) = \widetilde{\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}},$$

with $\mathrm{pr}(s) = S$ and $\mathrm{pr}(t) = T$. Then $s^2 = k(-\pi) = \zeta^{-1} \in \tilde{Z}$, so s and t generate $\tilde{\Gamma}_{\mathrm{mod}}$. To determine the relations $s^2 t = ts^2$ and $tstst = s$, one carries out a computation using (5-2). This implies that the linear space $\hom(\tilde{\Gamma}_{\mathrm{mod}}, \mathbb{C})$ has dimension 1, and is generated by $\alpha : t \mapsto \pi/6$, $\alpha : s \mapsto -\pi/2$. For reasons that will become clear later, we take this basis element, and not an integral-valued one.

5C. *Canonical generators.* The canonical generators of Γ induce canonical generators of $\tilde{\Gamma}$:

- Elements $\pi_1, \dots, \pi_{n_{\mathrm{par}}}$ of the form $\pi_j = \tilde{g}_{\kappa_j} n(1) \tilde{g}_{\kappa_j}^{-1}$ fixing a system of representatives $\kappa_1, \dots, \kappa_{n_{\mathrm{par}}}$ of the $\tilde{\Gamma}$-orbits of cusps.
- Elements $\varepsilon_1, \dots, \varepsilon_{n_{\mathrm{ell}}}$ conjugate in $\tilde{G}$ to $k(\pi/v_j)$ with $v_j \geq 2$.
- Elements $\eta_1, \dots, \eta_{2g}$ conjugate in $\tilde{G}$ to elements $a(t_j)$ with $t_j > 1$.
- The generator $\zeta = k(\pi)$ of the center $\tilde{Z}$ of $\tilde{\Gamma}$.

The relations are:

$$\begin{aligned} &\zeta \text{ is central,} \\ &\varepsilon_j^{v_j} = \zeta \text{ for } 1 \leq j \leq n_{\mathrm{ell}}, \\ &\pi_1 \dots \pi_{n_{\mathrm{par}}} \varepsilon_1 \dots \varepsilon_{n_{\mathrm{ell}}} [\eta_1, \eta_2] \dots [\eta_{2g-1}, \eta_{2g}] = \zeta^{2g-2+n_{\mathrm{par}}+n_{\mathrm{ell}}}. \end{aligned} \tag{5-6}$$

The integer $2g - 2 + n_{\mathrm{par}} + n_{\mathrm{ell}}$ is always positive. For these facts, see [Bruggeman 1994, §3.3]

If $n_{\mathrm{ell}} > 0$ or if $2g - 2 + n_{\mathrm{par}} = 1$ and $n_{\mathrm{ell}} = 0$, we do not need ζ as a generator. If $n_{\mathrm{ell}} = 0$, the group $\tilde{\Gamma}$ is free on $\pi_1, \dots, \pi_{n_{\mathrm{par}}-1}, \eta_1, \dots, \eta_{2g}, \zeta$.

Among the canonical generators we single out the following elements: $\alpha_1 = \pi_1$, $\dots, \alpha_{n_{\mathrm{par}}-1} = \pi_{n_{\mathrm{par}}-1}$, $\alpha_{n_{\mathrm{par}}} = \eta_1, \dots, \alpha_{t(\Gamma)-1} = \eta_{2g}$, $\alpha_{t(\Gamma)} = \zeta$. (We recall that $t(\Gamma) = n_{\mathrm{par}} + 2g$.) The α_j together with the ε_j generate $\tilde{\Gamma}$, with $\varepsilon_j^{v_j} = \zeta$ and the centrality of ζ as the sole relations.

For the *modular group* $\tilde{\Gamma}_{\mathrm{mod}}$, we have $n_{\mathrm{par}} = 1$, $n_{\mathrm{ell}} = 2$, $g = 0$, and $t(\Gamma_{\mathrm{mod}}) = 1$. We may take $\pi_1 = t = n(1)$, $\varepsilon_1 = t^{-1}s^{-1}$, and $\varepsilon_2 = s^{-1} = k(\pi/2) = p^{-1}k(\pi/3)p$, with $p = n\left(-\frac{1}{2}\right)a\left(\frac{\sqrt{3}}{2}\right)$.

By I we now denote the augmentation ideal of the group ring $\mathbb{C}[\tilde{\Gamma}]$. In $\mathbb{C}[\tilde{\Gamma}]$ we have the elements

$$\boldsymbol{b}(\boldsymbol{i}) = (\alpha_{\boldsymbol{i}(1)} - 1) \dots (\alpha_{\boldsymbol{i}(q)} - 1), \qquad \boldsymbol{i} \in \{1, \dots, t(\Gamma)\}^q. \tag{5-7}$$

We allow ourselves to use the same notation as in (3-8), since from now on we will use $\tilde{\Gamma}$. The centrality of ζ allows us to move $(\zeta - 1)$ through the product. So it suffices to consider only q-tuples $\boldsymbol{i}$ for which all $\boldsymbol{i}(l) = t(\Gamma)$ occur at the end. Such q-tuples we will call $\tilde{\Gamma}$-*q-tuples.*

Proposition 5.1. *A $\mathbb{C}$-basis of $I^{q+1}\backslash I^q$ is induced by the elements*

$$\boldsymbol{b}(\boldsymbol{i}) = (\alpha_{\boldsymbol{i}(1)} - 1) \dots (\alpha_{\boldsymbol{i}(q)} - 1), \tag{5-8}$$

where $\boldsymbol{i}$ runs over the $\tilde{\Gamma}$-q-tuples.

Proof. The ideal I^q is generated by the products of the form $(\gamma_1 - 1) \dots (\gamma_q - 1)$, with $\gamma_1, \dots, \gamma_q \in \tilde{\Gamma}$ [Deitmar 2009, Lemma 1.1]. With the relation

$$(\gamma\delta - 1) = (\gamma - 1)(\delta - 1) + (\gamma - 1) + (\delta - 1),$$

we can take the γ_j in a system of generators, for instance $\alpha_1, \dots, \alpha_{t(\Gamma)}, \varepsilon_1, \dots, \varepsilon_{n_{\mathrm{ell}}}$. For the elliptic elements ε_j, we use $\zeta - 1 = \sum_{k=0}^{v_j-1} \varepsilon_j^k(\varepsilon_j - 1) \equiv v_j(\varepsilon_j - 1) \mod I^2$ to see that the α_j suffice. (Note that v_j is invertible in $\mathbb{C}$.) Since $\alpha_{t(\Gamma)} = \zeta$ is central, we can move all occurrences of $\zeta - 1$ to the right to see that the $\boldsymbol{b}(\boldsymbol{i})$ in the proposition generate $I^{q+1}\backslash I^q$.

To see that the $\boldsymbol{b}(\boldsymbol{i})$ are linearly independent over $\mathbb{C}$, we proceed in rewriting terms $\xi(\alpha_{\boldsymbol{i}(1)} - 1) \dots (\alpha_{\boldsymbol{i}(q)} - 1)$ by replacing $\xi \in R := \mathbb{C}[\tilde{\Gamma}]$ by $n + \eta$, with $n \in \mathbb{C}$ and $\eta \in I$. In this way, we express each element of I^q as a $\mathbb{C}$-linear combination of products of q factors $\alpha_j - 1$ plus a term in I^N, with $N > q$. To eliminate I^N, we consider the I-adic completion $\hat{R}$ of $\mathbb{C}[\tilde{\Gamma}]$, with closure $\hat{I}^q$ of I^q. Each element of $\hat{I} \supset I$ is a countable sum of products of a complex number and finitely many factors $\alpha_j - 1$. Since $\hat{I}^{q+1}\backslash\hat{I}^q$ and $I^{q+1}\backslash I^q$ are isomorphic, it suffices to prove that the $\boldsymbol{b}(\boldsymbol{i})$ are linearly independent as elements of $\hat{I}^{q+1}\backslash\hat{I}^q$.

We suppose that there are $x_{\boldsymbol{i}} \in \mathbb{C}$ for all q-tuples $\boldsymbol{i}$ such that

$$\sum_{\boldsymbol{i}} x_{\boldsymbol{i}}(\alpha_{\boldsymbol{i}(1)} - 1) \dots (\alpha_{\boldsymbol{i}(q)} - 1) \ \in \ \hat{I}^{q+1}. \tag{5-9}$$

We can write this element of $\hat{I}^{q+1}$ as $\sum_{\boldsymbol{j}} c_{\boldsymbol{j}}\, \xi_{\boldsymbol{j}}$, with $c_{\boldsymbol{j}} \in \mathbb{C}$ and $\xi_{\boldsymbol{j}}$ running over the countably many products $(\alpha_{\boldsymbol{j}(1)} - 1) \dots (\alpha_{\boldsymbol{j}(m)} - 1)$ with m-tuples from $\{1, \dots, t(\Gamma)\}$ for all $m > q$.

We form the ring $N = \mathbb{C}\langle \Xi_1, \dots, \Xi_{t(\Gamma)}\rangle$ of power series in the noncommuting, algebraically independent (over $\mathbb{C}$) variables $\Xi_1, \dots, \Xi_t$, and the two-sided ideal

Z in N generated by the commutators

$$\Xi_j\,\Xi_{t(\Gamma)} - \Xi_{t(\Gamma)}\,\Xi_j, \qquad \text{for } 1 \le j \le t(\Gamma)\,.$$

The quotient ring $M := N/Z$ is noncommutative if $t(\Gamma) \ge 3$. The relations between the generators imply that there is a group homomorphism $\varphi : \tilde{\Gamma} \to M^*$ given by $\varphi(\alpha_j) = 1 + \Xi_j$ for $1 \le j \le t(\Gamma)$, and

$$\varphi(\varepsilon_j) = (1 + \Xi_{t(\Gamma)})^{1/v_j} = \sum_{l \ge 0} \binom{1/v_j}{l} \Xi_{t(\Gamma)}^l.$$

This group homomorphism induces a ring homomorphism $\hat{\varphi} : \hat{R} \to M$, for which

$$\hat{\varphi}(\xi_{\boldsymbol{i}}) = \hat{\varphi}(\alpha_{\boldsymbol{i}(1)} - 1)\,\hat{\varphi}(\alpha_{\boldsymbol{i}(2)} - 1) \ldots \hat{\varphi}(\alpha_{\boldsymbol{i}(|\boldsymbol{i}|)} - 1) = \Xi^{\boldsymbol{i}} := \Xi_{\boldsymbol{i}(1)}\,\Xi_{\boldsymbol{i}(2)} \ldots \Xi_{\boldsymbol{i}(|\boldsymbol{i}|)}.$$

Now we have

$$\sum_{\boldsymbol{i}} x_{\boldsymbol{i}}\,\Xi^{\boldsymbol{i}} = \hat{\varphi}\Big(\sum_{\boldsymbol{i}} x_{\boldsymbol{i}}\,\xi_{\boldsymbol{i}}\Big) = \hat{\varphi}\Big(\sum_{\boldsymbol{j}} c_{\boldsymbol{j}}\xi_{\boldsymbol{j}}\Big) = \sum_{\boldsymbol{j}} c_{\boldsymbol{j}}\,\Xi^{\boldsymbol{j}},$$

where $\boldsymbol{i}$ runs over q-tuples and $\boldsymbol{j}$ runs over countably many tuples with length strictly larger than q. Hence, all $x_{\boldsymbol{i}}$ (and $c_{\boldsymbol{j}}$) vanish. □

So for $\tilde{\Gamma}$ with cusps, the trivial $\tilde{\Gamma}$-module $I^{q+1}\backslash I^q$ is always nontrivial. The dimension is equal to the number of all $\tilde{\Gamma}$-q-tuples. Thus we have

$$\begin{aligned} \dim_{\mathbb{C}}(I^{q+1}\backslash I^q) &= n(\tilde{\Gamma}, q) \\ &= \sum_{m=0}^{q} (t(\Gamma) - 1)^m = \begin{cases} 1 & \text{if } t(\Gamma) = 1, \\ q+1 & \text{if } t(\Gamma) = 2, \\ \dfrac{(t(\Gamma)-1)^{q+1}-1}{t(\Gamma)-2} & \text{if } t(\Gamma) \ge 3. \end{cases} \end{aligned} \tag{5-10}$$

We obtain, for each $\tilde{\Gamma}$-module V, an exact sequence

$$0 \longrightarrow V^{\tilde{\Gamma},q} \longrightarrow V^{\tilde{\Gamma},q+1} \xrightarrow{\mathsf{m}_q} (V^{\tilde{\Gamma}})^{n(\tilde{\Gamma},q)},$$

with

$$(\mathsf{m}_q f)_{\boldsymbol{i}} = f\,|\,(\alpha_{\boldsymbol{i}(1)} - 1) \ldots (\alpha_{\boldsymbol{i}(q)} - 1). \tag{5-11}$$

For the modular group, we have $n_{\text{par}} = 1$, $n_{\text{ell}} = 2$, and $g = 0$, and hence $t(\Gamma_{\text{mod}}) = 1$ and $n(\tilde{\Gamma}_{\text{mod}}, q) = 1$ for all q. So in contrast to Γ_{mod}, for $\tilde{\Gamma}_{\text{mod}}$ we may hope for nontrivial higher-order automorphic forms.

6. Maass forms with generalized weight on the universal covering group

6A. ***The logarithm of the Dedekind eta function.*** In the Introduction, we mentioned that one of the motivating objects for the study of higher-order forms on the universal covering group is the logarithm of the Dedekind eta function. Its branch is fixed by the second of the following expressions:

$$\log \eta(z) = \frac{\pi i z}{12} + \sum_{n=1}^{\infty} \log(1 - e^{2\pi i n z}) = \frac{\pi i z}{12} - \sum_{n=1}^{\infty} \sigma_{-1}(n)\, e^{2\pi i n z}, \tag{6-1}$$

where $\sigma_u(n) = \sum_{d\,|\,n} d^u$. One can show that its behavior under Γ_{mod} is given by

$$\log \eta(z+1) = \log \eta(z) + \frac{\pi i}{12}, \qquad \log \eta\Bigl(-\frac{1}{z}\Bigr) = \log \eta(z) + \tfrac{1}{2} \log z - \frac{\pi i}{4}. \tag{6-2}$$

Except for the term $\frac{1}{2} \log z$, this looks like a second-order holomorphic modular form of weight 0. In the next few sections, we make this precise by generalizing the concept of *weight* of Maass forms, and replacing the group Γ_{mod} by the discrete subgroup $\tilde{\Gamma}_{\mathrm{mod}}$ of the universal covering group of $\mathrm{SL}_2(\mathbb{R})$, using the notation we introduced in the last section.

We first define the following function on $\mathfrak{H} \times \mathbb{R}$:

$$L(z, \vartheta) = \tfrac{1}{2} \log y + 2 \log \eta(z) + i\vartheta. \tag{6-3}$$

With (6-2), we check easily that $L(\gamma(z, \vartheta)) = L(z, \vartheta) + i\alpha(\gamma)$ for $\gamma = t$ and $\gamma = s$, where $\alpha : \tilde{\Gamma}_{\mathrm{mod}} \to (\pi/6)\mathbb{Z}$ is the group homomorphism at the end of Section 5B. Thus, L has the transformation behavior of a second-order invariant in the functions on $\tilde{G}$ for the action by left translation.

Routine computations show that L satisfies $\boldsymbol{E}^- L = 0$, $\boldsymbol{W} L = i$, and $\omega L = \frac{1}{2}$.

6B. ***General Maass forms on the universal covering group.*** The considerations on the function L on $\tilde{G}$ induced by the logarithm of the eta functions lead us to the definition of Maass forms on $\tilde{G}$.

We first establish appropriate notions of weight and holomorphicity. We say that a function f on $\tilde{G}$ has *(strict) weight* $r \in \mathbb{C}$ if $f(z, \vartheta) = e^{ir\vartheta} f(z, 0)$. Such a function is completely determined by the function $f_r(z) = f(z, 0)$ on $\mathfrak{H}$ and satisfies $\boldsymbol{W} f = irf$.

The left translation of f by $\tilde{g}$, with $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, induces an action $|$ of $\tilde{G}$ on the space of functions of strict weight on $\tilde{G}$. On the other hand, $\tilde{G}$ acts on the space of corresponding functions f_r on $\mathfrak{H}$ via

$$f_r \,|\, \tilde{g}(z) = e^{-ir \arg(cz+d)}\, f_r\Bigl(\frac{az+b}{cz+d}\Bigr).$$

The latter action corresponds to (4-7) when $r \in \mathbb{Z}$. In general, this is an action of $\tilde{G}$, not of $\mathrm{SL}_2(\mathbb{R})$. The map $f \mapsto f_r$ defined above on the space of functions of strict weight is then equivariant in terms of these actions.

Many important functions on $\tilde{G}$, such as L, are not eigenfunctions of the operator $\boldsymbol{W}$, but they are annihilated by a power of $\boldsymbol{W}$. This suggests the following definition.

Definition 6.1. An $f \in C^\infty(\tilde{G})$ has *generalized weight* $r \in \mathbb{C}$ if $(\boldsymbol{W} - ir)^n f = 0$ for some $n \in \mathbb{N}$.

Thus, L and all its powers have generalized weight 0.

Next, holomorphy of $F_r = y^{-r/2} f_r$ corresponds to the property $\boldsymbol{E}^- f = 0$.

Definition 6.2. We call any differentiable function f on $\tilde{G}$ *holomorphic* (resp. *antiholomorphic*) if $\boldsymbol{E}^- f = 0$ (resp. $\boldsymbol{E}^+ f = 0$). We call any twice differentiable function f on $\tilde{G}$ *harmonic* if it satisfies $\omega f = 0$.

Note that, for functions of nonzero weight, this definition of harmonicity does not correspond to the use of the word in "harmonic weak Maass forms" in [Bruinier et al. 2008], for example.

With these definitions, we set:

Definition 6.3. Let $k, \lambda \in \mathbb{C}$. Let $\tilde{\Gamma}$ be a discrete cofinite subgroup of $\tilde{G}$.

i) The space $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)$ consists of the smooth functions $f : \mathfrak{H} \times \mathbb{R} \to \mathbb{C}$ satisfying:
   a) (*eigenfunction Casimir operator*) $\omega f = \lambda f$.
   b) (*generalized weight*) $(\boldsymbol{W} - ik)^n f = 0$ for some $n \in \mathbb{N}$.
   c) (*exponential growth*) There exists $a \in \mathbb{R}$ such that for all compact sets X and $\Theta \subset \mathbb{R}$ and for all cusps κ of $\tilde{\Gamma}$, we have
$$f(\tilde{g}_\kappa(x + iy, \vartheta)) = \mathrm{O}(e^{ay}) \tag{6-4}$$
   as $y \to \infty$ uniformly in $x \in X$ and $\vartheta \in \Theta$.

ii) We set
$$\tilde{E}_k(\tilde{\Gamma}, \lambda) := \tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma}}$$
(where $\tilde{\Gamma}$ acts by left translation). The elements of $\tilde{E}_k(\tilde{\Gamma}, \lambda)$ are called *Maass forms on $\tilde{G}$ of generalized weight k and eigenvalue λ for $\tilde{\Gamma}$*.

The space $\tilde{E}_r(\tilde{\Gamma}, \lambda)$ is infinite-dimensional. Further, since ω and $\boldsymbol{W}$ commute with left translations in $\tilde{G}$, the space $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)$ is invariant under left translation by elements of $\tilde{\Gamma}$.

When $k \in 2\mathbb{Z}$, the space $E_k(\Gamma, \lambda)$ can be identified with $\tilde{E}_k(\tilde{\Gamma}, \lambda)$. We prove the following slightly stronger statement.

Theorem 6.4. *Let $\tilde\Gamma$ be a cofinite discrete subgroup of $\tilde G$, and let $k, \lambda \in \mathbb{C}$. If $\tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)^{\tilde Z}$ contains a nonzero element f, then $k \in 2\mathbb{Z}$ and $\partial_\vartheta f(z, \vartheta) = ikf(z, \vartheta)$.*

If $k \in 2\mathbb{Z}$, then the elements $f \in \tilde E_k(\tilde\Gamma, \lambda)$ correspond bijectively to the Maass forms $F \in E_k(\Gamma, \lambda)$ by

$$f(z, \vartheta) = y^{k/2} F(z) e^{ik\vartheta}.$$

So the condition of $\tilde Z$-invariance implies that the weight k is even, and that the weight is *strict*, that is, condition b) holds with $n = 1$.

Proof of Theorem 6.4. Any smooth function $f \in C^\infty(\mathfrak{H} \times \mathbb{R})$ satisfying b) in Definition 6.3 can be written in the form $f(z, \vartheta) = \sum_{j=0}^{n-1} \varphi_j(z)\, e^{ik\vartheta}\, \vartheta^j$, with $\varphi_j \in C^\infty(\mathfrak{H})$.

If such a function is left-invariant under $\tilde Z$, then the action of $k(\pi m) \in \tilde Z \subset \tilde\Gamma$ implies, for each $m \in \mathbb{Z}$,

$$e^{\pi ikm} \sum_j \varphi_j(z) e^{ik\vartheta} (\vartheta + \pi m)^j = \sum_j \varphi_j(z) e^{ik\vartheta} \vartheta^j, \qquad \text{for all } m \in \mathbb{Z}.$$

With induction, this gives $k \in 2\mathbb{Z}$ and $\varphi_j = 0$ for $j \geq 1$, and hence $f(z, \vartheta) = \varphi_0(z) e^{ik\vartheta}$. Moreover, the stronger condition $f \in \tilde E_k(\tilde\Gamma, \lambda) = \tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)^{\tilde\Gamma}$ can be checked to be equivalent to $F_k \in E_k(\Gamma, \lambda)$ for $F_k(z) = y^{-k/2} f(z, 0)$. □

We have the following generalization of Theorem 4.2.

Theorem 6.5. *Let $\tilde\Gamma$ be a cofinite discrete subgroup of $\tilde G$ with cusps. Then the $\tilde\Gamma$-module $\tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)$ is maximally perturbable for each $k \in 2\mathbb{Z}$ and each $\lambda \in \mathbb{C}$.*

In Section 8 we will prove this theorem. In this section we will show that it implies the corresponding result for $E_k(\Gamma, \lambda)$. We first give some facts that are of more general interest.

The map identifying $E_k(\Gamma, \lambda)$ and $\tilde E_k(\tilde\Gamma, \lambda)$ can be extended to an isomorphism

$$\mu : \mathscr{E}_k(\Gamma, \lambda) \to \tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)^{\tilde Z}.$$

Since the center $\tilde Z$ of $\tilde\Gamma$ acts trivially on $\tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)^{\tilde Z}$, it can be considered as a Γ-module. With this interpretation, we obtain an identification of the Γ-modules $\mathscr{E}_k(\Gamma, \lambda)$ and $\tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)^{\tilde Z}$. Specifically, for $F \in \mathscr{E}_k(\Gamma, \lambda)$, $g \in \tilde{\mathscr{E}}_k(\tilde\Gamma, \lambda)^{\tilde Z}$ we have

$$\begin{aligned} (\mu f)(z, \vartheta) &= y^{k/2} F(z)\, e^{ik\vartheta}, \\ (\mu^{-1} g)(z) &= y^{-k/2} g(z, 0), \\ \mu(F|_k\gamma) &= \mu(F) | \nu(\gamma) \qquad (\gamma \in \Gamma), \\ \mu^{-1}(g | \tilde Z\delta) &= \mu^{-1}(g)|_k \nu^{-1}(\tilde Z\delta) \quad (\delta \in \tilde\Gamma), \end{aligned} \tag{6-5}$$

where ν denotes the isomorphism identifying Γ with $\tilde Z\backslash\tilde\Gamma$.

Proposition 6.6. *Let Γ be a cofinite discrete subgroup of G with cusps, and let $\tilde\Gamma = \mathrm{pr}^{-1}\,\Gamma$. If the $\tilde\Gamma$-module V is maximally perturbable, then the subspace $V^{\tilde Z}$, considered as a Γ-module, is maximally perturbable.*

Proof. The projection $\mathrm{pr} : \tilde\Gamma \to \Gamma$ induces linear maps $\mathrm{pr} : \mathbb{C}[\tilde\Gamma] \to \mathbb{C}[\Gamma]$ between the group rings, $\mathrm{pr} : I_{\tilde\Gamma} \to I_\Gamma$ between the augmentation ideals, and $\mathrm{pr} : I_{\tilde\Gamma}^{q+1}\backslash I_{\tilde\Gamma}^q \to I_\Gamma^{q+1}\backslash I_\Gamma^q$ for all $q \in \mathbb{N}$. Since $\mathrm{pr}(A_i) = \alpha_i$, on the basis elements $\boldsymbol{b}_{\tilde\Gamma}(\boldsymbol{i})$ in Proposition 5.1 and $\boldsymbol{b}_\Gamma(\boldsymbol{i})$ in (3-8), we have for $\tilde\Gamma$-q-tuples:

$$\mathrm{pr}\ \boldsymbol{b}_{\tilde\Gamma}(\boldsymbol{i}) = \begin{cases} \boldsymbol{b}_\Gamma(\boldsymbol{i}) & \text{if } \boldsymbol{i}(l) < t(\Gamma) \text{ for } l = 1, \dots, q, \\ 0 & \text{if } \boldsymbol{i}(q) = t(\Gamma). \end{cases} \tag{6-6}$$

This means that we have the commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & V^{\tilde\Gamma,q} \arrow[r] & V^{\tilde\Gamma,q+1} \arrow[r, "\mathsf{m}_q"] & \hom\bigl(I_{\tilde\Gamma}^{q+1}\backslash I_{\tilde\Gamma}^q, V^{\tilde\Gamma}\bigr) \arrow[r] & 0 \\ 0 \arrow[r] & (V^{\tilde Z})^{\Gamma,q} \arrow[r] & (V^{\tilde Z})^{\Gamma,q+1} \arrow[r, "\mathsf{m}_q"] & \hom\bigl(I_\Gamma^{q+1}\backslash I_\Gamma^q, (V^{\tilde Z})^\Gamma\bigr), \arrow[u] \end{tikzcd}$$

where the vertical arrow sends $f : I_\Gamma^{q+1}\backslash I_\Gamma^q \to (V^{\tilde Z})^\Gamma = V^{\tilde\Gamma}$ to $\tilde f : I_{\tilde\Gamma}^{q+1}\backslash I_{\tilde\Gamma}^q \to V^{\tilde\Gamma}$ such that $\tilde f(\boldsymbol{b}_{\tilde\Gamma}(\boldsymbol{i})) = f(\boldsymbol{b}_\Gamma(\boldsymbol{i}))$ if $\boldsymbol{i} \in \{1, \dots, t(\Gamma)-1\}^q$, and $\tilde f(\boldsymbol{b}_{\tilde\Gamma}(\boldsymbol{i})) = 0$ otherwise.

We want to write a given $f : I_\Gamma^{q+1}\backslash I_\Gamma^q \to (V^{\tilde Z})^\Gamma$ as $\mathsf{m}_q v_0$ with $v_0 \in (V^{\tilde Z})^{\Gamma,q+1}$. By assumption, there is an element $v \in V^{\tilde\Gamma,q+1}$ such that $\mathsf{m}_q v = \tilde f$. If $v\,|\,(\zeta-1) = 0$, then $v \in V^{\tilde\Gamma,q+1} \cap V^{\tilde Z} = (V^{\tilde Z})^{\Gamma,q+1}$, and we are done.

Suppose that $w = v\,|\,(\zeta-1) \neq 0$. Take $r \in [1,q]$ minimal such that $w \in V^{\tilde\Gamma,r}$. We will show that we can replace v by another element $v_1 \in v + V^{\tilde\Gamma,q}$ with $v_1\,|\,(\zeta-1) \in V^{\tilde\Gamma,r_1}$ and $r_1 < r$. Repeating this process brings us eventually to $v_j\,|\,(\zeta-1) = 0$. For this v_j, we will have $\mathsf{m}_q v_j = \tilde f$ and $v_j\,|\,(\zeta -) = 0$, which, according to the remark of the last paragraph, suffices to prove the proposition.

From

$$\begin{aligned} w\,|\,(\gamma_1-1)\dots(\gamma_{q-1}-1) &= v\,|\,(\gamma_1-1)\dots(\gamma_{q-1}-1)(\zeta-1) \\ &= \tilde f(\gamma_1, \dots, \gamma_{q-1}, \zeta) = 0, \end{aligned}$$

we conclude that $r \leq q-1$. Define $\tilde g \in \hom(I_{\tilde\Gamma}^{r+1}\backslash I_{\tilde\Gamma}^r, V^{\tilde\Gamma})$ by

$$\tilde g(\boldsymbol{b}_{\tilde\Gamma}(\boldsymbol{j})) = w\,|\,(\alpha_{\boldsymbol{j}(1)}-1)\dots(\alpha_{\boldsymbol{j}(r-1)}-1)$$

if the $\tilde\Gamma$-r-tuple $\boldsymbol{j}$ satisfies $\boldsymbol{j}(r) = t(\Gamma)$, and $\tilde g(\boldsymbol{b}_{\tilde\Gamma}(\boldsymbol{j})) = 0$ otherwise. There is $u \in V^{\tilde\Gamma,r+1} \subset V^{\tilde\Gamma,q}$ with $\mathsf{m}_r u = \tilde g$. We take $v_1 = v - u \in v + V^{\tilde\Gamma,q}$. We check that for all $\tilde{\Gamma}$-$(r-1)$-tuples $\boldsymbol{j}$,

$$\begin{aligned} v_1 \,|\, (\zeta - 1)(\alpha_{\boldsymbol{j}(1)} - 1) \dots (\alpha_{\boldsymbol{j}(r-1)} - 1) \\ &= w \,|\, (\alpha_{\boldsymbol{j}(1)} - 1) \dots (\alpha_{\boldsymbol{j}(r-1)} - 1) - u \,|\, (\alpha_{\boldsymbol{j}(1)} - 1) \dots (\alpha_{\boldsymbol{j}(r-1)} - 1)(\zeta - 1) \\ &= 0. \end{aligned}$$

This shows that $v_1 \,|\, (\zeta - 1)$ has order less than r. □

Proof of Theorem 4.2. From Theorem 6.5, $V = \tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda_k)$ is maximally perturbable. Therefore, by Proposition 6.6, the space $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda_k)^{\tilde{Z}} \cong \mathscr{E}_k(\Gamma, \lambda_k)$ is maximally perturbable too. □

This proof illustrates the fact that, for groups with cusps, there are really more higher-order forms with generalized weight than with strict weight: the basis in Proposition 5.1 is for all such discrete groups larger than the corresponding basis in Section 3B.

6C. ***Holomorphic forms on the universal covering group.***

Definition 6.7. For $k \in 2\mathbb{Z}$, we define $\mathscr{H}_k(\tilde{\Gamma})$ as the space of elements of $C^\infty(\mathfrak{H} \times \mathbb{R})$ that satisfy:

i) *(Holomorphy)*: $\boldsymbol{E}^- f = 0$.

ii) *(Generalized weight)*: $(\boldsymbol{W} - ik)^n f = 0$, for some $n \in \mathbb{N}$.

iii) *(Exponential growth)*: as described in condition c) in Definition 6.3.

This is a $\tilde{\Gamma}$-module for the action by left translation. We denote by $\mathscr{H}_k^p(\tilde{\Gamma})$ (resp. $\mathscr{H}_k^c(\tilde{\Gamma})$) the space of $f \in \mathscr{H}_k(\tilde{\Gamma})$ satisfying $f(\tilde{g}_\kappa(x+iy, \vartheta)) = \mathrm{O}(y^C)$ for some $C \in \mathbb{R}$ (resp. $f(\tilde{g}_\kappa(x+iy, \vartheta)) = \mathrm{O}(e^{ay})$ for some $a < 0$) instead of (6-4).

We will prove:

Theorem 6.8. *Let $\tilde{\Gamma}$ be a cofinite discrete subgroup of $\tilde{G}$ with cusps. Then the $\tilde{\Gamma}$-module $\mathscr{H}_k(\tilde{\Gamma})$ is maximally perturbable for each $k \in 2\mathbb{Z}$.*

Proof of Theorem 4.3. As in the case of general Maass forms, we can show that, for $k \in 2\mathbb{Z}$, $\mathscr{E}_k^{\mathrm{hol}}(\Gamma, \lambda_k) \cong \mathscr{H}_k(\tilde{\Gamma})^{\tilde{Z}}$. Then Proposition 6.6 implies Theorem 4.3. □

Second-order forms and derivatives of L-functions. With this definition, L is a second-order invariant belonging to $\mathscr{H}_0(\tilde{\Gamma}_{\mathrm{mod}})^{\tilde{\Gamma}_{\mathrm{mod}},2}$. (Incidentally, this example shows that, for generalized weight k, the space $\mathscr{H}_k(\tilde{\Gamma})$ need not be contained in $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda_k)$.)

Based on L, we can construct a second-order form which is related to derivatives of classical modular forms. Specifically, for positive integer N, denote by G_N the group generated by $\tilde{g}$, $g \in \langle \Gamma_0(N), W_N \rangle$, where

$$W_N := \begin{pmatrix} 0 & -\sqrt{N}^{-1} \\ \sqrt{N} & 0 \end{pmatrix}.$$

Set
$$L_1(z,\vartheta)=L(z,\vartheta)+L(Nz,\vartheta).$$

Using the transformation law for L and the identity $\left(\begin{smallmatrix}N&0\\0&1\end{smallmatrix}\right)\left(\begin{smallmatrix}a&b\\Nc&d\end{smallmatrix}\right)=\left(\begin{smallmatrix}a&Nb\\c&d\end{smallmatrix}\right)\left(\begin{smallmatrix}N&0\\0&1\end{smallmatrix}\right)$, a routine calculation implies that, for some $\beta\in\mathrm{Hom}(G_N,\mathbb{C})$,
$$L_1(\gamma(z,\vartheta))=L_1(z,\vartheta)+i\beta(\gamma),\qquad\text{for all }\gamma\in G_N.$$

Now let f be in the space $S_2(G_N)$ of cusp forms of weight 2 for G_N. An example of such an f is a weight-2 newform for $\Gamma_0(N)$ of which the L-function $L_f(s)$ vanishes at 1, because then $f(W_Nw)d(W_Nw)=f(w)dw$. For all $\vartheta\in\mathbb{R}$,
$$\begin{aligned}\int_0^\infty f(iy)L_1(iy,\vartheta)\,diy&=-\int_{W_N0}^{W_N\infty}f(iy)L_1(iy,\vartheta)\,diy\\&=-\int_0^\infty f(W_Niy)L_1(W_Niy,\vartheta)\,d(W_Niy)\\&=-\int_0^\infty f(iy)L_1(W_Niy,\vartheta)\,diy.\end{aligned}\tag{6-7}$$
Since $L_1(z,\vartheta+x)=L_1(z,\vartheta)+2ix$ and $L_f(1)=2\pi\int_0^\infty f(iy)\,dy=0$, our integral is independent of ϑ. It further equals
$$\begin{aligned}-\int_0^\infty f(iy)L_1(\tilde{W}_N(iy,0))\,diy&=-\int_0^\infty f(iy)\big(L_1(iy,0)+i\beta(\tilde{W}_N)\big)\,diy\\&=-\int_0^\infty f(iy)L_1(iy,0)\,diy.\end{aligned}\tag{6-8}$$

Therefore, $\int_0^\infty f(iy)L_1(iy,0)\,dy=-\int_0^\infty f(iy)L_1(iy,0)\,dy$, that is,
$$\int_0^\infty f(iy)L_1(iy,0)\,dy=0,$$
and hence
$$\int_0^\infty f(iy)\log y\,dy+2\int_0^\infty f(iy)u(iy)\,dy=0,$$
where $u(z):=\log(\eta(z))+\log(\eta(Nz))$. From this we see that, since $L_f'(s)=2\pi\int_0^\infty f(iy)\log(y)dy$, we can retrieve, from an alternative perspective, the formula
$$L_f'(1)=-4\pi\int_0^\infty f(iy)u(iy)\,dy,$$
first derived in [Goldfeld 1995].

Thus, Goldfeld's expression of $L_f'(1)$ is equivalent to the orthogonality of $L_1\in\mathcal{H}_0^p(G_N)^{G_N,2}$ to the space $S_2(G_N)\hookrightarrow\mathcal{H}_2^c(G_N)^{G_N}$ in terms of the pairing
$$\langle\cdot,\cdot\rangle:\mathcal{H}_2^c(G_N)^{G_N}\times\mathcal{H}_0^p(G_N)^{G_N,2}\to\mathbb{C}$$

defined by

$$\langle g, h\rangle = \int_0^\infty g(iy,0)h(iy,0)\frac{dy}{y}.$$

6D. ***Examples of higher-order forms for the full modular group.*** Theorems 6.5 and 6.8 show that there are perturbations of 1 for the full original $\tilde{\Gamma}_{\mathrm{mod}}$ of $\mathrm{SL}_2(\mathbb{Z})$ in the universal covering group. Since $t(\Gamma_{\mathrm{mod}}) = 1$, all these perturbations are commutative (see (5-10)).

1) The function L can lead to second-order *harmonic perturbations of* 1. Specifically, although $L \notin \tilde{\mathscr{E}}_0(0)^{\tilde{\Gamma},2}$ (because $\omega L = \frac{1}{2}$), the imaginary part $\mathrm{Im}\, L : (z, \vartheta) \mapsto 2\,\mathrm{Im}\log\eta(z) + \vartheta$ is harmonic, has second order, and corresponds to the linear form $\alpha \in \mathrm{Mult}^1(\tilde{\Gamma}_{\mathrm{mod}}, \mathbb{C})$. It has generalized weight 0, and it is not holomorphic.

2) Set $\chi_r = e^{ir\alpha}$, $r \in \mathbb{C}$, where $\alpha \in \hom(\tilde{\Gamma}_{\mathrm{mod}}, \mathbb{C})$ is given by $\alpha(n(1)) = \pi/6$ and $\alpha(k(\pi/2)) = \pi/2$. The family

$$r \mapsto e^{rL(z,\vartheta)} = y^{r/2}\eta(z)^{2r}e^{ir\vartheta} \tag{6-9}$$

consists of elements of $\mathscr{H}_r(\tilde{\Gamma})$ that are $\tilde{\Gamma}_{\mathrm{mod}}$-invariant under the action given by

$$(f\,|\,\gamma)(z) = f(\gamma z)\overline{\chi_r(\gamma)}.$$

By Proposition 4.4, for $k \geq 1$ the derivative

$$\partial_r^k e^{rL(z,\vartheta)}\big|_{r=0} = L(z,\vartheta)^k$$

is a holomorphic perturbation of 1 of order $k+1$. The corresponding element of $\mathrm{Mult}^k(\tilde{\Gamma}_{\mathrm{mod}}, \mathbb{C})$ is $i^k k!\alpha^{\otimes k}$.

3) It is possible to obtain a more or less explicit description of a harmonic perturbation of 1 of order 3. We sketch how this can be done with the meromorphic continuation of the Eisenstein series in weight and spectral parameter jointly. This family is studied in [Bruggeman 1986]. In that work, automorphic forms are described as functions on $\mathfrak{H}$ transforming according to a multiplier system of Γ_{mod}. These correspond to functions on $\tilde{G}$ that transform according to a character of $\tilde{\Gamma}_{\mathrm{mod}}$. Carrying out the reformulation, we can rephrase §2.18 in [Bruggeman 1986] as stating that there is a meromorphic family of Maass forms on $U \times \mathbb{C}$, where U is some neighborhood of $(-12, 12)$ in $\mathbb{C}$. We retrieve the exact family studied in [Bruggeman 1986] by considering $z \mapsto E(r, s; z, 0)$. For each $(r, s) \in U \times \mathbb{C}$ at which E is not singular, it is an automorphic form of weight r for the character $\chi_r = e^{ir\alpha}$ of $\tilde{\Gamma}_{\mathrm{mod}}$ with eigenvalue $\lambda_s = \frac{1}{4} - s^2$. It is a meromorphic family of automorphic forms on $\tilde{\Gamma}_{\mathrm{mod}}$ with character χ_r with a Fourier expansion of the form

$$E(r,s) = \mu_r\Big(\frac{r}{12}, s\Big) + C_0(r,s)\mu_r\Big(\frac{r}{12}, -s\Big) + \sum_{n\neq 0} C_n(r,s)\omega_r\Big(n + \frac{r}{12}, s\Big), \tag{6-10}$$

where the $C_n(r,s)$ are meromorphic functions, and where we use the following notations:

$$\begin{aligned} \omega_r(\nu,s;z,\vartheta) &= e^{2\pi i\nu x} W_{r\,\mathrm{Sign}(\mathrm{Re}\,\nu)/2,s}(4\pi\nu\,\mathrm{Sign}(\mathrm{Re}\,\nu)y)e^{ir\vartheta},\\ \mu_r(\nu,s;z,\vartheta) &= e^{2\pi i\nu z} y^{1/2+s}\,{}_1F_1\Big(\frac{1}{2}+s-\frac{r}{2};1+2s;4\pi\nu y\Big)e^{ir\vartheta}. \end{aligned} \tag{6-11}$$

This family and its Fourier coefficient C_0 satisfy the following functional equations:

$$\begin{aligned} E(r,-s) &= C_0(r,-s)E(r,s),\\ E(r,s;-x+iy,-\vartheta) &= E(-r,s;x+iy,\vartheta). \end{aligned} \tag{6-12}$$

Further, the restriction of this family to the (complex) line $r=0$ exists, and gives a meromorphic family of automorphic forms depending on one parameter s. This is a family of weight 0, so it does not depend on the parameter ϑ on $\tilde{G}$. The resulting family on $\mathfrak{H}$ is the meromorphic continuation of the Eisenstein series for Γ_{mod} in weight 0, with Fourier expansion

$$\begin{aligned} E(0,s) = \mu_0(0,s) &+ \frac{\sqrt{\pi}\,\Gamma(s)\zeta(2s)}{\Gamma\big(s+\frac{1}{2}\big)\zeta(2s+1)}\mu_0(0,-s)\\ &+ \frac{\pi^{s+1/2}}{\Gamma\big(s+\frac{1}{2}\big)\zeta(2s+1)}\sum_{n\neq 0}\frac{\sigma_{2s}(|n|)}{|n|^{s+1/2}}\omega_0(n,s), \end{aligned} \tag{6-13}$$

where

$$\begin{aligned} \mu_0(0,s;z,\vartheta) &= y^{1/2+s},\\ \omega_0(n,s;z,\vartheta) &= e^{2\pi inx} W_{0,s}(4\pi|n|y) = e^{2\pi inx}\,2|n|^{1/2}K_s(2\pi|n|y). \end{aligned}$$

At $\big(0,-\frac{1}{2}\big)$, the family E is holomorphic in both variables r and s, with a constant as its value at $\big(0,-\frac{1}{2}\big)$; this is a consequence of Proposition 6.5 ii) in [Bruggeman 1986]. So in principle, we obtain higher-order harmonic perturbations of 1 by differentiating $r\mapsto E\big(r,-\frac{1}{2}\big)$. Here we encounter the problem that we have an explicit Fourier expansion (6-13) only for $E(0,s)$, and thus we cannot describe the derivatives in the direction of r directly. To overcome this problem, we use the fact that for r near 0, we have

$$\begin{aligned} E\Big(r,-\frac{1-r}{2};z,\vartheta\Big) &= H_r(z,\vartheta) = e^{rL(z,\vartheta)},\\ E\Big(r,-\frac{1+r}{2};z,\vartheta\Big) &= H_{-r}(-\bar{z},-\vartheta) = e^{-r\overline{L(z,\vartheta)}}. \end{aligned} \tag{6-14}$$

The proof of the first equality is contained in 6.10 in [Bruggeman 1986]. The second one follows from the second functional equation in (6-12). Now we use the

Taylor expansion of E of degree 2 at $(r,s)=\bigl(0,-\frac{1}{2}\bigr)$:

$$
\begin{aligned}
&E(r,s)\\
&=1+rA_{1,0}+\bigl(s+\tfrac{1}{2}\bigr)A_{0,1}+\tfrac{1}{2}r^2A_{2,0}+r\bigl(s+\tfrac{1}{2}\bigr)A_{1,1}+\tfrac{1}{2}\bigl(s+\tfrac{1}{2}\bigr)^2A_{0,2}+\dots.
\end{aligned}
\tag{6-15}
$$

By Proposition 4.4, the coefficients $A_{1,0}$ and $A_{2,0}$ are harmonic perturbations of 1 of order 2 and 3, respectively. From (6-14), we obtain the following results:

$$
\begin{aligned}
A_{1,0}&=i\ \mathrm{Im}\,L, &\qquad A_{0,1}&=2\,\mathrm{Re}\,L,\\
A_{2,0}+\tfrac{1}{4}A_{0,2}&=\mathrm{Re}\,L^2, &\qquad A_{1,1}&=i\ \mathrm{Im}\,L^2.
\end{aligned}
\tag{6-16}
$$

This confirms that $\mathrm{Im}\,L$ is a second-order harmonic perturbation of 1. Differentiation in the direction of s preserves $\tilde{\Gamma}_{\mathrm{mod}}$-invariance. So $A_{0,1}=2\,\mathrm{Re}\,L$ and $A_{0,2}$ are $\tilde{\Gamma}_{\mathrm{mod}}$-invariant. However, these functions are not in the kernel of ω.

Thanks to the identity $A_{2,0}+\frac{1}{4}A_{0,2}=\mathrm{Re}\,L^2$, to determine the third-order harmonic perturbation $A_{2,0}$ it suffices to explicitly compute $A_{0,2}$, because $\mathrm{Re}\,L^2$ is known in a fairly explicit way. The function $A_{0,2}$ can be obtained as the coefficient of $\frac{1}{2}\bigl(s+\frac{1}{2}\bigr)^2$ in the Taylor expansion of $E(0,s)$ at $s=-\frac{1}{2}$. As a by-product of this computation, we will also obtain the $\tilde{\Gamma}_{\mathrm{mod}}$-invariant function $A_{0,1}$ as the coefficient of $s+\frac{1}{2}$ in the same expansion.

We examine each term in the Fourier expansion (6-13) separately. We use the functional equation of the Riemann zeta function and its expansion at the point 0. We also use an integral representation of the Whittaker function $W_{0,s}$. This leads to

$$
\begin{aligned}
A_{0,1}(z,0)&=\log y-\frac{\pi}{3}y-2\sum_{n\ge1}\sum_{d\mid n}\frac{1}{d}(q^n+\bar{q}^n)\\
&=2\,\mathrm{Re}\biggl(\tfrac{1}{2}\log y+\frac{\pi i}{6}z-\sum_{n=1}^{\infty}\sigma_{-1}(n)\,q^n\biggr)=2\,\mathrm{Re}\,L(z,0),\\
A_{0,2}(z,0)&=(\log y)^2+\Bigl(8b_1-\frac{4\pi a_0}{3}+\frac{2\pi}{3}\log y\Bigr)y\\
&\quad+\sum_{n=1}^{\infty}\biggl(-4a_0\,\sigma_{-1}(n)(q^n+\bar{q}^n)+2\sigma_{-1}(n)\,(q^{-n}+\bar{q}^{-n})\,\Gamma(0,4\pi ny)\\
&\qquad\qquad-2(q^n+\bar{q}^n)\sum_{d\mid n}\frac{\log(d^2/n)}{d}\biggr),
\end{aligned}
\tag{6-17}
$$

with the notation $q=e^{2\pi iz}$.

A remarkable aspect of this computation is that we have used an explicit computation of the derivatives of the Eisenstein series in weight 0 to compute the second derivative in the r-direction of the more complicated Eisenstein family in two variables. The basic observation is (6-14), which shows that the Eisenstein family has easy derivatives in two directions. The Taylor expansion of E at $\bigl(0,-\frac{1}{2}\bigr)$ has three monomials in order 2. So it suffices to compute a second-order derivative in one more direction to get hold of all terms. Higher-order terms in the Taylor expansion have too many monomials for this method to work. We do not know how to compute all harmonic perturbations of 1 of higher order.

7. Higher-order Fourier expansions

This section is needed for the constructions on which the proofs of Theorems 6.5 and 6.8 are based, but it is also of independent interest. It provides a higher-order analogue of the classical Fourier expansions.

7A. ***Fourier expansion of Maass forms.*** If f is in $\tilde{E}_r(\tilde{\Gamma},\lambda)$, then for each cusp κ of Γ there is a Fourier expansion

$$f(\tilde{g}_\kappa g)=\sum_\nu F_{\kappa,\nu}\, f(g),\qquad F_{\kappa,\nu} f(g)=\int_0^1 e^{-2\pi i\nu x} f(\tilde{g}_\kappa n(x)g)\,dx, \tag{7-1}$$

where ν runs through a class in $\mathbb{C} \bmod \mathbb{Z}$ determined by χ and the cusp κ. The function $F_\nu f$ satisfies $F_{\kappa,\nu}\, f(z,\vartheta)=e^{2\pi i\nu x}F_{\kappa,\nu}\, f(iy,0)e^{ir\vartheta}$ and $\omega F_{\kappa,\nu}\, f=\lambda F_{\kappa,\nu}\, f$.

For each given ν, r, and s, set

$$\mathcal{W}_r(\nu,s):=\bigl\{f:\tilde{G}\to\mathbb{C};\ \omega f=\bigl(\tfrac14-s^2\bigr)f \text{ and } f(z,\theta)=e^{2\pi i\nu x+ir\theta}f(iy,0)\bigr\}. \tag{7-2}$$

Because of the second relation in the definition, $f\in\mathcal{W}_r(\nu,s)$ can be thought of as a function of y. Therefore, the space $\mathcal{W}_r(\nu,s)$ is isomorphic to the space of $f:\mathbb{R}\to\mathbb{C}$ satisfying

$$-y^2h''(y)+\bigl(4\pi^2\nu^2y^2-2\pi\nu r\,y-\tfrac14+s^2\bigr)\,h(y)=0. \tag{7-3}$$

It is convenient to write $\lambda=\lambda_s=\frac14-s^2$ with $s\in\mathbb{C}$. We can choose a fixed s with $\operatorname{Re} s\ge 0$ corresponding to the eigenvalue $\lambda=\lambda_s$ under consideration. The spaces $\mathcal{W}_r(\nu,s)$ are two-dimensional. We will use the basis elements in §4.2 of [Bruggeman 1994].

- For $\operatorname{Re}\nu\neq 0$, a basis of $\mathcal{W}_r(\nu,s)$ is formed by

$$\begin{aligned}\omega_r(\nu,s;z,\vartheta)&=e^{2\pi i\nu x}W_{r\,\mathrm{Sign}(\operatorname{Re}\nu)/2,s}(4\pi\nu\,\mathrm{Sign}(\operatorname{Re}\nu)y)e^{ir\vartheta},\\ \hat{\omega}_r(\nu,s;z,\vartheta)&=e^{2\pi i\nu x}W_{-r\,\mathrm{Sign}(\operatorname{Re}\nu)/2,s}(-4\pi\nu\,\mathrm{Sign}(\operatorname{Re}\nu)y)e^{ir\vartheta}.\end{aligned} \tag{7-4}$$

Here $W_{\mu,s}(t)$ is the Whittaker function that decreases exponentially as $t\to\infty$. We use the branch of $W_{\kappa,s}(z)$ that is holomorphic for $-\pi/2<\arg z<3\pi/2$. The asymptotic behavior as $y\to\infty$, by §4.2.1 in [Slater 1960], is:

$$\omega_r(\nu,s;z,\vartheta)\sim(4\pi\nu\varepsilon y)^{r\varepsilon/2}e^{2\pi\nu(ix-\varepsilon y)+ir\vartheta}, \tag{7-5}$$

$$\hat{\omega}_r(\nu,s;z,\vartheta)\sim e^{-\pi ir\varepsilon/2}(4\pi\varepsilon\nu y)^{-r\varepsilon/2}e^{2\pi\nu(ix+\varepsilon y)+ir\vartheta}, \tag{7-6}$$

where ε denotes Sign(Re ν). The subspace of $\mathcal{W}_r(\nu, s)$ generated by $\omega_r(\nu, s)$ is denoted by $\mathcal{W}_r^0(\nu, s)$.

• For $\nu = 0$, a basis is given by $\{y^{1/2+s}e^{ir\vartheta}, y^{1/2-s}e^{ir\vartheta}\}$ if $s \neq 0$ and $\{y^{1/2}e^{ir\vartheta}, y^{1/2}\log ye^{ir\vartheta}\}$ if $s = 0$.

The next proposition allows for Fourier expansions of functions with exponential growth. (See [Bruggeman 1994], §4.1–3 for a proof.)

Proposition 7.1. *Let $k \in 2\mathbb{Z}$, $\operatorname{Re} s \geq 0$. Suppose that the function $f \in C^\infty(\tilde{\Gamma}\backslash\tilde{G})$ satisfies $\omega f = \lambda_s f$ and $\mathbf{W} f = ikf$. Then it has at each cusp κ an absolutely converging Fourier expansion*

$$f(\tilde{g}_\kappa g) = \sum_{n\in\mathbb{Z}} F_{\kappa,n}\, f(g), \tag{7-7}$$

with $F_{\kappa,n}\, f \in \mathcal{W}_k(n, s)$. Moreover, $f \in \tilde{E}_k(\tilde{\Gamma}, \lambda_s)$ if and only if there exists $N > 0$ such that all Fourier terms $F_{\kappa,n}\, f$ with $|n| \geq N$ are in $\mathcal{W}_k^0(n, s)$ for all cusps κ.

7B. *Higher-order Fourier terms.* The higher-order invariants of $\mathcal{V}_k(n, s)$ that we will define now are the higher-order analogues of the classical Fourier terms.

Definition 7.2. Let $k \in 2\mathbb{Z}$, $n \in \mathbb{Z}$, and $s \in \mathbb{C}$. By $\mathcal{V}_k(n, s)$ we denote the space of functions f on $\tilde{G}$ that satisfy $\omega f = \lambda_s f$, have generalized weight k, and satisfy $(\partial_x - 2\pi i n)^m f = 0$ for some $m \in \mathbb{N}$ (which may depend on f).

For $n \neq 0$, we denote by $\mathcal{V}_k^0(n, s)$ the subspace of $f \in \mathcal{V}_k(n, s)$ satisfying $f(z, \vartheta) = \mathrm{O}(y^a\, e^{-2\pi|n|y})$ as $y \to \infty$ for some $a \in \mathbb{R}$.

The free Abelian group $\tilde{\Delta}$ generated by $\tau = n(1)$ and $\zeta = k(\pi)$ acts on these spaces by left translation.

Proposition 7.3. *Let k, n, s be as above. The $\tilde{\Delta}$-modules $\mathcal{V}_k(n, s)$ and $\mathcal{V}_k^0(n, s)$ are maximally perturbable.*

For each $q \in \mathbb{N}$, the elements $f \in \mathcal{V}_k(n, s)^{\tilde{\Delta},q}$ satisfy for each $\delta > 0$

$$f(z, \vartheta) \ll_\delta e^{(2\pi|n|+\delta)y}, \qquad (y \to \infty) \tag{7-8}$$

uniformly for x and ϑ in compact sets. If $n \neq 0$, then for each $q \in \mathbb{N}$, the elements $f \in \mathcal{V}_k^0(n, s)^{\tilde{\Delta},q}$ satisfy for each $\delta > 0$

$$f(z, \vartheta) \ll_\delta e^{(\delta-2\pi|n|)y}, \qquad (y \to \infty) \tag{7-9}$$

uniformly for x and ϑ in compact sets.

Proof. To prove that $\mathcal{V}_k(n,s)$ is maximally perturbable, we start with a characterization of the space $\mathcal{V}_k(n,s)^{\tilde\Delta}$. We first note that $\mathcal{W}_k(n,s) \subset \mathcal{V}_k(n,s)^{\tilde\Delta}$. Conversely, if $f \in \mathcal{V}_k(n,s)^{\tilde\Delta}$, then the reasoning in the proof of Theorem 6.4 shows that the weight of f is strict, and also that $\partial_x f = 2\pi i n f$, and hence $f(z,\vartheta) = e^{2\pi i n x} f(iy,\vartheta)$. So $f \in \mathcal{W}_k(n,s)$. If, for $n \neq 0$, the function f is also exponentially decreasing, it has to be a multiple of $\omega_k(n,s)$. Therefore, $\mathcal{V}_k^0(n,s)^{\tilde\Delta} = \mathcal{W}_k^0(n,s)$.

Let f be an arbitrary element of $\mathcal{W}_k(n,s)$. Since each of the basis elements of $\mathcal{W}_k(n,s)$ is a specialization of a holomorphic family of elements of $\mathcal{W}_r(\nu,s)$, there is a holomorphic family of $h(r,\nu) \in \mathcal{W}_r(\nu,s)$ such that $h(k,n) = f$. We have $h\big(r,\nu; n(\xi)k(\ell\pi)(z,\vartheta)\big) = e^{2\pi i\nu\xi + \pi i r\ell} h(r,\nu; z,\vartheta)$ for $\xi \in \mathbb{R}$ and $\ell \in \mathbb{Z}$.

Next, consider the polynomials $Q_q \in \mathbb{Q}[X]$ of degree q, defined by

$$\begin{aligned} Q_0 &= 1, \\ Q_{q+1}(X+1) - Q_{q+1}(X) &= Q_q(X), \\ Q_q(0) &= 0, \qquad \text{for } q \geq 1. \end{aligned} \tag{7-10}$$

Then for each $\boldsymbol{m} = (m_1, m_2)$, $m_j \geq 0$, set

$$h_k^{\boldsymbol{m}}(n,s) = Q_{m_1}\Big(\frac{1}{\pi i}\partial_r\Big) Q_{m_2}\Big(\frac{1}{2\pi i}\partial_\nu\Big) h(r,\nu)\big|_{\nu=n,\, r=k}. \tag{7-11}$$

Applying the differential operator $(1/2\pi i)\partial_\nu^a$ on $h(r,\nu)\,|\,(\tau-1) = (e^{2\pi i\nu}-1)h(r,\nu)$ we obtain

$$\begin{aligned} Q_{m_2}\Big(\frac{1}{2\pi i}\partial_\nu\Big)h(r,\nu)\,\big|\,(\tau-1) &= \Big(Q_{m_2}\Big(\frac{1}{2\pi i}\partial_\nu + 1\Big) - Q_{m_2}\Big(\frac{1}{2\pi i}\partial_\nu\Big)\Big)h(r,\nu) \\ &= Q_{m_2-1}\Big(\frac{1}{2\pi i}\partial_\nu\Big)h(r,\nu). \end{aligned} \tag{7-12}$$

Since τ, ζ commute, this implies $h_k^{\boldsymbol{m}}(n,s)\,|\,(\tau-1) = h_k^{(m_1,m_2-1)}(n,s)$. Likewise, we obtain the transformation law $h_k^{\boldsymbol{m}}(n,s)\,|\,(\zeta-1) = h_k^{(m_1-1,m_2)}(n,s)$. Therefore, for $l_1 + l_2 = m_1 + m_2$ $(l_1, l_2 \geq 0)$,

$$h_k^{(m_1,m_2)}(n,s)\,|\,(\zeta-1)^{l_1}(\tau-1)^{l_2} = \delta_{m_1,l_1}\delta_{m_2,l_2} f, \tag{7-13}$$

thus obtaining the maximal perturbability of $\mathcal{V}_k(n,s)$. For convenience, we shall call perturbations satisfying the transformation law (7-13) *perturbations of type* $\boldsymbol{m}$.

Based on $\mathcal{V}_k^0(n,s)^{\tilde\Delta} = \mathcal{W}_k^0(n,s)$, we deduce in an analogous way the maximal perturbability of $\mathcal{V}_k^0(n,s)$.

To prove (7-8) and (7-9), we first note that the maximal perturbability we have just shown implies that the functions $h^{\boldsymbol{m}}$ constructed from elements f ranging over a basis of $\mathcal{W}_k(n,s)$ (resp. $\mathcal{W}_k^0(n,s)$) induce a basis of the quotients $\mathcal{V}^{\tilde\Delta,q+1}/\mathcal{V}^{\tilde\Delta,q}$. Therefore, it suffices to show (7-8) and (7-9) for $h^{\boldsymbol{m}}$ only. In the case $n \neq 0$, the family h may be taken to be $\omega_r(\nu,s)$ or $\hat\omega_r(\nu,s)$ in (7-4). For these functions, the question reduces to the asymptotic behavior of $\partial_t^j \partial_\kappa^l W_{\kappa,s}(t)$, since the factors $e^{2\pi i \nu x}$ and $e^{ir\vartheta}$ produce polynomials in x and ϑ, which yield constants when they vary through compact sets. The differentiation of $4\pi\ \mathrm{Sign}(\mathrm{Re}\,\nu)\nu y$ yields only a power of y, which can be absorbed by the factor $e^{\delta y}$.

Differentiation of $W_{\kappa,s}(t)$ with respect to t does not change the exponential part of the asymptotic behavior, since derivatives of $W_{\kappa,s}(t)$ are linear combinations of $W_{\kappa,s}(t)$ and $W_{\kappa+1,s}(t)$ with powers of t in the factors [Slater 1960, (2.4.24)]. So we have to look only at differentiation with respect to κ. Then the bounds are a consequence of the integral representation

$$W_{\kappa,s}(t) = \frac{-1}{2\pi i}\Gamma\bigl(\kappa+\tfrac{1}{2}-s\bigr)e^{-t/2}t^{\kappa}\int_{(0+)}^{\infty} e^{-x}(-x)^{s-\kappa-1/2}\Bigl(1+\frac{x}{t}\Bigr)^{s+\kappa-1/2}dx, \quad (7\text{-}14)$$

for $t \in \mathbb{R}$ with $t > 0$, $\kappa - \frac{1}{2} - s \neq -1, -2, \dots$ (see (3.5.18) in [Slater 1960]). Here the contour comes from ∞ along a line slightly above the positive real axis, encircles 0 with radius $\delta < 1$, and then goes back to ∞ on a line slightly below the positive real axis. If $\kappa - \frac{1}{2} - s = -1, -2, \dots$, we use the representation

$$W_{\kappa,s}(t) = \frac{e^{-(1/2)t}\,t^{\kappa}e^{i\varphi(s-\kappa+1/2)}}{\Gamma\bigl(s+\frac{1}{2}-\kappa\bigr)}\int_0^{\infty} e^{-e^{i\varphi}u}u^{s-\kappa-1/2}\bigl(1+e^{i\varphi}u/t\bigr)^{s+\kappa-1/2}du, \quad (7\text{-}15)$$

for some $0 < \varphi < \pi/2$.

All these estimates, taken together, prove (7-8) and (7-9) (when $n \neq 0$). They further show that the derivatives of a family with exponential decay have exponential decay, and thus $\mathcal{V}_k^0(n,s)$ is also maximally perturbable.

If $n = 0$, we argue directly that we can find functions $h_k^{\boldsymbol{m}}(0,s)$ in $\mathcal{V}_k(0,s)$ of the form $p_{\boldsymbol{m}}(x,y,\vartheta)y^{1/2\pm 2}e^{ik\vartheta}$, where $p_{\boldsymbol{m}}$ is a polynomial in three variables with degree m_1 in ϑ and degree m_2 in x. If the coefficient of $\vartheta^{m_1}x^{m_2}$ in this polynomial does not depend on y, this leads to a perturbation of $y^{1/2\pm s}e^{ik\vartheta}$ of type $\boldsymbol{m}$. Such functions satisfy the required estimates, with a polynomial factor y^A instead of $e^{\delta y}$. The remaining task is to check that they can be chosen to satisfy

$$\bigl(\omega - \tfrac{1}{4} + s^2\bigr)h_k^{\boldsymbol{m}}(0,s) = 0.$$

We do this by induction in the degrees in ϑ and x. We check that

$$\begin{aligned}\bigl(\omega - \tfrac{1}{4} + s^2\bigr)&x^{m_2}y^{1/2\pm s+a}\vartheta^{m_1}e^{ik\vartheta}\\ &= -a(a\pm 2s)x^{m_2}y^{1/2\pm s+a}\vartheta^{m_1}e^{ik\vartheta} + \text{ terms of lower degree in } x \text{ or } \vartheta.\end{aligned}$$

With $a = 0$, this gives the top coefficient of $p_{\boldsymbol{m}}$. Moreover, the terms of lower degree all are multiples of $x^{\tilde{m}_2}y^{1/2\pm s+a}\vartheta^{\tilde{m}_1}e^{ik\vartheta}$ with $\tilde{m}_j \le m_j$, $\tilde{m}_1 < m_1$, or $\tilde{m}_2 < m_2$, and $a \in \mathbb{Z}_{\ge 0}$. Successively we can determine the lower-degree terms, and arrange for $h_k^{\boldsymbol{m}}(0,s)$ to be an eigenfunction of ω with eigenvalue $\frac{1}{4} - s^2$.

This takes care of the case $n = 0$, except if $s = 0$. It that case we also have to perform a computation involving $y^{1/2+a} \log y$, which we leave to the reader. $\square$

Holomorphic Fourier terms on $\tilde{G}$ are multiples of

$$\eta_r(\nu; z, \vartheta) = y^{r/2} e^{2\pi i \nu z} e^{ir\vartheta}. \tag{7-16}$$

Thus we have the spectral parameter $s = \pm(r-1)/2$. For real values of ν and r, we have

$$\eta_r(\nu) = \begin{cases} (4\pi\nu)^{-r/2} \omega_r\Big(\nu, \pm\frac{r-1}{2}\Big) & \text{if } \nu > 0, \\ \mu_r\Big(0, \frac{r-1}{2}\Big) & \text{if } \nu = 0, \\ e^{-\pi i r}(4\pi|\nu|)^{-r/2} \hat{\omega}_r\Big(\nu, \pm\frac{r-1}{2}\Big) & \text{if } \nu < 0, \end{cases} \tag{7-17}$$

with notations as in (7-4) and (6-11). The functions

$$\eta_k^{\boldsymbol{m}}(n; z, \vartheta) = Q_{m_1}\Big(\frac{2i\vartheta + \log y}{2\pi i}\Big) Q_{m_2}(z) \eta_k(n; z, \vartheta) \tag{7-18}$$

satisfy

$$\mathsf{m}_{m_1+m_2} \eta_k^{\boldsymbol{m}} : (\zeta - 1)^{l_1} (\tau - 1)^{l_2} \mapsto \delta_{m_1, l_1} \delta_{m_2, l_2} \eta_k(n) \tag{7-19}$$

for $l_1 + l_2 = m_1 + m_2$, and as $y \to \infty$, their growth is of order $\mathrm{O}(e^{(\delta - 2\pi n)y})$. For the commutative group $\tilde{\Delta}$ and for a fixed $\boldsymbol{m}$, they yield a basis of the space of forms of order $m_1 + m_2 + 1$ modulo lower-order forms.

