Algebra & Number Theory

Volume 6

2012

No. 4

mathematical sciences publishers

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 6:4(2012)

Spherical varieties and integral representations of L-functions

Yiannis Sakellaridis

We present a conceptual and uniform interpretation of the methods of integral representations of L-functions (period integrals, Rankin–Selberg integrals). This leads to (i) a way to classify such integrals, based on the classification of certain embeddings of spherical varieties (whenever the latter is available), (ii) a conjecture that would imply a vast generalization of the method, and (iii) an explanation of the phenomenon of "weight factors" in a relative trace formula. We also prove results of independent interest, such as the generalized Cartan decomposition for spherical varieties of split groups over p-adic fields (following an argument of Gaitsgory and Nadler).

1. Introduction

1.1. ***Goals.*** The study of automorphic L-functions (and their special values at distinguished points, or *L-values*) is very central in many areas of present-day number theory, and an incredible variety of methods has been developed in order to understand the properties of these mysterious objects and their deep links with seemingly unrelated arithmetic invariants. Oddly enough, notwithstanding their

MSC2000: primary 11F67; secondary 22E55, 11F70.

Keywords: automorphic L-functions, spherical varieties, Rankin–Selberg, periods of automorphic forms.

elegant and very general definition by Langlands in terms of Euler products, virtually all methods for studying them depart from an integral construction of the form:

> A suitable automorphic form (considered as a function on the automorphic quotient $[G] := G(k) \setminus G(\mathbb{A}_k)$), integrated against a suitable distribution on $G(k) \setminus G(\mathbb{A}_k)$, is equal to a certain L-value.

For "geometric" automorphic forms, such an integral can often be expressed as a pairing between elements in certain homology and cohomology groups, but the essence remains the same. Given the importance of such methods, it appears as a paradox that there is no general theory of integral representations of L-functions, and in fact, they are often considered as "accidents".

In this article, I present a uniform interpretation of a large array of such methods, which includes Tate integrals, period integrals and Rankin–Selberg integrals. This interpretation leads to the first systematic classification of such integrals, based on the classification of certain spherical varieties (see Sections 4 and 5). Moreover, it naturally gives rise to a very general conjecture (Conjecture 3.2.2), whose proof would lead to a vast extension of the method and would allow us to study many more L-functions than are within our reach at this moment. Finally, it explains phenomena that have been observed in the theory of the relative trace formula, in a way that is well suited to the geometric methods employed in the proof of the fundamental lemma by Ngô [2010]. In the course of the article we also prove some results that can be of independent interest, including results on the orbits of hyperspecial and congruence subgroups on the p-adic points of a spherical variety (Theorems 2.3.8 and 2.3.10).

The main idea is based on the well-known principle that a "multiplicity-freeness" property usually underlies integral constructions of L-functions. For our present purposes, a multiplicity-freeness property can be taken to mean that a suitable space of functions $\mathcal{S}(X)$ on a $G(\mathbb{A}_k)$-space X admits at most one, up to constants, morphism into any irreducible admissible representation π of $G(\mathbb{A}_k)$. Here G denotes a connected reductive algebraic group over a global field k, and $\mathbb{A}_k$ denotes the ring of adeles of k. Such spaces arise as the adelic points of spherical varieties. By definition, a spherical variety for G is a normal variety with a G-action such that, over the algebraic closure, the Borel subgroup of G has a dense orbit. Let X be an *affine* spherical variety, and denote by X^+ the open G-orbit on X. A second principle behind the main idea is based on ideas around the geometric Langlands program, according to which the correct "Schwartz space" $\mathcal{S}(X)$ of functions to consider (which are actually functions on $X^+(\mathbb{A}_k)$, not $X(\mathbb{A}_k)$) should be one reflecting the geometry and singularities of X. Then, for every cuspidal automorphic representation π of G with "sufficiently positive" central character, there is a natural pairing $\mathcal{P}_X : \mathcal{S}(X(\mathbb{A}_k)) \otimes \pi \to \mathbb{C}$. The weak version of our Conjecture 3.2.4 asserts that this pairing admits meromorphic continuation to all π. (A stronger version, 3.2.2, states that an "Eisenstein series" construction, obtained by summing over the k-points of X and integrating against characters of a certain torus acting on X, has meromorphic continuation.) Then, assuming the multiplicity-freeness property, one expects the pairing to be associated to some L-value of π.

If our variety is of the form $H \setminus G$ with H a reductive subgroup of G, then from this construction we recover in Section 4.2 the period integral of automorphic forms over $H(k) \setminus H(\mathbb{A}_k)$. More generally, if X is fibered over such a variety and the fibers are (related to) flag varieties, then we can prove meromorphic continuation using the meromorphic continuation of Eisenstein series, and we recover in Section 4.4 integrals of "Rankin–Selberg" type. Thus, we reduce the problem of finding Rankin–Selberg integrals to the problem of classifying affine spherical varieties with a certain geometry. For smooth affine spherical varieties, this geometric problem has been solved by Knop and Van Steirteghem [2006]. By inspection of their tables, we recover in Section 5 some of the best-known constructions, such as those of Rankin [1939] and Selberg [1940], Godement and Jacquet [1972], Bump and Friedberg [1990], all spherical period integrals, as well as some new ones.

In Section 4.5 we give an example, involving the tensor product L-function of n cuspidal representations on GL_2, to support the point of view that the basic object giving rise to an Eulerian integral related to an L-function is the spherical variety X and not a geometry related to flag varieties. Finally, in Section 6 we apply these ideas to the relative trace formula to show that certain "weight factors" that have appeared in examples of this theory and are often considered an "anomaly" can, in fact, be understood using the notion of Schwartz spaces.

1.2. ***Background on the methods.*** To an automorphic representation $\pi \simeq \bigotimes'_v \pi_v$ of a reductive group G over a global field k, and to an algebraic representation ρ of its Langlands dual group ${}^LG$, Langlands attached a complex L-function $L(\pi, \rho, s)$, defined for s in some right-half plane of the complex plane as the product, over all places v, of local factors $L_v(\pi_v, \rho, s)$.[1]

Despite the beauty of its generality, the definition is of little use when attempting to prove analytic properties of L-functions, such as their meromorphic continuation and functional equation. Such properties are usually obtained by integration techniques, namely presenting the L-function as some integral transform of an element in the space of the given automorphic representation. Such methods in fact predate Langlands by more than a century, but the most definitive construction (since every

[1] At ramified places and for most ρ, the definition still depends on the local functoriality conjectures.

automorphic L-function should be a GL_n L-function) was studied by Godement and Jacquet [1972] (generalizing Tate's construction [1967] for GL_1), who proved the analytic continuation and functional equation of $L(\pi, s) := L(\pi, \mathrm{std}, s)$, where π is an automorphic representation of $G = \mathrm{GL}_n$ and std is the standard representation of ${}^LG = \mathrm{GL}_n(\mathbb{C}) \times \mathrm{Gal}(\bar{k}/k)$. Their method relies on proving the equality

$$L(\pi, s - \tfrac{1}{2}(n-1)) = \int_{\mathrm{GL}_n(\mathbb{A}_k)} \langle \pi(g)\phi, \tilde{\phi} \rangle \Phi(g) |\det(g)|^s \, dg, \tag{1-1}$$

with ϕ a suitable vector in π, $\tilde{\phi}$ a suitable vector in its contragredient and Φ a suitable function in $\mathcal{S}(\mathrm{Mat}_n(\mathbb{A}_k))$, the Schwartz space of functions on $\mathrm{Mat}_n(\mathbb{A}_k)$. The main analytic properties of $L(\pi, \rho, s)$, then, follow from Fourier transform on the Schwartz space and the Poisson summation formula.

Several decades before, Hecke showed that the standard L-function of a cuspidal automorphic representation on GL_2 (with, say, trivial central character) has a presentation as a *period integral*, which in adelic language reads

$$L(\pi, s + \tfrac{1}{2}) = \int_{k^\times \backslash \mathbb{A}_k^\times} \phi\left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \right) |a|^s \, da, \tag{1-2}$$

where ϕ is again a suitable vector in the automorphic representation under consideration.

Period integrals (by which we mean integrals over the orbit of some subgroup on the automorphic space $G(k)\backslash G(\mathbb{A}_k)$, possibly against a character of that subgroup) have since been studied extensively, although there are still many open conjectures about their relation to L-functions; see, for instance, [Ichino and Ikeda 2010]. Still, they form perhaps the single class of examples where we have a general principle answering the question, How do we write down an integral with good analytic properties, which is related to some L-function (or L-value)? Piatetski-Shapiro discussed this in [1975], and suggested that the period integral of a cusp form on a group G over a subgroup H (against, perhaps, an analytic family δ_s of characters of H as in (1-2)) should always be related to some L-value if the subgroup H enjoys a "multiplicity-one" property: $\dim \mathrm{Hom}_{H(\mathbb{A}_k)}(\pi, \delta_s) \le 1$ for every irreducible representation π of $G(\mathbb{A}_k)$ and (almost) every s.

The method of periods usually fails when the subgroup H is nonreductive, the reason being that, typically, the group $H(\mathbb{A}_k)$ has no closed orbits on $G(k)\backslash G(\mathbb{A}_k)$. Therefore there is no a priori reason that the period integral should have nice analytic properties (as the character δ_s varies), and one can in fact check in examples (see, for instance, Example 3.2.1) that for values of s such that the period integral converges, it does not represent an L-function.

In a different vein, Rankin [1939] and Selberg [1940] independently discovered an integral representing the tensor product L-function of two cuspidal automorphic representations of GL_2. The integral uses as auxillary data an Eisenstein series on GL_2 and has the form

$$L(\pi_1 \times \pi_2, \otimes, s) = \int_{\mathrm{PGL}_2(k)\backslash\mathrm{PGL}_2(\mathbb{A}_k)} \phi_1(g)\phi_2(g)E(g,s)\,dg$$

with suitable $\phi_1 \in \pi_1$ and $\phi_2 \in \pi_2$.

Later, this method was taken up by Jacquet, Piatetski-Shapiro, Shalika, Rallis, Gelbart, Ginzburg, Bump, Friedberg and many others, in order to construct numerous examples of automorphic L-functions expressed as integrals of cusp forms against Eisenstein series, with important corollaries for every such expression discovered. Despite the abundance of examples, however, there has not been a systematic understanding of how to produce an integral representing an L-function.

1.3. ***Schwartz spaces and X-Eisenstein series.*** While the method of Godement and Jacquet can also be phrased in the language of Rankin–Selberg integrals (see [Gelbart et al. 1987]), the fact that no systematic theory of these constructions exists has led many authors to consider them as coincidental or to seek direct generalizations of [Godement and Jacquet 1972], as being a "more canonical" construction [Braverman and Kazhdan 2000]. We adopt a different point of view that treats Godement–Jacquet, Rankin–Selberg, and period integrals as parts of the same concept, in fact a concept that should be much more general!

The basic object here is an affine spherical variety X of the group G. The reason that such varieties are suitable is that they are related to the "multiplicity-free" property discussed above. For instance, in the category of *algebraic* representations, the ring of regular functions $k[X]$ of an affine G-variety is multiplicity-free if and only if the variety is spherical. In the p-adic setting and for unramified representations, questions of multiplicity were systematically examined in [Sakellaridis 2008; 2009], and of course in special cases such questions have been examined in much greater detail; see, for example, [Prasad 1990].

The main idea is to associate to every affine spherical variety a space of distributions on $G(k) \backslash G(\mathbb{A}_k)$ that should have "good analytic properties". For reasons of convenience we set up our formulations so that the analytic problem does not have to do with varying a character of some subgroup H (the isotropy subgroup of a "generic" point on X), but with varying a cuspidal automorphic representation of G. For instance, to the Hecke integral (for PGL_2) we do not associate the variety $\mathbb{G}_m \backslash \mathrm{PGL}_2$, but the variety $X = \mathrm{PGL}_2$ under the $G = \mathbb{G}_m \times \mathrm{PGL}_2$-action. Our distributions (in fact, smooth functions) on $G(k) \backslash G(\mathbb{A}_k)$ come from a "Schwartz space" of functions on $X^+(\mathbb{A}_k)$ via a theta series construction (that is, summation over k-points of X^+). Here X^+ denotes the open G-orbit on X. The main conjecture, 3.2.2, then states that the integral of these X-theta series against central idele class characters (I call this integral an X-Eisenstein series), originally defined in some domain of convergence, has meromorphic continuation everywhere. Under added assumptions on X (related to the multiplicity-freeness property mentioned above), the pairings of X-theta series with automorphic forms should be related, in a suitable sense, to automorphic L-functions or special values of those.

The geometric Langlands program provides ideas that allow us to speculate on the form of these Schwartz spaces, motivated also by the work of Braverman and Kazhdan [1999; 2002] on the special case that X is the affine closure of $[P, P]\backslash G$, where P is a parabolic subgroup. Let us discuss this work: The prototype here is the case $X^+ = U\backslash \mathrm{SL}_2 = \mathbb{A}^2 \smallsetminus \{0\}$ (where U denotes a maximal unipotent subgroup) and $X = \mathbb{A}^2$ (two-dimensional affine space). The Schwartz space is the usual Schwartz space on $X(\mathbb{A}_k)$ which, by definition, is the restricted tensor product $\mathcal{S}(X(\mathbb{A}_k)) := \bigotimes'_v (\mathcal{S}(k_v^2) : \Phi_v^0)$, where for finite places k_v with rings of integers $\mathfrak{o}_v$ the "basic vectors" Φ_v^0 are the characteristic functions of $X(\mathfrak{o}_v) = \mathfrak{o}_v^2$. There is a natural meromorphic family of morphisms $\mathcal{S}(X(\mathbb{A}_k)) \to I_{B(\mathbb{A}_k)}^{G(\mathbb{A}_k)}(\chi)$ (where I_P^G denotes normalized parabolic induction from the parabolic P and B denotes the Borel subgroup), and for idele class characters χ the composition with the Eisenstein series morphism $\mathrm{Eis}_\chi : I_{B(\mathbb{A}_k)}^{G(\mathbb{A}_k)}(\chi) \to C^\infty(G(k) \backslash G(\mathbb{A}_k))$ provides meromorphic sections of Eisenstein series, whose functional equation can be deduced from the Poisson summation formula on $\mathbb{A}_k^2$ — in particular, the L-factors that appear in the functional equation of "usual" (or "constant") sections are absent here.

This was found to be the case more generally in [Braverman and Kazhdan 1999; Braverman and Gaitsgory 2002; Braverman et al. 2002; Braverman and Kazhdan 2002]: One can construct normalized sections of Eisenstein series from certain Schwartz spaces of functions on $[P, P] \backslash G(\mathbb{A}_k)$ (or $U_P \backslash G(\mathbb{A}_k)$, where U_P is the unipotent radical of P). These Schwartz spaces should be defined as tensor products over all places, restricted with respect to some basic vector; and the basic vector should be the function-theoretic analog of the intersection cohomology sheaf of some geometric model for the space $X(\mathfrak{o}_v)$. For instance, if X is *smooth* then the intersection cohomology sheaf is constant, which means that Φ_v^0 is the characteristic function of $X(\mathfrak{o}_v)$; this explains the distributions in Tate's thesis, the work of Godement and Jacquet, and the case of period integrals. (In the latter, the characteristic function of $X(\mathfrak{o}_v) = H \backslash G(\mathfrak{o}_v)$ is obtained as the "smoothening" of the delta function at the point $H1 \in X$.)

Such geometric models were recently defined by Gaitsgory and Nadler [2010] for *every* affine spherical variety. They provide us with the data necessary to speculate on a generalization of the Rankin–Selberg method. It should be noted, however, that even to define the "correct" functions on $X^+(\mathbb{A}_k)$ out of these geometric models one has to rely on certain natural conjectures on them — therefore the problem of finding an independent or unconditional definition should be considered as one of the steps needed for establishing our conjecture.

1.4. *Comments.* Most of the ingredients in the present work are not new. Experts in the Rankin–Selberg method will recognize in our method, to a lesser or greater extent, the heuristics they have been using to find new integrals. The idea that geometric models and intersection cohomology should give rise to the "correct" space of functions on the p-adic points of a variety comes straight out of the geometric Langlands program and the work of Braverman and Kazhdan; I have nothing to offer in this direction.

However, the mixture of these ingredients is new and I think that there is enough evidence that it is the correct one. For the first time, a precise criterion is formulated on how to construct a "Rankin–Selberg" integral, reducing the problem to a purely geometric one — classifying certain embeddings of spherical varieties. And evidence shows that there should be a vast generalization that does not depend on such embeddings. I prove no "hard" theorems and, in particular, I do not know how to establish the meromorphic continuation of the X-Eisenstein series. Hence, I do not know whether I am putting the cart before the horse — however, as opposed to other conjectures that have appeared in the literature in the past, the distributions defined here are completely geometric and have nothing to do a priori with L-functions, which leaves a lot of room for hope. Finally, this point of view proves useful in explaining the phenomenon of "weight factors" in the relative trace formula.

2. Elements of the theory of spherical varieties

2.1. *Invariants associated to spherical varieties.* A *spherical variety* for a connected reductive group G over a field k is a normal variety X together with a G-action, such that over the algebraic closure the Borel subgroup of G has a dense orbit.

We denote throughout by k a number field and, unless otherwise stated, we make the following assumptions on G and X:

- G is a split, connected, reductive group.
- X is affine.

The open G-orbit in X will be denoted by X^+, and the open B-orbit by $\mathring{X}^+$, where B is a fixed Borel subgroup of G, whose unipotent radical we denote by U.[2]

The assumption that G is split is certainly very restrictive, but it is enough to demonstrate our point of view, and convenient because of many geometric and representation-theoretic results that have been established in this case. We will discuss *affine* spherical varieties in more detail later, but we just mention here that a common source of examples is when $X^+ = H\backslash G$, a quasiaffine homogeneous variety, and $X = \overline{H\backslash G}^{\text{aff}} = \operatorname{spec} k[H\backslash G]$, the *affine closure* of $H\backslash G$; see Section 2.2.

[2] Notice that this is different from that of [Gaitsgory and Nadler 2010], but compatible with the notation used in [Sakellaridis 2008; 2009; Sakellaridis and Venkatesh 2012].

We will be using standard and self-explanatory notation for varieties and algebraic groups; for example, $\mathcal{N}(H)$, $\mathcal{Z}(H)$, H^0 will be, respectively, the normalizer, center and connected component of a (sub)group H, and $\overline{Y}$ will be the closure of a subvariety Y, etc. The isotropy group of a point x under a G-action will be denoted by G_x and the fiber over $y \in Y$ of a morphism $X \to Y$ by X_y. The base change of an S-scheme Y with respect to a morphism $T \to S$ will be denoted by Y_T, but if v denotes a completion of a number field k and Y is defined over k then we will be denoting by Y_v the *set* $Y(k_v)$.

Let us discuss certain invariants associated to a spherical variety. First of all, for any algebraic group Γ we denote by $\mathcal{X}(\Gamma)$ its character group, and for any variety Y with an action of Γ we denote by $\mathcal{X}_\Gamma(Y)$ the group of Γ-eigencharacters appearing in the action of Γ on $k(Y)$. If Γ is our fixed Borel subgroup B, then we will denote $\mathcal{X}_B(Y)$ simply by $\mathcal{X}(Y)$. The multiplicative group of nonzero eigenfunctions (semiinvariants) for B on $k(Y)$ will be denoted by $k(Y)^{(B)}$. If Y has a dense B-orbit, then we have a short exact sequence $0 \to k^\times \to k(Y)^{(B)} \to \mathcal{X}(Y) \to 0$.

For a finitely generated $\mathbb{Z}$-module M, we denote the dual module $\mathrm{Hom}_{\mathbb{Z}}(M, \mathbb{Z})$ by M^*. For our spherical variety X, we let $\Lambda_X = \mathcal{X}(X)^*$ and $\mathcal{Q} = \Lambda_X \otimes_{\mathbb{Z}} \mathbb{Q}$. A B-invariant valuation on $k(X)$ that is trivial on $k^\times$ induces by restriction to $k(X)^{(B)}$ an element of Λ_X. We let $\mathcal{V} \subset \mathcal{Q}$ be the cone[3] generated by *G-invariant valuations* which are trivial on $k^\times$; see [Knop 1991, Corollary 1.8]. It is known that $\mathcal{V}$ is a polyhedral cone, and in fact that it is a fundamental domain for the action of a finite reflection group W_X on $\mathcal{Q}$. We denote by Λ_X^+ the intersection $\Lambda_X \cap \mathcal{V}$. Under the quotient map $\mathcal{X}(A)^* \otimes \mathbb{Q} \to \mathcal{Q}$, $\mathcal{V}$ contains the image of the *negative* Weyl chamber of G [Knop 1991, Corollary 5.3].

The *associated parabolic* of X is the standard parabolic

$$P(X) := \{p \in G \mid \mathring{X}^+ \cdot p = \mathring{X}^+\}.$$

Make once and for all a choice of a point $x_0 \in \mathring{X}^+(k)$ and let H denote its stabilizer; hence $X^+ = H \backslash G$, and HB is open in G. There is the following "good" way of choosing a Levi subgroup $L(X)$ of $P(X)$: Pick $f \in k[X]$, considered by restriction as an element of $k[G]^H$, such that the set-theoretic zero locus of f is $X \smallsetminus \mathring{X}^+$. Its differential df at $1 \in G$ defines an element in the coadjoint representation of G, and the centralizer $L(X)$ of df is a Levi subgroup of $P(X)$. We fix throughout a maximal torus A in $B \cap L(X)$. We define A_X to be the torus $L(X)/(L(X) \cap H) = A/(A \cap H)$; its cocharacter group is Λ_X. We consider A_X as a subvariety of $\mathring{X}^+$ via the orbit map on x_0.

[3]A *cone* in a $\mathbb{Q}$-vector space is a subset that is closed under addition and under multiplication by $\mathbb{Q}_{\geq 0}$, its *relative interior* is its interior in the vector subspace that it spans, and a *face* of it is the zero set, in the cone, of a linear functional that is nonnegative on the cone — hence, the whole cone is a face as well.

The finite reflection group $W_X \subset \mathrm{End}(\mathcal{Q})$ for which $\mathcal{V}$ is a fundamental domain is called the *little Weyl group* of X. The set of simple roots of G corresponding to B and the maximal torus $A \subset B$ will be denoted by Δ. Consider the (strictly convex) cone negative-dual to $\mathcal{V}$, that is, the set $\{\chi \in \mathcal{X}(X) \otimes \mathbb{Q} \mid \langle \chi, v \rangle \leq 0 \text{ for every } v \in \mathcal{V}\}$. The generators of the intersections of its extremal rays with $\mathcal{X}(X)$ are called the (simple) *spherical roots*[4] of X and their set is denoted by Δ_X. They are known to form the set of simple roots of a based root system with Weyl group W_X. We will denote by $\Delta(X)$ the subset of Δ consisting of simple roots in $L(X)$, and by $W_{L(X)} \subset W$ the Weyl groups of $L(X)$, respectively G. There is a canonical way [Knop 1994b, Theorem 6.5] to identify W_X with a subgroup of W, which normalizes and intersects trivially the Weyl group $W_{L(X)}$ of $L(X)$. The data $\mathcal{X}(X)$, W_X, $\mathcal{V}$ are usually easy to compute by finding a point on the open B-orbit and using Knop's action of the Borel subgroup on the set of B-orbits [Knop 1995b]; for a more systematic treatment, see [Losev 2008].

If $\mathcal{V}$ is equal to the image of the negative Weyl chamber, we say that the variety is a *wavefront* spherical variety. (This term is justified by the proof for asymptotics of generalized matrix coefficients in [Sakellaridis and Venkatesh 2012].) Symmetric varieties, for example, are all wavefront [Knop 1991, Section 5]. Motivated by the results of [Sakellaridis 2008], we will call *geometric multiplicity* of X the cardinality of the generic nonempty fiber of the map $\mathcal{X}(X)/W_X \to \mathcal{X}(A)/W$. While none implies the other, it is usually the case that varieties with geometric multiplicity one are wavefront. On the other hand, let us call *arithmetic multiplicity* of X the torsion subgroup of $\mathcal{X}(A)/\mathcal{X}(X)$. It was shown in [Sakellaridis 2008] that if F is a local nonarchimedean field, then for an irreducible unramified representation π of $G(F)$ that is in general position among X-distinguished ones (that is, with $\mathrm{Hom}_G(\pi, C^\infty(X(F))) \neq 0$), we have $\dim \mathrm{Hom}_G(\pi, C^\infty(X(F))) = 1$ if and only if both the geometric and arithmetic multiplicities of X are 1.