As an example, we note that the Fourier expansion (6-1) can be written in the following way:

$$L(z, \vartheta) = \pi i \eta_0^{(1,0)}(0; z, \vartheta) + \frac{\pi i}{6} \eta_0^{(0,1)}(0; z, \vartheta) - 2 \sum_{n \geq 1} \sigma_{-1}(n) \eta_0^{(0,0)}(n; z, \vartheta). \tag{7-20}$$

8. Proofs of Theorems 6.5 and 6.8

The method of the proof is highly inductive. At each step, we use the maximal perturbability of other spaces, which has been proved in a previous step. The starting point for this process is the space $\mathrm{Map}(\tilde{\Gamma}, \mathbb{C})$, whose maximal perturbability is proved based on general algebraic principles in Proposition 8.1. This implies directly the maximal perturbability of the $\tilde{\Gamma}$-module $\mathrm{Map}(\mathfrak{H} \times \mathbb{R}, \mathbb{C})$. We proceed by imposing increasingly stringent regularity conditions on the functions $\mathfrak{H} \times \mathbb{R} \to \mathbb{C}$. We consider $C^\infty(\mathfrak{H} \times \mathbb{R}) = C^\infty(\tilde{G})$, the subspace $C_k^\infty(\tilde{G})$ of functions in $C^\infty(\tilde{G})$ with generalized weight k, and the subspace $\mathscr{C}_k$ of $C_k^\infty(\tilde{G})$ of functions that have compact support modulo $\tilde{\Gamma}$. In Section 7, we considered higher-order invariant functions for the group $\tilde{\Delta}$ generated by $n(1)$ and $k(\pi)$. These functions are related to the Fourier expansions of Maass forms. After proving that some more auxiliary subspaces of $C_k^\infty(\mathfrak{H}\times\mathbb{R})$ are maximally perturbable, we finally prove in Section 8E the maximal perturbability of $\tilde{\mathscr{E}}_k(\tilde\Gamma,\lambda)$ and $\mathscr{H}_k(\tilde\Gamma)$.

8A. ***Higher-order invariants in maps on*** $\tilde{\boldsymbol{\Gamma}}$**.**

Proposition 8.1. *If $\tilde\Gamma$ is a discrete cofinite subgroup of $\tilde G$ with cusps, then the $\tilde\Gamma$-module* $\mathrm{Map}(\tilde\Gamma,\mathbb{C})$ (*with the action by left translation*) *is maximally perturbable.*

Proof. We first define $\boldsymbol{g_i}$ on the free subgroup $\tilde\Gamma_0$ of $\tilde\Gamma$ generated by $\alpha_1,\dots,\alpha_{t(\Gamma)-1}$ for $\boldsymbol{i}\in\{1,\dots,t(\Gamma)-1\}^q$ by the relations

$$\begin{aligned}\boldsymbol{g}_{()}&=1,\\ \boldsymbol{g}_{(j,\boldsymbol{i})}\,|\,(\alpha_j-1)&=\boldsymbol{g_i},\\ \boldsymbol{g_i}\,|\,(\alpha_j-1)&=0,\qquad \text{if } \boldsymbol{i}(1)\neq j,\\ \boldsymbol{g_i}(1)&=0,\qquad \text{if } |\boldsymbol{i}|\geq 1.\end{aligned}\tag{8-1}$$

By $|\boldsymbol{i}|$ we denote the length of the tuple $\boldsymbol{i}$.

Let $\varphi_0:\tilde\Gamma\to\tilde\Gamma_0$ be the surjective group homomorphism given by $\varphi_0(\alpha_j)=\alpha_j$ for $1\le j\le t(\Gamma)-1$, $\varphi_0(\zeta)=1$, and $\varphi_0(\varepsilon_j)=1$ for $1\le j\le n_{\mathrm{ell}}$. For $1\le j\le t(\Gamma)$, we define $\psi_j\in\hom(\tilde\Gamma,\mathbb{C})$ such that $\psi_j(\alpha_{j'})=\delta_{j,j'}$. This determines ψ_j completely, because values on elliptic generators are given by $\psi_j(\varepsilon_j)=(1/v_j)\psi_j(\zeta)$. For $\boldsymbol{i}=(\boldsymbol{i}',t(\Gamma),\dots,t(\Gamma))$, where there are m coordinates $t(\Gamma)$ at the end and where $\boldsymbol{i}'\in\{1,\dots,t(\Gamma)-1\}^{q-m}$, we put

$$\boldsymbol{f_i}(\gamma)=\boldsymbol{g_{i'}}(\varphi_0(\gamma))Q_m(\psi_{t(\Gamma)}(\gamma)),\tag{8-2}$$

where Q_n are the polynomials defined in (7-10). Now we can check the following properties of $\boldsymbol{f_i}$:

$$\boldsymbol{f}_{()}=1\qquad\text{(empty tuple, } q=0);\tag{8-3}$$

$$\boldsymbol{f_i}(1)=0\qquad\text{if } |\boldsymbol{i}|\geq 1;\tag{8-4}$$

$$\boldsymbol{f_i}\,|\,(\zeta-1)=\begin{cases}\boldsymbol{f_{i'}} & \text{if } \boldsymbol{i}=(\boldsymbol{i}',t(\Gamma)),\\ 0 & \text{if } \boldsymbol{i} \text{ does not end with a } t(\Gamma);\end{cases}\tag{8-5}$$

$$\boldsymbol{f_i}\,|\,(\alpha_j-1)=\begin{cases}\boldsymbol{f_{i'}} & \text{if } \boldsymbol{i}=(j,\boldsymbol{i}') \text{ with } j<t(\Gamma),\\ 0 & \text{if } j<t(\Gamma),\ j\neq\boldsymbol{i}(1).\end{cases}\tag{8-6}$$

Using this, we can see that

$$(\mathrm{m}_q\boldsymbol{f_i})(\boldsymbol{b}(\boldsymbol{j}))=\delta_{\boldsymbol{i},\boldsymbol{j}}.\tag{8-7}$$

Now, the choice of the basis $\boldsymbol{b}(\boldsymbol{i})$ in (5-7) for $\tilde\Gamma$-q-tuples $\boldsymbol{i}$ shows that to prove that $\mathrm{Map}(\tilde\Gamma,\mathbb{C})$ is maximally perturbable, it suffices to prove that for each $\boldsymbol{i}$ and for each function f on $\tilde\Gamma\backslash\tilde G$, a function $h_{\boldsymbol{i}}\in\mathrm{Map}(\tilde G,\mathbb{C})$ such that, for all $\tilde\Gamma$-q-tuples $\boldsymbol{j}$,

$$h_{\boldsymbol{i}}\,|\,(\alpha_{\boldsymbol{j}(1)}-1)\dots(\alpha_{\boldsymbol{j}(q)}-1)=\delta_{\boldsymbol{i},\boldsymbol{j}}\cdot f.\tag{8-8}$$

To construct such functions, we choose a *strict fundamental domain* $\mathfrak{F}_{\tilde\Gamma} \subset \tilde G$ for $\tilde\Gamma\backslash\tilde G$, that is, a set meeting each $\tilde\Gamma$-orbit exactly once. A choice for the sought function $h_{\boldsymbol{i}}$ is then

$$h_{\boldsymbol{i}}(\gamma g) = \boldsymbol{f_i}(\gamma) f(g), \qquad \gamma \in \Gamma,\ g \in \mathfrak{F}_{\tilde\Gamma} \tag{8-9}$$

or

$$h_{\boldsymbol{i}}(g) = \sum_{\gamma\in\tilde\Gamma} \boldsymbol{f_i}(\gamma) f(g)\psi(\gamma^{-1}g), \tag{8-10}$$

where ψ is the characteristic function of $\mathfrak{F}_{\tilde\Gamma}$. □

8B. *Higher-order invariants in smooth functions on $\tilde G$.* We will use essentially the same construction as in the last section to prove:

Proposition 8.2. *The $\tilde\Gamma$-module $C^\infty(\tilde G)$ is maximally perturbable.*

Proof. In order to show that $C^\infty(\tilde G)$ is a maximally perturbable $\tilde\Gamma$-module, we need to have (8-8) with $h_{\boldsymbol{i}} \in C^\infty(\tilde G)$ for each $f \in C^\infty(\tilde\Gamma\backslash\tilde G)$. We consider functions $\psi \in C^\infty(\mathfrak{H}\times\mathbb{R})$ such that $\sum_{\gamma\in\tilde\Gamma}\psi(\gamma^{-1}(z,\vartheta)) = 1$ for all $(z,\vartheta)\in\mathfrak{H}\times\mathbb{R}$ as a locally finite sum. If we define (8-10) with such a function ψ and $f \in C^\infty(\tilde\Gamma\backslash\tilde G)$, then the sum is locally finite, and the $h_{\boldsymbol{i}}$ are smooth. □

8C. *Higher-order invariants and generalized weight.* Set

$$C_k^\infty(\tilde G) = \{f \in C^\infty(\tilde G),\ \text{of generalized weight } k\}. \tag{8-11}$$

Proposition 8.3. *Let $k \in 2\mathbb{Z}$. Then the $\tilde\Gamma$-module $C_k^\infty(\tilde G)$ is maximally perturbable.*

Proof. As with the previous proofs, our approach is to show that for every $\tilde\Gamma$-q-tuple $\boldsymbol{i} = (\boldsymbol{i}', t(\Gamma), \dots, t(\Gamma))$ with exactly m occurrences of $t(\Gamma)$ at the end and for every $f \in C_k^\infty(\tilde\Gamma\backslash\tilde G)$, there exists $h_{\boldsymbol{i}} \in C_k^\infty(\tilde G)$ satisfying Equation (8-8) for all $\tilde\Gamma$-q-tuples $\boldsymbol{j}$. We note that, by Theorem 6.4, the $\tilde\Gamma$-invariance of f implies that its weight k is strict, that is, $f(gk(\vartheta)) = f(g)e^{ik\vartheta}$.

We will define the function $h_{\boldsymbol{i}}$ by an analogue of (8-10). We first define for each $g \in \tilde G$ the point $w(g) = \mathrm{pr}(g)i \in \mathfrak{H}$ and the real number $\Theta(g)\in\mathbb{R}$ such that $g = (w(g), \Theta(g)) \in \tilde G = \mathfrak{H}\times\mathbb{R}$. We also recall that $\Gamma = \tilde\Gamma/\tilde Z$. Since the group homomorphism ϕ_0 defined in the proof of Proposition 8.1 is trivial on $\tilde Z = \langle\zeta\rangle$, it induces a homomorphism on Γ. Now we take $\psi(z,\vartheta)$ to be a bounded locally finite partition of unity ψ_0 on $\mathfrak{H}$. (Compare Lemma 1 in §3 of [Kra 1969].) So the function $(z,\vartheta)\mapsto\psi(\gamma^{-1}(z,\vartheta))$ obtained by left translation depends only on the image of $\gamma\in\tilde\Gamma$ in $\Gamma\cong\tilde\Gamma/\tilde Z$. Let, as in the proof of Proposition 8.1, $\psi_{t(\Gamma)}$ be the function $\tilde\Gamma\to\mathbb{R}$ such that $\psi_{t(\Gamma)}(\alpha_{j'}) = \delta_{t(\Gamma),j'}$. For a given $\gamma\in\tilde\Gamma$, we have

$$\psi_{t(\Gamma)}(\zeta\gamma) = \psi_{t(\Gamma)}(\gamma) + 1$$

and $\Theta((\zeta\gamma)^{-1}g) = \Theta(\gamma^{-1}g) - \pi$. So $\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}g)/\pi$ is well-defined on $\Gamma = \tilde{\Gamma}/\tilde{Z}$. We can therefore set

$$h_{\boldsymbol{i}}(g) = \sum_{\gamma\in\Gamma} \boldsymbol{g}_{\boldsymbol{i}'}(\varphi_0(\gamma))\, Q_m\big(\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}g)/\pi\big) f(g)\psi(\gamma^{-1}g). \tag{8-12}$$

The support property of the partition of unity ψ ensures convergence; it is even a locally finite sum with a bounded number of nonzero terms. All factors depend smoothly on g. So $h_{\boldsymbol{i}} \in C^\infty(\tilde{G})$.

We consider $(\boldsymbol{W} - ik)h_{\boldsymbol{i}}$. Since $\boldsymbol{W}\psi = 0$, we need only consider

$$\begin{aligned}(\partial_\vartheta - ik)\, &Q_m\big(\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}gk(\vartheta))/\pi\big) f(gk(\vartheta)) \\ &= Q_m\big(\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}gk(\vartheta))/\pi\big)(\partial_\vartheta - ik) f(gk(\vartheta)) \\ &\qquad + f(gk(\vartheta))\partial_\vartheta Q_m\big(\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}g)/\pi + \vartheta/\pi\big) \\ &= 0 + \pi^{-1} Q_m'\big(\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}g)/\pi + \vartheta/\pi\big) f(gk(\vartheta)).\end{aligned} \tag{8-13}$$

Repeating this, we obtain

$$\begin{aligned}(\boldsymbol{W} - ik)^{m+1} Q_m\big(\psi_{t(\Gamma)}(\gamma) + \Theta(\gamma^{-1}g)/\pi\big) f(g) \\ = \pi^{-m-1} Q_m^{(m+1)}(\dots)\dots \\ = 0,\end{aligned} \tag{8-14}$$

since the degree of Q_m is m. So $h_{\boldsymbol{i}} \in C_k^\infty(\tilde{G})$. □

Remark. As A. Deitmar has pointed out, the last two propositions should also follow from [Deitmar 2008]. We have opted for explicit methods of proof because they are necessary for later parts of the paper.

8D. ***Higher-order invariants with support conditions.*** We discuss the motivation for the introduction of the invariants we will be dealing with. If Definition 6.3 of the space $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)$ did not include a growth condition at the cusps, we could consider $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)$ as the kernel $\mathscr{K}$ in the exact sequence

$$0 \longrightarrow \mathscr{K} \longrightarrow C_k^\infty(\tilde{G}) \xrightarrow{\omega-\lambda} C_k^\infty(\tilde{G}).$$

With exponential growth, one might want to try to replace $C_k^\infty(\tilde{G})$ by its subspace $C_l^\infty(\tilde{\Gamma})^{\mathrm{eg}}$ of functions with exponential growth at the cusps of $\tilde{\Gamma}$. This would lead to an exact sequence

$$0 \longrightarrow \tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda) \longrightarrow C_k^\infty(\tilde{\Gamma})^{\mathrm{eg}} \xrightarrow{\omega-\lambda} C_k^\infty(\tilde{\Gamma})^{\mathrm{eg}},$$

for which we might try to show that for each $q \in \mathbb{N}$,

$$0 \longrightarrow \tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma},q} \longrightarrow (C_k^\infty(\tilde{\Gamma})^{\mathrm{eg}})^{\tilde{\Gamma},q} \xrightarrow{\omega-\lambda} (C_k^\infty(\tilde{\Gamma})^{\mathrm{eg}})^{\tilde{\Gamma},q}$$

is exact. For this to be of use, it seems that we need surjectivity of the map

$$\omega-\lambda:(C_k^\infty(\tilde\Gamma)^{\mathrm{eg}})^{\tilde\Gamma}\to(C_k^\infty(\tilde\Gamma)^{\mathrm{eg}})^{\tilde\Gamma},$$

which we did not succeed in proving, and which may not hold. For this reason, we will instead work with other, better behaved subspaces of the spaces appearing in the exact sequence. We will therefore define subspaces $\mathscr{C}_k, \mathscr{D}_k(\lambda)\subset C_k^\infty(\tilde G)$ and $\mathscr{E}_k'(\lambda)\subset\mathscr{E}_k(\tilde\Gamma,\lambda)$, related by an exact sequence

$$0\longrightarrow\tilde{\mathscr{E}}_k'(\lambda)\longrightarrow\mathscr{D}_k(\lambda)\xrightarrow{\omega-\lambda}\mathscr{C}_k. \tag{8-15}$$

8D1. *The spaces $\mathscr{C}_k$.* For each cusp $\kappa=\tilde g_\kappa\infty$ and each $a>0$, we call

$$D_\kappa(a)=\tilde g_\kappa\{(z,\vartheta):\operatorname{Im}z\ge a,\ \vartheta\in\mathbb{R}\} \tag{8-16}$$

a *horocyclic set*. There is a number A_Γ such that for each $a\ge A_\Gamma$, the $D_\kappa(a)$ are disjoint for different cusps. The sets

$$\tilde G_a=\{(z,\vartheta)\in\mathfrak{H}\times\mathbb{R}:(z,\vartheta)\notin D_\kappa(a)\text{ for all }\kappa\} \tag{8-17}$$

satisfy $\tilde\Gamma\tilde G_a=\tilde G_a$. This follows from the fact that the g_κ have been chosen so that

$$\gamma\tilde\Gamma_\kappa\tilde g_\kappa=\tilde g_{\gamma\kappa}\tilde\Gamma_\infty, \tag{8-18}$$

for all cusps κ and for $\gamma\in\tilde\Gamma$. Here $\tilde\Gamma_\kappa:=\operatorname{pr}^{-1}\Gamma_\kappa=\{\gamma\in\tilde\Gamma:\gamma\kappa=\kappa\}$.

Definition 8.4. Let $k\in 2\mathbb{Z}$. The space $\mathscr{C}_k$ consists of the $f\in C_k^\infty(\tilde G)$ supported in $\tilde G_a$ for some $a\ge A_\Gamma$. (The a may depend on f).

So $\mathscr{C}_k$ consists of the smooth functions with generalized weight k whose supports project to compact subsets of $\Gamma\backslash\mathfrak{H}$. Clearly, the space $\mathscr{C}_k$ is $\tilde\Gamma$-invariant. If we apply the construction of h_i in the proof of Proposition 8.3 to functions $f\in\mathscr{C}_k^{\tilde\Gamma}\subset C_k^\infty(\tilde\Gamma\backslash\tilde G)$, then the support of each h_i is contained in the same set $\tilde G_a$ that contains $\operatorname{Supp}(f)$. This implies:

Proposition 8.5. *Let $k\in 2\mathbb{Z}$. Then the $\tilde\Gamma$-module $\mathscr{C}_k$ is maximally perturbable.*

8D2. *The spaces $\mathscr{D}_k(\lambda)$.* We will define $\mathscr{D}_k(\lambda)$ essentially as the space of functions that accept higher-order analogues of Fourier expansions at the cusps. To make this formal, we study spaces of functions defined for $y_0>0$ on regions of the form

$$S(y_0)=\{(x+iy,\vartheta)\in\mathfrak{H}\times\mathbb{R}:y>y_0\}. \tag{8-19}$$

Definition 8.6. Let $k\in 2\mathbb{Z}$, $\lambda\in\mathbb{C}$, and $y_0>0$. We denote by $\mathscr{E}_k(y_0,\lambda)$ the space of those $f\in C^\infty(S(y_0))$ that satisfy $\omega f=\lambda f$ and $(\boldsymbol{W}-ik)^n f=0$ for some $n\in\mathbb{N}$, and have at most exponential growth as $y\to\infty$, uniform for x and ϑ in compact sets. We denote by $\mathscr{E}_k^{\mathrm{hol}}(y_0)$ the space of holomorphic functions on $S(y_0)$ with generalized weight k and at most exponential growth as $y\to\infty$.

Proposition 8.7. *Let $k \in 2\mathbb{Z}$, $s \in \mathbb{C}$, and $y_0 > 0$. The spaces $\mathscr{E}_k(y_0, \lambda_s)$ and $\mathscr{E}_k^{\mathrm{hol}}(y_0)$ are maximally perturbable $\tilde{\Delta}$-modules.*

Let $q \in \mathbb{N}$. Each $f \in \mathscr{E}_k(y_0, \lambda_s)^{\tilde{\Delta}, q}$ has an absolutely convergent expansion

$$f(z, \vartheta) = \sum_{n \in \mathbb{Z}} f_n(z, \vartheta) \tag{8-20}$$

on $S(y_0)$ with $f_n \in \mathscr{V}_k(n, s)^{\tilde{\Delta}, q}$ for all n, and $f_n \in \mathscr{V}_k^0(n, s)^{\tilde{\Delta}, q}$ for almost all n.

Each $f \in \mathscr{E}_k^{\mathrm{hol}}(y_0)^{\tilde{\Delta}, q}$ has an absolutely convergent expansion on $S(y_0)$ of the form

$$f(z, \vartheta) = \sum_{\boldsymbol{m},\, m_1 + m_2 < q} \; \sum_n c_{\boldsymbol{m}}^n \eta_k^{\boldsymbol{m}}(n; z, \vartheta), \tag{8-21}$$

where the inner sum ranges from some, possible negative, integer to infinity.

Proof. We start with the holomorphic case. Let $f \in \mathscr{E}_k^{\mathrm{hol}}(y_0)^{\tilde{\Delta}}$. Then the function $z \mapsto y^{-k/2} f(z, 0)$ is holomorphic on $\{z \in \mathfrak{H} : y > y_0\}$ with period 1. So it has an expansion of the form $\sum_n a_n e^{2\pi i n z}$, finite to the left and converging absolutely on $y > y_0$. For each $y_1 > y_0$, we have $a_n = \mathrm{O}(e^{2\pi n y_1})$ as $n \to \infty$.

Hence, $f(z, \vartheta) = \sum_n a_n \eta_k(n; z, \vartheta)$ converges absolutely on $y > y_0$, and

$$f^{\boldsymbol{m}}(z, \vartheta) := \sum_{n \geq -N} a_n \eta_k^{\boldsymbol{m}}(n; z, \vartheta)$$

converges absolutely on $S(y_0)$, and the convergence is uniform on any set $y \geq y_1$ with $y_1 > y_0$, with x and ϑ in compact sets. These functions satisfy $f^{\boldsymbol{m}} | (\tau - 1) = f^{(m_1, m_2 - 1)}$, $f^{\boldsymbol{m}} | (\zeta - 1) = f^{(m_1 - 1, m_2)}$, and $f^{(0,0)} = f$, since all $\eta_k^{\boldsymbol{m}}$ have this property. Thus $f^{\boldsymbol{m}}$, with $\boldsymbol{m}$ such that $m_1 + m_2 < q$, is a perturbation of type $\boldsymbol{m}$, and we deduce that $\mathscr{E}_k^{\mathrm{hol}}(y_0)$ is maximally perturbable. An arbitrary element $h \in \mathscr{E}_k^{\mathrm{hol}}(y_0)^{\tilde{\Delta}, q}$ can be written as a finite linear combination of such $f^{\boldsymbol{m}}$, which all have expansions of the type given in (8-21).

For $f \in \mathscr{E}_k(y_0, \lambda_s)^{\tilde{\Delta}}$ we proceed similarly. By Proposition 7.1 and the integrality of k, there is an absolutely convergent Fourier expansion

$$f(z, \vartheta) = \sum_{n \in \mathbb{Z}} f_n(z, \vartheta)$$

on $S(y_0)$ with $f_n \in \mathscr{W}_k(n, s)$. By the exponential growth, $f_n \in \mathscr{W}_k^0(n, s)$ for $|n| > N$, for some $N \in \mathbb{N}$.

For $|n| > N$, we have $f_n = a_n \omega_k(n, s)$, and from (7-5) we conclude that $a_n = \mathrm{O}(e^{2\pi |n| y_1})$ as $|n| \to \infty$ for each $y_1 > y_0$. So by (7-5), the series

$$\sum_{n,\, |n| > N} a_n \omega_k^{\boldsymbol{m}}(n, s)$$

converges absolutely on $S(y_0)$ and uniformly on each set $y \geq y_1$ with $y_1 > y_0$, and gives an exponentially decreasing function as $y \to \infty$. It is a λ_s-eigenfunction of ω, since the decay allows differentiation inside the sum. To produce a perturbation $f^{\boldsymbol{m}}$ of f, we pick $f_n^{\boldsymbol{m}} \in \mathcal{V}_k(n,s)^{\tilde\Delta, m_1+m_2+1}$ such that $f_n^{\boldsymbol{m}} | (\tau - 1) = f_n^{(m_1,m_2-1)}$, $f_n^{\boldsymbol{m}} | (\zeta - 1) = f_n^{(m_1-1,m_2)}$, and $f_n^{(0,0)} = f_n$ for the finitely many n with $|n| \le N$. The estimate (7-8) shows that the growth of these terms is at most of the order $\mathrm{O}(e^{(2\pi N+\delta)y})$ as $y \to \infty$ for each $\delta > 0$. Thus we get (nonuniquely) a perturbation of type $\boldsymbol{m}$ in $\mathscr{E}_k(y_0, \lambda_s)$:

$$f^{\boldsymbol{m}} = \sum_{|n|\le N} f_n^{\boldsymbol{m}} + \sum_{|n|>N} a_n \omega_k^{\boldsymbol{m}}(n,s).$$

Thus we get (8-20) and the maximal perturbability of $\mathscr{E}_k(y_0, \lambda_s)$. □

We are now ready to define $\mathscr{D}_k(\lambda)$ and $\mathscr{D}_k^{\mathrm{hol}}$.

Definition 8.8. Let $k \in 2\mathbb{Z}$ and $\lambda \in \mathbb{C}$. We define $\mathscr{D}_k(\lambda)$ as the space of functions $f \in C_k^\infty(\tilde G)$ (hence with generalized weight k) for which there exist $b \ge A_\Gamma$, $a \in \mathbb{R}$, and $q \in \mathbb{N}$ such that for each cusp κ of $\tilde\Gamma$, the function $(z,\vartheta) \mapsto f(\tilde g_\kappa(z,\vartheta))$ is an element of $\mathscr{E}_k(b,\lambda)^{\tilde\Delta,q}$ and satisfies a bound $\mathrm{O}(e^{ay})$ as $y \to \infty$.

We define $\mathscr{D}_k^{\mathrm{hol}}$ similarly, with $(z,\vartheta) \mapsto f(\tilde g_\kappa(z,\vartheta))$ in $\mathscr{E}_k^{\mathrm{hol}}(b)^{\tilde\Delta,q}$, with bound $\mathrm{O}(e^{ay})$.

Remark 8.9. The numbers a, b, and q may depend on the function f.

Remark 8.10. Definition 8.6 of $\mathscr{E}_k(b,\lambda)$ implies that elements of $\mathscr{D}_k(\lambda)$ are λ-eigenfunctions of ω on the set $\bigsqcup_\kappa D_\kappa(b)$. Similarly, elements of $\mathscr{D}_k^{\mathrm{hol}}$ are holomorphic functions on $\bigsqcup_\kappa D_\kappa(b)$. In both cases, we have exponential growth at each cusp. The definition requires that the order of this exponential growth stay bounded when we vary the cusp.

The space $\mathscr{C}_k$ is contained in $\mathscr{D}_k(\lambda)$ and in $\mathscr{D}_k^{\mathrm{hol}}$. Indeed, for given $f \in \mathscr{C}_k$, we can take b large so that $\bigsqcup_\kappa D_\kappa(b)$ is outside the support of f. Elements f of $\tilde{\mathscr{E}}_k(\tilde\Gamma,\lambda)^{\tilde\Gamma}$ restricted to $D_\kappa(b)$ induce elements $(z,\vartheta) \mapsto f(\tilde g_\kappa(z,\vartheta))$ in $\mathscr{E}_k(b,\lambda)^{\tilde\Delta}$ for each cusp κ, and similarly in the holomorphic case. Hence

$$\tilde{\mathscr{E}}_k(\tilde\Gamma,\lambda)^{\tilde\Gamma} \subset \mathscr{D}_k(\lambda)^{\tilde\Gamma}, \qquad \mathscr{H}_k(\tilde\Gamma)^{\tilde\Gamma} \subset (\mathscr{D}_k^{\mathrm{hol}})^{\tilde\Gamma}. \tag{8-22}$$

Maximal perturbability of $\mathscr{D}_k(\lambda)$ and $\mathscr{D}_k^{\mathrm{hol}}$. We need a technical lemma in order to relate $\tilde\Delta$-invariants to $\tilde\Gamma$-invariants.

We first note that if ∞ is a cusp of $\tilde\Gamma$ and if $\tilde g_\infty = 1$, then $\tilde\Delta = \tilde\Gamma_\infty$. In general, the group $\tilde\Gamma_\kappa$ can be conjugated to $\tilde g_\kappa^{-1}\tilde\Gamma_\kappa \tilde g_\kappa = \tilde\Delta$ in $\tilde g_\kappa^{-1}\tilde\Gamma\tilde g_\kappa$. So we can assume here that $\tilde\Delta \subset \tilde\Gamma$.

The abelian group $\tilde\Delta$ is free on the generators $\tau = n(1)$ and $\zeta = k(\pi)$. The dimension of $\mathrm{Map}(\tilde\Delta,\mathbb{C})^{\tilde\Delta,q+1}$ is $(q+1)(q+2)/2$, with an explicit basis described as follows. Define a sequence of maps on $\tilde\Delta$ by setting

$$\begin{aligned}\varphi^{(l,m)}\,|\,(\zeta-1)&=\varphi^{(l-1,m)},\\ \varphi^{(l,m)}\,|\,(\tau-1)&=\varphi^{(l,m-1)},\end{aligned} \tag{8-23}$$

$$\varphi^{(0,0)}=1,\quad \varphi^{(l,m)}=0,\quad \text{for } l \text{ or } m \text{ negative},$$

and

$$\varphi^{(l,m)}(1)=0,\quad \text{for } l,m\ge 0,\ l+m>0.$$

Then

$$(\mathsf{m}_q\varphi^{(l,m)})\big((\zeta^r-1)(\tau^s-1)\big)=\delta_{l,r}\delta_{m,s}$$

for $l+m=r+s=q$, and therefore the $\varphi^{(l,m)}$ with $l,m\ge 0$, $l+m\le q$ is a basis of $\mathrm{Map}(\tilde\Delta,\mathbb{C})^{\tilde\Delta,q+1}$.

Let R be a system of representatives of $\tilde\Gamma/\tilde\Delta$; so $R\subset\tilde\Gamma$. Consider the system $\{\boldsymbol{f_j}\}_{|\boldsymbol{j}|=q}\subset\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{\tilde\Gamma,q+1}$ in the proof of Proposition 8.1. If $|\boldsymbol{j}|=q$, then for every $\gamma\in\tilde\Gamma$, $\delta\mapsto \boldsymbol{f_j}(\gamma\delta)$ is a function on $\tilde\Delta$ of order at most $q+1$. Hence, there are functions $a^{\boldsymbol{j}}_{l,m}$ on R such that for all $\rho\in R$ and $\delta\in\tilde\Delta$,

$$\boldsymbol{f_j}(\rho\delta)=\sum_{\substack{l,m\ge 0\\ l+m\le q}} a^{\boldsymbol{j}}_{l,m}(\rho)\varphi^{(l,m)}(\delta). \tag{8-24}$$

Lemma 8.11. *Let $a^{\boldsymbol{j}}_{l,m}$ be as in* (8-24), *and suppose that we have functions $\psi^{(l,m)}\in\mathrm{Map}(\tilde\Delta,\mathbb{C})$ satisfying*

$$\begin{aligned}\psi^{(0,0)}&=0,\\ \psi^{(l,m)}\,|\,(\tau-1)&=\psi^{(l-1,m)} && \text{for } l\ge 1,\\ \psi^{(l,m)}\,|\,(\zeta-1)&=\psi^{(l,m-1)} && \text{for } m\ge 1.\end{aligned} \tag{8-25}$$

Then

$$f(\rho\delta)=\sum_{\substack{l,m\ge 0\\ l+m\le q}} a^{\boldsymbol{j}}_{l,m}(\rho)\psi^{(l,m)}(\delta),\qquad (\rho\in R,\ \delta\in\tilde\Delta) \tag{8-26}$$

defines an element of $\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{\tilde\Gamma,q}$.

Proof. We proceed by induction in $q=|\boldsymbol{j}|$. If $q=0$, then $m=n=0$, so $f(\rho\delta)=a^{\boldsymbol{j}}_{0,0}(\rho)\cdot\psi^{(0,0)}=0\in\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{\tilde\Gamma,0}=\{0\}$.

It is clear that (8-26) gives a well-defined map on $\tilde\Gamma$. It suffices to prove that, for any generator α_j of $\tilde\Gamma$, $f\,|\,(\alpha_j-1)\in\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{q-1}$. Suppose first that $\boldsymbol{j}=(j,\boldsymbol{j}')$. For each $\rho\in R$, there are unique $\rho_1\in R$ and $\delta_1\in\tilde\Delta$ such that $\alpha_j\rho=\rho_1\delta_1$.

From (8-24), it follows that

$$\begin{aligned} \boldsymbol{f_j}\,|\,(\alpha_{\boldsymbol{j}(1)}-1)(\rho\delta) = \sum_{\substack{l,m\geq 0\\ l+m\leq q}} a^{\boldsymbol{j}}_{l,m}(\rho_1)\varphi^{(l,m)}\,|\,(\delta_1-1)(\delta) \\ + \sum_{\substack{l,m\geq 0\\ l+m\leq q}} \bigl(a^{\boldsymbol{j}}_{l,m}(\rho_1)-a^{\boldsymbol{j}}_{l,m}(\rho)\bigr)\varphi^{(l,m)}(\delta). \end{aligned} \tag{8-27}$$

By (8-6), the left-hand side equals

$$\sum_{\substack{l,m\geq 0\\ l+m\leq q-1}} a^{\boldsymbol{j}'}_{l,m}(\rho)\phi^{(l,m)}(\delta).$$

The function $\varphi^{(l,m)}\,|\,(\delta_1-1)$ is a linear combination, depending on ρ, of $\varphi^{(a,b)}$ with $0\leq a\leq l$, $0\leq b\leq m$, and $a+b\leq q-1$. Thus we get an expression for the $a^{\boldsymbol{j}'}_{l,m}(\rho)$ in terms of the $a^{\boldsymbol{j}}_{l,m}(\rho)$. The form of this expression depends on the relations (8-23), but not on the specific value of the constant basis element $\varphi^{(0,0)}$. The relations of (8-23) hold for $\psi^{(l,m)}$ too. Therefore, the right-hand side of (8-27), upon replacement of ϕ by ψ, equals

$$\sum_{\substack{l,m\geq 0\\ l+m\leq q-1}} a^{\boldsymbol{j}'}_{l,m}(\rho)\psi^{(l,m)}(\delta) \quad (\rho\in R,\ \delta\in\tilde\Delta),$$

which, by induction, is in $\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{\tilde\Gamma,q-1}$. Since it follows from (8-26) that the right-hand side of (8-27) with φ replaced by ψ equals $f\,|\,(\alpha_j-1)$ too, we deduce that $f\,|\,(\alpha_j-1)\in\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{\tilde\Gamma,q-1}$.

In the same way, we deduce that $f\,|\,(\alpha_j-1)\in\mathrm{Map}(\tilde\Gamma,\mathbb{C})^{\tilde\Gamma,q-1}$ when $j=t(\Gamma)$ or $j<t(\Gamma)$ and $j\neq\boldsymbol{j}(1)$. □

Proposition 8.12. *The $\tilde\Gamma$-modules $\mathcal{D}_k(\lambda)$ and $\mathcal{D}^{\mathrm{hol}}_k$ are maximally perturbable for all $k\in 2\mathbb{Z}$ and $\lambda\in\mathbb{C}$.*

Proof. It suffices to construct, for a given $f\in\mathcal{D}_k(\lambda)^{\tilde\Gamma}$, a given $q\in\mathbb{N}$, and a given $\tilde\Gamma$-q-tuple $\boldsymbol{i}$, an element $\eta_{\boldsymbol{i}}\in\mathcal{D}_k(\lambda)$ such that $\eta_{\boldsymbol{i}}\,|\,(\alpha_{\boldsymbol{i}'(1)}-1)\dots(\alpha_{\boldsymbol{i}'(q)}-1)=\delta_{\boldsymbol{i},\boldsymbol{i}'}f$ for all $\tilde\Gamma$-q-tuples $\boldsymbol{j}$.

We will write $f=f_{\mathrm{cpt}}+\sum_\kappa f_\kappa$, with κ running over a set C of representatives of the $\tilde\Gamma$-orbits of cusps, where $f_{\mathrm{cpt}}\in(\mathcal{C}_k)^{\tilde\Gamma}$ and $f_\kappa\in\mathcal{D}_k(\lambda)^{\tilde\Gamma}$. We will produce perturbations for each of these components.

We choose a strict fundamental domain $\mathfrak{F}_{\tilde\Gamma}$ for $\tilde\Gamma\backslash\tilde G$ such that

$$\mathfrak{F}_{\tilde\Gamma}\cap D_\infty(b)=\bigl\{(x+iy,\vartheta):0\leq x<1,\ y\geq b,\ 0\leq\vartheta<\pi\bigr\}.$$

Definition 8.8 provides $b\geq A_\Gamma$ and $r\in\mathbb{N}$ such that $v_\kappa(z,\vartheta)=f(\tilde g_\kappa(z,\vartheta))$ is in $\mathcal{E}_k(b,\lambda)^{\tilde\Delta,r}$ for each cusp κ. Furthermore, b can be chosen large enough for the sets $\mathfrak{F}_{\tilde\Gamma}\cap D_\kappa(b)$ $(\kappa\in C)$ to be pairwise disjoint. Since f is $\tilde\Gamma$-invariant, we even have $v_\kappa \in \mathscr{E}_k(b,\lambda)^{\tilde\Delta}$. We choose a function $\chi \in C^\infty(0,\infty)$ that is equal to 0 on $\left(0, b+\frac{1}{2}\right]$ and equal to 1 on $[b+1,\infty)$, and define for $\kappa \in C$

$$f_\kappa(z,\vartheta) = \begin{cases} 0 & \text{if } (z,\vartheta) \in \mathfrak{F}_{\tilde\Gamma} - D_\kappa(b), \\ \chi(\operatorname{Im}(z_1)) v_\kappa(z_1,\vartheta_1) & \text{if } (z,\vartheta) = \tilde g_\kappa(z_1,\vartheta_1) \in \mathfrak{F}_{\tilde\Gamma} \cap D_\kappa(b). \end{cases} \tag{8-28}$$

Extend to $\tilde G$ by $\tilde\Gamma$-linearity. So f_κ is equal to 0 outside $\tilde\Gamma D_\kappa(b)$, and equal to f on $\tilde\Gamma D_\kappa(b+1)$. We check in Definition 8.8 that $f_\kappa \in \mathscr{D}_\kappa(\lambda)$. The function

$$f_{\text{cpt}} = f - \sum_{\kappa \in C} f_\kappa$$

is $\tilde\Gamma$-invariant and vanishes on $D_\kappa(b+1)$ for all cusps κ; hence $f_{\text{cpt}} \in \mathscr{C}_k^{\tilde\Gamma}$.

Proposition 8.5 implies that there is $h_{\boldsymbol{i}} \in \mathscr{C}_k \subset \mathscr{D}_k(\lambda)$ satisfying the conditions $h_{\boldsymbol{i}} \,|\, (\alpha_{\boldsymbol{i}(1)} - 1) \dots (\alpha_{\boldsymbol{i}(q)} - 1) = f_{\text{cpt}}$ and $h_{\boldsymbol{i}'} \,|\, (\alpha_{\boldsymbol{i}'(1)} - 1) \dots (\alpha_{\boldsymbol{i}'(q)} - 1) = 0$ for $\tilde\Gamma$-q-tuples $\boldsymbol{i}' \neq \boldsymbol{i}$. So we can restrict our attention to the f_κ.

Since the supports of the f_κ with $\kappa \in C$ are disjoint, we can consider each of the f_κ separately. Without loss of generality, we can assume that ∞ is a cusp of $\tilde\Gamma$ with $\tilde g_\kappa = 1$, and take $\infty \in C$. Conjugation by the original $\tilde g_\kappa$ then gives the same result for a general $\kappa \in C$.

The function v_∞ used in (8-28) is an element of $\mathscr{E}_k(b,\lambda)^{\tilde\Delta}$. The proof of Proposition 8.7 shows that for each $\boldsymbol{m} \in \mathbb{N}_0^2$, there is a perturbation

$$v_\infty^{\boldsymbol{m}} \in \mathscr{E}_k(b,\lambda)^{\tilde\Delta, m_1+m_2+1}$$

of $(z,\vartheta) \mapsto f_\infty(z,\vartheta)$ of type $\boldsymbol{m}$. We define $\eta_{\boldsymbol{i}}$ by $\eta_{\boldsymbol{i}} = 0$ on $\tilde G_b$ and on all $\tilde\Gamma D_\kappa(b)$ for all $\kappa \in C \smallsetminus \{\infty\}$, and

$$\eta_{\boldsymbol{i}}(\rho(x+iy,\vartheta)) = \sum_{\substack{l,m \ge 0 \\ l+m \le q}} \chi(y) a_{l,m}^{\boldsymbol{i}}(\rho) v_\infty^{(l,m)}(x+iy,\vartheta) \tag{8-29}$$

for $y \ge b$ and ρ in a system of representatives R of $\tilde\Gamma/\tilde\Delta$. The functions $a_{l,m}^{\boldsymbol{i}}$ are as in (8-24). Since the sets $\rho D_\infty(b)$ are disjoint, this defines a smooth function, which can be checked to be an element of $\mathscr{D}_k(\lambda)$.

For each fixed $g = (x+iy,\vartheta)$ with $y \ge b$, the function $\delta \mapsto v_\infty^{(l,m)}(\delta g)$ on $\tilde\Delta$ satisfies the same relations as $\delta \mapsto \varphi^{(l,m)}(\delta)\, v_\infty(g)$ in (8-23). So their difference, as a function of δ, satisfies (8-25).

Ignoring smoothness for a moment, we have $f_\infty \in \operatorname{Map}(\tilde G, \mathbb{C})^{\tilde\Gamma}$. Equation (8-9) gives a function $h_{\boldsymbol{i}}$ on $\tilde G$ such that $h_{\boldsymbol{i}} \,|\, (\alpha_{\boldsymbol{i}'(1)} - 1) \dots (\alpha_{\boldsymbol{i}'(q)} - 1) = \delta_{\boldsymbol{i},\boldsymbol{i}'} f_\infty$ for all $\tilde\Gamma$-q-tuples $\boldsymbol{i}'$. With our choice of fundamental domain, and using (8-24), we find for $\rho \in R$, $\delta \in \tilde\Delta$, and $g = (x+iy,\vartheta)$ with $y \ge b$:

$$h_{\boldsymbol{i}}(\rho\delta g) = \sum_{\substack{l,m \ge 0 \\ l+m \le q}} a_{l,m}^{\boldsymbol{i}}(\rho) \varphi^{(l,m)}(\delta) \chi(y) v_\infty(g). \tag{8-30}$$

Outside $\tilde\Gamma D_\infty(b)$, the functions f_∞, h_i are zero. With Lemma 8.11, we conclude that the function induced by

$$(\eta_i - h_i)(\rho\delta g) = \sum_{\substack{l,m\geq 0\\ l+m\leq q}} a^{i}_{l,m}(\rho)\chi(y)\bigl(v_\infty^{(l,m)}(\delta g) - \varphi^{(l,m)}(\delta)v_\infty(g)\bigr) \tag{8-31}$$

is in $\mathrm{Map}(\tilde G,\mathbb{C})^{\tilde\Gamma,q}$. This implies that

$$\eta_i \in \bigl(h_i + \mathrm{Map}(\tilde G,\mathbb{C})^{\tilde\Gamma,q}\bigr)\cap \mathscr{D}_k(\lambda) = \mathscr{D}_k(\lambda)^{\tilde\Gamma,q+1},$$

and behaves in the desired way under $(\alpha_{i'(1)}-1)\dots(\alpha_{i'(q)}-1)$ for all $\tilde\Gamma$-q-tuples $\boldsymbol{i}'$. Thus, we have proved that $\mathscr{D}_k(\lambda)$ is maximally perturbable.

Everywhere in this proof, we can replace $\mathscr{E}_k(b,\lambda)$ by $\mathscr{E}_k^{\mathrm{hol}}(b)$, and $\mathscr{D}_k(\lambda)$ by $\mathscr{D}_k^{\mathrm{hol}}$. In that way, we also obtain the maximal perturbability of $\mathscr{D}_k^{\mathrm{hol}}$, thus completing the proof of Proposition 8.12. □

8D3. *Relations between the spaces $\mathscr{C}_k$ and $\mathscr{D}_k(\lambda)$.* By Remark 8.10, for each $f\in\mathscr{D}_k(\lambda)$, the support of $(\omega-\lambda)f$ is contained in some set $\tilde G_b$, and hence $(\omega-\lambda)f\in\mathscr{C}_k$. So the differential operator $\omega-\lambda$ maps $\mathscr{D}_k(\lambda)$ to $\mathscr{C}_k$. Since the operator ω commutes with the action of $\tilde\Gamma$, we have $(\omega-\lambda)\mathscr{D}_k(\lambda)^{\tilde\Gamma,q}\subset\mathscr{C}_k^{\tilde\Gamma,q}$ for all $q\geq 1$. Similarly, $\boldsymbol{E}^-(\mathscr{D}_k^{\mathrm{hol}})^{\tilde\Gamma,q}\subset\mathscr{C}_{k-2}^{\tilde\Gamma,q}$ for all $q\geq 1$.

Proposition 8.13. *Let $\lambda\in\mathbb{C}$ and $k\in 2\mathbb{Z}$. The following maps are surjective*:

i) $\omega-\lambda : \mathscr{D}_k(\lambda)^{\tilde\Gamma}\to\mathscr{C}_k^{\tilde\Gamma}$,

ii) $\boldsymbol{E}^- : (\mathscr{D}_k^{\mathrm{hol}})^{\tilde\Gamma}\to\mathscr{C}_{k-2}^{\tilde\Gamma}$.

Before presenting the proof we give a corollary:

Corollary 8.14. *For each $q\geq 1$, the maps $\omega-\lambda : \mathscr{D}_k(\lambda_s)^{\tilde\Gamma,q}\to\mathscr{C}_k^{\tilde\Gamma,q}$ and $\boldsymbol{E}^- : (\mathscr{D}_k^{\mathrm{hol}})^{\tilde\Gamma,q}\to\mathscr{C}_k^{\tilde\Gamma,q}$ are surjective.*

Proof. Proposition 8.13 gives the case $q=1$. The rows in the following commutative diagram are exact by Propositions 8.5 and 8.12. See (5-11) for m_q.

$$\begin{tikzcd}
0 \arrow[r] & \mathscr{D}_k(\lambda)^{\tilde\Gamma,q} \arrow[r]\arrow[d,"\omega-\lambda"] & \mathscr{D}_k(\lambda)^{\tilde\Gamma,q+1} \arrow[r,"\mathsf{m}_q"]\arrow[d,"\omega-\lambda"] & (\mathscr{D}_k(\lambda)^{\tilde\Gamma})^{n(\tilde\Gamma,q)} \arrow[r]\arrow[d,"\omega-\lambda"] & 0\\
0 \arrow[r] & \mathscr{C}_k^{\tilde\Gamma,q} \arrow[r]\arrow[d] & \mathscr{C}_k^{\tilde\Gamma,q+1} \arrow[r,"\mathsf{m}_q"]\arrow[d] & (\mathscr{C}_k^{\tilde\Gamma})^{n(\tilde\Gamma,q)} \arrow[r]\arrow[d] & 0\\
 & 0 & \mathrm{coker}(\omega-\lambda) & 0 &
\end{tikzcd} \tag{8-32}$$

The third column is exact by Proposition 8.13. With the exactness of the first column as induction hypothesis, we obtain the vanishing of $\mathrm{coker}(\omega-\lambda)$, and thus the surjectivity of $\omega-\lambda : \mathscr{D}_k(\lambda)^{\tilde\Gamma,q+1}\to\mathscr{C}_k^{\tilde\Gamma,q+1}$, by the snake lemma.

The case of $\boldsymbol{E}^- : (\mathcal{D}_k^{\mathrm{hol}})^{\tilde\Gamma,q} \to \mathcal{C}_k^{\tilde\Gamma,q}$ is similar. □

8D4. *Proof of* Proposition 8.13. We first note that the spaces $\mathcal{D}_k(\lambda)^{\tilde\Gamma}$ and $\mathcal{C}_k^{\tilde\Gamma}$ are invariant under $\tilde Z$. Hence, the weight k is strict and we are dealing with functions on $G = \mathrm{PSL}_2(\mathbb{R})$. (See the first statement in Theorem 6.4.) We use the spectral theory of automorphic forms to prove Proposition 8.13.

We work with the space of square integrable functions on $\tilde\Gamma\backslash\tilde G = \Gamma\backslash G$ of strict weight $k \in 2\mathbb{Z}$, where $G = \mathrm{PSL}_2(\mathbb{R})$. We can view the elements of the Hilbert space $H_k = L^2(\tilde\Gamma\backslash\tilde G)_k = L^2(\Gamma\backslash G)_k$ as functions $z \mapsto f(z,0)$ on $\mathfrak{H}$, transforming according to weight k as indicated in (4-7). The inner product in H_k is given by

$$(f, f_1) = \int_{\mathfrak{F}} f(z,0)\,\overline{f_1(z,0)}\,\frac{dx\,dy}{y^2}.$$

Here $\mathfrak{F}$ can be any fundamental domain for $\Gamma\backslash\mathfrak{H}$. We take it so that for each $b > A_\Gamma$, it has a decomposition

$$\mathfrak{F} = \mathfrak{F}_b \sqcup \bigsqcup_{\kappa\in C} V_\kappa, \qquad V_\kappa = \bigl\{g_\kappa(x+iy) : x_\kappa \le x \le x_\kappa + 1,\ y \ge b\bigr\}, \tag{8-33}$$

with C a system of representatives of the Γ-orbits of cusps, and $x_\kappa \in \mathbb{R}$ depending on $\mathfrak{F}$ and on the earlier choice of the g_κ. The set $\mathfrak{F}_b$ has compact closure in $\mathfrak{H}$.

The differential operator $\omega_k = -y^2\partial_y^2 - y^2\partial_x^2 + iky\partial_x$ in (4-8) determines a densely defined self-adjoint operator A_k in H_k. The spectral theory of automorphic forms gives the decomposition of this operator A_k in terms of Maass forms. One may consult Chapters 4 and 7 in [Iwaniec 1995] for weight 0. For other weights, the proofs are almost completely similar. (See [Roelcke 1966; 1967].)

The spectral decomposition gives the *Parseval formula*

$$(f, f_1) = \sum_\ell a_k^\ell(f)\,\overline{a_k^\ell(f_1)} + \sum_\kappa \frac{1}{2\pi}\int_0^\infty e_k^\kappa(f;it)\,\overline{e_k^\kappa(f_1;it)}\,dt, \tag{8-34}$$

with κ running through a set of representatives of the cuspidal orbit and ℓ indexing a maximal orthonormal system of eigenfunctions ψ_k^ℓ of A_k, with eigenvalue $\lambda_\ell \in \mathbb{R}$. (These eigenvalues are discrete in $\mathbb{R}$, with finite multiplicities.) For each $f \in H_k$, we have $a_k^\ell(f) = (f, \psi_k^\ell)$. If f is sufficiently regular, then the functions $e_k^\kappa(f;\cdot)$ are obtained by integration against the Eisenstein series $E_k^\kappa(it)$ at the cusp κ. The Parseval formula (8-34) shows that the Hilbert space H_k is isomorphic to the direct sum of the subspace spanned by the ψ_k and a number of copies of $L^2\bigl((0,\infty), dt/2\pi\bigr)$. The operator A_k corresponds to a multiplication operator. For f in its domain, we have $a_k^\ell(A_k f) = \lambda_k a_k(f)$ and

$$e_k^\kappa(A_k f; it) = \bigl(\tfrac14 + t^2\bigr)e_k^\kappa(t; it).$$

For the smooth, modulo Γ compactly supported elements $f \in \mathscr{C}_k^{\tilde{\Gamma}}$, we have $a_k^\ell(f) = \mathrm{O}((|\lambda_k|+1)^{-a})$ and $e_k^\kappa\kappa(f; it) = \mathrm{O}((1+t)^{-1})$ for each $a \in \mathbb{R}$. Moreover, $e_k(f; \cdot)$ extends as a holomorphic function on some neighborhood of $i\mathbb{R}$ in $\mathbb{C}$.

Solving $(A_k - \lambda) f_1 = f$ with unknown $f_1 \in H_k$ for a given $f \in \mathscr{C}_k^\Gamma$ can be done by dividing by the factor $\lambda_k - \lambda$, respectively $\frac{1}{2} + it^2 - \lambda_k$ in the spectral decomposition, if these divisions are possible. The requirements are:

i) $a_k^\ell(f) = 0$ if $\lambda_\ell = \lambda$. This condition occurs for at most finitely many indices ℓ.

ii) $e_k^\kappa(t; it_\lambda) = 0$ for all κ if $\lambda = \frac{1}{4} + t_\lambda^2$ with $t_\lambda \in (0, \infty)$.

iii) $s \mapsto e_k^\kappa(f; s)$ has a double zero at $s = 0$ for all κ if $\lambda = \frac{1}{4}$.

These requirements impose finitely many linear conditions on f. So there is a subspace $\mathscr{C}_k(\Gamma, \lambda)$ of $\mathscr{C}_k^\Gamma$ such that for $f \in \mathscr{C}_k(\Gamma, \lambda)$, the equation $(A_k - \lambda) f_1 = f$ can be solved with $f_1 \in H_k$. This means that $(\omega - \lambda) f_1 = f$ holds in the sense of distributions, and hence f_1 is in $C^\infty(\mathfrak{H})$ with the transformation behavior (4-7). It need not be in $\mathscr{C}_k^\Gamma$. However, we have $\omega_k f_1 = \lambda k_1$ on the sets $\{g_\kappa z \in \mathfrak{H} : \operatorname{Im} z > b\}$ for some b depending on the support of f. The square integrability of f_1 ensures that it has less than exponential growth at the cusps, and hence $f_1 \in \mathscr{D}_k(\lambda)^\Gamma$.

So we are done with the proof of part i) for the subspace $\mathscr{C}_k(\Gamma, \lambda)$ of $\mathscr{C}_k^{\tilde{\Gamma}}$ of finite codimension.

Lemma 8.15. *Let κ be the cusp that we keep fixed. Suppose that λ is in the spectrum of A_k. Then there is a finite set $X \subset \mathbb{Z}$ such that, for each $n \in X$, there exist $h_n \in \mathscr{C}_k^{\tilde{\Gamma}}$ of the form*

$$h_n(\gamma \tilde{g}_\kappa(z, \vartheta)) = \begin{cases} e^{2\pi i n x} \chi_n(y) e^{ik\vartheta} & \text{on } \tilde{\Gamma} D_\kappa(A_\Gamma), \\ 0 & \text{elsewhere}, \end{cases} \tag{8-35}$$

for some $\chi_n \in C_c^\infty(A_\Gamma, \infty)$, such that $\{h_n + C_k(\Gamma, \lambda)\}_n$ spans $\mathscr{C}_k^{\tilde{\Gamma}} / \mathscr{C}_k(\Gamma, \lambda)$.

Proof. We shall examine each of the three cases for the eigenvalues of A_k on H_k separately:

• $\lambda = \frac{1}{4} - s^2 \notin \left[\frac{1}{4}, \infty\right)$. Assume $\operatorname{Re} s > 0$. There are finitely many indices $\ell_1, \dots, \ell_m$ such that $\lambda_{\ell_j} = \lambda$. The $\psi_k^{\ell_j}$ form a basis of $\ker(A_k - \lambda)$. Each of these m linearly independent square integrable automorphic forms is given by its Fourier expansion at the fixed cusp κ. By Proposition 7.1, the Fourier terms of nonzero order are multiples of $\omega_k(n, s)$. The Fourier term of order zero is a multiple of $y^{1/2-s} e^{ik\vartheta}$. We choose a set X of m elements in $\mathbb{Z}$ such that the $m \times m$-matrix whose columns are the n-th Fourier coefficients of $\psi_k^{\ell_j}$ $(1 \le j \le m)$, with $n \in X$, is invertible. We choose the $\chi_n \in C_c^\infty$, $n \in X$, in the statement of the lemma, in such a way that $\int_{A_\Gamma}^\infty \chi_n(y) \overline{\omega_k(n, s)(iy, 0)}\, dy/y^2 \neq 0$ or $\int_{A_\Gamma}^\infty \chi_n(y) \overline{y^{1/2-s}}\, dy/y^2 \neq 0$, as the case may be. Consider the linear form on the space $A_k^2(\lambda)$ of square integrable automorphic forms with eigenvalue λ given by

$$\begin{aligned}\psi \mapsto (h_n, \psi) &= \int_{\mathfrak{F}} h_n(z,0)\,\overline{\psi(z,0)}\,\frac{dxdy}{y^2}\\ &= \int_{A_\Gamma}^{\infty}\int_{-1/2}^{1/2} \chi_n(y) e^{2\pi i n x} \bar{a}_0 y^{1/2-\bar{s}}\,\frac{dxdy}{y^2}\\ &\qquad + \sum_{m\neq 0} \bar{a}_m \int_{A_\Gamma}^{\infty}\int_{-1/2}^{1/2} \chi_n(y) e^{2\pi i n x}\,\overline{\omega_k(m,s)(iy,0)}\,\frac{dxdy}{y^2}.\end{aligned}$$

This depends only on the Fourier coefficient of ψ of order n in the expansion at κ. Therefore, the $m \times m$-matrix with the scalar product $(h_n, \psi_k^{\ell_j})$ at position (j, n) is invertible. (Here j runs from 1 to m, and n runs through X.) Hence, there are complex numbers $b_{j,p}$ (with $1 \le j \le m$, $p \in X$) such that $\sum_{n\in X} b_{j,n}(h_n, \psi_k^{\ell_{j'}}) = \delta_{j,j'}$. Setting

$$c_n(f) = \sum_{j'=1}^{m} (f, \psi_k^{\ell_{j'}})\, b_{j',n}$$

for $f \in \mathscr{C}_k^{\tilde{\Gamma}}$, we obtain for $1 \le j \le m$:

$$\sum_{n\in X} c_n(f)(h_n, \psi^{\ell_j}) = (f, \psi_k^{\ell_j}).$$

So $f - \sum_n c_n(f)\, h_n$ is indeed in $\mathscr{C}_k(\Gamma, \lambda)$.

• $\lambda = \frac{1}{4} + t^2$, $t \in \mathbb{R} \smallsetminus \{0\}$. A basis of $\ker(A_k - \lambda)$ in this case consists of Eisenstein series $E_k^{\nu}(it, \cdot)$ ($\nu \in C$) and possibly cusp forms $\psi_k^{\ell_j}$ with $\lambda_{\ell_j} = \lambda$. The proof of the previous case can be applied with the obvious adjustments (for example, replacing scalar products by integrals for the terms corresponding to E_k^{ν}) to give the result. The only essential modification is that we have to use the space $A_k^*(\lambda)$ of automorphic forms with polynomial growth and eigenvalue λ in place of $A_k^2(\lambda)$, because the Eisenstein series are not square integrable. This can be done because (conjugates of) elements of $A_k^*(\lambda)$ appear only integrated against elements of $\mathscr{C}_k^{\tilde{\Gamma}}$, which have compact support modulo $\tilde{\Gamma}$.

• $\lambda = \frac{1}{4}$. Now we have the condition that $e_k^{\kappa}(f - \sum_n h_n; it)$ should have a double zero at $t = 0$ or, equivalently, that the first two terms of the Taylor expansion at $s = 0$ should vanish. Since the first two Taylor terms of $E_k^{\kappa}(-\,; z)$ are linearly independent from the other functions in $A_k^*\left(\frac{1}{4}\right)$, a choice of χ_n with the desired properties is again possible. □

Now we turn to the task of solving $(\omega - \lambda_s) f_1 = h_n$ with $f_1 \in \mathscr{D}_k(\lambda)^{\tilde{\Gamma}}$ for h_n as in Lemma 8.15. We aim at f_1 with support in $\tilde{\Gamma} D_\kappa(A_\Gamma)$. Writing

$$f_1(\tilde{g}_\kappa(z, \vartheta)) = e^{2\pi i n x} h(y) e^{ik\vartheta},$$

the differential equation $(\omega-\lambda) f_1 = h_n$ becomes

$$-y^2h''(y)+\bigl(4\pi^2n^2y^2-2\pi nky-\tfrac14+s^2\bigr)h(y)=\chi_n(y).$$

(Compare (7-3).) This ordinary differential equation is regular on $y \geq A_\Gamma$. It has a unique solution for the initial conditions $h(A_\Gamma)=h'(A_\Gamma)=0$. It is zero below the support of χ_n. Since χ_n has compact support, the function h thus obtained is a solution of the homogeneous Equation (7-3) on (b,∞) for some $b > A_\Gamma$ depending on $\mathrm{Supp}(\chi_n)$. Thus, we see that $(z,\vartheta)\mapsto f_1(\tilde g_\kappa(z,\vartheta))$ is an element of $\mathcal{W}_k(n,s)$. Hence, it may have exponential growth of order $e^{(2\pi|n|+\delta)y}$. This is the point where the need to work with exponentially growing functions arises.

We extend f_1 by $\tilde\Gamma$-invariance, and check that it is an element of $\mathcal{D}_k(\lambda_s)$. This completes the proof of the first statement in Proposition 8.13.

Let us denote by $\boldsymbol{E}_k^-$ the unbounded operator $H_k\to H_{k-2}$ given by the differential operator $\boldsymbol{E}^-$, and similarly $\boldsymbol{E}_{k-2}^+ : H_{k-2}\to H_k$. For the surjectivity of $\boldsymbol{E}^- : (\mathcal{D}_k^{\mathrm{hol}})^{\tilde\Gamma}\to \mathcal{C}_{k-2}^{\tilde\Gamma}$, we first note that, on an eigenfunction of ω in weight $k-2$ with eigenvalue λ, the operator $\boldsymbol{E}_k^-\boldsymbol{E}_{k-2}^+$ acts as multiplication by $-4\bigl(\lambda-\frac k2+\frac{k^2}{4}\bigr)$. See (5-5). We will use $\boldsymbol{E}_{k-2}^+$ to "invert" $\boldsymbol{E}_k^-$.

We can arrange the choice of the orthonormal systems of square integrable eigenfunctions in H_{k-1} and H_k in such a way that $\boldsymbol{E}^-\psi_k^\ell = -\sqrt{k^2-2k+4\lambda_k}\,\psi_{k-2}^\ell$. We have $k^2-2k+4\lambda_k\geq 0$. This factor can be zero for finitely many ℓ, corresponding to a system of holomorphic automorphic forms of weight k. For these ℓ, there is no corresponding eigenfunction ψ_{k-2}^ℓ. Also there can be finitely many indices ℓ such that ψ_{k-2}^ℓ does not occur as image of some ψ_{k-2}^ℓ, corresponding to antiholomorphic automorphic forms. Anyhow, this leads to $a_k^\ell(\boldsymbol{E}^+f)=\sqrt{k^2-2k+4\lambda_\ell}\,a_{k-2}^\ell(f)$, with $k^2-2k+4\lambda_\ell\neq 0$ except for finitely many ℓ. For the Eisenstein series, we have $e_k^\kappa(\boldsymbol{E}^+f;s)=(2s+k-1)e_{k-2}^\kappa(f,s)$. Here the factor is nonzero for all $s\in i[0,\infty)$. So for f in a subspace of finite codimension in $\mathcal{C}_{k-2}^{\tilde\Gamma}$, we can find by the method used for the first part a smooth element $f_1\in H_k$ with $\boldsymbol{E}^-f_1=f$. It is smooth, and near all cusps it is annihilated by $\boldsymbol{E}^-$, and hence it is holomorphic near the cusps. So it is in $(\mathcal{D}_k^{\mathrm{hol}})^{\tilde\Gamma}$.

We are left with finitely many ℓ for which $\boldsymbol{E}_{k-2}^+\psi_{k-2}^\ell=0$. We form functions h_n as in Lemma 8.15, corresponding to a set X of Fourier term orders such that elements of H_{k-2}^a are determined by the Fourier coefficients in X. Solving $\boldsymbol{E}_k^- f_1 = h_n$ leads to the differential equation

$$(-2iy\partial_x+2y\partial_y-k)e^{2\pi inx}\varphi(y)=\chi(y),$$
$$\varphi(y_0)=\varphi'(y_0)=0,$$

with which we proceed as in the previous case.

This establishes the surjectivity of $\boldsymbol{E}^- : (\mathcal{D}_k^{\mathrm{hol}})^{\tilde\Gamma}\to\mathcal{C}_{k-2}^{\tilde\Gamma}$ in Proposition 8.13.

8E. ***Higher-order invariants and Maass forms.*** We now will derive the main results of this paper, Theorems 6.5 and 6.8, from the following result:

Proposition 8.16. *The $\tilde{\Gamma}$-modules*

$$\tilde{\mathscr{E}}'_k(\lambda) := \ker\bigl(\omega - \lambda : \mathscr{D}_k(\lambda) \to \mathscr{C}_k\bigr) \tag{8-36}$$

and

$$\mathscr{H}'_k := \ker\bigl(\boldsymbol{E}^- : \mathscr{D}_k^{\mathrm{hol}} \to \mathscr{C}_{k-2}\bigr) \tag{8-37}$$

are maximally perturbable.

Proof. We have the following extension of the commutative diagram (8-32):

$$\begin{tikzcd}
0 \arrow[r] & \tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma},q} \arrow[r] \arrow[d] & \tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma},q+1} \arrow[r, "\mathsf{m}_q"] \arrow[d] & (\tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma}})^{n(\tilde{\Gamma},q)} \arrow[d] & \\
0 \arrow[r] & \mathscr{D}_k(\lambda)^{\tilde{\Gamma},q} \arrow[r] \arrow[d, "\omega-\lambda"] & \mathscr{D}_k(\lambda)^{\tilde{\Gamma},q+1} \arrow[r, "\mathsf{m}_q"] \arrow[d, "\omega-\lambda"] & (\mathscr{D}_k(\lambda)^{\tilde{\Gamma}})^{n(\tilde{\Gamma},q)} \arrow[r] \arrow[d, "\omega-\lambda"] & 0 \\
0 \arrow[r] & \mathscr{C}_k^{\tilde{\Gamma},q} \arrow[r] \arrow[d] & \mathscr{C}_k^{\tilde{\Gamma},q+1} \arrow[r, "\mathsf{m}_q"] \arrow[d] & (\mathscr{C}_k^{\tilde{\Gamma}})^{n(\tilde{\Gamma},q)} \arrow[r] \arrow[d] & 0 \\
& 0 & 0 & 0 &
\end{tikzcd} \tag{8-38}$$

The exactness of the columns follows from the definition of $\tilde{\mathscr{E}}'_k(\lambda)$, (3-2), the left-exactness of the functor $\hom_{\mathbb{C}[\Gamma]}(I^q\backslash\mathbb{C}[\Gamma], -)$, and Corollary 8.14. Propositions 8.3 and 8.12 imply that the second and third row are exact. The snake lemma then implies that the first row is exact and that

$$\mathsf{m}_q : \tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma},q+1} \to (\tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma}})^{n(\tilde{\Gamma},q)}$$

is surjective.

Replacing in this diagram the space $\tilde{\mathscr{E}}'_l(\lambda)$ by $\mathscr{H}'_k$ and the map $\omega - \lambda$ by $\boldsymbol{E}^-$, we obtain the maximal perturbability of $\mathscr{H}'_k$. □

Proof of Theorems 6.5 and 6.8. The $\tilde{\Gamma}$-module $\tilde{\mathscr{E}}'_k(\lambda)$ is contained in $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)$. See Definition 6.3. It is a smaller space than $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)$, since elements of $\mathscr{D}_k(\lambda)$ have a special structure near the cusps. With (8-22), $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma}}$ is a subspace of $\tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma}}$. Therefore $\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma}} = \tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma}}$, and thus

$$\begin{tikzcd}
0 \arrow[r] & \tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma},q} \arrow[r] \arrow[d] & \tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma},q+1} \arrow[r] \arrow[d] & (\tilde{\mathscr{E}}'_k(\lambda)^{\tilde{\Gamma}})^{n(\tilde{\Gamma},n)} \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma},q} \arrow[r] & \tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma},q+1} \arrow[r] & (\tilde{\mathscr{E}}_k(\tilde{\Gamma}, \lambda)^{\tilde{\Gamma}})^{n(\tilde{\Gamma},q)} &
\end{tikzcd}$$

with exact rows. Induction with respect to q and the snake lemma show that $\tilde{\mathscr{E}}_k(\tilde\Gamma,\lambda)^{\tilde\Gamma,q}$ is equal to $\tilde{\mathscr{E}}'_k(\lambda)^{\tilde\Gamma,q}$ for all q. Hence, the space $\tilde{\mathscr{E}}_k(\tilde\Gamma,\lambda)$ is maximally perturbable.

The proof of Theorem 6.8 is completely similar. □

Index of commonly used notation

$a(y)$	Section 5A
$a_k^\ell(f)$	(8-34)
α	Section 5B
α_i	Section 5C
$\boldsymbol{b}(\boldsymbol{i})$	(3-8), (5-7)
$C_k^\infty(\tilde G)$	(8-11)
$\mathscr{C}_k$	Definition 8.4
$D_\kappa(a)$	(8-16)
$\mathscr{D}_k(\lambda),\mathscr{D}_k^{\mathrm{hol}}$	Definition 8.8
ε_i	Section 5C
E_i	Section 3B
$E_k(\Gamma,\lambda)$	Definition 4.1
$E_k^{\mathrm{hol}}(\Gamma,\lambda_k)$	Section 4B
$\boldsymbol{E}^\pm$	(5-4)
$e_k^\kappa(f;it)$	(8-34)
$\mathscr{E}_k(y_0,\lambda),\mathscr{E}_k^{\mathrm{hol}}(y_0)$	Definition 8.6
$\tilde{\mathscr{E}}_k(\tilde\Gamma,\lambda)$	Definition 6.3
$\tilde E_r(\tilde\Gamma,\chi,\lambda)$	Definition 6.3
$E_k(\Gamma,\lambda)$	Definition 4.1
$\boldsymbol{f_i}$	(8-2)
g_κ	(4-1)
$\tilde G$	Section 5A
$\tilde G_a$	(8-17)
$h_{\boldsymbol{i}}$	(8-9), (8-10)
H_i	Section 3B
$h_k^{\boldsymbol{m}}(n,s)$	(7-11)
$\mathscr{H}_k(\tilde\Gamma),\mathscr{H}_k^p(\tilde\Gamma),\mathscr{H}_k^c(\tilde\Gamma)$	Section 6C
η_i	Section 5C
$\eta_r(n;z,\vartheta)$	(7-16)
$\eta_k(n)$	(7-17)
$\eta_k^{\boldsymbol{m}}(n;z,\vartheta)$	(7-18)
$k(\vartheta)$	Section 5A
κ_i	Section 5C
L_k	(4-3)
$L(z,\vartheta)$	(6-3)
m_q	(3-3)
$\mathscr{M}_k(\Gamma,\lambda)$	Definition 4.1
$M_k(\Gamma,\lambda)$	Definition 4.1
$\mathscr{M}_k^{\mathrm{hol}}(\Gamma,\lambda_k)$	Section 4B
μ_f	(3-5)
$n_{\mathrm{ell}},n_{\mathrm{par}}$	Section 3B
$n(x)$	Section 5A
$n(\Gamma,q)$	(3-4)
P_i	Section 3B
pr, pr_2	Section 5A
π_i	Section 5C
Q_n	(7-10)
s	Section 5B
$S(y_0)$	(8-19)
t	Section 5B
$t(\Gamma)$	Section 3B
$\mathscr{V}_k(n,s),\mathscr{V}_k^0(n,s)$	Definition 7.2
$\mathscr{W}_r(\nu,s)$	(7-2)
$\boldsymbol{W}$	(5-4)
ζ	Section 5B
ω	(5-5)
$\omega_r,\hat\omega_r$	(7-4)

References

[Bruggeman 1986] R. W. Bruggeman, "Modular forms of varying weight, III", *J. Reine Angew. Math.* **371** (1986), 144–190. MR 88e:11040 Zbl 0588.10021

[Bruggeman 1994] R. W. Bruggeman, *Families of automorphic forms*, Monographs in Mathematics **88**, Birkhäuser, Boston, MA, 1994. MR 95k:11060 Zbl 0821.11029

[Bruinier et al. 2008] J. H. Bruinier, K. Ono, and R. C. Rhoades, "Differential operators for harmonic weak Maass forms and the vanishing of Hecke eigenvalues", *Math. Ann.* **342**:3 (2008), 673–693. MR 2009f:11046 Zbl 05375997

[Chen 1971] K.-t. Chen, "Algebras of iterated path integrals and fundamental groups", *Trans. Amer. Math. Soc.* **156** (1971), 359–379. MR 43 #1069 Zbl 0217.47705

[Chinta et al. 2002] G. Chinta, N. Diamantis, and C. O'Sullivan, "Second order modular forms", *Acta Arith.* **103**:3 (2002), 209–223. MR 2003b:11037 Zbl 1020.11025

[Deitmar 2008] A. Deitmar, "Higher order invariants, cohomology, and automorphic forms", preprint, 2008. arXiv 0811.1088v1

[Deitmar 2009] A. Deitmar, "Higher order group cohomology and the Eichler–Shimura map", *J. Reine Angew. Math.* **629** (2009), 221–235. MR 2011a:11107 Zbl 05551467

[Diamantis 1999] N. Diamantis, "Special values of higher derivatives of L-functions", *Forum Math.* **11**:2 (1999), 229–252. MR 2000a:11075 Zbl 0932.11032

[Diamantis and O'Sullivan 2008] N. Diamantis and C. O'Sullivan, "The dimensions of spaces of holomorphic second-order automorphic forms and their cohomology", *Trans. Amer. Math. Soc.* **360**:11 (2008), 5629–5666. MR 2009h:11067 Zbl 1158.11021

[Diamantis and Sim 2008] N. Diamantis and D. Sim, "The classification of higher-order cusp forms", *J. Reine Angew. Math.* **622** (2008), 121–153. MR 2010a:11067 Zbl 1230.11049

[Diamantis and Sreekantan 2006] N. Diamantis and R. Sreekantan, "Iterated integrals and higher order automorphic forms", *Comment. Math. Helv.* **81**:2 (2006), 481–494. MR 2007b:11050 Zbl 1156.11019

[Goldfeld 1995] D. Goldfeld, "Special values of derivatives of L-functions", pp. 159–173 in *Number theory* (Halifax, 1994), edited by K. Dilcher, CMS Conf. Proc. **15**, American Mathematical Society, Providence, RI, 1995. MR 96h:11039 Zbl 0845.11021

[Iwaniec 1995] H. Iwaniec, *Introduction to the spectral theory of automorphic forms*, Revista Matemática Iberoamericana, Madrid, 1995. MR 96f:11078 Zbl 0847.11028

[Koecher and Krieg 1998] M. Koecher and A. Krieg, *Elliptische Funktionen und Modulformen*, Springer, Berlin, 1998. MR 2000e:11047 Zbl 0895.11001

[Kra 1969] I. Kra, "On cohomology of kleinian groups", *Ann. of Math.* (2) **89** (1969), 533–556. MR 41 #8656a Zbl 0193.04003

[Lang 1975] S. Lang, $SL_2(\mathbf{R})$, Addison-Wesley, Reading, MA, 1975. MR 55 #3170 Zbl 0311.22001

[Lehner 1964] J. Lehner, *Discontinuous groups and automorphic functions*, Mathematical Surveys **8**, American Mathematical Society, Providence, R.I., 1964. MR 29 #1332 Zbl 0178.42902

[Maass 1983] H. Maass, *Lectures on modular functions of one complex variable*, 2nd ed., Tata Inst. Fund. Res. Lectures on Math. and Phys. **29**, Tata Institute of Fundamental Research, Bombay, 1983. MR 85g:11034 Zbl 0539.10021

[Manin 1972] Y. I. Manin, "Parabolic points and zeta functions of modular curves", *Izv. Akad. Nauk SSSR Ser. Mat.* **36** (1972), 19–66. In Russian; translated in *Math. USSR-Izv.* **6** (1972), 19–64. MR 47 #3396 Zbl 0243.14008

[Petersson 1948] H. Petersson, "Über den Bereich absoluter Konvergenz der Poincaréschen Reihen", *Acta Math.* **80** (1948), 23–63. MR 10,111d Zbl 0031.12502

[Petridis and Risager 2004] Y. N. Petridis and M. S. Risager, "Modular symbols have a normal distribution", *Geom. Funct. Anal.* **14**:5 (2004), 1013–1043. MR 2005h:11101 Zbl 1142.11332

[Roelcke 1966] W. Roelcke, "Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene, I", *Math. Ann.* **167**:4 (1966), 292–337. MR 0243062 Zbl 0152.07705

[Roelcke 1967] W. Roelcke, "Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene, II", *Math. Ann.* **168** (1967), 261–324. MR 0243062 Zbl 0152.07705

[Slater 1960] L. J. Slater, *Confluent hypergeometric functions*, Cambridge University Press, New York, 1960. MR 21 #5753 Zbl 0086.27502

[Sreekantan 2009] R. Sreekantan, "Higher order modular forms and mixed Hodge theory", *Acta Arith.* **139**:4 (2009), 321–340. MR 2010j:11068 Zbl 1225.11052

Communicated by Peter Sarnak
Received 2011-04-29 Revised 2011-09-20 Accepted 2011-10-27

r.w.bruggeman@uu.nl *Mathematisch Instituut, Universiteit Utrecht, Postbus 80010, 3508 TA Utrecht, Netherlands*

nikolaos.diamantis@maths.nottingham.ac.uk *School of Mathematical Science, University of Nottingham, Nottingham, NG7 2RD, United Kingdom*

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1459

Log canonical thresholds, F-pure thresholds, and nonstandard extensions

Bhargav Bhatt, Daniel J. Hernández,
Lance Edward Miller and Mircea Mustaţă

We present a new relation between an invariant of singularities in characteristic zero (the log canonical threshold) and an invariant of singularities defined via the Frobenius morphism in positive characteristic (the F-pure threshold). We show that the set of limit points of sequences of the form (c_p), where c_p is the F-pure threshold of an ideal on an n-dimensional smooth variety in characteristic p, coincides with the set of log canonical thresholds of ideals on n-dimensional smooth varieties in characteristic zero. We prove this by combining results of Hara and Yoshida with nonstandard constructions.

1. Introduction

The connection between invariants of singularities in characteristic zero and positive characteristic is a topic that has recently attracted a lot of attention. Typically, the invariants of singularities that arise in birational geometry are defined via divisorial valuations. In characteristic zero, one can use (log) resolutions of singularities to compute such invariants. On the other hand, in commutative algebra in positive characteristic one defines invariants using the action of the Frobenius morphism. It turns out that these invariants have subtle connections, some of them proven, and some still conjectural; see, for example, [Hara and Watanabe 2002; Hara and Yoshida 2003; Mustaţă et al. 2005]. The typical such connection involves reduction from characteristic zero to positive characteristic. In this note we describe a different, though related connection. We use nonstandard constructions to study limits of invariants in positive characteristic, where the characteristic tends to infinity, in terms of invariants in characteristic zero.

The invariants we study in this paper are the log canonical threshold (in characteristic zero) and the F-pure threshold (in positive characteristic). The log canonical threshold is an invariant that plays an important role in birational geometry; see

Hernández was partially supported by RTG grant 0502170. Mustaţă was partially supported by NSF grant DMS-0758454 and a Packard Fellowship.
MSC2010: primary 13A35; secondary 13L05, 14B05, 14F18.
Keywords: F-pure threshold, log canonical threshold, ultrafilters, multiplier ideals, test ideals.

[Kollár 1997; Ein and Mustaţă 2006]. Given an irreducible, smooth scheme X defined over a field k of characteristic zero, and a proper ideal $\mathfrak{a} \subset \mathcal{O}_X$, the log canonical threshold of $\mathfrak{a}$ is denoted by $\mathrm{lct}(\mathfrak{a})$. For the precise definition in terms of a log resolution of $(X, \mathfrak{a})$, we refer to Section 2. Given a point $x \in V(\mathfrak{a})$, one defines $\mathrm{lct}_x(\mathfrak{a})$ to be $\mathrm{lct}(\mathfrak{a}|_U)$, where U is a small enough open neighborhood of x in X.

On the other hand, suppose that W is a smooth scheme of finite type over a perfect field L of positive characteristic p. For a proper ideal $\mathfrak{a} \subset \mathcal{O}_W$, the F-pure threshold $\mathrm{fpt}(\mathfrak{a})$ was introduced and studied in [Takagi and Watanabe 2004]. Given $x \in V(\mathfrak{a})$, one defines as before the local version of this invariant, denoted $\mathrm{fpt}_x(\mathfrak{a})$. The original definition of the F-pure threshold involved notions and constructions from tight closure theory. However, since we always assume that the ambient scheme is smooth, one can use an alternative description, following [Mustaţă et al. 2005; Blickle et al. 2008] (see Section 2 below). Part of the interest in the study of the F-pure threshold comes from the fact that it shares many of the formal properties of the log canonical threshold.

Before stating our main result, let us recall the fundamental connection between log canonical thresholds and F-pure thresholds via reduction mod p. Suppose that X and $\mathfrak{a} \subset \mathcal{O}_X$ are defined over k, as above. We may choose a subring $A \subset k$, finitely generated over $\mathbb{Z}$, and models X_A and $\mathfrak{a}_A \subset \mathcal{O}_{X_A}$ for X and $\mathfrak{a}$, respectively, defined over A. In particular, given any closed point $s \in \operatorname{Spec} A$, we may consider the corresponding reductions X_s and $\mathfrak{a}_s \subset \mathcal{O}_{X_s}$ defined over the finite residue field of s denoted $k(s)$. One of the main results in [Hara and Yoshida 2003] implies the following relation between log canonical thresholds and F-pure thresholds: after possibly replacing A by a localization A_a for some nonzero $a \in A$,

i) $\mathrm{lct}(\mathfrak{a}) \geq \mathrm{fpt}(\mathfrak{a}_s)$ for every closed point in $s \in \operatorname{Spec} A$, and

ii) there is a sequence of closed points $s_m \in \operatorname{Spec} A$ with $\lim_{m\to\infty} \mathrm{char}(k(s_m)) = \infty$ and such that $\lim_{m\to\infty} \mathrm{fpt}(\mathfrak{a}_{s_m}) = \mathrm{lct}(\mathfrak{a})$.

It is worth pointing out that a fundamental open problem in the field predicts that in this setting there is a dense set of closed points $S \subset \operatorname{Spec} A$ such that $\mathrm{lct}(\mathfrak{a}) = \mathrm{fpt}(\mathfrak{a}_s)$ for every $s \in S$.

We now turn to the description of our main result. For every $n \geq 1$, let $\mathcal{L}_n$ be the set of all $\mathrm{lct}(\mathfrak{a})$, where the pair $(X, \mathfrak{a})$ is as above, with $\dim(X) = n$. Similarly, given n and a prime p, let $\mathcal{F}(p)_n$ be the set of all $\mathrm{fpt}(\mathfrak{a})$, where $(W, \mathfrak{a})$ is as above, with $\dim(W) = n$, and W defined over a field of characteristic p. The following is our main result.

Theorem 1.1. *For every $n \geq 1$, the set of limit points of all sequences (c_p), where $c_p \in \mathcal{F}(p)_n$ for every prime p, coincides with $\mathcal{L}_n$.*

A key ingredient in the proof of Theorem 1.1 is provided by ultraproduct constructions. Note that if $c \in \mathcal{L}_n$ is given as $c = \mathrm{lct}(\mathfrak{a})$, then the above mentioned results in [Hara and Yoshida 2003] (more precisely, property ii) above) imply that $c = \lim_{p\to\infty} c_p$, where for $p \gg 0$ prime, c_p is the F-pure threshold of a suitable reduction $\mathfrak{a}_s \subset \mathcal{O}_{X_s}$ with $\mathrm{char}(k(s)) = p$. Thus the interesting statement in the above theorem is the converse: given pairs $(W_m, \mathfrak{a}_m)$ over L_m with $\dim(W_m) = n$, $\lim_{m\to\infty} \mathrm{char}(L_m) = \infty$, and with $\lim_{m\to\infty} \mathrm{fpt}(\mathfrak{a}_m) = c$, there is a pair $(X, \mathfrak{a})$ in characteristic zero with $\dim(X) = n$ and such that $c = \mathrm{lct}(\mathfrak{a})$.

It is easy to see that we may assume that each $W_m = \mathrm{Spec}(L_m[x_1, \dots, x_n])$ and $c_m = \mathrm{fpt}_0(\mathfrak{a}_m)$ for some $\mathfrak{a}_m \subseteq (x_1, \dots, x_n)$. If we put $\mathfrak{a}_m^{(d)} = \mathfrak{a}_m + (x_1, \dots, x_n)^d$, we have $|\mathrm{fpt}(\mathfrak{a}_m^{(d)}) - \mathrm{fpt}(\mathfrak{a}_m)| \leq n/d$ for all m and d. Ultraproduct constructions give nonstandard extensions of our algebraic structures. In particular, we get a field $k = [L_m]$ of characteristic zero. Since all ideals $\mathfrak{a}_m^{(d)}$ are generated in degree less than or equal to d, they determine an ideal $\mathfrak{a}^{(d)}$ in $k[x_1, \dots, x_n]$. The key point is to show that for every $\varepsilon > 0$, we have $|\,\mathrm{lct}_0(\mathfrak{a}^{(d)}) - \mathrm{fpt}_0(\mathfrak{a}_m)| < \varepsilon$ for infinitely many m. This easily implies that $\lim_{d\to\infty} \mathrm{lct}_0(\mathfrak{a}^{(d)}) = c$, and since $\mathcal{L}_n$ is closed by [de Fernex and Mustață 2009, Theorem 1.3] (incidentally, this is proved in *loc. cit.* also by nonstandard arguments), we conclude that $c \in \mathcal{L}_n$.

As in [Hara and Yoshida 2003], the result relating the log canonical threshold of $\mathfrak{a}^{(d)}$ and the F-pure thresholds of $\mathfrak{a}_m^{(d)}$ follows from a more general result relating the multiplier ideals of $\mathfrak{a}^{(d)}$ and the test ideals of $\mathfrak{a}_m^{(d)}$ (see Theorem 4.1 below). We prove this by following, with some simplifications, the main line of argument in [ibid.] in our nonstandard setting.

The use of ultraproduct techniques in commutative algebra has been pioneered by Schoutens; see [Schoutens 2010] and the list of references therein. This point of view has been particularly effective for passing from positive characteristic to characteristic zero in an approach to tight closure theory and to its applications. Our present work combines ideas of Schoutens [2005] with the nonstandard approach to studying limits of log canonical thresholds and F-pure thresholds from [de Fernex and Mustață 2009] and [Blickle et al. 2009], respectively.

The paper is structured as follows. In Section 2 we review the definitions of multiplier ideals and test ideals, and recall how the log canonical threshold and the F-pure threshold appear as the first jumping numbers in these families of ideals. In Section 3 we review the basic definitions involving ultraproducts. For the benefit of the reader, we also describe in detail how to go from schemes, morphisms, and sheaves over an ultraproduct of fields to sequences of similar objects defined over the corresponding fields. The proof of Theorem 1.1 is given in Section 4.

2. Multiplier ideals and test ideals

In this section we review the basic facts that we will need about multiplier ideals and test ideals. Both these concepts can be defined under mild assumptions on the singularities of the ambient space. However, since our main result only deals with smooth varieties, we will restrict to this setting in order to simplify the definitions.

2A. ***Multiplier ideals and the log canonical threshold.*** In what follows we recall the definition and some basic properties of multiplier ideals and log canonical thresholds. For details and further properties, we refer the reader to [Lazarsfeld 2004, §9].

Let k be a field of characteristic zero, and X an irreducible and smooth scheme of finite type over k. Given a nonzero ideal[1] $\mathfrak{a}$ on X, its multiplier ideals are defined as follows. Let us fix a log resolution of the pair $(X, \mathfrak{a})$: this is a projective, birational morphism $\pi : Y \to X$ with Y smooth and $\mathfrak{a}\cdot\mathcal{O}_Y = \mathcal{O}_Y(-F)$ for an effective divisor F such that $F + \mathrm{Exc}(\pi)$ is a divisor with simple normal crossings. Here $\mathrm{Exc}(\pi)$ denotes the exceptional divisor of π. Such resolutions exist by Hironaka's theorem, since we are in characteristic zero. Recall that $K_{Y/X}$ denotes the relative canonical divisor of π: this is an effective divisor supported on $\mathrm{Exc}(\pi)$ such that $\mathcal{O}_Y(K_{Y/X}) \simeq \omega_Y \otimes f^*(\omega_X^{-1})$. With this notation, the multiplier ideal of $\mathfrak{a}$ of exponent $\lambda \in \mathbb{R}_{\geq 0}$ is defined by

$$\mathcal{J}(\mathfrak{a}^\lambda) := \pi_*\mathcal{O}_Y(K_{Y/X} - \lfloor \lambda F \rfloor). \tag{1}$$

Here, for a divisor with real coefficients $E = \sum_i a_i E_i$, we write $\lfloor E \rfloor = \sum_i \lfloor a_i \rfloor E_i$, where $\lfloor a_i \rfloor$ is the largest integer $\leq a_i$. It is a basic fact that the definition of multiplier ideals is independent of resolution.

Let us consider some easy consequences of the definition (1). If $\lambda < \mu$, then $\mathcal{J}(\mathfrak{a}^\mu) \subseteq \mathcal{J}(\mathfrak{a}^\lambda)$. Furthermore, given λ, there is $\varepsilon > 0$ such that $\mathcal{J}(\mathfrak{a}^\lambda) = \mathcal{J}(\mathfrak{a}^\mu)$ whenever $\lambda \leq \mu \leq \lambda + \varepsilon$. A positive λ is a *jumping number* of $\mathfrak{a}$ if $\mathcal{J}(\mathfrak{a}^\lambda) \neq \mathcal{J}(\mathfrak{a}^\mu)$ for all $\mu < \lambda$. If we write $F = \sum_i a_i E_i$, it follows from (1) that if λ is a jumping number, then $\lambda a_i \in \mathbb{Z}$ for some i. In particular, we see that the jumping numbers of $\mathfrak{a}$ form a discrete set of rational numbers.

Suppose now that $\mathfrak{a} \neq \mathcal{O}_X$. The smallest jumping number of $\mathfrak{a}$ is the *log canonical threshold* $\mathrm{lct}(\mathfrak{a})$. Note that if $0 \leq \lambda \ll 1$, then $\mathcal{J}(\mathfrak{a}^\lambda) = \mathcal{O}_X$, hence $\mathrm{lct}(\mathfrak{a}) = \min\{\lambda \mid \mathcal{J}(\mathfrak{a}^\lambda) \neq \mathcal{O}_X\}$ (this is finite since $\mathfrak{a} \neq \mathcal{O}_X$). If $\mathfrak{a} \subseteq \mathfrak{b}$, then $\mathcal{J}(\mathfrak{a}^\lambda) \subseteq \mathcal{J}(\mathfrak{b}^\lambda)$ for all λ; in particular, we have $\mathrm{lct}(\mathfrak{a}) \leq \mathrm{lct}(\mathfrak{b})$. We make the convention $\mathrm{lct}(0) = 0$ and $\mathrm{lct}(\mathcal{O}_X) = \infty$.

It is sometimes convenient to also have available a local version of the log canonical threshold. If $x \in X$, then we put $\mathrm{lct}_x(\mathfrak{a}) := \max_V \mathrm{lct}(\mathfrak{a}|_V)$, where the maximum ranges over all open neighborhoods V of x. Equivalently, we have

$$\mathrm{lct}_x(\mathfrak{a}) = \min\{\lambda \mid \mathcal{J}(\mathfrak{a}^\lambda) \cdot \mathcal{O}_{X,.x} \neq \mathcal{O}_{X,x}\}$$

(with the convention that this is 0 if $\mathfrak{a} = (0)$, and it is infinite if $x \notin V(\mathfrak{a})$). Note that given a proper ideal $\mathfrak{a}$ on X, there is a closed point $x \in X$ such that $\mathrm{lct}(\mathfrak{a}) = \mathrm{lct}_x(\mathfrak{a})$.

[1] Every ideal sheaf that we consider is assumed to be coherent.

The definition of multiplier ideals commutes with extension of the base field, as follows. For a proof, see the proof of [de Fernex and Mustaţă 2009, Propositions 2.9].

Proposition 2.1. *Let $\mathfrak{a}$ be an ideal on X. If $k \subset k'$ is a field extension, and $\varphi\colon X' = X \times_{\operatorname{Spec} k} \operatorname{Spec} k' \to X$ and $\mathfrak{a}' = \mathfrak{a} \cdot \mathcal{O}_{X'}$, then $\mathcal{J}(\mathfrak{a}'^{\lambda}) = \mathcal{J}(\mathfrak{a}^{\lambda}) \cdot \mathcal{O}_{X'}$ for every $\lambda \in \mathbb{R}_{\geq 0}$. In particular, $\operatorname{lct}_{x'}(\mathfrak{a}') = \operatorname{lct}_{\varphi(x')}(\mathfrak{a})$ for every $x' \in X'$.*

Recall from Section 1 that $\mathcal{L}_n$ consists of all nonnegative rational numbers of the form $\operatorname{lct}(\mathfrak{a})$, where $\mathfrak{a}$ is a proper ideal on an n-dimensional smooth variety over a field k of characteristic zero. It is clear that equivalently, we may consider the invariants $\operatorname{lct}_x(\mathfrak{a})$, where $(X, \mathfrak{a})$ is as above, and $x \in X$ is a closed point. Furthermore, by Proposition 2.1 we may assume that k is algebraically closed. One can show that in this definition we can fix the algebraically closed field k and assume that $X = \mathbb{A}^n_k$ and obtain the same set; see [ibid., Propositions 3.1 and 3.3]. Furthermore, we will make use of the fact that $\mathcal{L}_n$ is a closed set; see [ibid., Theorem 1.3].

2B. ***Test ideals and the F-pure threshold.*** In this section we assume that X is an irreducible, Noetherian, regular scheme of characteristic $p > 0$. We also assume that X is F-finite, that is, the Frobenius morphism $F\colon X \to X$ is finite (in fact, most of the time X will be a scheme of finite type over a perfect field, in which case this assumption is clearly satisfied). Recall that for an ideal J on X, the e-th Frobenius power $J^{[p^e]}$ is generated by u^{p^e}, where u varies over the (local) generators of J.

Suppose that $\mathfrak{b}$ is an ideal on X. Given a positive integer e, one can show that there is a unique smallest ideal J such that $\mathfrak{b} \subseteq J^{[p^e]}$. This ideal is denoted by $\mathfrak{b}^{[1/p^e]}$. Given a nonzero ideal $\mathfrak{a}$ and $\lambda \in \mathbb{R}_{\geq 0}$, one has

$$(\mathfrak{a}^{\lceil \lambda p^e \rceil})^{[1/p^e]} \subseteq (\mathfrak{a}^{\lceil \lambda p^{e+1} \rceil})^{[1/p^{e+1}]}$$

for all $e \geq 1$ (here $\lceil u \rceil$ denotes the smallest integer greater than or equal to u). By the Noetherian property, it follows that there is an ideal $\tau(\mathfrak{a}^{\lambda})$ that is equal to $(\mathfrak{a}^{\lceil \lambda p^e \rceil})^{[1/p^e]}$ for all $e \gg 0$. This is the *test ideal* of $\mathfrak{a}$ of exponent λ. For details and basic properties of test ideals, we refer to [Blickle et al. 2008].

It is again clear that if $\lambda < \mu$, then $\tau(\mathfrak{a}^{\mu}) \subseteq \tau(\mathfrak{a}^{\lambda})$. It takes a little argument to show that given any λ, there is $\varepsilon > 0$ such that $\tau(\mathfrak{a}^{\lambda}) = \tau(\mathfrak{a}^{\mu})$ whenever $\lambda \leq \mu \leq \lambda + \varepsilon$; see [ibid., Proposition 2.14]. We say that $\lambda > 0$ is an *F-jumping number* of $\mathfrak{a}$ if $\tau(\mathfrak{a}^{\lambda}) \neq \tau(\mathfrak{a}^{\mu})$ for every $\mu < \lambda$. It is proved in [ibid., Theorem 3.1] that if X is a scheme of finite type over an F-finite field, then the F-jumping numbers of $\mathfrak{a}$ form a discrete set of rational numbers.

The smallest F-jumping number of $\mathfrak{a}$ is the *F-pure threshold* $\operatorname{fpt}(\mathfrak{a})$. Since $\tau(\mathfrak{a}^{\lambda}) = \mathcal{O}_X$ for $0 \leq \lambda \ll 1$, the F-pure threshold is characterized by

$$\operatorname{fpt}(\mathfrak{a}) = \min\{\lambda \mid \tau(\mathfrak{a}^{\lambda}) \neq \mathcal{O}_X\}.$$

Note that this is finite if and only if $\mathfrak{a} \neq \mathcal{O}_X$. We make the convention that $\mathrm{fpt}(\mathfrak{a}) = 0$ if $\mathfrak{a} = (0)$.

We have a local version of the F-pure threshold: given $x \in X$, we put $\mathrm{fpt}_x(\mathfrak{a}) := \max_V \mathrm{fpt}(\mathfrak{a}|_V)$, where the maximum is over all open neighborhoods V of x. It can be also described by

$$\mathrm{fpt}_x(\mathfrak{a}) = \min\{\lambda \mid \tau(\mathfrak{a}^\lambda) \cdot \mathcal{O}_{X,x} \neq \mathcal{O}_{X,x}\},$$

and it is finite if and only if $x \in V(\mathfrak{a})$. Note that given any $\mathfrak{a}$, there is $x \in X$ such that $\mathrm{fpt}(\mathfrak{a}) = \mathrm{fpt}_x(\mathfrak{a})$.

We will make use of the following two properties of F-pure thresholds.

Proposition 2.2 [Blickle et al. 2008, Proposition 2.13]. *If $\mathfrak{a}$ is an ideal on X and $S = \widehat{\mathcal{O}_{X,x}}$ is the completion of the local ring of X at a point $x \in X$, then $\tau(\mathfrak{a}^\lambda) \cdot S = \tau((\mathfrak{a} \cdot S)^\lambda)$ for every $\lambda \geq 0$. In particular*, $\mathrm{fpt}_x(\mathfrak{a}) = \mathrm{fpt}(\mathfrak{a} \cdot S)$.

Proposition 2.3 [Blickle et al. 2009, Corollary 3.4]. *If $\mathfrak{a}$ and $\mathfrak{b}$ are ideals on X, and $x \in V(\mathfrak{a}) \cap V(\mathfrak{b})$ is such that $\mathfrak{a} \cdot \mathcal{O}_{X,x} + \mathfrak{m}^r = \mathfrak{b} \cdot \mathcal{O}_{X,x} + \mathfrak{m}^r$ for some $r \geq 1$, where $\mathfrak{m}$ is the maximal ideal in $\mathcal{O}_{X,x}$, then*

$$|\mathrm{fpt}_x(\mathfrak{a}) - \mathrm{fpt}_x(\mathfrak{b})| \leq \frac{\dim(\mathcal{O}_{X,x})}{r}.$$

The local F-pure threshold admits the following alternative description, following [Mustaţă et al. 2005]. If $\mathfrak{a}$ is an ideal on X and $x \in V(\mathfrak{a})$, let $\nu(e)$ denote the largest r such that $\mathfrak{a}^r \cdot \mathcal{O}_{X,x} \not\subseteq \mathfrak{m}^{[p^e]}$, where $\mathfrak{m}$ is the maximal ideal in $\mathcal{O}_{X,x}$ (we make the convention $\nu(e) = 0$ if $\mathfrak{a} = 0$). One can show that

$$\mathrm{fpt}_x(\mathfrak{a}) = \lim_{e \to \infty} \frac{\nu(e)}{p^e} \tag{2}$$

(see [Blickle et al. 2008, Proposition 2.29]). This immediately implies the assertion in the following proposition.

Proposition 2.4. *Let $L \subset L'$ be a field extension of F-finite fields of positive characteristic. If $\mathfrak{a} \subseteq L[x_1, \ldots, x_n]$ is an ideal vanishing at the origin, and $\mathfrak{a}' = \mathfrak{a} \cdot L'[x_1, \ldots, x_n]$, then* $\mathrm{fpt}_0(\mathfrak{a}) = \mathrm{fpt}_0(\mathfrak{a}')$.

Recall that we have introduced in Section 1 the set $\mathcal{F}(p)_n$ consisting of all invariants of the form $\mathrm{fpt}(\mathfrak{a})$, where $\mathfrak{a}$ is a proper ideal on an irreducible, n-dimensional smooth scheme of finite type over L, with L a perfect field of characteristic p. We can define two other related subsets of $\mathbb{R}_{\geq 0}$. Let $\mathcal{F}(p)'_n$ be the set of invariants $\mathrm{fpt}_0(\mathfrak{a})$, where $\mathfrak{a} \subset L[x_1, \ldots, x_n]$ is an ideal vanishing at the origin, and L is an algebraically closed field of characteristic p. We also put $\mathcal{F}(p)''_n$ for the set of all $\mathrm{fpt}(\mathfrak{a})$, where $\mathfrak{a}$ is a proper ideal on an irreducible, regular, n-dimensional F-finite scheme of characteristic p. We clearly have the following inclusions:

$$\mathcal{F}(p)'_n \subseteq \mathcal{F}(p)_n \subseteq \mathcal{F}(p)''_n. \tag{3}$$

Proposition 2.5. *$\mathcal{F}(p)'_n$ is dense in $\mathcal{F}(p)''_n$ (hence also in $\mathcal{F}(p)_n$).*

This implies that in Theorem 1.1 we may replace the sets $\mathcal{F}(p)_n$ by $\mathcal{F}(p)'_n$ or by $\mathcal{F}(p)''_n$.

Proof of Proposition 2.5. Suppose that $\mathfrak{a}$ is a proper ideal on X, where X is irreducible, regular, F-finite, n-dimensional, and of characteristic p. Let $c = \mathrm{fpt}(\mathfrak{a})$. We can find $x \in X$ such that $c = \mathrm{fpt}_x(\mathfrak{a})$. By Proposition 2.2, we have

$$c = \mathrm{fpt}(\mathfrak{a} \cdot \widehat{\mathcal{O}_{X,x}}).$$

Note that by Cohen's theorem, we have an isomorphism $\widehat{\mathcal{O}_{X,x}} \simeq L[\![x_1, \ldots, x_d]\!]$, with L an F-finite field, and $d \le n$. If $\mathfrak{m}$ is the maximal ideal in $\widehat{\mathcal{O}_{X,x}}$ and $c_i = \mathrm{fpt}(\mathfrak{a} \cdot \widehat{\mathcal{O}_{X,x}} + \mathfrak{m}^i)$, then Proposition 2.3 gives $c = \lim_{i\to\infty} c_i$. On the other hand, there are ideals $\mathfrak{b}_i \subset L[x_1, \ldots, x_d]$ such that $\mathfrak{b}_i \cdot \widehat{\mathcal{O}_{X,x}} = \mathfrak{a} \cdot \widehat{\mathcal{O}_{X,x}} + \mathfrak{m}^i$, and another application of Proposition 2.2 gives $c_i = \mathrm{fpt}_0(\mathfrak{b}_i)$. It is easy to see (for example, from formula (2)) that $c_i = \mathrm{fpt}_0(\mathfrak{b}_i \cdot L[x_1, \ldots, x_n])$. It now follows from Proposition 2.4 that $c_i = \mathrm{fpt}_0(\mathfrak{b}_i \cdot \overline{L}[x_1, \ldots, x_n])$, where $\overline{L}$ is an algebraic closure of L. Therefore all c_i lie in $\mathcal{F}(p)'_n$, which proves the proposition. $\square$

In Section 4 we will use a slightly different description of the test ideals that we now present. More precisely, we give a different description of $\mathfrak{b}^{[1/p^e]}$, when $\mathfrak{b}$ is an arbitrary ideal on X. Suppose that X is an irreducible, smooth scheme of finite type over a perfect field L of characteristic p. Let $\omega_X = \wedge^n \Omega_{X/L}$, where $n = \dim(X)$. Recall that the Cartier isomorphism (see [Deligne and Illusie 1987]) gives in particular an isomorphism $\omega_X \simeq \mathcal{H}^n(F_*\Omega^\bullet_{X/L})$, where F is the (absolute) Frobenius morphism, and $\Omega^\bullet_{X/L}$ is the de Rham complex of X. In particular, we get a surjective $\mathcal{O}_X$-linear map $t_X\colon F_*\omega_X \to \omega_X$. This can be explicitly described in coordinates, as follows. Suppose that $u_1, \ldots, u_n \in \mathcal{O}_{X,x}$ form a regular system of parameters, where $x \in X$ is a closed point. We may assume that $u_1, \ldots, u_n$ are defined in an affine open neighborhood U of x, and that $du = du_1 \wedge \cdots \wedge du_n$ gives a basis of ω_X on U. Furthermore, we may assume that $\mathcal{O}_U$ is free over $\mathcal{O}_U^p$, with basis

$$\{u_1^{i_1} \cdots u_n^{i_n} \mid 0 \le i_j \le p-1 \text{ for } 1 \le j \le n\}$$

(note that the residue field of $\mathcal{O}_{X,x}$ is a finite extension of L, hence it is perfect). In this case t_X is characterized by the fact that $t_X(h^p w) = h \cdot t_X(w)$ for every $h \in \mathcal{O}_X(U)$, and on the above basis over $\mathcal{O}_X(U)^p$ it is described by

$$t_X(u_1^{i_1} \cdots u_n^{i_n} du) = \begin{cases} du & \text{if } i_j = p-1 \text{ for all } j, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Iterating e times t_X gives $t_X^e\colon F_*^e\omega_X \to \omega_X$. These maps are functorial in the following sense. If $\pi : Y \to X$ is a proper birational morphism between irreducible smooth varieties as above, then we have a commutative diagram

$$\begin{array}{ccc} \pi^*(F^e_*(\omega_X)) & \xrightarrow{\pi^*(t^e_X)} & \pi^*(\omega_X) \\ \big\downarrow & & \big\downarrow{\scriptstyle\psi} \\ F^e_*(\omega_Y) & \xrightarrow{t^e_Y} & \omega_Y, \end{array} \tag{5}$$

where ψ is the canonical morphism induced by pulling-back n-forms, and the left vertical map is the composition

$$\pi^*(F^e_*(\omega_X)) \longrightarrow F^e_*(\pi^*(\omega_X)) \xrightarrow{F^e_*(\psi)} F^e_*(\omega_Y).$$

Suppose now that X is as above, and $\mathfrak{b}$ is an ideal on X. Since ω_X is a line bundle, it follows that the image of $F^e_*(\mathfrak{b}\cdot\omega_X)$ by t^e_X can be written as $J\cdot\omega_X$ for a unique ideal J on X. It is an easy consequence of the description of $\mathfrak{b}^{[1/p^e]}$ in [Blickle et al. 2008, Proposition 2.5] and of formula (4) that in fact $J=\mathfrak{b}^{[1/p^e]}$; see also [Blickle et al. 2010, Proposition 3.10].

3. A review of nonstandard constructions

We begin by reviewing some general facts about ultraproducts. For a detailed introduction to this topic, the reader is referred to [Goldblatt 1998]. We then explain how geometric objects over an ultraproduct of fields correspond to sequences of such geometric objects over the fields we are starting with, up to a suitable equivalence relation. Most of this material is well-known to the experts, and can be found, for example, in [Schoutens 2005, §2]. However, we prefer to give a detailed presentation for the benefit of those readers having little or no familiarity with nonstandard constructions.

3A. ***Ultrafilters and ultraproducts.*** Recall that an *ultrafilter* on the set of positive integers $\mathbb{N}$ is a nonempty collection $\mathcal{U}$ of subsets of $\mathbb{N}$ that satisfies the following properties:

(i) If A and B lie in $\mathcal{U}$, then $A\cap B$ lies in $\mathcal{U}$.

(ii) If $A\subseteq B$ and A is in $\mathcal{U}$, then B is in $\mathcal{U}$.

(iii) The empty set does not belong to $\mathcal{U}$.

(iv) Given any $A\subseteq\mathbb{N}$, either A or $\mathbb{N}\smallsetminus A$ lies in $\mathcal{U}$.

An ultrafilter $\mathcal{U}$ is *nonprincipal* if no finite subsets of $\mathbb{N}$ lie in $\mathcal{U}$. It is an easy consequence of Zorn's Lemma that nonprincipal ultrafilters exist, and we fix one such ultrafilter $\mathcal{U}$. Given a property $\mathcal{P}(m)$, where $m\in\mathbb{N}$, we say that $\mathcal{P}(m)$ holds *for almost all m* if $\{m\in\mathbb{N}\mid\mathcal{P}(m)\text{ holds}\}$ lies in $\mathcal{U}$.

Given a sequence of sets $(A_m)_{m\in\mathbb{N}}$, the *ultraproduct* $[A_m]$ is the quotient of $\prod_{m\in\mathbb{N}} A_m$ by the equivalence relation given by $(a_m) \sim (b_m)$ if $a_m = b_m$ for almost all m. We write the class of (a_m) in $[A_m]$ by $[a_m]$. Note that the element $[a_m]$ is well-defined even if a_m is only defined for almost all m. Similarly, the set $[A_m]$ is well-defined if we give A_m for almost all m.

If $A_m = A$ for all m, then one writes ${}^*A$ instead of $[A_m]$. This is the *nonstandard extension* of A. Note that there is an obvious inclusion $A \hookrightarrow {}^*A$ that takes $a \in A$ to the class of the constant sequence (a).

The general principle is that if all A_m have a certain algebraic structure, then so does $[A_m]$, by defining the corresponding structure component-wise on $\prod_{m\in\mathbb{N}} A_m$. For example, if we consider fields $(L_m)_{m\in\mathbb{N}}$, then $k := [L_m]$ is a field. In particular, the nonstandard extension ${}^*\mathbb{R}$ of $\mathbb{R}$ is an ordered field. Furthermore, it is easy to see that if all L_m are algebraically closed, then so is k. Note also that if $\lim_{m\to\infty} \operatorname{char}(L_m) = \infty$, then $\operatorname{char}(k) = 0$.

Given a sequence of maps $f_m \colon A_m \to B_m$ for $m \in \mathbb{N}$, we get a map $[f_m] : [A_m] \to [B_m]$ that takes $[a_m]$ to $[f_m(a_m)]$. In particular, given a map $f \colon A \to B$, we get a map ${}^*f : {}^*A \to {}^*B$ that extends f. If each A_m is a subset of B_m, we can identify $[A_m]$ to a subset of $[B_m]$ via the corresponding map. The subsets of $[B_m]$ of this form are called *internal*.

We will use in Section 4 the following notion. Suppose that $u = [u_m] \in {}^*\mathbb{R}$ is bounded (this means that there is $M \in \mathbb{R}_{>0}$ such that ${}^*|u| \le M$, that is, $|u_m| \le M$ for almost all m). In this case, there is a unique real number, the *shadow* $\operatorname{sh}(u)$ of u, with the property that for every positive real number ε, we have ${}^*|u - \operatorname{sh}(u)| < \varepsilon$, that is, $|\operatorname{sh}(u) - u_m| < \varepsilon$ for almost all m. We refer to [Goldblatt 1998, §5.6] for a discussion of shadows. A useful property is that if $(c_m)_{m\in\mathbb{N}}$ is a convergent sequence, with $\lim_{m\to\infty} c_m = c$, then $\operatorname{sh}([c_m]) = c$; see [ibid., Theorem 6.1]. On the other hand, it is a consequence of the definition that $\operatorname{sh}([c_m])$ is the limit of a suitable subsequence of $(c_m)_{m\in\mathbb{N}}$.

3B. ***Schemes, morphisms, and sheaves over an ultraproduct of fields.*** Suppose that $\mathcal{U}$ is a nonprincipal ultrafilter on $\mathbb{N}$ as in the previous section, and suppose that $(L_m)_{m\in\mathbb{N}}$ is a sequence of fields. We denote the corresponding ultraproduct by $k = [L_m]$. Let us temporarily fix $n \ge 1$, and consider the polynomial rings $R_m = L_m[x_1, \dots, x_n]$. We write $k[x_1, \dots, x_n]_{\mathrm{int}}$ for the ring $[R_m]$, the ring of *internal polynomials* in n variables (we emphasize, however, that the elements of this ring are *not* polynomials). Given a sequence of ideals $(\mathfrak{a}_m \subseteq R_m)_{m\in\mathbb{N}}$, we get the *internal ideal* $[\mathfrak{a}_m]$ in $k[x_1, \dots, x_n]_{\mathrm{int}}$.

We have an embedding $k[x_1, \dots, x_n] \hookrightarrow k[x_1, \dots, x_n]_{\mathrm{int}}$. Its image consists of those $g = [g_m] \in k[x_1, \dots, x_n]_{\mathrm{int}}$ for which there is $d \in \mathbb{N}$ such that $\deg(g_m) \le d$ for almost all m (in this case we say that g *has bounded degree*). We say that an ideal $\mathfrak{b} \subseteq k[x_1, \dots, x_n]_{\mathrm{int}}$ is generated in bounded degree if it is generated by an ideal in $k[x_1, \dots, x_n]$ (in which case $\mathfrak{b}$ is automatically an internal ideal). Given an ideal $\mathfrak{a}$ in $k[x_1, \dots, x_n]$, we put $\mathfrak{a}_{\mathrm{int}} := \mathfrak{a} \cdot k[x_1, \dots, x_n]_{\mathrm{int}}$.

The connection between $k[x_1, \dots, x_n]$ and $k[x_1, \dots, x_n]_{\mathrm{int}}$ is studied in [van den Dries and Schmidt 1984]. In particular:

Theorem 3.1 [van den Dries and Schmidt 1984, Theorem 1.1]. *The extension $k[x_1, \dots, x_n] \hookrightarrow k[x_1, \dots, x_n]_{\mathrm{int}}$ is faithfully flat. In particular, given any ideal $\mathfrak{a}$ in $k[x_1, \dots, x_n]$, we have $\mathfrak{a}_{\mathrm{int}} \cap k[x_1, \dots, x_n] = \mathfrak{a}$.*

It follows from the theorem that ideals of $k[x_1, \dots, x_n]_{\mathrm{int}}$ generated in bounded degree are in order-preserving bijection with the ideals in $k[x_1, \dots, x_n]$. Furthermore, note that every such ideal of $k[x_1, \dots, x_n]_{\mathrm{int}}$ is of the form $[\mathfrak{a}_m]$ for a sequence $(\mathfrak{a}_m)_{m \in \mathbb{N}}$ that is *generated in bounded degree*, that is, such that for some d, $\mathfrak{a}_m \subseteq L_m[x_1, \dots, x_n]$ is generated by polynomials of degree less than or equal to d for almost all m. Of course, we have $[\mathfrak{a}_m] = [\mathfrak{b}_m]$ if and only $\mathfrak{a}_m = \mathfrak{b}_m$ for almost all m. Given such a sequence $(\mathfrak{a}_m)_{m \in \mathbb{N}}$, we call $[\mathfrak{a}_m] \cap k[x_1, \dots, x_n]$ the *ideal of polynomials* corresponding to the sequence.

Our next goal is to describe how to associate to a geometric object over k a sequence of corresponding objects over each of L_m (in fact, an equivalence class of such sequences). Given a separated scheme X of finite type over k, we will associate to it an *internal scheme* $[X_m]$, by which we mean the following: we have schemes X_m of finite type over L_m for almost all m; furthermore, two such symbols $[X_m]$ and $[Y_m]$ define the same equivalence class if $X_m = Y_m$ for almost all m. An *internal morphism* $[f_m] \colon [X_m] \to [Y_m]$ between internal schemes consists of an equivalence class of sequences of morphisms of schemes $f_m \colon X_m \to Y_m$ (defined for almost all m), where $[f_m] = [g_m]$ if $f_m = g_m$ for almost all m.

We want to define a functor $X \to X_{\mathrm{int}}$ from separated schemes of finite type over k to internal schemes. We first consider the case when X is affine. In this case let us choose a closed embedding $X \hookrightarrow \mathbb{A}_k^N$, defined by the ideal $\mathfrak{a} \subseteq k[x_1, \dots, x_N]$. If $\mathfrak{a}_{\mathrm{int}} = [\mathfrak{a}_m]$, then we take X_m to be defined in $\mathbb{A}_{L_m}^N$ by $\mathfrak{a}_m$. Note that $X_{\mathrm{int}} := [X_m]$ is well defined. We also put $\mathbb{O}(X)_{\mathrm{int}} := [L_m[x_1, \dots, x_N]/\mathfrak{a}_m]$, and note that we have a canonical ring homomorphism $\eta_X \colon \mathbb{O}(X) \to \mathbb{O}(X)_{\mathrm{int}}$. Suppose now that we have a morphism $f \colon Y \to X$ of affine schemes as above, and closed embeddings $Y \hookrightarrow \mathbb{A}_k^N$ and $X \hookrightarrow \mathbb{A}_k^M$. We have a homomorphism $\varphi \colon k[x_1, \dots, x_M] \to k[x_1, \dots, x_N]$ that induces f, and that extends to an internal morphism $k[x_1, \dots, x_M]_{\mathrm{int}} \to k[x_1, \dots, x_N]_{\mathrm{int}}$. This induces morphisms $f_m \colon Y_m \to X_m$ for almost all m, hence an internal morphism $Y_{\mathrm{int}} \to X_{\mathrm{int}}$. It is easy to see that this is independent of the choice of the lifting φ and that it is functorial.

The first consequence is that if we replace $X \hookrightarrow \mathbb{A}_k^N$ by a different embedding $X \hookrightarrow \mathbb{A}_k^M$, then the two internal schemes that we obtain are canonically isomorphic.

We use this to extend the above definition to the case when X is not necessarily affine, as follows. Note first that if $\overline{L_m}$ is an algebraic closure of L_m, and if $\bar{k} = [\overline{L_m}]$, then $\bar{k}$ is an algebraically closed field containing k, and for every affine X as above, with $X_{\text{int}} = [X_m]$, we have a natural bijection of sets $X(\bar{k}) \simeq [X_m(\overline{L_m})]$.

Lemma 3.2. *Let X be an affine scheme as above, $U \subset X$ an affine open subset, and write $X_{\text{int}} = [X_m]$ and $U_{\text{int}} = [U_m]$.*

(i) *The induced maps $U_m \to X_m$ are open immersions for almost all m.*

(ii) *If $X = U_1 \cup \cdots \cup U_r$ is an affine open cover, and $(U_i)_{\text{int}} = [U_{i,m}]$ for every i, then $X_m = U_{1,m} \cup \cdots \cup U_{r,m}$ for almost all m.*

Proof. The first assertion is clear in the case when U is a principal affine open subset corresponding to $f \in \mathcal{O}(X)$: if the image of f in $\mathcal{O}(X)_{\text{int}}$ is $[f_m]$, then for almost all m we have that U_m is the principal affine open subset of X_m corresponding to U_m. The assertion in (ii) is clear, too, when all U_i are principal affine open subsets in X: once we know that the $U_{i,m}$ are open in X_m, to get the assertion we want it is enough to look at the $\bar{k}$-valued points of X.

We now obtain the assertion in (i) in general, since we may cover U by finitely many principal affine open subsets in X (hence also in U). We then deduce (ii) in general from (i) by considering the $\bar{k}$-valued points of X. □

Given any scheme X, separated and of finite type over k, consider an affine open cover $X = U_1 \cup \cdots \cup U_r$, and let $(U_i)_{\text{int}} = [U_{i,m}]$. The intersection $U_i \cap U_j$ is affine and open in both U_i and U_j, hence by Lemma 3.2, $U_{i,m} \cap U_{j,m}$ is affine and open in both $U_{i,m}$ and $U_{j,m}$ for almost all m. We get X_m by gluing, for all i and j, the open subsets $U_{i,m}$ and $U_{j,m}$ along $(U_i \cap U_j)_m$, and put $X_{\text{int}} = [X_m]$. It is straightforward to check that X_{int} is independent of the choice of cover (up to a canonical isomorphism). Similarly, given a morphism of schemes $f : Y \to X$ we get an internal morphism $f_{\text{int}} = [f_m]$: $Y_{\text{int}} \to X_{\text{int}}$ by gluing the internal morphisms obtained by restricting f to suitable affine open subsets. Therefore we have a functor from the category of separated schemes of finite type over k to the category of internal schemes and internal morphisms. This has the property that given L_m-algebras A_m for almost all m, if $A = [A_m]$, then we have a natural bijection of sets

$$\operatorname{Hom}(\operatorname{Spec} A, X) \simeq [\operatorname{Hom}(\operatorname{Spec} A_m, X_m)], \tag{6}$$

where $X_{\text{int}} = [X_m]$. In particular, we have a bijection $X(\bar{k}) \simeq [X_m(\overline{L_m})]$.

We do not attempt to give a comprehensive account of the properties of this construction, but list in the following proposition a few that we will need.

Proposition 3.3. *Let X and Y be separated schemes of finite type over k, and $X_{\text{int}} = [X_m]$ and $Y_{\text{int}} = [Y_m]$ the corresponding internal schemes.*

(i) *For every affine open subset U of X, the ring homomorphism*

$$\eta_U : \mathcal{O}_X(U) \to \mathcal{O}_X(U)_{\mathrm{int}}$$

is faithfully flat.

(ii) *X is reduced or integral if and only if X_m has the same property for almost all m.*

(iii) *The internal scheme corresponding to $X \times Y$ is $[X_m \times Y_m]$.*

(iv) *If $f : Y \to X$ is an open or closed immersion, then the induced morphisms $f_m : Y_m \to X_m$ have the same property for almost all m. In particular, X_m is separated for almost all m.*

(v) *If $X^{(1)}, \ldots, X^{(r)}$ are the irreducible components of X, and $X^{(i)}_{\mathrm{int}} = [X^{(i)}_m]$, then $X^{(1)}_m, \ldots, X^{(r)}_m$ are the irreducible components of X_m for almost all m.*

(vi) *If X is affine and $f : Y \to X$ is a projective morphism, then $f_m : Y_m \to X_m$ is projective for almost all m.*

(vii) *If $\dim(X) = d$, then $\dim(X_m) = d$ for almost all m.*

Proof. The assertion in (i) follows from definition and Theorem 3.1. The assertions in (ii) follow from definition and the fact that if $\mathfrak{a}$ is an ideal in $k[x_1, \ldots, x_N]$, then $\mathfrak{a}$ is prime or radical if and only if $\mathfrak{a} \cdot k[x_1, \ldots, x_n]_{\mathrm{int}}$ has the same property; see [van den Dries and Schmidt 1984, Theorem 2.5, Corollary 2.7]. Properties (iii) and (iv) are easy consequences of the definition (note that we have already checked the assertion regarding open immersions when both X and Y are affine). The second assertion in (iv) follows from the fact that the diagonal map $X \to X \times X$ being a closed immersion implies that $X_m \to X_m \times X_m$ is a closed immersion for almost all m. We obtain (v) from (ii), (iv), and the fact that $X^{(1)}_m, \ldots, X^{(r)}_m$ cover X_m for almost all m. This follows by computing the $\bar{k}$-points of X, and using (6).

In order to prove (vi), note that $(\mathbb{P}^N_k)_{\mathrm{int}} \simeq [\mathbb{P}^N_{L_m}]$. Therefore a closed embedding $\iota : Y \hookrightarrow X \times \mathbb{P}^N_k$ induces by (iii) and (iv) closed embeddings $\iota_m : Y_m \hookrightarrow X_m \times \mathbb{P}^N_{L_m}$ for almost all m.

We prove (vii) by induction on $\dim(X)$. Using (v), we reduce to the case when X is irreducible. After replacing X by X_{red}, we see that we may assume, in fact, that X is integral, hence by (ii), for almost all m we have X_m integral. It is enough to prove the assertion for an affine open subset U of X, hence we may assume that $X = \operatorname{Spec} A$ is affine, and let us write $X_m = \operatorname{Spec} A_m$. If $f \in \mathcal{O}(X)$ is nonzero, then $\dim(A/(f)) = \dim(A) - 1$. Let $[f_m] = \eta_X(f) \in \mathcal{O}(X)_{\mathrm{int}}$, hence for almost all m we have $f_m \neq 0$ and $\dim(A_m/(f_m)) = \dim(A_m) - 1$. Since the internal scheme corresponding to $\operatorname{Spec} A/(f)$ is $[\operatorname{Spec} A_m/(f_m)]$, we conclude by induction. $\square$

Remark 3.4. We emphasize that to an arbitrary internal scheme $[X_m]$ we do *not* associate a scheme over k. In order to illustrate the problems that arise when trying to do this, consider the following two examples.

1) Let $X_m = \mathrm{Spec}(R_m)$, where $R_m = L_m[x_1, \ldots, x_m]$.
2) Let Y_m be the closed subscheme of $\mathrm{Spec}(L_m[x, y])$ defined by (f_m), where $f_m = x^2 + y^m$.

The only reasonable schemes to associate to $[X_m]$ and $[Y_m]$ are $X = \mathrm{Spec}([R_m])$ and $Y = \mathrm{Spec}(k[x, y]_{\mathrm{int}}/(f))$, respectively, where $f = [f_m] = x^2 + y^{\omega}$, with ω being the nonstandard integer corresponding to $(1, 2, 3, \ldots)$. In this case, the *internal* k-valued points of X and Y are in natural bijection with $[X_m(L_m)]$ and $[Y_m(L_m)]$, respectively. However, since both X and Y are far from being of finite type over k, we will not further consider such general constructions.

Suppose now that X is a scheme over k as above, and $\mathcal{F}$ is a coherent sheaf on X. If $X_{\mathrm{int}} = [X_m]$, we define an internal coherent sheaf on $[X_m]$ to be a symbol $[\mathcal{F}_m]$, where $\mathcal{F}_m$ is defined for almost all m and is a coherent sheaf of X_m. Furthermore, two such symbols $[\mathcal{F}_m]$ and $[\mathcal{F}'_m]$ are identified precisely when $\mathcal{F}_m = \mathcal{F}'_m$ for almost all m. A morphism of internal coherent sheaves is defined in a similar way, and we get an abelian category consisting of internal coherent sheaves on $[X_m]$.

We now define a functor $\mathcal{F} \to \mathcal{F}_{\mathrm{int}}$ from the category of coherent sheaves on X to that of internal coherent sheaves on X_{int}. Given an affine open subset U of X and the corresponding internal scheme $U_{\mathrm{int}} = [U_m]$, we consider the $\mathcal{O}_X(U)_{\mathrm{int}}$-module $\mathcal{T}_U := \mathcal{F}(U) \otimes_{\mathcal{O}_X(U)} \mathcal{O}_X(U)_{\mathrm{int}}$. We claim that this is equal to $[M_m]$ for suitable $\mathcal{O}_{X_m}(U_m)$-modules M_m. Indeed, this follows by considering a finite free presentation

$$\mathcal{O}_X(U)^{\oplus r} \xrightarrow{\varphi} \mathcal{O}_X(U)^{\oplus s} \to \mathcal{F}(U) \to 0.$$

If φ is defined by a matrix $(a_{i,j})_{i,j}$ and if we write $\eta_U(a_{i,j}) = [a_{i,j,m}]$, then we may take each M_m to be the cokernel of the map $\mathcal{O}_{X_m}(U_m)^{\oplus r} \to \mathcal{O}_{X_m}(U_m)^{\oplus s}$ defined by the matrix $(a_{i,j,m})_{i,j}$. We put $\mathcal{F}_m(U) = M_m$ for almost all m. It is now easy to see that the $\mathcal{F}_m(U)$ glue together for almost all m to give coherent sheaves $\mathcal{F}_m$ on X_m. Therefore we get an internal coherent sheaf $\mathcal{F}_{\mathrm{int}}$ on X_{int}. Given a morphism of coherent sheaves on X, we clearly get a corresponding morphism of internal coherent sheaves. It follows from definition and Proposition 3.3 (i) that this functor is exact in a strong sense: a bounded complex of coherent sheaves on X is acyclic if and only if the corresponding complexes of coherent sheaves on X_m are acyclic for almost all m. Note also that the functor is compatible with tensor product: if $\mathcal{F}_{\mathrm{int}} = [\mathcal{F}_m]$ and $\mathcal{G}_{\mathrm{int}} = [\mathcal{G}_m]$, then $(\mathcal{F} \otimes_{\mathcal{O}_X} \mathcal{G})_{\mathrm{int}}$ is canonically isomorphic to $[\mathcal{F}_m \otimes_{\mathcal{O}_{X_m}} \mathcal{G}_m]$. We collect in the following proposition a few other properties of this functor that we will need.

Proposition 3.5. *Let X be a separated scheme of finite type over k, and $\mathcal{F}$ a coherent sheaf on X. Consider $X_{\mathrm{int}} = [X_m]$ and $\mathcal{F}_{\mathrm{int}} = [\mathcal{F}_m]$.*

(i) *$\mathcal{F}$ is locally free of rank r if and only if $\mathcal{F}_m$ has the same property for almost all m.*

(ii) *If $\mathcal{F}$ is an ideal in $\mathcal{O}_X$ defining the closed subscheme Z of X, and $Z_{\mathrm{int}} = [Z_m]$, then $\mathcal{F}_m$ is (isomorphic to) the ideal defining Z_m in X_m for almost all m.*

(iii) *If $f\colon Y \to X$ is a morphism of schemes as above, and $f_{\mathrm{int}} = [f_m]$, then we have a canonical isomorphism $f^*(\mathcal{F})_{\mathrm{int}} \simeq [f_m^*(\mathcal{F}_m)]$.*

(iv) *If $g\colon Y \to X$ is a projective morphism of schemes as above, and $g_{\mathrm{int}} = [g_m]\colon [Y_m] \to [X_m]$, then for every $i \geq 0$ we have a canonical isomorphism*

$$R^i f_*(\mathcal{F})_{\mathrm{int}} \simeq [R^i(f_m)_*(\mathcal{F}_m)].$$

(v) *If f is as in* (iv), *X is affine, and $\mathcal{F}$ is a line bundle on X that is* (*very*) *ample over X, then $\mathcal{F}_m$ is* (*very*) *ample over X_m for almost all m.*

Proof. The first assertion follows from Proposition 3.3 (i) and the fact that given a faithfully flat ring homomorphism $A \to B$, a finitely generated A-module M is locally free of rank r if and only if the B-module $M \otimes_A B$ is locally free of rank r. Assertion (ii) is an immediate consequence of the definitions. In order to prove (iii) it is enough to consider the case when both X and Y are affine. In this case the assertion follows from the natural isomorphism $[M_m] \otimes_{[A_m]} [B_m] \simeq [M_m \otimes_{A_m} B_m]$ whenever $A_m \to B_m$ are ring homomorphisms and the M_m are finitely generated A_m-modules.

Let us now prove (iv). Suppose first that X is affine. The first step is to construct canonical morphisms

$$H^i(Y, \mathcal{F})_{\mathrm{int}} \to [H^i(Y_m, \mathcal{F}_m)]. \tag{7}$$

This can be done by computing the cohomology as Čech cohomology with respect to a finite affine open cover of Y, and the corresponding affine open covers of Y_m (and by checking that the definition is independent of the cover). It is enough to prove that the maps (7) are isomorphisms: if X is not affine, then we simply glue the corresponding isomorphisms over a suitable affine open cover of X. Since Y is isomorphic to a closed subscheme of some $X \times \mathbb{P}_k^N$, it is enough to prove that the morphisms (7) are isomorphisms when $Y = \mathbb{P}_X^N$. Explicit computation of cohomology implies that (7) is an isomorphism when $\mathcal{F} = \mathcal{O}_{\mathbb{P}_X^N}(\ell)$ (note that $\mathcal{O}_{\mathbb{P}_X^N}(\ell)_{\mathrm{int}} \simeq [\mathcal{O}_{\mathbb{P}_{X_m}^N}(\ell)]$).

We now prove that (7) is an isomorphism by descending induction on i, the case $i > N$ being trivial. Given any $\mathcal{F}$, there is an exact sequence

$$0 \to \mathcal{G} \to \mathcal{O}_{\mathbb{P}_X^N}(\ell)^{\oplus r} \to \mathcal{F} \to 0,$$

for some ℓ and r. We use the induction hypothesis, the long exact sequence in cohomology and the 5-lemma to show first that (7) is surjective for all $\mathcal{F}$. Applying this for $\mathcal{G}$, we then conclude that (7) is also injective for all $\mathcal{F}$. This completes the proof of (iv). The assertion in (v) follows using (iii) and Proposition 3.3 (iv), from the fact that if $Y = \mathbb{P}^N_X$ and $\mathcal{F} = \mathcal{O}_Y(1)$, then $Y_m \simeq \mathbb{P}^N_{X_m}$ and $\mathcal{F}_m \simeq \mathcal{O}_{Y_m}(1)$ for almost all m. □

We will need the following uniform version of asymptotic Serre vanishing; see also [Schoutens 2005, Corollary 2.16].