The G-automorphism group of a homogeneous G-variety $X^+ = H \backslash G$ is equal to the quotient $\mathcal{N}(H)/H$. It is known [Losev 2008, Lemma 7.17] that for X^+ spherical the G-automorphisms of X^+ extend to any affine completion X of X^+. Moreover, it is known that $\mathrm{Aut}^G(X)$ is diagonalizable; the cocharacter group of its connected component can be canonically identified (by considering the scalars by which an automorphism acts on rational B-eigenfunctions) with $\Lambda_X \cap \mathcal{V} \cap (-\mathcal{V})$. We will be denoting $\mathcal{Z}(X) := (\mathrm{Aut}^G(X))^0$. It will be convenient many times to replace the group G by a central extension thereof and then divide by the subgroup

[4]The work of Gaitsgory and Nadler [2010] and Sakellaridis and Venkatesh [2012] suggests that for representation-theoretic reasons one should slightly modify this definition of spherical roots. However, the lines on which the modified roots lie are still the same, and for the purposes of the present article this is enough.

of $\mathfrak{X}(G)^0$ that acts trivially on X, so that the map $\mathfrak{X}(G)^0 \to \mathfrak{X}(X)$ becomes an isomorphism.

2.2. ***Spherical embeddings and affine spherical varieties.*** We will use the words "embedding", "completion" or "compactification" of a spherical G-variety X for a spherical G-variety $\overline{X}$ (not necessarily complete) with an open equivariant embedding $X \to \overline{X}$. A spherical embedding is called *simple* if it contains a unique closed G-orbit. Spherical embeddings have been classified by Luna and Vust [1983]; our basic reference for this theory will be [Knop 1991]. We will now recall the main theorem classifying simple spherical embeddings.

For now we assume that k is an algebraically closed field in characteristic zero. However, for Theorem 2.2.1 below the assumption on the characteristic is unnecessary, and any result that does not involve "colors" holds verbatim without the assumption of algebraic closedness when the group G is split. Let X be a spherical variety and let X^+ be its open G-orbit. The *colors* of X are the closures of the B-stable prime divisors of X^+; their set will be denoted by $\mathfrak{D}$. For every B-stable divisor D in any completion X of X^+, we denote by $\rho(D)$ the element of $\mathcal{Q}$ induced by the valuation defined by D. A *strictly convex colored cone* is a pair $(\mathscr{C}, \mathscr{F})$ with $\mathscr{C} \subset \mathcal{Q}$, $\mathscr{F} \subset \mathfrak{D}$ such that

(1) $\mathscr{C}$ is a strictly (that is, not containing lines) convex cone generated by $\rho(\mathscr{F})$ and finitely many elements of $\mathscr{V}$,

(2) the intersection of $\mathscr{V}$ with the relative interior of $\mathscr{C}$ is nonempty, and

(3) $0 \notin \rho(\mathscr{F})$.

If X is a simple embedding of X^+ with closed orbit Y, we let $\mathscr{F}(X)$ denote the set of $D \in \mathfrak{D}$ such that $\overline{D} \supset Y$, and we let $\mathscr{C}(X)$ denote the cone in $\mathcal{Q}$ generated by all $\rho(D)$, where D is a B-invariant divisor (possibly also G-invariant) in X containing Y.

Theorem 2.2.1 [Knop 1991, Theorem 3.1]. *The association $X \to (\mathscr{C}(X), \mathscr{F}(X))$ is a bijection between isomorphism classes of simple embeddings of X^+ and strictly convex colored cones.*

Now let us focus on affine and quasiaffine spherical varieties. We recall from [Knop 1991, Theorem 6.7]:

Theorem 2.2.2. *A spherical variety X is affine if and only if X is simple and there exists a $\chi \in \mathscr{X}(X)$ with $\chi|_{\mathscr{V}} \geq 0$, $\chi|_{\mathscr{C}(X)} = 0$ and $\chi|_{\rho(\mathfrak{D} \smallsetminus \mathscr{F}(X))} < 0$. In particular, $H \backslash G$ is affine if and only if $\mathscr{V}$ and $\rho(\mathfrak{D})$ are separated by a hyperplane, while it is quasiaffine if and only if $\rho(\mathfrak{D})$ does not contain zero and spans a strictly convex cone.*

Recall from [Braverman and Gaitsgory 2002, Section 1.1] that a variety Y over a field k is called *strongly quasiaffine* if the algebra $k[Y]$ of global functions on Y is finitely generated and the natural map $Y \to \operatorname{spec} k[Y]$ is an open embedding. Then the variety $\overline{Y}^{\mathrm{aff}} := \operatorname{spec} k[Y]$ is called the *affine closure* of Y.

Proposition 2.2.3. *A homogeneous quasiaffine spherical variety $Y = H \backslash G$ is strongly quasiaffine. If $X := \overline{H \backslash G}^{\mathrm{aff}}$, then the data $(\mathscr{C}(X), \mathscr{F}(X))$ can be described as follows: Consider the cone $\mathscr{R} \subset \mathscr{X}(X) \otimes \mathbb{Q}$ generated by the set of $\chi \in \mathscr{X}(X)$ such that $\chi|_{\mathscr{V}} \geq 0$ and $\chi|_{\rho(\mathscr{D})} \leq 0$. Choose a point χ in the relative interior of $\mathscr{R}$. Then $\mathscr{F}(X) = \{D \in \mathscr{D} \mid \rho(D)(\chi) = 0\}$ and $\mathscr{C}(X)$ is the cone generated by $\mathscr{F}(X)$.*

Remark 2.2.4. The first statement of the proposition generalizes a result of Hochschild and Mostow [1973] for the variety $U_P \backslash G$, where U_P is the unipotent radical of a parabolic subgroup P of G. Indeed, this variety is spherical under the action of $M \times G$, where M is the reductive quotient of P.

Proof. As a representation of G, $k[Y]$ is locally finite and decomposes as

$$k[Y] = \bigoplus_{\lambda} V_\lambda, \tag{2-1}$$

where V_λ is the isotypic component corresponding to the representation with highest weight λ, and the sum is taken over all λ with $V_\lambda \neq 0$. Since the variety is spherical, each V_λ is isomorphic to one copy of the representation with highest weight λ. Moreover, the multiplicative monoid of nonzero highest-weight vectors $k[Y]^{(B)}$ is the submonoid of $k(Y)^{(B)}$ (the group of nonzero rational B-eigenfunctions) consisting of regular functions. Regular B-eigenfunctions are precisely those whose eigencharacter satisfies $\chi|_{\rho(\mathscr{D})} \geq 0$; since the set $\mathscr{D}$ is finite, the monoid of λ that appears in the decomposition (2-1) is finitely generated. Since the multiplication map $V_\mu \otimes V_\nu$ has image in the sum of V_λ with $\lambda \leq \mu + \nu$, and since its composition with the projection $k[Y] \to V_{\mu+\nu}$ is surjective, the sum of the V_λ, for λ in a set of generators for the monoid of λ's appearing in (2-1), generates $k[Y]$.

The second condition, namely that $Y \to X$ is an open embedding, follows from the assumption that Y is quasiaffine and from the homogeneity of Y. Hence, Y is strongly quasiaffine.

The affine closure X has the property that for every affine completion X' of Y, there is a morphism $X \to X'$. The description of $(\mathscr{C}(X), \mathscr{F}(X))$ now follows from Theorem 2.2.2 and [Knop 1991, Theorem 4.1], which describes morphisms between spherical embeddings. The cone $\mathscr{C}(X)$, as described, will necessarily contain the intersection of $\mathscr{V}$ with the cone generated by $\rho(\mathscr{D})$ in its relative interior; therefore its relative interior will have nonempty intersection with $\mathscr{V}$. □

Let us now discuss the geometry of affine spherical varieties. The following is a corollary of Luna's slice theorem:

Theorem 2.2.5 [LuS 1973, III.1, Corollaire 2]. *If G is a reductive group over an algebraically closed field k in characteristic zero, acting on an affine variety X so that $k[X]^G = k$, then X contains a closed G-homogeneous affine subvariety Y such that the embedding $Y \hookrightarrow X$ admits an equivariant splitting $X \twoheadrightarrow Y$. If G is smooth, then the fiber over any (closed) point $y \in Y$ is G_y-equivariantly isomorphic to the vector space of a linear representation of G_y.*

Luna's theorem also states that Y is contained in the closure of any G-orbit, which is easily seen to be true in the spherical case since affine spherical varieties are simple. The G-automorphism group "retracts" X onto Y:

Proposition 2.2.6. *Let X be an affine spherical G-variety and let Y be as in the theorem above, considered both as a quotient and as a subvariety of X. Let T be the maximal torus in $\mathrm{Aut}^G(X)$ that acts trivially on Y. Then the closure of the T-orbit of every point on X meets Y. Equivalently, $k[X]^T = k[Y]$.*

Proof. This is essentially [Knop 1994a, Corollary 7.9]. More precisely, let us assume that G has a fixed point on X, that is, Y is a point. (The question is easily reduced to this case, since every G_y-automorphism of the fiber of $X \to Y$ over y extends uniquely to a G-automorphism of X.) The proof of [loc. cit.] shows that for a generic point $x \in X$ there is a one-parameter subgroup H of $\mathrm{Aut}^G(X)$ such that $x \cdot H$ contains the fixed point in its closure. Hence $k[X]^T = k$ and therefore X contains a unique closed T-orbit. □

If G has a fixed point on X, we can embed X into a finite sum $V = \bigoplus_i V_i$ of finite-dimensional representations of G, such that the fixed point is the origin in V and there is a subtorus T of $\prod_i \mathrm{Aut}^G(V_i)$ acting on X with the origin as its only closed orbit. (Simply take V to be the dual of a G-stable, generating subspace of $k[X]$.)

2.3. ***Generalized Cartan decomposition.*** Let $\mathcal{K} = \mathbb{C}((t))$, the field of formal Laurent series over $\mathbb{C}$, and $\mathfrak{O} = \mathbb{C}[[t]]$, the ring of formal power series. Let X^+ be a homogeneous spherical variety over $\mathbb{C}$.

Theorem 2.3.1 [Luna and Vust 1983]. *$G(\mathfrak{O})$-orbits on $X^+(\mathcal{K})$ are parametrized by Λ_X^+, where to $\check{\lambda} \in \Lambda_X^+$ corresponds the orbit through $\check{\lambda}(t) \in A_X(\mathcal{K})$.*

A new proof given by Gaitsgory and Nadler [2010] can be used to prove the analogous statement over p-adic fields. We revisit their argument, adapt it to the p-adic case, and extend it to determine the set of $G(\mathfrak{o}_F)$-orbits on $X(\mathfrak{o}_F)$, when G and X are affine and defined over a number field and F is a nonarchimedean completion (outside of a finite set of places).

Remark 2.3.2. In the case of symmetric spaces, similar statements on the set of $G(\mathfrak{o}_F)$-orbits on $X(F)$ and in a more general setting — without assuming that G is split — have been proven by Benoist and Oh [2007] and Delorme and Sécherre [2011].

The argument uses compactification results of Brion, Luna and Vust. We first need to recall a few more elements of the theory of spherical varieties. The results below have appeared in the literature for k an algebraically closed field in characteristic zero, but the proofs hold verbatim when k is any field in characteristic zero and the groups in question are split over k. (The basic observation being, here, that in all proofs one gets to choose B-eigenfunctions in $k(X)$, and since the variety is spherical and the group is split, the eigenspaces of B are one-dimensional and defined over k, and therefore the chosen eigenfunctions are k-rational up to $\bar{k}$-multiple.)

A *toroidal embedding* of X^+ is an embedding X^c of X^+ in which no color (B-stable divisor which is not G-stable) contains a G-orbit. Theorem 2.2.1 implies that *simple* toroidal embeddings are classified by strictly convex, finitely generated subcones of $\mathcal{V}$. Moreover, the simple toroidal embedding X^c obtained from a simple embedding X by taking the cone $\mathcal{C}(X^c) = \mathcal{C}(X) \cap \mathcal{V}$ comes with a proper equivariant morphism $X^c \to X$ [Knop 1991, Theorem 4.1] that is surjective [ibid., Lemma 3.2].

The local structure of a simple toroidal embedding is given by the following theorem of Brion, Luna and Vust:

Theorem 2.3.3 [Brion et al. 1986, Théorème 3.5]. *Let X^c be a simple toroidal embedding of X^+ and let X^c_B denote the complement of all colors. Then X^c_B is an open, $P(X)$-stable, affine variety with the following properties*:

(1) *X^c_B meets every G-orbit.*

(2) *If we let Y^c be the closure of A_X in X^c_B, then the action map $Y^c \times U_{P(X)} \to X^c_B$ is an isomorphism.*

We emphasize the structure of the affine toric variety Y^c: Its cone of regular characters is precisely $\mathcal{C}(X^c)^\vee := \{\chi \in \mathcal{X}(X) \otimes \mathbb{Q} \mid \langle \chi, v \rangle \geq 0 \text{ for all } v \in \mathcal{C}(X^c)\}$; in other words,

$$Y^c = \operatorname{spec} k[\mathcal{C}(X^c)^\vee \cap \mathcal{X}(X)].$$

By the theory of toric varieties, the theorem also implies that X^c is smooth if and only if the monoid $\mathcal{C}(X^c) \cap \Lambda_X$ is generated by primitive elements in its "extremal rays" (that is, is a free abelian monoid).

When $\mathcal{V}$ is strictly convex (equivalently $\mathrm{Aut}^G(X^+)$ is finite), then X^+ admits a canonical toroidal embedding $\overline{X}$, with $\mathcal{C}(\overline{X}) = \mathcal{V}$, which is complete. This is sometimes called the *wonderful completion* of X^+, although often the term "wonderful" is reserved for the case that this completion is smooth. If $\mathcal{V}$ is not strictly convex, then X^+ still admits a (nonunique) complete toroidal embedding $\overline{X}$, which is not simple, but as remarked in [Gaitsgory and Nadler 2010, 8.2.7], Theorem 2.3.3 still holds, with Y^c a suitable (nonaffine) toric variety containing A_X. The fan of Y^c depends on the chosen embedding $\overline{X}$, but its support is precisely the dual cone of $\mathcal{V}$ (that is, the set of cocharacters λ of A_X such that $\lim_{t\to 0}\lambda(t)\in Y^c$ is equal to Λ_X^+).

We will use Theorem 2.3.3 for two toroidal varieties: First, for a complete toroidal embedding $\overline{X}$ of X^+. Secondly, for the variety $\hat{X}$ obtained from our affine spherical variety X by taking $\mathcal{C}(\hat{X})=\mathcal{C}(X)\cap\mathcal{V}$. Before we proceed, we discuss models of these varieties over rings of integers.

2.3.4. *Models over rings of integers.* We start with toric varieties. Let $\mathfrak{o}$ be an integral domain with fraction field k, and let Y be a simple (equivalently, affine) toric variety for a split torus T over k. We endow T with its smooth model $\mathcal{T}=\mathfrak{o}[\mathcal{X}(T)]$ over $\mathfrak{o}$. Since $Y=\operatorname{spec} k[M]$ for some saturated monoid $M\subset\mathcal{X}(T)$, the $\mathfrak{o}$-scheme $\mathcal{Y}=\operatorname{spec}\mathfrak{o}[M]$ is a model for Y over $\mathfrak{o}$ with an action of $\mathcal{T}$, and we will call it the *standard model*. The notion easily extends to the case where Y is not necessarily affine, but defined by a fan. If T and Y are defined over a number field k and endowed with compatible models over the S-integers $\mathfrak{o}_S$ for a finite set S of places of k, then these models will coincide with the standard models over $\mathfrak{o}_{S'}$, for some finite $S'\supset S$.

Now we return to the setting where k is a number field, G, X, X^+, $\overline{X}$, $\hat{X}$ are as before (over k), and let us also fix a point $x_0\in\mathring{X}^+(k)$. Then we can choose compatible integral models outside of a finite set of places, such that the structure theory of Brion, Luna and Vust continues to hold for these models:

Proposition 2.3.5. *There are a finite set of places S_0 of k and compatible flat models $\mathcal{G}$, $\mathfrak{X}$, $\overline{\mathfrak{X}}$ and $\hat{\mathfrak{X}}$ for G, X, $\overline{X}$ and $\hat{X}$, respectively, over the S_0-integers $\mathfrak{o}_{S_0}$ of k, such that*

- *S_0 contains all archimedean places*;
- *the chosen point x_0 is in $in\mathring{\mathcal{X}}^+(\mathfrak{o}_{S_0})$*;
- *$\mathcal{G}$ is reductive over $\mathfrak{o}_{S_0}$, and $\mathcal{X}^+\to\operatorname{spec}\mathfrak{o}_{S_0}$ is smooth and surjective*;
- *the statement of Theorem 2.3.3 holds for $\overline{\mathfrak{X}}$ and $\hat{\mathfrak{X}}$ over $\mathfrak{o}_{S_0}$, namely, if we denote any either of them by $\mathfrak{X}^c$, then there is an open, $\mathcal{P}(X)$-stable subscheme $\mathfrak{X}^c_B$ and a toric $\mathcal{A}$-subscheme $\mathcal{Y}^c$* of standard type *such that the subscheme $\mathfrak{X}^c_B$ meets every $\mathcal{G}$-orbit on $\mathfrak{X}^c$ and the action map $\mathcal{Y}^c\times\mathcal{U}_{P(X)}\to\mathfrak{X}^c_B$ is an isomorphism of $\mathfrak{o}_{S_0}$-schemes.*
- *$\overline{\mathfrak{X}}$ is proper over $\mathfrak{o}_{S_0}$, and the morphism $\hat{\mathfrak{X}}\to\mathfrak{X}$ is proper.*

Remarks 2.3.6. (1) By $\mathfrak{X}^+$ and $\mathring{\mathfrak{X}}^+$ we denote the complement of the closure, in any of the above schemes, of the complement of X^+ and $\mathring{X}^+$, respectively, in the generic fiber.

(2) It is implicitly part of the "compatibility" of the models that the scheme structures on $\mathfrak{X}^+$ and $\mathring{\mathfrak{X}}^+$ do not depend on which of the ambient schemes we choose to define them.

(3) We understand the statement "meets every orbit" as follows: Let $|\mathscr{Z}|$ denote the set of scheme-theoretic points of a scheme $\mathscr{Z}$. Consider the two maps $p : \mathscr{G} \times \mathfrak{X} \to \mathfrak{X}$ (projection to the second factor) and $a : \mathscr{G} \times \mathfrak{X} \to \mathfrak{X}$ (action map). Then for every $x \in |\mathfrak{X}^c|$ the set $a(p^{-1}\{x\})$ intersects $|\mathfrak{X}^c_B|$ nontrivially.

Proof. For a finite set S of places and a flat model $\mathfrak{X}^c$ of X^c over $\mathfrak{o}_S$ (assumed proper if $X^c = \overline{X}$), let D denote the union of all colors over the generic point of $\operatorname{spec} \mathfrak{o}_S$, let $\mathfrak{D}$ denote the closure of D in $\mathfrak{X}^c$, and let $\mathfrak{X}^c_B$ be the complement of $\mathfrak{D}$ in $\mathfrak{X}^c$. Let $\mathscr{G}$ denote a compatible reductive model for G over $\mathfrak{o}_S$. (All these choices are possible by sufficiently enlarging S.) The image of $\mathscr{G} \times \mathfrak{X}^c_B \to \mathfrak{X}^c$ is open and contains the generic fiber; hence by enlarging the set S, if necessary, we can make it surjective.

Now define $\mathscr{Y}^c$ as the closure of Y^c in $\mathfrak{X}^c_B$. By enlarging the set S, if necessary, we may assume that $\mathscr{Y}^c$ is of standard type. The action map $\mathscr{Y}^c \times \mathscr{U}_{P(X)} \to \mathfrak{X}^c_B$ being an isomorphism over the generic fiber, it is an isomorphism over $\mathfrak{o}_S$ by enlarging S, if necessary. □

From now on we fix such a finite set of places S_0 and such models. The combinatorial invariants of the schemes above are the same at all places of S_0:

Proposition 2.3.7. *Each of the data*[5] $\mathscr{X}(X), \mathscr{V}, \mathscr{C}(X), \mathscr{C}(\overline{X}), \mathscr{C}(\hat{X})$ *is the same for the reductions of* $\mathfrak{X}, \overline{\mathfrak{X}}, \hat{\mathfrak{X}}$ *at all closed points of* $\mathfrak{o}_{S_0}$. *The set of G-orbits on each of these varieties is in natural bijection with the set of $\mathscr{G}$-orbits on each of their reductions.*

Proof. The toric scheme $\mathscr{Y}^c$ being of the standard type, it means that $\mathscr{X}(X) = \mathscr{X}_A(Y^c)$ is the same at all reductions. For every place v of $\mathfrak{o}_S$ the reductions $\overline{\mathfrak{X}}_{\mathbb{F}_v}$ and $\hat{\mathfrak{X}}_{\mathbb{F}_v}$ are toroidal: Indeed, denoting by $\mathfrak{X}^c$ either of them, the complement of $(\mathfrak{X}^c_B)_{\mathbb{F}_v}$ is a $\mathscr{B}_{\mathbb{F}_v}$-stable union of divisors that does not contain any $\mathscr{G}_{\mathbb{F}_v}$-orbit, since $(\mathfrak{X}^c_B)_{\mathbb{F}_v}$ meets every $\mathscr{G}_{\mathbb{F}_v}$-orbit. Moreover, $\mathfrak{X}^c_B$ meets no colors, for if it did, then a nonopen $\mathscr{A}_{\mathbb{F}_v}$-orbit on $\mathscr{Y}^c_{\mathbb{F}_v}$ would belong to the open $\mathscr{G}_{\mathbb{F}_v}$-orbit, and hence the open $\mathscr{G}_{\mathbb{F}_v}$-orbit would belong to the closure of a nonopen G-orbit over the generic point, a contradiction since by assumption $\mathfrak{X}^+$ is smooth and surjective. Therefore,

[5] Since $\overline{X}$ is not necessarily simple, it is not described by a cone but by a fan. However, we slightly abuse the common notation here and write $\mathscr{C}(\overline{X})$ for the set of invariant valuations whose center is in $\overline{X}$ — that is, for the support of the fan associated to $\overline{X}$.

the complement of $(\mathfrak{X}^c_B)_{\mathbb{F}_v}$ is the union of all colors of $\mathfrak{X}^c_{\mathbb{F}_v}$, and $\mathfrak{X}^c_{\mathbb{F}_v}$ is toroidal. Moreover, the $\mathscr{G}_{\mathbb{F}_v}$-invariant valuations on $\mathbb{F}_v(\mathfrak{X}^+{}_{\mathbb{F}_v})$ whose centers are in $\mathfrak{X}^c_{\mathbb{F}_v}$ are precisely those of $\Lambda_X \cap \mathscr{C}(X^c)$ (which proves the equality of $\mathscr{C}(\mathfrak{X}^c_{\mathbb{F}_v})$ with $\mathscr{C}(X^c)$ at all $v \notin S_0$), and from the fact that $\overline{\mathfrak{X}}_{\mathbb{F}_v}$ is complete and $\mathscr{C}(\overline{\mathfrak{X}}_{\mathbb{F}_v}) = \Lambda^+_X$, it follows that $\mathscr{V}$ is precisely the cone of invariant valuations on $\mathbb{F}_v(\mathfrak{X}^+)$. □

Now we are ready to apply the argument of [Gaitsgory and Nadler 2010, Theorem 8.2.9] to describe representatives for the set of $\mathscr{G}(\mathfrak{o}_F)$-orbits on $\mathfrak{X}^+(\mathfrak{o}_F)$, for every completion F of k outside of S_0, and also extend it to a description of the set of orbits that are contained in $\mathfrak{X}(\mathfrak{o}_F)$. Notice that since $\mathscr{G}$ is reductive, $\mathscr{G}(\mathfrak{o}_F)$ is a hyperspecial maximal compact subgroup of $G(F)$. From now on we denote our fixed models over $\mathfrak{o}_{S_0}$ by regular script, since there will be no possibility of confusion. There is a canonical $A_X(\mathfrak{o}_F)$-invariant homomorphism $A_X(F) \to \Lambda_X$ (under which an element of the form $\lambda(\varpi)$, where ϖ is a uniformizer for F, maps to λ) and we denote by $A_X(F)^+$ the preimage of Λ^+_X.