Corollary 3.6. *Let $f\colon Y \to X$ be a projective morphism of schemes over k as above, with X affine. If $\mathcal{F}$ is a coherent sheaf on Y and $\mathcal{L}$ is a line bundle on Y that is ample over X and such that $H^i(Y, \mathcal{F} \otimes \mathcal{L}^j) = 0$ for all $i \geq 1$ and all $j \geq j_0$, then for almost all m we have $H^i(Y_m, \mathcal{F}_m \otimes \mathcal{L}_m^j) = 0$ for all $i \geq 1$ and $j \geq j_0$, where $Y_{\text{int}} = [Y_m]$, $\mathcal{F}_{\text{int}} = [\mathcal{F}_m]$, and $\mathcal{L}_{\text{int}} = [\mathcal{L}_m]$.*

Proof. Note first that we may assume that $\mathcal{L}$ is very ample. Indeed, if N is such that $\mathcal{L}^N$ is very ample, then we may apply the very ample case to the line bundle $\mathcal{L}^N$ and to the sheaves $\mathcal{F}, \mathcal{F} \otimes \mathcal{L}, \ldots, \mathcal{F} \otimes \mathcal{L}^{N-1}$ to obtain the assertion in the corollary. Let $r = \dim(Y)$. It follows from Proposition 3.5 (iv) that if $i \geq 1$ and $j \geq j_0$ are fixed, then $H^i(Y_m, \mathcal{F}_m \otimes \mathcal{L}_m^j) = 0$ for almost all m. In particular, for almost all m we have $H^i(Y_m, \mathcal{F}_m \otimes \mathcal{L}_m^j) = 0$ for $1 \leq i \leq \dim(Y_m) = r$ and $j_0 \leq j \leq j_0 + r - 1$. For every such m, it follows that $\mathcal{F}_m$ is $(j_0 + r)$-regular in the sense of Castelnuovo–Mumford regularity, hence $H^i(Y_m, \mathcal{F}_m \otimes \mathcal{L}_m^j) = 0$ for every $i \geq 1$ and $j \geq j_0 + r - i$; see [Lazarsfeld 2004, Chapter 1.8.A]. This completes the proof of the corollary. □

Proposition 3.7. *If X is a separated scheme of finite type over k and $X_{\text{int}} = [X_m]$, then there is a canonical isomorphism $(\Omega_{X/k})_{\text{int}} \simeq [\Omega_{X_m/L_m}]$. In particular, X is smooth of pure dimension n if and only if X_m is smooth of pure dimension n for almost all m.*

Proof. It is enough to give a canonical isomorphism $(\Omega_{X/k})_{\text{int}} = [\Omega_{X_m/L_m}]$ when X is affine. Note that we have such an isomorphism when $X = \mathbb{A}^N_k$. In general, if X is a closed subscheme of $\mathbb{A}^N_k$ defined by the ideal $\mathfrak{a}$, the sheaf $\Omega_{X/k}$ is the cokernel of a morphism $\mathfrak{a}/\mathfrak{a}^2 \to \Omega_{\mathbb{A}^N_k}|_X$. If $\mathfrak{a} = [\mathfrak{a}_m]$, then for almost all m we have an analogous description of each Ω_{X_m/L_m} in terms of the embedding $X_m \hookrightarrow \mathbb{A}^N_{L_m}$ given by $\mathfrak{a}_m$. Therefore we obtain the desired isomorphism, and one can then check that this is independent of the embedding.

Recall that X is smooth of pure dimension n if and only if $\dim(X) = n$ and $\Omega_{X/k}$ is locally free of rank n. The second assertion in the proposition now follows from the first one, together with Proposition 3.3 (vii) and Proposition 3.5 (i). □

Suppose that X is a smooth scheme over k as above, and $D = a_1 D^{(1)} + \cdots + a_r D^{(r)}$ is a divisor on X. It follows from Proposition 3.3 (ii), (vii) that if $D^{(i)}_{\text{int}} = [D^{(i)}_m]$, then $D_m^{(i)}$ is a prime divisor on X_m for almost all m. For all such m we put $D_m = a_1 D_m^{(1)} + \cdots + a_r D_m^{(r)}$.

Remark 3.8. Note that in the case when D is effective, and thus can be considered as a subscheme of X, the above convention is compatible with our previous definition via $D_{\text{int}} = [D_m]$. Indeed, if we define the D_m via the latter formula, then it follows from definition that since D is locally defined by one nonzero element, the same holds for D_m for almost all m. Furthermore, Proposition 3.3 (v) implies that $D_m^{(1)}, \ldots, D_m^{(r)}$ are the irreducible components of D_m for almost all m. We also see that the coefficient of $D_m^{(i)}$ in D_m is equal to a_i for almost all m: this follows from the fact that this coefficient is the largest nonnegative integer d_i such that $d_i D_m^{(i)}$ is a subscheme of D_m.

We thus see that for every divisor D, we have $\mathcal{O}(D)_{\text{int}} = [\mathcal{O}(D_m)]$. Indeed, when $-D$ is effective, this follows from the above remark and Proposition 3.5 (ii). The general case follows easily by reducing to the case when X is affine, and replacing D by $D + \text{div}(f)$ for a suitable $f \in \mathcal{O}(X)$ such that $-D - \text{div}(f)$ is effective.

Proposition 3.9. *Let X be a smooth, separated scheme over k, and $D = \sum_{i=1}^N D^{(i)}$ an effective divisor on X, with simple normal crossings, where the $D^{(i)}$ are distinct prime divisors. If $X_{\text{int}} = [X_m]$ and $D_{\text{int}} = [D_m]$, then D_m has simple normal crossings for almost all m.*

Proof. Note that X_m is smooth over L_m for almost all m by Proposition 3.7. Since D has simple normal crossings, for every r and every $1 \le i_1 < \cdots < i_r \le N$ the subscheme $D^{(i_1)} \cap \cdots \cap D^{(i_r)}$ is smooth over k (possibly empty). It follows from definition that we have

$$(D^{(i_1)} \cap \cdots \cap D^{(i_r)})_{\text{int}} = [D_m^{(i_1)} \cap \cdots \cap D_m^{(i_r)}],$$

hence $D_m^{(i_1)} \cap \cdots \cap D_m^{(i_r)}$ is smooth over L_m for almost all m, by another application of Proposition 3.7. Thus D_m has simple normal crossings for almost all m. $\square$

4. Limits of F-pure thresholds

The following is our main result. As we will see, it easily implies the theorem stated in Section 1.

Theorem 4.1. *Let $(L_m)_{m\in\mathbb{N}}$ be a sequence of fields of positive characteristic such that $\lim_{m\to\infty} \text{char}(L_m) = \infty$. We fix a nonprincipal ultrafilter on $\mathbb{N}$, and let $k = [L_m]$. If $\mathfrak{a}_m \subseteq L_m[x_1, \ldots, x_n]$ are nonzero ideals generated in bounded degree, and if $\mathfrak{a} \subseteq k[x_1, \ldots, x_n]$ is the ideal of polynomials corresponding to $(\mathfrak{a}_m)_{m\ge 1}$, then for every $\lambda \in \mathbb{R}_{\ge 0}$ we have*

$$\mathcal{J}(\mathfrak{a}^\lambda)_{\text{int}} = [\tau(\mathfrak{a}_m^\lambda)].$$

Corollary 4.2. *If $(\mathfrak{a}_m)_{m\in\mathbb{N}}$ and $\mathfrak{a}$ are as in the above theorem, and $\mathfrak{a}_m$ vanishes at the origin for almost all m, then*

$$\mathrm{lct}_0(\mathfrak{a}) = \mathrm{sh}([\mathrm{fpt}_0(\mathfrak{a}_m)]).$$

Proof. Note first that since $\mathfrak{a}_m \subseteq (x_1, \ldots, x_n)L_m[x_1, \ldots, x_n]$ for almost all m, we have $\mathfrak{a} \subseteq (x_1, \ldots, x_n)k[x_1, \ldots, x_n]$. By definition, we have

$$\mathrm{lct}_0(\mathfrak{a}) = \min\{\lambda \in \mathbb{R}_{\geq 0} \mid \mathcal{J}(\mathfrak{a}^\lambda) \subseteq (x_1, \ldots, x_n)\}.$$

Since $\mathcal{J}(\mathfrak{a}^\lambda) \subseteq (x_1, \ldots, x_n)$ if and only if $\mathcal{J}(\mathfrak{a}^\lambda)_{\mathrm{int}} \subseteq (x_1, \ldots, x_n)_{\mathrm{int}}$, it follows from Theorem 4.1 that this is the case if and only if $\tau(\mathfrak{a}_m^\lambda) \subseteq (x_1, \ldots, x_n)$ for almost all m. This is further equivalent to $\lambda \geq \mathrm{fpt}_0(\mathfrak{a}_m)$ for almost all m. We conclude that $\mathrm{lct}_0(\mathfrak{a}) \geq \mathrm{fpt}_0(\mathfrak{a}_m)$ for almost all m. In addition, for every $\varepsilon \in \mathbb{R}_{>0}$, we have $\mathcal{J}(\mathfrak{a}^{\mathrm{lct}_0(\mathfrak{a})-\varepsilon})_{\mathrm{int}} \not\subseteq (x_1, \ldots, x_n)_{\mathrm{int}}$, and using again Theorem 4.1 we deduce that $\tau(\mathfrak{a}_m^{\mathrm{lct}_0(\mathfrak{a})-\varepsilon}) \not\subseteq (x_1, \ldots, x_n)$ for almost all m. By definition, this means that $\mathrm{fpt}_0(\mathfrak{a}_m) \geq \mathrm{lct}_0(\mathfrak{a}) - \varepsilon$ for almost all m. This proves the assertion in the corollary. $\square$

The result stated in Section 1 is an easy consequence of the above corollary.

Proof of Theorem 1.1. Suppose first that we have a sequence $(c_m)_{m\in\mathbb{N}}$ with $c_m \in \mathcal{F}(p_m)_n$ for all m, and such that $\lim_{m\to\infty} p_m = \infty$ and $c = \lim_{m\to\infty} c_m$. We need to show that $c \in \mathcal{L}_n$. By Proposition 2.5, we may assume that there are algebraically closed fields L_m of characteristic p_m, and ideals $\mathfrak{a}_m \subseteq L_m[x_1, \ldots, x_n]$ vanishing at the origin, such that $c_m = \mathrm{fpt}_0(\mathfrak{a}_m)$. For every d, let $\mathfrak{a}_m^{(d)} = \mathfrak{a}_m + (x_1, \ldots, x_n)^d$. It follows from Proposition 2.3 that $|\mathrm{fpt}_0(\mathfrak{a}_m) - \mathrm{fpt}_0(\mathfrak{a}_m^{(d)})| \leq \frac{n}{d}$.

Let $\mathcal{U}$ be a nonprincipal ultrafilter on $\mathbb{N}$. We put $k = [L_m]$, and for every d, we denote by $\mathfrak{a}^{(d)} \subseteq k[x_1, \ldots, x_n]$ the ideal of polynomials associated to the sequence of ideals generated in bounded degree $(\mathfrak{a}_m^{(d)})_{m\in\mathbb{N}}$. Given any $\varepsilon \in \mathbb{R}_{>0}$, let $d \gg 0$ be such that $\frac{n}{d} < \varepsilon$. By Corollary 4.2, we have $|\mathrm{fpt}_0(\mathfrak{a}_m^{(d)}) - \mathrm{lct}_0(\mathfrak{a}^{(d)})| < \varepsilon - \frac{n}{d}$ for almost all m. Therefore $|\mathrm{fpt}_0(\mathfrak{a}_m) - \mathrm{lct}_0(\mathfrak{a}^{(d)})| < \varepsilon$ for infinitely many m. Since this holds for every $\varepsilon \in \mathbb{R}_{>0}$, we conclude that c lies in the closure of $\{\mathrm{lct}_0(\mathfrak{a}^{(d)}) \mid d \geq 1\}$. As we have mentioned in Section 2A, $\mathcal{L}_n$ is closed, hence $c \in \mathcal{L}_n$.

In order to prove the converse, let us consider $c \in \mathcal{L}_n$. Consider a sequence of prime integers $(p_m)_{m\in\mathbb{N}}$ with limit infinity, and let L_m be an algebraically closed field of characteristic p_m. We fix, as above, a nonprincipal ultrafilter on $\mathbb{N}$, and let $k = [L_m]$. As pointed out in Section 2A, since k is algebraically closed, we can find an ideal $\mathfrak{b} \subset k[x_1, \ldots, x_n]$ vanishing at the origin, such that $c = \mathrm{lct}_0(\mathfrak{b})$. Let us write $\mathfrak{b}_{\mathrm{int}} = [\mathfrak{b}_m]$. It follows from Corollary 4.2 that c is the limit of a suitable subsequence of $(\mathrm{fpt}_0(\mathfrak{b}_m))_{m\in\mathbb{N}}$. This completes the proof of the theorem. Note that the second implication also follows from the results of [Hara and Yoshida 2003] discussed in the introduction. $\square$

Before giving the proof of Theorem 4.1, we describe the approach from [Hara and Yoshida 2003] for proving the equality of multiplier ideals with test ideals in a fixed positive characteristic. The main ingredients are due independently to Hara [1998] and Mehta and Srinivas [1997]. We simplify somewhat the approach in [Hara and Yoshida 2003], avoiding the use of local cohomology, which is important in our nonlocal setting.

Suppose that L is a perfect field of positive characteristic p, and W is a smooth, irreducible, n-dimensional affine scheme over L. We consider a nonzero ideal $\mathfrak{b}$ on W, and suppose that we have given a log resolution $\pi : \widetilde{W} \to W$ of $\mathfrak{b}$. Let Z be the effective divisor on $\widetilde{W}$ such that $\widetilde{\mathfrak{b}} := \mathfrak{b} \cdot \mathcal{O}_{\widetilde{W}} = \mathcal{O}_{\widetilde{W}}(-Z)$, and let $E = E_1 + \cdots + E_N$ be a simple normal crossings divisor on $\widetilde{W}$ such that both $K_{\widetilde{W}/W}$ and Z are supported on E. For every $\lambda \geq 0$, we put $\mathcal{J}(\mathfrak{b}^\lambda) = \pi_* \mathcal{O}_{\widetilde{W}}(K_{\widetilde{W}/W} - \lfloor \lambda Z \rfloor)$ (it is irrelevant for us whether this is independent of the given resolution). In this setting, it is shown in [Hara and Yoshida 2003] that the test ideals are always contained in the multiplier ideals.

Proposition 4.3. *With the above notation, we have* $\tau(\mathfrak{b}^\lambda) \subseteq \mathcal{J}(\mathfrak{b}^\lambda)$ *for all* $\lambda \in \mathbb{R}_{\geq 0}$.

Proof. We give a proof using the description of test ideals at the end of Section 2, since the approach will be relevant also when considering the reverse inclusion. We show that

$$(\mathfrak{b}^m)^{[1/p^e]} \subseteq \mathcal{J}(\mathfrak{b}^{m/p^e}) \tag{8}$$

for every $m \geq 0$ and $e \geq 1$. This is enough: given $\lambda \in \mathbb{R}_{\geq 0}$, we have for $e \gg 0$

$$\tau(\mathfrak{b}^\lambda) = (\mathfrak{b}^{\lceil \lambda p^e \rceil})^{[1/p^e]} \subseteq \mathcal{J}(\mathfrak{b}^{\lceil \lambda p^e \rceil / p^e}) = \mathcal{J}(\mathfrak{b}^\lambda).$$

Note that the last equality follows from the fact that $0 \leq (\lceil \lambda p^e \rceil / p^e) - \lambda \ll 1$ for $e \gg 0$.

The commutative diagram (5) induces a commutative diagram

$$\begin{tikzcd} F^e_*(\omega_W) \arrow[r, "t^e_W"] \arrow[d, "\eta = F^e_*(\rho)"'] & \omega_W \arrow[d, "\rho"] \\ F^e_* \pi_*(\omega_{\widetilde{W}}) \arrow[r, "\pi_*(t^e_{\widetilde{W}})"] & \pi_*(\omega_{\widetilde{W}}), \end{tikzcd} \tag{9}$$

where the vertical maps are isomorphisms. Note that $t^e_{\widetilde{W}}$ induces a (surjective) map $F^e_*(\omega_{\widetilde{W}}(-mZ)) \to \omega_{\widetilde{W}}(-\lfloor (m/p^e) Z \rfloor)$, and thus a map

$$F^e_* \pi_*(\omega_{\widetilde{W}}(-mZ)) \to \pi_*\Big(\omega_{\widetilde{W}}\Big(-\Big\lfloor \frac{m}{p^e} Z \Big\rfloor\Big)\Big).$$

Since

$$(\mathfrak{b}^m)^{[1/p^e]} \omega_W = t^e_W(F^e_*(\mathfrak{b}^m \omega_W))$$

and $\eta(F^e_*(\mathfrak{b}^m\omega_W)) \subseteq F^e_*\pi_*(\omega_{\widetilde{W}}(-mZ))$, while

$$\rho^{-1}\Big(\pi_*\Big(\omega_{\widetilde{W}}\Big(-\Big\lfloor\frac{m}{p^e}Z\Big\rfloor\Big)\Big)\Big) = \mathcal{J}\big(\mathfrak{b}^{m/p^e}\big)\omega_W,$$

we see that (8) follows from the fact that $t^e_W F^e_*(\mathfrak{b}^m\omega_W) = (\mathfrak{b}^m)^{[1/p^e]}\omega_W$ and the commutativity of (9). □

We now explain a criterion for the reverse inclusion $\mathcal{J}(\mathfrak{b}^\lambda) \subseteq \tau(\mathfrak{b}^\lambda)$ to hold. We start with the following proposition.

Proposition 4.4. *Suppose that $\widetilde{W}$ is a smooth, irreducible, n-dimensional variety over the perfect field L of positive characteristic p. If E is a simple normal crossings divisor on $\widetilde{W}$, and G is a $\mathbb{Q}$-divisor supported on E such that $-G$ is effective, then the canonical morphism*

$$\Gamma(\widetilde{W}, F^e_*(\omega_{\widetilde{W}}(\lceil p^e G\rceil))) \to \Gamma(\widetilde{W}, \omega_{\widetilde{W}}(\lceil G\rceil)) \tag{10}$$

is surjective for every $e \geq 1$, provided that the following two conditions hold:

(A) $H^i(\widetilde{W}, \Omega^{n-i}_{\widetilde{W}}(\log E)(-E + \lceil p^\ell G\rceil)) = 0$ *for all $i \geq 1$ and $\ell \geq 1$.*

(B) $H^{i+1}(\widetilde{W}, \Omega^{n-i}_{\widetilde{W}}(\log E)(-E + \lceil p^\ell G\rceil)) = 0$ *for all $i \geq 1$ and $\ell \geq 0$.*

This is applied as follows. Suppose that $\lambda \in \mathbb{R}_{\geq 0}$ is fixed, and we have a rational number $\mu > \lambda$ such that $\mathcal{J}(\mathfrak{b}^\lambda) = \mathcal{J}(\mathfrak{b}^\mu)$ (note that if $Z = \sum_i a_i E_i$, then it is enough to take μ such that $\mu - \lambda < (\lfloor\lambda a_i\rfloor + 1 - \lambda a_i)/a_i$ for all i with $a_i > 0$). Let us consider now a $\mathbb{Q}$-divisor D on $\widetilde{W}$ such that D is ample over W, and $-D$ is effective.[2] We will apply the above proposition with $G = \mu(D - Z)$. We may and will assume that $\lceil G\rceil = \lceil -\mu Z\rceil$ (again this condition only depends on μ and the coefficients of Z; since $-D$ is effective, it is always satisfied if we replace D by εD, with $0 < \varepsilon \ll 1$).

Proposition 4.5. *With the above notation, if* (10) *is surjective for every $e \geq 1$, then $\mathcal{J}(\mathfrak{b}^\lambda) \subseteq \tau(\mathfrak{b}^\lambda)$.*

Proof. We use again the commutative diagram (9). This induces a commutative diagram

$$\begin{tikzcd}
F^e_*\pi_*(\omega_{\widetilde{W}}(\lceil p^e G\rceil)) \arrow[r, "\pi_*(t^e_{\widetilde{W}})"] \arrow[d, "\eta^{-1}"'] & \pi_*(\omega_{\widetilde{W}}(\lceil G\rceil)) = \pi_*(\omega_{\widetilde{W}}(-\lfloor\mu Z\rfloor)) \arrow[d, "\rho^{-1}"] \\
F^e_*\omega_W \arrow[r, "t^e_W"] & \omega_W
\end{tikzcd} \tag{11}$$

in which the top horizontal map is surjective by assumption (recall that W is affine), and the image of the right vertical map is $\mathcal{J}(\mathfrak{b}^\mu)\omega_W$. The image of the left vertical

[2] Such a divisor always exists: if we express $\widetilde{W}$ as the blow-up of W along a suitable ideal, then we may take D to be the negative of the exceptional divisor.

map can be written as $F_*^e(J_e\omega_W)$, where $J_e = \pi_*\mathcal{O}_{\widetilde{W}}(K_{\widetilde{W}/W} + \lceil p^e G\rceil)$, and we deduce from the commutativity of (11) that

$$\mathcal{J}(\mathfrak{b}^\mu) \subseteq J_e^{[1/p^e]}.$$

By Lemma 4.6 below, there is r such that $\mathcal{J}(\mathfrak{b}^m) \subseteq \mathfrak{b}^{m-r}$ for every $m \geq r$. Since $-D$ is effective, by letting $e \gg 0$, we get

$$J_e = \pi_*\mathcal{O}_{\widetilde{W}}(K_{\widetilde{W}/W} - \lfloor p^e\mu(Z-D)\rfloor) \subseteq \mathcal{J}\big(\mathfrak{b}^{\mu p^e}\big) \subseteq \mathfrak{b}^{\lfloor \mu p^e\rfloor - r},$$

and therefore

$$\mathcal{J}(\mathfrak{b}^\lambda) = \mathcal{J}(\mathfrak{b}^\mu) \subseteq \big(\mathfrak{b}^{\lfloor \mu p^e\rfloor - r}\big)^{[1/p^e]} \subseteq \tau\Big(\mathfrak{b}^{\frac{\lfloor \mu p^e\rfloor - r}{p^e}}\Big) \subseteq \tau(\mathfrak{b}^\lambda),$$

since $\lim_{e\to\infty}(\lfloor \mu p^e\rfloor - r)/p^e = \mu > \lambda$. This completes the proof. □

Lemma 4.6. *With the above notation, there is r such that $\mathcal{J}(\mathfrak{b}^m) \subseteq \mathfrak{b}^{m-r}$ for every integer $m \geq r$.*

Proof. It is enough to prove, more generally, that for every coherent sheaf $\mathcal{F}$ on $\widetilde{W}$, the graded module $M := \oplus_{m\geq 0}\Gamma(\widetilde{W}, \mathcal{F}(-mZ))$ is finitely generated over the Rees algebra $S := \oplus_{m\geq 0}\mathfrak{b}^m$. We may factor π as

$$\widetilde{W} \xrightarrow{g} B \xrightarrow{f} W,$$

where B is the normalized blow-up of W along $\mathfrak{b}$ (that is, $B = \mathrm{Proj}(S')$, where S' is the normalization of S). The line bundle $\mathfrak{b}\cdot\mathcal{O}_B = \mathcal{O}_B(-T)$ is ample over W, and using the projection formula we see that $M = \oplus_{m\geq 0}\Gamma(B, \pi_*(\mathcal{F})\otimes\mathcal{O}_B(-mT))$ is finitely generated over $S' = \oplus_{m\geq 0}\Gamma(B, \mathcal{O}_B(-mT))$. Since S' is a finite S-algebra, it follows that M is a finitely generated S-module. □

We recall, for completeness, the proof of Proposition 4.4, which makes use of the de Rham complex $\Omega^\bullet_{\widetilde{W}}(\log(E))$ with log poles along the simple normal crossings divisor E. Note that while this complex does not have $\mathcal{O}_{\widetilde{W}}$-linear differentials, its Frobenius push-forward $F_*\Omega^\bullet_{\widetilde{W}}(\log(E))$ does have this property. In particular, we may tensor this complex with line bundles. If $\mathcal{L}$ is a line bundle, then by the projection formula we have

$$(F_*\Omega^i_{\widetilde{W}}(\log E)) \otimes \mathcal{L} \simeq F_*(\Omega^i_{\widetilde{W}}(\log E)\otimes\mathcal{L}^p).$$

The following facts are the key ingredients in the proof of Proposition 4.4.

(1) The Cartier isomorphism: There is a canonical isomorphism (see [Deligne and Illusie 1987, Theorem 1.2])

$$C^{-1}\colon \Omega^i_{\widetilde{W}}(\log E) \simeq \mathcal{H}^i F_*(\Omega^\bullet_{\widetilde{W}}(\log E)).$$

(2) Insensitivity to small effective twists: Suppose that B is an effective divisor supported on E, with all coefficients less than p. We have a twisted de Rham complex with log poles $\Omega^{\bullet}_{\widetilde{W}}(\log E)(B)$ (it is enough to check that the differential of the de Rham complex of meromorphic differential forms on X preserves these subsheaves). In this case, the natural inclusion

$$\Omega^{\bullet}_{\widetilde{W}}(\log E) \hookrightarrow \Omega^{\bullet}_{\widetilde{W}}(\log E)(B)$$

is a quasiisomorphism; see [Hara 1998, Lemma 3.3; Mehta and Srinivas 1997, Corollary 4.2] for a proof. Combining this with the Cartier isomorphism, we find

$$\Omega^{i}_{\widetilde{W}}(\log E) \simeq \mathcal{H}^{i}(F_*(\Omega^{\bullet}_{\widetilde{W}}(\log E)(B))). \tag{12}$$

Proof of Proposition 4.4. Note first that it is enough to prove the case $e = 1$. Indeed, if $\alpha_{G,e}$ is the morphism (10), we see that $\alpha_{G,e} = \alpha_{G,1} \circ \alpha_{pG,1} \circ \cdots \circ \alpha_{p^{e-1}G,1}$, and the hypothesis implies that we may apply the condition for $e = 1$ to each of $G, pG, \ldots, p^{e-1}G$. Therefore from now on we assume that $e = 1$ (and in this case we will only need condition (A) for $\ell = 1$ and condition (B) for $\ell = 0$).

Let

$$B := (p-1)E + \lceil pG \rceil - p\lceil G \rceil = (p-1)E + p\lfloor -G \rfloor - \lfloor -pG \rfloor.$$

Since $-G$ is effective, it follows from the second expression that B is effective, and its coefficients are less than p. Let $K^{\bullet} := F_*\Omega^{\bullet}_{\widetilde{W}}(\log E)(-E + \lceil pG \rceil)$. By tensoring (12) with $\mathcal{O}_{\widetilde{W}}(-E + \lceil G \rceil)$, and using the projection formula, we get

$$\Omega^{i}_{\widetilde{W}}(\log E)(-E + \lceil G \rceil) \simeq \mathcal{H}^{i}(F_*(\Omega^{\bullet}_{\widetilde{W}}(\log E)(B - pE + p\lceil G \rceil))) = \mathcal{H}^{i}(K^{\bullet}).$$

Note that the morphism $\alpha_{G,1}$ is identified to $\Gamma(\widetilde{W}, K^n) \to \Gamma(\widetilde{W}, \mathcal{H}^n(K^{\bullet}))$. It is then straightforward to show, by breaking $K^{\bullet}$ into short exact sequences, and using the corresponding long exact sequences for cohomology, that $\alpha_{G,1}$ is surjective if $H^i(\widetilde{W}, K^{n-i}) = 0$ and $H^{i+1}(\widetilde{W}, \mathcal{H}^{n-i}(K^{\bullet})) = 0$ for all $i \geq 1$. By what we have seen, these are precisely conditions (A) with $\ell = 1$ and (B) with $\ell = 0$. □

We will also make use of the following version of the Kodaira–Akizuki–Nakano vanishing theorem (in characteristic zero).

Theorem 4.7. *Let Y be a smooth, irreducible variety over a field k of characteristic zero. If Y is projective over an affine scheme X, E is a reduced simple normal crossings divisor on Y, and G is a $\mathbb{Q}$-divisor on Y such that $G - \lfloor G \rfloor$ is supported on E and G is ample over X, then*

$$H^i(Y, \Omega^j_Y(\log E)(-E + \lceil G \rceil)) = 0 \quad \textit{if } i + j > \dim(X).$$

Proof. This is proved when $\mathrm{char}(k) = p > 0$ in [Hara 1998, Corollary 3.8] under the assumption that $p > \dim(X)$ and that both Y and E admit a lifting to the second ring of Witt vectors $W_2(k)$ of k. The proof relies on an application of the results from [Deligne and Illusie 1987]. It is then standard to deduce the assertion in characteristic zero; see, for example, the proof of [ibid., Corollary 2.7]. $\square$

We can now give the proof of our main result.

Proof of Theorem 4.1. Let $p_m = \mathrm{char}(L_m)$. We have by hypothesis $\lim_{m\to\infty} p_m = \infty$, hence $\mathrm{char}(k) = 0$. In particular, there is a log resolution $\pi : Y \to X = \mathbb{A}^n_k$ of $\mathfrak{a}$. We write $\mathfrak{a} \cdot \mathcal{O}_Y = \mathcal{O}_Y(-Z)$, and let E be a simple normal crossings divisor on Y such that both Z and $K_{Y/X}$ are supported on E. Let $[\pi_m]\colon [Y_m] \to [X_m] = [\mathbb{A}^n_{L_m}]$ be the corresponding morphism of internal schemes. It follows from Proposition 3.5 (iii) that if $Z_{\mathrm{int}} = [Z_m]$, then $\mathfrak{a}_m \cdot \mathcal{O}_{Y_m} = \mathcal{O}(-Z_m)$ for almost all m. On the other hand, it is easy to deduce from Proposition 3.7 that $(K_{Y/X})_{\mathrm{int}} = [K_{Y_m/X_m}]$. If $E_{\mathrm{int}} = [E_m]$, then E_m has simple normal crossings for almost all m by Proposition 3.9, and we conclude that π_m is a log resolution of $\mathfrak{a}_m$ for almost all m. Moreover, if we use π_m to define $\mathcal{J}(\mathfrak{a}_m^{\lambda})$ on X_m, then we have $\mathcal{J}(\mathfrak{a}^{\lambda})_{\mathrm{int}} = [\mathcal{J}(\mathfrak{a}_m^{\lambda})]$ by Proposition 3.5 (iv).

For every m such that π_m gives a log resolution of $\mathfrak{a}_m$ we have $\tau(\mathfrak{a}_m^{\lambda}) \subseteq \mathcal{J}(\mathfrak{a}_m^{\lambda})$ by Proposition 4.3. We now choose a rational number $\mu > \lambda$ such that $\mathcal{J}(\mathfrak{a}^{\lambda}) = \mathcal{J}(\mathfrak{a}^{\mu})$, so that $\mathcal{J}(\mathfrak{a}_m^{\lambda}) = \mathcal{J}(\mathfrak{a}_m^{\mu})$ for almost all m. We also choose a $\mathbb{Q}$-divisor D supported on E such that $-D$ is effective, D is ample over X, and $\lceil \mu(D - Z)\rceil = \lceil -\mu Z\rceil$. We write $G = \mu(D - Z)$, and denote by D_m and respectively G_m the corresponding divisors on Y_m. It is clear that for almost all m the divisor $-D_m$ is effective, D_m is ample over X_m (see Proposition 3.3 (v)), and $\lceil G_m\rceil = \lceil -\mu Z_m\rceil$. We deduce from Propositions 4.4 and 4.5 that $\mathcal{J}(\mathfrak{a}_m^{\lambda}) \subseteq \tau(\mathfrak{a}_m^{\lambda})$ if the following conditions hold:

(A_m) $\quad H^i(Y_m, \Omega^{n-i}_{Y_m}(\log E_m)(-E_m + \lceil p_m^{\ell} G_m\rceil)) = 0$ for all $i \geq 1$ and $\ell \geq 1$.

(B_m) $\quad H^{i+1}(Y_m, \Omega^{n-i}_{Y_m}(\log E_m)(-E_m + \lceil p_m^{\ell} G_m\rceil)) = 0$ for all $i \geq 1$ and $\ell \geq 0$.

It follows that in order to complete the proof, it is enough to show that conditions (A_m) and (B_m) hold for almost all m.

Note first that by Theorem 4.7, we have $H^{i+1}(Y, \Omega^{n-i}_Y(\log E)(-E + \lceil G\rceil)) = 0$ for all $i \geq 0$. Using Proposition 3.5 (iv), we deduce that

$$H^{i+1}(Y_m, \Omega^{n-i}_{Y_m}(\log E_m)(-E_m + \lceil G_m\rceil)) = 0$$

for all $i \geq 0$ and almost all m (since these groups vanish automatically when $i \geq n$, we only need to consider finitely many such i). This takes care of the condition (B_m) for $\ell = 0$.

We now treat the remaining conditions. Let us fix a positive integer d such that dG is an integral divisor. Let $\mathcal{F}_1, \dots, \mathcal{F}_M$ denote the sheaves $\Omega^t_Y(\log E)(-E + \lceil sG\rceil)$, for integers $0 \leq s \leq d-1$ and $0 \leq t \leq n$. Since dG is ample over the affine variety

X, there is j_0 such that $H^i(Y, \mathcal{F}_t(jdG)) = 0$ for every $j \geq j_0$, every $i \geq 1$ and every $t \leq M$. If m is such that $p_m \geq (j_0+1)d$, and if for $\ell \geq 1$ we take s with $0 \leq s \leq d-1$ such that $p_m^\ell \equiv s \pmod d$, then

$$\lceil p_m^\ell G \rceil = \frac{p_m^\ell - s}{d}(dG) + \lceil sG \rceil \quad \text{and} \quad \frac{p_m^\ell - s}{d} \geq \frac{p_m - s}{d} \geq j_0.$$

We deduce from Corollary 3.6 that the vanishings in (A_m) and (B_m) hold when $\ell \geq 1$ for almost all m (note that for such m we may assume that $p_m \geq (j_0+1)d$). This completes the proof of the theorem. □

Acknowledgments

Our project started during the "Mathematics Research Community" on Commutative Algebra in Snowbird, Utah, 2010. We are grateful to the AMS for setting up this program, and for providing a stimulating research environment.

References

[Blickle et al. 2008] M. Blickle, M. Mustaţă, and K. E. Smith, "Discreteness and rationality of F-thresholds", *Michigan Math. J.* **57** (2008), 43–61. MR 2010c:13003 Zbl 1177.13013

[Blickle et al. 2009] M. Blickle, M. Mustaţă, and K. E. Smith, "F-thresholds of hypersurfaces", *Trans. Amer. Math. Soc.* **361**:12 (2009), 6549–6565. MR 2011a:13006 Zbl 1193.13003

[Blickle et al. 2010] M. Blickle, K. Schwede, S. Takagi, and W. Zhang, "Discreteness and rationality of F-jumping numbers on singular varieties", *Math. Ann.* **347**:4 (2010), 917–949. MR 2011k:13008 Zbl 1198.13007

[Deligne and Illusie 1987] P. Deligne and L. Illusie, "Relèvements modulo p^2 et décomposition du complexe de de Rham", *Invent. Math.* **89**:2 (1987), 247–270. MR 88j:14029 Zbl 0632.14017

[van den Dries and Schmidt 1984] L. van den Dries and K. Schmidt, "Bounds in the theory of polynomial rings over fields: A nonstandard approach", *Invent. Math.* **76**:1 (1984), 77–91. MR 85i:12016 Zbl 0539.13011

[Ein and Mustaţă 2006] L. Ein and M. Mustaţă, "Invariants of singularities of pairs", pp. 583–602 in *International Congress of Mathematicians* (Madrid, 2006), vol. 2, edited by M. Sanz-Solé et al., European Mathematical Society, Zürich, 2006. MR 2007m:14050 Zbl 1096.14030

[de Fernex and Mustaţă 2009] T. de Fernex and M. Mustaţă, "Limits of log canonical thresholds", *Ann. Sci. Éc. Norm. Supér.* (4) **42**:3 (2009), 491–515. MR 2010i:14030 Zbl 1186.14007

[Goldblatt 1998] R. Goldblatt, *Lectures on the hyperreals: An introduction to nonstandard analysis*, Graduate Texts in Mathematics **188**, Springer, New York, 1998. MR 2000a:03113 Zbl 0911.03032

[Hara 1998] N. Hara, "A characterization of rational singularities in terms of injectivity of Frobenius maps", *Amer. J. Math.* **120**:5 (1998), 981–996. MR 99h:13005 Zbl 0942.13006

[Hara and Watanabe 2002] N. Hara and K.-I. Watanabe, "F-regular and F-pure rings vs. log terminal and log canonical singularities", *J. Algebraic Geom.* **11**:2 (2002), 363–392. MR 2002k:13009 Zbl 1013.13004

[Hara and Yoshida 2003] N. Hara and K.-I. Yoshida, "A generalization of tight closure and multiplier ideals", *Trans. Amer. Math. Soc.* **355**:8 (2003), 3143–3174. MR 2004i:13003 Zbl 1028.13003

[Kollár 1997] J. Kollár, "Singularities of pairs", pp. 221–287 in *Algebraic geometry, Part 1* (Santa Cruz, CA, 1995), edited by J. Kollár et al., Proc. Sympos. Pure Math. **62**, Amer. Math. Soc., Providence, RI, 1997. MR 99m:14033 Zbl 0905.14002

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, II: Positivity for vector bundles, and multiplier ideals*, Ergeb. Math. Grenzgeb. (3) **49**, Springer, Berlin, 2004. MR 2005k:14001b Zbl 1093.14500

[Mehta and Srinivas 1997] V. B. Mehta and V. Srinivas, "A characterization of rational singularities", *Asian J. Math.* **1**:2 (1997), 249–271. MR 99e:13009 Zbl 0920.13020

[Mustaţă et al. 2005] M. Mustaţă, S. Takagi, and K.-i. Watanabe, "F-thresholds and Bernstein–Sato polynomials", pp. 341–364 in *European Congress of Mathematics* (Stockholm, 2004), edited by A. Laptev, European Mathematical Society, Zürich, 2005. MR 2007b:13010 Zbl 1092.32014

[Schoutens 2005] H. Schoutens, "Log-terminal singularities and vanishing theorems via non-standard tight closure", *J. Algebraic Geom.* **14**:2 (2005), 357–390. MR 2006e:13005 Zbl 1070.14005

[Schoutens 2010] H. Schoutens, *The use of ultraproducts in commutative algebra*, Lecture Notes in Mathematics **1999**, Springer, Berlin, 2010. MR 2012g:13046 Zbl 1205.13002

[Takagi and Watanabe 2004] S. Takagi and K.-i. Watanabe, "On F-pure thresholds", *J. Algebra* **282**:1 (2004), 278–297. MR 2006a:13010 Zbl 1082.13004

Communicated by Craig Huneke
Received 2011-06-01 Revised 2011-11-16 Accepted 2011-12-20

bhargav.bhatt@gmail.com *Department of Mathematics University of Michigan, Ann Arbor, MI 48109, United States*

Current address: *School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, United States*

dhernan@umn.edu *Department of Mathematics, University of Minnesota, Minneapolis, MN 55455, United States*

lmiller@math.utah.edu *Department of Mathematics, University of Utah, Salt Lake City, UT 84112, United States*

mmustata@umich.edu *Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, United States*

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1483

The semistable reduction problem for the space of morphisms on $\mathbb{P}^n$

Alon Levy

We restate the semistable reduction theorem from geometric invariant theory in the context of spaces of morphisms from $\mathbb{P}^n$ to itself. For every complete curve C downstairs, we get a $\mathbb{P}^n$-bundle on an abstract curve D mapping finite-to-one onto C, whose trivializations correspond to not necessarily complete curves upstairs with morphisms corresponding to identifying each fiber with the morphism the point represents. Finding a trivial bundle is equivalent to finding a complete D upstairs mapping finite-to-one onto C; we prove that in every space of morphisms, there exists a curve C for which no such D exists. In the case when D exists, we bound the degree of the map from D to C in terms of C for C rational and contained in the stable space.

1. Introduction and the statement of the problem

The moduli spaces of dynamical systems on $\mathbb{P}^n$ are the spaces of morphisms, and more generally rational maps, defined by polynomials of degree d; the case we will study is $d > 1$, in which case the morphisms are not automorphisms (that is, they do not have inverses that are morphisms). For each n and d, we write each rational map φ as $(\varphi_0 : \cdots : \varphi_n)$, so that the space is parametrized by the coefficients of the monomials of each φ_i and is naturally isomorphic to a large projective space, $\mathbb{P}^N$. By an elementary computation, $N = (n+1)\binom{n+d}{d} - 1$. As we will not consider more than one of these moduli spaces at a time, there is no ambiguity in writing just N, without explicit dependence on n and d. Thus, in the remainder of this paper, N will invariably be used for the dimension of the moduli space of self-maps on $\mathbb{P}^n$ defined by polynomials of degree d.

Within the space of rational maps, the space of morphisms is an affine open subvariety, denoted Hom_d^n. The group $\mathrm{PGL}(n+1)$ acts on $\mathbb{P}^N$ by conjugation, corresponding to coordinate change, that is, A maps φ to $A\varphi A^{-1}$; this action preserves Hom_d^n, since the property of being a morphism is independent of coordinate change.

MSC2010: primary 14L24; secondary 37P45, 37P55.
Keywords: semistable reduction, moduli space, dynamical system, GIT, geometric invariant theory.

We study the quotient of the action using geometric invariant theory [Mumford and Fogarty 1982]. To do this, we need to replace $\mathrm{PGL}(n+1)$ with $\mathrm{SL}(n+1)$, which projects onto $\mathrm{PGL}(n+1)$ finite-to-one. Geometric invariant theory defines stable and semistable loci for the $\mathrm{SL}(n+1)$-action. To take the quotient, we need to remove the unstable locus, defined as the complement of the semistable locus. The quotient of Hom_d^n by $\mathrm{SL}(n+1)$ is denoted M_d^n, and parametrizes morphisms on $\mathbb{P}^n$ up to coordinate change. The stable and semistable loci for the action of $\mathrm{SL}(n+1)$ on $\mathbb{P}^N$ are denoted by $\mathrm{Hom}_d^{n,s}$ and $\mathrm{Hom}_d^{n,ss}$, and their quotients are denoted by $\mathrm{M}_d^{n,s}$ and $\mathrm{M}_d^{n,ss}$.

It is a fact that every regular map is in the stable locus. More precisely, we have the following prior results [Silverman 1998; Petsche et al. 2009; Levy 2011]:

Theorem 1.1. $\mathrm{Hom}_d^{n,s}$ *and* $\mathrm{Hom}_d^{n,ss}$ *are open subvarieties of* $\mathbb{P}^N$ *such that* $\mathrm{Hom}_d^n \subsetneq \mathrm{Hom}_d^{n,s} \subseteq \mathrm{Hom}_d^{n,ss} \subsetneq \mathbb{P}^N$. *The middle containment is an equality if and only if* $n = 1$ *and* d *is even.*

Theorem 1.2. *The stabilizer group in* $\mathrm{PGL}(n+1)$ *of each element of* Hom_d^n *is finite and bounded in terms of* d *and* n.

$\mathrm{M}_d^{n,ss}$ is a proper variety, as it is the quotient of the largest semistable subspace of $\mathbb{P}^N$ for the action of $\mathrm{SL}(n+1)$. We make the following simplifying definition.

Definition 1.3. A rational map $\varphi \in \mathbb{P}^N$ is called semistable if it is in the semistable space $\mathrm{Hom}_d^{n,ss}$.

The semistable reduction theorem states the following, answering in the affirmative a conjecture for $\mathbb{P}^1$ in [Szpiro et al. 2010]:

Theorem 1.4. *If* C *is a complete curve with* $K(C)$ *its function field, and if* $\varphi_{K(C)}$ *is a semistable rational map on* $\mathbb{P}^n_{K(C)}$, *then there exists a curve* D *mapping finite-to-one onto* C *with a* $\mathbb{P}^n$*-bundle* $\boldsymbol{P}(\mathscr{E})$ *on* D *with a self-map* Φ *such that*:

(1) *The restriction* φ_x *of* Φ *to the fiber of each* $x \in D$ *is a semistable rational self-map.*

(2) Φ *is a semistable map over* $K(D)$, *and is equivalent to* $\varphi_{K(D)}$ *under coordinate change.*

This is a classical result of geometric invariant theory; for one proof of a result that implies it, see [Zhang 1996]. We will include the proof in Section 2, along with other general facts about geometric invariant theory, including a description of the stable and semistable spaces $\mathrm{Hom}_d^{n,s}$ and $\mathrm{Hom}_d^{n,ss}$.

Theorem 1.4 leads to the natural question of which vector bundle classes can occur for each $C \subseteq \mathrm{M}_d^{n,ss}$, and more generally, for each choice of n and d. One interesting subquestion is whether, for every C, we can choose the bundle to be trivial. Equivalently, given C, it asks whether we can find a proper $D \subseteq \mathrm{Hom}_d^{n,ss}$

that maps finite-to-one onto C. For most curves upstairs, the answer should be positive, by simple dimension counting: as demonstrated in [Silverman 1998] and [Levy 2011], the complement of $\mathrm{Hom}_d^{n,ss}$ has high codimension, equal to about half of N. However, it turns out that the answer is sometimes negative, and in fact, for every n and d, we can find a C with only nontrivial bundle classes. More precisely:

Theorem 1.5. *For every n and d, there exists a curve with no trivial bundle class satisfying semistable reduction.*

Remark 1.6. An equivalent formulation for Theorem 1.5 is that for every n and d, we can find a curve $C \subseteq \mathrm{M}_d^{n,ss}$ such that there does not exist a curve $D \subseteq \mathrm{Hom}_d^{n,ss}$ mapping onto C under π.

Although most curves in $\mathrm{Hom}_d^{n,ss}$ can be completed, this does not imply that we can find a nontrivial bundle on an open dense set of the Chow variety of $\mathrm{M}_d^{n,ss}$. In fact, as we will see in Section 5, there exist components of the Chow variety of $\mathrm{M}_d^{n,ss}$ where, at least generically, a nontrivial bundle is required.

Our study of bundle classes now splits into two cases. In the case of curves satisfying semistable reduction with a trivial bundle, the reformulation of Remark 1.6, in its positive form, means that we can study D directly as a curve in $\mathbb{P}^N$. We can bound the degree of the map from D to C in terms of the stabilizer groups that occur on D. More precisely:

Proposition 1.7. *Let X be a projective variety over an algebraically closed field with an action by a geometrically reductive linear algebraic group G. Using the terminology of geometric invariant theory, let D be a complete curve in the stable space X^s whose quotient by G is a complete curve C; say the map from D to C has degree m. Suppose the stabilizer is generically finite of size h, and either D or C is normal. Then there exists a finite subgroup $S_D \subseteq G$, of order equal to mh, such that for all $x \in D$ and $g \in G$, $gx \in D$ if and only if $g \in S_D$.*

Corollary 1.8. *With the same notation and conditions as in Proposition 1.7, the map from D to C is ramified precisely at points $x \in D$ where the stabilizer group is larger than h, and intersects S_D in a larger subgroup than in the generic case.*

If the genus of C is 0, then the only way the map from D to C could have high degree is if it ramifies over many points; therefore, Corollary 1.8 forces the degree to be small, at least as long as C is contained in the stable locus.

In the case of curves that only satisfy semistable reduction with a nontrivial bundle, we do not have a description purely in terms of coordinates. Instead, we will study which bundle classes can be attached to every curve C. The question of which bundles occur is an invariant of C; therefore, it is essentially an invariant that we can use to study the scheme $\mathrm{Hom}(C, \mathrm{M}_d^{n,ss})$. In the sequel, we will study the scheme using the bundle class set and height invariants.

For the study of which nontrivial bundle classes can occur, first observe that fixing a D for which a bundle exists, we can apply the reformulation of Theorem 2.11 to obtain a unique extension of φ locally. This can be done at every point, so it is true globally, so we have:

Proposition 1.9. *Using the notation of Theorem 1.4, the bundle class $\boldsymbol{P}(\mathscr{E})$ depends only on D and its trivialization U_i, $U_i \hookrightarrow \mathrm{Hom}_d^{n,ss}$.*

Note that the bundle class does not necessarily depend only on D, regarded as an abstract curve with a map to C. The reason is that a point of D may not be stable, which means it may correspond to one of several different orbits, whose closures intersect. However, there are only finitely many orbits corresponding to each point, so the bundle class depends on D up to a finite amount; if C happens to be contained in the stable locus, then it depends only on D.

Thus we can study which bundle classes occur for a given C. We will content ourselves with rational curves, for which there is a relatively easy description of all projective bundles. Recall that every vector bundle over $\mathbb{P}^1$ splits as a direct sum of line bundles, and that the bundle $\bigoplus_i \mathbb{O}(m_i)$ is projectively equivalent to $\bigoplus_i \mathbb{O}(l + m_i)$ for all $l \in \mathbb{Z}$. In other words, a $\mathbb{P}^n$-bundle over $\mathbb{P}^1$ can be written as $\mathbb{O} \oplus \mathbb{O}(m_1) \oplus \cdots \oplus \mathbb{O}(m_n)$; if the m_i's are in nondecreasing order, then the expression uniquely determines the bundle's class. We will show that:

Proposition 1.10. *There exists a curve C for which multiple nonisomorphic bundle classes can occur. In fact, suppose C is isomorphic to $\mathbb{P}^1$, and there exists $U \subseteq \mathrm{Hom}_d^{n,ss}$ mapping finite-to-one into C such that U is a projective curve minus a point. Then there are always infinitely many possible classes: if the class of U is thought of as splitting as $\boldsymbol{P}(\mathscr{E}) = \mathbb{O} \oplus \mathbb{O}(m_1) \oplus \cdots \oplus \mathbb{O}(m_n)$, where $m_i \in \mathbb{N}$, then for every integer l the class $\mathbb{O} \oplus \mathbb{O}(lm_1) \oplus \cdots \oplus \mathbb{O}(lm_n)$ also occurs.*

Proposition 1.10 frustrated our initial attempt to obtain an easy classification of bundles based on curves. However, it raises multiple interesting questions instead. First, the construction uses a rational D mapping finite-to-one onto C, and going to higher m involves raising the degree of the map $D \to C$. It may turn out that bounding the degree bounds the bundle class; we conjecture that if we fix the degree of the map, then we obtain only finitely many bundle classes. Furthermore, in analogy with the consequences of Corollary 1.8, we should conversely be able to bound the degree of the map in terms of C and the bundle class, at least for rational C.

Second, it is nontrivial to find the minimal m_i's for which a bundle splitting as $\mathbb{O} \oplus \mathbb{O}(m_1) \oplus \cdots \oplus \mathbb{O}(m_n)$ would satisfy semistable reduction; the case of $n = 1$ could be stated particularly simply, as the question would be about the minimal m for which $\mathbb{O} \oplus \mathbb{O}(m)$ occurs.

In Sections 3 and 4 we will illustrate Theorem 1.5: in Section 3 we will give some examples and compute the bundle classes that occur, proving Proposition 1.10 on the way, while in Section 4 we will prove Theorem 1.5. In Section 5 we will focus on the trivial bundle case, proving Proposition 1.7 and defining the height function, which will impose constraints on which curves admit a trivial bundle; this will allow us to obtain a large family of curves C in M_2^{ss} with no trivial bundle.

2. A description of the stable and semistable spaces

Unless another reference is given, the general geometric invariant theory results given in this section are all from [Mumford and Fogarty 1982].

Recall that when a geometrically reductive linear algebraic group G has a linear action on a projectivized vector space $\mathbb{P}(V)$, we have:

Definition 2.1. A point $x \in V$ is called semistable (resp. stable) if any of the following equivalent conditions hold:

(1) There exists a G-invariant homogeneous section s such that $s(x) \neq 0$ (resp. same condition, and the action of G on x is closed).

(2) The closure of $G \cdot x$ does not contain 0 (resp. $G \cdot x$ is closed).

(3) Every one-parameter subgroup T acts on x with both nonnegative and nonpositive weights (resp. negative and positive weights).

Remark 2.2. The last condition in the definition is equivalent to having nonpositive (resp. negative) weights. This is because if we can find a subgroup acting with only negative weights, then we can take its inverse and obtain only positive weights.

Observe that for every nonzero scalar k, x is stable (resp. semistable) if and only if kx is. So the same definitions of stability and semistability hold for points of $\mathbb{P}(V)$. The definitions also descend to every G-invariant projective variety $X \subseteq \mathbb{P}(V)$; in fact, in [Mumford and Fogarty 1982] they are defined for X in terms of a G-equivariant line bundle L. When L is ample, as in the case of the space under discussion in this paper, this reduces to the above definition.

The importance of stability is captured in the following results:

Proposition 2.3. *The space of all stable points, X^s, and the space of all semistable points, X^{ss}, are both open and G-invariant.*

Theorem 2.4. *There exists a quotient $Y = X^{ss}//G$, called a good categorical quotient* (*in the category of separated schemes*), *with a natural map $\pi : X \to Y$, satisfying the following properties*:

(1) *π is a G-equivariant map, where G acts on Y trivially.*

(2) *Every G-equivariant map $X \to W$, where G acts on W trivially, factors through π.*

(3) *π is an open submersion.*

(4) *$\pi(x_1) = \pi(x_2)$ if and only if the closures of $G \cdot x_1$ and $G \cdot x_2$ intersect.*

(5) *For every open $U \subseteq Y$, $\mathbb{O}_U = \mathbb{O}(\pi^{-1}(U))^G$.*

In addition, Y is proper.

Theorem 2.5. *There exists a quotient $Z = X^s /\!/ G$, called a good geometric quotient, with a natural map $\pi : X \to Z$ satisfying all enumerated conditions of a good categorial quotient, as well as the following:*

(1) *$\pi(x_1) = \pi(x_2)$ if and only if $G \cdot x_1 = G \cdot x_2$.*

(2) *Z is naturally an open subset of $X^{ss} /\!/ G$.*

Theorem 2.6. *On X^s, the dimension of the stabilizer group* $\operatorname{Stab}_G(x)$ *is constant.*

Returning to our case of self-maps of $\mathbb{P}^n$, we write the stable and semistable spaces for the conjugation action as $\operatorname{Hom}_d^{n,s}$ and $\operatorname{Hom}_d^{n,ss}$. This involves a fair amount of abuse of notation, since those two spaces are open subvarieties of $\mathbb{P}^N$ and in fact properly contain Hom_d^n, which consists only of regular maps.

In [Levy 2011] we proved the fact that $\operatorname{Hom}_d^n \subsetneq \operatorname{Hom}_d^{n,s}$ by describing $\operatorname{Hom}_d^{n,s}$ and $\operatorname{Hom}_d^{n,ss}$ more or less explicitly. We will recapitulate the results, which are very technical but help us answer the question of when we can obtain a trivial bundle class in the semistable reduction problem and when we cannot.

We use the Hilbert–Mumford criterion, the last condition in Definition 2.1. In more explicit terms, the criterion for semistability (resp. stability) states that for every one-parameter subgroup $T \leq \mathrm{SL}(n+1)$, the action of T on φ can be diagonalized with eigenvalues t^{a_I} and at least one a_I is nonpositive (resp. negative). Now, assume by conjugation that this one-parameter subgroup is in fact diagonal, with diagonal entries $t^{a_0}, \ldots, t^{a_n}$, and that $a_0 \geq \cdots \geq a_n$; we may also assume that the a_i's are coprime, as dividing throughout by a common factor would not change the underlying group. Note also that $a_0 + \cdots + a_n = 0$. Our task is made easy by the fact that our standard coordinates for $\mathbb{A}^{N+1}$ are the monomials, on which T already acts diagonally. Throughout this analysis, we fix $\boldsymbol{a} = (a_0, \ldots, a_n)$, and similarly for $\boldsymbol{x}$ and $\boldsymbol{d}$.

Now, T acts on the $x_0^{d_0} \ldots x_n^{d_n}$ monomial of the i-th polynomial, φ_i, with weight $a_i - \boldsymbol{a} \cdot \boldsymbol{d}$. A map $\varphi \in \mathbb{P}^N$ is unstable (resp. not stable) if and only if, after conjugation, there exists a choice of a_i's such that whenever the $\mathbf{x}^{\mathbf{d}}$-coefficient of φ_i satisfies $\boldsymbol{a} \cdot \boldsymbol{d} \leq a_i$ (resp. $<$), it is equal to zero.

Remark 2.7. While in principle there are infinitely many possible T's, parametrized by a hyperplane in $\mathbb{P}^n(\mathbb{Q})$, in practice there are up to conjugation only finitely many. This is because each diagonal T imposes conditions of the form "the $\mathbf{x}^{\mathbf{d}}$-coefficient of φ_i is zero," and there are only finitely many such conditions. Thus the stable and semistable spaces are indeed open in $\mathbb{P}^N$.

Remark 2.8. The conjugation conditions we have chosen for T are such that the conditions they impose for φ to be unstable (or merely not stable) are the most stringent on φ_n and least stringent on φ_0, and are the most stringent on monomials with high x_0-degrees and least stringent on monomials with high x_n-degrees.

If $n = 1$, we have a simpler description:

Theorem 2.9 [Silverman 1998]. *$\varphi \in \mathbb{P}^N$ is unstable (resp. not stable) if and only if it is equivalent under coordinate change to a map*

$$x \mapsto \frac{a_0 x^d + \cdots + a_d y^d}{b_0 x^d + \cdots + b_d y^d},$$

such that:

(1) *$a_i = 0$ for all $i \leq (d-1)/2$ (resp. $<$).*

(2) *$b_i = 0$ for all $i \leq (d+1)/2$ (resp. $<$).*

The description for $n = 1$ can be thought of as giving a dynamical criterion for stability and semistability. A point $\varphi \in \mathbb{P}^N$ is unstable if there exists a point $x \in \mathbb{P}^1$ where φ has a bad point of degree more than $(d+1)/2$, or φ has a bad point of degree more than $(d-1)/2$ where it in addition has a fixed point. Following Rahul Pandharipande's unpublished reinterpretation of [Silverman 1998], we define "bad point" as a vertical component of the graph $\Gamma_\varphi \subseteq \mathbb{P}^1 \times \mathbb{P}^1$, and "fixed point" as a fixed point of the unique nonvertical component of Γ_φ. When $n = 1$, $d = 2$, this condition reduces to having a fixed point at a bad point, or alternatively, a repeated bad point.

The conditions for higher n are not as geometric. However, if we interpret fixed points liberally enough, there are still strong parallels with the $n = 1$ case. One can show that the unstable space for $n = 2$ and $d = 2$ consists of two irreducible components, which roughly generalize the $n = 1$, $d = 2$ condition of having a fixed point at a bad point; in this case, one needs to define a limit of the value of $\varphi(x)$ as x approaches the bad point, though this limit can be defined purely in terms of degrees of polynomials, without needing to resort to a specific metric on the base field.

Finally, let us prove semistable reduction. Let us restate Theorem 1.4:

Theorem 2.10. *If C is a complete curve with $K(C)$ its function field, and if $\varphi_{K(C)}$ is a semistable rational map on $\mathbb{P}^n_{K(C)}$, then there exists a curve D mapping finite-to-one onto C with a $\mathbb{P}^n$-bundle $\boldsymbol{P}(\mathscr{E})$ on D with a self-map Φ such that:*

(1) *The restriction φ_x of Φ to the fiber of each $x \in D$ is a semistable rational self-map.*

(2) *Φ is a semistable map over $K(D)$, and is equivalent to $\varphi_{K(D)}$ under coordinate change.*

Semistable reduction can be thought of as extending a rational map defined over a field K to a rational map defined over a discrete valuation ring R whose fraction field is K, in a way that is not too degenerate. The reason a discrete valuation ring suffices is that once we know we can extend to a discrete valuation ring, we can extend to some larger integral domain. In other words, it suffices to show the following, more general statement:

Theorem 2.11. *Let G be a geometrically reductive group acting on a projective variety X whose stable and semistable spaces are X^s and X^{ss}, respectively. Let R be a discrete valuation ring with fraction field K, and let $x_K \in X^s_K$. Then for some finite extension K' of K, with R' the integral closure of R in K', x_K has an integral model over R' with semistable reduction modulo the maximal ideal. In other words, we can find some $A \in G(\overline{K})$ such that $A \cdot x_K$ has semistable reduction. If $x_K \in X^{ss}_K$, then the same result is true, except that $x_{R'}$ could be an integral model for some $x'_{K'}$ mapping to the same point of $X^{ss}//G$ such that $x'_{K'} \notin G \cdot x_K$.*

Proof. We follow the method used in [Zhang 1996]. Let C be the Zariski closure of x_K in $X^{ss}_R//G$, and reduce it modulo the maximal ideal to obtain x_k, where k is the residue field of R. Observe that C is a one-dimensional subscheme of $X^{ss}_R//G$ and is isomorphic to $\operatorname{Spec} R$, and is as a result connected. Since G is connected, the preimage $\pi^{-1}(C)$ is also connected: when x_K is stable, this follows from the fact that $\pi^{-1}(C)$ is the Zariski closure of $G \cdot x_K$ in X^{ss}, and even when it is not, $\pi^{-1}(C)$ is the union of connected orbits whose closures intersect. Further, since $\pi^{-1}(C)$ surjects onto C, we can find an integral one-dimensional subscheme mapping surjectively to C. This subscheme necessarily maps finite-to-one onto C by dimension counting, so it is isomorphic to some finite extension ring R', giving us K' as its fraction field. □

Remark 2.12. Theorem 2.11 can also be proven in a much more explicit way, producing for each $\varphi_K \in \mathrm{Hom}_d^{n,ss}$ a sequence of A's conjugating it to a model with semistable reduction.

Remark 2.13. Szpiro et al. [2010] study semistable reduction for the moduli space of self-maps of $\mathbb{P}^1$ and raise a conjecture that Theorems 1.4 and 2.11 answer in the affirmative.

3. Examples of nontrivial bundles

In the case $n = 1$, we follow [Silverman 1998] and write Rat_d for Hom_d^1 and M_d for M_d^1. The space Rat_2 and its quotient M_2 have been analyzed with more success than the larger spaces, yielding the following prior structure result:

Theorem 3.1 [Milnor 2006; Silverman 1998]. $\mathrm{M}_2 = \mathbb{A}^2$; $\mathrm{M}_2^s = \mathrm{M}_2^{ss} = \mathbb{P}^2$. *The first two elementary symmetric polynomials in the multipliers of the fixed points realize both isomorphisms.*

Recall that within $\mathbb{P}^N = \mathbb{P}^5$, a map $(a_0x^2 + a_1xy + a_2y^2)/(b_0x^2 + b_1xy + b_2y^2)$ is unstable if and only if it is in the closure of the PGL(2)-orbit of the subvariety $a_0 = b_0 = b_1 = 0$. In other words, it is unstable if and only if there the map is degenerate and has a double bad point, or a fixed point at a bad point.

Definition 3.2. A map on $\mathbb{P}^1$ is a polynomial if and only if there exists a totally invariant fixed point. Taking such a point to infinity turns the map into a polynomial in the ordinary sense. In Rat_d, or generally in $\mathbb{P}^N = \mathbb{P}^{2d+1}$, a map is polynomial if and only if it is in the closure of the PGL(2)-orbit of the subvariety defined by zeros in all coefficients in the denominator except the y^d-coefficient.

Remark 3.3. A totally invariant fixed point is not necessarily a totally fixed point. A totally invariant fixed point is one that is totally ramified. A totally fixed point is the root of the fixed point polynomial when it is unique, that is, when the polynomial is a power of a linear term. In fact by an easy computation, a map has a totally invariant, totally fixed point x if and only if it is degenerate linear with bad point of multiplicity $d-1$ at x, in which case it is necessarily unstable.

The polynomial maps define a curve in M_2^{ss}; we will show:

Proposition 3.4. *The polynomial curve in* M_2^{ss} *only satisfies semistable reduction with nontrivial bundles.*

Proof. First, note that in $\mathbb{P}^5$, the polynomial maps are those that can be conjugated to the form $(a_0x^2 + a_1xy + a_2y^2)/b_2y^2$, in which case the totally invariant fixed point is $\infty = (1 : 0)$. We will call the polynomial map locus X. If $a_0 = 0$, then the map is unstable; we will show that every curve in X contains a map for which $a_0 = 0$. Clearly, the set of all maps with a given totally invariant fixed point is isomorphic to $\mathbb{P}^3$, and the unstable locus within it is isomorphic to $\mathbb{P}^2$ as a linear subvariety, so for there to be any hope of a trivial bundle, a curve in X cannot lie entirely over one totally invariant point.

Now, the fixed point equation for a map of the form f/g is $fy - gx = 0$; the homogeneous roots of this equation are the fixed points, with the correct multiplicities. For our purposes, when the totally invariant point is ∞, the fixed point equation is $a_0x^2y + (a_1 - b_2)xy^2 + a_2y^3 = 0$. We get that $a_0 = 0$ if and only if the totally invariant point is a repeated root of the fixed point equation.

There exists a map from X to $\mathbb{P}^1 \times \mathbb{P}^2$, mapping φ to its totally invariant point in $\mathbb{P}^1$, and to the two elementary symmetric polynomials in the two other fixed points in $\mathbb{P}^2$. Write $(x : y)$ for the image in $\mathbb{P}^1$ and $(a : b : c)$ for the image in $\mathbb{P}^2$. Now $(x : y)$ is a repeated root if $ax^2 + bxy + cy^2 = 0$. The equation defines an ample divisor, so every curve in $\mathbb{P}^1 \times \mathbb{P}^2$ will meet it. Finally, a curve in X maps either to a single point in $\mathbb{P}^1 \times \mathbb{P}^2$, in which case it must contain points with $a_0 = 0$ as above, or to a curve, in which case it intersects the divisor $ax^2 + bxy + cy^2 = 0$. In both cases, the curve contains unstable points. Thus there is no global semistable curve D in Rat_2^{ss} mapping down to C. □

Note that in the above proof, maps conjugate to x^2 have two totally invariant points, so a priori the map from X to $\mathbb{P}^1 \times \mathbb{P}^2$ is not well-defined at them. However, for any curve D in X, there is a well-defined completion of this map, whose value at x^2 on the $\mathbb{P}^1$ factor is one of the two totally invariant points. Thus this complication does not invalidate the above proof.

Let us now compute the vector bundle classes that do occur for the polynomial curve. We work with the description $x^2 + c$, which yields an affine curve that maps one-to-one into C, missing only the point at infinity, which is conjugate to $(x^2 - x)/0$. To hit the point at infinity, we choose the alternative parametrization $cx^2 - cx + 1$, which, when $c = \infty$, corresponds to the unique (up to conjugation) semistable degenerate constant map. For any c, this map is conjugate to $x^2 - cx + c$ and thence $x^2 + c/2 - c^2/4$, using the transition function $[c, -\frac{1}{2}; 0, 1]$. Thus the bundle splits as $\mathbb{O} \oplus \mathbb{O}(1)$.

This bundle depends on the choice of D. In fact, if we choose another parametrization for D, for example $c^2x^2 - c^2x + 1$, then the transition function is $[c^2, -\frac{1}{2}; 0, 1]$, which leads to the bundle $\mathbb{O} \oplus \mathbb{O}(2)$. This is not equivalent to $\mathbb{O} \oplus \mathbb{O}(1)$. This then leads to the question of which classes of bundles can occur over each C. In the example we have just done, the answer is every nontrivial class: for every positive integer m, we can use $c^m x^2 - c^m x + 1$ as a parametrization, leading to $\mathbb{O} \oplus \mathbb{O}(m)$, which exhausts all nontrivial projective bundle classes.

Recall the result of Proposition 1.10:

Proposition 3.5. *Suppose C is isomorphic to $\mathbb{P}^1$, and there exists $U \subseteq \mathrm{Hom}_d^{n,ss}$ mapping finite-to-one into C such that U is a projective curve minus a point. Then there are always infinitely many possible classes: if the class of U is thought of as splitting as $\boldsymbol{P}(\mathscr{E}) = \mathbb{O} \oplus \mathbb{O}(m_1) \oplus \cdots \oplus \mathbb{O}(m_n)$, where $m_i \in \mathbb{N}$, then for every integer l, the class $\mathbb{O} \oplus \mathbb{O}(lm_1) \oplus \cdots \oplus \mathbb{O}(lm_n)$ also occurs.*

Proof. Imitating the analysis of the polynomial curve above, we can parametrize C by one variable, say c, and choose coordinates such that the sole bad point in the closure of U corresponds to $c = \infty$. Now, we can by assumption find a piece U' above the infinite point with a transition function determining the vector bundle $\mathbb{O} \oplus \mathbb{O}(m_1) \oplus \cdots \oplus \mathbb{O}(m_n)$. Now let V be the composition of U' with the map $c \mapsto c^l$. Then U and V determine a vector bundle satisfying semistable reduction, of class $\mathbb{O} \oplus \mathbb{O}(lm_1) \oplus \cdots \oplus \mathbb{O}(lm_n)$, as required. □

The example in Proposition 3.4, of polynomial maps, is equivalent to a multiplier condition. When $d = 2$, a map is polynomial if and only if it has a superattracting fixed point, that is, one whose multiplier is zero; see the description in the first chapter of [Silverman 2007]. One can imitate the proof that semistable reduction does not hold for a more general curve, defined by the condition that there exists a fixed point of multiplier $t \neq 1$. In that case, the condition $b_1 = 0$ is replaced by $b_1 = ta_0$, and the point is a repeated root of the fixed point equation if and only if $a_0 = b_1$, in which case we clearly have $a_0 = b_1 = 0$ and the point is unstable.

When the multiplier is 1, the fixed point in question is automatically a repeated root, with $b_1 = a_0$. The condition that the point be the only fixed point corresponds to $b_2 = a_1$, which by itself does not imply that the map fails to be a morphism, let alone that it is unstable.

Instead, the condition that gives us $b_1 = a_0 = 0$ is the condition that the fixed point be totally invariant. Specifically, the fixed point's two preimages are itself and one more point; when the fixed point is ∞, the extra point is $-b_2/b_1$. Now we can map X to $\mathbb{P}^1 \times \mathbb{P}^1$, where the first coordinate is the fixed point and the second is its preimage. This map is well-defined on all of X because only one point can be a double root of a cubic. Now the diagonal is ample in $\mathbb{P}^1 \times \mathbb{P}^1$, so the only way a curve D can avoid it is by mapping to a single point; but in that case, D lies in a fixed variety isomorphic to $\mathbb{P}^3$ where the unstable locus is $\mathbb{P}^2$, so it will intersect the unstable locus.

The fact that any condition of the form "there exists a fixed point of multiplier t" induces a curve for which semistable reduction requires a nontrivial bundle means that there is no hope of enlarging the semistable space in a way that ensures we always have a trivial bundle. We really do need to think of semistable reduction as encompassing nontrivial bundle classes as well as trivial ones.

Specifically: it is trivial to show that the closure of the polynomial locus in Rat_2 includes all the unstable points (fix ∞ to be the totally invariant point and let a_0 go to zero). At least some of those unstable points will also arise as closures of other multiplier-t conditions. However, different multiplier-t conditions limit to different points in $\mathrm{M}_2^{ss} \setminus \mathrm{M}_2$.

4. The general case

So far we have talked about nontrivial classes in M_2. But we have a stronger result, restating Theorem 1.5:

Theorem 4.1. *For all n and d, over any base field, there exists a curve with no trivial bundle class satisfying semistable reduction.*

Proof. In all cases, we will focus on ***polynomial maps***, which we will define to be maps that are $\mathrm{PGL}(n+1)$-conjugate to maps for which the last polynomial φ_n has zero coefficients in every monomial except possibly x_n^d.

Lemma 4.2. *The set of polynomial maps, defined above, is closed in* $\overline{\mathrm{Hom}_d^n} = \mathbb{P}^N$.

Proof. Clearly, the set of polynomial maps with respect to a particular hyperplane — for example, $x_n = 0$ — is closed. Now, for each hyperplane $a_0x_0 + \cdots + a_nx_n = 0$, we can check by conjugation to see that the condition that the map be polynomial corresponds to the condition that $a_0\varphi_0 + \cdots + a_n\varphi_n = c(a_0x_0 + \cdots + a_nx_n)^d$, where c may be zero. As $\mathbb{P}^n$ is proper, it suffices to show that the condition "φ is polynomial with respect to $a_0x_0 + \cdots + a_nx_n = 0$" is closed in $(\mathbb{P}^n)^* \times \mathbb{P}^N$.

Now, we may construct a rational function f from $(\mathbb{P}^n)^* \times \mathbb{P}^N$ to $\mathrm{Sym}^d(\mathbb{P}^n) \times \mathrm{Sym}^d(\mathbb{P}^n)$ by $\big((a_0x_0 + \ldots + a_nx_n), \varphi\big) \mapsto \big((a_0x_0 + \cdots + a_nx_n)^d, a_0\varphi_0 + \cdots + a_n\varphi_n\big)$. The map φ is polynomial with respect to $a_0x_0 + \cdots + a_nx_n = 0$ if and only if f is ill-defined at $\big((a_0x_0 + \cdots + a_nx_n), \varphi\big)$ or $f\big((a_0x_0 + \cdots + a_nx_n), \varphi\big) \in \Delta$, the diagonal subvariety. The ill-defined locus of f is closed, and the preimage of Δ is closed in the well-defined locus. $\square$

In fact, the condition of φ being polynomial with respect to r distinct hyperplanes in general position, where r is a fixed integer — in other words, the condition that φ be conjugate to a map for which $\varphi_i = c_i x_i^d$ for all $i > d - r$ — is more or less closed as well. It is not closed, but a sufficiently good condition is closed. Namely:

Lemma 4.3. *For each* $1 \leq i \leq n$, *consider the* $\mathrm{PGL}(n+1)$*-orbit of the space of maps in which, for each* $j \geq i$, φ_j *has zero coefficients in every monomial containing any term* x_k *with* $k < j$. *This orbit is closed in* $\mathbb{P}^N$.

Proof. Observe that the above-defined space of maps consists of maps that are polynomial with respect to $x_n = 0$, such that the induced map on the totally invariant hyperplane $x_n = 0$ is polynomial with respect to $x_{n-1} = 0$, and so on until we reach the induced map on the totally invariant subspace $x_{i+1} = \cdots = x_n = 0$.

Now we use descending induction. Lemma 4.2 is the base case, when $i = n$. Now suppose it is true down to i. Then for $i - 1$, the condition of having no nonzero x_k term in φ_{i-1} with $k < i - 1$ is equivalent to the condition that the induced map on the totally invariant subspace $x_i = x_{i+1} = \ldots = x_n = 0$ be polynomial; this condition is closed in the space of all maps that are polynomial down to x_i, which we assume closed by the induction hypothesis. $\square$

Definition 4.4. We call maps of the form in Lemma 4.3 ***polynomial with respect to B***, where B is the Borel subgroup preserving the ordered basis of conditions. In the case above, B is the upper triangular matrices.

We need one final result to make computations easier:

Lemma 4.5. *Let* X *be a curve of polynomial maps, all with respect to a Borel subgroup* B, *and let* φ *be a semistable map in* $\overline{\mathrm{PGL}(n+1) \cdot X}$. *Then* $\varphi \in \overline{B \cdot X}$.

Proof. Let C be the closure of the image of X in $\mathrm{M}_d^{n,ss}$. By semistable reduction, there exists some affine curve $Y \ni \varphi$ mapping finite-to-one to C, that is, dominantly. We need to find some open $Z \subseteq Y$ containing φ and some $f : Z \to \mathrm{PGL}(n+1)$ such that $f(\varphi)$ is the identity matrix, and $Z' = \{(f(z) \cdot z)\}$ consists of maps which are polynomial with respect to B. Such a map necessarily exists: we have a map h from Y to the flag variety of $\mathbb{P}^n$ sending each y to the subgroup with respect to which it is polynomial (possibly involving some choice if generically y is polynomial with respect to more than one flag), which then lifts to G, possibly after deleting finitely many points. Generically, a point of X maps to a point of C that is in the image of Z; therefore, picking the correct points in X, we get that $\varphi \in \overline{B \cdot X}$. □

With the above lemmas, let us now prove Theorem 4.1 for $n = 1$, which is slightly easier than the higher-n case, where the more complicated Lemma 4.3 is required. We will use the family $x^d + c$, where $c \in \mathbb{A}^1$. In projective notation, this is $(a_0 x^d + a_d y^d)/b_d y^d$, which is a one-dimensional family modulo conjugation.

Lemma 4.6. *Let V be the closure of the* PGL(2)*-orbit of the family*

$$\frac{a_0 x^d + a_d y^d}{b_d y^d}$$

in $\mathbb{P}^N$. Then:

(1) *In characteristic* 0 *or $p \nmid d$, every $\varphi \in V$ is actually in the* PGL(2)*-orbit of the family, or else it is a degenerate linear map, conjugate to*

$$\frac{a_{d-1} x y^{d-1} + a_d y^d}{b_d y^d}.$$

(2) *In characteristic $p \mid d$, with $p^m \parallel d$ and $p^m \neq d$, every $\varphi \in V$ is in the* PGL(2)*-orbit of the family or is a degenerate map conjugate to*

$$\frac{a_{d-p^m} x y^{d-p^m} + a_d y^d}{b_d y^d}.$$

(3) *In characteristic p with $d = p^m$, set V to be the closure of the orbit of the family $(a_0 x^d + a_{d-1} x y^{d-1})/b_d y^d$; then every $\varphi \in V$ is actually in the orbit of the family, or else it is a degenerate linear map, conjugate to*

$$\frac{a_{d-1} x y^{d-1} + a_d y^d}{b_d y^d},$$

and furthermore, $a_{d-1} = b_d$.

Proof. Observe that the first two cases are really the same: case (2) is reduced to case (1) viewed as a degree-(d/p^m) map in $(x^{p^m} : y^{p^m})$. So it suffices to prove case (1) to prove (2); we will start with the family $(a_0 x^d + a_d y^d)/b_d y^d$ and see what algebraic equations its orbit satisfies. As polynomials are closed in $\overline{\mathrm{Rat}_d}$, every point in the closure of the orbit is a polynomial. We may further assume it is polynomial with respect to $y = 0$; therefore, by Lemma 4.5, it suffices to look at the action of upper triangular matrices. Further, the condition of being within the family $(a_0x^d + a_d y^d)/b_d y^d$ is stabilized by diagonal matrices; therefore, it suffices to look at the action of matrices of the form $[1, t; 0, 1]$.

Now, the conjugation action of $[1, t; 0, 1]$ fixes $b_d y^d$ and maps $a_0x^d + a_d y^d$ to $a_0(x-ty)^d+(a_d+tb_d)y^d$. Clearly, there is no hope of obtaining any condition on b_d or a_d. Now, the conditions on the terms $a_0, \dots, a_{d-1}$ are that for some t, they fit into the pattern $a_0(x^d-dtx^{d-1}y+\cdots\pm dt^{d-1}xy^{d-1})$, that is, $a_i = (-t)^i\binom{d}{i}a_0$. To remove the dependence on t, note that when $i + j = k+l$, we have $\binom{d}{i}\binom{d}{j}a_ia_j = \binom{d}{k}\binom{d}{l}a_ka_l$, as long as $i, j, k, l < d$.

Let us now look at what those conditions imply. Setting $j = i, k = i-1, l = i+1$, we get conditions of the form $\binom{d}{i}^2 a_i^2 = \binom{d}{i-1}\binom{d}{i+1}a_{i-1}a_{i+1}$, whenever $i+1 < d$. If $a_0 \neq 0$, then the value of a_1 uniquely determines the value of a_2 by the condition with $i = 1$; the value of a_2 uniquely determines a_3 by the condition with $i = 2$; and so on, until we uniquely determine a_{d-1}. In this case, choosing $t = -a_1/da_0$ will conjugate this map back to the family $(a_0x^d + a_d y^d)/b_d y^d$. If $a_0 = 0$, then the equation with $i = 1$ will imply that $a_1 = 0$; then the equation with $i = 2$ will imply that $a_2 = 0$; and so on, until we set $a_{d-2} = 0$. We cannot ensure $a_{d-1} = 0$ because a_{d-1} always appears in those equations multiplied by a different a_i, instead of squared. Hence we could get a degenerate-linear map.

In case (3), we again look at the action of matrices of the form $[1, t; 0, 1]$. Such matrices map $(a_0x^d + a_{d-1}xy^{d-1})/b_d y^d$ to

$$\frac{a_0x^d + a_{d-1}xy^{d-1} + (-a_0t^d - a_{d-1}t + b_dt)y^d}{b_d y^d}.$$

Now the only way a map of the form $(a_0x^d + a_{d-1}xy^{d-1} + a_d y^d)/b_d y^d$ could degenerate is if the image of the polynomial map $t \mapsto -a_0t^d - a_{d-1}t + b_dt$ misses a_d, which could only happen if the polynomial were constant, that is, $a_0 = 0$ and $a_{d-1} = b_d$, giving us a degenerate-linear map. □

Remark 4.7. The importance of the lemma is that in all degenerate cases, the map is necessarily unstable, since $d-1$ (or, in case (2), $d - p^m$) is always at least as large as $d/2$.

We can now prove Theorem 4.1 when $n = 1$. So if we can always find a $D \subseteq \mathrm{Hom}_d^{n,ss}$ that works globally, we can find one over a family in which every map is conjugate to $(a_0x^d + a_d y^d)/b_d y^d$, or, in characteristic p with $d = p^m$,

$$\frac{a_0x^d + a_{d-1}xy^{d-1}}{b_d y^d}.$$

It suffices to show that there exists a map with $a_0 = 0$. For this, we use the fixed point polynomial, which is well-defined on this family. If the polynomial is fixed, then all maps in the family may be simultaneously conjugated to the form

$$\frac{a_0 x^d + a_d y^d}{b_d y^d}$$

(or $(a_0 x^d + a_{d-1} x y^{d-1})/b_d y^d$), and then one map must have $a_0 = 0$. If the polynomial varies, then some map will have the point at infinity colliding with another fixed point. This will force the map to be ill-defined at infinity; recall that totally invariant points are simple roots of the fixed point polynomial, unless they are bad. This will force a_0 to be zero, again.

For higher n, the proof is similar. The lemma we need is similar to Lemma 4.6, but is somewhat more complicated:

Lemma 4.8. *Let V be the closure of the* $\mathrm{PGL}(n+1)$*-orbit of the family*

$$(c_0 x_0^d + b x_1^d : \varphi_1 : \ldots : \varphi_n),$$

where φ_i is x_j-free for all $j < i$.

(1) *If the characteristic does not divide d, then every $\varphi \in V$ is actually in the* $\mathrm{PGL}(n+1)$*-orbit of the family, or else it is a degenerate map, whose only possible nonzero coefficients in φ_0 are those without an x_0 term and those of the form $x_0 p_0$, where there is no nonzero x_0-term in p_0.*

(2) *If the characteristic p satisfies $p \mid d$, with $d \neq p^m \,\|\, d$, then the same statement as in case* (1) *holds as long as each φ_i is in terms of $x_j^{p^m}$, but with $x_0 p_0$ replaced by $x_0^{p^m} p_0$.*

(3) *If the characteristic p satisfies $d = p^m$, then changing the family to*

$$(c_0 x_0^d + b x_0 x_1^{d-1} : \varphi_1 : \ldots : \varphi_n),$$

with φ_i in terms of x_j^d as in case (2), *the same statement as in case* (1) *holds.*

Proof. As in the one-dimensional case, case (2) is reducible to case (1) with d replaced with d/p^m and x_i with $x_i^{p^m}$. By Lemma 4.5, we only need to conjugate by upper triangular matrices. Further, we only need to conjugate by matrices of the family E, with first row $(1, t_1, \ldots, t_n)$ and other rows the same as the identity matrix. This is because we can control the diagonal elements because the condition of being in the family is diagonal matrix-invariant, and we can control the rest by projecting any curve Z of unipotent upper triangular matrices onto E.

Set $a_{\boldsymbol{d}}$ to be the $\boldsymbol{x^d}$-coefficient in φ_0. For all vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}, \boldsymbol{l}$ with $\boldsymbol{i} + \boldsymbol{j} = \boldsymbol{k} + \boldsymbol{l}$, we have $\binom{d}{\boldsymbol{i}}\binom{d}{\boldsymbol{j}} a_{\boldsymbol{i}} a_{\boldsymbol{j}} = \binom{d}{\boldsymbol{k}}\binom{d}{\boldsymbol{l}} a_{\boldsymbol{k}} a_{\boldsymbol{l}}$, as long as none of $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, or $\boldsymbol{l}$ is in the span of $\boldsymbol{e}_i$

for $i > 0$. Note that i and $\boldsymbol{i}$ are two separate quantities, one an index of coordinates and one an index of monomials.

As in the one-dimensional case, we may set $\boldsymbol{j} = \boldsymbol{i}$ and $\boldsymbol{k} = \boldsymbol{i} - \boldsymbol{e}_0 + \boldsymbol{e}_i$. If $c_0 = a_{(d,0,\dots,0)} \neq 0$, then by the same argument as before, the values of the $x_0^{d-1}x_i$-coefficients determine all the rest, and we can conjugate the map back to the desired form. And if $c_0 = 0$, then the value of every coefficient that can occur as $\boldsymbol{i}$ in the above construct is zero; the only coefficients that cannot are those with no x_0 component and those with a linear x_0 component.

In case (3), we restrict to matrices of the same form as in case (1), and observe that those matrices only generate extra x_i^d and $x_i x_1^{d-1}$ in φ_0. The statement is vacuous if $c_0 = 0$, so assume $c_0 \neq 0$. For $i = 1$, this is identical to the one-dimensional case, so if $c_0 \neq 0$, then we can find an appropriate t_1. For higher i, if $b \neq 0$, then we can extract t_i from the $x_i x_1^{d-1}$ coefficient, which will necessarily work for the x_i^d coefficient as well, making the map conjugate to the family; if $b = 0$, then the same equations as for $i = 1$ hold for higher i, and we can again find t_i's conjugating the map to the family. □

While we could also control the terms involving a linear (or p-power) x_0 coefficient in the above construction, it is not necessary for our purposes.

To finish the proof of Theorem 4.1, first note that in the closure of the family above, any map for which $c_0 = 0$ is unstable. Indeed, the one-parameter subgroup of $\mathrm{PGL}(n+1)$ with diagonal coefficients $t_0 = n$, $t_i = -1$ for $i > 0$, shows instability. Recall that a map is unstable with respect to such a family if $t_i > t_0 d_0 + \cdots + t_n d_n$ whenever the $x_0^{d_0} \dots x_n^{d_n}$-coefficient of φ_i is nonzero. With the above one-parameter subgroup, we have $t_0 d_0 + \cdots + t_n d_n = -d < -1$ for the only nonzero monomials in φ_i with $i > 0$; in φ_0, the maximal value of $t_0 d_0 + \cdots + t_n d_n$ is $t_0 + t_i(d-1) = n - (d-1) < n$.

Now we need to show only that for some map in the family, c_0 will indeed be zero. So suppose on the contrary that c_0 is never zero. Then all maps are, after conjugation, in the family $(c_0 x_0^d + b x_1^d : \varphi_1 \dots : \varphi_n)$, where the linear subvariety $\varphi_i = \varphi_{i+1} = \cdots = \varphi_n$ is totally invariant. Now look at the action on the line $x_2 = \cdots = x_n = 0$. Every morphism will induce a morphism on this line, so there will be three fixed points on it, counting multiplicity. We now imitate the proof in the one-dimensional case: the totally invariant fixed point on this line, $(1 : 0 : \dots : 0)$, will collide with another fixed point, so the map will be ill-defined at it. This means that $(1 : 0 : \dots : 0)$ is a bad point, which cannot happen unless $c_0 = 0$. □

Trivially, the above theorem for curves shows the same for higher-dimensional families in $\mathrm{M}_d^{n,ss}$. An interesting question could be to generalize semistable reduction to higher-dimensional families, for which we may get projective vector bundles just like in the case of curves. Trivially, if we have two proper subvarieties of $\mathrm{M}_d^{n,ss}$,

$V_1 \subseteq V_2$, and a bundle class occurs for V_2, then its restriction to V_1 occurs for V_1. In particular, if we have the trivial class over V_2, then we also have it over V_1, as well as any other subvariety of V_2. This leads to the following question: if the trivial class occurs for every proper closed subvariety of V_2, does it necessarily occur for V_2? What if we weaken the condition and only require the trivial class to occur for subvarieties that cover V_2?

5. The trivial bundle case

For most curves $C \subseteq \mathrm{M}_d^{n,ss}$, there occurs a trivial bundle. Since the complement of $\mathrm{Hom}_d^{n,ss}$ in $\mathbb{P}^N$ has high codimension, this is true by simple dimension counting. Therefore, it is useful to analyze those curves separately, as we have more tools to work with. Specifically, we can use more machinery from geometric invariant theory. We will start by proving Proposition 1.7, restated below:

Proposition 5.1. *Let X be a projective variety over an algebraically closed field with an action by a geometrically reductive linear algebraic group G. Using the terminology of geometric invariant theory, let D be a complete curve in the stable space X^s whose quotient by G is a complete curve C; say the map from D to C has degree m. Suppose the stabilizer is generically finite, of size h, and either D or C is normal. Then there exists a finite subgroup $S_D \subseteq G$, of order equal to mh, such that for all $x \in D$ and $g \in G$, $gx \in D$ if and only if $g \in S_D$.*

Proof. For $x \in D$, we define $S_D(x) = \{g \in G : gx \in D\}$. This is a map of sets from an open dense subset of D to $\mathrm{Sym}^{mh}(G)$, and is regular on an open dense subset. We have:

Lemma 5.2. *The map from* $\mathrm{Sym}^{mh}(G) \times X^s$ *to* $\mathrm{Sym}^{mh}(X^s) \times X^s$ *defined by sending each* $\big(\{g_1, \dots, g_{mh}\}, x\big)$ *to* $\big(\{g_1 \cdot x, \dots, g_{mh} \cdot x\}, x\big)$ *is proper.*

Proof. By standard geometric invariant theory, the map from $G \times X^s$ to $X^s \times X^s$, $(g, x) \mapsto (g \cdot x, x)$, is proper. Thus the map from $G^{mh} \times (X^s)^{mh}$ to $(X^s)^{mh} \times (X^s)^{mh}$ defined by $(g_i, x_i) \mapsto (g_i \cdot x_i, x_i)$ is also proper, as the product of proper maps. Now closed immersions are proper, so the map remains proper if we restrict it to $G^{mh} \times X^s$, where we embed X^s into $(X^s)^{mh}$ diagonally; the image of this map lands in $(X^s)^{mh} \times X^s$. Finally, we quotient out by the symmetric group S_{mh}, obtaining

$$\begin{tikzcd}
G^{mh} \times X^s \arrow[r] \arrow[d, "\pi"'] & (X^s)^{mh} \times X^s \arrow[d, "\pi"] \\
\mathrm{Sym}^{mh}(G) \times X^s \arrow[r] & \mathrm{Sym}^{mh}(X^s) \times X^s
\end{tikzcd}$$

The map on the bottom is already separated and finite-type; we will show it is universally closed. Extend it by some arbitrary scheme Y. If

$$V \subseteq \mathrm{Sym}^{mh}(G) \times X^s \times Y$$

is closed, then so is $\pi^{-1}(V) \subseteq G^{mh} \times X^s \times Y$. The map on top is universally closed, so its image is closed in $(X^s)^{mh} \times X^s \times Y$. But the map on the right is proper, so the image of V is also closed in $\mathrm{Sym}^{mh}(X^s) \times X^s \times Y$. □

Now, the rational map $f_D(x) = S_D(x) \cdot x \in \mathrm{Sym}^{mh}(D)$ can be extended to a morphism on all of D, since both D and $\mathrm{Sym}^{mh}(D)$ are proper. This is trivial if D is normal; if it is not normal, but C is normal, then observe that the map factors through C since it is constant on orbits, and then analytically extend it through C. But now $(f_D(x), x)$ embeds into $\mathrm{Sym}^{mh}(X^s) \times X^s$ as a proper curve. The preimage in $\mathrm{Sym}^{mh}(G) \times X^s$ of this curve is also proper; for each $(f_D(x), x)$, it is a finite set of points of the form (S, x) satisfying $S \cdot x = f_D(x)$, including $(S_D(x), x)$. Projecting onto the $\mathrm{Sym}^{mh}(G)$ factor, we still get a proper set, which means it must be a finite set of points, as $\mathrm{Sym}^{mh}(G)$ is affine. One of these points will be S_D, which is then necessarily finite.

Finally, if $g, h \in S_D$ and $x \in D$, then $g \cdot h \cdot x \in g \cdot D = D$; therefore S_D is a group. □

Remark 5.3. The proposition essentially says that the cover $D \to C$ is necessarily Galois. The generic stabilizer is necessarily a group H, normal in S_D.

Corollary 5.4. *With the same notation and conditions as in Proposition 5.1, the map from D to C ramifies precisely at points $x \in D$ such that* $\mathrm{Stab}(x)$ *intersects S_D in a strictly larger group than H. Furthermore, the ramification degree is exactly* $[\mathrm{Stab}(x) \cap S_D : H]$.

For high n or d, the stabilized locus of Hom_d^n is of high codimension. Furthermore, most curves in $\mathrm{Hom}_d^{n,ss}$ lie in $\mathrm{Hom}_d^{n,s}$. Therefore, generically not only is H trivial, but also there are no points on D with nontrivial stabilizer. Thus for most C and D, the map $D \to C$ must be unramified. Thus, when C is rational, generically the degree is 1.

It's based on this observation that we conjecture the bounds for the nontrivial bundle case in both directions — that is, that if we fix C and the bundle class $\boldsymbol{P}(\mathscr{E})$, then the degree of the map $\pi : D \to C$ is bounded.

Using the structure result on $\mathrm{M}_2^{ss} = \mathbb{P}^2$, we can prove much more:

Proposition 5.5. *If C is a generic line in M_2^{ss}, then it requires a nontrivial bundle.*

Proof. Generically, C is not the line consisting of the resultant locus, $\mathrm{M}_2^{ss} \setminus \mathrm{M}_2$. So it intersects this line at exactly one point. Furthermore, since the resultant Res_2 is an $\mathrm{SL}(2)$-invariant section, we have $D.\mathrm{Res}_2 = m \cdot C.\mathrm{Res}_2$; we abuse notation and use Res_d^n to refer to the resultant divisor both upstairs and downstairs. Since the degree of the resultant upstairs is $(n+1)d^n = 4$ [Jouanolou 1991], we obtain $4 \cdot D.\mathbb{O}(1) = m$. In other words, $m \geq 4$.

However, using Proposition 5.1, we will show $m \leq 2$ generically. The generic stabilizer is trivial, and the stabilized locus is a cuspidal cubic in $\mathbb{P}^2$, on which the stabilizer is isomorphic to $\mathbb{Z}/2\mathbb{Z}$, except at the cusp, where it is S_3. The generic line C will intersect this cuspidal curve at three points, none of which is the cusp. Therefore, $h = 1$, and there are at most three points of ramification, with ramification degree 2. By Riemann–Hurwitz, the maximum m is 2, contradicting $m \geq 4$. □

Acknowledgements

The author would like to thank his Ph.D. advisor Shouwu Zhang for proposing this problem, Johan de Jong for many helpful discussions about Section 5, and the anonymous reviewer for fixing the presentation and proof of Theorem 1.4.

References

[Jouanolou 1991] J.-P. Jouanolou, “Le formalisme du résultant”, *Adv. Math.* **90**:2 (1991), 117–263. MR 93g:14003 Zbl 0747.13007

[Levy 2011] A. Levy, “The space of morphisms on projective space”, *Acta Arith.* **146**:1 (2011), 13–31. MR 2012d:37211 Zbl 05834866

[Milnor 2006] J. Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies **160**, Princeton University Press, 2006. MR 2006g:37070 Zbl 1085.30002

[Mumford and Fogarty 1982] D. Mumford and J. Fogarty, *Geometric invariant theory*, 2nd ed., Ergeb. Math. Grenzgeb. **34**, Springer, Berlin, 1982. MR 86a:14006 Zbl 0504.14008

[Petsche et al. 2009] C. Petsche, L. Szpiro, and M. Tepper, “Isotriviality is equivalent to potential good reduction for endomorphisms of $\mathbb{P}^N$ over function fields”, *J. Algebra* **322**:9 (2009), 3345–3365. MR 2011e:14023 Zbl 1190.14013

[Silverman 1998] J. H. Silverman, “The space of rational maps on $\mathbb{P}^1$”, *Duke Math. J.* **94**:1 (1998), 41–77. MR 2000m:14010 Zbl 0966.14031

[Silverman 2007] J. H. Silverman, *The arithmetic of dynamical systems*, Graduate Texts in Mathematics **241**, Springer, New York, 2007. MR 2008c:11002 Zbl 1130.37001

[Szpiro et al. 2010] L. Szpiro, M. Tepper, and P. Williams, “Resultant and conductor of geometrically semi-stable self maps of the projective line over a function field”, preprint, 2010. arXiv 1010.5030v1

[Zhang 1996] S. Zhang, “Heights and reductions of semi-stable varieties”, *Compositio Math.* **104**:1 (1996), 77–105. MR 97m:14027 Zbl 0924.11055

Communicated by Bjorn Poonen
Received 2011-06-15 Revised 2011-08-02 Accepted 2011-09-11

levy@math.columbia.edu *Department of Mathematics, Brown University, Providence, RI 02912, United States*

Department of Mathematics, Columbia University, New York, NY 10027, United States

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1503

Grothendieck's trace map for arithmetic surfaces via residues and higher adèles

Matthew Morrow

We establish the reciprocity law along a vertical curve for residues of differential forms on arithmetic surfaces, and describe Grothendieck's trace map of the surface as a sum of residues. Points at infinity are then incorporated into the theory and the reciprocity law is extended to all curves on the surface. Applications to adelic duality for the arithmetic surface are discussed.

1. Introduction

Grothendieck's trace map for a smooth, projective curve over a finite field can be expressed as a sum of residues over all closed points of the curve; see [Hartshorne 1977, III.7.14]. This result was generalised to algebraic surfaces by A. Parshin [1976] using his theory of two-dimensional adèles and residues for two-dimensional local fields. The theory for arbitrary-dimensional algebraic varieties is essentially contained in A. Beilinson's short paper [1980] on higher-dimensional adèles, with considerable additional work by J. Lipman [1984], V. Lomadze [1981], D. Osipov [1997], A. Yekutieli [1992], et al. In all these existing cases one restricts to varieties over a field. The purpose of this paper (together with [Morrow 2010]) is to provide the first extension of the theory to nonvarieties, namely to arithmetic surfaces, even taking into account the points "at infinity".

In the standard approach to Grothendieck duality of algebraic varieties using residues, there are three key steps. Firstly one must define suitable local residue maps, either on spaces of differential forms or on local cohomology groups (the latter approach is followed by E. Kunz [2008] using Grothendieck's residue symbol [Hartshorne 1966, III.§9]). Secondly, the local residue maps are used to define the dualising sheaf, and finally the local residue maps must be patched together to define Grothendieck's trace map on the cohomology of the dualising sheaf. In [Morrow 2010], we carried out most of the first two steps for arithmetic surfaces, as we now explain.

MSC2010: primary 14H25; secondary 14B15, 14F10.
Keywords: reciprocity laws, higher adèles, arithmetic surfaces, Grothendieck duality, residues.

Section 2 provides a detailed summary of the required results from [Morrow 2010], while also establishing several continuity and vanishing results which are required later. Briefly, given a two-dimensional local field F of characteristic zero and a fixed local field $K \leq F$, we introduced (see Section 2A) a relative residue map

$$\mathrm{Res}_F : \Omega^{\mathrm{cts}}_{F/K} \to K,$$

where $\Omega^{\mathrm{cts}}_{F/K}$ is a suitable space of "continuous" relative differential forms. In the case $F \cong K((t))$, this is the usual residue map; but if F is of mixed characteristic, then this residue map is new (though versions of it appear in I. Fesenko's two-dimensional adelic analysis [2010, §27, Proposition] and in D. Osipov's geometric counterpart [1997, Definition 5] to this paper). Then the reciprocity law for two-dimensional local rings was proved, justifying our definition of the relative residue map for mixed characteristic fields. For example, suppose A is a characteristic zero, two-dimensional, normal, complete local ring with finite residue field, and fix the ring of integers of a local field $\mathcal{O}_K \leq A$. To each height-one prime $y \subset A$, one associates the two-dimensional local field $\mathrm{Frac}\,\widehat{A_y}$ and thus obtains a residue map $\mathrm{Res}_y : \Omega^1_{\mathrm{Frac}\,A/K} \to K$ (see Section 2B). We showed

$$\sum_y \mathrm{Res}_y\,\omega = 0$$

for all $\omega \in \Omega^1_{\mathrm{Frac}\,A/K}$. The main new result in Section 2 is Lemma 2.8, stating that the residue map Res_y is continuous with respect to the $\mathfrak{m}$-adic topology on A.

Geometrically, if $\pi : X \to \mathrm{Spec}\,\mathcal{O}_K$ is an arithmetic surface and one chooses a closed point $x \in X$ and an irreducible curve $y \subset X$ passing through x, then one obtains a residue map

$$\mathrm{Res}_{x,y} : \Omega^1_{K(X)/K} \to K_{\pi(x)},$$

where $K_{\pi(x)}$ is the completion of K at the prime sitting under x (see Section 2D for details). The established reciprocity law now takes the form

$$\sum_{y\,:\,y \ni x} \mathrm{Res}_{x,y}\,\omega = 0,$$

where one fixes $\omega \in \Omega^1_{K(X)/K}$ and the summation is taken over all curves y passing through a fixed point x.

As discussed, the second step in a residue-theoretic approach to Grothendieck duality is a suitable description of the dualising sheaf. This was also given in [Morrow 2010]: if $\pi : X \to \mathrm{Spec}\,\mathcal{O}_K$ is an arithmetic surface (the precise requirements are those given at the start of Section 3), then the dualising sheaf ω_π of π

can be described as follows:

$$\omega_\pi(U)=\{\omega\in\Omega^1_{K(X)/K}:\operatorname{Res}_{x,y}(f\omega)\in\widehat{\mathcal{O}_{K,\pi(x)}}\text{ for all }x\in y\subset U\text{ and }f\in\mathcal{O}_{X,y}\}$$

where x runs over all closed points of X inside U and y runs over all curves of U containing x.

This paper treats the third step of the process. In order to patch the local residues together to define the trace map on cohomology, one must, just as in the basic case of a smooth, projective curve, establish certain reciprocity laws. For an arithmetic surface, these take the form

$$\sum_{y\,:\,y\ni x}\operatorname{Res}_{x,y}\omega=0,\qquad\sum_{x\,:\,x\in y}\operatorname{Res}_{x,y}\omega=0.$$

In both cases one fixes $\omega\in\Omega^1_{K(X)/K}$, but the first summation is taken over all curves passing through a fixed point x while the second summation is over all closed points of a fixed vertical curve y. The first of these laws, namely reciprocity around a point, has already been discussed, while Section 3 establishes the reciprocity law along a vertical curve: the key idea of the proof is to reduce to the case when $\mathcal{O}_K$ is a complete discrete valuation ring and then combine the reciprocity law around a point with the usual reciprocity law along the generic fibre.

Section 4 uses the Parshin–Beilinson higher adèles for coherent sheaves to express Grothendieck's trace map

$$\operatorname{tr}_\pi:H^1(X,\omega_\pi)\to\mathcal{O}_K$$

as a sum of the residue maps $(\operatorname{Res}_{x,y})_{x,y}$. Indeed, the reciprocity laws imply that our residue maps descend to cohomology: the argument is analogous to the case of a smooth, projective curve, except we must work with adèles for two-dimensional schemes rather than the more familiar adèles of a curve. Remark 4.11 explains the basic framework of the theory in arbitrary dimensions.

Whereas the material discussed above is entirely scheme-theoretic, the final part of the paper is the most important and interesting from an arithmetic perspective as it incorporates archimedean points (points at infinity). It is natural to ask whether there exists a reciprocity law for *all* curves on X, not merely the vertical ones, when $\mathcal{O}_K$ is the ring of integers of a number field. By compactifying $\operatorname{Spec}\mathcal{O}_K$ and X to include archimedean points in Section 5, we indeed prove a reciprocity law for any horizontal curve y on X. Owing to the nonexistence (at least naïvely) of $\operatorname{Spec}\mathbb{F}_1$, this takes the form

$$\prod_{x\,:\,x\in y}\psi_{x,y}(\omega)=1,$$

where $\psi_{x,y}:\Omega^1_{K(X)/K}\to S^1$ are *absolute residue maps* (additive characters) and ω lies in $\Omega^1_{K(X)/K}$. This provides detailed proofs of various claims made in [Fesenko

2010, §27, §28] concerning the foundations of harmonic analysis and adelic duality for arithmetic surfaces, and extends Parshin's absolute reciprocity laws for algebraic surfaces to the arithmetic case. Essentially this yields a framework which encodes both arithmetic duality of K and Grothendieck duality of $X \to S$, and which would be equivalent to Serre duality were X a geometric surface; a comparison of these results with Arakelov theory has yet to be carried out but there is likely an interesting connection.

Combined with [Morrow 2010], which should be seen as a companion to this article and which contains a much more extensive introduction to the subject, these results provide a theory of residues and explicit duality for arithmetic surfaces. The analogous theory for an algebraic surface fibred smoothly over a curve is due to Osipov [1997], who proved, using Parshin's reciprocity laws for an algebraic surface, the analogues of our reciprocity laws around a point and along a vertical curve, and also showed that the sum of residues induces the trace map on cohomology.

Notation. When differential forms appear in this paper, they will be 1-forms, so we write $\Omega_{A/R}$ in place of $\Omega^1_{A/R}$ to ease notation. Frac denotes the total ring of fractions; that is, if R is a commutative ring then $\operatorname{Frac} R = S^{-1}R$, where S is the set of regular elements in R. The maximal ideal of a local ring A is usually denoted $\mathfrak{m}_A$; an exception to this rule is when $A = \mathcal{O}_F$ is a discrete valuation ring with fraction field F, in which case we prefer the notation $\mathfrak{p}_F$.

When X is a scheme and $n \geq 0$, we write X^n for the set of codimension-n points of X. X_0 denotes the closed points of X. Typically, X will be two-dimensional, in which case we will often identify any $y \in X^1$ with the corresponding irreducible subscheme $\overline{\{y\}}$; moreover, "$x \in y$" then more precisely means that x is a codimension-1 point of $\overline{\{y\}}$. "Curve" usually means "irreducible curve". Given $z \in X$, the maximal ideal of the local ring $\mathcal{O}_{X,z}$ is written $\mathfrak{m}_{X,z}$.

$I \subset^1 A$ means that I is a height-one ideal of the ring A.

2. Relative residue maps in dimension two

In [Morrow 2010], a theory of residues on arithmetic surfaces was developed; we repeat here the main definitions and properties, also verifying several new results which will be required later.

2A. ***Two-dimensional local fields.*** Suppose first that F is a two-dimensional local field (that is, a complete discrete valuation field whose residue field $\overline{F}$ is a local field[1]) of characteristic zero, and that $K \leq F$ is a local field (this local field K will

[1] In this paper our local fields always have finite residue fields, though many of the calculations continue to hold in the case of perfect residue fields.

appear naturally in the geometric applications); write

$$\Omega^{\mathrm{cts}}_{F/K} = \Omega^{\mathrm{sep}}_{\mathcal{O}_F/\mathcal{O}_K} \otimes_{\mathcal{O}_F} F$$

(for a module over a local ring A, we write $M^{\mathrm{sep}} = M/\bigcap_{n\geq 0} \mathfrak{m}_A^n M$ for the maximal separated quotient of M). Let k_F be the algebraic closure of K inside F; this is a finite extension of K and hence is also a local field.

If F has equal characteristic then any choice of a uniformiser $t \in F$ induces a unique k_F-isomorphism $F \cong k_F((t))$, and $\Omega^{\mathrm{sep}}_{\mathcal{O}_F/\mathcal{O}_K} = \mathcal{O}_F\, dt$. The *relative residue map*, which does not depend on t, is the usual residue map which appears in the theory of curves over a field (e.g., [Serre 1988, II.7]):

$$\mathrm{res}_F : \Omega^{\mathrm{cts}}_{F/K} \to k_F, \quad f\, dt \mapsto \mathrm{coeft}_{t^{-1}} f,$$

where the notation means that f is to be expanded as a series in powers of t and the coefficient of t^{-1} is to be taken.

If F is a mixed characteristic two-dimensional local field then F/k_F is an infinite extension of complete discrete valuation fields, and F is called *standard* if and only if $e(F/k_F) = 1$. If F is standard then any choice of a first local parameter $t \in \mathcal{O}_F$ (that is, $\bar{t}$ is a uniformiser in the local field $\bar{F}$) induces a unique k_F-isomorphism $F \cong k_F\{\{t\}\}$ (defined to be the completion of $\mathrm{Frac}(\mathcal{O}_{k_F}[[t]])$ at the discrete valuation corresponding to the prime ideal $\mathfrak{p}_{k_F}\mathcal{O}_{k_F}[[t]]$; see [Morrow 2010, Example 2.10]), and $\Omega^{\mathrm{sep}}_{\mathcal{O}_F/\mathcal{O}_K} = \mathcal{O}_F\, dt \oplus \mathrm{Tors}\, \Omega^{\mathrm{sep}}_{\mathcal{O}_F/\mathcal{O}_K}$; so we may define

$$\mathrm{res}_F : \Omega^{\mathrm{cts}}_{F/K} \to k_F, \quad f\, dt \mapsto -\mathrm{coeft}_{t^{-1}} f,$$

which was shown in [ibid., Proposition 2.19] not to depend on the choice of t. (The notation again means that f is to be expanded as a series in powers in t, but this time in the field $k_F\{\{t\}\}$, and the coefficient of t^{-1} taken). If F is not necessarily standard, then choose a subfield $M \leq F$ which is a standard two-dimensional local field such that F/M is a finite extension, and which satisfies $k_M = k_F$. The *relative residue map* in this case is defined by

$$\mathrm{res}_F = \mathrm{res}_M \circ \mathrm{Tr}_{F/M} : \Omega^{\mathrm{cts}}_{F/K} \to k_F,$$

which was shown in [ibid., Lemma 2.21] not to depend on M.

In both cases, it is also convenient to write $\mathrm{Res}_F = \mathrm{Tr}_{k_F/K} \circ \mathrm{res}_F : \Omega^{\mathrm{cts}}_{F/K} \to K$. Also note that res_F is k_F-linear, and that therefore Res_F is K-linear. The expected functoriality result holds:

Lemma 2.1. *Let L be a finite extension of K. Then $\Omega^{\mathrm{cts}}_{L/K}$ is naturally isomorphic to $\Omega^{\mathrm{cts}}_{F/K} \otimes_F L$, so that there is a trace map $\mathrm{Tr}_{L/F} : \Omega^{\mathrm{cts}}_{L/K} \to \Omega^{\mathrm{cts}}_{F/K}$. If $\omega \in \Omega^{\mathrm{cts}}_{L/K}$, then*

$$\mathrm{Res}_F(\mathrm{Tr}_{L/F}\, \omega) = \mathrm{Res}_L\, \omega \quad \text{in } K.$$

Proof. In the equal characteristic case this is classical; see, for example, [Serre 1988, II.12 Lemma 5]. For the mixed characteristic case, see [Morrow 2010, Proposition 2.22]. □

Next we show a couple of results on the continuity of residues which, though straightforward, will be frequently employed. Lemma 2.8 is a stronger, similar result.

Lemma 2.2. *Suppose that $\omega \in \Omega^{\mathrm{cts}}_{F/K}$ is integral, that is, belongs to the image of $\Omega^{\mathrm{sep}}_{\mathbb{O}_F/\mathbb{O}_K}$. Then $\mathrm{res}_F\,\omega \in \mathbb{O}_{k_F}$ and so $\mathrm{Res}_F\,\omega \in \mathbb{O}_K$; in fact, if F is equal characteristic, then $\mathrm{res}_F\,\omega = 0$.*

Proof. In the equal characteristic or standard case this follows immediately from the definitions. In the nonstandard, mixed characteristic case, one picks a standard subfield M as above and uses a classical formula for the different of $\mathbb{O}_F/\mathbb{O}_M$ to show that the trace map $\Omega^{\mathrm{cts}}_{F/K} \to \Omega^{\mathrm{cts}}_{M/K}$ may be pulled back to $\Omega^{\mathrm{sep}}_{\mathbb{O}_F/\mathbb{O}_K} \to \Omega^{\mathrm{sep}}_{\mathbb{O}_M/\mathbb{O}_K}$, from which the result follows. See [Morrow 2010, §2.3.4] for the details. □

Remark 2.3. It was also shown in [ibid., Corollary 2.23] that, when F has mixed characteristic, the following diagram commutes:

$$\begin{tikzcd} \Omega^{\mathrm{sep}}_{\mathbb{O}_F/\mathbb{O}_K} \arrow[r, "\mathrm{Res}_F"] \arrow[d] & \mathbb{O}_K \arrow[d] \\ \Omega_{\overline{F}/\overline{K}} \arrow[r, "e(F/K)\,\mathrm{Res}_{\overline{F}}"'] & \overline{K} \end{tikzcd}$$

The top horizontal arrow here makes sense by the previous lemma, and the lower horizontal arrow is the ramification degree $e(F/K)$ times the residue map for the local field $\overline{F}$ of finite characteristic, which contains the finite field $\overline{K}$.

Corollary 2.4. *Fix $\omega \in \Omega^{\mathrm{cts}}_{F/K}$. Then*

$$F \to K, \quad f \mapsto \mathrm{Res}_F(f\omega)$$

is continuous with respect to the discrete valuation topologies on F and K; in fact, if F is equal characteristic, then it is even continuous with respect to the discrete topology on K.

Proof. After multiplying ω by a nonzero element of F, we may assume that ω is integral in the sense of the previous lemma. If F is equal characteristic then $\mathrm{Ker}(f \mapsto \mathrm{res}_F(f\omega))$ contains the open set $\mathbb{O}_F$, proving continuity with respect to the discrete topology on K. Now assume F has mixed characteristic and let π be a uniformiser of K; since F/K is an extension of complete discrete valuation fields, we may put $e = e(F/K) = \nu_F(\pi) > 0$. Then the previous lemma implies

$$\mathrm{Res}_F(\mathfrak{p}_F^{es}\omega) = \mathrm{Res}_F(\pi^s\mathbb{O}_F\omega) = \pi^s\,\mathrm{Res}_F(\mathbb{O}_F\omega) \subseteq \mathfrak{p}_K^s$$

for all $s \in \mathbb{Z}$, proving continuity with respect to the discrete valuation topologies. $\square$

2B. *Two-dimensional complete rings.* Let A be a two-dimensional, normal, complete, local ring of characteristic zero, with a finite residue field of characteristic p; set $F = \operatorname{Frac} A$. Then there is a unique ring homomorphism $\mathbb{Z}_p \to A$ and it is a closed embedding; let $\mathbb{O}_K$ be a finite extension of $\mathbb{Z}_p$ inside A; that is, $\mathbb{O}_K$ is the ring of integers of K, which is a finite extension of $\mathbb{Q}_p$.

If $y \subset A$ is a height-one prime (we often write $y \subset^1 A$), then $\widehat{A}_y$ is a complete discrete valuation ring; its field of fractions $F_y := \operatorname{Frac} \widehat{A}_y$ is a two-dimensional local field containing K. Moreover, there is a natural isomorphism

$$\Omega^{\mathrm{sep}}_{A/\mathbb{O}_K} \otimes_A \widehat{A}_y \cong \Omega^{\mathrm{sep}}_{\widehat{A}_y/K}$$

(see [ibid., Lemma 3.8]); so we define $\operatorname{Res}_y : \Omega^{\mathrm{sep}}_{A/\mathbb{O}_K} \otimes_A F \to K$ to be the composition

$$\Omega^{\mathrm{sep}}_{A/\mathbb{O}_K} \otimes_A F \longrightarrow \Omega^{\mathrm{sep}}_{A/\mathbb{O}_K} \otimes_A F_y \cong \Omega^{\mathrm{cts}}_{F_y/K} \xrightarrow{\operatorname{Res}_{F_y}} K.$$

The definition of the residue maps is justified by the following reciprocity law:

Theorem 2.5. *Let $\omega \in \Omega^{\mathrm{sep}}_{A/\mathbb{O}_K} \otimes_A F$; then for all but finitely many height-one primes $y \subset A$ the residue $\operatorname{Res}_y \omega$ is zero, and*

$$\sum_{y \subset^1 A} \operatorname{Res}_y \omega = 0.$$

Proof. See [ibid., Theorem 3.10]. $\square$

As is often the case, the residue law was reduced to a special case by taking advantage of functoriality:

Lemma 2.6. *Suppose that C is a finite extension of A which is also normal; set $L = \operatorname{Frac} C$. Then for any $\omega \in \Omega^{\mathrm{sep}}_{C/\mathbb{O}_K} \otimes_C L$ and any height-one prime $y \subset A$, we have*

$$\operatorname{Res}_y(\operatorname{Tr}_{L/F} \omega) = \sum_{Y|y} \operatorname{Res}_Y \omega,$$

where Y varies over the finitely many height-one primes of C which sit over y.

Proof. See [ibid., Theorem 3.9]. $\square$

The proof of the reciprocity theorem also required certain results on the continuity of the residues whose proofs were omitted in [ibid.]; we shall require similar such results several times in this article and now is a convenient opportunity to establish them:

Lemma 2.7. *Set* $B = \mathbb{O}_K[\![t]\!]$, $M = \operatorname{Frac} B$ *and let* $\omega \in \Omega^{\mathrm{sep}}_{B/\mathbb{O}_K} \otimes_B M$; *then, for any height-one prime* $y \subset B$, *the map*

$$B \to K, \quad f \mapsto \operatorname{Res}_y f\omega$$

is continuous with respect to the $\mathfrak{m}_B$*-adic topology on* B *and the discrete valuation topology on* K.

Proof. We first consider the case when $y = \rho B$ is generated by an irreducible Weierstrass polynomial $\rho(t) \in \mathbb{O}_K[t]$. Let K' be a sufficiently large finite extension of K such that ρ splits into linear factors in K'; the decomposition has the form $\rho(t) = \prod_{i=1}^d (t-\lambda_i)$ with $d = \deg \rho$ and $\lambda_i \in \mathfrak{p}_{K'}$ since h is a Weierstrass polynomial. Put $B' = \mathbb{O}_{K'}[\![t]\!]$ and $M' = \operatorname{Frac} B'$. According to functoriality of residues (the previous lemma), we have

$$\operatorname{Res}_y \operatorname{Tr}_{M'/M} \omega = \sum_{i=1}^d \operatorname{Res}_{Y_i} \omega$$

for all $\omega \in \Omega^{\mathrm{sep}}_{B'/\mathbb{O}_K} \otimes_{B'} M'$, where $Y_i = (t - \lambda_i)B'$. Since multiplication by $f \in B$ commutes with the trace map, it is now enough to prove that

$$B' \to K, \quad f \mapsto \operatorname{Res}_{Y_i} f\omega$$

is continuous for all i and all $\omega \in \Omega^{\mathrm{sep}}_{B'/\mathbb{O}_K} \otimes_{B'} M'$. In other words, replacing K by K' and B by B', we have reduced to the case when $\rho(t)$ is a linear polynomial: $\rho(t) = t - \lambda$, with $\lambda \in \mathfrak{p}_K$. After another reduction, we will prove the continuity claim in this case.

Let π be a uniformiser for K. It is well-known that $\Omega^{\mathrm{sep}}_{B/\mathbb{O}_K} = B\,dt$ and that any element of M can be written as a finite sum of terms of the form

$$\frac{\pi^n g}{h^r},$$

with $h \in \mathbb{O}_K[t]$ an irreducible Weierstrass polynomial, $r > 0$, $n \in \mathbb{Z}$, and $g \in B$ (a proof was given in [Morrow 2010, Lemma 3.4]). By continuity of addition $K \times K \xrightarrow{+} K$ and of the multiplication maps $B \xrightarrow{\times g} B$, $K \xrightarrow{\times \pi^n} K$, it is enough to treat the case $\omega = h^{-r}\,dt$, where $h \in \mathbb{O}_K[t]$ is an irreducible Weierstrass polynomial.

Now return to $y = \rho B$, $\rho = t - \lambda$. If $h \neq \rho$, then $h^{-r}dt \in \Omega^{\mathrm{sep}}_{B/\mathbb{O}_K} \otimes_B B_y$, and so $\operatorname{Res}_y(B\omega) = 0$ by Lemma 2.2, which is certainly enough. Else $h = \rho$, which we now consider. To obtain more suggestive notation, we write $t_y := \rho(t) = t - \lambda$; thus

$$\omega = h^{-r}\,dt = t_y^{-r}\,dt_y.$$

Let $m \geq 0$; we claim that if $n \geq m + r$ then $\operatorname{Res}_y(\mathfrak{m}_B^n \omega) \subseteq \mathfrak{p}_K^m$. Since λ is divisible by π, the maximal ideal of B is generated by π and t_y: $\mathfrak{m}_B = \langle \pi, t \rangle = \langle \pi, t_y \rangle$. Therefore an arbitrary element of $\mathfrak{m}_B^n$ is a sum of terms of the form $\pi^\alpha t_y^\beta g$, with

$g \in B$, $\alpha, \beta \geq 0$, and $\alpha+\beta \geq n$, and so it is enough to consider such an element. Moreover, again since π divides λ, there is a unique continuous isomorphism

$$\mathbb{O}_K[\![t_y]\!] \xrightarrow{\sim} \mathbb{O}_K[\![t]\!], \quad t_y \mapsto t-\lambda,$$

and therefore $g \in B$ may be written as $g = \sum_{j=0}^{r-1} a_j t_y^j + t_y^r g_1$ with $a_j \in \mathbb{O}_K$ and $g_1 \in B$ (we could extend this expansion to infinity, of course, but since we are trying to prove continuity, it is better not to risk confusion between "formal series" and "convergent series"). Then

$$\operatorname{Res}_y(\pi^\alpha t_y^\beta g\omega) = \pi^\alpha \operatorname{Res}_y\left(t_y^{\beta-r}\sum_{j=0}^{r-1} a_j t_y^j \, dt_y\right) + \pi^\alpha \operatorname{Res}_y(t_y^\beta g_1 \, dt_y). \qquad (\dagger)$$

The second residue is zero by Lemma 2.2 again since $t_y^\beta g_1 \in B$. If $\beta \geq r$ then the first residue is zero for the same reason; but if $\beta < r$ then it follows that $\alpha > m$, whence the first residue is $\pi^\alpha a_{r-\beta-1} \in \mathfrak{p}_K^\alpha \subseteq \mathfrak{p}_K^m$. So in any case, ($\dagger$) belongs to $\mathfrak{p}_K^m$, completing the proof of our claim and thereby showing the desired continuity result for $y = \rho B$.

Having treated the case of a prime y generated by a Weierstrass polynomial, we must secondly consider $y = \pi B$. By exactly the same argument as above, we may assume that $\omega = h^{-r}\, dt$, with h an irreducible Weierstrass polynomial. Then $M_y = K\{\{t\}\}$ and $h^{-r} \in B_y$; hence h^{-r} may be written as a series

$$h^{-r} = \sum_{j \in \mathbb{Z}} a_j t^j \in \mathbb{O}_K\{\{t\}\}$$

where $a_j \to 0$ in $\mathbb{O}_K$ as $j \to -\infty$. Let $m \geq 0$ be fixed, and pick $J > 2$ such that $a_j \in \mathfrak{p}_K^m$ whenever $j \leq -J$. We claim that if $n \geq J-2+m$ then $\operatorname{Res}_y(\mathfrak{m}_B^n \omega) \subseteq \mathfrak{p}_K^m$. Since an arbitrary element of $\mathfrak{m}_B^n$ is a sum of terms of the form $\pi^\alpha t^\beta g$, with $g \in B$, $\alpha, \beta \geq 0$, and $\alpha+\beta \geq n$, it is enough it consider such an element; write $g = \sum_{i=0}^\infty b_i t^i$. Then

$$\begin{aligned}\operatorname{Res}_y(\pi^\alpha t^\beta g\omega) = \operatorname{Res}_y(\pi^\alpha t^\beta g h^{-r}\, dt) &= -\pi^\alpha \operatorname{coeft}_{t^{-1}}\left(t^\beta \sum_{i=0}^\infty b_i t^i \sum_{j\in\mathbb{Z}} a_j t^j\right)\\ &= -\pi^\alpha \sum_{i=0}^\infty b_i a_{-i-\beta-1} \in \begin{cases}\mathfrak{p}_K^{\alpha+m} & \text{if } \beta \geq J-2\\ \mathfrak{p}_K^\alpha & \text{in any case.}\end{cases}\end{aligned}$$

But $\alpha+\beta \geq J-2+m$ and so if it is not the case that $\beta \geq J-2$, then it follows that $\alpha \geq m$; so, regardless of which inequality holds, we obtain $\operatorname{Res}_y(\pi^\alpha t^\beta g\omega) \in \mathfrak{p}_K^m$, as required. $\square$

Now we extend the lemma to the general case of our two-dimensional, normal, complete, local ring A. This result is a significant strengthening of Corollary 2.4, since the $\mathfrak{m}_A$-adic topology on A is considerably finer than the y-adic topology, for any $y \subset^1 A$.

Lemma 2.8. *Let $\omega \in \Omega^{\mathrm{sep}}_{A/\mathbb{O}_K} \otimes_A F$; then, uniformly in y, the map*

$$A \to K, \quad f \mapsto \mathrm{Res}_y f\omega$$

is continuous with respect to the $\mathfrak{m}_A$-adic topology on A and the discrete valuation topology on K.

Proof. Firstly, it is enough to prove that the given map is continuous for any fixed y; the uniformity result then follows from the fact that, for almost all $y \subset^1 A$, ω belongs to $\Omega^{\mathrm{sep}}_{A_y/\mathbb{O}_K}$ and y does not contain $\mathfrak{p}_K$; for such primes, $\mathrm{Res}_y A\omega = 0$ by Lemma 2.2.

By Cohen structure theory [1946] (the details of the argument are in [Morrow 2010, Lemma 3.3]), there is a subring $B \leq A$ containing $\mathbb{O}_K$ which is isomorphic to $\mathbb{O}_K[[t]]$ and such that A is a finitely generated B-module; set $M = \mathrm{Frac}\, B$. Write $\omega = g\omega_0$ for some $g \in F$ and $\omega_0 \in \Omega^{\mathrm{sep}}_{B/\mathbb{O}_K} \otimes_B M$.

Now we make some remarks on continuity of the trace map. $\mathrm{Tr}_{F/M}(Ag)$ is a finitely generated B-module and so there exists $g_0 \in M^\times$ such that $\mathrm{Tr}_{F/M}(Ag) \subseteq Bg_0$. Moreover, since A/B is a finite extension of local rings, one has $\mathfrak{m}_A^s \subseteq \mathfrak{m}_B A$ for some $s > 0$. Hence $\mathrm{Tr}_{F/M}(\mathfrak{m}_A^{ns} g) \subseteq \mathfrak{m}_B^n g_0$ for all $n \geq 0$, meaning that the restriction of the trace map to $Ag \to Bg_0$ is continuous with respect to the $\mathfrak{m}$-adic topologies on each side. It immediately follows that

$$\tau : A \to B, \quad f \mapsto \mathrm{Tr}_{F/M}(fg)g_0^{-1}$$

is both well defined and continuous.

Functoriality (Lemma 2.6) implies that for any $y \subset^1 B$,

$$\sum_{Y|y} \mathrm{Res}_Y f\omega = \mathrm{Res}_y \mathrm{Tr}_{F/M}(f\omega)$$

for all $f \in A$, where Y varies over the finitely many height-one primes of A which sit over y. The right side may be rewritten as

$$\mathrm{Res}_y(\tau(f)\, g_0\omega_0)$$

where $g_0\omega_0 \in \Omega^{\mathrm{sep}}_{B/\mathbb{O}_K} \otimes_B M$; according to the previous lemma, this is a continuous function of f. In conclusion,

$$A \to K, \quad f \mapsto \sum_{Y|y} \mathrm{Res}_Y f\omega \qquad (\dagger)$$

is continuous, which we will now use to show that each map $f \mapsto \mathrm{Res}_Y f\omega$ is individually continuous, thereby completing the proof. Fix $m \geq 0$.

Let $Y_1, \dots, Y_l$ be the height-one primes of A sitting over y, and let $\nu_1, \dots, \nu_l$ denote the corresponding discrete valuations of F. If $l = 1$ then there is nothing more to show, so assume $l > 1$. Since the map

$$F_{Y_i} \to K, \quad f \mapsto \mathrm{Res}_{Y_i}(f\omega)$$

is continuous with respect to the discrete valuation topologies on each side (Corollary 2.4), there exists $S > 0$ (which we may obviously assume is independent of i) such that $\mathrm{Res}_{Y_i}(f\omega) \subseteq \mathfrak{p}_K^m$ whenever $\nu_i(f) \geq S$. According to the approximation theorem for discrete valuations, there exists an element $e \in F$ which satisfies $\nu_1(e-1) \geq S$ and $\nu_i(e) \geq S$ for $i = 2, \dots, l$. Now, since (†) remains continuous if we replace ω by $e\omega$, there also exists $J > 0$ such that

$$\sum_{Y|y} \mathrm{Res}_Y(fe\omega) \in \mathfrak{p}_K^m \quad \text{whenever } f \in \mathfrak{m}_A^J.$$

So, if $f \in \mathfrak{m}_A^J$ then

$$\mathrm{Res}_{Y_1}(f\omega) = \mathrm{Res}_{Y_1}(f(1-e)\omega) - \sum_{i=2}^{l} \mathrm{Res}_{Y_i}(fe\omega) + \sum_{i=1}^{l} \mathrm{Res}_{Y_i}(fe\omega)$$

belongs to $\mathfrak{p}_K^m$ since $\nu_1(f(1-e)) \geq S$ and $\nu_i(fe) \geq S$ for $i = 2, \dots, l$. That is, $\mathrm{Res}_{Y_1}(\mathfrak{m}_A^J\omega) \subseteq \mathfrak{p}_K^m$, which proves the desired continuity result. □

Remark 2.9. Lemma 2.8 can be reformulated as saying that the residue map

$$\mathrm{Res}_{F_y} : \Omega^{\mathrm{cts}}_{F_y/K} \to K$$

is continuous with respect to the valuation topology on K and the vector space topology on $\Omega^{\mathrm{cts}}_{F_y/K}$, where F_y is equipped with its two-dimensional local field topology [Madunts and Zhukov 1995].

Finally, regarding vanishing of the residue of a differential form:

Lemma 2.10. *Suppose that $\omega \in \Omega^{\mathrm{sep}}_{A/\mathcal{O}_K} \otimes_A F$ is integral, in the sense that it belongs to the image of $\Omega^{\mathrm{sep}}_{A/\mathcal{O}_K}$, and let $y \subset^1 A$. Then $\mathrm{Res}_y\, \omega \in \mathfrak{p}_K$. If y does not contain p or if y is the only height-one prime of A containing p, then $\mathrm{Res}_y\, \omega = 0$.*

Proof. If y does not contain p then F_y is equal characteristic and we have already proved a stronger result in Lemma 2.2: Res_y vanishes on the image of $\Omega^{\mathrm{sep}}_{A/\mathcal{O}_K} \otimes_A A_y$. If instead y is the only height-one prime of A containing p, then the vanishing claim follows from the reciprocity law and the previous case.

Finally, suppose y contains p but do not assume that it is the only height-one prime to do so. Using functoriality of differential forms and Remark 2.3, we have a commutative diagram:

$$\begin{tikzcd}
\Omega^{\mathrm{sep}}_{A/\mathcal{O}_K} \arrow[r] \arrow[d] & \Omega^{\mathrm{sep}}_{\mathcal{O}_{F_y}/\mathcal{O}_K} \arrow[r, "\mathrm{Res}_{F_y}"] \arrow[d] & \mathcal{O}_K \arrow[d] \\
\Omega_{(A/y)/\overline{K}} \arrow[r] & \Omega_{\overline{F}_y/\overline{K}} \arrow[r, "e(F_y/K)\,\mathrm{Res}_{\overline{F}_y}"'] & \overline{K}
\end{tikzcd}$$

The residue map $\mathrm{Res}_{\overline{F}_y}$ on the characteristic p local field $\overline{F}_y$ vanishes on integral differential forms; since A/y belongs to the ring of integers of $\overline{F}_y$, it follows immediately from the diagram that $\mathrm{Res}_y\,\omega \in \mathfrak{p}_K$. □

Example 2.11. This example will show that the previous lemma cannot be improved. We consider the "simplest" A in which p splits. Set $B = \mathbb{Z}_p[[T]]$, with field of fractions M, and let $A = B[\alpha]$ where α is a root of $f(X) = X^2 - TX - p$, with field of fractions F. Since $f(X)$ does not have a root in $B/TB = \mathbb{Z}_p$, it does not have a root in B, and so F/M is a degree two extension. Since A is a finitely generated B-module, it is also a two-dimensional, complete local ring, and we leave it to the reader to check that A is regular, hence normal.