Theorem 2.3.8. *For F a completion of k outside of S_0, each $G(\mathfrak{o}_F)$-orbit on $X^+(F)$ contains an element of $A_X(F)^+$, and elements of $A_X(F)^+$ with different image in Λ^+_X belong to distinct $G(\mathfrak{o}_F)$-orbits. If the quotient $\mathscr{X}(A)/\mathscr{X}(X)$ is torsion-free, then the map from $G(\mathfrak{o}_F)$-orbits on $X^+(F)$ to Λ^+_X is a bijection. The orbits contained in $X(\mathfrak{o}_F)$ are precisely those mapping to $\Lambda^+_X \cap \mathscr{C}(X)$.*

Remark 2.3.9. The torsion of the quotient $\mathscr{X}(A)/\mathscr{X}(X)$ is the "arithmetic multiplicity" defined in Section 2.1. It is trivial if and only if the map $A_X(F)/A(\mathfrak{o}) \to \Lambda_X$ is bijective; hence the statement about bijectivity in that case is straightforward. In general, elements in different $A(\mathfrak{o}_F)$-orbits may belong to the same $G(\mathfrak{o}_F)$-orbit; for instance, if $X^+ = H\backslash G$ with H connected then the map $G(\mathfrak{o}_F) \ni g \mapsto x_0 \cdot g \in X^+(\mathfrak{o}_F)$ will be surjective by an application of Lang's theorem (the vanishing of Galois cohomology of H over a finite field). But it is also not always the case that elements corresponding to the same λ will always be in the same $G(\mathfrak{o}_F)$-orbit — for instance, when H is not connected.

We will prove this theorem together with a theorem about orbits of the first congruence subgroup, which will not be used here but will be useful elsewhere. Let $\mathbb{F}$ denote the residue field of F.

Theorem 2.3.10. *Let $K_1, A_{X,1}, U_1$ be the preimages of $1 \in G(\mathbb{F})$, $1 \in A_X(\mathbb{F})$, $1 \in U(\mathbb{F})$ in $G(\mathfrak{o}_F)$, $A_X(\mathfrak{o}_F)$, $U(\mathfrak{o}_F)$, respectively. Then for every $x \in A_X(F)^+$, we have $x \cdot K_1 \subset x \cdot A_{X,1} \cdot U_1$.*

Proof of Theorems 2.3.8 and 2.3.10. Denote $\mathfrak{o}_F$ by $\mathfrak{o}$. We use the notation X^c, X^c_B, Y^c, etc. as above for the scheme $\overline{X}$. The $\mathfrak{o}$-scheme X^c is proper and hence $X^c(\mathfrak{o}) = X^c(F)$. We will first show that $Y^c(\mathfrak{o})$ contains representatives for all $G(\mathfrak{o})$-orbits on $X^c(\mathfrak{o})$. Let $x \in X^c(\mathfrak{o})$ and denote by $\bar{x} \in X^c(\mathbb{F})$ its reduction. The open, $P(X)$-stable subvariety X^c_B meets every G-orbit; for a spherical variety for a split reductive group over an arbitrary field (denoted $\mathbb{F}$, since we will apply it to this field) the $\mathbb{F}$-points of the open B-orbit meet every $G(\mathbb{F})$-orbit. (This is proven following the argument of [Sakellaridis 2008, Lemma 3.7.3], that is, reducing to the case of rank one groups, and by inspection of the spherical varieties for SL_2, classified in [Knop 1995a, Theorem 5.1].) This means that there is a $\bar{g} \in G(\mathbb{F})$ (which we can lift to a $g \in G(\mathfrak{o})$) such that $\overline{x \cdot g} \in X^c_B(\mathbb{F})$. Since X^c_B is open, this means that $x \cdot g \in X^c_B(\mathfrak{o}) = Y^c(\mathfrak{o}) \times U_{P(X)}(\mathfrak{o})$. Acting by a suitable element of $U_{P(X)}(\mathfrak{o})$, we get a representative for the $G(\mathfrak{o})$-orbit of x in $Y^c(\mathfrak{o})$. Hence, $G(\mathfrak{o})$-orbits on $X^+(F)$ are represented by elements of $A_X(F)^+ = Y^c(\mathfrak{o}) \cap A_X(F)$.

To prove that elements mapping to distinct $\lambda, \lambda' \in \Lambda^+_X$ belong to different $G(\mathfrak{o})$-orbits, the argument of Gaitsgory and Nadler carries over verbatim: If λ and λ' are not $\mathbb{Q}$-multiples of each other, we can construct as in [Knop 1991] a toroidal embedding X^t of X^+ over $\mathfrak{o}$ such that $\lambda(\varpi) \in X^t(\mathfrak{o})$ but $\lambda'(\varpi) \notin X^t(\mathfrak{o})$. Finally, if λ and λ' are $\mathbb{Q}$-multiples of each other (without loss of generality, $\lambda \neq 0$), then we can find a toroidal compactification X^t such that $\lim_{t\to 0} \lambda(t)$ belongs to some G-orbit D of codimension one, and then the intersection numbers of $\lambda(\varpi)$ and $\lambda'(\varpi)$ (considered as 1-dimensional subschemes of X^t) with D are different. (Notice that the constructions of [Knop 1991] are over a field of arbitrary characteristic, and based on Proposition 2.3.7 one can carry them over the ring $\mathfrak{o}_F$.)

To finish the proof of Theorem 2.3.8, if we now consider $\hat{X}$ then we have a proper morphism $\hat{X} \to X$ that is an isomorphism on X^+. By the valuative criterion for properness, every point in $X(\mathfrak{o}) \cap X^+(F)$ lifts to a point on $\hat{X}(\mathfrak{o})$; therefore for the last statement it suffices to determine the set of $G(\mathfrak{o})$-orbits on $\hat{X}(\mathfrak{o}) \cap X^+(F)$. By the same argument as before, every $G(\mathfrak{o})$-orbit meets $\hat{Y}(\mathfrak{o})$, and the latter intersects $A_X(F)$ precisely in the union of $A_X(\mathfrak{o})$-orbits represented by $\Lambda_X \cap \mathscr{C}(X)$.

For Theorem 2.3.10, we first notice that $X^c_B(\mathfrak{o})$ (where X^c still denotes $\overline{X}$) is K_1-stable; indeed, for any $x \in X^c_B(\mathfrak{o})$ and $g \in K_1$ the reduction of $x \cdot g$ belongs to $X^c_B(\mathbb{F})$, and since X^c_B is open this implies that $x \cdot g \in X^c_B(\mathfrak{o})$. Now we claim that $Y^c(\mathfrak{o}) \cdot U_1$ is also K_1-stable; indeed, this is the preimage in $X^c_B(\mathbb{F})$ of $Y^c(\mathbb{F})$, and for every $x \in Y^c(\mathfrak{o}) \cdot U_1$ and $g \in K_1$, the reduction of $x \cdot g$ belongs to $Y^c(\mathbb{F})$. We have already argued that elements of $A_X(F)^+$ with different images in Λ^+_X belong to distinct $G(\mathfrak{o})$-orbits, hence to distinct K_1-orbits; hence, $x \cdot K_1$ belongs to the set of elements of $A_X(F)^+ \cdot U_1$ with the same image $\lambda_x \in \Lambda^+_X$ as x.

To distinguish between those elements, we assign to them some invariants that will be preserved by the K_1-action. First of all, if $\lambda_x = 0$, then the reduction of x modulo $\mathfrak{p}$ is an element of $X^+(\mathbb{F})$ that is preserved by K_1, and the elements of $A_X(F)^+ \cdot U_1$ having the same reduction are precisely the elements in the same

$A_{X,1} \cdot U_1$-orbit as x. Assume now that $\lambda_x \neq 0$ and fix as above a spherical embedding X^t of X^+ over $\mathfrak{o}$ such that $\lim_{t\to 0} \lambda(t)$ belongs to a G-orbit of codimension one, whose closure we denote by D. Let n be the intersection number of $x \in X^t(\mathfrak{o}) \cap X^+(F)$ with D; then $x : \operatorname{spec} \mathfrak{o} \to X^t$ has reductions $\bar{x} : \operatorname{spec} \mathbb{F} \to D$, $\bar{x}^n : \operatorname{spec}(\mathfrak{o}/\mathfrak{p}^n) \to D$ and $\bar{x}^{n+1} : \operatorname{spec}(\mathfrak{o}/\mathfrak{p}^{n+1}) \to X^t$, which give rise to an $\mathbb{F}$-linear map from the fiber at $\bar{x}$ of the conormal bundle of D in X^t to $\mathfrak{p}^n/\mathfrak{p}^{n+1}$. The group K_1 preserves the reduction of x and acts trivially on the fiber of the conormal bundle of D over it; therefore preserves this map. It is straightforward to see that for elements of $A_X(F)^+ \cdot U_1$ with the same image in Λ_X^+ this invariant characterizes the $A_{X,1} \cdot U_1$-orbit of x. □

3. Conjectures on Schwartz spaces and automorphic distributions

This section is highly conjectural and only aims at fixing ideas. We speculate on the existence of some "Schwartz space" of functions on the points of an affine spherical variety over a local field, and explain how to construct from it distributions on the automorphic quotient $[G] := G(k) \setminus G(\mathbb{A}_k)$ that should have good analytic properties. At almost every place this space of functions should come equipped with a distinguished, unramified element that should be related (in a rather ad hoc way, using the generalized Cartan decomposition) to intersection cohomology sheaves on spaces defined by Gaitsgory and Nadler. In subsequent sections we will specialize to the case where X has a certain geometry (which we call a "preflag bundle"), and these distinguished functions will be described explicitly, in order to understand the Rankin–Selberg method.

3.1. *Formalism of Schwartz spaces and theta series.*

3.1.1. *Schwartz space.* We fix an affine spherical variety X for a (split) reductive group G over a global field k, and for every place v of k, we denote by X_v^+ the space of k_v-points of X^+. We assume as given, for every v, a G_v-invariant "Schwartz space" of functions $\mathcal{S}(X_v) \subset C^\infty(X_v^+)$, and for almost every (finite) v a distinguished unramified element $\Phi_v^0 \in \mathcal{S}(X_v)^{G(\mathfrak{o}_v)}$ (called "basic vector" or "basic function") such that

$$\Phi_v^0|_{X^+(\mathfrak{o}_v)} = 1. \tag{3-1}$$

(Clearly, the integral model that is implicit in the definitions will not play any role.) We also assume the following regarding the support of Schwartz functions and their growth close to the complement of X^+:

- The closure in X_v of the support of any element of $\mathcal{S}(X_v)$ is compact.
- There exist a finite set $\{f_1, \dots, f_n\}$ of elements of $k[X]$, whose common zeroes lie in $X \smallsetminus X^+$, and a natural number n, such that for any place v and any

$\Phi_v \in \mathcal{S}(X_v)$, there is a constant c_v, equal to 1 for $\Phi_v = \Phi_v^0$, such that for all $x \in X^+(k_v)$ we have $|\Phi_v(x)| \le c_v \cdot (\max_i |f_i(x)|)^{-1}$.

At archimedean places the requirement of compact support is far from ideal, but for our present purposes it is enough. One should normally impose similar growth conditions on the derivatives (at archimedean places) of elements of the Schwartz space, but we will not need them here.

The corresponding *global Schwartz space* is, by definition, the restricted tensor product

$$\mathcal{S}(X(\mathbb{A}_k)) := \bigotimes_v{}' \mathcal{S}(X_v) \tag{3-2}$$

with respect to the basic vectors Φ_v^0.

Despite the notation, the elements of $\mathcal{S}(X(\mathbb{A}_k))$ cannot be interpreted as functions on $X(\mathbb{A}_k)$. They *can* be considered, though, as functions on $X^+(\mathbb{A}_k)$, because of the requirement (3-1).

We may require, without serious loss of generality, that $X^+(\mathbb{A}_k)$ carries a positive $G(\mathbb{A}_k)$-eigenmeasure dx whose eigencharacter ψ is the absolute value of an algebraic character. We normalize the regular representation of $G(\mathbb{A}_k)$ on functions on $X^+(\mathbb{A}_k)$ so that it is unitary when restricted to $L^2(X) = L^2(X, dx)$:

$$g \cdot \Phi(x) := \sqrt{\psi(g)}\Phi(x \cdot g).$$

The *X-theta series* is the following functional on $\mathcal{S}(X(\mathbb{A}_k))$:

$$\theta(\Phi) := \sum_{\gamma \in X^+(k)} \Phi(\gamma). \tag{3-3}$$

Translating by $G(\mathbb{A}_k)$, we can also consider it as a morphism

$$\mathcal{S}(X(\mathbb{A}_k)) \to C^\infty([G]), \tag{3-4}$$

which will be denoted by the same letter, that is,

$$\theta(\Phi, g) = \sum_{\gamma \in X^+(k)} (g \cdot \Phi)(\gamma). \tag{3-5}$$

This sum is absolutely convergent, by the first growth assumption. (Notice that X is affine and hence $X(k)$ is discrete in $X(\mathbb{A}_k)$.)

3.1.2. *Mellin transform.* Now recall (Proposition 2.2.6) that, unless X is affine homogeneous, it has a positive-dimensional group of G-automorphisms, that is, $\mathfrak{X}(X) \neq 0$. By enlarging G and dividing by the subgroup of $\mathfrak{X}(G)^0$ that acts trivially, we will from now on assume that $\mathfrak{X}(G)^0 \simeq \mathfrak{X}(X)$ under its action on X. An algebraic character of $\mathfrak{X}(X)$ will be called *X-positive* if it extends to the closure of a generic orbit of $\mathfrak{X}(X)$, that is, $\chi : \mathfrak{X}(X) \to \mathbb{G}_m$ is positive if for $Y = \overline{\mathfrak{X}(X) \cdot x}$, where

x is a generic point (say, a point on the open G-orbit), the function $z \cdot x \mapsto \chi(z) \in \mathbb{G}_m \subset \mathbb{G}_a$ extends to a morphism $Y \to \mathbb{G}_a$. Obviously, X-positive characters span a polyhedral cone in $\mathcal{X}(\mathfrak{Z}(X)) \otimes \mathbb{Q}$, and we will use the expression "sufficiently X-positive characters" to refer to characters in the translate of this cone by an element belonging to its relative interior. This notion will also be used for complex-valued characters: A sufficiently X-positive character is one whose absolute value can be written as the product of the absolute values sufficiently X-positive algebraic characters, raised to powers ≥ 1. Similar notions will be used for the dual cone, in the space of cocharacters into $\mathfrak{Z}(X)$; for example, a cocharacter $\check{\lambda}$ is X-positive if and only if for a generic point $x \in X$ we have $\lim_{t\to 0} x \cdot \check{\lambda}(t) \in X$. Finally, since by our assumption, $\mathcal{X}(G) \otimes \mathbb{Q} = \mathcal{X}(\mathfrak{Z}(X)) \otimes \mathbb{Q}$, we can use the notion of X-positive characters for characters of G, as well.

Proposition 3.1.3. *The function $\theta(\Phi, g)$ on $G(k) \backslash G(\mathbb{A}_k)$ is of moderate growth. Moreover, it is compactly supported in the direction of X-positive cocharacters into $\mathfrak{Z}(G)$; that is, for every $g \in G(\mathbb{A}_k)$ we have*

$$\theta(\Phi, g \cdot \check{\lambda}(a)) = 0$$

if $\check{\lambda}$ is a nontrivial X-positive cocharacter into $\mathfrak{Z}(X) = \mathfrak{Z}(G)^0$ and the norm of $a \in \mathbb{A}_k^\times$ is sufficiently large.

The statement about the support is an obvious corollary of the compact support of Φ, and the statement on moderate growth will be proven in the next subsections. Assuming it for now, we may consider the *Mellin transform* of $\theta(\Phi, g)$ with respect to the action of $\mathfrak{Z}(G)$:

$$E(\Phi, \omega, g) = \int_{\mathfrak{Z}(X)(\mathbb{A}_k)} \theta(z \cdot \Phi, g)\omega(z)\, dz, \tag{3-6}$$

originally defined for sufficiently X-positive idele class characters ω. We will call this an *X-Eisenstein series*.

Proposition 3.1.4. *For sufficiently X-positive ω, the integral (3-6) converges and the function $E(\Phi, \omega, g)$ is of moderate growth in g.*

Proof. The convergence statement follows immediately from Proposition 3.1.3; the statement on moderate growth is proven in the same way as Proposition 3.1.3, and we will not comment on it separately. □

3.1.5. *Adelic distance functions.* Let $Z \subset X$ be a closed subvariety of an affine variety, and let X^+ denote the complement of Z. We would like to define some "natural" notion of distance from Z (denoted d_Z) for the adelic points of X^+. The distance function will be an Euler product

$$d_Z(x) = \prod_v d_{Z,v}(x_v),$$

where, for $x \in X^+(\mathbb{A}_k)$, almost all factors will be equal to one.

We do it in the following way: First, we fix a finite set S of places, including the archimedean ones, and an affine flat model for X over the S-integers $\mathfrak{o}_S$. The closure of Z in this model defines an ideal $J \subset \mathfrak{o}_S[X]$. We can choose a finitely generated $\mathfrak{o}_S$-submodule M of J such that M generates J as an $\mathfrak{o}_S[X]$-module. In the case when X carries the action of a group G and Z is G-stable, we also choose a compatible flat model for G over $\mathfrak{o}_S$ and require that M be G-stable (that is, the action map maps M to $M \otimes_{\mathfrak{o}_S} \mathfrak{o}_S[G]$).

Finally, let $\{f_i\}_i$ be a finite set of generators of M over $\mathfrak{o}_S$. Then for a point $x \in X^+(\mathbb{A}_k)$, we define

$$d_{Z,v}(x_v) = \max_i \{|f_i(x_v)|_v\} \tag{3-7}$$

and

$$d_Z(x) = \prod_v d_{Z,v}(x_v). \tag{3-8}$$

We will call this an *adelic distance function* from Z. Notice that almost all factors of this product are 1 since $x \in X^+(\mathbb{A}_k)$. Moreover, the function extends by zero to a continuous function on $X(\mathbb{A}_k)$.

Remark 3.1.6. For $v \notin S$, the local factor $d_{Z,v}$ depends only on M and not the choices of the f_i: It is the absolute value of the fractional ideal generated by the image of M under $x_v : \mathfrak{o}_S[X] \to \mathfrak{o}_v$. Moreover, the restriction of $d_{Z,v}$ to $X(\mathfrak{o}_v)$ does not depend on M, either, since the image of J generates the same fractional ideal. (The restriction of $d_{Z,v}$ to $X(\mathfrak{o}_v)$ is a height function, that is, q_v raised to the intersection number of $x \in X(\mathfrak{o}_v)$ with Z.)

Finally, the restriction of d_Z to any compact subset of $X(\mathbb{A}_k)$ is up to a constant multiple independent of choices. Indeed, such a compact subset is the product of $X(\mathfrak{o}_v)$, for v outside of a finite number of places $S' \supset S$, with a compact subset of $\prod_{v \in S'} X(k_v)$; therefore it suffices to prove independence for the $d_{Z,v}$ when $v \in S'$. For any two sets of functions $\{f_j\}_j, \{f'_i\}_i$ as above, we can write $f'_i = \sum_j h_{ij} f_j$ with $h_{ij} \in \mathfrak{o}_S[X]$ and for each $v \in S'$ there is a constant C_v such that $|h_{ij}(x_v)|_v \le C_v$ when x is in the given compact set. Then $\max_i |f'_i(x_v)|_v \le C_v \max_j |f'_j(x_v)|_v$, and therefore $d'_Z(x) \le C d_Z(x)$ in the given compact set, where $C = \prod_{v \in S'} C_v$.

For two complex valued functions f_1 and f_2 we will write $f_1 \ll^p f_2$ (where the exponent p stands for "polynomially") if there exists a polynomial P such that $|f_1| \le P(|f_2|)$. We will say that f_1 and f_2 are *polynomially equivalent* if $f_1 \ll^p f_2$ and $f_2 \ll^p f_1$.

In this language, it is easy to see that the assumption of Section 3.1.1 on growth of Schwartz functions close to the complement of X^+ is equivalent to the following:

If Z denotes the complement of X^+ in X, then for any adelic distance function d_Z from Z and any $\Phi \in \mathcal{S}(X(\mathbb{A}_k))$, we have

$$|\Phi(x)| \ll^p d_Z(x)^{-1} \tag{3-9}$$

for every $\Phi \in \mathcal{S}(X(\mathbb{A}_k))$.

Indeed, let the functions f_i be as in the assumption of Section 3.1.1 and let the functions f'_j define an adelic distance function as above. By enlarging S we may assume that $f_i \in \mathfrak{o}_S[X]$ for all i, and by enlarging it further we may assume that the support of Φ is the product of $\prod_{v \notin S} X(\mathfrak{o}_v)$ with a compact subset of $\prod_{v \in S} X(k_v)$. By the assumption, the functions f_i generate an ideal whose radical contains J. Therefore, $(f_i)_i \supset J^n$ for some J and hence for each j there are $h_{ij} \in \mathfrak{o}_S[X]$ such that

$$(f'_j)^n = \sum_i h_{ij} f_i.$$

Therefore for $v \notin S$ and $x_v \in X(\mathfrak{o}_v)$ we have

$$d_{Z,v}(x_v)^n \le \max_i |f_i(x)|,$$

and for $v \in S$ we can find C_v such that $|h_{ij}(x_v)|_v \le C_v$ if x is in the support of Φ. Therefore, for x in the support of Φ, we have

$$\prod_v (\max_i |f_i(x_v)|_v)^{-1} \le \prod_{v \in S} C_v^{-1} \cdot d_Z(x)^{-n}.$$

Vice versa, if Φ is known to be polynomially bounded by $d_Z(x)^{-1}$, then it is bounded by a constant times $d_Z(x)^{-n}$ for some n (since $d_Z(x)$ is bounded in the support of Φ), which implies the bound of the assumption.

3.1.7. *Proof of Proposition 3.1.3.* Recall that an automorphic function ϕ is "of moderate growth" if $\phi \ll^p \|g\|$ for some natural norm $\|\cdot\|$ on G_∞. Recall that a "natural norm" is a positive function on G_∞ that is polynomially equivalent to $\|\rho(g)\|$, where ρ denotes an algebraic embedding $G \hookrightarrow \mathrm{GL}_n$, and $\|g\| := \max\{|g|_{l^\infty}, |g^{-1}|_{l^\infty}\}$ on $\mathrm{GL}_n(k_\infty)$ (where $|\cdot|_{l^\infty}$ denotes the operator norm for the standard representation of GL_n on $l^\infty(\{1, \dots, n\})$).

Assume without loss of generality that $\Phi = \bigotimes_v \Phi_v$, with $\Phi_v \in \mathcal{S}(X_v)$, and let $S_\Phi = \prod S_{\Phi_v}$, where S_{Φ_v} is the support of Φ_v in $X(k_v)$ (a compact subset).

The claim of the proposition will follow from (3-9) if, in addition, we establish that (for $g \in G_\infty$ and $x \in X^+(\mathbb{A}_k)$)

- $\#(X^+(k) \cap S_\Phi g) \ll^p \|g\|$, and
- $(\inf d_Z(X^+(k)g))^{-1} \ll^p \|g\|$.

Indeed, assuming these properties we have

$$\theta(\Phi, g) = \sum_{\gamma \in X^+(k)} (g \cdot \Phi)(\gamma) \le \#(X^+(k) \cap S_\Phi g^{-1}) \cdot \sup_{x \in X^+(k)} |\Phi(xg)| \ll^p \\ \ll^p \|g\| \cdot \Big(\inf_{x \in X^+(k)} d_Z(xg)\Big)^{-1} \ll^p \|g\| \cdot \|g\|.$$

The first property is standard, and follows from the analogous claim for GL_n (after fixing an equivariant embedding of X in the vector space of a representation of G), since S_Φ is a compact subset of $X(\mathbb{A}_k)$.

To prove the second property, we may assume that the elements $f_i \in k[X]$ defining d_Z span over k a G-invariant space $M \subset k[X]$ and that the norm on G_∞ is induced by the $l^\infty(\{f_i\}_i)$-operator norm on $\mathrm{GL}(M_\infty)$. (If the homomorphism $G \to \mathrm{GL}(M)$ is not injective, then this l^∞ norm is bounded by some natural norm on G_∞, which is enough for the proof of this property.) Then for every $x \in X_\infty$ and $g \in G_\infty$, we have

$$\|g\|^{-1} \cdot d_{Z,\infty}(x) \le d_{Z,\infty}(x \cdot g) \le \|g\| \cdot d_{Z,\infty}(x)$$

(where we keep assuming that d_Z is defined by a basis for M).