In A, p completely splits as $p = \alpha(T-\alpha)$, and therefore, setting $y = \alpha A$, the natural map $\mathbb{Q}_p\{\{T\}\} = M_{pB} \to F_y$ is an isomorphism. Indeed, $f(X)$ splits in the residue field $B_{pB}/pB_{pB} = \mathbb{F}_p((T))$ into distinct factors and so Hensel's lemma implies that $f(X)$ splits in $\widehat{B_{pB}}$; that is, $\alpha \in \widehat{B_{pB}} \subset M_{pB}$.

One readily checks that $\alpha \equiv -pT^{-1} \bmod p^2$ in $\widehat{B_{pB}} = \widehat{A}_y$, which implies that $\mathrm{Res}_y(\alpha\, dT) \equiv -p \bmod p^2$. In particular, $\mathrm{Res}_y(\alpha\, dT) \neq 0$ even though $\alpha\, dT$ is integral.

2C. *Two-dimensional, finitely generated rings.* Next suppose that $\mathcal{O}_K$ is a Dedekind domain of characteristic zero and with finite residue fields, and that B is a two-dimensional, normal, local ring, which we assume is the localisation of a two-dimensional, finitely generated $\mathcal{O}_K$-algebra. Set $A = \widehat{B_{\mathfrak{m}_B}}$ and $s = \mathfrak{m}_B \cap \mathcal{O}_K$. Then A satisfies all the conditions introduced at the start of the previous subsection and contains $\mathcal{O}_s := \widehat{\mathcal{O}_{K,s}}$, which is the ring of integers of the local field $K_s := \mathrm{Frac}\,\widehat{\mathcal{O}_{K,s}}$. Moreover, there is a natural identification $\Omega_{B/\mathcal{O}_K} \otimes_B A = \Omega^{\mathrm{sep}}_{A/\mathcal{O}_s}$ (see [Morrow 2010, Lemma 3.11]). For each height-one prime $y \subset B$, we may therefore define

$$\mathrm{Res}_y : \Omega_{\mathrm{Frac}\,B/K} \to K_s$$

to be the composition

$$\Omega_{\mathrm{Frac}\,B/K} \longrightarrow \Omega_{\mathrm{Frac}\,B/K} \otimes_{\mathrm{Frac}\,B} \mathrm{Frac}\,A \cong \Omega^{\mathrm{sep}}_{A/\mathcal{O}_s} \otimes_A \mathrm{Frac}\,A \xrightarrow{\sum_{y'|y} \mathrm{Res}_{y'}} K_s,$$

where y' varies over the finitely many primes of A, necessarily of height one, which sit over y.

The reciprocity law remains true in this setting:

Theorem 2.12 [Morrow 2010, Theorem 3.13]. *Let $\omega \in \Omega_{\operatorname{Frac} B/K}$; then for all but finitely many height-one primes $y \subset B$ the residue* $\operatorname{Res}_y \omega$ *is zero, and*

$$\sum_{y \subset^1 B} \operatorname{Res}_y \omega = 0.$$

The following vanishing identity will be useful:

Lemma 2.13. *Let $y \subset^1 B$ and suppose that $\omega \in \Omega_{\operatorname{Frac} B/K}$ belongs to the image of $\Omega_{B_y/\mathcal{O}_K}$. Then* $\operatorname{Res}_y \omega \in \mathcal{O}_s$. *In fact,* $\operatorname{Res}_y \omega = 0$ *in either of the following two cases: if y is horizontal (that is, $y \cap \mathcal{O}_K = 0$); or if y is the only height-one prime of B which is vertical (that is, containing s) and ω is in the image of $\Omega_{A/\mathcal{O}_K}$.*

Proof. The first claims follow from Lemma 2.2, since y being horizontal is equivalent to the two-dimensional local fields $\operatorname{Frac} \widehat{A_{y'}}$, with $y' \subset A$ sitting over y, being equicharacteristic. The second claim follows from the previous reciprocity law since any prime is either vertical or horizontal. □

2D. *Geometrisation.* Continue to let $\mathcal{O}_K$ be a Dedekind domain of characteristic zero and with finite residue fields. Let X be a two-dimensional, normal scheme, flat and of finite type over $S = \operatorname{Spec} \mathcal{O}_K$, and let $\Omega_{X/S} = \Omega^1_{X/S}$ be the relative sheaf of one forms. Let $x \in X^2$ be a closed point sitting over a closed point $s \in S_0$, and let $y \subset X$ be an irreducible curve containing x. Identify y with its local equation (that is, corresponding prime ideal) $y \subset^1 \mathcal{O}_{X,x}$ and note that $\mathcal{O}_{X,x}$ satisfies all the conditions which B did in the previous subsection. Define the residue map $\operatorname{Res}_{x,y} : \Omega_{K(X)/K} \to K_s$ $(= \operatorname{Frac} \widehat{\mathcal{O}_{K,s}})$ to be

$$\operatorname{Res}_y : \Omega_{\operatorname{Frac} \mathcal{O}_{X,x}/K} \longrightarrow K_s.$$

The reciprocity law now states that, for any fixed $\omega \in \Omega_{K(X)/K}$,

$$\sum_{\substack{y \subset X \\ y \ni x}} \operatorname{Res}_{x,y} \omega = 0$$

in K_s, where the sum is taken over all curves in X which pass through x. For a few more details, see [Morrow 2010, §4].

3. Reciprocity along vertical curves

As explained in the introduction, residues on a surface should satisfy *two* reciprocity laws, one as we vary curves through a fixed point, and another as we vary points along a fixed curve. The first was explained immediately above and now we will prove the second.

Let $\mathcal{O}_K$ be a Dedekind domain of characteristic zero and with finite residue fields; denote by K its field of fractions. Let X be an $\mathcal{O}_K$-curve; more precisely, X is a normal scheme, proper and flat over $S = \operatorname{Spec} \mathcal{O}_K$, whose generic fibre is a smooth, geometrically connected curve.

The aim of this section is to establish the following reciprocity law for vertical curves on X:

Theorem 3.1. *Let $\omega \in \Omega_{K(X)/K}$, and let $y \subset X$ be an irreducible component of a special fibre X_s, where $s \in S_0$. Then*

$$\sum_{x \in y} \operatorname{Res}_{x,y} \omega = 0$$

in K_s, where the sum is taken over all closed points x of y.

Here, as usual, $\mathcal{O}_s = \widehat{\mathcal{O}_{K,s}}$ and $K_s = \operatorname{Frac} \mathcal{O}_s$. The proof will consist of several steps. We begin with a short proof of a standard adelic condition:

Lemma 3.2. *Let $y \subset X$ be an irreducible curve, let $f \in \mathcal{O}_{X,y}$, and let $r \geq 1$. Then $f \in \mathcal{O}_{X,x} + \mathfrak{m}_{X,y}^r$ for all but finitely many closed points $x \in y$.*

The result also holds after completion: if $f \in \widehat{\mathcal{O}_{X,y}}$, then $f \in \mathcal{O}_{X,x} + \mathfrak{m}_{X,y}^r \widehat{\mathcal{O}_{X,y}}$ for almost all x.

Proof. Let $U = \operatorname{Spec} A$ be an open affine neighbourhood of (the generic point of) y, let $\mathfrak{p} \subset A$ be the prime ideal defining y, and set $P = A \cap \mathfrak{p}^r A_{\mathfrak{p}}$, $B = A/P$. If $b \in B$ is not a zero divisor, then B/bB is zero-dimensional and so has only finitely many primes; hence only finitely many primes of B contain b. Set

$$\bar{f} := f \bmod \mathfrak{m}_{X,y}^r \in A_{\mathfrak{p}}/\mathfrak{p}^r A_{\mathfrak{p}} = \operatorname{Frac} B;$$

by what we have just proved, $\bar{f}$ belongs to $B_{\mathfrak{q}}$ for all but finitely many primes $\mathfrak{q} \subset B$, that is $f \in \mathcal{O}_{X,x} + \mathfrak{m}_{X,y}^r$ for all but finitely many $x \in y \cap U$. Since U contains all but finitely many points of y, we have finished.

The complete version now follows from the identity

$$\widehat{\mathcal{O}_{X,y}}/\mathfrak{m}_{X,y}^r \widehat{\mathcal{O}_{X,y}} = \mathcal{O}_{X,y}/\mathfrak{m}_{X,y}^r. \qquad \square$$

The lemma lets us prove that the theorem makes sense:

Lemma 3.3. *Let $\omega \in \Omega_{K(X)/K}$, and let $y \subset X$ be an irreducible component of a special fibre X_s, where $s \in S_0$. Then the sum $\sum_{x \in y} \operatorname{Res}_{x,y} \omega$ converges in the s-adic valuation topology on K_s (we will see that only countably many terms are nonzero).*

Moreover, the map

$$K(X) \to K_s, \quad h \mapsto \sum_{x \in y} \operatorname{Res}_{x,y}(h\omega)$$

is continuous with respect to the topology on $K(X)$ induced by the discrete valuation ν associated to y, and the s-adic valuation topology on K_s.

Proof. For any point $z \in X$, let Ω_z denote the image of $\Omega_{\mathbb{O}_{X,z}/\mathbb{O}_K}$ inside $\Omega_{K(X)/K}$. Let $r \geq 0$.

Let $\pi \in \mathbb{O}_K$ be a uniformiser at s, fix $\omega \in \Omega_{K(X)/K}$ and pick $a \geq 0$ such that $\pi^a \omega \in \Omega_y$. Then it easily follows from the previous lemma that $\pi^a\omega$ lies in $\Omega_x + \pi^r \Omega_y$ for almost all closed points $x \in y$. But Lemma 2.13 implies that if x is any closed point of y then $\mathrm{Res}_{x,y}(\Omega_y) \subseteq \mathbb{O}_s$, and moreover that if x does not lie on any other irreducible component of the fibre X_s then $\mathrm{Res}_{x,y}(\Omega_x) = 0$. We deduce that

$$\mathrm{Res}_{x,y}\, \pi^a \omega \in \pi^r \mathbb{O}_s$$

for almost all closed points $x \in y$. So $\mathrm{Res}_{x,y}\, \omega \in \pi^{r-a}\mathbb{O}_s$ for almost all $x \in y$; since this holds for all $r \geq 0$ we see that

$$\sum_{x \in y} \mathrm{Res}_{x,y}\, \omega$$

converges and also that $\sum_{x\in y} \mathrm{Res}_{x,y}\, \omega \in \pi^{-a}\mathbb{O}_s$.

If $h \in K(X)$ satisfies $\nu(h) \geq b$ for some $b \in \mathbb{Z}$, then we may write $h = \pi^b u$ for some $u \in \mathbb{O}_{X,y}$. This implies that $\pi^{a-b} h\omega \in \Omega_y$ and so, by what we have just shown, $\sum_{x\in y} \mathrm{Res}_{x,y}\, h\omega \in \pi^{b-a}\mathbb{O}_s$. This proves that $h \mapsto \sum_{x\in y} \mathrm{Res}_{x,y}\, h\omega$ is continuous. □

Remark 3.4. The analogous vertical reciprocity law in the geometric setting is [Osipov 1997, Proposition 6], where Osipov gives an example to show that it really is possible for the sum of residues along the points of $y \subset X_s$ to contain infinitely many nonzero terms.

We aim to reduce the vertical reciprocity law to the case of $\mathbb{O}_K$ being a complete discrete valuation ring by using several lemmas on the functoriality of residues.

Let s be a nonzero prime of $\mathbb{O}_K$, and set $\mathbb{O}_s = \widehat{\mathbb{O}_{K,s}}$, $K_s = \mathrm{Frac}\, \mathbb{O}_s$ as usual. Set $\widehat{X} = X \times_{\mathbb{O}_K} \mathbb{O}_s$ and let $p : \widehat{X} \to X$ be the natural map. Then p induces an isomorphism of the special fibres $\widehat{X}_s \cong X_s$ and, for any point $x \in X_s$, p induces an isomorphism of the completed local rings $\widehat{\mathbb{O}_{X,p(x)}} \cong \widehat{\mathbb{O}_{\widehat{X},x}}$ (see, e.g., [Liu 2002, Lemma 8.3.49]). From the excellence of X it follows that $\mathbb{O}_{\widehat{X},x}$ is normal for all $x \in \widehat{X}_s$, and therefore $\widehat{X}$ is normal. So $\widehat{X}$ is a $\mathbb{O}_s$-curve, in the same sense as at the start of the section.

Lemma 3.5. *Let $y \subset X$ be an irreducible curve and suppose x is a closed point of y over s. Then the following diagram commutes*:

$$\begin{tikzcd}
\Omega_{K(\widehat{X})/K_s} \arrow[dr, "\sum_{y'|y} \mathrm{Res}_{x',y'}"] & \\
\Omega_{K(X)/K} \arrow[u] \arrow[r, "\mathrm{Res}_{x,y}"'] & K_s
\end{tikzcd}$$

where y' varies over the irreducible curves of $\widehat{X}$ whose generic point sits over the generic point of y, and x' is the unique closed point sitting over x (that is, $p(x')=x$).

Proof. This essentially follows straight from the original definitions of the residue maps in sections 2C and 2D. Indeed, set $B = \mathcal{O}_{X,x}$ and let $y \subset B$ be the local equation for y at x, so that

$$\mathrm{Res}_{x,y} = \sum_{\substack{y'' \subset^1 \widehat{B} \\ y'' | y}} \mathrm{Res}_{y''} : \Omega^{\mathrm{sep}}_{\widehat{B}/\mathcal{O}_s} \otimes_{\widehat{B}} \operatorname{Frac} \widehat{B} \to K_s,$$

where y'' varies over the height-one primes of $\widehat{B}$ sitting over y.

But we remarked above that there is a natural $\mathcal{O}_s$-isomorphism $\widehat{\mathcal{O}_{\widehat{X},x'}} \cong \widehat{B}$, and this expression for the residues remains valid if B is replaced by $\mathcal{O}_{\widehat{X},x'}$ and y is replaced by some y' sitting over y. Therefore

$$\mathrm{Res}_{x,y} = \sum_{\substack{y'' \subset^1 \widehat{B} \\ y'' | y}} \mathrm{Res}_{y''} = \sum_{\substack{y' \subset^1 \mathcal{O}_{\widehat{X},x'} \\ y' | y}} \sum_{\substack{y'' \subset^1 \widehat{B} \\ y'' | y'}} \mathrm{Res}_{y''} = \sum_{\substack{y' \subset^1 \mathcal{O}_{\widehat{X},x'} \\ y' | y}} \mathrm{Res}_{y'} = \sum_{y'|y} \mathrm{Res}_{x',y'},$$

as required. □

Corollary 3.6. *Let $y \subset X$ be an irreducible component of the special fibre X_s and let x be a closed point of y; let $x' = p^{-1}(x)$, $y' = p^{-1}(y)$ be the corresponding point and curve on $\widehat{X}_s \cong X_s$. Then the following diagram commutes:*

$$\begin{tikzcd}
\Omega_{K(\widehat{X})/K_s} \arrow[dr, "\mathrm{Res}_{x',y'}"] & \\
\Omega_{K(X)/K} \arrow[u] \arrow[r, "\mathrm{Res}_{x,y}"'] & K_s
\end{tikzcd}$$

Informally, this means that residues along the special fibre X_s may be computed after completing $\mathcal{O}_K$.

Proof. The unique irreducible curve of $\widehat{X}$ sitting over y is y', so this follows from the previous lemma. □

Corollary 3.7. *If the vertical reciprocity law holds for $\widehat{X}/\mathcal{O}_s$ (for all $s \in S_0$), then it holds for $X/\mathcal{O}_K$.*

Proof. This immediately follows from the previous corollary. □

In the remainder of the section (except Remark 3.9), we replace X by $\widehat{X}$ and $\mathcal{O}_K$ by $\mathcal{O}_s$, so that the base is a now a complete, discrete valuation ring (of characteristic zero, with finite residue field, with field of fractions K being a local field).

The horizontal curves on X are all of the form $\overline{\{z\}}$ for a uniquely determined closed point z of the generic fibre X_η. Moreover, because our base ring is now complete, $\overline{\{z\}}$ meets the special fibre X_s at a unique point $\mathfrak{r}(z)$, which is necessarily closed and is called the reduction of z.

Lemma 3.8. *For any $\omega \in \Omega_{K(X)/K} = \Omega_{K(X_\eta)/K}$,*

$$\operatorname{Res}_{\mathfrak{r}(z),\overline{\{z\}}} \omega = \operatorname{Res}_z \omega,$$

where the left residue is the two-dimensional residue on X associated to the point and curve $\mathfrak{r}(z) \in \overline{\{z\}}$, and the right residue is the usual residue for the K-curve X_η at its closed point z.

Proof. This is a small exercise in chasing the definitions of the residue maps. Set $B = \mathcal{O}_{X,\mathfrak{r}(z)}$ and let $\mathfrak{p}$ be the local equation for $\overline{\{z\}}$ at $\mathfrak{r}(z)$. For any $n \geq 0$, $B/\mathfrak{p}^n$ is a finite $\mathcal{O}_K$-algebra, hence is complete. This implies that

$$\widehat{B}/\mathfrak{p}\widehat{B} = B/\mathfrak{p},$$

whence $\mathfrak{p}' = \mathfrak{p}\widehat{B}$ is prime in $\widehat{B}$, and also that

$$\widehat{B}_{\mathfrak{p}'}/\mathfrak{p}'^n \widehat{B}_{\mathfrak{p}'} = B_{\mathfrak{p}}/\mathfrak{p}^n B_{\mathfrak{p}}.$$

Therefore

$$\widehat{\widehat{B}_{\mathfrak{p}'}} = \varprojlim_n \widehat{B}_{\mathfrak{p}'}/\mathfrak{p}'^n \widehat{B}_{\mathfrak{p}'} = \varprojlim_n B_{\mathfrak{p}}/\mathfrak{p}^n B_{\mathfrak{p}} = \widehat{B}_{\mathfrak{p}} = \widehat{\mathcal{O}_{X_\eta,z}}.$$

Then $F := \operatorname{Frac} \widehat{\widehat{B}_{\mathfrak{p}'}}$ is the two-dimensional local field used to define the residue at the flag $\mathfrak{r}(z) \in \overline{\{z\}}$; it has equal characteristic, and we have just shown it is equal to $\operatorname{Frac} \mathcal{O}_{X_\eta,z}$. But the residue map on a two-dimensional local field of equal characteristic was exactly *defined* to be the familiar residue map for a curve. □

Remark 3.9. If $\mathcal{O}_K$ is not necessarily a complete, discrete valuation ring, as at the start of the section, then the above lemma remains valid when reformulated as follows: Let z be a closed point of the generic fibre, and X_s a special fibre. For any $\omega \in \Omega_{K(X)/K} = \Omega_{K(X_\eta)/K}$,

$$\sum_{x \in \overline{\{z\}} \cap X_s} \operatorname{Res}_{x,\overline{\{z\}}} \omega = \operatorname{Res}_z \omega$$

where the left is the sum of two-dimensional residues on X associated to the flags $x \in \overline{\{z\}}$ where x runs over the finitely many points in $\overline{\{z\}} \cap X_s$, and the right residue is the usual residue at the closed point z on the curve X_η. This may easily be deduced from the previous lemma using Lemma 5.1 below.

Proof of Theorem 3.1. We may now prove the vertical reciprocity law. Let

$$y_1 (= y), y_2, \ldots, y_l$$

be the irreducible components of the fibre X_s.

Firstly, combining the usual reciprocity law for the curve X_η with the previous lemma yields

$$\sum_{z \in (X_\eta)_0} \operatorname{Res}_{\mathfrak{r}(z), \overline{\{z\}}} \omega = 0,$$

where the sum is taken over closed points of the generic fibre and only finitely many terms of the summation are nonzero. Since $\overline{\{z\}}$, for $z \in (X_\eta)_0$, are all the irreducible horizontal curves of X, we may rewrite this as

$$\sum_{x \in X_0} \Bigg(\sum_{\substack{Y \subset X \text{ horiz.} \\ Y \ni x}} \operatorname{Res}_{x,Y} \omega \Bigg) = 0.$$

Moreover, according to the reciprocity law around a point from Section 2D, if $x \in X_0$ is a closed point then

$$\sum_{\substack{Y \subset X \\ Y \ni x}} \operatorname{Res}_{x,Y} \omega = 0,$$

where only finitely many terms in the summation are nonzero. We deduce that

$$\sum_{x \in X_0} \Bigg(\sum_{\substack{Y \subset X \text{ vert.} \\ Y \ni x}} \operatorname{Res}_{x,Y} \omega \Bigg) = 0,$$

where the sum is now taken over the irreducible vertical curves in X. That is,

$$\sum_{i=1}^{l} \sum_{x \in y_i} \operatorname{Res}_{x,y_i} \omega = 0, \tag{$\dagger$}$$

where the rearrangement of the double summation is justified by Lemma 3.3, which says that each internal sum of (†) converges in K.

If X_s is irreducible, this is exactly the sum over the closed points of $y_1 = y$ and we have finished. Else we must proceed by a "weighting" argument as in Lemma 2.8. Let $\nu_1, \dots, \nu_l$ be the discrete valuations on $K(X)$ associated to $y_1, \dots, y_l$ respectively. For $m > 0$, pick $f_m \in K(X)$ such that $\nu_1(f_m - 1) \geq m$ and $\nu_i(f_m) \geq m$ for $i = 2, \dots, l$; this exists because the $(\nu_i)_i$ are inequivalent discrete valuations. Replacing ω by $f_m\omega$ in (†) yields

$$\sum_{i=1}^{l} \sum_{x \in y_i} \operatorname{Res}_{x,y_i} f_m\omega = 0.$$

Letting $m \to \infty$ and applying the continuity part of Lemma 3.3 yields

$$\sum_{i=1}^{l}\sum_{x\in y_i}\mathrm{Res}_{x,y_i}\, f_m\omega = 0 \longrightarrow \sum_{x\in y_1}\mathrm{Res}_{x,y_1}\,\omega \quad \text{as } m \longrightarrow \infty.$$

This completes the proof of Theorem 3.1. □

4. Trace map via residues on higher adèles

We are now ready to adelically construct Grothendieck's trace map

$$H^1(X, \omega) \to \mathcal{O}_K$$

as a sum of our residues, where $\pi : X \to \operatorname{Spec}\mathcal{O}_K$ is an arithmetic surface and $\omega = \omega_\pi$ is its relative dualising sheaf. The key idea is to use the reciprocity laws to show that sums of residues descend to cohomology.

Remark 4.1. Passing from local constructions to global or cohomological objects is always the purpose of reciprocity laws. Compare with the reciprocity law around a point in K. Kato and S. Saito's [1983, §4] two-dimensional class field theory. Sadly, using reciprocity laws for the reciprocity map of two-dimensional local class field theory to construct two-dimensional global class field theory has not been written down in detail anywhere, but a sketch of how it should work in the geometric case was given by Parshin [1978]. More details, which are also valid in the arithmetic case, can be found in [Fesenko 2010, Chapter 2].

4A. ***Adèles of a curve.*** We begin with a quick reminder of adèles for curves. Let X be a one-dimensional, Noetherian, integral scheme with generic point η; we will be interested in both the case when X is smooth over a field and when X is the spectrum of the ring of integers of a number field. If E is a coherent sheaf on X, then the *adelic resolution* of E is the following flasque resolution:

$$0 \to E \to i_\eta(E_\eta) \oplus \prod_{x\in X_0} i_x(\widehat{E_x}) \to \prod_{x\in X_0}{}' i_x(\widehat{E_x} \otimes_{\mathcal{O}_{X,x}} K(X)) \to 0.$$

Here $i_\eta(E_\eta)$ is the constant E_η sheaf on X, $\widehat{E_x}$ is the $\mathfrak{m}_{X,x}$-adic completion of E_x and $i_x(\widehat{E_x})$ is the corresponding skyscraper sheaf at x, and the "restricted product" term $\prod'$ is the sheaf whose sections on an open set $U \subseteq X$ are

$$\prod_{x\in U_0}{}' \widehat{E_x}\otimes_{\mathcal{O}_{X,x}} K(X) = \left\{ (f_x) \in \prod_{x\in U_0} \widehat{E_x} \otimes_{\mathcal{O}_{X,x}} K(X) \;:\; \begin{array}{l} f_x \text{ is in the image of } \widehat{E_x} \text{ for} \\ \text{all but finitely many } x \in U_0 \end{array} \right\}.$$

The Zariski cohomology of E is therefore exactly the cohomology of the *adelic complex* $\mathbb{A}(X, E)$:

$$0 \to E_\eta \oplus \prod_{x \in X_0} \widehat{E}_x \to \prod_{x \in X_0}{}' \widehat{E}_x \otimes_{\mathbb{O}_{X,x}} K(X) \to 0$$
$$(g, (f_x)) \mapsto (g - f_x)$$

These observations remain valid if we do not bother completing E at each point x, leading to the *rational adelic complex* $a(X, E)$ (classically called repartitions, see for example [Serre 1988, II.5]):

$$0 \to E_\eta \oplus \prod_{x \in X_0} E_x \to \prod_{x \in X_0}{}' E_\eta \to 0$$

whose cohomology also equals the Zariski cohomology of E.

Remark 4.2. The reader who is about to encounter adelic spaces for surfaces for the first time may find it useful to see the following equality for the curve X:

$$\begin{aligned}\prod_{x \in X_0}{}' E_\eta & \\ &:= \{(f_x) \in \prod_{x \in X_0} E_\eta : f_x \text{ is in the image of } E_x \text{ for all but finitely many } x \in X_0\} \\ &= \{(f_x) \in \prod_{x \in X_0} E_\eta : \exists \text{ a coherent submodule } M \subseteq i_\eta(E_\eta) \text{ such that } f_x \in M_x \\ &\qquad\qquad \text{for all } x \in X_0\}\end{aligned}$$

4B. ***Rational adelic spaces for surfaces.*** The theory of adèles for curves was generalised to algebraic surfaces by Parshin (see [Parshin 1976], for example) and then to arbitrary Noetherian schemes by Beilinson [1980]. The main source of proofs is A. Huber's paper [1991]. We will describe the rational (that is, no completions are involved) adelic spaces, defined in [Huber 1991, §5.2], associated to a coherent sheaf E on a surface X. More precisely, X is any two-dimensional, Noetherian, integral scheme, with generic point η and function field $F = K(X)$. The quasicoherent sheaf which is constantly F will be denoted $\underline{F}$.

Remark 4.3. We choose to use the rational, rather than completed, adelic spaces to construct the trace map only for the sake of simplicity of notation. There is no substantial difficulty in extending the material of this section to the completed adèles, which becomes essential for the dualities discussed in Remark 5.6.

Adelic groups 0, 1*, and* 2*.* The first rational adelic groups are defined as follows:

$$a(0) = F, \qquad a(1) = \prod_{y \in X^1} \mathbb{O}_{X,y}, \qquad a(2) = \prod_{x \in X^2} \mathbb{O}_{X,x}.$$

More generally, if E is a coherent sheaf on X, then we define

$$a(0, E) = E_\eta, \qquad a(1, E) = \prod_{y \in X^1} E_y, \qquad a(2, E) = \prod_{x \in X^2} E_x.$$

Adelic group 01. Next we have the 01 adelic group:

$$
\begin{aligned}
&a(01)\\
&\quad = \Big\{(f_y) \in \prod_{y\in X^1} F : \exists \text{ a coherent submodule } M \subseteq \underline{F} \text{ such that } f_y \in M_y \text{ for all } y\Big\}\\
&\quad = \varinjlim_{M\subseteq \underline{F}} a(1, M)
\end{aligned}
$$

where the limit is taken over all coherent submodules M of the constant sheaf $\underline{F}$. This ring is commonly denoted using restricted product notation: $a(01) = \prod'_{y\in X^1} F$. Again more generally, if E is an arbitrary coherent sheaf, we put

$$
\begin{aligned}
&a(01, E)\\
&\quad = \Big\{(f_y) \in \prod_{y\in X^1} E_\eta : \exists \text{ a coherent submodule } M \subseteq \underline{E}_\eta \text{ such that } f_y \in M_y \text{ for all } y\Big\}\\
&\quad = \varinjlim_{M\subseteq \underline{E}_\eta} a(1, M),
\end{aligned}
$$

where the limit is taken over all coherent submodules M of the constant sheaf associated to E_η.

Adelic group 02. Next,

$$
\begin{aligned}
&a(02)\\
&\quad = \Big\{(f_x) \in \prod_{x\in X^2} F : \exists \text{ a coherent submodule } M \subseteq \underline{F} \text{ such that } f_x \in M_x \text{ for all } x\Big\}\\
&\quad = \varinjlim_{M\subseteq \underline{F}} a(2, M),
\end{aligned}
$$

where the limit is taken over all coherent submodules M of $\underline{F}$. This ring is commonly denoted $\prod'_{x\in X^2} F$. We leave it to the reader to write down the definition of $a(02, E)$, for E an arbitrary coherent sheaf.

Adelic group 12.

Remark 4.4. We first require some notation. If $z \in X$ is any point and N is a $\mathcal{O}_{X,z}$ module, then we write

$$[N]_z = j_{z*}(\widetilde{N}),$$

where $j_z : \operatorname{Spec} \mathcal{O}_{X,z} \hookrightarrow X$ is the natural morphism and $\widetilde{N}$ is the quasicoherent sheaf on $\operatorname{Spec} \mathcal{O}_{X,z}$ induced by N. For example, $\underline{F} = [\mathcal{O}_{X,\eta}]_\eta$.

We may now introduce

$$a(12) = \prod_{y\in X^1} a_y(12),$$

where

$$\begin{aligned} a_y(12) &= \Big\{ (f_x) \in \prod_{x \in y} \mathcal{O}_{X,y} : \exists \text{ a coherent submodule } M \subseteq [\mathcal{O}_{X,y}]_y \text{ such that} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad f_x \in M_x \text{ for all } x \in y \Big\} \\ &= \varinjlim_{M \subseteq [\mathcal{O}_{X,y}]_y} a(2, M), \end{aligned}$$

where the limit is taken over all coherent submodules M of $[\mathcal{O}_{X,y}]_y$. Recall our convention that if $y \in X^1$ then "$x \in y$" means that x is a codimension-one point of the closure of y; more precisely, $x \in X^2 \cap \overline{\{y\}}$.

We again leave it to the reader to write down the definition of $a(12, E)$ for an arbitrary coherent sheaf E (just replace $\mathcal{O}_{X,y}$ by E_y everywhere in the construction).

This is a convenient place to make one observation concerning an adelic condition which holds for $a(12, E)$:

Lemma 4.5. *Let E be a coherent sheaf on X, fix $y \in X^1$, $r \geq 0$, and let $(f_x)_{x \in y} \in a_y(12, E)$; then $f_x \in E_x + \mathfrak{m}_{X,y}^r E_y$ for all but finitely many $x \in y$.*

Proof. There is a coherent submodule $M \subseteq [E_y]_y$ such that $f_x \in M_x$ for all $x \in y$. Let $U = \operatorname{Spec} A$ be an affine open neighbourhood of (the generic point of) y, and let $\mathfrak{p} \subset A$ be the prime ideal defining y. Then $M(U)$ is a finitely generated A-submodule of $E_\mathfrak{p}$ and therefore $M(U) \subseteq fE$ for some $f \in A_\mathfrak{p}$. For any $r \geq 0$, the argument of Lemma 3.2 shows that $f \in A_\mathfrak{m} + \mathfrak{p}^r A_\mathfrak{p}$ for all but finitely many of the maximal ideals $\mathfrak{m}$ of A containing $\mathfrak{p}$; for such maximal ideals we have $M_\mathfrak{m} \subseteq E_\mathfrak{m} + \mathfrak{p}^r E_\mathfrak{p}$. Since U contains all but finitely many of the points of $\overline{\{y\}}$, this is enough. $\square$

Adelic group 012. Finally,

$$a(012) = \varinjlim_{M \subseteq \underline{F}} a(12, M) \subseteq \prod_{y \in X^1} \prod_{x \in y} F.$$

(and we similarly define $a(012, E)$ for any coherent E by taking the limit over coherent submodules M of the constant sheaf $\underline{E_\eta}$).

Simplicial structure and cohomology. Consider the following homomorphisms of rings:

$$\begin{tikzcd}
 & & F \arrow[dl] \arrow[dr] & & \\
 & \prod_{y \in X^1} F \arrow[r] & \prod_{y \in X^1} \prod_{x \in y} F & \prod_{x \in X^2} F \arrow[l] & \\
\prod_{y \in X^1} \mathcal{O}_{X,y} \arrow[ur] \arrow[rr] & & \prod_{y \in X^1} \prod_{x \in y} \mathcal{O}_{X,y} \arrow[u] & & \prod_{x \in X^2} \mathcal{O}_{X,x} \arrow[ll] \arrow[ul]
\end{tikzcd}$$

where the three ascending arrows are the obvious inclusions and the remaining arrows are diagonal embeddings. These homomorphisms restrict to the rational adelic groups just defined to give a commutative diagram of ring homomorphisms:

$$\begin{tikzcd}
 & & a(0) \arrow[dl, "\partial^0_{01}"'] \arrow[dr, "\partial^0_{02}"] & & \\
 & a(01) \arrow[r, "\partial^{01}_{012}"] & a(012) & a(02) \arrow[l, "\partial^{02}_{012}"'] & \\
a(1) \arrow[ur, "\partial^1_{01}"] \arrow[rr, "\partial^1_{12}"'] & & a(12) \arrow[u, "\partial^{12}_{012}"'] & & a(2) \arrow[ul, "\partial^2_{02}"'] \arrow[ll, "\partial^2_{12}"]
\end{tikzcd}$$

(and similarly with any coherent sheaf E in place of $\mathcal{O}_X$). For example, to see that ∂^1_{12} is well defined, once must check that if $f \in \mathcal{O}_{X,y}$ then there is a coherent submodule M of $[\mathcal{O}_{X,y}]_y$ such that $f_x \in M_x$ for all $x \in y$; but f may be viewed as a global section of $[\mathcal{O}_{X,y}]_y$ and therefore $M := \mathcal{O}_X f \subseteq [\mathcal{O}_{X,y}]_y$ suffices.

We reach the analogue for X of the rational adelic complex which we saw for a curve in Section 4A above:

Theorem 4.6. *Let E be a coherent sheaf on X; then the Zariski cohomology of E is equal to the cohomology of the complex*

$$0 \longrightarrow a(0, E) \oplus a(1, E) \oplus a(2, E) \\ \longrightarrow a(01, E) \oplus a(02, E) \oplus a(12, E) \longrightarrow a(012, E) \longrightarrow 0,$$

where the nontrivial arrows are given respectively by

$$(f_0, f_1, f_2) \mapsto (\partial^0_{01} f_0 - \partial^1_{01} f_1, \partial^2_{02} f_2 - \partial^0_{02} f_0, \partial^1_{12} f_1 - \partial^2_{12} f_2),$$
$$(g_{01}, g_{02}, g_{12}) \mapsto \partial^{01}_{012} g_{01} + \partial^{02}_{012} g_{02} + \partial^{12}_{012} g_{12}.$$

(This is the total complex associated to the simplicial group given above.)

Proof. This is due to Parshin [1976]; the general case of higher-dimensional X is due to Beilinson [1980] and Huber [1991]. □

4C. ***Construction of the trace map.*** Let $\mathcal{O}_K$ be a Dedekind domain of characteristic zero with finite residue fields; its field of fractions is K. Let $\pi : X \to S = \operatorname{Spec} \mathcal{O}_K$ be an $\mathcal{O}_K$-curve as at the start of Section 3. According to the main result of [Morrow 2010], the relative dualising sheaf ω of π is explicitly given by, for open $U \subseteq X$,

$$\omega(U) = \{\omega \in \Omega_{K(X)/K} : \operatorname{Res}_{x,y}(f\omega) \in \widehat{\mathcal{O}_{K,\pi(x)}} \text{ for all } x \in y \subset U \text{ and } f \in \mathcal{O}_{X,y}\} \quad (\dagger)$$

where x runs over all closed points of X inside U and y runs over all curves of U containing x.

As previously, closed points of S are denoted s, and we put $\mathcal{O}_s = \widehat{\mathcal{O}_{K,s}}$ and $K_s = \operatorname{Frac} \mathcal{O}_s$.

Proposition 4.7. *If* $\underline{\omega} = (\omega_{x,y})_{x \in y} \in a(012, \omega)$ *and* $s \in S_0$, *then*

$$\operatorname{Res}_s(\underline{\omega}) := \sum_{\substack{x,y \\ x \in y \cap X_s}} \operatorname{Res}_{x,y} \omega_{x,y} \tag{‡}$$

converges in K_s, *where the sum is taken over all points* x *and curves* y *in* X *for which* $x \in y \cap X_s$. *Moreover*, $\operatorname{Res}_s(\underline{\omega}) \in \mathcal{O}_s$ *for all but finitely many* $s \in S_0$.

If $\underline{\omega} \in \partial_{012}^{12} a(12, \omega)$ *then all terms of the sum*, *hence also* $\operatorname{Res}_s(\underline{\omega})$, *belong to* $\mathcal{O}_s$.

Proof. Let E be a coherent submodule of the constant sheaf $\underline{\omega}_\eta = \underline{\omega}_{K(X)/K}$ such that $\underline{\omega} \in a(12, E)$; then E and ω are equal at the generic point (replacing E by $E + \omega$, if necessary), hence on an open set, and therefore $E_y = \omega_y$ for all but finitely many $y \in X^1$. We call the remaining finitely many y *bad*.

If y is a horizontal curve which is not bad and $x \in y$, then $\omega_{x,y} \in E_y = \omega_y$ and so $\operatorname{Res}_{x,y} \omega_{x,y} = 0$ (indeed, if $\pi \in \mathcal{O}_{K,s}$ is a uniformiser at s then $\pi^{-1} \in \mathcal{O}_{X,y}$ and so the definition of ω implies that $\pi^{-m} \operatorname{Res}_{x,y} \omega_{x,y} \in \mathcal{O}_s$ for all $m \geq 0$; this is only possible if $\operatorname{Res}_{x,y} \omega_{x,y} = 0$). Therefore, only finitely many horizontal curves contribute to the summation in (‡); so it is enough to prove that if y is an irreducible component of X_s then

$$\sum_{x \in y} \operatorname{Res}_{x,y} \omega_{x,y}$$

converges. This is straightforward, using Lemma 4.5 and arguing exactly as in Lemma 3.3, and completes the proof that $\operatorname{Res}_s(\underline{\omega})$ is well defined.

Secondly, for any curve y, each of ω_y and E_y are (nonzero) finitely generated $\mathcal{O}_{X,y}$ submodules of $\Omega_{K(X)/K}$, and therefore there exists $r \geq 0$ such that $\mathfrak{m}_{X,y}^r E_y \subseteq \omega_y$; clearly we may pick r so that this inclusion holds for all bad y. Then Lemma 4.5 tells us that for all but finitely many x in any bad curve y, we have

$$E_y \subseteq E_x + \mathfrak{m}_{X,y}^r E_y \subseteq E_x + \omega_y.$$

Next, if y_1, y_2 are two horizontal curves, then y_1 and y_2 will have a common point of intersection on a vertical curve Y for only finitely many Y (for else $y_1 \cap y_2$ would be infinite). It follows that there is an open set $U \subseteq X$ consisting of fibres such that any $x \in U$ satisfies one of the following conditions:

(i) x sits on no bad curve, or

(ii) x sits on exactly one bad curve y; y is horizontal and $E_y \subseteq E_x + \omega_y$.

Note that U contains all but finitely many of the fibres X_s, for $s \in S_0$, and to prove our second claim it is enough to show that for any closed point x on a fibre X_s belonging to U, and curve y passing through x, one has $\operatorname{Res}_{x,y} \omega_{x,y} \in \mathcal{O}_s$. There are two cases to consider:

(i) y is not bad. Then $\omega_{x,y} \in E_y = \omega_y$, whence $\operatorname{Res}_{x,y} \omega_{x,y} \in \mathcal{O}_s$ by (†).

(ii) y is bad. Then y is horizontal by construction of U and so $\mathrm{Res}_{x,y}\,\omega_y = 0$ (as argued in the previous paragraph); therefore condition (ii) on U implies that $\mathrm{Res}_{x,y}\,\omega_{x,y} = \mathrm{Res}_{x,y}\,\zeta$ for some $\zeta \in E_x$. If Y is any curve through x apart from y then $\zeta \in E_x \subseteq E_Y = \omega_Y$ and so (†) now implies that $\mathrm{Res}_{x,Y}\,\zeta \in \mathbb{O}_s$. But the reciprocity law about a point from Section 2D shows that

$$\mathrm{Res}_{x,y}\,\zeta = -\sum_{Y} \mathrm{Res}_{x,Y}\,\zeta,$$

where the sum is taken over all curves Y passing through x apart from y; therefore $\mathrm{Res}_{x,y}\,\zeta \in \mathbb{O}_s$.

This completes the proof that $\mathrm{Res}_s\,\underline{\omega}$ belongs to $\mathbb{O}_s$ for all but finitely many $s \in S_0$.

Finally, if $\underline{\omega}$ is in the image of the boundary map ∂_{012}^{12} then $\omega_{x,y} \in \omega_y$ for all flags $x \in y$; so (†) implies that $\mathrm{Res}_{x,y}\,\omega_{x,y} \in \mathbb{O}_s$. This proves the final claim. □

Let

$$\mathbb{A}_S = \prod_{s\in S_0}{}' K_s = \left\{(a_s) \in \prod_{s\in S_0} K_s : a_s \in \mathbb{O}_s \text{ for all but finitely many } s\right\}$$

and

$$\mathbb{A}_S(0) = \prod_{s\in S_0} \mathbb{O}_s$$

be the rings of (finite) adèles and integral adèles of K respectively (we will incorporate archimedean information in the final section). The adelic complex for S, as discussed in Section 4A, is

$$0 \longrightarrow K \oplus \mathbb{A}_S(0) \longrightarrow \mathbb{A}_S \longrightarrow 0$$
$$(\lambda, (a_s)) \mapsto (\lambda - a_s)$$

Corollary 4.8. *The map*

$$\mathrm{Res} : a(012, \omega) \to \mathbb{A}_S, \quad \underline{\omega} \mapsto (\mathrm{Res}_s(\underline{\omega}))_{s\in S_0}$$

is well defined, and restricts to $\mathrm{Res}\circ\partial_{012}^{12} : a(12, \omega) \to \mathbb{A}_S(0)$.

Proof. This is exactly the content of the previous proposition. □

Define a map

$$\mathrm{Res}' : a(01, \omega) \oplus a(02, \omega) \oplus a(12, \omega) \to K \oplus \mathbb{A}_S(0)$$
$$(\underline{\omega}', \underline{\omega}'', \underline{\omega}) \mapsto \left(\sum_{z\in X_\eta} \mathrm{Res}_z\,\omega'_z, \mathrm{Res}(\partial_{012}^{12}\underline{\omega})\right)$$

where the first sum is taken over closed points z of X_η or, equivalently, horizontal curves in X, and Res_z denotes the usual residue for X_η as a smooth curve over K

(note that this makes sense as $\omega_\eta = \Omega_{K(X_\eta)/K}$). In the remainder of the paper, z will always denote a closed point of X_η.

The key application of the reciprocity laws is to deduce that taking sums of residues induces a morphism of adelic complexes:

Proposition 4.9. *The following maps give a homomorphism of adelic complexes from X to S:*

$$\begin{tikzcd}
0 \arrow[r] & a(0,\omega)\oplus a(1,\omega)\oplus a(2,\omega) \arrow[r] \arrow[d] & a(01,\omega)\oplus a(02,\omega)\oplus a(12,\omega) \arrow[r] \arrow[d, "\mathrm{Res}'"] & a(012,\omega) \arrow[r] \arrow[d, "\mathrm{Res}"] & 0 \\
& 0 \arrow[r] & K\oplus \mathbb{A}_S(0) \arrow[r] & \mathbb{A}_S \arrow[r] & 0
\end{tikzcd}$$

Proof. Commutativity of the first square is equivalent to the following results:

(i) If $\omega \in a(0, \omega) = \Omega_{K(X)/K}$ then $\sum_{z\in X_\eta} \mathrm{Res}_z\, \omega = 0$.

(ii) If $\underline{\omega} = (\omega_y)_{y\in X^1} \in a(1, \omega)$ then $\sum_{z\in X_\eta} \mathrm{Res}_z\, \omega_z = 0$ and $\mathrm{Res}(\partial^{12}_{012}\partial^{1}_{12}\underline{\omega}) = 0$.

(iii) If $\underline{\omega} \in a(2, \omega)$ then $\mathrm{Res}(\partial^{12}_{012}\partial^{2}_{12}\underline{\omega}) = 0$.

(i) is the usual reciprocity law for the curve X_η/K. The first vanishing claim in (ii) holds since $\omega_z \in \omega_z = \Omega_{X_\eta/K,z}$ and the residue of a differential form on X_η at a point where it is regular is zero. For the second vanishing claim in (ii), note that if $s \in S_0$ then

$$\mathrm{Res}_s(\partial^{01}_{012}\partial^{1}_{01}\underline{\omega}) = \sum_{y\subseteq X_s}\sum_{x\in y} \mathrm{Res}_{x,y}\,\omega_y + \sum_{\text{horiz. } y}\ \sum_{x\in X_s\cap y} \mathrm{Res}_{x,y}\,\omega_y,$$

where we have split the summation (‡) (of Proposition 4.7) depending on whether y is an irreducible component of X_s or is horizontal. But the first double summation is zero, according to the reciprocity law along a vertical curve (Theorem 3.1), while every term in the second double summation is zero since they are residues along horizontal curves y of forms in ω_y (see the second paragraph of the previous proof). We will return to (iii) in a moment.

Commutativity of the second square is almost automatic since Res' was obtained by restricting Res to $a(01, \omega)$ and $a(12, \omega)$; it remains only to check that if $\underline{\omega} \in a(02, \omega)$ then $\mathrm{Res}\,\partial^{02}_{012}\underline{\omega} = 0$. This follows immediately from the reciprocity law around a point from Section 2D. This also establishes (iii), since if $\underline{\omega} \in a(2, \omega)$ then $\partial^{12}_{012}\partial^{2}_{12}\underline{\omega} = \partial^{02}_{012}\partial^{2}_{02}\underline{\omega} \in \partial^{02}_{012}a(02, \omega)$. □

Noting that H^0 of the adelic complex for S is simply $\mathcal{O}_K$ and that H^1 of the adelic complex for X is $H^1(X, \omega)$ (by Theorem 4.6), the proposition implies that there is an induced map

$$\mathrm{Res} : H^1(X, \omega) \to \mathcal{O}_K.$$

Our construction would be irrelevant without the final theorem:

Theorem 4.10. Res *is equal to Grothendieck's trace map* tr_π.

Proof. There is a natural morphism from the rational adelic complex of X for the coherent sheaf ω to the rational adelic complex of X_η for the coherent sheaf $\Omega_{X_\eta/K}$:

$$\begin{tikzcd} 0 \arrow[r] & a(0,\omega)\oplus a(1,\omega)\oplus a(2,\omega) \arrow[r] \arrow[d, "(\omega_0,\omega_1,\omega_2)\mapsto(\omega_0,p_1(\omega_1))"] & a(01,\omega)\oplus a(02,\omega)\oplus a(12,\omega) \arrow[r] \arrow[d, "(\omega_{01},\omega_{02},\omega_{12})\mapsto p_{01}(\omega_{01})"] & a(012,\omega) \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & \Omega_{K(X)/K}\oplus\prod_{z\in X_\eta}\Omega_{X_\eta/K,z} \arrow[r] & \prod'_{z\in X_\eta}\Omega_{K(X)/K} \arrow[r] & 0 & \end{tikzcd}$$

This is given by the identity $a(0,\omega)=\Omega_{K(X)/K}$, the projection

$$a(1,\omega)=\prod_{y\in X^1}\omega_y=\prod_{z\in X_\eta}\Omega_{X_\eta/K,z}\times\prod_{\substack{y\in X^1\\ \text{vertical}}}\omega_y\overset{p_1}{\twoheadrightarrow}\prod_{z\in X_\eta}\Omega_{X_\eta/K,z},$$

and the restriction of the projection

$$\prod_{y\in X^1}\omega_\eta=\prod_{z\in X_\eta}\Omega_{K(X)/K}\times\prod_{\substack{y\in X^1\\ \text{vertical}}}\Omega_{K(X)/K}\twoheadrightarrow\prod_{z\in X_\eta}\Omega_{K(X)/K}$$

to the adelic spaces $a(01,\omega)\overset{p_{01}}{\twoheadrightarrow}\prod'_{z\in X_\eta}\Omega_{K(X)/K}$.

By the functoriality of adèles, the resulting map $H^*(X,\omega)\to H^*(X_\eta,\Omega_{X_\eta/K})$ is the natural map on cohomology induced by the restriction $\omega|_{X_\eta}=\Omega_{X_\eta/K}$. Using this, we will now show that

$$\begin{tikzcd} H^1(X,\omega) \arrow[r] \arrow[d, "\mathrm{Res}"'] & H^1(X_\eta,\Omega_{X_\eta/K}) \arrow[d, "\mathrm{tr}"] \\ \mathbb{O}_K \arrow[r] & K \end{tikzcd} \tag{$*$}$$

commutes, where the right vertical arrow is the trace map for the K-curve X_η. Indeed, from the definition of Res' above, the following diagram certainly commutes:

$$\begin{tikzcd} \mathrm{Ker}\langle a(01,\omega)\oplus a(02,\omega)\oplus a(12,\omega)\to a(012,\omega)\rangle \arrow[r, "(\omega_{01},\omega_{02},\omega_{12})\mapsto p_{01}(\omega_{01})"] \arrow[d, "\mathrm{Res}'"'] & \prod'_{z\in X_\eta}\Omega_{K(X)/K} \arrow[d, "(\omega_z)\mapsto\sum_{z\in X_\eta}\mathrm{Res}_z\omega_z"'] \\ \mathrm{Ker}\langle K\oplus\mathbb{A}_S(0)\to\mathbb{A}_S\rangle=\mathbb{O}_K \arrow[r] & K \end{tikzcd}$$

Passing to cohomology groups, we deduce that

$$\begin{array}{ccccc} H^1(X,\omega) & \longrightarrow & H^1(X_\eta,\Omega_{X_\eta/K}) & = & \mathrm{Coker}\left\langle \Omega_{K(X)/K}\oplus\prod_{z\in X_\eta}\Omega_{X_\eta/K,z}\to\prod'_{z\in X_\eta}\Omega_{K(X)/K}\right\rangle \\ {\scriptstyle\mathrm{Res}}\big\downarrow & & \big\downarrow{\scriptstyle(\omega_z)\mapsto\sum_{z\in X_\eta}\mathrm{Res}_z\,\omega_z} & & \\ \mathbb{O}_K & \longrightarrow & K & & \end{array}$$

commutes; but the vertical map on the right is the trace map for X_η, by the familiar result (which we are generalising!) that the trace map of a smooth projective curve is represented by the sum of residues. This completes the proof that $(*)$ commutes.

Finally, the diagram $(*)$ also commutes if Res is replaced by tr_π, since trace maps commute with localisation of the base ring. Therefore $\mathrm{Res}=\mathrm{tr}_\pi$. □

Remark 4.11. Before complicating matters by incorporating archimedean data, this is a convenient opportunity to explain how the previous material should fit into a general framework.

A *flag of points* on a scheme X is a sequence of points $\xi=(x_0,\dots,x_n)$ such that $x_{i-1}\in\overline{\{x_i\}}$ for $i=1,\dots,n$. By a process of successive completions and localisations, the flag ξ yields a ring F_ξ. More generally, to any quasicoherent sheaf E, one obtains a module E_ξ over F_ξ; for details, see [Huber 1991, §3.2].

Now let $f:X\to Y$ be a morphism of S-schemes, where S is a Noetherian scheme (perhaps Cohen–Macaulay), and notice that we may push forward any flag from X to Y,

$$f_*(\xi):=(f(x_0),\dots,f(x_n)),$$

resulting in an inclusion of rings $F_{f_*(\xi)}\subseteq F_\xi$. Let ω_X, ω_Y denote the dualising sheaves of X, Y over S. If f is proper (and probably Cohen–Macaulay) of fibre dimension d, then we expect there to exist a residue map

$$\mathrm{Res}_\xi:\omega_{X,\xi}\to\omega_{Y,f_*(\xi)}$$

which is the trace map when f is a finite morphism and which is transitive when given another proper, CM morphism $Y\to Z$. Globally, taking sums of these residue maps will induce a morphism of degree $-d$ on the adelic complexes

$$\mathrm{Res}_{X/Y}:\mathbb{A}(X,\omega_X)\to\mathbb{A}(Y,\omega_Y).$$

The patching together of the local residue maps to induce a morphism of complexes is equivalent to a collection of reciprocity laws being satisfied. In turn, this induces maps on the cohomology

$$H^*(X,\omega_X)=H^*(\mathbb{A}(X,\omega_X))\longrightarrow H^{*-d}(\mathbb{A}(Y,\omega_Y))=H^{*-d}(Y,\omega_Y),$$

which will be nothing other than Grothendieck's trace map.

When S is a field this framework more or less follows from [Lomadze 1981] and [Yekutieli 1992], though it has not been written down carefully. This article and the author's previous [Morrow 2010] focus on the case where $Y = S = \operatorname{Spec} \mathbb{O}_K$ and X is a surface.

The fully general case requires a rather careful development of relative residue maps in arbitrary dimensions, and becomes a technically difficult exercise quite quickly. The Hochschild homology-theoretic description of residue maps [Hübl 1989; Lipman 1987] may be the key to a smoother approach.

5. Archimedean reciprocity along horizontal curves

We continue to study an $\mathbb{O}_K$-curve X in the sense introduced at the start of Section 3, but we now assume that K is a number field and $\mathbb{O}_K$ its ring of integers (with generic point η). If ∞ is an infinite place of K then we write $X_\infty = X \times_{\mathbb{O}_K} K_\infty$ where K_∞ is the completion of K at ∞; so X_∞ is a smooth projective curve over $\mathbb{R}$ or $\mathbb{C}$.

The natural morphism

$$X_\infty = X \times_{\mathbb{O}_K} K_\infty \xrightarrow{\rho} X_\eta = X \times_{\mathbb{O}_K} K$$

can send a closed point to the generic point; but there are only finitely many points over any closed point. Indeed, let $z \in X_\eta$ be a closed point; then the fibre over z is

$$X_\infty \times_{X_\eta} k(z) = (K_\infty \times_K X_\eta) \times_{X_\eta} k(z) = \operatorname{Spec}(K_\infty \otimes_K k(z)),$$

which is a finite reduced scheme.

If y is a horizontal curve on X then $y = \overline{\{z\}}$ for a unique closed point $z \in X_\eta$. We say that a closed point $x \in X_\infty$ sits on y if and only if $\rho(x) = z$. Hence there are only finitely many points on X_∞ which sit on y, and we will allow ourselves to denote this set of points by $y \cap X_\infty$. Such points are the primes of $K_\infty \otimes_K k(z)$ and therefore correspond to the infinite places of the number field $k(z)$ extending the place ∞ on K. Note that each $x \in X_\infty$ sits on at most one horizontal curve, which may seem strange at first.

In this situation, we define the *archimedean residue map* $\operatorname{Res}_{x,y} : \Omega_{K(X)/K} \to K_\infty$ to be

$$\Omega_{K(X)/K} \longrightarrow \Omega_{K(X_\infty)/K_\infty} \xrightarrow{\operatorname{Res}_x} K_\infty,$$

where Res_x is the usual one-dimensional residue map associated to the closed point x on the smooth curve X_∞ over K_∞.

The following easy lemma was used in Remark 3.9; since we need it again, let's state it accurately:

Lemma 5.1. *Let C be a smooth, geometrically connected curve over a field K of characteristic zero, let L be an arbitrary extension of K, and let z be a closed point of C.*

(i) *Let $x \in C_L$ be a closed point sitting over z; then the following diagram commutes*:

$$\begin{tikzcd} \Omega_{K(C_L)/L} \arrow[r, "\mathrm{res}_x"] & k(x) \\ \Omega_{K(C)/K} \arrow[u] \arrow[r, "\mathrm{res}_z"] & k(z) \arrow[u] \end{tikzcd}$$

(*Notation*: res_x *is the residue map* $\Omega_{K(C)/K} \to k(x)$, *and* $\mathrm{Res}_x = \mathrm{Tr}_{k(x)/K} \circ \mathrm{res}_x$; *similarly for other points.*)

(ii) *With x now varying over all the closed points of C_L sitting over z, the following diagram commutes*:

$$\begin{tikzcd} \Omega_{K(C_L)/L} \arrow[r, "\sum_{x|z} \mathrm{Res}_x"] & L \\ \Omega_{K(C)/K} \arrow[u] \arrow[r, "\mathrm{Res}_z"] & K \arrow[u] \end{tikzcd}$$

Proof. If $t \in K(C)$ is a local parameter at z then it is also a local parameter at x, and the characteristic zero assumption implies that there are compatible isomorphisms $K(C_L)_x \cong k(x)((t))$, $K(C)_z \cong k(z)((t))$; the first claim easily follows. Secondly $k(z) \otimes_K L \cong \bigoplus_{x|z} k(x)$, so that $\mathrm{Tr}_{k(z)/K} = \sum_{x|z} \mathrm{Tr}_{k(x)/L}$; hence, for $\omega \in \Omega_{K(C)/K}$, part (i) lets us use the usual argument:

$$\begin{aligned} \sum_{x|z} \mathrm{Res}_x(\omega) = \sum_{x|z} \mathrm{Tr}_{k(x)/L} \mathrm{res}_x(\omega) = \sum_{x|z} \mathrm{Tr}_{k(x)/L} \mathrm{res}_z(\omega) \\ = \mathrm{Tr}_{k(z)/K} \mathrm{res}_z(\omega) = \mathrm{Res}_z(\omega) \quad \square \end{aligned}$$

We obtain an analogue of Remark 3.9:

Corollary 5.2. *Returning to the notation before the lemma, if ∞ and $y = \overline{\{z\}}$ are fixed, and $\omega \in \Omega_{K(X)/K}$, then*

$$\sum_{x \in y \cap X_\infty} \mathrm{Res}_{x,y}\, \omega = \mathrm{Res}_z\, \omega.$$

Proof. Apply the previous lemma with $C = X_\eta$ and $L = K_\infty$. □

Write $\bar{S} = \operatorname{Spec} \mathcal{O}_K \cup \{\infty\text{'s}\}$ for the "compactification" of $S = \operatorname{Spec} \mathcal{O}_K$ by the infinite places (in fact, the notation $s \in \bar{S}$ will always mean that s is a place of K, never the generic point of S) and let

$$\mathbb{A}_{\bar{S}} = \prod_{s \in \bar{S}}{}' K_s = \mathbb{A}_S \times \prod_{\infty} K_\infty$$

be the usual ring of adèles of the number field K. Let

$$\psi = \otimes_{s\in\bar{S}} \psi_s : \mathbb{A}_{\bar{S}} \to S^1 \quad (= \text{the circle group}^2)$$

be a continuous additive character which is trivial on the global elements $K \subset \mathbb{A}_{\bar{S}}$ [Tate 1967, Lemma 4.1.5].

Note that, if y is a horizontal curve on X, then even with our definition of points at infinity, it does not make sense to consider a reciprocity law

$$\text{“}\sum_{x\in y} \operatorname{Res}_{x,y} \omega = 0\text{”}$$

since the residues appearing live in different local fields. This problem is fixed by using the "absolute base" S^1:

Definition 5.3. Let y be a curve on X and $x \in y$ a closed point sitting over $s \in \bar{S}$ (this includes the possibility that y is horizontal and s is an infinite place). Define the *absolute residue map*

$$\psi_{x,y} : \Omega_{K(X)/K} \to S^1$$

to be the composition

$$\Omega_{K(X)/K} \xrightarrow{\operatorname{Res}_{x,y}} K_s \xrightarrow{\psi_s} S^1.$$

We may now establish the reciprocity law on X along *any* curve, including the horizontal ones:

Theorem 5.4. *Let y be a curve on X and $\omega \in \Omega_{K(X)/K}$. Then for all but finitely many closed points $x \in y$ the absolute residue $\psi_{x,y}(\omega)$ is* 1, *and*

$$\prod_{x\in y} \psi_{x,y}(\omega) = 1 \quad \textit{in } S^1.$$

Proof. First consider the case that y is an irreducible component of a special fibre X_s (here $s \in S_0$). Then $\operatorname{Ker}\psi_s$ is an open subgroup of K_s, and so the proof of Lemma 3.3 shows that $\operatorname{Res}_{x,y}\omega \in \operatorname{Ker}\psi_s$ for all but finitely many $x \in y$. Also,

$$\prod_{x\in y} \psi_{x,y}(\omega) = \psi_s\Big(\sum_{x\in y} \operatorname{Res}_{x,y}(\omega)\Big),$$

which is $\psi_s(0) = 1$ according to the reciprocity law along the vertical curve y (Theorem 3.1).

Secondly suppose that $y = \overline{\{z\}}$ is a horizontal curve; here z is a closed point of X_η. The proof of Proposition 4.7 shows that $\operatorname{Res}_{x,y}\omega \in \mathbb{O}_{\pi(x)}$ for all but finitely many $x \in y$ (here x is a genuine schematic point on X); since $\operatorname{Ker}\psi_s$ contains $\mathbb{O}_s$

[2] We never consider the set of codimension-one points of $S = \operatorname{Spec}\mathbb{O}_K$, so this shouldn't cause confusion.

for all but finitely many $s \in S_0$, it follows that $\psi_{x,y}(\omega) = 1$ for all but finitely many $x \in y$. It also follows that

$$\underline{f} := \Big(\sum_{x \in y \cap X_s} \operatorname{Res}_{x,y} \omega \Big)_{s \in \bar{S}}$$

belongs to $\mathbb{A}_{\bar{S}}$, and clearly

$$\prod_{x \in y} \psi_{x,y}(\omega) = \prod_{s \in \bar{S}} \psi_s \Big(\sum_{x \in y \cap X_s} \operatorname{Res}_{x,y} \omega \Big) = \psi(\underline{f}).$$

But Remark 3.9 (for $s \in S_0$) and the previous corollary (for s infinite) imply that $\underline{f}$ is the global adèle $\operatorname{Res}_z \omega \in K$. As ψ was chosen to be trivial on global elements, the proof is complete. □

Remark 5.5. The reciprocity law around a point $x \in X^2$ stated in Section 2D obviously implies that the absolute residue maps satisfy a similar law:

$$\prod_{y \subset X : y \ni x} \psi_{x,y}(\omega) = 1.$$

Therefore we have absolute reciprocity laws for all points and for all curves, which are analogues for an arithmetic surface of the reciprocity laws established by Parshin [1976] for an algebraic surface.

Remark 5.6. Let $F_{x,y}$ be the finite product of two-dimensional local fields attached to a flag $x \in y$; that is, $F_{x,y} = \operatorname{Frac} \widehat{A_{\mathfrak{p}}}$, where $A = \widehat{\mathcal{O}_{X,x}}$, $\mathfrak{p} = y\mathcal{O}_{X,x}$, and $y \subset \mathcal{O}_{X,x}$ also denotes the local equation for y at x; so $F_{x,y} = \prod_{y' | y} F_{x,y'}$ where y' varies over the finitely many height-one primes of A over y, and $F_{x,y'} = \operatorname{Frac} \widehat{A_{y'}}$.

By the local construction of the residue maps we see that $\psi_{x,y}$ is really the composition

$$\Omega_{K(X)/K} \longrightarrow \Omega_{K(X)/K} \otimes_{K(X)} F_{x,y} = \bigoplus_{y' | y} \Omega^{\text{cts}}_{F_{x,y'}/K_s} \xrightarrow{\sum_{y'|y} \operatorname{Res}_{F_{x,y'}}} K_s \xrightarrow{\psi_s} S^1$$

($s \in S_0$ is the point under x as usual), and each $\psi_s \circ \operatorname{Res}_{F_{x,y'}} : F_{x,y'} \to S^1$ is a continuous (with respect to the two-dimensional topology; see Remark 2.9) character on the two-dimensional local field $F_{x,y'}$. This character will induce self-duality of the topological group $F_{x,y'}$, which in turn will induce various dualities on the (complete) adelic groups; for some results in this direction, see [Fesenko 2010, §27, §28].

Remark 5.7. Taking $S = \operatorname{Spec} \mathbb{Z}$, it would be very satisfying to have an extension of the framework discussed in Remark 4.11 to include archimedean points. The main existing problem is the lack at present of a good enough theory of adèles in arbitrary dimensions which includes the points at infinity. The author is currently trying to develop such a theory and hopes that this will allow the dualities discussed in the previous remark to be stated more precisely and in greater generality (in all dimensions and including points at infinity).

Acknowledgements

I am thankful to I. Fesenko, A. Beilinson, A. Yekutieli and A. Cámara for discussions about this work. Parts of this research were funded by the Simons foundation and the EPSRC, and I am grateful to both organisations for their support.

I heartily thank the referee for suggestions and for reading the article in considerable depth, bringing to my attention several important issues which would have otherwise escaped my notice.

References

[Beilinson 1980] A. A. Beilinson, "Residues and adèles", *Funktsional. Anal. i Prilozhen.* **14**:1 (1980), 44–45. In Russian; translated in *Functional Anal. Appl.* **14** (1980), 34?35. MR 81f:14010 Zbl 0509.14018

[Cohen 1946] I. S. Cohen, "On the structure and ideal theory of complete local rings", *Trans. Amer. Math. Soc.* **59** (1946), 54–106. MR 7,509h Zbl 0060.07001

[Fesenko 2010] I. Fesenko, "Analysis on arithmetic schemes, II", *J. K-Theory* **5**:3 (2010), 437–557. MR 2011k:14019 Zbl 1225.14019

[Hartshorne 1966] R. Hartshorne, *Residues and duality*, Lecture Notes in Mathematics **20**, Springer, Berlin, 1966. MR 36 #5145 Zbl 0212.26101

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Huber 1991] A. Huber, "On the Parshin–Beilinson adèles for schemes", *Abh. Math. Sem. Univ. Hamburg* **61** (1991), 249–273. MR 92k:14024 Zbl 0763.14006

[Hübl 1989] R. Hübl, *Traces of differential forms and Hochschild homology*, Lecture Notes in Mathematics **1368**, Springer, Berlin, 1989. MR 92a:13010 Zbl 0675.13019

[Kato and Saito 1983] K. Kato and S. Saito, "Two-dimensional class field theory", pp. 103–152 in *Galois groups and their representations* (Nagoya, 1981), edited by Y. Ihara, Adv. Stud. Pure Math. **2**, North-Holland, Amsterdam, 1983. MR 87a:11060 Zbl 0544.12011

[Kunz 2008] E. Kunz, *Residues and duality for projective algebraic varieties*, University Lecture Series **47**, American Mathematical Society, Providence, RI, 2008. MR 2010b:14030 Zbl 1180.14002

[Lipman 1984] J. Lipman, *Dualizing sheaves, differentials and residues on algebraic varieties*, Astérisque **117**, Société Mathématique de France, Paris, 1984. MR 86g:14008 Zbl 0562.14003

[Lipman 1987] J. Lipman, *Residues and traces of differential forms via Hochschild homology*, Contemporary Mathematics **61**, American Mathematical Society, Providence, RI, 1987. MR 88b:14017 Zbl 0606.14015

[Liu 2002] Q. Liu, *Algebraic geometry and arithmetic curves*, Oxford Graduate Texts in Mathematics **6**, Oxford University Press, Oxford, 2002. MR 2003g:14001 Zbl 1103.14001

[Lomadze 1981] V. G. Lomadze, "On residues in algebraic geometry", *Izv. Akad. Nauk SSSR Ser. Mat.* **45**:6 (1981), 1258–1287. In Russian; translated in *Math. USSR, Izv.* **19**:(1982), 495–520. MR 83g:14021

[Madunts and Zhukov 1995] A. I. Madunts and I. B. Zhukov, "Multidimensional complete fields: topology and other basic constructions", pp. 1–34 in *Proceedings of the St. Petersburg Mathematical Society*, vol. 3, edited by O. A. Ladyzhenskaya, Amer. Math. Soc. Transl. Ser. 2 **166**, American Mathematical Society, Providence, RI, 1995. MR 97b:12009 Zbl 0874.11078

[Morrow 2010] M. Morrow, "An explicit approach to residues on and dualizing sheaves of arithmetic surfaces", *New York J. Math.* **16** (2010), 575–627. MR 2012a:14061 Zbl 05935390

[Osipov 1997] D. V. Osipov, "Adelic constructions of direct images of differentials and symbols", *Mat. Sb.* **188**:5 (1997), 59–84. In Russian; translated in *Sb. Math.* **188**:5 (1997), 697–723. MR 98k:11093

[Parshin 1976] A. N. Parshin, "On the arithmetic of two-dimensional schemes, I: Distributions and residues", *Izv. Akad. Nauk SSSR Ser. Mat.* **40**:4 (1976), 736–773. In Russian; translated in *Math. USSR Izv.* **10**:4 (1976), 695–730. MR 54 #7479

[Parshin 1978] A. N. Parshin, "Abelian coverings of arithmetic schemes", *Dokl. Akad. Nauk SSSR* **243**:4 (1978), 855–858. In Russian; translated in *Soviet Math. Dokl.* **19**:6 (1978), 1438–1442. MR 80b:14014 Zbl 0443.12006

[Serre 1988] J.-P. Serre, *Algebraic groups and class fields*, Graduate Texts in Mathematics **117**, Springer, New York, 1988. MR 88i:14041 Zbl 0703.14001

[Tate 1967] J. T. Tate, "Fourier analysis in number fields, and Hecke's zeta-functions", pp. 305–347 in *Algebraic number theory* (Brighton, 1965), edited by J. Cassels and A. Fröhlich, Thompson, Washington, D.C., 1967. MR 36 #121

[Yekutieli 1992] A. Yekutieli, *An explicit construction of the Grothendieck residue complex*, Astérisque **208**, 1992. MR 94e:14026 Zbl 0788.14011

Communicated by Hélène Esnault
Received 2011-06-26 Revised 2011-10-25 Accepted 2011-12-13

mmorrow@math.uchicago.edu *Department of Mathematics, University of Chicago, 5734 University Avenue, Chicago, IL 60637-1546, United States*
http://math.uchicago.edu/~mmorrow/

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1537

Crystalline extensions and the weight part of Serre's conjecture

Toby Gee, Tong Liu and David Savitt

Let $p > 2$ be prime. We complete the proof of the weight part of Serre's conjecture for rank-two unitary groups for mod p representations in the totally ramified case by proving that any Serre weight which occurs is a predicted weight. This completes the analysis begun by Barnet-Lamb, Gee, and Geraghty, who proved that all predicted Serre weights occur. Our methods are a mixture of local and global techniques, and in the course of the proof we use global techniques (as well as local arguments) to establish some purely local results on crystalline extension classes. We also apply these local results to prove similar theorems for the weight part of Serre's conjecture for Hilbert modular forms in the totally ramified case.

1. Introduction

The weight part of generalisations of Serre's conjecture has seen significant progress in recent years, particularly for (forms of) GL_2. Conjectural descriptions of the set of Serre weights were made in increasing generality in [Buzzard et al. 2010; Schein 2008; Gee et al. 2012], and cases of these conjectures were proved in [Gee 2011; Gee and Savitt 2011a]. Most recently, significant progress was made towards completely establishing the conjecture for rank-two unitary groups in [Barnet-Lamb et al. 2011]. We briefly recall this result. Let $p > 2$ be prime, F a CM field, and

The authors were partially supported by NSF grants DMS-0841491 (Gee), DMS-0901360 (Liu), and DMS-0901049 (Savitt).
MSC2010: 11F33.
Keywords: Serre's conjecture, p-adic Hodge theory, automorphy lifting theorems.

$\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ a modular representation (see [Barnet-Lamb et al. 2011] for the precise definition of "modular", which is in terms of automorphic forms on compact unitary groups). There is a conjectural set $W^?(\bar{r})$ of Serre weights in which $\bar{r}$ is predicted to be modular, which is defined in Section 2, following [Gee et al. 2012]. Then the main result of [Barnet-Lamb et al. 2011] is that under mild technical hypotheses, $\bar{r}$ is modular of every weight in $W^?(\bar{r})$. We note that this result is rather more general than anything that has been proved for inner forms of GL_2 over totally real fields, where there is a parity obstruction due to the unit group; algebraic Hilbert modular forms must have paritious weight. This problem does not arise for the unitary groups considered here, which is why we use them, rather than making use of the more obvious choice of an inner form. In the absence of a mod p functoriality principle, it is not known that the results for inner and outer forms of GL_2 are equivalent, and at present the theory for outer forms is in a better state.

It remains to show that if $\bar{r}$ is modular of some Serre weight, then this weight is contained in $W^?(\bar{r})$. It had been previously supposed that this was the easier direction; indeed, just as in the classical case, the results of [Barnet-Lamb et al. 2011] reduce the weight part of Serre's conjecture for these unitary groups to a purely local problem in p-adic Hodge theory. However, this problem has proved to be difficult, and so far only fragmentary results are known. In the present paper we resolve the problem in the totally ramified case, so that in combination with [ibid.] we resolve the weight part of Serre's conjecture in this case, proving the following theorem (see Theorem 6.1.2).

Theorem A. *Let F be an imaginary CM field with maximal totally real subfield F^+, and suppose that F/F^+ is unramified at all finite places, that $\zeta_p \notin F$, and that $[F^+ : \mathbb{Q}]$ is even. Suppose that $p > 2$, and that $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is an irreducible modular representation with split ramification such that $\bar{r}(G_{F(\zeta_p)})$ is adequate. Assume that for each place $w|p$ of F, $F_w/\mathbb{Q}_p$ is totally ramified.*

Let $a \in (\mathbb{Z}^2_+)^S_0$ be a Serre weight. Then $a_w \in W^?(\bar{r}|_{G_{F_w}})$ if and only if $\bar{r}$ is modular of weight a.

(See the body of the paper, especially Section 2.2, for any unfamiliar notation and terminology.) While [Barnet-Lamb et al. 2011] reduced this result to a purely local problem, our methods are not purely local; in fact we use the main result of [ibid.], together with potential automorphy theorems, as part of our proof.

In the case that $\bar{r}|_{G_{F_w}}$ is semisimple for each place $w|p$, the result was established (in a slightly different setting) in [Gee and Savitt 2011a]. The method of proof was in part global, making use of certain potentially Barsotti–Tate lifts to obtain conditions on $\bar{r}|_{G_{F_w}}$. We extend this analysis in the present paper to the case that $\bar{r}|_{G_{F_w}}$ is reducible but nonsplit, obtaining conditions on the extension classes that can occur; we show that (other than in one exceptional case) they lie in a certain set L_{flat}, defined in terms of finite flat models. We are also able to apply our final local results to improve on the global theorems proved in [Gee and Savitt 2011a]; see Theorem 6.1.3 below.

In the case that $\bar{r}|_{G_{F_w}}$ is reducible the definition of $W^?$ also depends on the extension class; it is required to lie in a set L_{crys}, defined in terms of reducible crystalline lifts with specified Hodge–Tate weights. To complete the proof, we show that $L_{\mathrm{crys}} = L_{\mathrm{flat}}$, except in one exceptional case that we handle separately in Proposition 5.2.9. An analogous result was proved in generic unramified cases in Section 3.4 of [Gee 2011] by means of explicit calculations with Breuil modules; our approach here is less direct, but has the advantage of working in nongeneric cases, and requires far less calculation.

We use a global argument to show that $L_{\mathrm{crys}} \subset L_{\mathrm{flat}}$. Given a class in L_{crys}, we use potential automorphy theorems to realise the corresponding local representation as part of a global modular representation, and then apply the main result of [Barnet-Lamb et al. 2011] to show that this representation is modular of the expected weight. Standard congruences between automorphic forms then show that this class is also contained in L_{flat}.

To prove the converse inclusion, we make a study of different finite flat models to show that L_{flat} is contained in a vector space of some dimension d. A standard calculation shows that L_{crys} contains a space of dimension d, so equality follows. As a byproduct, we show that both L_{flat} and L_{crys} are vector spaces. We also show that various spaces defined in terms of crystalline lifts are independent of the choice of lift (see Corollary 5.2.8). The analogous property was conjectured in the unramified case in [Buzzard et al. 2010].

It is natural to ask whether our methods could be extended to handle the general case, where $F_w/\mathbb{Q}_p$ is an arbitrary extension. Unfortunately, this does not seem to be the case, because in general the connection between being modular of some Serre weight and having a potentially Barsotti–Tate lift of some type is less direct. We expect that our methods could be used to reprove the results of Section 3.4 of [Gee 2011], but we do not see how to extend them to cover the unramified case completely. In particular, we are unsure as to when the equality $L_{\mathrm{flat}} = L_{\mathrm{crys}}$ holds in general.

We now explain the structure of the paper. In Section 2 we recall the definition of $W^?$, and the global results from [Barnet-Lamb et al. 2011] that we will need. In Section 3 we recall (and give a concise proof of) a potential automorphy result from [Gee and Kisin 2012], allowing us to realise a local mod p representation globally. Section 4 contains the definitions of the spaces L_{crys} and L_{flat} and the proof that $L_{\mathrm{crys}} \subset L_{\mathrm{flat}}$, and in Section 5 we carry out the necessary calculations with Breuil modules to prove our main local results. All of these results are in the reducible case, the irreducible case being handled in [Gee and Savitt 2011a]. Finally, in Section 6 we combine our local results with the techniques of [ibid.] and the main result of [Barnet-Lamb et al. 2011] to prove Theorem A, and we deduce a similar result in the setting of [Gee and Savitt 2011a].

Notation. If M is a field, we let G_M denote its absolute Galois group. Let ϵ denote the p-adic cyclotomic character, and $\bar{\epsilon}$ the mod p cyclotomic character. If M is a global field and v is a place of M, let M_v denote the completion of M at v. If M is a finite extension of $\mathbb{Q}_l$ for some l, we write I_M for the inertia subgroup of G_M. If R is a local ring we write $\mathfrak{m}_R$ for the maximal ideal of R.

Let K be a finite extension of $\mathbb{Q}_p$, with ring of integers $\mathcal{O}_K$ and residue field k. We write $\mathrm{Art}_K : K^\times \to W_K^{\mathrm{ab}}$ for the isomorphism of local class field theory, normalised so that uniformisers correspond to geometric Frobenius elements. For each $\sigma \in \mathrm{Hom}(k, \overline{\mathbb{F}}_p)$ we define the fundamental character ω_σ corresponding to σ to be the composite

$$I_K \longrightarrow W_K^{\mathrm{ab}} \xrightarrow{\mathrm{Art}_K^{-1}} \mathcal{O}_K^\times \longrightarrow k^\times \xrightarrow{\sigma} \overline{\mathbb{F}}_p^\times.$$

In the case that $k \cong \mathbb{F}_p$, we will sometimes write ω for ω_σ. Note that in this case we have $\omega^{[K:\mathbb{Q}_p]} = \bar{\epsilon}$.

We fix an algebraic closure $\overline{K}$ of K. If W is a de Rham representation of G_K over $\overline{\mathbb{Q}}_p$ and τ is an embedding $K \hookrightarrow \overline{\mathbb{Q}}_p$ then the multiset $\mathrm{HT}_\tau(W)$ of Hodge–Tate weights of W with respect to τ is defined to contain the integer i with multiplicity

$$\dim_{\overline{\mathbb{Q}}_p}(W \otimes_{\tau,K} \widehat{\overline{K}}(-i))^{G_K},$$

with the usual notation for Tate twists. Thus for example $\mathrm{HT}_\tau(\epsilon) = \{1\}$.

2. Serre weight conjectures: definitions

2.1. *Local definitions.* We begin by recalling some generalisations of the weight part of Serre's conjecture. We begin with some purely local definitions. Let K be a finite totally ramified extension of $\mathbb{Q}_p$ with absolute ramification index e, and let $\bar{\rho} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation.

Definition 2.1.1. A *Serre weight* is an irreducible $\overline{\mathbb{F}}_p$-representation of $\mathrm{GL}_2(\mathbb{F}_p)$. Up to isomorphism, any such representation is of the form

$$F_a := \det{}^{a_2} \otimes \mathrm{Sym}^{a_1-a_2}\, \overline{\mathbb{F}}_p^2$$

where $0 \le a_1 - a_2 \le p-1$. We also use the term Serre weight to refer to the pair $a = (a_1, a_2)$.

We say that two Serre weights a and b are *equivalent* if and only if $F_a \cong F_b$ as representations of $\mathrm{GL}_2(\mathbb{F}_p)$. This is equivalent to demanding that we have $a_1 - a_2 = b_1 - b_2$ and $a_2 \equiv b_2 \pmod{p-1}$.

We write $\mathbb{Z}^2_+$ for the set of pairs of integers (n_1, n_2) with $n_1 \geq n_2$, so that a Serre weight a is by definition an element of $\mathbb{Z}^2_+$. We say that an element $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)}$ is a *lift* of a Serre weight $a \in \mathbb{Z}^2_+$ if there is an element $\tau \in \mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)$ such that $\lambda_\tau = a$, and for all other $\tau' \in \mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)$ we have $\lambda_{\tau'} = (0, 0)$.

Definition 2.1.2. Let $K/\mathbb{Q}_p$ be a finite extension, let $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)}$, and let $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ be a de Rham representation. Then we say that ρ has *Hodge type* λ if for each $\tau \in \mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)$ we have $\mathrm{HT}_\tau(\rho) = \{\lambda_{\tau,1} + 1, \lambda_{\tau,2}\}$.

In particular, we will say that ρ has "Hodge type $\underline{0}$" if its Hodge–Tate weights are (0, 1) with respect to each embedding. Following [Gee et al. 2012] (which in turn follows [Buzzard et al. 2010; Schein 2008]), we define an explicit set of Serre weights $W^?(\bar{\rho})$.

Definition 2.1.3. If $\bar{\rho}$ is reducible, then a Serre weight $a \in \mathbb{Z}^2_+$ is in $W^?(\bar{\rho})$ if and only if $\bar{\rho}$ has a crystalline lift of the form

$$\begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$$

which has Hodge type λ for some lift $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)}$ of a.

In particular, if $a \in W^?(\bar{\rho})$ then by Lemma 6.2 of [Gee and Savitt 2011a] it is necessarily the case that there is a decomposition $\mathrm{Hom}(\mathbb{F}_p, \overline{\mathbb{F}}_p) = J \amalg J^c$ and an integer $0 \leq \delta \leq e - 1$ such that

$$\bar{\rho}|_{I_K} \cong \begin{pmatrix} \omega^\delta \prod_{\sigma \in J} \omega_\sigma^{a_1+1} \prod_{\sigma \in J^c} \omega_\sigma^{a_2} & * \\ 0 & \omega^{e-1-\delta} \prod_{\sigma \in J^c} \omega_\sigma^{a_1+1} \prod_{\sigma \in J} \omega_\sigma^{a_2}. \end{pmatrix}$$

We remark that this definition in terms of crystalline lifts is hard to work with concretely, and this is the reason for most of the work in this paper. We also remark that while it may seem strange to consider the single element set $\mathrm{Hom}(\mathbb{F}_p, \overline{\mathbb{F}}_p)$, this notation will be convenient for us (note that we always assume that the residue field of K is $\mathbb{F}_p$).

Definition 2.1.4. Let K' denote the quadratic unramified extension of K inside $\overline{K}$, with residue field k' of order p^2.

If $\bar{\rho}$ is irreducible, then a Serre weight $a \in \mathbb{Z}^2_+$ is in $W^?(\bar{\rho})$ if and only if there is a subset $J \subset \mathrm{Hom}(k', \overline{\mathbb{F}}_p)$ of size 1, and an integer $0 \leq \delta \leq e - 1$ such that if we write $\mathrm{Hom}(k', \overline{\mathbb{F}}_p) = J \amalg J^c$, then

$$\bar{\rho}|_{I_K} \cong \begin{pmatrix} \prod_{\sigma \in J} \omega_\sigma^{a_1+1+\delta} \prod_{\sigma \in J^c} \omega_\sigma^{a_2+e-1-\delta} & 0 \\ 0 & \prod_{\sigma \in J^c} \omega_\sigma^{a_1+1+\delta} \prod_{\sigma \in J} \omega_\sigma^{a_2+e-1-\delta} \end{pmatrix}.$$

We remark that by Lemma 4.1.19 of [Barnet-Lamb et al. 2011], if $a \in W^?(\bar{\rho})$ and $\bar{\rho}$ is irreducible then $\bar{\rho}$ necessarily has a crystalline lift of Hodge type λ for any lift $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)}$ of a. Note also that if a and b are equivalent and $a \in W^?(\bar{\rho})$ then $b \in W^?(\bar{\rho})$.

Remark 2.1.5. If $\bar{\theta} : G_K \to \overline{\mathbb{F}}_p^\times$ is an unramified character, then

$$W^?(\bar{r}) = W^?(\bar{r} \otimes \bar{\theta}).$$

2.2. ***Global conjectures.*** The point of the local definitions above is to allow us to formulate global Serre weight conjectures. Following [Barnet-Lamb et al. 2011], we work with rank-two unitary groups which are compact at infinity. As we will not need to make any arguments that depend on the particular definitions made in that article, and our main results are purely local, we simply recall some notation and basic properties of the definitions, referring the reader to [Barnet-Lamb et al. 2011] for precise formulations.

We emphasise that our conventions for Hodge–Tate weights are the opposite of the ones there; for this reason, we must introduce a dual into the definitions.

Fix an imaginary CM field F, and let F^+ be its maximal totally real subfield. We assume that each prime of F^+ over p has residue field $\mathbb{F}_p$ and splits in F. We define a global notion of Serre weight by taking a product of local Serre weights in the following way.

Definition 2.2.1. Let S denote the set of places of F above p. If $w \in S$ lies over a place v of F^+, write $v = ww^c$. Let $(\mathbb{Z}^2_+)^S_0$ denote the subset of $(\mathbb{Z}^2_+)^S$ consisting of elements $a = (a_w)_{w \in S}$ such that $a_{w,1} + a_{w^c,2} = 0$ for all $w \in S$. We say that an element $a \in (\mathbb{Z}^2_+)^S_0$ is a *Serre weight* if for each $w|p$ we have

$$p - 1 \geq a_{w,1} - a_{w,2}.$$

Let $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous irreducible representation. Definition 2.1.9 of [Barnet-Lamb et al. 2011] states what it means for $\bar{r}$ to be modular, and more precisely for $\bar{r}$ to be modular of some Serre weight a; roughly speaking, $\bar{r}$ is modular of weight a if there is a cohomology class on some unitary group with coefficients in the local system corresponding to a whose Hecke eigenvalues are determined by the characteristic polynomials of $\bar{r}$ at Frobenius elements. Since our conventions for Hodge–Tate weights are the opposite of those of Barnet-Lamb et al., we make the following definition.

Definition 2.2.2. Suppose that $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is a continuous irreducible modular representation. Then we say that $\bar{r}$ *is modular of weight* $a \in (\mathbb{Z}^2_+)^S_0$ if $\bar{r}^\vee$ is modular of weight a in the sense of Definition 2.1.9 of [Barnet-Lamb et al. 2011].

We remark that the definition of "modular" in that reference includes the hypotheses that F/F^+ is unramified at all finite places, that every place of F^+ dividing p splits in F, and that $[F^+ : \mathbb{Q}]$ is even.

If $\bar{r}$ is modular then $\bar{r}^c \cong \bar{r}^\vee \otimes \bar{\epsilon}$. We globalise the definition of the set $W^?(\bar{\rho})$ in the following natural fashion.

Definition 2.2.3. If $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is a continuous representation, then we define $W^?(\bar{r})$ to be the set of Serre weights $a \in (\mathbb{Z}^2_+)^S_0$ such that for each place $w|p$ the corresponding Serre weight $a_w \in \mathbb{Z}^2_+$ is an element of $W^?(\bar{r}|_{G_{F_w}})$.

One then has the following conjecture.

Conjecture 2.2.4. Suppose that $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is a continuous irreducible modular representation, and that $a \in (\mathbb{Z}^2_+)^S_0$ is a Serre weight. Then $\bar{r}$ is modular of weight a if and only if $a \in W^?(\bar{r})$.

If $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is a continuous representation, then we say that $\bar{r}$ has *split ramification* if any finite place of F at which $\bar{r}$ is ramified is split over F^+. We will frequently place ourselves in the following situation.

Hypothesis 2.2.5. Let F be an imaginary CM field with maximal totally real subfield F^+, and let $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Assume that:

- $p > 2$,
- $[F^+ : \mathbb{Q}]$ is even,
- F/F^+ is unramified at all finite places,
- $F_w/\mathbb{Q}_p$ is totally ramified for each place $w|p$ of F, and
- $\bar{r}$ is an irreducible modular representation with split ramification.

We point out that the condition that any place above p in F^+ splits in F, which is assumed throughout [ibid.], is implied by the third and fourth conditions above. The following result is Theorem 5.1.3 of [ibid.], one of the main theorems of that paper, specialised to the case of interest to us where $F_w/\mathbb{Q}_p$ is totally ramified for each place $w|p$ of F. (Note that in [ibid.], the set of Serre weights $W^?(\bar{r})$ is referred to as $W^{\mathrm{explicit}}(\bar{r})$.)

Theorem 2.2.6. *Suppose that Hypothesis 2.2.5 holds. Suppose further that $\zeta_p \notin F$ and $\bar{r}(G_{F(\zeta_p)})$ is adequate. Let $a \in (\mathbb{Z}^2_+)^S_0$ be a Serre weight. Assume that $a \in W^?(\bar{r})$. Then $\bar{r}$ is modular of weight a.*

Here *adequacy* is a group-theoretic condition, introduced in [Thorne 2011], that for subgroups of $\mathrm{GL}_2(\overline{\mathbb{F}}_p)$ with $p>5$ is equivalent to the usual condition that $\bar{r}|_{G_{F(\zeta_p)}}$ is irreducible. For a precise definition, see [Barnet-Lamb et al. 2011, Definition A.1.1].

Theorem 2.2.6 establishes one direction of Conjecture 2.2.4, and we are left with the problem of "elimination," that is, the problem of proving that if $\bar{r}$ is modular of weight a, then $a \in W^?(\bar{r})$. We believe that this problem should have a purely local resolution, as we now explain.

The key point is the relationship between being modular of weight a, and the existence of certain de Rham lifts of the local Galois representations $\bar{r}|_{G_{F_w}}$ with $w|p$. The link between these properties is provided by local-global compatibility for the Galois representations associated to the automorphic representations under consideration; rather than give a detailed development of this connection, we simply summarise the key results of [Barnet-Lamb et al. 2011].

Proposition 2.2.7 [Barnet-Lamb et al. 2011, Corollary 4.1.8]. *Suppose Hypothesis 2.2.5 holds. Let $a \in (\mathbb{Z}^2_+)^S_0$ be a Serre weight. If $\bar{r}$ is modular of weight a, then for each place $w|p$ of F, there is a crystalline representation $\rho_w : G_{F_w} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ lifting $\bar{r}|_{G_{F_w}}$ such that ρ_w has Hodge type λ_w for some lift $\lambda_w \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(F_w,\overline{\mathbb{Q}}_p)}$ of a.*

We stress that Proposition 2.2.7 does not complete the proof of Conjecture 2.2.4 because the representation ρ_w may be irreducible (compare with Definition 2.1.3). However, in light of this result, it is natural to make the following purely local conjecture, which together with Theorem 2.2.6 would essentially resolve Conjecture 2.2.4.

Conjecture 2.2.8. Let $K/\mathbb{Q}_p$ be a finite totally ramified extension, and let $\bar{\rho} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Let $a \in \mathbb{Z}^2_+$ be a Serre weight, and suppose that for some lift $\lambda \in (\mathbb{Z}^2_+)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,\overline{\mathbb{Q}}_p)}$, there is a continuous crystalline representation $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ lifting $\bar{\rho}$ such that ρ has Hodge type λ.

Then $a \in W^?(\bar{r})$.

We do not know how to prove this conjecture, and we do not directly address the conjecture in the rest of this paper. Instead, we proceed more indirectly. Proposition 2.2.7 is a simple consequence of lifting automorphic forms of weight a to forms of weight λ; we may also obtain nontrivial information by lifting to forms of weight 0 and nontrivial type. In this paper, we will always consider principal series types. Recall that if $K/\mathbb{Q}_p$ is a finite extension the *inertial type* of a potentially semistable Galois representation $\rho : G_K \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$ is the restriction to I_K of the corresponding Weil–Deligne representation. In this paper we normalise this definition as in the appendix to [Conrad et al. 1999], so that, for example, the inertial type of a finite order character is just the restriction to inertia of that character. We refer the reader to Definition 2.1.2 and the discussion immediately following it for our definition of "Hodge type $\underline{0}$."

Proposition 2.2.9. *Suppose that Hypothesis 2.2.5 holds. Let $a \in (\mathbb{Z}^2_+)^S_0$ be a Serre weight. If $\bar{r}$ is modular of weight a, then for each place $w|p$ of F, there is a continuous potentially semistable representation $\rho_w : G_{F_w} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ lifting $\bar{r}|_{G_{F_w}}$, such that ρ_w has Hodge type $\underline{0}$ and inertial type $\widetilde{\omega}^{a_1} \oplus \widetilde{\omega}^{a_2}$. (Here $\widetilde{\omega}$ is the Teichmüller lift of ω.) Furthermore, ρ_w is potentially crystalline unless*

$$a_1 - a_2 = p - 1 \quad \textit{and} \quad \bar{r}|_{G_{F_w}} \cong \begin{pmatrix} \overline{\chi}\bar{\epsilon} & * \\ 0 & \overline{\chi} \end{pmatrix} \textit{ for some character } \overline{\chi}.$$

Proof. This is proved in exactly the same way as [Gee and Savitt 2011a, Lemma 3.4], working in the setting of [Barnet-Lamb et al. 2011] (cf. the proof of Lemma 3.1.1 there). Note that if ρ_w is not potentially crystalline, then it is necessarily a twist of an extension of the trivial character by the cyclotomic character. □

3. Realising local representations globally

3.1. We now recall a result from [Gee and Kisin 2012], which allows us to realise local representations globally, in order to apply the results of Section 2.2 in a purely local setting.

Theorem 3.1.1. *Suppose that $p > 2$, that $K/\mathbb{Q}_p$ is a finite extension, and let $\bar{r}_K : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Then there is an imaginary CM field F and a continuous irreducible representation $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ such that, if F^+ denotes the maximal totally real subfield of F,*

- *each place $v|p$ of F^+ splits in F and has $F^+_v \cong K$,*
- *for each place $v|p$ of F^+, there is a place $\tilde{v}$ of F lying over F^+ with $\bar{r}|_{G_{F_{\tilde{v}}}}$ isomorphic to an unramified twist of $\bar{r}_K$,*
- *$\zeta_p \notin F$,*
- *$\bar{r}$ is unramified outside of p,*
- *$\bar{r}$ is modular in the sense of* [Barnet-Lamb et al. 2011], *and*
- *$\bar{r}(G_{F(\zeta_p)})$ is adequate.*

Proof. We give a brief (but complete) proof; a more detailed version appears in [Gee and Kisin 2012, Appendix A.1.5]. The argument is a straightforward application of potential modularity techniques. First, an application of Proposition 3.2 of [Calegari 2012] supplies a totally real field L^+ and a continuous irreducible representation $\bar{r} : G_{L^+} \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ such that

- for each place $v|p$ of L^+, $L^+_v \cong K$ and $\bar{r}|_{L^+_v} \cong \bar{r}_K$,
- for each place $v|\infty$ of L^+, $\det \bar{r}(c_v) = -1$, where c_v is a complex conjugation at v, and
- there is a nontrivial finite extension $\mathbb{F}/\mathbb{F}_p$ such that $\bar{r}(G_{L^+}) = \mathrm{GL}_2(\mathbb{F})$.

By a further base change one can also arrange that $\bar{r}|_{G_{L_v^+}}$ is unramified at each finite place $v \nmid p$ of L^+.

By Lemma 6.1.6 of [Barnet-Lamb et al. 2012] and the proof of Proposition 7.8.1 of [Snowden 2009], $\bar{r}_K$ admits a potentially Barsotti–Tate lift, and one may then apply Proposition 8.2.1 of [Snowden 2009] to deduce that there is a finite totally real Galois extension F^+/L^+ in which all primes of L^+ above p split completely, such that $\bar{r}|_{G_{F^+}}$ is modular in the sense that it is congruent to the Galois representation associated to some Hilbert modular form of parallel weight 2.

By the theory of base change between GL_2 and unitary groups (see Section 2 of [Barnet-Lamb et al. 2011]), it now suffices to show that there is a totally imaginary quadratic extension F/F^+ and a character $\bar{\theta} : G_F \to \overline{\mathbb{F}}_p^\times$ such that $\bar{r}|_{G_F} \otimes \bar{\theta}$ has multiplier $\bar{\epsilon}^{-1}$ and such that for each place $v|p$ of F^+, there is a place $\tilde{v}$ of F lying over v with $\bar{\theta}|_{G_{F_{\tilde{v}}}}$ unramified. The existence of such a character is a straightforward exercise in class field theory, and follows for example from Lemma 4.1.5 of [Clozel et al. 2008]. □

4. Congruences

4.1. Having realised a local mod p representation globally, we can now use the results explained in Section 2 to deduce nontrivial local consequences.

Proposition 4.1.1. *Let $p > 2$ be prime, let $K/\mathbb{Q}_p$ be a finite totally ramified extension, and let $\bar{\rho} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Let $a \in W^?(\bar{\rho})$ be a Serre weight. Then there is a continuous potentially semistable representation $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ lifting $\bar{\rho}$, such that ρ has Hodge type $\underline{0}$ and inertial type $\widetilde{\omega}^{a_1} \oplus \widetilde{\omega}^{a_2}$. Furthermore, ρ is potentially crystalline unless*

$$a_1 - a_2 = p - 1 \quad \text{and} \quad \bar{\rho} \cong \begin{pmatrix} \bar{\chi}\bar{\epsilon} & * \\ 0 & \bar{\chi} \end{pmatrix}$$

for some character $\bar{\chi}$.

Proof. By Theorem 3.1.1, there is an imaginary CM field F and a modular representation $\bar{r} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ such that

- for each place $v|p$ of F^+, v splits in F as $\tilde{v}\tilde{v}^c$, and we have $F_{\tilde{v}} \cong K$, and $\bar{r}|_{G_{F_{\tilde{v}}}}$ is isomorphic to an unramified twist of $\bar{\rho}$,
- $\bar{r}$ is unramified outside of p,
- $\zeta_p \notin F$, and
- $\bar{r}(G_{F(\zeta_p)})$ is adequate.

Now, since the truth of the result to be proved is obviously unaffected by making an unramified twist (if $\bar{\rho}$ is replaced by a twist by an unramified character $\bar{\theta}$, one may replace ρ by a twist by an unramified lift of $\overline{\theta}$), we may without loss of generality suppose that $\bar{r}|_{G_{F_w}} \cong \bar{\rho}$. Let $b \in (\mathbb{Z}^2_+)^S_0$ be the Serre weight such that $b_{\tilde{v}} = a$ for each place $v|p$ of F^+, where S denotes the set of places of F above p. By Remark 2.1.5, $b \in W^?(\bar{r})$. Then by Theorem 2.2.6, $\bar{r}$ is modular of weight b. The result now follows from Proposition 2.2.9. $\square$

4.2. ***Spaces of crystalline extensions.*** We now specialise to the reducible setting of Definition 2.1.3. As usual, we let $K/\mathbb{Q}_p$ be a finite totally ramified extension with residue field $k = \mathbb{F}_p$, ramification index e, and uniformiser π. We fix a Serre weight $a \in \mathbb{Z}^2_+$. Note that all the subsequent constructions that we make (such as the definitions of the spaces L_{flat} and L_{crys} below) will depend on this choice. We fix a continuous representation $\bar{\rho} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$, and we assume that there is:

- a decomposition $\mathrm{Hom}(\mathbb{F}_p, \overline{\mathbb{F}}_p) = J \amalg J^c$, and
- an integer $0 \le \delta \le e-1$ such that

$$\bar{\rho}|_{I_K} \cong \begin{pmatrix} \omega^{\delta} \prod_{\sigma\in J} \omega_\sigma^{a_1+1} \prod_{\sigma\in J^c} \omega_\sigma^{a_2} & * \\ 0 & \omega^{e-1-\delta} \prod_{\sigma\in J^c} \omega_\sigma^{a_1+1} \prod_{\sigma\in J} \omega_\sigma^{a_2} \end{pmatrix}.$$

Note that in general there might be several choices of J, δ. Consider pairs of characters $\chi_1, \chi_2 : G_K \to \overline{\mathbb{Q}}_p^\times$ with the properties that:

(1) $\bar{\rho} \cong \begin{pmatrix} \overline{\chi}_1 & * \\ 0 & \overline{\chi}_2 \end{pmatrix}$,

(2) χ_1 and χ_2 are crystalline, and

(3) if we let S denote the set of $\mathrm{Hom}_{\mathbb{Q}_p}(K, \overline{\mathbb{Q}}_p)$, then there exist J, δ as above such that either
   (i) J is nonempty, and there is one embedding $\tau \in S$ with $\mathrm{HT}_\tau(\chi_1) = a_1+1$ and $\mathrm{HT}_\tau(\chi_2) = a_2$, there are δ embeddings $\tau \in S$ with $\mathrm{HT}_\tau(\chi_1) = 1$ and $\mathrm{HT}_\tau(\chi_2) = 0$, and for the remaining $e-1-\delta$ embeddings $\tau \in S$ we have $\mathrm{HT}_\tau(\chi_1) = 0$ and $\mathrm{HT}_\tau(\chi_2) = 1$, or
   (ii) $J = \varnothing$, and there is one embedding $\tau \in S$ with $\mathrm{HT}_\tau(\chi_1) = a_2$ and $\mathrm{HT}_\tau(\chi_2) = a_1+1$, there are δ embeddings $\tau \in S$ with $\mathrm{HT}_\tau(\chi_1) = 1$ and $\mathrm{HT}_\tau(\chi_2) = 0$, and for the remaining $e-1-\delta$ embeddings $\tau \in S$ we have $\mathrm{HT}_\tau(\chi_1) = 0$ and $\mathrm{HT}_\tau(\chi_2) = 1$.

Note that these properties do not uniquely determine the characters χ_1 and χ_2, even in the unramified case, as one is always free to twist either character by an unramified character which is trivial mod p. We point out that the Hodge type of any de Rham extension of χ_2 by χ_1 will be a lift of a. Conversely, by Lemma 6.2 of [Gee and Savitt 2011a] any χ_1, χ_2 satisfying (1) and (2) such that the Hodge type of $\chi_1 \oplus \chi_2$ is a lift of a will satisfy (3) for a valid choice of J and δ (unique unless $a = 0$).

Suppose now that we have fixed two such characters χ_1 and χ_2, and we now allow the (line corresponding to the) extension class of $\bar{\rho}$ in $\mathrm{Ext}_{G_K}(\bar{\chi}_2, \bar{\chi}_1)$ to vary. We naturally identify $\mathrm{Ext}_{G_K}(\bar{\chi}_2, \bar{\chi}_1)$ with $H^1(G_K, \bar{\chi}_1\bar{\chi}_2^{-1})$ from now on.

Definition 4.2.1. Let L_{χ_1,χ_2} be the subset of $H^1(G_K, \bar{\chi}_1\bar{\chi}_2^{-1})$ such that the corresponding representation $\bar{\rho}$ has a crystalline lift ρ of the form

$$\begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}.$$

We have the following variant of Lemma 3.12 of [Buzzard et al. 2010].

Lemma 4.2.2. *L_{χ_1,χ_2} is an $\overline{\mathbb{F}}_p$-vector subspace of $H^1(G_K, \bar{\chi}_1\bar{\chi}_2^{-1})$ of dimension $|J| + \delta$ unless $\bar{\chi}_1 = \bar{\chi}_2$, in which case it has dimension $|J| + \delta + 1$.*

Proof. Let $\chi = \chi_1\chi_2^{-1}$. Recall $H^1_f(G_K, \overline{\mathbb{Z}}_p(\chi))$ is the preimage of $H^1_f(G_K, \overline{\mathbb{Q}}_p(\chi))$ under the natural map $\eta : H^1(G_K, \overline{\mathbb{Z}}_p(\chi)) \to H^1(G_K, \overline{\mathbb{Q}}_p(\chi))$, so that L_{χ_1,χ_2} is the image of $H^1_f(G_K, \overline{\mathbb{Z}}_p(\chi))$ in $H^1(G_K, \bar{\chi})$. The kernel of η is precisely the torsion part of $H^1(G_K, \overline{\mathbb{Z}}_p(\chi))$. Since $\chi \neq 1$, e.g., by examining Hodge–Tate weights, this torsion is nonzero if and only if $\bar{\chi} = 1$, in which case it has the form $\kappa^{-1}\overline{\mathbb{Z}}_p/\overline{\mathbb{Z}}_p$ for some $\kappa \in \mathfrak{m}_{\overline{\mathbb{Z}}_p}$. (To see this, note that if $\chi \neq 1$ is defined over E, then the long exact sequence associated to $0 \to \mathcal{O}_E(\chi) \to \mathcal{O}_E(\chi) \to k_E(\bar{\chi}) \to 0$ identifies $k_E(\bar{\chi})^{G_K}$ with the ϖ-torsion in $\ker(\eta)$.)

By Proposition 1.24(2) of [Nekovář 1993] we see that

$$\dim_{\overline{\mathbb{Q}}_p} H^1_f(G_K, \overline{\mathbb{Q}}_p(\chi)) = |J| + \delta,$$

again using $\chi \neq 1$. Since $H^1(G_K, \overline{\mathbb{Z}}_p(\chi))$ is a finitely generated $\overline{\mathbb{Z}}_p$-module, the result follows. □

Definition 4.2.3. If $\bar{\chi}_1$ and $\bar{\chi}_2$ are fixed, we define L_{crys} to be the subset of $H^1(G_K, \bar{\chi}_1\bar{\chi}_2^{-1})$ given by the union of the L_{χ_1,χ_2} over all χ_1 and χ_2 as above.

Note that L_{crys} is a union of subspaces of possibly varying dimensions, and as such it is not clear that L_{crys} is itself a linear subspace. Note also that the representations $\bar{\rho}$ corresponding to elements of L_{crys} are by definition precisely those for which $a \in W^?(\bar{\rho})$. Note also that L_{crys} depends only on $\bar{\rho}^{\mathrm{ss}}$ and a.

Definition 4.2.4. Let L_{flat} be the subset of $H^1(G_K, \bar{\chi}_1\bar{\chi}_2^{-1})$ consisting of classes with the property that if

$$\bar{\rho} \cong \begin{pmatrix} \bar{\chi}_1 & * \\ 0 & \bar{\chi}_2 \end{pmatrix}$$

is the corresponding representation, then there is a finite field $k_E \subset \overline{\mathbb{F}}_p$ and a finite flat k_E-vector space scheme over $\mathcal{O}_{K(\pi^{1/(p-1)})}$ with generic fibre descent data to K of type $\omega^{a_1} \oplus \omega^{a_2}$ (see Definition 5.1.1) whose generic fibre is $\bar{\rho}$.

Note that L_{flat} depends only on $\bar{\rho}^{\text{ss}}$ and a.

Proposition 4.2.5. *Provided that $a_1 - a_2 \neq p-1$ or that $\overline{\chi}_1\overline{\chi}_2^{-1} \neq \bar{\epsilon}$, $L_{\text{crys}} \subset L_{\text{flat}}$.*

Proof. Take a class in L_{crys}, and consider the corresponding representation

$$\bar{\rho} \cong \begin{pmatrix} \overline{\chi}_1 & * \\ 0 & \overline{\chi}_2 \end{pmatrix}.$$

As remarked above, $a \in W^?(\bar{\rho})$, so by Proposition 4.1.1, $\bar{\rho}$ has a crystalline lift of Hodge type $\underline{0}$ and inertial type $\widetilde{\omega}^{a_1} \oplus \widetilde{\omega}^{a_2}$, and this representation can be taken to have coefficients in the ring of integers $\mathcal{O}_E$ of a finite extension $E/\mathbb{Q}_p$. Let ϖ be a uniformiser of $\mathcal{O}_E$, and k_E the residue field. Such a representation corresponds to a p-divisible $\mathcal{O}_E$-module with generic fibre descent data, and taking the ϖ-torsion gives a finite flat k_E-vector space scheme with generic fibre descent data whose generic fibre is $\bar{\rho}$. By Corollary 5.2 of [Gee and Savitt 2011b] this descent data has type $\omega^{a_1} \oplus \omega^{a_2}$. □

In the next section we will make calculations with finite flat group schemes in order to relate L_{flat} and L_{crys}.

5. Finite flat models

5.1. We work throughout this section in the following setting:

- $K/\mathbb{Q}_p$ is a finite extension with ramification index e, ring of integers $\mathcal{O}_K$, uniformiser π and residue field $\mathbb{F}_p$.
- $\overline{\chi}_1, \overline{\chi}_2$ are characters $G_K \to \overline{\mathbb{F}}_p^\times$.
- $a \in \mathbb{Z}_+^2$ is a Serre weight.
- There is a decomposition $\mathrm{Hom}(\mathbb{F}_p, \overline{\mathbb{F}}_p) = J \amalg J^c$, and an integer $0 \le \delta \le e-1$ such that

$$\overline{\chi}_1|_{I_K} = \omega^{\delta} \prod_{\sigma \in J} \omega^{a_1+1} \prod_{\sigma \in J^c} \omega^{a_2},$$
$$\overline{\chi}_2|_{I_K} = \omega^{e-1-\delta} \prod_{\sigma \in J^c} \omega^{a_1+1} \prod_{\sigma \in J} \omega^{a_2}.$$

Note in particular that $(\overline{\chi}_1\overline{\chi}_2)|_{I_K} = \omega^{a_1+a_2+e}$.

Let $K_1 := K(\pi^{1/(p-1)})$. Let k_E be a finite extension of $\mathbb{F}_p$ such that $\overline{\chi}_1, \overline{\chi}_2$ are defined over k_E; for the moment k_E will be fixed, but eventually it will be allowed to vary.

We wish to consider the representations

$$\bar{\rho} \cong \begin{pmatrix} \overline{\chi}_1 & * \\ 0 & \overline{\chi}_2 \end{pmatrix}$$

such that there is a finite flat k_E-vector space scheme $\mathcal{G}$ over $\mathcal{O}_{K_1}$ with generic fibre descent data to K of type $\omega^{a_1} \oplus \omega^{a_2}$ (see Definition 5.1.1), whose generic fibre is $\bar{\rho}$. In order to do so, we will work with Breuil modules with descent data from K_1 to K. We recall the necessary definitions from [Gee and Savitt 2011b].

Fix π_1, a $(p-1)$-st root of π in K_1. Write $e' = e(p-1)$. The category $\mathrm{BrMod}_{\mathrm{dd}}$ consists of quadruples $(\mathcal{M}, \mathrm{Fil}^1\mathcal{M}, \phi_1, \{\widehat{g}\})$ where:

- $\mathcal{M}$ is a finitely generated free $k_E[u]/u^{e'p}$-module,
- $\mathrm{Fil}^1\mathcal{M}$ is a $k_E[u]/u^{e'p}$-submodule of $\mathcal{M}$ containing $u^{e'}\mathcal{M}$,
- $\phi_1 : \mathrm{Fil}^1\mathcal{M} \to \mathcal{M}$ is k_E-linear and ϕ-semilinear (where

  $$\phi : \mathbb{F}_p[u]/u^{e'p} \to \mathbb{F}_p[u]/u^{e'p}$$

  is the p-th power map) with image generating $\mathcal{M}$ as a $k_E[u]/u^{e'p}$-module, and
- $\widehat{g} : \mathcal{M} \to \mathcal{M}$ for each $g \in \mathrm{Gal}(K_1/K)$ are additive bijections that preserve $\mathrm{Fil}^1\mathcal{M}$, commute with the ϕ_1-, and k_E-actions, and satisfy $\widehat{g_1} \circ \widehat{g_2} = \widehat{g_1 \circ g_2}$ for all $g_1, g_2 \in \mathrm{Gal}(K_1/K)$; furthermore $\widehat{1}$ is the identity, and if $a \in k_E$, $m \in \mathcal{M}$ then $\widehat{g}(au^i m) = a((g(\pi)/\pi)^i)u^i\widehat{g}(m)$.

The category $\mathrm{BrMod}_{\mathrm{dd}}$ is equivalent to the category of finite flat k_E-vector space schemes over $\mathcal{O}_{K_1}$ together with descent data on the generic fibre from K_1 to K (this equivalence depends on π_1); see [Savitt 2008], for instance. We obtain the associated G_K-representation (which we will refer to as the generic fibre) of an object of $\mathrm{BrMod}_{\mathrm{dd},K_1}$ via the covariant functor $T^K_{\mathrm{st},2}$ (which is defined immediately before Lemma 4.9 of [Savitt 2005]).

Definition 5.1.1. Let $\mathcal{M}$ be an object of $\mathrm{BrMod}_{\mathrm{dd}}$ such that the underlying k_E-module has rank two. We say that the finite flat k_E-vector space scheme corresponding to $\mathcal{M}$ *has descent data of type* $\omega^{a_1} \oplus \omega^{a_2}$ if $\mathcal{M}$ has a basis e_1, e_2 such that $\widehat{g}(e_i) = \omega^{a_i}(g)e_i$. (Here we abuse notation by identifying an element of G_K with its image in $\mathrm{Gal}(K_1/K)$.)

We now consider a finite flat group scheme with generic fibre descent data $\mathcal{G}$ as above. By a standard scheme-theoretic closure argument, $\overline{\chi}_1$ corresponds to a finite flat subgroup scheme with generic fibre descent data $\mathcal{H}$ of $\mathcal{G}$, so we begin by analysing the possible finite flat group schemes corresponding to characters.

Suppose now that $\mathcal{M}$ is an object of $\mathrm{BrMod}_{\mathrm{dd}}$. The rank one objects of $\mathrm{BrMod}_{\mathrm{dd}}$ are classified as follows.

Proposition 5.1.2. *With our fixed choice of uniformiser π, every rank-one object of* $\mathrm{BrMod}_{\mathrm{dd}}$ *has the form*:

- $\mathcal{M} = (k_E[u]/u^{e'p}) \cdot v$,
- $\mathrm{Fil}^1\mathcal{M} = u^{x(p-1)}\mathcal{M}$,
- $\phi_1(u^{x(p-1)}v) = cv$ *for some* $c \in k_E^\times$, *and*
- $\widehat{g}(v) = \omega(g)^k v$ *for all* $g \in \mathrm{Gal}(K_1/K)$,

where $0 \le x \le e$ *and* $0 \le k < p-1$ *are integers.*

Then $T_{\mathrm{st},2}^K(\mathcal{M}) = \omega^{k+x} \cdot \mathrm{ur}_{c^{-1}}$, *where* $\mathrm{ur}_{c^{-1}}$ *is the unramified character taking an arithmetic Frobenius element to* c^{-1}.

Proof. This is a special case of Proposition 4.2 and Corollary 4.3 of [Gee and Savitt 2011b]. □

Let $\mathcal{M}$ (or $\mathcal{M}(x)$) be the rank-one Breuil module with k_E-coefficients and descent data from K_1 to K corresponding to $\mathcal{H}$, and write $\mathcal{M}$ in the form given by Proposition 5.1.2. Since $\mathcal{G}$ has descent data of type $\omega^{a_1} \oplus \omega^{a_2}$, we must have $\omega^k \in \{\omega^{a_1}, \omega^{a_2}\}$.

5.2. *Extensions.* Having determined the rank-one objects, we now go further and compute the possible extension classes. By a scheme-theoretic closure argument, the Breuil module $\mathcal{P}$ corresponding to $\mathcal{G}$ is an extension of $\mathcal{N}$ by $\mathcal{M}$, where $\mathcal{M}$ is as in the previous section, and $\mathcal{N}$ (or $\mathcal{N}(y)$) is defined by

- $\mathcal{N} = (k_E[u]/u^{e'p}) \cdot w$,
- $\mathrm{Fil}^1\mathcal{N} = u^{y(p-1)}\mathcal{N}$,
- $\phi_1(u^{y(p-1)}v) = dw$ for some $d \in k_E^\times$, and
- $\widehat{g}(v) = \omega(g)^l v$ for all $g \in \mathrm{Gal}(K_1/K)$,

where $0 \le y \le e$ and $0 \le l < p-1$ are integers. Now, as noted above, the descent data for $\mathcal{G}$ is of type $\omega^{a_1} \oplus \omega^{a_2}$, so we must have that either $\omega^k = \omega^{a_1}$ and $\omega^l = \omega^{a_2}$, or $\omega^k = \omega^{a_2}$ and $\omega^l = \omega^{a_1}$. Since by definition we have $(\overline{\chi}_1\overline{\chi}_2)|_{I_K} = \omega^{a_1+a_2+e}$, we see from Proposition 5.1.2 that

$$x + y \equiv e \pmod{p-1}.$$

We have the following classification of extensions of $\mathcal{N}$ by $\mathcal{M}$.

Proposition 5.2.1. *Every extension of $\mathcal{N}$ by $\mathcal{M}$ is isomorphic to exactly one of the form*

- $\mathcal{P} = (k_E[u]/u^{e'p}) \cdot v + (k_E[u]/u^{e'p}) \cdot w$,
- $\mathrm{Fil}^1\mathcal{P} = (k_E[u]/u^{e'p}) \cdot u^{x(p-1)}v + (k_E[u]/u^{e'p}) \cdot (u^{y(p-1)}w + \nu v)$,
- $\phi_1(u^{x(p-1)}v) = cv$, $\phi_1(u^{y(p-1)}w + \nu v) = dw$,

- $\widehat{g}(v) = \omega^k(g)v$ *and* $\widehat{g}(w) = \omega^l(g)w$ *for all* $g \in \mathrm{Gal}(K_1/K)$,

where $\nu \in u^{\max\{0,(x+y-e)(p-1)\}}k_E[u]/u^{e'p}$ *has all nonzero terms of degree congruent to* $l-k$ *modulo* $p-1$, *and has all terms of degree less than* $x(p-1)$, *unless* $\overline{\chi}_1 = \overline{\chi}_2$ *and* $x \geq y$, *in which case it may additionally have a term of degree* $px - y$.

Proof. This is a special case of Theorem 7.5 of [Savitt 2004] with the addition of k_E-coefficients. When K (in the notation of *loc. cit.*) is totally ramified, the proof of that theorem with coefficients added proceeds in the same manner with only the following changes, where l corresponds to p in our present paper.

- Replace Lemma 7.1 of *loc. cit.* (i.e., Lemma 5.2.2 of [Breuil et al. 2001]) with Lemma 5.2.4 of [Breuil et al. 2001] (with $k' = k_E$ and $k = \mathbb{F}_p$ in the notation of that lemma). In particular replace t^l with $\phi(t)$ wherever it appears in the proof, where ϕ is the k_E-linear endomorphism of $k_E[u]/u^{e'l}$ sending u^i to u^{li}.
- Instead of applying Lemma 4.1 of [Savitt 2004], note that the cohomology group $H^1(\mathrm{Gal}(K_1/K), k_E[u]/u^{e'l})$ vanishes because $\mathrm{Gal}(K_1/K)$ has prime-to-l order while $k_E[u]/u^{e'l}$ has l-power order.
- Every occurrence of T_i^l in the proof (for any subscript i) should be replaced with T_i.
- The coefficients of h, t are permitted to lie in k_E (that is, they are not constrained to lie in any particular proper subfield). □

The formulas for $(\mathcal{P}, \mathrm{Fil}^1\mathcal{P}, \phi_1, \{\widehat{g}\})$ in the statement of Proposition 5.2.1 define a Breuil module with descent data provided that $\mathrm{Fil}^1\mathcal{P}$ contains $u^{e'}\mathcal{P}$ and is preserved by each $\widehat{g}$. This is the case as long as ν lies in $u^{\max\{0,(x+y-e)(p-1)\}}k_E[u]/u^{e'p}$ and has all nonzero terms of degree congruent to $l-k$ modulo $p-1$ (compare the discussion in Section 7 of [Savitt 2004]); denote this Breuil module by $\mathcal{P}(x, y, \nu)$. Note that c is fixed while x determines k, since we require $\omega^{k+x} \cdot \mathrm{ur}_{c^{-1}} = \overline{\chi}_1$; similarly d is fixed and y determines l. So this notation is reasonable.

We would like to compare the generic fibres of extensions of different choices of $\mathcal{M}$ and $\mathcal{N}$. To this end, we have the following result. Write $\overline{\chi}_1|_{I_K} = \omega^\alpha$, $\overline{\chi}_2|_{I_K} = \omega^\beta$.

Proposition 5.2.2. *The Breuil module* $\mathcal{P}(x, y, \nu)$ *has the same generic fibre as the Breuil module* $\mathcal{P}'$, *where*

- $\mathcal{P}' = (k_E[u]/u^{e'p}) \cdot v' + (k_E[u]/u^{e'p}) \cdot w'$,
- $\mathrm{Fil}^1\mathcal{P}' = (k_E[u]/u^{e'p}) \cdot u^{e(p-1)}v' + (k_E[u]/u^{e'p}) \cdot (w' + u^{p(e-x)+y}\nu v')$,
- $\phi_1(u^{e(p-1)}v') = cv'$, $\phi_1(w' + u^{p(e-x)+y}\nu v') = dw'$,
- $\widehat{g}(v') = \omega^{\alpha-e}(g)v'$ *and* $\widehat{g}(w') = \omega^\beta(g)w'$ *for all* $g \in \mathrm{Gal}(K_1/K)$.

Proof. Consider the Breuil module $\mathcal{P}''$ defined by

- $\mathcal{P}'' = (k_E[u]/u^{e'p}) \cdot v'' + (k_E[u]/u^{e'p}) \cdot w''$,

- $\mathrm{Fil}^1\mathcal{P}'' = (k_E[u]/u^{e'p}) \cdot u^{e(p-1)}v'' + (k_E[u]/u^{e'p}) \cdot (u^{y(p-1)}w'' + u^{p(e-x)}\nu v'')$,
- $\phi_1(u^{e(p-1)}v'') = cv''$, $\phi_1(u^{y(p-1)}w'' + u^{p(e-x)}\nu v'') = dw''$,
- $\widehat{g}(v'') = \omega^{k+x-e}(g)v''$ and $\widehat{g}(w'') = \omega^l(g)w''$ for all $g \in \mathrm{Gal}(K_1/K)$.

(One checks without difficulty that this *is* a Breuil module. For instance the condition on the minimum degree of terms appearing in ν guarantees that $\mathrm{Fil}^1\mathcal{P}''$ contains $u^{e'}\mathcal{P}''$.) Note that $k + x \equiv \alpha \pmod{p-1}$, $l + y \equiv \beta \pmod{p-1}$. We claim that $\mathcal{P}$, $\mathcal{P}'$ and $\mathcal{P}''$ all have the same generic fibre. To see this, one can check directly that there is a morphism $\mathcal{P} \to \mathcal{P}''$ given by

$$v \mapsto u^{p(e-x)}v'', \;\; w \mapsto w'',$$

and a morphism $\mathcal{P}' \to \mathcal{P}''$ given by

$$v' \mapsto v'', \;\; w' \mapsto u^{py}w''.$$

By Proposition 8.3 of [Savitt 2004], it is enough to check that the kernels of these maps do not contain any free $k_E[u]/(u^{e'p})$-submodules, which is an immediate consequence of the inequalities $p(e-x)$, $py < e'p$. □

Remark 5.2.3. For future reference, while the classes in $H^1(G_K, \overline{\chi}_1\overline{\chi}_2^{-1})$ realised by $\mathcal{P}(x, y, \nu)$ and $\mathcal{P}'$ may not coincide, they differ at most by multiplication by a k_E-scalar. To see this, observe that the maps $\mathcal{P} \to \mathcal{P}''$ and $\mathcal{P}' \to \mathcal{P}''$ induce k_E-isomorphisms on the one-dimensional sub- and quotient characters.

We review the constraints on the integers x, y: they must lie between 0 and e, and if we let k, l be the residues of $\alpha - x$, $\beta - y \pmod{p-1}$ in the interval $[0, p-1)$ then we must have $\{\omega^k, \omega^l\} = \{\omega^{a_1}, \omega^{a_2}\}$. Call such a pair x, y *valid*.

Corollary 5.2.4. *Let x', y' be another valid pair. Suppose that $x' + y' \le e$ and $p(x' - x) + (y - y') \ge 0$. Then $\mathcal{P}(x, y, \nu)$ has the same generic fibre as $\mathcal{P}(x', y', \nu')$, where $\nu' = u^{p(x'-x)+(y-y')}\nu$.*

Proof. The Breuil module $\mathcal{P}(x', y', \nu')$ is well-defined: one checks, for example from the relation $l - k \equiv \beta - \alpha + x - y \pmod{p-1}$, that the congruence condition on the degrees of the nonzero terms in ν' is satisfied, while since $x' + y' \le e$ there is no condition on the lowest degrees appearing in ν'. Now the result is immediate from Proposition 5.2.2, since $u^{p(e-x)+y}\nu = u^{p(e-x')+y'}\nu'$. □

Recall that $x + y \equiv e \pmod{p-1}$, so that x and $e - y$ have the same residue modulo $p-1$. It follows that if x, y is a valid pair of parameters, then so is $e - y, y$; and similarly for $x, e - x$. Let X be the largest value of x over all valid pairs x, y, and similarly Y the smallest value of y. Then on the one hand $X \ge e - Y$ by definition of X, while on the other hand $e - X \ge Y$ by definition of Y. It follows that $X + Y = e$.

Corollary 5.2.5. *The module $\mathcal{P}(x, y, \nu)$ has the same generic fibre as $\mathcal{P}(X, Y, \mu)$ where $\mu \in k_E[u]/u^{e'p}$ has all nonzero terms of degree congruent to $\beta - \alpha + X - Y$ modulo $p-1$, and has all terms of degree less than $X(p-1)$, unless $\overline{\chi}_1 = \overline{\chi}_2$, in which case it may additionally have a term of degree $pX - Y$.*

Proof. Since $X + Y = e$ and $p(X - x) + (y - Y) \geq 0$ from the choice of X, Y, Corollary 5.2.4 shows that $\mathcal{P}(x, y, \nu)$ has the same generic fibre as some $\mathcal{P}(X, Y, \nu')$; by Proposition 5.2.1 there exists μ as in the statement such that $\mathcal{P}(x, y, \mu)$ has the same generic fibre as $\mathcal{P}(X, Y, \nu')$. (Note that if $\overline{\chi}_1 = \overline{\chi}_2$ then automatically $X \geq Y$, because in this case if (x, y) is a valid pair then so is (y, x).) $\square$

Proposition 5.2.6. *Let X be as above, that is, X is the maximal integer such that*

- $0 \leq X \leq e$, *and*
- *either* $\overline{\chi}_1|_{I_K} = \omega^{a_1+X}$ *or* $\overline{\chi}_1|_{I_K} = \omega^{a_2+X}$.

Then L_{flat} is an $\overline{\mathbb{F}}_p$-vector space of dimension at most X, unless $\overline{\chi}_1 = \overline{\chi}_2$, in which case it has dimension at most $X + 1$.

Proof. Let $L_{\text{flat},k_E} \subset L_{\text{flat}}$ consist of the classes η such that the containment $\eta \in L_{\text{flat}}$ is witnessed by a k_E-vector space scheme with generic fibre descent data. By Corollary 5.2.5 and Remark 5.2.3, these are exactly the classes arising from the Breuil modules $\mathcal{P}(X, Y, \mu)$ with k_E-coefficients as in Corollary 5.2.5. These classes form a k_E-vector space (since they are *all* the extension classes arising from extensions of $\mathcal{N}(Y)$ by $\mathcal{M}(X)$), and by counting the (finite) number of possibilities for μ we see that $\dim_{k_E} L_{\text{flat},k_E}$ is at most X or, when $\overline{\chi}_1 = \overline{\chi}_2$, $X + 1$.

Since $L_{\text{flat},k_E} \subset L_{\text{flat},k'_E}$ if $k_E \subset k'_E$ it follows easily that $L_{\text{flat}} = \cup_{k_E} L_{\text{flat},k_E}$ is an $\overline{\mathbb{F}}_p$-vector space of dimension at most X or $X + 1$, respectively. $\square$

We can now prove our main local result, the relation between L_{flat} and L_{crys}.

Theorem 5.2.7. *If either $a_1 - a_2 \neq p - 1$ or $\overline{\chi}_1 \overline{\chi}_2^{-1} \neq \overline{\epsilon}$, we have $L_{\text{flat}} = L_{\text{crys}}$.*

Proof. Before we begin the proof, we remind the reader that the spaces L_{crys} and L_{flat} depend on the fixed Serre weight a and the fixed representation $\bar{\rho}^{\text{ss}}$, and that we are free to vary J and δ in our arguments. By Proposition 4.2.5, we know that $L_{\text{crys}} \subset L_{\text{flat}}$, so by Proposition 5.2.6 it suffices to show that L_{crys} contains an $\overline{\mathbb{F}}_p$-subspace of dimension X (respectively $X + 1$ if $\overline{\chi}_1 = \overline{\chi}_2$). Since L_{crys} is the union of the spaces L_{χ_1,χ_2}, it suffices to show that one of these spaces has the required dimension. Let X be as in the statement of Proposition 5.2.6, so that X is maximal in $[0, e]$ with the property that either $\overline{\chi}_1|_{I_K} = \omega^{a_1+X}$ or $\overline{\chi}_1|_{I_K} = \omega^{a_2+X}$. Note that by the assumption that there is a decomposition $\text{Hom}(\mathbb{F}_p, \overline{\mathbb{F}}_p) = J \amalg J^c$, and an integer $0 \leq \delta \leq e - 1$ such that

$$\bar{\rho}|_{I_K} \cong \begin{pmatrix} \omega^\delta \prod_{\sigma \in J} \omega_\sigma^{a_1+1} \prod_{\sigma \in J^c} \omega_\sigma^{a_2} & * \\ 0 & \omega^{e-1-\delta} \prod_{\sigma \in J^c} \omega_\sigma^{a_1+1} \prod_{\sigma \in J} \omega_\sigma^{a_2} \end{pmatrix},$$

we see that if $X = 0$ then $\overline{\chi}_1|_{I_K} = \omega^{a_2}$.

If $\overline{\chi}_1|_{I_K} = \omega^{a_2+X}$ then we can take J to be empty and we take $\delta = X$; otherwise $X > 0$ and $\overline{\chi}_1|_{I_K} = \omega^{a_1+X}$, and we can take J^c to be empty and $\delta = X - 1$. In either case, we may define characters χ_1 and χ_2 as in Section 4.2, and we see from Lemma 4.2.2 that $\dim_{\overline{\mathbb{F}}_p} L_{\chi_1,\chi_2} = X$ unless $\overline{\chi}_1 = \overline{\chi}_2$, in which case it is $X + 1$. The result follows. □

As a consequence of this result, we can also address the question of the relationship between the different spaces L_{χ_1,χ_2} for a fixed Serre weight $a \in W^?(\bar{\rho})$. If e is large, then these spaces do not necessarily have the same dimension, so they cannot always be equal. However, it is usually the case that the spaces of maximal dimension coincide, as we can now see.

Corollary 5.2.8. *If either $a_1 - a_2 \neq p - 1$ or $\overline{\chi}_1\overline{\chi}_2^{-1} \neq \bar{\epsilon}$, then the spaces L_{χ_1,χ_2} of maximal dimension are all equal.*

Proof. In this case $\dim_{\overline{\mathbb{F}}_p} L_{\chi_1,\chi_2} = \dim_{\overline{\mathbb{F}}_p} L_{\text{crys}}$ by the proof of Theorem 5.2.7, so we must have $L_{\chi_1,\chi_2} = L_{\text{crys}}$. □

Finally, we determine L_{crys} in the one remaining case, where the spaces L_{χ_1,χ_2} of maximal dimension no longer coincide.

Proposition 5.2.9. *If $a_1 - a_2 = p - 1$ and $\overline{\chi}_1\overline{\chi}_2^{-1} = \bar{\epsilon}$, then $L_{\text{crys}} = H^1(G_K, \bar{\epsilon})$.*

Proof. We adapt the proof of [Barnet-Lamb et al. 2012, Lemma 6.1.6]. By twisting we can reduce to the case $(a_1, a_2) = (p-1, 0)$. Let L be a given line in $H^1(G_K, \bar{\epsilon})$, and choose an unramified character ψ with trivial reduction. Let χ be some fixed crystalline character of G_K with Hodge–Tate weights $p, 1, \dots, 1$ such that $\bar{\chi} = \bar{\epsilon}$. Let $E/\mathbb{Q}_p$ be a finite extension with ring of integers $\mathbb{O}$, uniformiser ϖ and residue field $\mathbb{F}$, such that ψ and χ are defined over E and L is defined over $\mathbb{F}$ (that is, there is a basis for L which corresponds to an extension defined over $\mathbb{F}$). Since any extension of 1 by $\chi\psi$ is automatically crystalline, it suffices to show that we can choose ψ so that L lifts to $H^1(G_K, \mathbb{O}(\psi\chi))$.