We apply this to points $x \in X^+(F)$. For every $x \in X^+(k)$, $f_i(x)$ is in k and is nonzero for at least one i; hence $d_Z(x) = \prod_v \max_i |f_i(x)|_v \ge \max_i \prod_v |f_i(x)|_v = 1$. Therefore, we have $d_{Z,\infty}(x \cdot g) \ge \|g\|^{-1} \cdot d_{Z,\infty}(x) \ge \|g\|^{-1}$. □

3.2. ***Conjectural properties of the Schwartz space.*** We saw in Proposition 3.1.4 that, under very mild assumptions on the basic functions Φ_v^0, the Mellin transform of the corresponding X-theta series converges for sufficiently X-positive characters ω. However, there is no reason to expect in general that it admits meromorphic continuation to the set of all ω. Indeed, this often fails for the most naive choice of basic functions, namely the characteristic functions of $X^+(\mathfrak{o}_v)$. We discuss an example, which will be encountered again in Section 4.5:

Example 3.2.1. Let $G = (\mathrm{PGL}_2)^3 \times \mathbb{G}_m$, and let H denote the subgroup

$$\left\{ \begin{pmatrix} a & x_1 \\ & 1 \end{pmatrix} \times \begin{pmatrix} a & x_2 \\ & 1 \end{pmatrix} \times \begin{pmatrix} a & x_3 \\ & 1 \end{pmatrix} \times a \;\middle|\; x_1 + x_2 + x_3 = 0 \right\}.$$

If we defined the local Schwartz space to be equal to $C_c^\infty(H_v \backslash G_v)$, with basic function Φ_v equal to the characteristic function of $(H \backslash G)(\mathfrak{o}_v)$ (which is equal to the characteristic function of a single $G(\mathfrak{o}_v)$-orbit), then, as we will explain in more detail in Section 4.2, the integral of a cusp form against an X-Eisenstein series is equal to the period integral of a cusp form on G over $H(k) \backslash H(\mathbb{A}_k)$, and the usual “unfolding” method shows that this can be written as

$$\int_{\mathbb{A}_k^\times} W_1\begin{pmatrix} a & \\ & 1\end{pmatrix} W_2\begin{pmatrix} a & \\ & 1\end{pmatrix} W_3\begin{pmatrix} a & \\ & 1\end{pmatrix} |a|^s\, da,$$

where the W_i are Whittaker functions of cusp forms on PGL_2 and the parameter s depends on the restriction of the given representation to $\mathbb{G}_m(\mathbb{A}_k)$ (assumed to factor through the absolute value map, for simplicity). For $\Re(s)$ large this integral can be written as a convergent Euler product of the analogous local integrals.

An explicit but lengthy computation shows that, if the $W_i(1)$ are normalized to be equal to 1, if a, b, c denote the Satake parameters of the three PGL_2-cusp forms (considered as elements in $\mathbb{C}^\times$, well-defined up to inverse), and if we set $Q = q^{-3/2-s}$, then the local unramified factors of this Euler product are equal, for a certain choice of measure on $\mathbb{A}_k^\times$, equal to

$$\frac{(-1+3Q^2+3Q^4-Q^6)+(Q^2+Q^4)(a^2+a^{-2}+b^2+b^{-2}+c^2+c^{-2})}{\prod_{\sigma=(\sigma_1,\sigma_2,\sigma_3)\in\{\pm1\}^3}(1-Qa^{\sigma_1}b^{\sigma_2}c^{\sigma_3})} - \frac{2Q^3(a+a^{-1})(b+b^{-1})(c+c^{-1})}{\prod_{\sigma=(\sigma_1,\sigma_2,\sigma_3)\in\{\pm1\}^3}(1-Qa^{\sigma_1}b^{\sigma_2}c^{\sigma_3})}.$$

The denominator of this expression is very pleasant (it is equal to the denominator of the tensor product L-function of the three cuspidal representations), but the numerator does not represent an L-function and it would be unreasonable to expect that its Euler product admits meromorphic continuation. Therefore, this was not the correct Schwartz space.

The conjectures that follow are very speculative, but will provide the suitable ground for unifying various methods of integral representations of L-functions. There are several reasonable assumptions that one could impose on the spherical variety, the strongest of which would be that for every irreducible admissible representation π of $G(\mathbb{A}_k)$, we have $\dim_{G(\mathbb{A}_k)}(\pi, C^\infty(X^+(\mathbb{A}_k))) \le 1$. At the very minimum, we require from now on that the arithmetic multiplicity (Section 2.1) of X is trivial. Equivalently, at every place v there is a unique open $B(k_v)$-orbit, and this also implies that generic G-stabilizers are connected[6] and therefore, at almost every (finite) place v, the space $X^+(\mathfrak{o}_v)$ is homogeneous under $G(\mathfrak{o}_v)$.

Conjecture 3.2.2. *Given an affine spherical variety X over k with trivial arithmetic multiplicity, there exists a Schwartz space $\mathcal{S}(X(\mathbb{A}_k))$, in the sense described above, such that*

[6] If H is not connected then we have a finite cover $H^0 \backslash G \to H \backslash G$ that gives rise to a finite cover of the associated open B-orbits. But this implies that the B-stabilizer B_x of a generic point is not connected; hence $H^1(k, B_x) \neq 0$, and therefore $(B_x \backslash B)(k) \supsetneq B_x(k) \backslash B(k)$.

- *the basic functions* Φ_v^0 *factor through the map of the generalized Cartan decomposition*

$$\{G(\mathfrak{o}_v)\text{-orbits on } X_v^+\} \to \Lambda_X^+$$

and as functions on Λ_X^+ *are equal to the functions obtained via the function-sheaf correspondence from the "basic sheaf" of Gaitsgory and Nadler*, *as will be explained in* 3.3.3; *and*

- *for every* $\Phi \in \mathcal{S}(X(\mathbb{A}_k))$, *the* X*-Eisenstein series* $E(\Phi, \omega, g)$, *originally defined for sufficiently* X*-positive characters*, *admits a meromorphic continuation everywhere.*

Remarks 3.2.3. (1) The first property could be taken as the definition of the basic function, if one knew that the functions obtained from the Gaitsgory–Nadler sheaf are independent of some choices, which we will explain in Section 3.3.3. In any case, such a definition would be very ad hoc and not useful; one hopes that there exists an alternative construction of the Schwartz space, as in [Braverman and Kazhdan 1999].

(2) The property of meromorphic continuation is mostly dependent on the basic vectors and not on the whole Schwartz space; for instance, at a finite number of places we may replace any function with a function whose (local) Mellin transform is a meromorphic multiple of the Mellin transform of the original function without affecting the meromorphicity property. Therefore, the properties do not determine the Schwartz space uniquely; they should hold, for instance, if we take $\mathcal{S}(X_v)$ to be the G-space generated by the basic vector and $C_c^\infty(X_v^+)$.

(3) The fact that the theta series is defined with reference to the group G (since we are summing over the k-points of its open orbit) certainly seems unnatural; it would be more "geometric" to sum over the k-points of the open subvariety where $\mathfrak{Z}(X)$ acts faithfully. However, this does not affect the validity of Conjecture 3.2.2, since one case can be inferred from the other by induction on the dimension of X.

The conjecture about meromorphic continuation of the Mellin transform is a very strong one (see Section 4.5 for an example) and, in fact, is not even known in the case of usual Eisenstein series, that is, the case of $X = \overline{U_P \backslash G}^{\text{aff}}$, where U_P is the unipotent radical of a parabolic P (except when P is a Borel subgroup). We now formulate a weaker conjecture that says that the X-Eisenstein series can be continued meromorphically "as functionals on the space of automorphic forms". In fact, the precise interpretation of them as functionals on the whole space of automorphic forms would require a theory similar to the spectral decomposition of the relative trace formula, that lies beyond the scope of the present paper. Therefore, we confine ourselves to the cuspidal component of this functional. (Notice, however, that there are a lot of interesting examples which have zero cuspidal contribution, e.g., $X = \mathrm{Sp}_{2n} \setminus \mathrm{GL}_{2n}$.)

Conjecture 3.2.4 (weak form). *Same assumptions as in Conjecture 3.2.2, but the second property is replaced by the following*:

- *For every cusp form ϕ on $G(k) \setminus G(\mathbb{A}_k)$, the integral*

$$\int_{[G]} \phi \cdot \omega(g)\theta(\Phi, g)\, dg \tag{3-10}$$

originally defined for sufficiently X-positive idele class characters ω of G, admits meromorphic continuation to the space of all idele class characters of G.

Remark 3.2.5. Following up on the third part of Remarks 3.2.3, we will see in Proposition 4.4.3 that for the large class of *wavefront* spherical varieties (see Section 2.1), the integral (3-10) is the same whether the theta series is defined by summation over $X^+(k)$ or over the largest subvariety where $\mathcal{Z}(X)$ acts faithfully. The reason is a phenomenon that has frequently been observed in the Rankin–Selberg method, namely that the stabilizers of points in all but the open orbit contain unipotent radicals of proper parabolics. Although this is not a feature of the Rankin–Selberg method only, we present the proof there in order not to interrupt the exposition here.

3.3. ***Geometric models and the basic function.*** We now discuss the geometric models and explain the first requirement of Conjecture 3.2.2. The models we are about to discuss are relevant to a spherical variety X over an *equal-characteristic* local field F, and are not local, but global in nature.

3.3.1. *The Gaitsgory–Nadler spaces* [Gaitsgory and Nadler 2010]. Let X be an affine spherical variety over $\mathbb{C}$, and let C be a smooth complete complex algebraic curve. Consider the ind-stack $\mathcal{Z}$ of *meromorphic quasimaps* which, by definition, classifies data

$$(c, \mathcal{P}_G, \sigma),$$

where $c \in C$, $\mathcal{P}_G$ is a principal G-bundle on C, and σ is a section $C \smallsetminus \{c\} \to \mathcal{P}_G \times^G X$ whose image is not contained in $X \smallsetminus X^+$. Clearly, $\mathcal{Z}$ is fibered over C (projection to the first factor). It is a stack of infinite type; however it is a union of open substacks of finite type, each being the quotient of a scheme by an affine group, and therefore one can define intersection cohomology sheaves on it without a problem.

The same definitions can be given if G and X are defined over a finite field $\mathbb{F}$.

To any quasimap one can associate an element of $X^+(\mathcal{K})/G(\mathfrak{O})$ (where $\mathfrak{O} = \mathbb{C}[[t]]$, $\mathcal{K} = \mathbb{C}((t))$) as follows: Choose a trivialization of $\mathcal{P}_G$ in a formal neighborhood of c and an identification of this formal neighborhood with $\operatorname{spec}(\mathfrak{O})$ — then the section σ defines a point in $X^+(\mathcal{K})$, that depends on the choices made. The corresponding coset in $X^+(\mathcal{K})/G(\mathfrak{O})$ is independent of choices.

This allows us to stratify our space according to the stratification, provided by Theorem 2.3.1, of $X^+(\mathcal{K})/G(\mathfrak{O})$. We only describe some of the strata here: For $\theta \in \Lambda_X^+$, let $\mathscr{Z}^\theta$ denote the quasimaps of the form $(c, \mathcal{P}_G, \sigma : C \smallsetminus \{c\} \to \mathcal{P}_G \times^G X^+)$ that correspond to the coset $\theta \in X^+(\mathcal{K})/G(\mathfrak{O})$ at c. Then $\mathscr{Z}^\theta$ can be thought of as a (global) geometric model for that coset. The *basic stratum* $\mathscr{Z}^0$ consists of quasimaps of the form $(c \in C, \mathcal{P}_G, \sigma : C \to \mathcal{P}_G \times^G X^+)$. Notice that these substacks do not depend on the compactification X of X^+. Their *closure*, though, does. For instance, the closure of $\mathscr{Z}^0$ can be identified with an open substack in the quotient stack X_C/G_C over C, namely the stack whose S-objects are S-objects of X_C/G_C but not of $(X \smallsetminus X^+)_C/G_C$. These are the quasimaps for which the corresponding point in $X^+(\mathcal{K})/G(\mathfrak{O})$ lies in the image of $X^+(\mathcal{K}) \cap X(\mathfrak{O})$. Hence, the closure of $\mathscr{Z}^0$ should be thought of as a geometric model for $X^+(\mathcal{K}) \cap X(\mathfrak{O})$.

Since the spaces of Gaitsgory and Nadler are global in nature, it is in fact imprecise to say that they are geometric models for local spaces. However, *their singularities* are expected to model the singularities of $G(\mathfrak{O})$-invariant subsets of $X^+(\mathcal{K})$.

3.3.2. *Drinfeld's compactifications.* The spaces of Gaitsgory and Nadler described above are (slightly modified) generalizations of spaces introduced by Drinfeld in the cases $X = \overline{U_P \backslash G}^{\text{aff}}$ or $X = \overline{[P, P] \backslash G}^{\text{aff}}$, where $P \subset G$ is a proper parabolic and U_P its unipotent radical. The corresponding spaces are denoted by $\widetilde{\mathrm{Bun}}_P$ and $\overline{\mathrm{Bun}}_P$, respectively. Our basic references here are [Braverman and Gaitsgory 2002; Braverman et al. 2002]. The only differences between the definition of these stacks and the stacks $\mathscr{Z}$ of Gaitsgory and Nadler are that the section σ has to be defined on all C, and it does not have a distinguished point c. Therefore, for a quasimap in Drinfeld's spaces and any point $c \in C$, the corresponding element of $X^+(\mathcal{K})/G(\mathfrak{O})$ has to belong to the cosets that belong to $X(\mathfrak{O})$. (These will be described later when we review the computations of [Braverman et al. 2002].)

This particular case is very important to us because it is related to Eisenstein series, and moreover the intersection cohomology sheaf of the "basic stratum" has been computed (when G, X are defined over $\mathbb{F}$).

3.3.3. *The basic function.* We return to the setting where X is an affine spherical variety for a split group G over a local, nonarchimedean field F whose ring of integers we denote by $\mathfrak{o}$ and whose (finite) residue field we denote by $\mathbb{F}$. We assume that X, G and the completions $\overline{X}$, $\hat{X}$ introduced before have the properties of Proposition 2.3.5 over $\mathfrak{o}$, and denote $K = G(\mathfrak{o})$. The goal is to define the "basic function" Φ^0 on $X^+(F)$, which will be K-invariant and supported in $X(\mathfrak{o})$. This function will factor through the map $X^+(F)/K \to \Lambda_X^+$ of Theorem 2.3.8. The idea is to define a function on Λ_X^+ using equal-characteristic models of X.

Define the Gaitsgory–Nadler stack $\mathscr{Z}$ as in Section 3.3.1 over $\mathbb{F}$. Since, by assumption, $X_{\mathbb{F}}$ has a completion $\overline{X}_{\mathbb{F}}$ with the properties of Proposition 2.3.5 (and, hence, the same holds for the base change $X_{\mathbb{F}[[t]]}$), the generalized Cartan decomposition 2.3.8 holds for $G(\mathbb{F}[[t]])$-orbits on $X^+(F((t)))$: They admit a natural map onto Λ_X^+. Hence the strata $\mathscr{Z}^\theta$ of $\mathscr{Z}$ are well-defined over $\mathbb{F}$. Let IC^0 denote the intersection cohomology sheaf of the closure of the basic stratum $\mathscr{Z}^0$ (how exactly to normalize it is not important at this point, since we will normalize the corresponding function). We will obtain the value of our function at $\check{\lambda} \in \Lambda_X^+$ as the trace of Frobenius acting on the stalk of IC^0 at an $\mathbb{F}$-object $x_{\check{\lambda}}$ in the stratum $\mathscr{Z}^{\check{\lambda}}$. However, since these strata are only locally of finite type, and not of pure dimension, we must be careful to make compatible choices of points as $\check{\lambda}$ varies. (It is expected that IC^0 is locally constant on the strata — this will be discussed below.)

The compatibility condition is related to the natural requirement that the action of the unramified Hecke algebra on the functions that will be obtained from sheaves is compatible, via the function-sheaf correspondence, with the action of its geometric counterpart on sheaves. First of all, let us fix a quasimap $x_0 = (c_0, \mathscr{P}_0, \sigma_0)$ in the $\mathbb{F}$-objects of the basic stratum $\mathscr{Z}^0$. Now consider the subcategory $\mathscr{Z}_{x_0}$ of $\mathscr{Z}$ consisting of $\mathbb{F}$-quasimaps $(c_0, \mathscr{P}_G, \sigma)$ with the property that there exists an isomorphism $\iota : \mathscr{P}_0|_{C\smallsetminus\{c_0\}} \xrightarrow{\sim} \mathscr{P}_G|_{C\smallsetminus\{c_0\}}$ (inducing isomorphisms between $\mathscr{P}_0 \times^G X$ and $\mathscr{P}_G \times^G X$, also to be denoted by ι) such that $\sigma = \iota \circ \sigma_0$. Hence, the objects in $\mathscr{Z}_{x_0}$ are those obtained from x_0 via *meromorphic Hecke modifications* at the point c_0 [Gaitsgory and Nadler 2010, §4].

For each $\check{\lambda} \in \Lambda_X^+$, pick an object $x_{\check{\lambda}} \in \mathscr{Z}_{x_0}$ that belongs to the stratum $\mathscr{Z}^{\check{\lambda}}$. We define the *basic function* Φ^0 on Λ_X^+ to be

$$\Phi^0(\check{\lambda}) = c \cdot \sum_i (-1)^i \operatorname{tr}(\mathrm{Fr}, H^i(IC^0_{x_{\check{\lambda}}})), \tag{3-11}$$

where $IC^0_{x_{\check{\lambda}}}$ denotes the stalk of IC^0 at $x_{\check{\lambda}}$ and Fr denotes the geometric Frobenius. The constant c (independent of $\check{\lambda}$) is chosen so that $\Phi^0(0) = 1$.

Now we return to $X(F)$ and we identify Φ^0 with a K-invariant function on $X^+(F)$ (also to be denoted by Φ^0) via the stratification of Theorem 2.3.8.

This is the "basic function" of Conjecture 3.2.2 at the given place. The definition implies that the support of the basic function is contained in $X(\mathfrak{o})$, since the closure of the basic stratum includes the stratum $\mathscr{Z}^\theta$ only if θ corresponds to a $G(\mathfrak{o})$-orbit belonging to $X(\mathfrak{o})$. The independence of choices of the basic function is widely expected but, in the absence of suitable finite-dimensional geometric models, not known. We impose it as an assumption, together with other properties that should naturally follow from the properties of intersection cohomology if one had suitable local models. Notice that for $X = \overline{U_P \backslash G}^{\mathrm{aff}}$ or $X = \overline{[P, P] \backslash G}^{\mathrm{aff}}$, one could have used instead the Drinfeld models of 3.3.2 to define the basic function.

Assumption 3.3.4. (1) The basic function Φ^0 on $X^+(F)$ is well-defined and independent of

- the choices of objects $x_{\check{\lambda}}$;
- (if $X = \overline{U_P \backslash G}^{\mathrm{aff}}$ or $X = \overline{[P, P] \backslash G}^{\mathrm{aff}}$) which model of Section 3.3 one uses to define them;
- the group G acting on X; more precisely, if G_1, G_2 act on X and we denote by X_1^+, X_2^+ the open orbits, then the restriction of Φ^0 to $X_1^+(F) \cap X_2^+(F)$ should be the same.

(2) If Z is an affine homogeneous spherical G-variety and $p : X \to Z$ a surjective equivariant morphism, then the basic function on X, evaluated at any point $x \in X^+(F) \cap X(\mathfrak{o})$, is equal to the basic function of the fiber of p over $p(x)$ (considered as a $G_{p(x)}$-spherical variety).

We discuss how to deduce the growth assumption on elements of the Schwartz space (Section 3.1) for the basic function. Assume now that X is defined globally over a number field k, and fix a finite set of places S_0 and suitable $\mathfrak{o}_{S_0}$-models as in Proposition 2.3.5. Recall (Section 3.1.5) that these models define a distance function $d_Z = \prod_{v \notin S_0} d_{Z,v}$ from $Z = X \smallsetminus X^+$ on $\prod_{v \notin S_0} X(\mathfrak{o}_v)$.

Proposition 3.3.5. *Assume that there are a $\chi \in \mathcal{X}(X) \otimes \mathbb{R}$ such that*

$$|\Phi_v^0(\check{\lambda})| \le q_v^{\langle \chi, \check{\lambda} \rangle}$$

for all places v and all $\check{\lambda} \in \Lambda_X^+$ (where $q_v = |\mathbb{F}_v|$). Then there is a natural number n such that

$$\left| \prod_{v \notin S_0} \Phi_v^0(x) \right| \le (d_Z(x))^{-n} \quad \textit{for all } x \in X^+(\mathbb{A}_k^{S_0}).$$

Here $\mathbb{A}_k^{S_0}$ denotes the adeles outside of S_0. Of course, the function is zero off $\prod_{v \notin S_0} X(\mathfrak{o}_v)$ so the extension of the distance function off integral points of X plays no role in the statement.

Proof. First of all, we claim:

The local distance function $d_{Z,v}$ on $X(\mathfrak{o}_v)$ is $G(\mathfrak{o}_v)$-invariant.

Indeed, $G(\mathfrak{o}_v)$ preserves the ideal of Z in $\mathfrak{o}_v[X]$ and therefore its image in $\mathfrak{o}_v$ under any $\mathfrak{o}_v$-point.

Hence, since both d_Z and $\prod_{v\notin S_0}\Phi_v^0$ are $\prod_{v\notin S_0} G(\mathfrak{o}_v)$-invariant, it suffices to prove the proposition for a set of representatives of $\prod_{v\notin S_0} G(\mathfrak{o}_v)$-orbits in the support of $\prod_{v\notin S_0}\Phi_v^0$, namely elements of $A_X(\mathbb{A}_k^{S_0})$ that at every place v have image in $\check\Lambda_X^+\cap\mathscr{C}(X)$.

Let Y denote the "standard $\mathfrak{o}_{S_0}$-model" of the affine toric embedding of A_X corresponding to the cone $\check\Lambda_X^+\cap\mathscr{C}(X)$. By assumption (see Proposition 2.3.5), there is a morphism $Y\to X$. Therefore, if Y_1 denotes the complement of the open orbit on Y, the corresponding distance functions on $A_X(k_v)$, for every $v\notin S_0$, compare as $d_{Z,v}\le d_{Y_1,v}$. On the other hand, clearly, for every $\chi\in\mathscr{X}(X)\otimes\mathbb{R}$ there is a natural number n such that

$$d_{Y_1,v}^{-n}\ge q_v^{\langle\chi,\check\lambda\rangle}$$

on $A_X(k_v)\cap Y(\mathfrak{o}_v)$ for all $v\notin S_0$. The claim follows. □

4. Periods and the Rankin–Selberg method

4.1. ***Preflag bundles.*** We are about to describe the geometric structure that gives rise to Rankin–Selberg integrals. We hasten to clarify, and it will probably be clear to the reader, that it is not a very natural structure from the general point of view that we have taken thus far, and its occurrence should be seen as a coincidence. Indeed, the structure is not defined in terms of the original group G, but in terms of a possibly different group $\tilde G$, and relies on being able to decompose the variety by a sequence of maps with simple, easily identifiable fibers.

We keep assuming that $\mathfrak{X}(G)^0 \xrightarrow{\sim} \mathfrak{X}(X)$. We will say that an affine spherical G-variety X has the structure of a *preflag bundle* if it is the affine closure of a G-stable subvariety $\tilde X^+$, which has the following structure:

(1) $\tilde X^+$ is homogeneous under a reductive group $\tilde G$;

(2) there is a diagram of homogeneous $\tilde G$-varieties with surjective morphisms

$$\begin{array}{l} \tilde X^+ \\ \Big\downarrow {\scriptstyle L\text{-torsor}} \\ \tilde Y \\ \Big\downarrow {\scriptstyle \text{fiber over } y\in Y \text{ is a flag variety for } \tilde G_y} \\ Y\ (\simeq G'_y\backslash G'\simeq \tilde G_y\backslash\tilde G \text{ with } G'_y,\ \tilde G_y \text{ reductive}), \end{array}$$

where

- Y is an affine, $\tilde G$-homogeneous variety (called the *base* of the preflag bundle);

- $\tilde{Y}$ is proper over Y (hence the fiber over $y \in Y$ is a flag variety for $\tilde{G}_y$);
- $\tilde{Y}$ is the quotient of $\tilde{X}^+$ by the free, $\tilde{G}$-equivariant action of a reductive group L that contains $\mathcal{Z}(X)$; and
- L is an almost direct factor of G.