Let H be the hyperplane in $H^1(G_K, \mathbb{F})$ which annihilates L under the Tate pairing. Let $\delta_1 : H^1(G_K, \mathbb{F}(\bar{\epsilon})) \to H^2(G_K, \mathbb{O}(\psi\chi))$ be the map coming from the exact sequence

$$0 \to \mathbb{O}(\psi\chi) \xrightarrow{\varpi} \mathbb{O}(\psi\chi) \to \mathbb{F}(\bar{\epsilon}) \to 0$$

of G_K-modules. We need to show that $\delta_1(L) = 0$ for some choice of ψ.

Let δ_0 be the map

$$H^0(G_K, (E/\mathbb{O})(\psi^{-1}\chi^{-1}\epsilon)) \to H^1(G_K, \mathbb{F})$$

coming from the exact sequence

$$0 \to \mathbb{F} \to (E/\mathbb{O})(\psi^{-1}\chi^{-1}\epsilon) \xrightarrow{\varpi} (E/\mathbb{O})(\psi^{-1}\chi^{-1}\epsilon) \to 0$$

of G_K-modules. By Tate local duality, the condition that L vanishes under the map δ_1 is equivalent to the condition that the image of the map δ_0 is contained in H. Let $n \geq 1$ be the largest integer with the property that $\psi^{-1}\chi^{-1}\epsilon \equiv 1 \pmod{\varpi^n}$. Then we can write $\psi^{-1}\chi^{-1}\epsilon(x) = 1 + \varpi^n \alpha_\psi(x)$ for some function $\alpha_\psi : G_K \to \mathbb{O}$. Let $\overline{\alpha}_\psi$ denote $\alpha_\psi \pmod{\varpi} : G_K \to \mathbb{F}$. Then $\overline{\alpha}_\psi$ is a group homomorphism (that is a 1-cocycle), and the choice of n ensures that it is nontrivial. It is straightforward to check that the image of the map δ_0 is the line spanned by $\overline{\alpha}_\psi$. If $\overline{\alpha}_\psi$ is in H for some ψ, we are done. Suppose this is not the case. We break the rest of the proof into two cases.

Case 1: L is très ramifié. To begin, we observe that it is possible to have chosen ψ so that $\overline{\alpha}_\psi$ is ramified. To see this, let m be the largest integer with the property that $(\psi^{-1}\chi^{-1}\epsilon)|_{I_K} \equiv 1 \pmod{\varpi^m}$. Note that m exists since the Hodge–Tate weights of $\psi^{-1}\chi^{-1}\epsilon$ are not all 0. If $m = n$ then we are done, so assume instead that $m > n$. Let $g \in G_K$ be a fixed lift of Frob_K. We claim that

$$\psi^{-1}\chi^{-1}\epsilon(g) = 1 + \varpi^n \alpha_\psi(g) \quad \text{such that } \alpha_\psi(g) \not\equiv 0 \pmod{\varpi}.$$

In fact, if

$$\alpha_\psi(g) \equiv 0 \pmod{\varpi} \quad \text{then } \psi^{-1}\chi^{-1}\epsilon(g) \in 1 + \varpi^{n+1}\mathbb{O}_K.$$

Since $m > n$ we see that $\psi^{-1}\chi^{-1}\epsilon(G_K) \subset 1 + \varpi^{n+1}\mathbb{O}_K$ and this contradicts the selection of n. Now let ψ' be the unramified character sending our fixed g to $1 + \varpi^n \alpha_\psi(g)$. Then ψ' has trivial reduction, and after replacing ψ by $\psi\psi'$ we see that n has increased but m has not changed. After finitely many iterations of this procedure we have $m = n$, completing the claim.

Suppose, then, that $\overline{\alpha}_\psi$ is ramified. The fact that L is très ramifié implies that H does not contain the unramified line in $H^1(G_K, \mathbb{F})$. Thus there is a unique $\overline{x} \in \mathbb{F}^\times$ such that $\overline{\alpha}_\psi + u_{\overline{x}} \in H$ where $u_{\overline{x}} : G_K \to \mathbb{F}$ is the unramified homomorphism sending Frob_K to $\overline{x}$. Replacing ψ with ψ times the unramified character sending Frob_K to $(1 + \varpi^n x)^{-1}$, for x a lift of $\overline{x}$, we are done.

Case 2: L is peu ramifié. Making a ramified extension of $\mathbb{O}$ if necessary, we can and do assume that $n \geq 2$ (for example, replacing E by $E(\varpi^{1/2})$ has the effect of replacing n by $2n$). The fact that L is peu ramifié implies that H contains the unramified line. It follows that if we replace ψ with ψ times the unramified character sending Frob_K to $1 + \varpi$, then we are done (as the new $\overline{\alpha}_\psi$ will be unramified). □

6. Global consequences

6.1. We now deduce our main global results, using the main theorems of [Barnet-Lamb et al. 2011] together with our local results to precisely determine the set of Serre weights for a global representation in the totally ramified case.

Theorem 6.1.1. *Suppose that Hypothesis 2.2.5 holds. Let $a \in (\mathbb{Z}^2_+)^S_0$ be a Serre weight such that $\bar{r}$ is modular of weight a. Let w be a place of F dividing p, write $a_w = (a_1, a_2)$, and write ω for the unique fundamental character of I_{F_w} of niveau one. Then $a_w \in W^?(\bar{r}|_{G_{F_w}})$.*

Proof. Let e be the ramification degree of F_w. Suppose first that $\bar{r}|_{G_{F_w}}$ is irreducible. Then the proof of Lemma 5.5 of [Gee and Savitt 2011a] goes through unchanged, and gives the required result. So we may suppose that $\bar{r}|_{G_{F_w}}$ is reducible. In this case the proof of Lemma 5.4 of [ibid.] goes through unchanged, and shows that

$$\bar{r}|_{G_{F_w}} \cong \begin{pmatrix} \overline{\chi}_1 & * \\ 0 & \overline{\chi}_2 \end{pmatrix}$$

where $(\overline{\chi}_1\overline{\chi}_2)|_{I_K} = \omega^{a_1+a_2+e}$, and either $\overline{\chi}_1|_{I_K} = \omega^{a_1+z}$ or $\overline{\chi}_1|_{I_K} = \omega^{a_2+e-z}$ for some $1 \le z \le e$, so we are in the situation of Section 4.2. Consider the extension class in $H^1(G_{F_w}, \overline{\chi}_1\overline{\chi}_2^{-1})$ corresponding to $\bar{r}|_{G_{F_w}}$. By Proposition 2.2.9, either $a_1 - a_2 = p - 1$ and $\overline{\chi}_1\overline{\chi}_2^{-1} = \bar{\epsilon}$, or this extension class is in L_{flat}. In either case, by Theorem 5.2.7 and Proposition 5.2.9, the extension class is in L_{crys}, so that $a_w \in W^?(\bar{r}|_{G_{F_w}})$, as required. □

We remark that we have stated Theorem 6.1.1 only when $F_w/\mathbb{Q}_p$ is totally ramified for all places $w|p$ of F in order to avoid recalling the definition of Serre weights in any greater generality; however, the above argument would prove essentially the same result at any totally ramified place $w|p$ of F, even if not all places $w|p$ are totally ramified (just modify Proposition 2.2.9 suitably).

Combining Theorem 6.1.1 with Theorem 5.1.3 of [Barnet-Lamb et al. 2011], we obtain our main global result.

Theorem 6.1.2. *Suppose that Hypothesis 2.2.5 holds. Suppose further that $\zeta_p \notin F$ and $\bar{r}(G_{F(\zeta_p)})$ is adequate. Let $a \in (\mathbb{Z}^2_+)^S_0$ be a Serre weight. Then $a_w \in W^?(\bar{r}|_{G_{F_w}})$ for all places $w|p$ of F if and only if $\bar{r}$ is modular of weight a.*

Finally, we may apply our local results to the case of inner forms of GL_2, as considered in [Gee and Savitt 2011a]. Here is an example of the kind of theorem that one can prove. We refer the reader to [ibid.] for the notion of $\bar{\rho}$ as below being modular (of some weight).

Theorem 6.1.3. *Let F be a totally real field, let $p \ge 7$ be prime, and suppose that p is totally ramified in F, and that $[F(\zeta_p) : F] > 4$. Let $\bar{\rho} : G_F \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous modular representation, and suppose that $\bar{\rho}|_{G_{F(\zeta_p)}}$ is irreducible. Let $a \in \mathbb{Z}^2$ be a Serre weight. Let v be the unique place of F lying over p, and assume that $\bar{\rho}|^{\text{ss}}_{G_{F_v}} \not\cong \bar{\epsilon}\omega^{a_1} \oplus \omega^{a_2}, \bar{\epsilon}\omega^{a_2} \oplus \omega^{a_1}$. Then $\bar{\rho}$ is modular of weight a if and only if $a \in W^?(\bar{\rho}|_{G_{F_v}})$, where v is the unique place of F lying over p.*

Proof. This follows easily from Theorem 5.2.7 together with (the proof of) Corollary 7.3 of [ibid.], replacing the use of Theorem 7.1 of [ibid.] with an appeal to Theorem 6.1.9 of [Barnet-Lamb et al. 2012]. □

Acknowledgments

We would like to thank the anonymous referee for an extremely thorough reading of the paper, and for their helpful suggestions which have improved the exposition in many places. One of us (DS) thanks Fred Diamond for valuable discussions on closely related questions.

References

[Barnet-Lamb et al. 2011] T. Barnet-Lamb, T. Gee, and D. Geraghty, "Serre weights for rank two unitary groups", preprint, 2011. arXiv 1106.5586

[Barnet-Lamb et al. 2012] T. Barnet-Lamb, T. Gee, and D. Geraghty, "Congruences between Hilbert modular forms: constructing ordinary lifts", *Duke Math. J.* **161**:8 (2012), 1521–1580. Zbl 06050950

[Breuil et al. 2001] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, "On the modularity of elliptic curves over $\mathbb{Q}$: Wild 3-adic exercises", *J. Amer. Math. Soc.* **14**:4 (2001), 843–939. MR 2002d:11058 Zbl 0982.11033

[Buzzard et al. 2010] K. Buzzard, F. Diamond, and F. Jarvis, "On Serre's conjecture for mod ℓ Galois representations over totally real fields", *Duke Math. J.* **155**:1 (2010), 105–161. MR 2730374 Zbl 1227.11070

[Calegari 2012] F. Calegari, "Even Galois representations and the Fontaine–Mazur conjecture, II", *J. Amer. Math. Soc.* **25**:2 (2012), 533–554. MR 2869026 Zbl 06022636

[Clozel et al. 2008] L. Clozel, M. Harris, and R. Taylor, "Automorphy for some l-adic lifts of automorphic mod l Galois representations", *Publ. Math. Inst. Hautes Études Sci.* **108** (2008), 1–181. MR 2010j:11082 Zbl 1169.11020

[Conrad et al. 1999] B. Conrad, F. Diamond, and R. Taylor, "Modularity of certain potentially Barsotti–Tate Galois representations", *J. Amer. Math. Soc.* **12**:2 (1999), 521–567. MR 99i:11037 Zbl 0923.11085

[Gee 2011] T. Gee, "On the weights of mod p Hilbert modular forms", *Invent. Math.* **184**:1 (2011), 1–46. MR 2012g:11088 Zbl 05883620

[Gee and Kisin 2012] T. Gee and M. Kisin, "The Breuil–Mézard conjecture for potentially Barsotti–Tate representations", preprint, 2012. arXiv 1208.3179

[Gee and Savitt 2011a] T. Gee and D. Savitt, "Serre weights for mod p Hilbert modular forms: The totally ramified case", *J. Reine Angew. Math.* **660** (2011), 1–26. MR 2855818 Zbl 05987921

[Gee and Savitt 2011b] T. Gee and D. Savitt, "Serre weights for quaternion algebras", *Compos. Math.* **147**:4 (2011), 1059–1086. MR 2012h:11070 Zbl 05938390

[Gee et al. 2012] T. Gee, F. Herzig, and D. Savitt, "Explicit Serre weight conjectures", in preparation, 2012.

[Nekovář 1993] J. Nekovář, "On p-adic height pairings", pp. 127–202 in *Séminaire de théorie des nombres* (Paris, 1990–1991), edited by S. David, Progr. Math. **108**, Birkhäuser, Boston, MA, 1993. MR 95j:11050 Zbl 0859.11038

[Savitt 2004] D. Savitt, "Modularity of some potentially Barsotti–Tate Galois representations", *Compos. Math.* **140**:1 (2004), 31–63. MR 2004h:11051 Zbl 1053.11048

[Savitt 2005] D. Savitt, "On a conjecture of Conrad, Diamond, and Taylor", *Duke Math. J.* **128**:1 (2005), 141–197. MR 2006c:11060 Zbl 1101.11017

[Savitt 2008] D. Savitt, "Breuil modules for Raynaud schemes", *J. Number Theory* **128**:11 (2008), 2939–2950. MR 2010a:14083 Zbl 1222.14098

[Schein 2008] M. M. Schein, "Weights in Serre's conjecture for Hilbert modular forms: The ramified case", *Israel J. Math.* **166** (2008), 369–391. MR 2009e:11090 Zbl 1197.11063

[Snowden 2009] A. Snowden, "On two dimensional weight two odd representations of totally real fields", preprint, 2009. arXiv 0905.4266

[Thorne 2011] J. Thorne, "On the automorphy of l-adic Galois representations with small residual image", preprint, 2011. To appear in *J. Inst. Math. Jussieu.* arXiv 1107.5989

Communicated by Brian Conrad
Received 2011-07-02 Revised 2011-10-04 Accepted 2011-11-02

tobygee@gmail.com *Imperial College London, London, SW7 2AZ, United Kingdom*

tongliu@math.purdue.edu *Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, United States*

savitt@math.arizona.edu *Department of Mathematics, University of Arizona, 617 N. Santa Rita Ave., PO Box 210089, Tucson, AZ 85721-0089, United States*

ALGEBRA AND NUMBER THEORY 6:7 (2012)
dx.doi.org/10.2140/ant.2012.6.1561

Annihilating the cohomology of group schemes

Bhargav Bhatt

Our goal in this note is to show that cohomology classes with coefficients in finite flat group schemes can be killed by finite covers of the base scheme, and similarly for abelian schemes with "finite covers" replaced by "proper covers." We apply this result to commutative algebra to give by a conceptual proof of Hochster–Huneke's theorem on the existence of big Cohen–Macaulay algebras in positive characteristic; all previous proofs of this result were equational or cocycle-theoretic in nature.

1. Introduction

Given a scheme S with a sheaf G and class $\alpha \in H^n(S, G)$ for $n > 0$, a natural question one may ask is if there exist covers $\pi : T \to S$ such that $\pi^*\alpha = 0$? Of course, as stated, the answer is trivially yes as we may take T to be a disjoint union of suitable opens occurring in a Čech cocycle representing α. However, the question becomes interesting if we require geometric conditions on π, such as properness or even finiteness. Our goal is to study such questions for fppf cohomology in the case that G is either a finite flat commutative group scheme or an abelian scheme. Our main results are:

Theorem 1.1. *Let S be a noetherian excellent scheme, and let G be a finite flat commutative group scheme over S. Then classes in $H^n_{\mathrm{fppf}}(S, G)$ can be killed by finite surjective maps to S for $n > 0$.*

Theorem 1.2. *Let S be a noetherian excellent scheme, and let A be an abelian scheme over S. Then classes in $H^n_{\mathrm{fppf}}(S, A)$ can be killed by proper surjective maps to S for $n > 0$. Moreover, there exists an example of a normal affine scheme S that is essentially of finite type over $\mathbb{C}$, and an abelian scheme $A \to S$ with a class in $H^1_{\mathrm{fppf}}(S, A)$ that cannot be killed by finite surjective maps to S.*

MSC2010: primary 14L15; secondary 13D45, 14K05, 14F20.
Keywords: group schemes, abelian varieties, étale cohomology, fppf cohomology, big Cohen–Macaulay algebras.

We stress that there are no assumptions on the residue characteristics of S in either theorem above.

Our primary motivation for proving the preceding results was to obtain a better understanding of the Hochster–Huneke proof of the existence of big Cohen–Macaulay algebras in positive characteristic commutative algebra; see [Hochster and Huneke 1992]. We have succeeded in this endeavour as we can give a new and essentially *topological* proof of the Hochster–Huneke result by using the cohomology-annihilation results discussed above in lieu of the more traditional equational approaches; see Section 5 for more. We are hopeful that a similar approach, coupled with the constructions in [Fontaine 1994] of mixed characteristic rings admitting Frobenius actions, will eventually provide an approach to Hochster's homological conjectures in mixed characteristic commutative algebra; we refer the interested reader to [Hochster 2007] for further information.

An informal summary of the proofs: To prove Theorem 1.1, we first use a theorem of Raynaud to embed a finite flat group scheme into an abelian scheme; this permits a reduction to from fppf cohomology to étale cohomology by a theorem of Grothendieck. Next, using an observation due to Gabber, we reduce from étale cohomology to Zariski cohomology, and then we solve the problem by hand. For Theorem 1.2, we reduce as before to Zariski cohomology, and then solve the problem using de Jong's alterations results combined with an observation concerning rational sections of an abelian scheme over a regular base scheme. The example referred to in Theorem 1.2 is discussed in Section 6, and relies on a construction of Raynaud. Lastly, the Hochster–Huneke theorem is reproved by first reformulating it as a suitable cohomology-annihilation statement for the higher local cohomology of the structure sheaf, and then deducing this statement from Theorem 1.1 by using finite flat subgroup schemes of $\mathbb{G}_a$ defined by additive polynomials in Frobenius.

Notations and conventions. All group schemes occurring in this note are commutative; all the cohomology groups occurring in this note are computed in the fppf topology unless otherwise specified. For a scheme X, the big site of X equipped with the étale topology is denoted $(\mathrm{Sch}/X)_{\text{ét}}$, while the small site is denoted $X_{\text{ét}}$; similarly for other topologies like the fppf and Zariski topologies.

Organisation of this note. In Section 2 we recall Gabber's observation alluded to above. Using this observation, we prove Theorem 1.1 in Section 3, and the first half of Theorem 1.2 in Section 4. Next, in Section 5, we explain how to use Theorem 1.1 to give a new proof of the Hochster–Huneke theorem. We close in Section 6 by giving an example that illustrates the necessity of "proper" in the first half of Theorem 1.2 and finishes its proof.

2. An observation of Gabber

In this section, we recall a result of Gabber concerning the local structure of the étale topology. This observation permits reduction of étale cohomological considerations to those in finite flat cohomology and those in Zariski cohomology. We begin with an elementary lemma on extending covers that will be used repeatedly in the sequel.

Lemma 2.1. *Fix a noetherian scheme X. Given an open dense subscheme $U \to X$ and a finite (surjective) morphism $f : V \to U$, there exists a finite (surjective) morphism $\overline{f} : \overline{V} \to X$ such that $\overline{f}_U$ is isomorphic to f. Given a Zariski open cover $\mathcal{U} = \{j_i : U_i \to X\}$ with a finite index set, and finite (surjective) morphisms $f_i : V_i \to U_i$, there exists a finite (surjective) morphism $f : Z \to X$ such that f_{U_i} factors through f_i. The same claims hold if "finite (surjective)" is replaced by "proper (surjective)" everywhere.*

Proof. We first explain how to deal with the claims for finite morphisms. For the first part, Zariski's main theorem [Grothendieck 1966, Théorème 8.12.6] applied to the morphism $V \to X$ gives a factorisation $V \hookrightarrow W \to X$ where $V \hookrightarrow W$ is an open immersion, and $W \to X$ is a finite morphism. The scheme-theoretic closure $\overline{V}$ of V in W provides the required compactification in view of the fact that finite morphisms are closed.

For the second part, by the above, we may extend each $j_i \circ f_i : V_i \to X$ to a finite surjective morphism $\overline{f}_i : \overline{V}_i \to X$ such that $\overline{f}_i$ restricts to f_i over $U_i \hookrightarrow X$. Setting W to be the fibre product over X of all the $\overline{V}_i$ is then seen to solve the problem.

To deal with the case of proper (surjective) morphisms instead of finite (surjective), we repeat the same argument as above replacing the reference to Zariski's main theorem by one to Nagata's compactification theorem; see [Conrad 2007, Theorem 4.1]. $\square$

Next, we state Gabber's result (see [Hoobler 1982, Lemma 5; Stacks, 02LH]):

Lemma 2.2. *Let $f : U \to X$ be a surjective étale morphism of affine schemes. Then there exists a finite flat map $g : X' \to X$, and a finite Zariski open cover $\{U_i \hookrightarrow X'\}$ such that the natural map $\bigsqcup_i U_i \to X$ factors through $U \to X$.*

For completeness, we sketch a proof when X is *local*; this will be enough for applications.

Sketch of proof. We only explain the proof when $X = \mathrm{Spec}(A)$ is the spectrum of a local ring A, and $U = \mathrm{Spec}(B)$ is the spectrum of a local étale A-algebra B. The structure theorem for étale morphisms (see [Grothendieck 1962, Exposé I, Théorème 7.6]) implies that $B = C_{\mathfrak{m}}$ where

$$C = A[x]/(f(x)) \quad \text{with } f(x) = x^n + a_1 x^{n-1} + \cdots + a_n$$

a monic polynomial, and $\mathfrak{m} \subset C$ a maximal ideal with $f'(x) \notin \mathfrak{m}$. We define

$$D = A[x_1, \ldots, x_n]/(\sigma_i(x_1, \ldots, x_n) - (-1)^{n-i} a_i)$$

where $\sigma_1, \ldots, \sigma_n$ are the elementary symmetric polynomials in the x_i. This ring is finite free over A of rank $n!$, admits an action of S_n that is transitive on the maximal ideals, and formalises the idea that the coefficients of $f(x)$ can be written as elementary symmetric functions in its roots. In particular, there is a natural morphism $C \to D$ sending x to x_1. As both C and D are finite free over A, there is a maximal ideal $\mathfrak{m}_1 \subset D$ lying over $\mathfrak{m} \subset C$. Thus, there is a natural map $a : B \to D_{\mathfrak{m}_1}$. By the S_n-action, for every maximal ideal $\mathfrak{n} \subset D$, there is an automorphism $D \to D$ sending $\mathfrak{m}_1$ to $\mathfrak{n}$. Composing such an automorphism with a, we see that for every maximal ideal $\mathfrak{n} \subset D$, the structure map $A \to D_{\mathfrak{n}}$ factorises through $A \to B$ for some map $B \to D_{\mathfrak{n}}$; the claim follows. □

Actually, we use a slight weakening of Gabber's result — relaxing finite flat to finite surjective — that remains true when the schemes under consideration are no longer assumed to be affine.

Lemma 2.3. *Let $f : U \to X$ be a surjective étale morphism of noetherian schemes. Then there exists a finite surjective map $g : X' \to X$, and a finite Zariski open cover $\{U_i \hookrightarrow X'\}$ such that the natural map $\bigsqcup_i U_i \to X$ factors through $U \to X$.*

Proof. We can solve the problem locally on X by Lemma 2.2 and a "smearing out" argument. This means that there exists a Zariski open cover $\{V_i \hookrightarrow X\}$, finite surjective (even flat) maps $W_i \to V_i$, and Zariski covers $\{Y_{ij} \hookrightarrow W_i\}$ such that $\bigsqcup Y_{ij} \to V_i$ factors through $U \times_X V_i \to V_i$. By Lemma 2.1, we may find a single finite surjective map $W \to X$ such that $W \times_X V_i \to V_i$ factors through $W_i \to V_i$. Setting $X' = W$ and pulling back the covers $\{Y_{ij} \to W_i\}$ to $W \times_X V_i$ then solves the problem. □

3. The theorem for finite flat commutative group schemes

In this section we prove Theorem 1.1 following the plan explained in the introduction. To carry that program out, we first explain how to relate the fppf cohomology of finite flat group schemes to étale cohomology; it turns out that they are almost the same.

Proposition 3.1. *Let S be the spectrum of a strictly henselian local ring, and let G be a finite flat commutative group scheme over S. Then $H^i(S, G) = 0$ for $i > 1$.*

Proof. We first explain the idea informally. Using a theorem of Raynaud, we can embed G into an abelian scheme, which allows us to express the cohomology of G in terms of that of abelian schemes. As abelian schemes are smooth, a result of Grothendieck ensures that their fppf cohomology coincides with their étale cohomology. As the latter vanishes when S is strictly henselian, we obtain the desired conclusion.

Now for the details: a construction of Raynaud (see [Berthelot et al. 1982, Théorème 3.1.1]) gives the existence of an abelian scheme $A \to S$ and an S-closed immersion $G \hookrightarrow A$ of group schemes. By Deligne's theorem [Tate and Oort 1970, §1, Theorem], we have $G \subset A[n]$ where n is the order of G. The quotient map $A/G \to A/A[n] \simeq A$ of fppf sheaves is an $A[n]/G$-torsor. Since $A[n]/G$ is a finite group scheme [Raynaud 1967, Théorème 1.1 (v)], the map $A/G \to A$ is fppf locally representable by a finite morphism of schemes. Since the quotient A is a scheme, fppf descent for finite morphisms shows that A/G is also a scheme. The map $A/G \to A$ is finite, so A/G is proper over S and acquires the structure of an S-group scheme by functoriality. Using the faithful flatness of $A \to A/G$ (as it is a G-torsor) and $A \to S$, one concludes:

- $A/G \to S$ is faithfully flat by an elementary flatness argument.
- $A/G \to S$ has geometrically regular fibres as these fibres admit a finite flat cover that is smooth.
- $A/G \to S$ has geometrically connected fibres as these fibres are dominated by those of $A \to S$.

These properties show that $A/G \to S$ is an abelian scheme. Hence, we have a short exact sequence

$$0 \to G \to A \to A/G \to 0$$

of abelian sheaves on the fppf site of S relating the finite flat commutative group scheme G to the abelian schemes A and A/G. This gives rise to a long exact sequence

$$\cdots \to H^{n-1}(S, A/G) \to H^n(S, G) \to H^n(S, A) \to \cdots$$

of fppf cohomology groups. By Grothendieck's theorem [1968b, Théorème 11.7], fppf cohomology coincides with étale cohomology when the coefficients are smooth group schemes. Applying this to A and A/G shows $H^i(S, A) = H^i(S, A/G) = 0$ for $i > 0$ as S is strictly henselian. The claim about G now follows from the preceding exact sequence. □

Remark 3.2. Proposition 3.1 may be reformulated topologically to say for a scheme X and a finite flat group scheme $G \to X$, we have $\mathrm{R}^i f_* G = 0$ for $i \geq 2$, where $f : (\mathrm{Sch}/X)_{\mathrm{fppf}} \to (\mathrm{Sch}/X)_{\text{ét}}$ is the morphism of (big) topoi defined by viewing étale covers as fppf covers. The Leray spectral sequence then reduces to a long exact sequence

$$\cdots \to H^i_{\text{ét}}(X, G) \to H^i_{\mathrm{fppf}}(X, G) \to H^{i-1}_{\text{ét}}(X, \mathrm{R}^1 f_* G) \to \cdots .$$

Next, we explain how to deal with Zariski cohomology with coefficients in a finite flat group scheme.

Proposition 3.3. *Let S be a normal noetherian scheme, and let $G \to S$ be a finite flat commutative group scheme. Then $H^n_{\mathrm{Zar}}(S, G) = 0$ for $n > 0$.*

Proof. We may assume that S is connected. As constant sheaves on irreducible topological spaces are acyclic, it will suffice to show that G restricts to a constant sheaf on the small Zariski site of S, that is, that the restriction maps $G(S) \to G(U)$ are bijective for any nonempty open subset $U \hookrightarrow S$. Injectivity follows from the density of $U \hookrightarrow S$ and the separatedness of $G \to S$. To show surjectivity, we note that given a section $U \to G$ of G over U, we can simply take the scheme-theoretic closure of U in G to obtain an integral closed subscheme $S' \hookrightarrow G$ such that the projection map $S' \to S$ is finite and an isomorphism over U. By the normality of S, this forces $S' = S$. Thus, G restricts to a constant sheaf on S, as claimed. □

We can now complete the proof of Theorem 1.1 by following the outline sketched in the introduction.

Proof of Theorem 1.1. Let S be a noetherian excellent scheme, and let $G \to S$ be a finite flat commutative group scheme. We need to show that classes in $H^n(S, G)$ can be killed by finite covers for S for $n > 0$. We deal with the $n = 1$ case on its own, and then proceed inductively.

For $n = 1$, note that classes in $H^1(S, G)$ are represented by fppf G-torsors T over S. By faithfully flat descent for finite flat morphisms, such schemes $T \to S$ are also finite flat. Passing to the total space of T trivialises the G-torsor T. Therefore, classes in $H^1(S, G)$ can be killed by finite flat covers of S.

We now fix an integer $n > 1$ and a cohomology class $\alpha \in H^n(S, G)$. By Proposition 3.1, we know that there exists an étale cover of S over which α trivialises. By Lemma 2.3, after replacing S by a finite cover, we may assume that there exists a Zariski cover $\mathcal{U} = \{U_i \hookrightarrow S\}$ such that $\alpha|_{U_i}$ is Zariski locally trivial. The Čech spectral sequence for this cover is

$$H^p(\mathcal{U}, H^q(G)) \Rightarrow H^{p+q}(S, G)$$

where $H^q(G)$ is the Zariski presheaf $V \mapsto H^q(V, G)$. By construction, the class α comes from some $\alpha' \in H^{n-q}(\mathcal{U}, H^q(G))$ with $q < n$. The group $H^{n-q}(\mathcal{U}, H^q(G))$ is the $(n-q)$-th cohomology group of the standard Čech complex

$$\prod_i H^q(U_i, G) \to \prod_{i<j} H^q(U_{ij}, G) \to \cdots$$

By the inductive assumption and the fact that $q < n$, terms of this complex can be annihilated by finite covers of the corresponding schemes. By Lemma 2.1, we may refine these finite covers by one that comes from all of S. In other words, we can find a finite surjective cover $S' \to S$ such that $\alpha'|_{S'} = 0$. After replacing S with S', the Čech spectral sequence then implies that α comes from some $H^{n-q'}(\mathcal{U}, H^{q'}(G))$ with $q' < q$. Proceeding in this manner, we can reduce the second index q all the way down to 0, that is, assume that the class α lies in the image of the map

$$H^n(\mathcal{U}, G) \to H^n(S, G).$$

Now we are reduced to the situation in Zariski cohomology that was tackled in Proposition 3.3. □

Remark 3.4. The proof given above for Theorem 1.1 used the intermediary of abelian schemes to connect fppf cohomology and étale cohomology with coefficients in a finite flat commutative group scheme G (see Proposition 3.1). When the coefficient group scheme G is smooth (or equivalently étale), this reduction follows directly from Grothendieck's theorem. In general, one can avoid abelian schemes by using a trick due to Messing to embed the group scheme in a smooth *affine* group: any commutative finite flat S-group scheme G may be realised as a closed subgroup of $A = \mathrm{Res}_{G^\vee/S}(\mathbb{G}_m)$ where $G^\vee$ denotes the Cartier dual of G; the map $G \to A$ is the tautological one coming from the definition $G^\vee = \mathrm{Hom}(G, \mathbb{G}_m)$; see [Messing 1972, §IV.1] for more. One can then show that A and A/G are S-smooth and representable, so the rest of the proof of Proposition 3.1 goes through. We thank Brian Conrad for pointing this out.

Remark 3.5. If G is a finite flat group scheme over S which is not necessarily abelian, the H^1 part of Theorem 1.1 remains valid since one can trivialise a G-torsor $\pi : T \to S$ using the finite flat morphism π.

Example 3.6. We give an example showing that Zariski, étale, and fppf cohomologies can differ. Let $k = \mathbb{F}_p$, and $G = \mu_p \times \mu_n$ where n is prime to p.

- $H^1_{\mathrm{Zar}}(\mathrm{Spec}(k), G) = 0$. Indeed, $\mathrm{Spec}(k)$ is a Zariski point, so the higher (Zariski) cohomology of all sheaves vanishes.
- $H^1_{\text{ét}}(\mathrm{Spec}(k), G) = k^*/(k^*)^n$. This follows from the Kummer sequence

  $$0 \to \mu_n \to \mathbb{G}_m \to \mathbb{G}_m \to 0,$$

  Hilbert's theorem 90, and the fact that $\mu_p \simeq 0$ on the small étale site of k.
- $H^1_{\mathrm{fppf}}(\mathrm{Spec}(k), G) = k^*/(k^*)^n \times k^*/(k^*)^p$. This follows from the Kummer sequence for both μ_n and μ_p; we need the flat topology to get right exactness of the Kummer sequence for μ_p.

4. The theorem for abelian schemes

Our goal in this section is to prove the first half of Theorem 1.2. The arguments here essentially mirror those for finite flat commutative group schemes presented in

Section 3. The key difference is that annihilating Zariski cohomology requires more complicated constructions when the coefficients are abelian schemes. We handle this by proving a generalisation of Weil's extension lemma (see Proposition 4.2). This generalisation requires strong regularity assumptions on S and is one of the two places in our proof of Theorem 1.2 that we need *proper* covers instead of finite ones; the other is the case of H^1.

We begin by recording an elementary criterion for a map to an abelian variety to be constant.

Lemma 4.1. *Let A be an abelian variety over an algebraically closed field k, and let C be a reduced variety over k. Fix an integer ℓ invertible on k. A map $g : C \to A$ is constant if and only if it induces the* 0 *map $H^1_{ét}(A, \mathbb{Q}_\ell) \to H^1_{ét}(C, \mathbb{Q}_\ell)$.*

Proof. It suffices to show that a map like g that induces the 0 map on H^1 is trivial. As any k-variety is covered by curves, it suffices to show that the map g is constant on all curves in C. Thus, we reduce to the case that C is a curve. We may also clearly assume that C is normal, that is, smooth. Let $\overline{C}$ denote the canonical smooth projective model of C. Since A is proper, the map g factors through a map $\overline{g} : \overline{C} \to A$. Since C and $\overline{C}$ are normal, the map $\pi_1(C) \to \pi_1(\overline{C})$ is surjective. Hence, the map $H^1_{ét}(\overline{C}, \mathbb{Q}_\ell) \to H^1_{ét}(C, \mathbb{Q}_\ell)$ is injective. Thus, to answer the question, we may assume that $C = \overline{C}$ is a smooth projective curve.

Let $A \hookrightarrow \mathbb{P}^n$ be a closed immersion corresponding to a very ample line bundle $\mathscr{L}$. The map $g : C \to A$ will be constant if we can show that $g^*\mathscr{L}$ is not ample, that is, has degree 0. As the ℓ-adic cohomology of an abelian variety is generated in degree 1 (see [Milne 2008, §12]), the hypothesis on H^1 implies that the map $H^2_{ét}(A, \mathbb{Q}_\ell) \to H^2(C, \mathbb{Q}_\ell)$ is also 0. In particular, $g^*(c_1\mathscr{L}) = 0$, where $c_1(\mathscr{L}) \in H^2(A, \mathbb{Q}_\ell(1)) \simeq H^2(A, \mathbb{Q}_\ell)$ is the first Chern class of the line bundle $\mathscr{L}$. Since applying g^* commutes with taking the first Chern class, it follows that $c_1(g^*\mathscr{L}) = 0$, hence $g^*\mathscr{L}$ has degree 0 as desired. □

We now prove the promised extension theorem for maps into abelian schemes.

Proposition 4.2. *Let S be a regular connected excellent noetherian scheme, and let $f : A \to S$ be an abelian scheme. For any nonempty open $U \subset S$, the restriction map $A(S) \to A(U)$ is bijective.*

Proof. Let $j : U \to S$ denote the open immersion defined by U. The bijectivity of $A(S) \to A(U)$ will follow by taking global sections if we can show that the natural map of presheaves $a : A \to j_*(A|_U)$ is an isomorphism on the small Zariski site of S. As both the source and the target of a are actually sheaves for the étale topology on S, we may localise to assume that S is the spectrum of a strictly henselian local ring R. In this setting, we will show that $A(S) \to A(U)$ is bijective using ℓ-adic cohomology.

The injectivity of $A(S) \to A(U)$ follows from the density of $U \subset S$ and the separatedness of $A \to S$. To show surjectivity, by the valuative criterion of properness, we may assume that the complement $S \setminus U$ has codimension at least 2 in S. Let $s : U \to A$ be a section of A over U. By taking the normalised scheme-theoretic closure of $s(U) \subset A$, we obtain a proper birational map $p : S' \to S$ that is an isomorphism over U, and an S-map $i : S' \to A$ extending s over U. The desired surjectivity then reduces to showing that i is constant on the fibres of p. Since $p_*\mathcal{O}_{S'} = \mathcal{O}_S$, the rigidity lemma (see [Mumford et al. 1994, Proposition 6.1]) shows that it suffices to show that i collapses the reduced special fibre S'_s, where $s \in S$ is the closed point. By Lemma 4.1, it is enough to check that the induced map $H^1(A_s, \mathbb{Q}_\ell) \to H^1(S'_s, \mathbb{Q}_\ell)$ is trivial for some integer ℓ invertible on S. Note that we have the following commutative diagram:

$$\begin{array}{ccc} H^1(A, \mathbb{Q}_\ell) & \longrightarrow & H^1(A_s, \mathbb{Q}_\ell) \\ \downarrow & & \downarrow \\ H^1(S', \mathbb{Q}_\ell) & \longrightarrow & H^1(S'_s, \mathbb{Q}_\ell). \end{array}$$

The horizontal maps are isomorphisms by the proper base change theorem in étale cohomology (see [Deligne 1977, Arcata IV-1, Théorème 1.2]) as S is a strictly henselian local scheme. Hence, it suffices to show that $H^1(A, \mathbb{Q}_\ell) \to H^1(S', \mathbb{Q}_\ell)$ is 0. Since $H^1(S', \mathbb{Q}_\ell) = \mathrm{Hom}_{\mathrm{conts}}(\pi_1(S'), \mathbb{Q}_\ell)$, it suffices to check that $\pi_1(S') = 0$. As S' is normal, we know that $\pi_1(U) \to \pi_1(S')$ is surjective. Moreover, by Zariski–Nagata purity (see [Grothendieck 1968a, Exposé X, Théorème 3.4]), we know that $\pi_1(U) \simeq \pi_1(S)$ since $S \setminus U$ has codimension ≥ 2 in S. Since S is strictly henselian, we have $\pi_1(S) = 0$ and hence $\pi_1(S') = 0$ as desired. $\square$

Remark 4.3. The main idea for the proof of Proposition 4.2 comes from obstruction theory in topology. Consider the universal family $\pi : \mathcal{U}_g \to \mathcal{A}_g$ of abelian varieties over the stack $\mathcal{A}_g$ of abelian varieties. Proposition 4.2 can be rephrased as asking if every map $S \to \mathcal{A}_g$ with a specified lift $U \to \mathcal{U}_g$ over a dense open $U \subset S$ admits an extension $S \to \mathcal{U}_g$ provided S is smooth. Since the stack $\mathcal{U}_g$ is a classifying space for its fundamental group (since the same is true for $\mathcal{A}_g$ and the fibres of π), the answer at the level of homotopy types would be yes if and only if $\pi_1(U) \to \pi_1(\mathcal{U}_g)$ factors through $\pi_1(U) \to \pi_1(S)$. This is essentially what is verified above using purity; Lemma 4.1 allows us to go from this homotopy-theoretic conclusion to a geometric one.

Remark 4.4. Proposition 4.2 can be considered a generalisation of Weil's extension lemma when applied to abelian varieties. Recall that this lemma says that the domain of definition of rational maps from a smooth variety to a group variety has pure codimension 1. In case the target is proper, that is, an abelian variety A, this reduces to the statement that $A(X) \simeq A(U)$ for any smooth variety X, and dense open $U \hookrightarrow X$.

Remark 4.5. Our proof of Proposition 4.2 is topological as explained in Remark 4.3. As pointed out to us by János Kollár after the present work was completed, one can also give a more geometric proof of Proposition 4.2 as follows: a theorem of Abhyankar (see [Kollár 1996, §VI.1, Theorem 1.2]) implies that for any proper modification $p : S' \to S$ with S noetherian regular excellent, the positive dimensional fibres of p contain nonconstant rational curves. Applying this theorem to the closure S' of the graph of a rational map defined by a section $U \to A$ over an open $U \subset S$ gives our desired claim as abelian varieties do not contain rational curves. We prefer the cohomological approach as a slight variation on it (using cohomology of the structure sheaf $\mathbb{O}_S$ instead of the constant sheaf in the proof of Proposition 4.2 and Lemma 4.1) shows that Proposition 4.2 remains valid in characteristic 0 if S has rational singularities. This also suggests a question to which we do not know the answer: if S is a scheme in positive characteristic satisfying some definition of rational singularities (such F-rationality), does Proposition 4.2 hold for S?

Example 4.6. We give an example to show that the regularity condition on S cannot be weakened too much in Proposition 4.2. Let $(E, e) \subset \mathbb{P}^2$ be an elliptic curve, and let S be the affine cone on E with origin s. Note that S is a hypersurface singularity of dimension 2 with 0 dimensional singular locus. In particular, it is normal. Let $A = S \times E$ denote the constant abelian scheme on E over S. Then $U = S \setminus \{s\}$ can be identified with the total space of the $\mathbb{G}_m$-torsor $\mathbb{O}(-1)|_E - 0(E)$ over E. Thus, there exists a nonconstant section of $A(U)$. On the other hand, all sections $S \to A$ are constant. Indeed, every point in S lies on an $\mathbb{A}^1$ containing s. As all maps $\mathbb{A}^1 \to E$ are constant, the claim follows. Thus, we obtain an example of a normal hypersurface singularity S and an abelian scheme $A \to S$ such that the conclusion of Proposition 4.2 fails for S. Of course, S is not a rational singularity, a fact supported by Remark 4.5.

Next, we point out how to use Proposition 4.2 to prove the version of Theorem 1.2 involving Zariski cohomology under strong regularity assumptions on the base scheme S; the proof is trivial.

Corollary 4.7. *Let S be a regular excellent noetherian scheme, and let $f : A \to S$ be an abelian scheme. Then $H^n_{\mathrm{Zar}}(S, A) = 0$ for $n > 0$.*

Proof. By Proposition 4.2, we know that A restricts to a constant sheaf on the small Zariski site of each connected component of S. By the vanishing of the cohomology of a constant sheaf on an irreducible topological space, the claim follows. □

We are now in a position to complete the proof of Theorem 1.2.

Proof of Theorem 1.2. Let S be a noetherian excellent scheme, and let $A \to S$ be an abelian scheme. We will show that cohomology classes in $H^n(S, A)$ are killed by proper surjective maps by induction on n provided $n > 0$. We may assume that S is integral.

For $n = 1$, classes in $H^1(S, A)$ are represented by étale A-torsors T over S. As T is an fppf S-scheme, there exists a quasifinite dominant morphism $U \to S$ such that $T(U)$ is nonempty. By picking an S-map $U \to T$ and taking the closure of the image, we obtain a proper surjective cover $S' \to S$ such that $T(S')$ is not empty. This implies that the cohomology class associated to T dies on passage to S', proving the claim.

We next proceed exactly as in the proof of Theorem 1.1 to reduce down to the case of a Čech cohomology class associated to a Zariski cover. The only difference is that the references to Proposition 3.1 should be replaced by references to Grothendieck's theorem [1968b, Théorème 11.7] which, in particular, implies that cohomology classes in $H^n_{\mathrm{fppf}}(S, A)$ trivialise over an étale cover; we omit the details.

To show the claim for a Čech cohomology class associated to a Zariski cover, assume first that S is of finite type over $\mathbb{Z}$. In this case, thanks to de Jong's theorems [1997], we can find a proper surjective cover of S with regular total space. Passing to this cover and applying Corollary 4.7 then solves the problem. In the case that S is no longer of finite type over $\mathbb{Z}$, we reduce to the finite type case using approximation. Indeed, the data (S, A, α) comprising of the base scheme S, the abelian scheme $A \to S$, and a Čech cohomology class $\alpha \in H^n(\mathcal{U}, A)$ associated to a finite Zariski open cover $\mathcal{U}$ of S can be approximated by similar data with all schemes involved of finite type over $\mathbb{Z}$. Given such an approximating triple (S', A', α') with S' of finite type over $\mathbb{Z}$, we can find a proper surjective map $S'' \to S'$ killing α' by the earlier argument. By functoriality, the pullback $S'' \times_{S'} S \to S$ is a proper surjective cover of S killing α. □

Remark 4.8. Theorem 1.2 admits a topological reformulation as follows. Given a noetherian scheme S and an abelian scheme G over S, let $(\mathrm{Sch}/S)_{\mathrm{prop}}$, $(\mathrm{Sch}/S)_{\mathrm{fppf}}$ and $(\mathrm{Sch}/S)_{\mathrm{prop,fppf}}$ denote the (big) topoi associated to the category of schemes over S equipped with the topology generated respectively by proper surjective maps, fppf maps, and both proper surjective and fppf maps. There are natural forgetful maps of topoi $a : (\mathrm{Sch}/S)_{\mathrm{prop,fppf}} \to (\mathrm{Sch}/S)_{\mathrm{prop}}$ and $b : (\mathrm{Sch}/S)_{\mathrm{prop,fppf}} \to (\mathrm{Sch}/S)_{\mathrm{fppf}}$ of topoi. Given an abelian scheme $G \to S$, let G also denote the sheafification of the representable presheaf associated to G in all of the above topologies. Theorem 1.2 can be reformulated as saying that the sheaves $\mathrm{R}^i a_* G$ vanish for $i > 0$. Since schemes are sheaves for the fppf topology, one can easily show that $a_* a^* G = G$. Thus, Theorem 1.2 can be reformulated saying that $G \simeq \mathrm{R}a_* G$. Note a consequence: since cohomology on sites is computed using hypercovers by Verdier's theorem

[Artin and Mazur 1969, Theorem 8.16], we see that for a class $\alpha \in H^n(S_{\mathrm{fppf}}, G)$ with $n > 0$, there exists a proper hypercover $f_\bullet : T_\bullet \to S$ and a map of simplicial schemes $\phi : T_\bullet \to K(G, n)$ representing $b^*\alpha$. If G is instead a finite flat group scheme, then the same remarks apply for Theorem 1.1, except that we replace proper maps by finite ones.

5. An application: big Cohen–Macaulay algebras in positive characteristic

Let $(R, \mathfrak{m})$ be an excellent noetherian local domain containing $\mathbb{F}_p$. A fundamental theorem of Hochster–Huneke [1992] asserts that the absolute integral closure R^+ (the integral closure of R in a fixed algebraic closure of its fraction field) is a Cohen–Macaulay algebra. This result and the ideas informing it form the bedrock of tight closure theory and huge swathes of positive characteristic commutative algebra.

Our goal in this section is to give a new proof of the Hochster–Huneke theorem using Theorem 1.1. We hasten to remark that there already exist alternative proofs in the literature, all cocycle-theoretic or equational at the core. The approach adopted here follows closely the relatively recent approach from [Huneke and Lyubeznik 2007], the essential new feature being the use of cohomology-annihilation result proven in Theorem 1.1 in place of explicit cocycle manipulations.

We begin by recording a coherent cohomology-annihilation result one can deduce from Theorem 1.1; this can be considered as the analogue of the "equational lemma" of [Hochster and Huneke 1992]; see also [Huneke and Lyubeznik 2007, Lemma 2.2].

Proposition 5.1. *Let $(R, \mathfrak{m})$ be a noetherian excellent local $\mathbb{F}_p$-algebra, and let $M \subset H^i_{\mathfrak{m}}(R)$ be a Frobenius stable finite length R-submodule for some $i > 0$. Then there exists a module-finite extension $f : R \to S$ such that $f^*(M) = 0$ where $f^* : H^i_{\mathfrak{m}}(R) \to H^i_{\mathfrak{m}}(S)$ is the induced map.*

Proof. After normalising R, we may assume that $i > 1$. With $U = \mathrm{Spec}(R) - \{\mathfrak{m}\}$, we have a Frobenius equivariant identification

$$H^{i-1}(U, \mathcal{O}) \simeq H^i_{\mathfrak{m}}(R)$$

which allows us to view M as a submodule of $H^{i-1}(U, \mathcal{O})$. The Frobenius action endows $H^{i-1}(U, \mathcal{O})$ with the structure of a $R\{X^p\}$-module, where $R\{X^p\}$ is the noncommutative polynomial ring over R with one generator X^p satisfying the commutation relation $X^p r = r^p X^p$ for $r \in R$. The finite length assumption implies that for each $m \in M$, there exists some monic additive polynomial $g(X^p) \in R\{X^p\}$ such that $g(m) = 0$. As g is additive and monic, we have a short exact sequence

$$0 \to \ker(g) \to \mathcal{O} \xrightarrow{g} \mathcal{O} \to 0$$

of abelian sheaves on $\mathrm{Spec}(R)_{\mathrm{fppf}}$. Moreover, the monicity of g also shows that the sheaf $\ker(g)$ is representable by a finite flat commutative group scheme over $\mathrm{Spec}(R)$. As $g(m) = 0$, we see that m comes from a cohomology class $m' \in H^{i-1}(U, \ker(g))$. Since $i - 1 > 0$, Theorem 1.1 shows that there exists a finite surjective map $\pi : V \to U$ such that $\pi^* m' = 0$. Setting S to be the (global sections of the) normalisation of R in V is then seen to solve the problem. $\square$

Using Proposition 5.1, we can give a proof that R^+ is Cohen–Macaulay. The argument given below is based entirely on [Huneke and Lyubeznik 2007, Theorem 2.1] and simply recorded here for convenience.

Theorem 5.2. *Let $(R, \mathfrak{m})$ be a noetherian excellent local $\mathbb{F}_p$-domain that admits a dualising complex, and let R^+ be its absolute integral closure. Then R^+ is Cohen–Macaulay.*

Proof. We first briefly review local duality and set up some notation. The local duality functor D (sometimes referred to as Matlis duality) is defined by $\mathrm{Hom}_R(-, E)$ where E is an injective hull of the residue field $R/\mathfrak{m}$. Once a dualising complex $\omega_R^\bullet$ has been fixed (normalised as usual to have the dualising sheaf ω_R in homological degree $\dim(R)$), the hull E may be identified with $\mathrm{R}\Gamma_{\mathfrak{m}}(\omega_R^\bullet)$ in $D^b(R)$. This functor is exact, contravariant, length preserving (on finite length R-modules), and transforms ind-artinian R-modules to pro-artinian R-modules. Moreover, local duality asserts that for finite R-modules M, one has $D(\mathrm{Ext}_R^{-i}(M, \omega_R^\bullet)) \simeq H^i_{\mathfrak{m}}(M)$ and

$$D(H^i_{\mathfrak{m}}(M)) \simeq \widehat{\mathrm{Ext}_R^{-i}(M, \omega_R^\bullet)},$$

where $\widehat{N}$ denotes the $\mathfrak{m}$-adic completion of N. For more details, see [Hartshorne 1966, Chapter V, §6; Brodmann and Sharp 1998, §10].

We need the compatibility between duality and localisation, which we recall next. Let $d = \dim(R)$, and for a prime $\mathfrak{p} \in \mathrm{Spec}(R)$, we set $d_{\mathfrak{p}} = \dim(R_{\mathfrak{p}})$ and $c_{\mathfrak{p}} = d - d_{\mathfrak{p}}$. One has $(\omega_R^\bullet)_{\mathfrak{p}} \simeq \omega_{R_{\mathfrak{p}}}^\bullet[c_{\mathfrak{p}}]$, which leads to the formula

$$\mathrm{Ext}_R^{-i}(M, \omega_R^\bullet)_{\mathfrak{p}} \simeq \mathrm{Ext}_{R_{\mathfrak{p}}}^{c_{\mathfrak{p}}-i}(M_{\mathfrak{p}}, \omega_{R_{\mathfrak{p}}}^\bullet)$$

for finite R-modules M. This gives a direct connection between $H^i_{\mathfrak{m}}(R)$ and $H^{i-c_{\mathfrak{p}}}_{\mathfrak{p}}(R_{\mathfrak{p}})$ which will be exploited in the proof below; see [Grothendieck 1968a, Exposé VIII, Théorème 2.1] for another application.

To show that R^+ is Cohen–Macaulay, we will show by induction on d that there exists a module-finite extension $R \to S$ which kills local cohomology outside degree d. By the local duality as explained above, it suffices to find a module-finite extension $R \to S$ such that the induced map $\mathrm{Ext}_R^{-i}(S, \omega_R^\bullet) \to \mathrm{Ext}_R^{-i}(R, \omega_R^\bullet)$ is 0 for $i < d$. The case $d = 0$ being vacuous, we assume $d > 0$ and pick a nonnegative integer $i < d$. Let $\mathfrak{p}_1, \dots, \mathfrak{p}_n$ be the set of all of nonmaximal associated primes of $\mathrm{Ext}_R^{-i}(R, \omega_R^\bullet)$. For each such prime $\mathfrak{p}_j$, induction constructs a module-finite extension $R_{\mathfrak{p}_j} \to S_j$ that kills $H_{\mathfrak{p}_j}^{i-c_{\mathfrak{p}_j}}(R_{\mathfrak{p}_j})$; note that $i - c_{\mathfrak{p}_j} < d_{\mathfrak{p}_j}$ since $i < d$. Setting S to be the normalisation of R in a compositum of all the S_j then shows that the map $R \to S$ induces a map $f_* : \mathrm{Ext}_R^{-i}(S, \omega_R^\bullet) \to \mathrm{Ext}_R^{-i}(R, \omega_R^\bullet)$ whose localisation at $\mathfrak{p}_j$ is the 0 map (as it is a map of finite $R_{\mathfrak{p}_j}$-modules that is 0 after $\mathfrak{p}_j$-adic completion). Since the only other possible associated prime of $\mathrm{Ext}_R^{-i}(R, \omega_R^\bullet)$ is $\mathfrak{m}$, it follows that $\mathrm{im}(f_*)$ is a finite length submodule of $\mathrm{Ext}_R^{-i}(R, \omega_R^\bullet)$. By duality, the image of $f^* : H_\mathfrak{m}^i(R) \to H_\mathfrak{m}^i(S)$ is a finite length R-submodule M of $H_\mathfrak{m}^i(S)$ which is moreover Frobenius stable. Hence, the S-submodule of $H_\mathfrak{m}^i(S) \simeq H_{\mathfrak{m}S}^i(S)$ generated by M is also Frobenius stable with finite length. Since $\mathfrak{m}S$ is a finite colength ideal in S, Proposition 5.1 gives a module-finite extension $S \to T$ killing M; the composite $R \to T$ then kills $H_\mathfrak{m}^i(R)$. □

6. An example of a torsor not killed by finite covers

Theorem 1.2 lets one construct proper covers annihilating cohomology with coefficients in an abelian scheme. We will show that "proper" cannot be replaced by "finite" in the preceding statement.

Example 6.1. In [Raynaud 1970, Example 3.2, Chapter XIII], one finds an example of a semilocal normal connected noetherian affine S of dimension 2, an abelian scheme $A \to S$, and an A-torsor $X \to S$ (in the category of fppf sheaves) that is Zariski locally trivial, and defines an infinite order element $\alpha \in H^1(S, A)$. By transfers (see Corollary 6.3), it follows that α cannot be trivialised by passing to a finite cover $T \to S$.

Example 6.1 relies on the existence of certain "transfer" maps whose construction we now explain. Fix a normal connected base scheme S, and a locally finitely presented algebraic space $A \to S$ that represents an abelian sheaf on the category of S-schemes. Our goal is to explain why étale cohomology with coefficients in A carries natural pushforward maps. As a corollary, when $A \to S$ is smooth, we obtain pushforwards maps in fppf cohomology as well.

Proposition 6.2. *Let $f : T \to S$ be a finite surjective map with T normal and equidimensional. There exists a map of abelian sheaves $\mathrm{Nm}(f) : f_* f^* A \to A$ on the small étale site of S such that the composite*

$$A \xrightarrow{f^*} f_* f^* A \xrightarrow{\mathrm{Nm}(f)} A$$

is multiplication by $d = \deg(f)$.

Proposition 6.2 is well-known, but we do not know a reference, so we sketch a proof.

Sketch of proof. We first construct the map on global sections, and then show it sheafifies.

Assume first that T is connected, and $f : T \to S$ is generically Galois with group G of cardinality d. By normality of S, we identify $T/G \simeq S$, where the quotient T/G is computed in the category of algebraic spaces (or schemes). Given a T-point $a \in A(T)$, we obtain a natural map $T \to \mathrm{Map}(G, A) \simeq A^d$ given by $t \mapsto (g \mapsto a(g(t)))$. The group $S_d = S_{\#G}$ acts on $\mathrm{Map}(G, A)$, and the preceding map $T \to \mathrm{Map}(G, A)$ is equivariant for the natural embedding $G \to S_d$ given by left translation. Taking quotients as algebraic spaces, we get a map

$$b : S \simeq T/G \to A^d/S_d = \mathrm{Sym}^d(A).$$

The d-fold multiplication map $A^d \to A$ is an S_d-equivariant map to an algebraic space, and hence factors as $A^d \to \mathrm{Sym}^d(A) \xrightarrow{m} A$. Composing m with b gives a map $S \to A$ that we declare to be the norm $\mathrm{Nm}(f)(S)(a)$.

Assume now that T is connected and $f : T \to S$ has degree d, but is not necessarily generically Galois. Then the arguments of [Suslin and Voevodsky 1996, Theorem 6.7, page 81] ensure that $\mathrm{Sym}^d(T) \to S$ has a natural section s constructed from f. Given $a \in A(T)$, we obtain a map

$$S \xrightarrow{s} \mathrm{Sym}^d(T) \xrightarrow{\mathrm{Sym}^d(a)} \mathrm{Sym}^d(A) \xrightarrow{m} A$$

that we declare to be the norm $\mathrm{Nm}(f)(S)(a)$.

In the general case, we perform the preceding construction on each connected component of T. The composite map $A(S) \to A(T) \to A(S)$ is easily seen to be multiplication by d using the fact that $A \xrightarrow{\Delta} A^d \to \mathrm{Sym}^d(A) \xrightarrow{m} A$ is multiplication by d. Finally, we observe that all hypotheses are stable under étale base change on S, so the preceding construction gives map of sheaves on the small étale site of S. $\square$

Corollary 6.3. *Let $f : T \to S$ be a finite surjective morphism with T normal and equidimensional. Assume that $A \to S$ is* smooth. *Then there exist pushforward maps $H^i_{\mathrm{fppf}}(T, A) \to H^i_{\mathrm{fppf}}(S, A)$ such that the composite*

$$H^i_{\mathrm{fppf}}(S, A) \to H^i_{\mathrm{fppf}}(T, A) \to H^i_{\mathrm{fppf}}(S, A)$$

is multiplication by d.

Proof. Let $f^{\text{ét}} : T_{\text{ét}} \to S_{\text{ét}}$ be the induced map of small étale sites. Acyclicity for finite morphisms shows that $H^i_{\text{ét}}(S, f^{\text{ét}}_* f^{\text{ét},*}A) = H^i_{\text{ét}}(T, f^{\text{ét},*}A)$. Grothendieck's theorem [1968b, Théorème 11.7] and the smoothness of A show that

$$H^i_{\text{ét}}(T, f^{\text{ét},*}A) = H^i_{\mathrm{fppf}}(T, A),$$

where the right hand side is defined by viewing A as a sheaf on the big fppf site of $(\mathrm{Sch}/S)_{\mathrm{fppf}}$. The claim now follows from Proposition 6.2. $\square$

Acknowledgements

The results presented here formed a part of my doctoral thesis written under Johan de Jong, and I would like to thank him for stimulating discussions. I would also like to thank János Kollár for a very useful conversation, and Brian Conrad for suggesting simplifications to some convoluted arguments (especially in Proposition 3.1) in an earlier version of the note. Finally, thanks are due to the anonymous referee for spotting gaps and errors in an earlier version of this preprint.

References

[Artin and Mazur 1969] M. Artin and B. Mazur, *Étale homotopy*, Lecture Notes in Mathematics **100**, Springer, Berlin, 1969. MR 39 #6883 Zbl 0182.26001

[Berthelot et al. 1982] P. Berthelot, L. Breen, and W. Messing, *Théorie de Dieudonné cristalline, II*, Lecture Notes in Mathematics **930**, Springer, Berlin, 1982. MR 85k:14023 Zbl 0516.14015

[Brodmann and Sharp 1998] M. P. Brodmann and R. Y. Sharp, *Local cohomology: an algebraic introduction with geometric applications*, Cambridge Studies in Advanced Mathematics **60**, Cambridge University Press, 1998. MR 99h:13020 Zbl 0903.13006

[Conrad 2007] B. Conrad, "Deligne's notes on Nagata compactifications", *J. Ramanujan Math. Soc.* **22**:3 (2007), 205–257. MR 2009d:14002 Zbl 1142.14001

[Deligne 1977] P. Deligne, *Cohomologie étale* (SGA $4\frac{1}{2}$), Lecture Notes in Math. **569**, Springer, Berlin, 1977. MR 57 #3132

[Fontaine 1994] J.-M. Fontaine, "Le corps des périodes p-adiques", pp. 59–111 Astérisque **223**, Société Mathématique de France, Paris, 1994. MR 95k:11086 Zbl 0940.14012

[Grothendieck 1962] A. Grothendieck, *Revêtements étales et groupe fondamental: Séminaire de géométrie algébrique du Bois Marie* 1960–61 (SGA 1), Lecture Notes in Math. **224**, Springer, Berlin, 1962. Reprinted Soc. math. de France, Paris, 2003. MR 50 #7129 Zbl 1039.14001

[Grothendieck 1966] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III", *Inst. Hautes Études Sci. Publ. Math.* 28 (1966), 255. MR 36 #178 Zbl 0144.19904

[Grothendieck 1968a] A. Grothendieck, *Cohomologie locale des faisceaux cohérents et théorèmes de Lefschetz locaux et globaux* (SGA 2), Advanced Stud. in Pure Math., North-Holland, Amsterdam, 1968. MR 57 #16294 Zbl 0159.50402

[Grothendieck 1968b] A. Grothendieck, "Le groupe de Brauer, III: exemples et compléments", pp. 88–188 in *Dix exposés sur la cohomologie des schémas*, North-Holland, Amsterdam, 1968. MR 39 #5586c Zbl 0198.25901

[Hartshorne 1966] R. Hartshorne, *Residues and duality*, Lecture Notes in Mathematics **20**, Springer, Berlin, 1966. MR 36 #5145 Zbl 0212.26101

[Hochster 2007] M. Hochster, "Homological conjectures, old and new", *Illinois J. Math.* **51**:1 (2007), 151–169. MR 2008j:13034 Zbl 1127.13010

[Hochster and Huneke 1992] M. Hochster and C. Huneke, "Infinite integral extensions and big Cohen–Macaulay algebras", *Ann. of Math.* (2) **135**:1 (1992), 53–89. MR 92m:13023 Zbl 0753.13003

[Hoobler 1982] R. T. Hoobler, "When is $\mathrm{Br}(X) = \mathrm{Br}'(X)$?", pp. 231–244 in *Brauer groups in ring theory and algebraic geometry* (Wilrijk, 1981), edited by F. M. J. van Oystaeyen and A. H. Verschoren, Lecture Notes in Math. **917**, Springer, Berlin, 1982. MR 83g:14006 Zbl 0491.14013

[Huneke and Lyubeznik 2007] C. Huneke and G. Lyubeznik, "Absolute integral closure in positive characteristic", *Adv. Math.* **210**:2 (2007), 498–504. MR 2008d:13005 Zbl 1109.13003

[de Jong 1997] A. J. de Jong, "Families of curves and alterations", *Ann. Inst. Fourier* (*Grenoble*) **47**:2 (1997), 599–621. MR 98f:14019 Zbl 0868.14012

[Kollár 1996] J. Kollár, *Rational curves on algebraic varieties*, Ergebnisse der Math. (3) **32**, Springer, Berlin, 1996. MR 98c:14001 Zbl 0877.14012

[Messing 1972] W. Messing, *The crystals associated to Barsotti–Tate groups: with applications to abelian schemes*, Lecture Notes in Mathematics **264**, Springer, Berlin, 1972. MR 50 #337 Zbl 0243.14013

[Milne 2008] J. Milne, "Abelian varieties", preprint, 2008, http://www.jmilne.org/math/CourseNotes/av.html.

[Mumford et al. 1994] D. Mumford, J. Fogarty, and F. Kirwan, *Geometric invariant theory*, 3rd ed., Ergebnisse der Math. (2) **34**, Springer, Berlin, 1994. MR 95m:14012 Zbl 0797.14004

[Raynaud 1967] M. Raynaud, "Sur le passage au quotient par un groupoïde plat", *C. R. Acad. Sci. Paris Sér. A-B* **265** (1967), A384–A387. MR 36 #5147 Zbl 0193.48901

[Raynaud 1970] M. Raynaud, *Faisceaux amples sur les schémas en groupes et les espaces homogènes*, Lecture Notes in Mathematics **119**, Springer, Berlin, 1970. MR 41 #5381 Zbl 0195.22701

[Stacks] Stacks, The Stacks project, http://math.columbia.edu/algebraic_geometry/stacks-git/.

[Suslin and Voevodsky 1996] A. Suslin and V. Voevodsky, "Singular homology of abstract algebraic varieties", *Invent. Math.* **123**:1 (1996), 61–94. MR 97e:14030 Zbl 0896.55002

[Tate and Oort 1970] J. Tate and F. Oort, "Group schemes of prime order", *Ann. Sci. École Norm. Sup.* (4) **3** (1970), 1–21. MR 42 #278 Zbl 0195.50801

Communicated by Brian Conrad
Received 2011-09-26 Revised 2011-12-06 Accepted 2012-01-03

bhattb@umich.edu *Department of Mathematics, University of Michigan, 2074 East Hall, 530 Church Street, Ann Arbor, MI 48109, United States*

Current address: *School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, United States*
http://www-personal.umich.edu/~bhattb/

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.

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

Volume 6 No. 7 2012

1937-0652(2012)6:7;1-8