Remark 4.1.1. The group G' has been inserted in the diagram for later reference. It is supposed to belong to an almost direct decomposition $G = L \cdot G'$ and it necessarily acts transitively on Y, since $\mathcal{Z}(X)$ acts trivially on Y while, on the other hand, it retracts all points onto a homogeneous subvariety by Proposition 2.2.6.

Hence, the notion of a preflag bundle combines the notion of a homogeneous affine variety (which here is the base Y), with the notion of a *preflag variety*, that is, a quasiaffine quotient of $N'' \backslash G''$ by a subgroup of M'', where $M''N''$ is the Levi decomposition of a parabolic of G'' (here, the fibers over Y are such,[7] setting G'' equal to the stabilizer of a point on Y). Of course, each of these constituents can be trivial; for instance Y can be a point (in which case we are dealing with a preflag variety, but possibly for a different group than G), or X could be equal to Y (in which case we are dealing with affine homogeneous varieties).

In this paper we will additionally impose the condition, without mentioning it further, that the fiber $\tilde{X}_y^+$ over $y \in Y$ is a product of varieties $[P_i, P_i] \backslash G_i$ or is of the form $U_{P_i} \backslash G_i$, where $\prod_i G_i = \tilde{G}_y$. This condition will allow us to restrict our discussion to Eisenstein series induced either from cusp forms or from characters of parabolic subgroups, and to use the computations of [Braverman et al. 2002]. Notice that the dual group of L acts on the unipotent radical of the dual parabolic to $\tilde{P}_y$ inside of the dual group of $\tilde{G}_y$; indeed the quotient $\tilde{P}_y \twoheadrightarrow L$ gives rise to a homomorphism

$$\check{L} \to \check{\tilde{L}}_y,$$

where $\check{\tilde{L}}_y$ is the standard Levi dual to $\tilde{P}_y$. We let $\check{\mathfrak{u}}_{\tilde{P}}$ denote[8] the Lie algebra of the unipotent radical of the parabolic dual to $\tilde{P}_y$, considered as a representation of $\check{L}$.

[7] Notice that L is necessarily a quotient of a Levi subgroup of $\tilde{G}_y$. Indeed, if we write as $\tilde{X}_y^+ = \tilde{H}_y \backslash \tilde{G}_y \to \tilde{P}_y \backslash \tilde{G}_y$ the map between the fibers of $\tilde{X}^+$, resp. $\tilde{X}^+/L$ over $y \in Y$, where $\tilde{P}_y$ is a parabolic of $\tilde{G}_y$, then L can be identified with a subgroup of $\mathrm{Aut}^{\tilde{G}_y}(X_y)$ preserving the fiber of this map, that is with a subgroup of $\mathcal{N}_{\tilde{P}_y}(\tilde{H}_y)/\tilde{H}_y$. Since it acts transitively on the fibers of this map, it follows that $\tilde{H}_y$ must be normal in $\tilde{P}_y$, and L must be the quotient $\tilde{P}_y/\tilde{H}_y$. Since L is reductive, this also implies that $\tilde{H}_y$ contains the unipotent radical of $\tilde{P}_y$.

[8] It would be more correct to consider only what will later be denoted by $\check{\mathfrak{u}}_{\tilde{P}}^f$ for those factors of $\tilde{X}_y^+$ that are of the form $[P_i, P_i] \backslash G_i$, but that does not make any difference for the statement of Theorem 4.1.3 below, since we are only using $\check{\mathfrak{u}}_{\tilde{P}}$ to require the meromorphic continuation of an L-function, and the difference if we took $\check{\mathfrak{u}}_{\tilde{P}}^f$ instead would just be some abelian L-function.

The requirement that $\tilde{G}$ commutes with the action of $\mathfrak{X}(X)$ (by the condition $\mathfrak{X}(X) \subset L$) is meant to allow us to relate the $\mathfrak{X}(X)$-Mellin transforms of X-theta series to usual Eisenstein series on $\tilde{G}_y$ induced from $\tilde{P}_y$.

Example 4.1.2. The variety Mat_n for $\mathrm{GL}_n \times \mathrm{GL}_n$ ($n \geq 2$) is a preflag variety, and more generally so is any N-dimensional vector space (here $N = n^2$) with a linear G-action, as it is equal to the affine closure of $P_N \backslash \mathrm{GL}_N$ (with P_N the mirabolic subgroup). Notice, however, that an $(n+m)$-dimensional vector space ($n, m \geq 2$) can be considered as a preflag variety for both $\tilde{G} = \mathrm{GL}_{n+m}$ and $\tilde{G} = \mathrm{GL}_n \times \mathrm{GL}_m$; which one we will choose will depend on which torus action we will consider (that is, what is $\mathfrak{X}(X)$). For instance, for the second possibility, decomposing the given vector space as $X = V = V_n \oplus V_m$ we find that

(1) Y is a point;
(2) $\tilde{X}^+ = (V_n \smallsetminus \{0\}) \times (V_m \smallsetminus \{0\})$;
(3) $\tilde{G} = \mathrm{GL}(V_n) \times GL(V_m)$;
(4) $L = \mathfrak{X}(X) = \mathbb{G}_m \times \mathbb{G}_m$, the two copies acting on V_n and V_m, respectively; and
(5) we can take $G = \tilde{G}$ (with L identified as its center), or any subgroup thereof that contains the center and acts spherically.

From our point of view, whether a spherical variety is a preflag bundle or not is a matter of "chance" and in fact should be irrelevant as far as properties of X-theta series and their applications go — the fundamental object is just X as a G-variety, and not its structure of a preflag bundle. We will try to provide support for this point of view in Section 4.5. However, in absence of a general proof of Conjecture 3.2.2, this is the only case where its validity, in the weaker form of Conjecture 3.2.4, can be proven. Moreover, the concept of preflag bundles is enough to explain a good part of the Rankin–Selberg method.

We assume throughout in this section that the local Schwartz spaces $\mathcal{S}(X_v)$ are the G-spaces generated by the "basic function", which we extract from computations on Drinfeld spaces (outside of a finite number of places), and by functions in $C_c^\infty(X_v^+)$ obtained as convolutions of delta functions with smooth, compactly supported measures on G_v. (At nonarchimedean places, such functions span $C_c^\infty(X_v)$.) The main result of this section is the following:

Theorem 4.1.3. *Assume that X is a wavefront spherical variety with trivial arithmetic multiplicity that has the structure of a preflag bundle, and let τ vary over a holomorphic family of cuspidal automorphic representations of G (that is, an irreducible cuspidal representation twisted by idele class characters of the group). Let τ_1 denote the isomorphism class of the restriction of τ to L, and assume that for some finite set of places S, the partial L-function $L^S(\tau_1, \check{\mathfrak{u}}_{\tilde{P}}, 1)$ has meromorphic continuation everywhere (as τ varies in this family).*

Then Conjecture 3.2.4 holds for $\phi \in \tau$ and $\mathcal{S}(X_v)$ as described above.

We prove this theorem in Section 4.4 by appealing to the meromorphic continuation of usual Eisenstein series, after explicitly describing the basic vectors according to the computations of intersection cohomology sheaves on Drinfeld spaces in [Braverman et al. 2002]. However, the application of the meromorphic continuation of Eisenstein series is not completely trivial as in some cases we have to use the theory of spherical varieties to show that as we "unfold" this integral certain summands vanish (in the language often used in the theory of Rankin–Selberg integrals, certain G-orbits on X are "negligible"). We start by demonstrating an extreme case, which gives rise to period integrals.

4.2. ***Period integrals.*** First consider the extreme case of a preflag bundle with trivial fibers: Namely, choosing a point $x_0 \in X(k)$, we have $X = H\backslash G$ with $H = G_{x_0}$ *reductive*. Then at each place $v \notin S_0$ the basic function is the characteristic function of $X(\mathfrak{o}_v)$, and we may assume that $\mathcal{S}(X(\mathbb{A}_k)) = C_c^\infty(X(\mathbb{A}_k))$. The multiplicity-freeness assumption of Section 3.2 implies, in particular, that at almost every place $G(\mathfrak{o}_v)$ acts transitively on $X(\mathfrak{o}_v)$. Then we can take $\Phi \in \mathcal{S}(X(\mathbb{A}_k))$ of the form $\Phi = h \star \delta_{x_0}$, where $h \in \mathcal{H}(G(\mathbb{A}_k))$, the Hecke algebra of compactly supported smooth measures on $G(\mathbb{A}_k)$, and δ_{x_0} is the delta function at x_0 (considered as a generalized function).

Then, if $\check{h}$ denotes the element of $\mathcal{H}(G(\mathbb{A}_k))$ adjoint to h, the integral

$$\int_{G(k)\backslash G(\mathbb{A}_k)} \phi \cdot \omega(g)\theta(\Phi, g)\, dg$$

of Conjecture 3.2.4 is equal to

$$\int_{H(k)\backslash H(\mathbb{A}_k)} (\check{h} \star \phi) \cdot \omega(g)\, dg. \tag{4-1}$$

This is called a *period integral*, and such integrals have been studied extensively. Hence *period integrals are the special case of the pairing of Conjecture 3.2.4 that is obtained from preflag bundles with trivial fibers* (that is, affine homogeneous spherical varieties).

For example, when $X = \mathrm{GL}_2$ and $G = \mathbb{G}_m \times \mathrm{GL}_2$, with $\mathbb{G}_m$ acting as a noncentral torus of GL_2 by multiplication on the left, we get the period integral of Hecke (1-2), discussed in the introduction. All spherical period integrals are included in the lists of Knop and van Steirteghem [2006] which we will discuss in the next section.

4.3. ***Connection to usual Eisenstein series.***

4.3.1. *Certain stacks and sheaves related to flag varieties.* The goal of this subsection is to explicate the basic functions Φ_v^0 for preflag bundles, based on the computations of [Braverman et al. 2002]. We work with the varieties $X = \overline{[P,P]\backslash G}^{\mathrm{aff}}$

or $X = \overline{U_P \backslash G}^{\mathrm{aff}}$ and use the notation of Section 3.3.2. We do not aim to give complete definitions of the constructions of [ibid.], but to provide a guide for the reader who would like to extract from it the parts most relevant to our present discussion. The final result will be the following formula for the basic function Φ^0 (locally at a nonarchimedean place, which we suppress from the notation):

Theorem 4.3.2. *Let $X = \overline{H \backslash G}$ in each of the following cases.*

- *If $H = U_P$, then*

$$\begin{aligned}\Phi^0 &= \sum_{i \geq 0} q^{-i} \breve{\mathrm{Sat}}_M \Big(\mathrm{Sym}^i (\check{\mathfrak{u}}_P) \Big) \star 1_{HK} \\ &= \breve{\mathrm{Sat}}_M \left(\frac{1}{\bigwedge^{\mathrm{top}} (1 - q^{-1} \check{\mathfrak{u}}_P)} \right) \star 1_{HK}. \end{aligned} \tag{4-2}$$

- *If $H = [P, P]$, then*

$$\begin{aligned}\Phi^0 &= \sum_{i \geq 0} q^{-i} \breve{\mathrm{Sat}}_{M^{\mathrm{ab}}} \Big(\mathrm{Sym}^i (\check{\mathfrak{u}}_P^f) \Big) \star 1_{HK} \\ &= \breve{\mathrm{Sat}}_{M^{\mathrm{ab}}} \left(\frac{1}{\bigwedge^{\mathrm{top}} (1 - q^{-1} \check{\mathfrak{u}}_P^f)} \right) \star 1_{HK}. \end{aligned} \tag{4-3}$$

Here $\breve{\mathrm{Sat}}$ denotes the power series in the Hecke algebra associated by the Satake isomorphism to the given power series in the representation ring of the dual group — it will be explained in detail in Section 4.3.5.

We denote by $\Lambda_{G,P}$ the lattice of cocharacters of the torus $M/[M, M]$ and by $\Lambda_{G,P}^{\mathrm{pos}}$ the subsemigroup spanned by the images of $\check{\Delta} \smallsetminus \check{\Delta}_M$. For every $\theta \in \Lambda_{G,P}^{\mathrm{pos}}$ we have a canonical locally closed embedding $j_\theta : C \times \mathrm{Bun}_P \to \overline{\mathrm{Bun}}_P$ [Braverman et al. 2002, Proposition 1.5]. The image will be denoted by ${}_{(\theta)}\overline{\mathrm{Bun}}_P$. (Notice: This is not the same as what is denoted in [loc. cit.] by ${}_{\theta}\overline{\mathrm{Bun}}_P$, but rather what is denoted by ${}_{\mathfrak{U}(\theta)}\overline{\mathrm{Bun}}_P$, when $\mathfrak{U}(\theta)$ is the trivial partition of θ.) Its preimage in $\widetilde{\mathrm{Bun}}_P$ will be denoted by ${}_{(\theta)}\widetilde{\mathrm{Bun}}_P$. We have a canonical isomorphism

$${}_{(\theta)}\widetilde{\mathrm{Bun}}_P \simeq \mathrm{Bun}_P \times_{\mathrm{Bun}_M} \mathscr{H}_M^{(\theta)},$$

where $\mathscr{H}_M^{(\theta)}$ is a stack that will be described below.

Remarks 4.3.3. (i) If $X = \overline{[P, P] \backslash G}^{\mathrm{aff}}$ under the $M^{\mathrm{ab}} = M/[M, M] \times G$-action, then Λ_X^+ can be identified with $\Lambda_{G,P}$, and ${}_{(\theta)}\overline{\mathrm{Bun}}_P$ is precisely the analog of what we denoted by $\mathfrak{X}^{w_0\theta}$ on the Gaitsgory–Nadler stacks, where w_0 is the longest element in the Weyl group of G. The reason that only $\theta \in \Lambda_{G,P}^{\mathrm{pos}}$ appear is that, as we remarked in Section 3.3.2, the quasimaps on Drinfeld spaces are, by definition, not allowed to have poles. For the reader who would like to trace this back to the combinatorics of quasiaffine varieties and their affine closures of Section 2.2, we mention that the cone spanned by $\rho(\mathcal{D})$ is the cone spanned by the images of $\check{\Delta} \smallsetminus \check{\Delta}_M$.

(ii) If $X = \overline{U_P \backslash G}^{\text{aff}}$ under the $M \times G$-action, then

$$\Lambda_X^+ \simeq \{\check{\lambda} \in \Lambda_A \mid \langle \check{\lambda}, \alpha \rangle \leq 0 \text{ for all } \alpha \in \Delta_M\}$$

(where we denote by A the maximal torus of G and by Λ_A its cocharacter lattice). There is a map $\Lambda_X \to \Lambda_{G,P}$, and ${}_{(\theta)}\widetilde{\text{Bun}}_P$ corresponds to the union of the strata $\mathfrak{X}^{w_0\check{\lambda}}$ of Gaitsgory–Nadler, with $\check{\lambda}$ ranging over all the M-dominant preimages of θ.

We have the geometric Satake isomorphism, that is, a functor $\text{Loc} : \text{Rep}(\check{G}) \to \text{Perv}(\mathcal{G}_G)$ such that the irreducible representation of $\check{G}$ with highest weight $\check{\lambda}$ goes to the intersection cohomology sheaf of a $G(\mathfrak{o})$-equivariant closed, finite-dimensional subscheme $\overline{\mathcal{G}_G}^{\check{\lambda}}$. We will make use of this functor for M, rather than G. If V is a representation of $\check{M}$ — assumed "positive" (this has to do with the fact that we don't allow poles, but there's no need to explain it here) — and $\theta \in \Lambda_{G,P}^{\text{pos}}$, then we define $\text{Loc}^{(\theta)}(V)$ to be $\text{Loc}(V_\theta)$, where V_θ is the θ-isotypic component of V. (We ignore a twist by $\overline{\mathbb{Q}_l}[1](\frac{1}{2})^{-1}$ introduced in [Braverman et al. 2002], and modify the results accordingly.)

We now introduce relative, global versions of the spaces above. We denote by $\mathcal{H}_M$ the *Hecke stack* of M. It is related to $\mathcal{G}_M$ as follows: If we fix a curve C and a point $x \in \mathcal{C}$ then, by definition, $\mathcal{G}_M$ is the functor Schemes $\to$ Sets that associates to every scheme S the set of pairs $(\mathcal{F}_G, \beta)$, where $\mathcal{F}_M$ is a principal M-bundle over $C \times S$ and β is an isomorphism of it outside of $(C \smallsetminus \{x\}) \times S$ with the trivial M-bundle. The relative version of this, as we allow the point x to move over the curve, is denoted by $\mathcal{G}_{M,C}$, and the relative version of the latter, as we replace the trivial M-bundle with an arbitrary M-bundle, is $\mathcal{H}_M$. It is fibered over $C \times \text{Bun}_M$.

3pt In [ibid., p. 389], certain closed, finite-dimensional subschemes $\mathcal{G}_M^{+,\theta}$ of $\mathcal{G}_M$ are defined for every $\theta \in \Lambda_{G,P}^{\text{pos}}$, which at the level of reduced schemes are isomorphic to $\overline{\mathcal{G}_M}^{\flat(\theta)}$, where $\flat(\theta)$ is an M-dominant coweight associated to θ — the "least dominant" coweight mapping to θ. The relative versions of those give rise to substacks $\mathcal{H}_M^{(\theta)}$ of $\mathcal{H}_M$.

For these relative versions we have: Functors $\text{Loc}_{\text{Bun}_M,C}$ (resp. $\text{Loc}^{(\theta)}_{\text{Bun}_M,C}$) from $\text{Rep}(\check{M})$ to perverse sheaves on $\mathcal{H}_M$ (resp. $\mathcal{H}_M^{(\theta)}$) and $\text{Loc}_{\text{Bun}_P,C}$ (resp. $\text{Loc}^{(\theta)}_{\text{Bun}_P,C}$) to perverse sheaves on $\text{Bun}_P \times_{\text{Bun}_M} \mathcal{H}_M$ (resp. $\text{Bun}_P \times_{\text{Bun}_M} \mathcal{H}_M^{(\theta)}$), the latter being $IC_{\text{Bun}\,P}$ along the base Bun_P.

Then the main theorem of Braverman et al. [Theorem 1.12] is a description of the $*$-restriction of $IC_{\widetilde{\text{Bun}}_P}$ to ${}_{(\theta)}\widetilde{\text{Bun}}_P \simeq \text{Bun}_P \times_{\text{Bun}_M} \mathcal{H}_M^{(\theta)}$. Moreover, [Theorem 7.3] does the same thing for $IC_{\overline{\text{Bun}}_P}$ and ${}_{(\theta)}\overline{\text{Bun}}_P \simeq C \times \text{Bun}_P$. The normalization of IC sheaves is that they are pure of weight 0; that is, for a smooth variety Y of dimension n we have $IC_Y \simeq (\overline{\mathbb{Q}_l}(\frac{1}{2})[1])^{\otimes n}$, where $\overline{\mathbb{Q}_l}(\frac{1}{2})$ is a fixed square root of q.

Theorem 4.3.4 [Braverman et al. 2002, Theorems 1.12 and 7.3]. *The $*$-restriction of $IC_{\widetilde{\mathrm{Bun}}_P}$ to ${}_{(\theta)}\widetilde{\mathrm{Bun}}_P \simeq \mathrm{Bun}_P \times_{\mathrm{Bun}_M} \mathcal{H}_M^{(\theta)}$ is equal to*

$$\mathrm{Loc}^{(\theta)}_{\mathrm{Bun}_P, C}\big(\bigoplus_{i\ge 0} \mathrm{Sym}^i(\check{\mathfrak{u}}_P) \otimes \overline{\mathbb{Q}}_l(i)[2i]\big). \tag{4-4}$$

The $*$-restriction of $IC_{\overline{\mathrm{Bun}}_P}$ to ${}_{(\theta)}\overline{\mathrm{Bun}}_P \simeq C \times \mathrm{Bun}_P$ is equal to

$$IC_{{}_{(\theta)}\overline{\mathrm{Bun}}_P} \otimes \overline{\mathrm{Loc}}\big(\bigoplus_{i\ge 0} \mathrm{Sym}^i(\check{\mathfrak{u}}_P^f)_\theta \otimes \overline{\mathbb{Q}}_l(i)[2i]\big). \tag{4-5}$$

Here $\check{\mathfrak{u}}_P$ denotes the adjoint representation of $\check{M}$ on the unipotent radical of the parabolic dual to P. Moreover, $\check{\mathfrak{u}}_P^f$ denotes the subspace that is fixed under the nilpotent endomorphism f of a principal $\mathfrak{sl}_2$-triple (h, e, f) in the Lie algebra of $\check{M}$. For the definition of $\overline{\mathrm{Loc}}(V)$, which takes into account the grading on V arising from the h-action, see [ibid., §7.1].

4.3.5. *The corresponding functions.* Let us fix certain normalized Satake isomorphisms. As before, our local, nonarchimedean field is denoted by F, its ring of integers by $\mathfrak{o}_F$, and our groups are assumed to have reductive models over $\mathfrak{o}_F$. As usual, we normalize the action of $M(F)$ (resp. $M^{\mathrm{ab}}(F)$) on functions on $(H\backslash G)(F)$ where $H = U_P$ (resp. $[P, P]$) so that it is unitary on $L^2((H \backslash G)(F))$:

$$m \cdot f(H(F)g) = \delta_P^{1/2}(m) f(H(F)m^{-1}g), \tag{4-6}$$

where δ_P is the modular character of P. We let $M_0 = M(\mathfrak{o}_F)$, and normalize the (classical) Satake isomorphism as follows:

- For the Hecke algebra $\mathcal{H}(M, M_0)$ in the usual way,

  $$\mathrm{Sat}_M : \mathbb{C}[\check{M}]^{\check{M}} \simeq \mathbb{C}[\mathrm{Rep}\,\check{M}] \xrightarrow{\sim} \mathcal{H}(M, M_0),$$

  where $\mathbb{C}[\mathrm{Rep}\,\check{M}]$ is the Grothendieck algebra over $\mathbb{C}$ of the category of algebraic representations of $\check{M}$.

- For the Hecke algebra $\mathcal{H}(M^{\mathrm{ab}}, M_0^{\mathrm{ab}})$ we shift the usual Satake isomorphism $\mathcal{H}(M^{\mathrm{ab}}, M_0^{\mathrm{ab}}) \simeq \mathbb{C}[\mathfrak{Z}(\check{M})] \simeq \mathbb{C}[\mathrm{Rep}\,\mathfrak{Z}(\check{M})]$ by $e^{-\rho_M}$, where ρ_M denotes half the sum of positive roots of M. In other words, if h is a compactly supported measure on $M(F)/M_0$, considered (canonically) as a linear combination of cocharacters of M^{ab} and hence as a regular function f on the center $\mathfrak{Z}(\check{M})$ of its dual group, then we will assign to h the function $z \mapsto f(e^{\rho_M} z)$ on the subvariety $e^{-\rho_M}\mathfrak{Z}(\check{M})$ of $\check{G}$:

  $$\mathrm{Sat}_{M^{\mathrm{ab}}} : \mathbb{C}[e^{-\rho_M}\mathfrak{Z}(\check{M})] \xrightarrow{\sim} \mathcal{H}(M^{\mathrm{ab}}, M_0^{\mathrm{ab}}).$$

Let 1_{HK} denote the characteristic function of $H \backslash HK$ (where $K = G(\mathfrak{o}_F)$), and consider the action map $\mathscr{H}(M, M_0) \to C_c^\infty((U_P \backslash G)(F))^{M_0 \times K}$, respectively $\mathscr{H}(M^{\mathrm{ab}}, M_0^{\mathrm{ab}}) \to C_c^\infty(([P, P] \backslash G)(F))^K$ given by $h \mapsto h \star 1_{HK}$. The map is bijective, and identifies the module $C_c^\infty((H \backslash G)(F))^{M_0 \times K}$ with $\mathbb{C}[\check{M}]^{\check{M}}$, resp. $\mathbb{C}[e^{-\rho_M} Z(\check{M})]$. Our normalization of the Satake isomorphism is such that this is compatible with the Satake isomorphism $\mathrm{Sat}_G : \mathscr{H}(G, K) = \mathbb{C}[\check{G}]^{\check{G}} = \mathbb{C}[\mathrm{Rep}(\check{G})]$ for G, in the sense that for $f \in \mathbb{C}[\check{G}]^{\check{G}}$ we have

$$\mathrm{Sat}_G(f) \star 1_{HK} = \widecheck{\mathrm{Sat}}_{M \text{ or } M^{\mathrm{ab}}}(f) \star 1_{HK}.$$

Here and later, by the symbol $\check{h}$ we will be denoting the adjoint of the element h in a Hecke algebra. Its appearance is due to the definition (4-6) of the action of M as a right action on the space and a left action on functions. We extend the "Sat" notation to the fraction field of $\mathbb{C}[\mathrm{Rep}\,\check{M}]$ (and, respectively, of $\mathbb{C}[e^{-\rho_M}\mathfrak{Z}(\check{M})]$), where $\mathrm{Sat}_{M \text{ or } M^{\mathrm{ab}}}(R)$ (with R in the fraction field) is thought of as a power series in the Hecke algebra.

Returning to the Drinfeld spaces discussed in the previous subsection, we let $\mathrm{Ff}(E)(x) := \sum_i (-1)^i \operatorname{tr}(\mathrm{Fr}, H^i(E_x))$ denote the alternating sum of the trace of Frobenius acting on the homology of the stalks of a perverse sheaf (Ff stands for "faisceaux-fonctions"). As in Section 3.3.3, we fix an object x_0 on the basic stratum, a point $c_0 \in C$ (recall that in the definition of Drinfeld's spaces, quasimaps do not have distinguished points) and we evaluate $\mathrm{Ff}(E)$, where $E = IC_{\widetilde{\mathrm{Bun}}_P}$ or $IC_{\overline{\mathrm{Bun}}_P}$, only at objects $x_{\check{\lambda}}$ that are obtained by $M \times G$-Hecke modifications at c_0. This way, and using the Iwasawa decomposition, we obtain our basic function Φ^0, which is an $M_0 \times K$-invariant function on $(H \backslash G)(F)$. Recall that it is by definition normalized such that $\Phi^0(H \backslash H1) = 1$.

The study of the Hecke correspondences in [Braverman and Gaitsgory 2002] implies that

$$\mathrm{Ff}(\mathrm{Loc}_{\mathrm{Bun}_P, C}(V)) = \widecheck{\mathrm{Sat}}_M(V) \star \mathrm{Ff}(\mathrm{Loc}_{\mathrm{Bun}_P, C}(1)) \quad \text{if } H = U_P,$$

and

$$\mathrm{Ff}(\overline{\mathrm{Loc}}(V)) = \widecheck{\mathrm{Sat}}_{M^{\mathrm{ab}}}(V) \star \mathrm{Ff}(\overline{\mathrm{Loc}}(1)) \quad \text{if } H = [P, P].$$

Remark 4.3.6. The "unitary" normalization of the action of M is already present in the sheaf-theoretic setting as follows: Suppose that an object $x_{\check{\lambda}}$ belongs to ${}_{(\check{\lambda})}\overline{\mathrm{Bun}}_P$ and can be obtained from x_0 via Hecke modifications at the distinguished object of x_0. Then the dimension of ${}_{(\check{\lambda})}\overline{\mathrm{Bun}}_P \simeq C \times \mathrm{Bun}_P$ at $x_{\check{\lambda}}$ is $\langle \check{\lambda}, 2\rho_P \rangle$ less than that of ${}_{(0)}\overline{\mathrm{Bun}}_P$ around x_0, where ρ_P denotes the half-sum of roots in the unipotent radical of P, that is, $\delta_P = e^{2\rho_P}$. Hence, by the aforementioned normalization of IC sheaves, the contribution of the factor $IC_{(\check{\lambda})}\overline{\mathrm{Bun}}_P$ (via Theorem 4.3.4) to $\Phi^0(\check{\lambda})$

will be $q^{\langle\check\lambda,\rho_P\rangle}$ times the contribution of the factor $IC_{(0)}\overline{\mathrm{Bun}}_P$ to $\Phi^0(0)$. Similarly for the strata of $\widetilde{\mathrm{Bun}}_P$.

Thus, Theorem 4.3.4 translates to the statement of Theorem 4.3.2:

- If $H = U_P$, then $$\Phi^0 = \sum_{i\ge 0} q^{-i}\widecheck{\mathrm{Sat}}_M(\mathrm{Sym}^i(\check{\mathfrak{u}}_P)) \star 1_{HK} = \widecheck{\mathrm{Sat}}_M\left(\frac{1}{\bigwedge^{\mathrm{top}}(1-q^{-1}\check{\mathfrak{u}}_P)}\right)\star 1_{HK}.$$
- If $H = [P, P]$, then $$\Phi^0 = \sum_{i\ge 0} q^{-i}\widecheck{\mathrm{Sat}}_{M^{\mathrm{ab}}}(\mathrm{Sym}^i(\check{\mathfrak{u}}_P^f)) \star 1_{HK} = \widecheck{\mathrm{Sat}}_{M^{\mathrm{ab}}}\left(\frac{1}{\bigwedge^{\mathrm{top}}(1-q^{-1}\check{\mathfrak{u}}_P^f)}\right)\star 1_{HK}.$$

Notice that in the last expression $\check{\mathfrak{u}}_P^f$ is considered as a representation of the maximal torus $\check A$ of $\check M$ determined by the principal $\mathfrak{sl}_2$-triple (h, e, f) and, by restricting its character to the subvariety $e^{-\rho_M}\mathfrak{X}(\check M)$, as an element of $\mathcal{H}(M^{\mathrm{ab}}, M_0^{\mathrm{ab}})$. This is the case studied in [Braverman and Kazhdan 2002], and Φ^0 is the function denoted by $c_{P,0}$ there.

4.3.7. *Connection to Eisenstein series.* Now we discuss our main conjecture when the variety is $X = \overline{U_P\backslash G}^{\mathrm{aff}}$ or $X = \overline{[P, P]\backslash G}^{\mathrm{aff}}$ under the (normalized) action of $M\times G$, resp. $M^{\mathrm{ab}}\times G$. In the latter case, our Eisenstein series $E(\Phi, \omega, g)$ are the usual (degenerate, if P is not the Borel) principal Eisenstein series normalized as in [Braverman and Kazhdan 1999; 2002], and hence $E(\Phi, \omega, g)$ is indeed meromorphic for all ω.

It will be useful to recall how these meromorphic sections are related to the more usual sections $E(f, \omega, g)$, which are defined in the same way but with $f \in C_c^\infty(([P, P]\backslash G)(\mathbb{A}_k))$. We assume that $\Phi = \prod_v \Phi_v$, $f = \prod_v f_v$ and S is a finite set of places (including S_0) such that $\Phi_v = \Phi_v^0$ and $f_v = f_v^0 := 1_{U\backslash G(\mathfrak{o}_v)}$ for $v \notin S$. Let us also assume for simplicity that $\Phi_v = f_v$ for $v \in S$ (a finite number of places certainly do not affect meromorphicity properties). Clearly, for $E(\Phi, \omega, g)$ and $E(f, \omega, g)$ to be nonzero, the character ω must be unramified outside of S. Then by the results of the previous paragraph we have

$$E(\Phi, \omega, g) = L^S(e^{-\rho_M}\omega, \check{\mathfrak{u}}_P^f, 1)E(f, \omega, g), \tag{4-7}$$

where $L^S(e^{-\rho_M}\omega, \check{\mathfrak{u}}_P^f, 1)$ denotes the value at 1 of the partial (abelian) L-function corresponding to the representation $\check{\mathfrak{u}}_P^f$, whose local factors (at each place v) are considered as rational functions on the maximal torus $\check A \subset \check M$ and evaluated at the point $e^{-\rho_M}\omega_v \in e^{-\rho_M}\mathfrak{X}(\check M) \subset \check A$.

Now let us consider the case $X = \overline{U_P\backslash G}^{\mathrm{aff}}$. We let τ vary over a holomorphic family of cuspidal representations of $M\times G$ and let $\tau \mapsto \phi_\tau$ be a meromorphic section; write $\tau = \tau_1 \otimes \tau_2$ according to the decomposition of the group $M \times G$, and assume that, accordingly, $\phi_\tau = \phi_{\tau_1} \otimes \phi_{\tau_2}$, a pure tensor. Assume momentarily that the central character of τ is sufficiently X-positive. If in the notation of Conjecture 3.2.4 we replace the group G by the group $M \times G$, and perform the integration of the conjecture, but only over the factor $M(k) \setminus M(\mathbb{A}_k)$, then this integral can be written as

$$\int_{M(k)\setminus M(\mathbb{A}_k)} \phi_\tau(m, g)\theta(\Phi, (m, g))dm = \phi_{\tau_2}(g) \int_{M(k)\setminus M(\mathbb{A}_k)} \phi_{\tau_1}(m)\theta(\Phi, (m, g))dm. \tag{4-8}$$

It is valued in the space of functions on $G(k) \setminus G(\mathbb{A}_k)$. If $\mathrm{Eis} : I_{P(\mathbb{A}_k)}^{G(\mathbb{A}_k)}(\tau_1) \to C^\infty(G(k) \setminus G(\mathbb{A}_k))$ denotes the usual Eisenstein operator, then by unfolding the last integral we see that it is equal to the Eisenstein series

$$E_M(\Phi, \phi_1, g) := \mathrm{Eis}\Big(\int_{M(\mathbb{A}_k)} \phi_{\tau_1}(m)(m \cdot \Phi)dm\Big)(g), \tag{4-9}$$

hence the connection to usual Eisenstein series.

Proposition 4.3.8. *Assume that the partial L-function $L^S(\tau_1, \check{\mathfrak{u}}_P, 1)$ (for some large enough S) has meromorphic everywhere as τ_1 is twisted by characters of M. Then the expression (4-8) admits meromorphic continuation to all τ_1.*

Proof. By the meromorphic continuation of Eisenstein series, it is enough to show that the integral $(\Phi, \phi_{\tau_1}) \mapsto \int_{M(\mathbb{A}_k)} \phi_{\tau_1}(m)(m\cdot\Phi)dm$, which represents a morphism

$$\iota_{\tau_1} : \mathcal{S}(U_P \setminus G(\mathbb{A}_k)) \to I_{P(\mathbb{A}_k)}^{G(\mathbb{A}_k)}(\tau_1),$$

admits meromorphic continuation in τ_1. This would be the case if Φ was in $C_c^\infty(U_P \setminus G(\mathbb{A}_k))$. The analogous morphism $C_c^\infty(U_P \setminus G(\mathbb{A}_k)) \to I_{P(\mathbb{A}_k)}^{G(\mathbb{A}_k)}(\tau_1)$ will also be denoted by ι_{τ_1}.

Again, we let S be a finite set of places containing S_0 and take functions $\Phi = \prod \Phi_v \in \mathcal{S}(U_P \setminus G(\mathbb{A}_k))$ and $f = \prod_v f_v \in C_c^\infty(U_P \setminus G(\mathbb{A}_k))$ such that for $v \notin S$ $\Phi_v = \Phi_v^0$ is the basic $M_0 \times K$-invariant function of the previous paragraph, $f_v = f_v^0 = 1_{U_P K}$ and for $v \in S$ we have $\Phi_v = f_v$ (for simplicity). Moreover, we assume that τ_1 is unramified for $v \notin S$, otherwise the integral will be zero.

We saw previously that

$$\Phi_v^0 = \widecheck{\mathrm{Sat}}_M\left(\frac{1}{\bigwedge^{\mathrm{top}}(1-q^{-1}\check{\mathfrak{u}}_P)}\right) \star f_v^0.$$

By definition of the Satake isomorphism and the equivariance of ι_τ, in the domain of convergence we have $\iota_{\tau_1}(\Phi) = L^S(\tau_1, \check{\mathfrak{u}}_P, 1)\iota_{\tau_1}(f)$.

Therefore $\mathrm{Eis}(\iota_{\tau_1}(\Phi)) = L^S(\tau_1, \check{\mathfrak{u}}_P, 1)\,\mathrm{Eis}(\iota_{\tau_1}(f))$, and the claim follows from the meromorphic continuation of $\mathrm{Eis}(\iota_{\tau_1}(f))$. $\square$

Remarks 4.3.9. (1) The meromorphic continuation of $L^S(\tau_1, \check{\mathfrak{u}}_P, 1)$ is known in many cases, e.g., for G a classical group and τ generic, by the work of Langlands, Shahidi and Kim; see [Cogdell et al. 2004].

(2) Notice that, as was also observed in [Braverman and Kazhdan 1999; 2002], the Eisenstein series (4-9) has normalized functional equations without L-factors.

4.4. ***The Rankin–Selberg method.*** According to [Bump 2005, §5], the Rankin–Selberg method involves a cusp form on G and an Eisenstein series on a group $\tilde{G}$, where we have either an embedding $G \hookrightarrow \tilde{G}$ or an embedding $\tilde{G} \hookrightarrow G$, or "something more complicated". We certainly do not claim to explain all constructions that have been called "Rankin–Selberg integrals", but let us see how a large part[9] of this method is covered by our constructions.

Let X be a preflag bundle; we will use the notation of Section 4.1. For notational simplicity (the arguments do not change), let us also assume that L is a direct factor of G, that is, $G = L \times G'$. Let $\Phi \in \mathcal{S}(X(\mathbb{A}_k))$. Recall that the X-theta series $\theta(\Phi, g)$ has been defined via a sum over $X^+(k)$, where X^+ denotes the open G-orbit on X. On the other hand, to relate our integrals to usual Eisenstein series, we need to sum over $\tilde{X}^+(k)$, where $\tilde{X}^+$ is the open $\tilde{G}$-orbit. Hence, we define

$$\tilde{\theta}(\Phi, g) = \sum_{\gamma \in \tilde{X}^+(k)} \Phi(\gamma \cdot g).$$

We compare the integral of Conjecture 3.2.4 with the corresponding integral when θ is substituted by $\tilde{\theta}$:

Proposition 4.4.1. *Suppose that X is a wavefront spherical variety with the structure of a preflag bundle. If ϕ is a cusp form on G* (*with sufficiently X-positive central character, so that the following integrals converge*), *then*

$$\int_{G(k)\backslash G(\mathbb{A}_k)} \phi(g)\theta(\Phi, g)\, dg = \int_{G(k)\backslash G(\mathbb{A}_k)} \phi(g)\tilde{\theta}(\Phi, g)\, dg. \tag{4-10}$$

Assume this proposition for now, and let us prove Theorem 4.1.3; at the same time, we will see that the integral of Conjecture 3.2.4 is equal to a Rankin–Selberg integral.

[9] The multiplicity-one property that seems to underlie almost every integral representation of an L-function can be achieved by nonspherical subgroups if we put extra restrictions on the representations we are considering. For example, in the construction of the symmetric square L-function by Bump and Ginzburg [1992], we have $H =$ the diagonal copy of GL_n in $\mathrm{GL}_n \times$(a central quotient of) $\widetilde{\mathrm{GL}_n}^2$, where $\widetilde{\mathrm{GL}_n}$ denotes a metaplectic cover, but one restricts to certain "exceptional" (and induced-from-exceptional) representations on $\widetilde{\mathrm{GL}_n}^2$. The examples that our method covers should be seen as the part of the method where such restrictions do not enter.

Without loss of generality, $\Phi = \prod_v \Phi_v$, and $\phi = \phi_1(l)\phi_2(g)$ according to the decomposition $G = L \times G'$. By Assumption 3.3.4, and repeating the argument of Section 4.2, we may write Φ as the convolution with an element $h \in \mathcal{H}(G'(\mathbb{A}_k))$ of a Schwartz function Φ^y on $X_y(\mathbb{A}_k)$, where $y \in Y(k)$ and the Schwartz function on $X_y(\mathbb{A}_k)$ is considered as a generalized function on $\tilde{X}^+(\mathbb{A}_k)$. Then, as in Section 4.2,

$$\int_{G(k)\backslash G(\mathbb{A}_k)} \phi(g)\tilde{\theta}(\Phi, g)\, dg = \int_{G_y(k)\backslash G_y(\mathbb{A}_k)} \check{h} \star \phi(h)\tilde{\theta}_{\tilde{X}_y^+}(\Phi^y, h),$$

where $\tilde{\theta}_{\tilde{X}_y^+}(\Phi, g)$ denotes the theta series for the $\tilde{G}_y$-spherical variety X_y.

By the decomposition $G = L \times G'$ this is equal to

$$\int_{G'_y(k)\backslash G'_y(\mathbb{A}_k)} \check{h} \star \phi_2(g) \int_{L(k)\backslash L(\mathbb{A}_k)} \phi_1(l)\tilde{\theta}_{\tilde{X}_y^+}(\Phi^y, lg)\, dl\, dg.$$

The inner integral is equal to the Eisenstein series $E_L(\Phi, \phi_1, g')$ on the group $\tilde{G}'_y$, in the notation of (4-9), or a degenerate Eisenstein series as in (4-7), or a product of such,[10] and it has meromorphic continuation under the assumption that $L^S(\tau_1, \check{\mathfrak{u}}_{\tilde{P}}, 1)$ does. Hence, we see that *the integral of Conjecture 3.2.4 is equal to the Rankin–Selberg integral*:

$$\int_{G'_y(k)\backslash G'_y(\mathbb{A}_k)} \check{h} \star \phi_2(g) E_L(\Phi, \phi_1, g)\, dg \tag{4-11}$$

and this also completes the proof of Theorem 4.1.3. In the language of [Bump 2005, §5], our formalism combines the appearance of a subgroup $G_y \subset G$ with an embedding of it into another group: $G_y \hookrightarrow \tilde{G}_y$.

4.4.2. *Proof of Proposition 4.4.1: Negligible orbits.* Proposition 4.4.1 will follow from the following statement on the structure of certain spherical varieties:

Proposition 4.4.3. *If X is a wavefront spherical variety for G with* $\mathrm{Aut}^G(X)$ *finite, then the isotropy groups of all nonopen G-orbits contain the unipotent radical of a proper parabolic of G.*

From this, Proposition 4.4.1 follows easily; in the domain of convergence we have

$$\int_{G(k)\backslash G(\mathbb{A}_k)} \phi(g)\tilde{\theta}(\Phi, g) = \sum_{\xi \in [\tilde{X}^+(k)/G(k)]} \int_{G_\xi(k)\backslash G(\mathbb{A}_k)} \phi(g) g \cdot \Phi(\xi)\, dg,$$

where $[\tilde{X}^+(k)/G(k)]$ denotes any set of representatives for the set of $G(k)$-orbits on $\tilde{X}^+(k)$. Notice that, by the multiplicity-freeness assumption on X, the k-points

[10] Rankin–Selberg constructions with products of Eisenstein series have often been encountered in the literature, e.g., [Bump et al. 1999; Ginzburg and Hundley 2004].

of the open G-orbit form a unique $G(k)$-orbit. The summand corresponding to ξ can be written

$$\int_{G_\xi(\mathbb{A}_k)\backslash G(\mathbb{A}_k)} g\cdot\Phi(\xi)\int_{G_\xi(k)\backslash G_\xi(\mathbb{A}_k)}\phi(hg)\,dh\,dg$$

Since $\mathrm{Aut}^G(\tilde{X}^+/\mathfrak{X}(X))$ is finite, for ξ in the nonopen orbit the stabilizer G_ξ contains the unipotent radical of a proper parabolic by Proposition 4.4.3, and since ϕ is cuspidal the inner integral will vanish. Therefore, only the summand corresponding to the open orbit survives, which folds back to the integral

$$\int_{G(k)\backslash G(\mathbb{A}_k)}\phi(g)\theta(\Phi,g).$$

Proposition 4.4.3, in turn, rests on the following result of Luna. A G-homogeneous variety Y is said to be *induced* from a parabolic P if it is of the form $Y'\times^P G$, where Y' is a homogeneous spherical variety for the Levi quotient of P; equivalently, $Y=H\backslash G$, where $H\subset P$ contains the unipotent radical of P.

Proposition 4.4.4 [Luna 2001, Proposition 3.4]. *A homogeneous spherical variety Y for G is induced from a parabolic $\overline{P}$ (assumed opposite to a standard parabolic P) if and only if the union of $\Delta(Y)$ with the support*[11] *of the spherical roots of Y is contained in the set of simple roots of the Levi subgroup of P.*

Proof of Proposition 4.4.3. For every G-orbit Y in a spherical variety X, there is a simple toroidal variety $\tilde{X}$ with a morphism $\tilde{X}\to X$ that is birational and whose image contains Y. Therefore, it suffices to assume that X is a simple toroidal variety.

Moreover, if $\overline{X}$ denotes the wonderful compactification of X^+ (that is, the simple toroidal compactification with $\mathscr{C}(\overline{X})=\mathscr{V}$), then every simple toroidal variety X admits a morphism $X\to\overline{X}$ which, again, is birational and has the property that every nonopen G-orbit on X goes to a nonopen G-orbit in $\overline{X}$. Indeed, any nonopen G-orbit $Y\subset X$ corresponds to a nontrivial face of $\mathscr{C}(X)$, and its character group $\mathscr{X}(Y)$ is the orthogonal complement of that face in $\mathscr{X}(X)$, which is of lower rank than $\mathscr{X}(X)$; therefore Y has to map to an orbit of lower rank. Moreover, Y is a torus bundle over its image. This reduces the problem to the case where X is a wonderful variety, which we will now assume.

By Proposition 4.4.4, it suffices to show that the union of $\Delta(X)$ and the support of the spherical roots of Y is not the whole set Δ of simple roots. The spherical roots of Y are a proper subset of the spherical roots of X, and $\Delta(Y)=\Delta(X)$. It therefore suffices to prove that for any proper subset $\Theta\subset\Delta_X$, there exists a simple root $\alpha\in\Delta\smallsetminus\Delta(X)$ such that α is not contained in the support of Θ.

[11] The support of a subset in the span of Δ is the smallest set of elements of Δ in the span of which it lies.

Define $\mathfrak{a}^* := \mathcal{X}(A)^* \otimes \mathbb{Q}$ and $\mathfrak{a}^*_{P(X)} = (\Delta(X))^\perp \subset \mathfrak{a}^*$, and consider the canonical quotient map $q : \mathfrak{a} \to \mathcal{Q}$. Denote by $\mathfrak{f}_\varnothing \subset \mathfrak{a}^*$ the antidominant Weyl chamber in $\mathfrak{a}$. Every set of spherical roots $s \subset \Delta_X$ corresponds to a face $\mathcal{V}_s \subset \mathcal{V} = \mathcal{V}_\varnothing \subset \mathcal{Q}$ (more precisely, $\mathcal{V}_s$ is the face spanning the orthogonal complement of s), and similarly every set $r \subset \Delta$ of simple roots of G corresponds to a face $\mathfrak{f}_r \subset \mathfrak{f}_\varnothing$. The simple roots in the support of $\gamma \in \Delta_X$ are those corresponding to the largest face $\mathfrak{f}$ of $\mathfrak{f}_\varnothing$ that is contained in $q^{-1}(\mathcal{V}_{\{\gamma\}})$. Notice that the maximal vector subspace $\mathfrak{f}_\Delta$ of $\mathfrak{f}_\varnothing$ maps into the maximal vector subspace $\mathcal{V}_{\Delta_X}$ of $\mathcal{V}$.

By assumption, $\mathfrak{f}_\varnothing$ surjects onto $\mathcal{V}$. Moreover, since every element of $\mathfrak{f}_\varnothing$ can be written as a sum of an element in $\mathfrak{f}_{\Delta(X)}$ and a nonnegative linear combination of $\check{\Delta}(X) := \{\check{\alpha} \mid \alpha \in \Delta(X)\}$, and since $\check{\Delta}(X)$ is in the kernel of $\mathfrak{a} \to \mathcal{Q}$, it follows that $\mathfrak{f}_{\Delta(X)}$ surjects onto $\mathcal{V}$. Now let $\Theta \subset \Delta_X$ be a proper subset. Let $\mathfrak{f}_s$ be a face of $\mathfrak{f}_{\Delta(X)}$ that surjects onto $\mathcal{V}_\Theta$. Since $\mathfrak{f}_s \neq \mathfrak{f}_\Delta$, there is an $\alpha \in \Delta \smallsetminus \Delta(X)$ that is not in the support of Θ. □

4.5. ***Tensor product L-functions of* $\mathbf{GL_2}$ *cusp forms.*** In Section 3 we proposed a general conjecture involving distributions that are obtained from the geometry of an affine spherical variety X, and in this section we saw how this conjecture is true, and gives rise to period- and Rankin–Selberg integrals, in the case that X admits the structure of a "preflag bundle". It was written above that such a structure should be considered essentially irrelevant and a matter of "chance". We now wish to provide some evidence for this point of view by recalling the known constructions of n-fold tensor product L-functions for GL_2, where $n \le 3$. The point is that while these constructions seem completely different from the point of view of Rankin–Selberg integrals, from the point of view of spherical varieties they are completely analogous!

Before we consider the specific example, let us become a bit more precise about what it means that *a period integral is related to some L-value*. Let $\pi = \bigotimes' \pi_v$ be an (abstract) unitary representation of $G(\mathbb{A}_k)$, the tensor product of unitary irreducible representations π_v of $G(k_v)$ with respect to distinguished unramified vectors u_v^0 (for almost every place v) of norm 1. Let $\mathcal{P}$ be a functional on π. In our applications the functional $\mathcal{P}$ will arise as the composition of a cuspidal automorphic embedding $\nu : \pi \to L^2_{\mathrm{cusp}}(G(k) \backslash G(\mathbb{A}_k))$, assumed unitary, with a period integral or, more generally, the pairing (3-10) with a fixed X-theta series. Let ρ be a representation of the dual group, and let $L(\pi, \rho, s)$ denote the value of the corresponding L-function at the point s. We say that $|\mathcal{P}|^2$ is *related to* $L(\pi, \rho, s)$ if there exist nonzero skew-symmetric forms $\Lambda_v : \pi_v \otimes \bar{\pi}_v \to \mathbb{C}$ for every v such that for any large enough set of places S, and for a vector $u = \bigotimes_{v \in S} u_v^0 \otimes_{v \notin S} u_v$, one has $|\mathcal{P}(u)|^2 = L^S(\pi, \rho, s) \cdot \prod_{v \in S} \Lambda_v(u_v, \bar{u}_v)$. (Of course, for this to happen we must have $\Lambda_v(u_v^0, \bar{u}_v^0) = L_v(\pi_v, \rho_v, s)$.) Moreover, it is required that each Λ_v has a definition that has no reference to any other representation but π_v. The reader will notice that the last condition does not stand the test of mathematical rigor; however, not imposing it would make the rest of the statement void up to whether $\mathcal{P}$ is zero or not. In practice, the Λ_v will be given by reference to some nonarithmetic model for π_v. See [Ichino and Ikeda 2010] for a precise conjecture in a specific case, and [Sakellaridis and Venkatesh 2012] for a more general but less precise conjecture.[12]

Example 4.5.1. If $\mathcal{P}$ denotes the Whittaker period

$$\phi \mapsto \int_{U(k)\backslash U(\mathbb{A}_k)} \phi(u)\psi^{-1}(u)\,du$$

(where ψ is a generic idele class character of the maximal unipotent subgroup) on cusp forms for $G = \mathrm{GL}_n$, then $|\mathcal{P}|^2$ is related to the L-value

$$\frac{1}{L(\pi, \mathrm{Ad}, 1)}.$$

see [Jacquet 2001; Sakellaridis and Venkatesh 2012]. Notice that the examples we are about to discuss admit "Whittaker unfolding" and this factor will enter, although most references introduce a different normalization and ignore this factor.

Now we are ready to discuss our example: Let n be a positive integer, let $G = (\mathrm{GL}_2)^n \times \mathbb{G}_m$, and let H be the subgroup: We let $X = \overline{H \backslash G}^{\mathrm{aff}}$. As usual, we normalize the action of G on functions on X^+ so that it is unitary with respect to the natural measure. Let us see that for $n = 1, 2, 3$, the variety X admits the structure of a preflag bundle, and hence the integral of Conjecture 3.2.4 can be interpreted as a Rankin–Selberg integral, as discussed above:

- $n = 1$. Here $\overline{H \backslash G}^{\mathrm{aff}} = H \backslash G$ and we get the integral (1-2) of Hecke. If $\tau_s = \tau \otimes |\cdot|^s$, where τ is a cuspidal representation of GL_2 (for simplicity, with trivial central character), the square of the absolute value of the corresponding linear functional on $\tau_s \otimes \widetilde{\tau_s}$ is related to the L-value
$$\frac{L(\tau, \frac{1}{2}+s)L(\tilde{\tau}, \frac{1}{2}-s)}{L(\tau, \mathrm{Ad}, 1)}.$$
- $n = 2$. Here the projection of H to GL_2^2 is conjugate to the mirabolic subgroup of GL_2 embedded diagonally. Therefore, the affine closure of $H \backslash G$ is equal to the bundle over $\mathrm{GL}_2^{\mathrm{diag}} \backslash (\mathrm{GL}_2)^2$ with fiber equal to the affine closure of $U_2 \backslash \mathrm{GL}_2$, where U_2 denotes a maximal unipotent subgroup of GL_2. Corresponding to this preflag bundle is a Rankin–Selberg integral "with the

[12] For the sake of completeness, we should mention that when $\mathcal{P}$ comes from a period integral one should in general modify the conjecture above by some "mild" arithmetic factors, such as the sizes of centralizers of Langlands parameters — see [Ichino and Ikeda 2010]. However, in the example we are about to discuss there is no such issue since the group is GL_2.

Eisenstein series on the smaller group" $\mathrm{GL}_2^{\mathrm{diag}}$, namely the classical integral of Rankin and Selberg. If $\tau = \tau_1 \otimes \tau_2 \otimes |\cdot|^s$ is a cuspidal automorphic representation of G (for simplicity, with trivial central character), the square of the absolute value of the corresponding integral is related to the L-value

$$\frac{L(\tau_1 \otimes \tau_2, \frac{1}{2} + s)L(\tilde{\tau}_1 \otimes \tilde{\tau}_2, \frac{1}{2} - s)}{L(\tau, \mathrm{Ad}, 1)}.$$

- $n = 3$. In this case there is a structure of a preflag variety not on X, but on X^0, the corresponding spherical variety for the subgroup

$$G^0 = \{(g_1, g_2, g_3, a) \in G \mid \det(g_1) = \det(g_2) = \det(g_3)\}.$$

The structure of a preflag variety involves the group $\tilde{G} = \mathrm{GSp}_6$ and the subgroup $\tilde{H} = [\tilde{P}, \tilde{P}]$, where $\tilde{P}$ is the Siegel parabolic — this is a construction of Garrett [1987]. The group $(\mathrm{GL}_2^3)^0$ is embedded in GSp_6 as $(\mathrm{GSp}_2^3)^0$. Then, according to [Piatetski-Shapiro and Rallis 1987, Corollary 1 to Lemma 1.1], the group G^0 has an open orbit in $[\tilde{P}, \tilde{P}] \backslash \tilde{G}$ with stabilizer equal to H.

Lemma 4.5.2. *The affine closure X^0 of $H \backslash G^0$ is equal to the affine closure of $[\tilde{P}, \tilde{P}] \backslash \tilde{G}$.*

Proof. Denote by Y the affine closure of $[\tilde{P}, \tilde{P}] \backslash \tilde{G}$. We have an open embedding $X^0 \hookrightarrow Y$. By [Piatetski-Shapiro and Rallis 1987, Lemma 1.1], all nonopen G-orbits have codimension at least two. Therefore, the embedding is an isomorphism. □

Hence, our integral for X^0 coincides with the Rankin–Selberg integral of Garrett. The only thing that remains to do is to compare the normalizations for the sections of Eisenstein series. From [Piatetski-Shapiro and Rallis 1987, Theorem 3.1], one sees that the square of the absolute value of our integral is related to the L-value

$$\frac{L(\tau_1 \otimes \tau_2 \otimes \tau_3, \frac{1}{2} + s)L(\tilde{\tau}_1 \otimes \tilde{\tau}_2 \otimes \tilde{\tau}_3, \frac{1}{2} - s)}{L(\tau, \mathrm{Ad}, 1)}.$$

(Again, for simplicity, we assume trivial central characters. Notice that the zeta factors in [Piatetski-Shapiro and Rallis 1987, Theorem 3.1] disappear because of the correct normalization of the Eisenstein series!)

It is completely natural to expect the corresponding integral for $n = 4$ or higher to be related to the n-fold tensor product L-function. It becomes obvious from the example above that the point of view of the spherical variety is the natural setting for such integrals, while at the same time the structure of a preflag bundle may not exist and, even if it exists, it has a completely different form in each case, which conceals the uniformity of the construction.

5. Smooth affine spherical varieties

Given that we do not know how to prove Conjecture 3.2.4, except in the cases of wavefront preflag bundles, it is natural to ask the purely algebro-geometric question, Which spherical varieties admit the structure of a preflag bundle? An answer would amount to a complete classification of Rankin–Selberg integrals, in the restricted sense that "Rankin–Selberg" has been used here. Such an answer has been given in the special case of *smooth* affine spherical varieties: These varieties automatically have the structure of a preflag bundle, and they have been classified by Knop and Van Steirteghem [2006], and hence can be used to produce Eulerian integrals of automorphic forms! There seems to be little point in computing every single example in the tables of [Knop and Van Steirteghem 2006], and my examination of most of the cases has not produced any striking new examples. However, we get some of the best-known integral constructions, as well as some new ones (which do not produce any interesting new L-functions).

5.1. ***Smooth affine spherical triples.*** By Theorem 2.2.5 of Luna, every smooth affine spherical variety of G (over an algebraically closed field in characteristic zero) is of the form $V \times^H G$, where H is a reductive subgroup (so that $H \backslash G$ is affine) and V is an H-module. As we have seen in Example 4.1.2, vector spaces are preflag varieties, and therefore all smooth affine spherical varieties are preflag bundles. We check the details carefully:

Lemma 5.1.1. *Every smooth affine spherical variety admits the structure of a preflag bundle.*[13]

Proof. If $X = V \times^H G$ as above, we set $Y = (\mathcal{N}(H)^0 \cdot H) \backslash G$. We let $\tilde{X}^+$ be the subvariety on which $\mathfrak{Z}(X)$ acts freely, and take $\tilde{G} = G$. Clearly, $\mathfrak{Z}(X)$ contains the connected centralizer of H in $\mathrm{GL}(V)$ (which is a torus, since X is spherical), so if $V = \bigoplus_i V_i$ is the decomposition into irreducible H-representations according to $\mathfrak{Z}(H)^0$, then $\tilde{X}^+ = \prod_i (V_i \smallsetminus \{0\}) \times^H G$, and G acts transitively on $\tilde{X}^+$. By the assumption $\mathfrak{Z}(X) = \mathfrak{Z}(G)^0$, $\mathfrak{Z}(X)$ is the connected center of $\mathcal{N}(H)$, and hence $\tilde{Y} := \tilde{X}^+/\mathfrak{Z}(X)$ has fibers $\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_n$ over Y and is therefore proper over Y. □

The corresponding integrals include all period integrals over reductive subgroups, as well as Rankin–Selberg integrals involving *mirabolic* Eisenstein series (that is, those induced from the mirabolic subgroup of GL_n).

[13]Strictly speaking, the "affine closure" condition is not satisfied when the fibers have one-dimensional summands under the action of $\mathfrak{Z}(X)$; one should modify the definition of a preflag bundle to allow this case, but in order not to over complicate things we prefer not to do so. Notice that after integrating by characters of $\mathfrak{Z}(X)$ the "basic function" of $\mathbb{G}_m$ differs from that of $\mathbb{G}_a$ only by a Dirichlet L-function, so the meromorphic properties of the integrals we are considering are not affected by whether we compactify $\mathbb{G}_m$ or not.

In [2006], Knop and Van Steirteghem classify all smooth affine spherical *triples* $(\mathfrak{g}, \mathfrak{h}, V)$, which amounts to a classification of smooth affine spherical varieties up to coverings, central tori and $\mathbb{G}_m$-fibrations. We recall their definitions:

Definition 5.1.2. (1) Let $\mathfrak{h} \subset \mathfrak{g}$ be semisimple Lie algebras and let V be a representation of $\mathfrak{h}$. For $\mathfrak{s}$, a Cartan subalgebra of the centralizer $c_{\mathfrak{g}}(\mathfrak{h})$ of $\mathfrak{h}$, put $\bar{\mathfrak{h}} := \mathfrak{h} \oplus \mathfrak{s}$, a maximal central extension of $\mathfrak{h}$ in $\mathfrak{g}$. Let $\mathfrak{z}$ be a Cartan subalgebra of $\mathfrak{gl}(V)^{\mathfrak{h}}$ (the centralizer of $\mathfrak{h}$ in $\mathfrak{gl}(V)$). We call $(\mathfrak{g}, \mathfrak{h}, V)$ a *spherical triple* if there exists a Borel subalgebra $\mathfrak{b}$ of $\mathfrak{g}$ and a vector $v \in V$ such that

   (a) $\mathfrak{b} + \bar{\mathfrak{h}} = \mathfrak{g}$ and
   (b) $[(\mathfrak{b} \cap \bar{\mathfrak{h}}) + \mathfrak{z}]v = V$, where $\mathfrak{s}$ acts via any homomorphism $\mathfrak{s} \to \mathfrak{z}$ on V.

(2) Two triples $(\mathfrak{g}_i, \mathfrak{h}_i, V_i)$ for $i = 1, 2$ are *isomorphic* if there exist isomorphisms of Lie algebras resp. vector spaces $\alpha : \mathfrak{g}_1 \to \mathfrak{g}_2$ and $\beta : V_1 \to V_2$ such that

   (a) $\alpha(\mathfrak{h}_1) = \mathfrak{h}_2$ and
   (b) $\beta(\xi v) = \alpha(\xi)\beta(v)$ for all $\xi \in \mathfrak{h}_1$ and $v \in V_1$.

(3) Triples of the form $(\mathfrak{g}_1 \oplus \mathfrak{g}_2, \mathfrak{h}_1 \oplus \mathfrak{h}_2, V_1 \oplus V_2)$ with $(g_i, h_i, V_i) \neq (0, 0, 0)$ are called *decomposable*.

(4) Triples of the form $(\mathfrak{k}, \mathfrak{k}, 0)$ and $(0, 0, V)$ are said to be *trivial*. A pair $(\mathfrak{g}, \mathfrak{h})$ of semisimple Lie algebras is called spherical if $(\mathfrak{g}, \mathfrak{h}, 0)$ is a spherical triple.

(5) A spherical triple (or pair) is *primitive* if it is nontrivial and indecomposable.

Clearly, every smooth affine spherical variety gives rise to a spherical triple. Conversely, each spherical triple is obtained from a (not necessarily unique) smooth affine spherical variety, as follows by an a posteriori inspection of all spherical triples. (The nonobvious step here is that the $\mathfrak{h}$-module V integrates to an H-module, where H is the corresponding subgroup.)

The classification of all primitive spherical triples is given in [ibid., Tables 1, 2, 4 and 5], modulo the inference rules described in [Table 3]. The diagrams are read in the following way: The nodes in the first row correspond to the simple direct summands $\mathfrak{g}_i$ of $\mathfrak{g}$, the ones in the second row to the simple direct summands $\mathfrak{h}_i$ of $\mathfrak{h}$ and the ones in the third row to the simple direct summands V_i of V. If $(\mathfrak{g}, \mathfrak{h})$ contains a direct summand of the form $(\mathfrak{h}_1, \mathfrak{h}_1)$, then the $\mathfrak{h}_1$ summand is omitted from the first row. There is an edge between $\mathfrak{g}_i$ and $\mathfrak{h}_j$ if $\mathfrak{h}_j \hookrightarrow \mathfrak{g} \twoheadrightarrow \mathfrak{g}_i$ is nonzero and an edge between h_j and V_k if V_k is a nontrivial h_j-module. The edges are labeled to describe the inclusion of $\mathfrak{h}$ in $\mathfrak{g}$, resp. the action of $\mathfrak{h}$ on V; the labels are omitted when those are the "natural" ones.

We number the cases appearing in the list of Knop and Van Steirteghem as follows: First, according to the table on which they appear [Tables 1, 2, 4, 5]; and for each table, numbered from left to right, top to bottom.

5.2. *Eulerian integrals arising from smooth affine varieties.* In what follows we will discuss a sample of the global integrals obtained from varieties in the list of Knop and Van Steirteghem [2006]. At this point it is more convenient not to normalize the action of G unitarily. We allow ourselves to choose the spherical variety corresponding to a given spherical triple as is most convenient, and in fact we sometimes replace semisimple groups by reductive ones. Of course, the classification in [ibid.] is over an algebraically closed field, which leaves a lot of freedom for choosing the precise form of the spherical variety over k, even when G is split. In the discussion that follows we will always take both the group and generic stabilizer to be split. Many of the varieties of Knop and Van Steirteghem have zero cuspidal contribution (that is, the integral (3-10) is zero for every cusp form) or are not multiplicity-free. Still, this list contains some of the best-known examples of integral representations of L-functions. It contains also some new ones.

In Section 4.5 we explained what it means for a period integral $\mathcal{P}$ to be "related to" an L-value, namely by considering the value of $\mathcal{P}|_\pi \cdot \mathcal{P}|_{\tilde{\pi}}$, assuming that π is an abstract unitary representation of an adelic group, embedded unitarily into the space of cuspidal automorphic forms for that group. For the examples that we are about to see, we will adopt a language that describes the value of $\mathcal{P}|_\pi$ itself, divided by the value of a period integral that does not depend on a continuous parameter, such as the Whittaker period. For example, for the Hecke integral (1-2) we say that it is related to $L(\pi, s+\frac{1}{2})$ with respect to Whittaker normalization, while for the Godement–Jacquet integral (1-1) we say that it is related to $L(\pi, s-\frac{1}{2}(n-1))$ with respect to the "inner product" period on $\pi \otimes \tilde{\pi}$.

5.2.1. *Table 1.* In this table the group H is equal to G, that is, the data consists of a group and a spherical representation of it. This table contains the following interesting integrals (numbered according to their occurrence in the tables of Knop and Van Steirteghem):

1. The integrals of Godement and Jacquet. Here the group is $\mathrm{GL}_n \times \mathrm{GL}_m$ with the tensor product representation (that is, on $\mathrm{Mat}_{n\times m}$). It is easy to see that if $m \neq n$, then the stabilizer is parabolically induced; hence the only interesting case (as far as cusp forms are concerned) is $m = n$. In this case, our integral (3-10) is that of Godement and Jacquet:

$$\int_{Z^{\mathrm{diag}}(\mathbb{A}_k)\,\mathrm{GL}_n^{\mathrm{diag}}(k)\backslash \mathrm{GL}_n(\mathbb{A}_k)\times \mathrm{GL}_n(\mathbb{A}_k)} \phi_1(g_1)\phi_2(g_2)\Phi(g_1^{-1}g_2)|\det(g_1^{-1}g_2)|^s d(g_1, g_2).$$

15. Two new integrals. (Here there is a choice between the first and the last fundamental representation of GL_n. It can easily be seen that they amount to the same integral, so we will consider only ω_1.)

The group is $\mathrm{GL}_m \times \mathrm{GL}_n$ and the representation is the direct sum $\mathrm{Mat}_{m\times n}$ with the standard representation for GL_n. If $m \neq n, n-1$ then we can easily see that the stabilizer is parabolically induced. Hence there are two interesting cases:

(i) $m = n$. We let $\phi_1 \in \pi_1$, $\phi_2 \in \pi_2$ be two cusp forms on GL_n. Then the integral is

$$\int_{P_n^{\mathrm{diag}}(k)\backslash \mathrm{GL}_n(\mathbb{A}_k)\times \mathrm{GL}_n(\mathbb{A}_k)} \phi_1(g_1)\phi_2(g_2)\Phi(g_1^{-1}g_2)\Phi'([0,\dots,0,1]\cdot g_1) \\ \cdot |\det(g_1^{-1}g_2)|^{s_1}|\det(g_1)|^{s_2}\,dg_1\,dg_2.$$

Here Φ is a Schwartz function on $\mathrm{Mat}_n(\mathbb{A}_k)$ and Φ' is a Schwartz function on $\mathbb{A}_k^n$.

Theorem 5.2.2. *The integral above is Eulerian and with respect to Whittaker normalization is related to the L-value*

$$L(\pi_1\otimes\pi_2, s_2)\cdot L(\pi_2, s_1 - \tfrac{1}{2}(n-1)). \tag{5-1}$$

Proof. It follows from the standard "unfolding" technique that the integral above, in the domain of convergence, is equal to

$$\int_{(U_n(\mathbb{A}_k)\backslash \mathrm{GL}_n(\mathbb{A}_k))^2} W_1(g_1)W_2'(g_2)\Phi(g_1^{-1}g_2)\Phi'([0,\dots,0,1]\cdot g_1) \\ \cdot |\det(g_1^{-1}g_2)|^{s_1}|\det(g_1)|^{s_2}\,dg_1\,dg_2,$$

where $W_1(g) = \int_{U_n(k)\backslash U_n(\mathbb{A}_k)} \phi_1(ug)\psi(u)\,du$ and W_2' the is same but with ϕ_1 replaced by ϕ_2 and ψ replaced by ψ^{-1}.

The last integral is (for "factorizable data") a product of local factors:

$$\int_{(U_n(k_v)\backslash \mathrm{GL}_n(k_v))^2} W_{1,v}(g_1)W_{2,v}'(g_2)\Phi_v(g_1^{-1}g_2)\Phi_v'([0,\dots,0,1]\cdot g_1) \\ \cdot |\det(g_1^{-1}g_2)|^{s_1}|\det(g_1)|^{s_2}\,dg_1\,dg_2.$$

Assume that $\Phi_v = \Phi_v^0$, the basic function of $\mathcal{S}(\mathrm{Mat}_n(k_v))$. By considering the action of the spherical Hecke algebra of G_2 (that is, the second copy of GL_n) on $\mathcal{S}(\mathrm{Mat}_n(k_v))$, the work of Godement and Jacquet [1972, Lemma 6.10] proves that

$$\Phi_v^0(x)|\det(x)|^{s_1} = \widecheck{\mathrm{Sat}}_{G_2}\left(\frac{1}{\bigwedge^{\top}\left(1-q_v^{-s_1+\frac{1}{2}(n-1)}\cdot \mathrm{std}\right)}\right)\star 1_{\mathrm{GL}_n(\mathfrak{o})} \tag{5-2}$$

Therefore for unramified data, the last integral is equal to

$$\begin{aligned}
&L(\pi_2, s_1 - \tfrac{1}{2}(n-1)) \\
&\quad\cdot \int_{(U_n(k_v)\backslash \mathrm{GL}_n(k_v))^2} W_{1,v}(g_1) W'_{2,v}(g_2) 1_{\mathrm{GL}_n(\mathfrak{o}_v)}(g_1^{-1} g_2) \Phi'_v([0,\dots,0,1]\cdot g_1) \\
&\qquad\qquad \cdot |\det(g_1^{-1} g_2)|^{s_1} |\det(g_1)|^{s_2}\, dg_1\, dg_2 \\
&\quad = L(\pi_2, s_1 - \tfrac{1}{2}(n-1)) \\
&\qquad\qquad \cdot \int_{(U_n(k_v)\backslash \mathrm{GL}_n(k_v))} W_{1,v}(g) W'_{2,v}(g) \Phi'_v([0,\dots,0,1]\cdot g) |\det(g)|^{s_2}\, dg.
\end{aligned}$$

The latter is the classical Rankin–Selberg integral, which with respect to Whittaker normalization is related to $L(\pi_1 \otimes \pi_2, s_2)$; see, for instance, [Cogdell 2003]. □

(ii) $m = n-1$. Notice that if V denotes the standard representation of GL_n, then the space $\mathrm{Mat}_{(n-1)\times n} \oplus V$ can be identified under the $G_1 \times G_2 := \mathrm{GL}_{n-1} \times \mathrm{GL}_n$-action with the space $X = \mathrm{Mat}_n$, where $g \in G_1$ acts as $\mathrm{diag}(g^{-1}, 1)$ on the left Let $\phi_1 \in \pi_1$ be a cusp form on GL_{n-1} and $\phi_2 \in \pi_2$ a cusp form in GL_n. Then the integral is

$$\int_{\mathrm{GL}_n^{\mathrm{diag}}(k)\backslash \mathrm{GL}_{n+1}(\mathbb{A}_k)\times \mathrm{GL}_n(\mathbb{A}_k)} \phi_1(g_1)\phi_2(g_2) \cdot \Phi\big(\mathrm{diag}(g_1^{-1}, 1) g_2\big) \left| \frac{\det(g_2)}{\det(g_1)} \right|^{s_1} |\det(g_1)|^{s_2}\, dg_1\, dg_2,$$

where $\Phi \in \mathcal{S}(\mathrm{Mat}_n(\mathbb{A}_k))$.

Theorem 5.2.3. *The integral above is Eulerian and with respect to Whittaker normalization related to the L-value*

$$L(\pi_1 \otimes \pi_2, s_2 + \tfrac{1}{2}) \cdot L(\pi_2, s_1 - \tfrac{1}{2} n). \tag{5-3}$$

5.2.4. *Table 2.* In this table H is smaller than G and the representation V of H is nontrivial. This table contains the following interesting integrals:

1. The Bump–Friedberg integral. The group is GL_{m+n}, where $m = n$ or $n+1$, the subgroup H is $\mathrm{GL}_m \times \mathrm{GL}_n$ and the representation is the standard representation of the second factor. This is the integral examined in [Bump and Friedberg 1990]:

$$\int_{\mathrm{GL}_m(k)\times \mathrm{GL}_n(k)\backslash \mathrm{GL}_m(\mathbb{A}_k)\times \mathrm{GL}_n(\mathbb{A}_k)} \phi\, \mathrm{diag}(g_1, g_2) \left| \frac{\det(g_1)}{\det(g_2)} \right|^{s_1} \cdot \Phi([0,\dots,0,1]\cdot g_2) |\det g_2|^{s_2}\, dg_1\, dg_2.$$

It is related with respect to Whittaker normalization to the L-value

$$L(\pi, s_1 + \tfrac{1}{2}) L(\pi, \wedge^2, s_2).$$

3. A new integral. The group is $\mathrm{GL}_{m+1} \times \mathrm{GL}_n$, and $G' = \mathrm{GL}_m \times \mathrm{GL}_n$ with the tensor product of the standard representations (that is, on $\mathrm{Mat}_{m\times n}$). The only interesting case is $m = n$. If $n > m$, then the stabilizer is parabolically induced, and when $m > n$ it unfolds to a parabolically induced model.

If $m = n$, we get

$$\int_{\mathrm{GL}^{\mathrm{diag}}(k)\backslash \mathrm{GL}_n(\mathbb{A}_k)\times \mathrm{GL}_n(\mathbb{A}_k)} \phi_1 \operatorname{diag}(g_1, 1)\phi_2(g_2)\Phi(g_1^{-1}g_2) \cdot \left|\frac{\det(g_2)}{\det(g_1)}\right|^{s_1} |\det(g_1)|^{s_2} d(g_1, g_2).$$

The next result is proved as before:

Theorem 5.2.5. *The integral above is Eulerian and with respect to Whittaker normalization related to the L-value*

$$L(\pi_1 \otimes \pi_2, s_2 + \tfrac{1}{2}) \cdot L(\pi_2, s_1 - \tfrac{1}{2}(n-1)). \tag{5-4}$$

5. The classical Rankin–Selberg integral. The group is $\mathrm{GL}_n \times \mathrm{GL}_n$ and the subgroup G' is $\mathrm{GL}_n^{\mathrm{diag}}$ with the standard representation. This is the classical Rankin–Selberg integral,

$$\int_{\mathrm{GL}_n(k)\backslash \mathrm{GL}_n(\mathbb{A}_k)} \phi_1(g)\phi_2(g)\Phi([0, \dots, 0, 1]\cdot g)|\det g|^s \, dg.$$

It is related with respect to Whittaker normalization to the L-value $L(\pi_1\otimes\pi_2, s)$; see [Cogdell 2003].

5.2.6. *Tables 4 and 5.* Here the representation V is trivial; hence we get period integrals over reductive algebraic subgroups (Section 4.2). All known cases of multiplicity-free period integrals are contained in these tables.

6. A remark on a relative trace formula

At this point we drop our assumptions on the group G, in order to discuss nonsplit examples. We will assume the existence of Schwartz spaces with similar properties in this setting, in order to give a conceptual explanation to the phenomenon of "weight factors" in a relative trace formula.

The relative trace formula is a method that was devised by Jacquet and his coauthors to study period integrals of automorphic forms. In its most simplistic form, it can be described as follows: Let H_1 and H_2 be two reductive spherical subgroups of G (a reductive group defined over a global field k) and let $f \in C_c^\infty(G(\mathbb{A}_k))$. Then one builds the usual kernel function

$$K_f(x, y) = \sum_{\gamma\in G(k)} f(x^{-1}\gamma y)$$

for the action of f on the space of automorphic functions and (ignoring analytic difficulties) defines the functional

$$\mathrm{RTF}^G_{H_1,H_2}(f) = \int_{H_1(k)\backslash H_1(\mathbb{A}_k)} \int_{H_2(k)\backslash H_2(\mathbb{A}_k)} K_f(h_1,h_2)\, dh_1\, dh_2. \tag{6-1}$$

The functional can be decomposed in two ways, one geometric and one spectral, and the spectral expansion involves period integrals of automorphic forms. By comparing two such RTFs (that is, made with different choices of H_1, H_2, maybe even different groups G) one can deduce properties of those period integrals, such as that their nonvanishing characterizes certain functorial lifts.

The presentation above is too simplistic for several reasons: First, the correct functional has something to do with the stack-theoretic quotient $H_1 \backslash G / H_2$, which sometimes forces one to take a sum over certain inner forms of G and H_i. We will not discuss stack-theoretic quotients or inner forms here, but at first approximation we observe that from this algebro-geometric point of view the variety $H_i\backslash G$ is more natural than the space $H_i(k)\backslash G(k)$; hence, if $G(k)$ does not surject onto $(H_i\backslash G_i)(k)$ one should take the sum of the expressions above over stabilizers $H_{i,\epsilon}$ of a set of representatives of $G(k)$-orbits. (This will become clearer in a reformulation we will present below.) Moreover, one can consider an idele class character η of H_i and integrate against this character; we will adjust our notation accordingly, for instance, $\mathrm{RTF}^G_{H_1,(H_2,\eta)}$. There are often analytic difficulties in making sense of the integrals above. And one does not have to restrict to reductive subgroups, but can consider parabolically induced subgroups together with a character on their unipotent radical (such as in the Whittaker period). However, we will ignore most of these issues and focus on another one, first noticed in [Jacquet et al. 1993]: It seems that in certain cases, in order for the relative trace formula $\mathrm{RTF}^G_{H_1,H_2}$ to be comparable to some other relative trace formula, the functional (6-1) is not the correct one and one has to add a "weight factor" in the definition, such as

$$\mathrm{RTF}^G_{H_1,H_2}(f) = \int_{H_1(k)\backslash H_1(\mathbb{A}_k)} \int_{H_2(k)\backslash H_2(\mathbb{A}_k)} K_f(h_1,h_2)\theta(h_1)\, dh_1\, dh_2, \tag{6-2}$$

where θ is a suitable automorphic form on H_1.

Our goal here is to explain how, under the point of view developed in this paper, the expression above is not a relative trace formula for H_1, H_2 but represents a relative trace formula for some *other* subgroups. We will discuss this in the context of [Jacquet et al. 1993], though our starting point will not be (6-2) but another formula of [ibid.] from which the identities for (6-2) are derived, and which is closer to our point of view.

More precisely, let E/F be a quadratic extension of number fields with corresponding idele class character η, $G = \mathrm{Res}_{E/F}\,\mathrm{PGL}_2$, $G' = \mathrm{PGL}_2 \times \mathrm{PGL}_2$ (over F),

$H \subset G$ the projectivization of the quasisplit unitary group (which is in fact split, that is, isomorphic to PGL_2 over F), H' = the diagonal copy of PGL_2 in G'. (Compared to [Jacquet et al. 1993], we restrict to PGL_2 for simplicity.) We consider η as a character of H in the natural way. Naively, one would like to compare the functional $\mathrm{RTF}^G_{H,(H,\eta)}$ to the functional $\mathrm{RTF}^{G'}_{H',H'}$ (usual trace formula for G'). However, it turns out that the correct comparison is between the functionals

$$f \mapsto \int_{(H(k)\backslash H(\mathbb{A}_k))^2} K_f(h_1,h_2)E(h_1,s)\eta(h_1)\,dh_1\,dh_2, \tag{6-3}$$

$$f' \mapsto \int_{(H'(k)\backslash H'(\mathbb{A}_k))^2} K_{f'}(h'_1,h'_2)E'(h'_1,s)\,dh'_1\,dh'_2 \tag{6-4}$$

on G and G' respectively, where E, E' are suitable Eisenstein series on H, H'. (More precisely, in the first case one takes the sum over the unitary groups of all $G(k)$-conjugacy classes of nondegenerate hermitian forms for E/F, as we mentioned above, *but only in the second variable*.)

We have already made a modification to the formulation of [Jacquet et al. 1993], namely in the second case they let $G' = \mathrm{PGL}_2$ and consider the integral

$$\int_{\mathrm{PGL}_2(k)\backslash \mathrm{PGL}_2(\mathbb{A}_k)} K_{f'}(x,x)E'(x,s)\,dx,$$

but this is easily seen to be equivalent to our present formulation.

Claim. *The functionals (6-3) and (6-4) can naturally be understood as pairings*

$$\mathrm{RTF}^{\mathbb{G}_m\times G,\omega}_{X_1,X_2} : \mathcal{S}(X_1(\mathbb{A}_k))\otimes \mathcal{S}(X_2(\mathbb{A}_k)) \to \mathbb{C}$$

and

$$\mathrm{RTF}^{\mathbb{G}_m\times G',\omega'}_{X'_1,X'_2} : \mathcal{S}(X'_1(\mathbb{A}_k))\otimes \mathcal{S}(X'_2(\mathbb{A}_k)) \to \mathbb{C},$$

respectively, where $X_2 = H\backslash G$, $X'_2 = H'\backslash G'$ and X_1, X'_1 are the affine closures of the varieties $U_F\backslash G$ and $U'_F\backslash G'$, respectively, where U_F and U'_F are maximal unipotent subgroups of H and H' respectively.

The varieties X_1 and X'_1 are considered here as spherical varieties under $\mathbb{G}_m\times G$ (resp. $\mathbb{G}_m\times G'$), where $\mathbb{G}_m = B_2/U_2$, and we extend the $\mathbb{G}_m$-action to the varieties X_2, X'_2 in the trivial way. The exponent ω in $\mathrm{RTF}^{\mathbb{G}_m\times G,\omega}_{X_1,X_2}$ will be explained below.

Before we explain the claim, let us go back to the simpler formula (6-1) and explain how it can be considered as a pairing between $\mathcal{S}(X_1(\mathbb{A}_k))$ and $\mathcal{S}(X_2(\mathbb{A}_k))$ (where $X_i = H_i\backslash G_i$). Here we will identify Hecke algebras with spaces of functions, by choosing Haar measures. Assume that $f = \check{f}_1 \star f_2$ with $f_i \in C_c^\infty(G(\mathbb{A}_k))$. Then we set $\Phi_i(g) = \int_{H_i(\mathbb{A}_k)} f_i(hg)\,dh$. By the definition of $\mathcal{S}(X_i(\mathbb{A}_k))$ when H_i is reductive, it follows that $\Phi_i \in \mathcal{S}(X_i(\mathbb{A}_k))$. (It is at this point that one should add over representatives for $G_i(k)$-orbits on $X_i(k)$, since in general the map $C_c^\infty(G(\mathbb{A}_k)) \to \mathcal{S}(X_i(\mathbb{A}_k))$ is not surjective.) The functional $\mathrm{RTF}^G_{H_1,H_2}(f_1 \star f_2)$ clearly does not depend on f_1 and f_2 but only on Φ_1 and Φ_2. Hence, it defines a G^{diag}-invariant functional

$$\mathcal{S}(X_1(\mathbb{A}_k)) \otimes \mathcal{S}(X_2(\mathbb{A}_k)) \to \mathbb{C}.$$

Now let us return to the setting of the claim, and of equations (6-3) and (6-4). The product $E(h_1, s)\eta(h_1)$ in (6-3) will be considered as an Eisenstein series on $H(k) \setminus H(\mathbb{A}_k)$. We have seen that suitable sections of Eisenstein series can be obtained from integrating X-theta series $\theta_{U_2}^{\mathbb{G}_m \times H}(\Phi, g)$, where $\Phi \in \mathcal{S}(U_2 \setminus H(\mathbb{A}_k))$, against a character ω of $\mathbb{G}_m$. Now consider $\Phi \in \mathcal{S}(U_2 \setminus H(\mathbb{A}_k))$ as a generalized function on $U_2 \setminus G(\mathbb{A}_k)$. Assume again that $f = \check{f}_1 \star f_2 \in C_c^\infty(G(\mathbb{A}_k))$. Then $\Phi_1 := f_1 \star \Phi \in \mathcal{S}(U_2 \setminus H(\mathbb{A}_k))$ and $\Phi_2(g) := \int_{H_2(\mathbb{A}_k)} f(hg)\, dg \in \mathcal{S}(H \setminus G(\mathbb{A}_k))$. Again, of course, we must take many f's and sum over representatives for orbits of $G(k)$ on $X_2(k)$ — incidentally, our point of view explains why there is no need to sum over representatives for orbits in the first variable: because $G(k)$ surjects on $X_1(k)$!

Similarly, one can explain (6-4) as a pairing between $\mathcal{S}(X_1'(\mathbb{A}_k)) \otimes \mathcal{S}(X_2'(\mathbb{A}_k))$, and this completes the explanation of our claim. (We have introduced the exponents ω and ω' in the notation, because we have already integrated against the corresponding character of $\mathbb{G}_m$ in order to form Eisenstein series.) Hence, by viewing the Jacquet–Lai–Rallis trace formulas as being attached to the spaces X_1, X_2 and X_1', X_2' instead of the original $H \setminus G$ and $H' \setminus G'$, the weight factors do not appear as corrections any more, but as a natural part of the setup.

This point of view is very close to the geometric interpretation of the fundamental lemma which led to its proof by Ngô [2010] in the case of the Arthur–Selberg trace formula. Indeed, by the geometric methods of Ngô (see also [Gaitsgory and Nadler 2010]), one naturally gets a hold on the orbital integrals of unramified functions arising from intersection cohomology, not the "naive" ones defined as characteristic functions of $G(\mathfrak{o}_v)$-orbits. I hope that this point of view will lead to a more systematic study of the relative trace formula — at least by alleviating the impression created by weight factors that it is something "less canonical" than the Arthur–Selberg trace formula.

Acknowledgments

This work started in the fall of 2004 during a semester at New York University and was put aside for most of the time since. I am very grateful to Joseph Bernstein, Daniel Bump, Dennis Gaitsgory, David Ginzburg, Hervé Jacquet, David Nadler and Akshay Venkatesh for many useful discussions and encouragement. I also thank the referee for many useful comments.

References

[Benoist and Oh 2007] Y. Benoist and H. Oh, "Polar decomposition for p-adic symmetric spaces", *Int. Math. Res. Not.* **2007**:24 (2007), Art. ID rnm121, 20. MR 2009a:22006

[Braverman and Gaitsgory 2002] A. Braverman and D. Gaitsgory, "Geometric Eisenstein series", *Invent. Math.* **150**:2 (2002), 287–384. MR 2003k:11109 Zbl 1046.11048

[Braverman and Kazhdan 1999] A. Braverman and D. Kazhdan, "On the Schwartz space of the basic affine space", *Selecta Math.* (*N.S.*) **5**:1 (1999), 1–28. MR 2000g:22026 Zbl 0924.22016

[Braverman and Kazhdan 2000] A. Braverman and D. Kazhdan, "γ-functions of representations and lifting", *Geom. Funct. Anal.* Special Volume, Part I (2000), 237–278. MR 2002g:11064 Zbl 1004.11026

[Braverman and Kazhdan 2002] A. Braverman and D. Kazhdan, "Normalized intertwining operators and nilpotent elements in the Langlands dual group", *Mosc. Math. J.* **2**:3 (2002), 533–553. MR 2004f:22016 Zbl 1022.22015

[Braverman et al. 2002] A. Braverman, M. Finkelberg, D. Gaitsgory, and I. Mirković, "Intersection cohomology of Drinfeld's compactifications", *Selecta Math.* (*N.S.*) **8**:3 (2002), 381–418. MR 2003h:14060 Zbl 1031.14019

[Brion et al. 1986] M. Brion, D. Luna, and T. Vust, "Espaces homogènes sphériques", *Invent. Math.* **84**:3 (1986), 617–632. MR 87g:14057 Zbl 0604.14047

[Bump 2005] D. Bump, "The Rankin–Selberg method: an introduction and survey", pp. 41–73 in *Automorphic representations, L-functions and applications: progress and prospects*, edited by J. W. Cogdell et al., Ohio State Univ. Math. Res. Inst. Publ. **11**, de Gruyter, Berlin, 2005. MR 2006k:11097 Zbl 1188.11022

[Bump and Friedberg 1990] D. Bump and S. Friedberg, "The exterior square automorphic L-functions on $\mathrm{GL}(n)$", pp. 47–65 in *Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part II* (Ramat Aviv, 1989), edited by S. Gelbart et al., Israel Math. Conf. Proc. **3**, Weizmann, Jerusalem, 1990. MR 93d:11050 Zbl 0712.11030

[Bump and Ginzburg 1992] D. Bump and D. Ginzburg, "Symmetric square L-functions on $\mathrm{GL}(r)$", *Ann. of Math.* (2) **136**:1 (1992), 137–205. MR 93i:11058 Zbl 0753.11021

[Bump et al. 1999] D. Bump, S. Friedberg, and D. Ginzburg, "Rankin–Selberg integrals in two complex variables", *Math. Ann.* **313**:4 (1999), 731–761. MR 2000a:11078 Zbl 1055.11516

[Cogdell 2003] J. W. Cogdell, "Analytic theory of L-functions for GL_n", pp. 197–228 in *An introduction to the Langlands program* (Jerusalem, 2001), edited by J. Bernstein and S. Gelbart, Birkhäuser, Boston, MA, 2003. MR 1990380 Zbl 1111.11303

[Cogdell et al. 2004] J. W. Cogdell, H. H. Kim, and M. R. Murty, *Lectures on automorphic L-functions*, Fields Institute Monographs **20**, American Mathematical Society, Providence, RI, 2004. MR 2005h:11104 Zbl 1066.11021

[Delorme and Sécherre 2011] P. Delorme and V. Sécherre, "An analogue of the Cartan decomposition for p-adic symmetric spaces of split p-adic reductive groups", *Pacific J. Math.* **251**:1 (2011), 1–21. MR 2012f:22011 Zbl 1220.22003

[Gaitsgory and Nadler 2010] D. Gaitsgory and D. Nadler, "Spherical varieties and Langlands duality", *Mosc. Math. J.* **10**:1 (2010), 65–137, 271. MR 2011h:22018 Zbl 1207.22013

[Garrett 1987] P. B. Garrett, "Decomposition of Eisenstein series: Rankin triple products", *Ann. of Math.* (2) **125**:2 (1987), 209–235. MR 88m:11033 Zbl 0625.10020

[Gelbart et al. 1987] S. Gelbart, I. Piatetski-Shapiro, and S. Rallis, *Explicit constructions of automorphic L-functions*, Lecture Notes in Mathematics **1254**, Springer, Berlin, 1987. MR 89k:11038 Zbl 0612.10022

[Ginzburg and Hundley 2004] D. Ginzburg and J. Hundley, "Multivariable Rankin–Selberg integrals for orthogonal groups", *Int. Math. Res. Not.* **2004**:58 (2004), 3097–3119. MR 2005j:11036 Zbl 1062.22036

[Godement and Jacquet 1972] R. Godement and H. Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Mathematics **260**, Springer, Berlin, 1972. MR 49 #7241 Zbl 0244.12011

[Hochschild and Mostow 1973] G. Hochschild and G. D. Mostow, "Unipotent groups in invariant theory", *Proc. Nat. Acad. Sci. U.S.A.* **70** (1973), 646–648. MR 47 #8713 Zbl 0262.14004

[Ichino and Ikeda 2010] A. Ichino and T. Ikeda, "On the periods of automorphic forms on special orthogonal groups and the Gross–Prasad conjecture", *Geom. Funct. Anal.* **19**:5 (2010), 1378–1425. MR 2011a:11100 Zbl 1216.11057

[Jacquet 2001] H. Jacquet, "Factorization of period integrals", *J. Number Theory* **87**:1 (2001), 109–143. MR 2002a:11050 Zbl 1040.11035

[Jacquet et al. 1993] H. Jacquet, K. F. Lai, and S. Rallis, "A trace formula for symmetric spaces", *Duke Math. J.* **70**:2 (1993), 305–372. MR 94d:11033 Zbl 0795.22008

[Knop 1991] F. Knop, "The Luna–Vust theory of spherical embeddings", pp. 225–249 in *Proceedings of the Hyderabad Conference on Algebraic Groups* (Hyderabad, 1989), Manoj Prakashan, Madras, 1991. MR 92m:14065 Zbl 0812.20023

[Knop 1994a] F. Knop, "The asymptotic behavior of invariant collective motion", *Invent. Math.* **116**:1-3 (1994), 309–328. MR 94m:14063 Zbl 0802.58024

[Knop 1994b] F. Knop, "A Harish-Chandra homomorphism for reductive group actions", *Ann. of Math.* (2) **140**:2 (1994), 253–288. MR 95h:14045 Zbl 0828.22017

[Knop 1995a] F. Knop, "Homogeneous varieties for semisimple groups of rank one", *Compositio Math.* **98**:1 (1995), 77–89. MR 96i:14042 Zbl 0860.14040

[Knop 1995b] F. Knop, "On the set of orbits for a Borel subgroup", *Comment. Math. Helv.* **70**:2 (1995), 285–309. MR 96c:14039 Zbl 0828.22016

[Knop and Van Steirteghem 2006] F. Knop and B. Van Steirteghem, "Classification of smooth affine spherical varieties", *Transform. Groups* **11**:3 (2006), 495–516. MR 2007g:14057 Zbl 1120.14042

[Losev 2008] I. Losev, "Combinatorial invariants of algebraic Hamiltonian actions", *Mosc. Math. J.* **8**:3 (2008), 493–519. MR 2009k:14094 Zbl 1168.14031

[Luna 2001] D. Luna, "Variétés sphériques de type A", *Publ. Math. Inst. Hautes Études Sci.* 94 (2001), 161–226. MR 2003f:14056 Zbl 1085.14039

[Luna and Vust 1983] D. Luna and T. Vust, "Plongements d'espaces homogènes", *Comment. Math. Helv.* **58**:2 (1983), 186–245. MR 85a:14035 Zbl 0545.14010

[LuS 1973] *Sur les groupes algébriques*, Bull. Soc. Math. France Mém. **33**, Société Mathématique de France, Paris, 1973. Contributions de Sivaramakrishna Anantharaman et Domingo Luna, Supplément au Bull. Soc. Math. France Tome 101. MR 47 #6714 Zbl 0688.14040

[Ngô 2010] B. C. Ngô, "Le lemme fondamental pour les algèbres de Lie", *Publ. Math. Inst. Hautes Études Sci.* 111 (2010), 1–169. MR 2011h:22011 Zbl 1200.22011

[Piatetski-Shapiro 1975] I. I. Pjateckij-Šapiro, "Euler subgroups", pp. 597–620 in *Lie groups and their representations* (Budapest, 1971), Halsted, New York, 1975. MR 53 #10720 Zbl 0329.20028

[Piatetski-Shapiro and Rallis 1987] I. Piatetski-Shapiro and S. Rallis, "Rankin triple L functions", *Compositio Math.* **64**:1 (1987), 31–115. MR 89k:11037 Zbl 0637.10023

[Prasad 1990] D. Prasad, "Trilinear forms for representations of GL(2) and local ϵ-factors", *Compositio Math.* **75**:1 (1990), 1–46. MR 91i:22023 Zbl 0731.22013

[Rankin 1939] R. A. Rankin, "Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions, I and II", *Proc. Cambridge Philos. Soc.* **35** (1939), 351–372. MR 1,69d

[Sakellaridis 2008] Y. Sakellaridis, "On the unramified spectrum of spherical varieties over p-adic fields", *Compos. Math.* **144**:4 (2008), 978–1016. MR 2010d:22026 Zbl 1167.22011

[Sakellaridis 2009] Y. Sakellaridis, "Spherical functions on spherical varieties", preprint, 2009. arXiv 0905.4244

[Sakellaridis and Venkatesh 2012] Y. Sakellaridis and A. Venkatesh, "Periods and harmonic analysis on spherical varieties", preprint, 2012. arXiv 1203.0039

[Selberg 1940] A. Selberg, "Bemerkungen über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist", *Arch. Math. Naturvid.* **43** (1940), 47–50. MR 2,88a

[Tate 1967] J. T. Tate, "Fourier analysis in number fields, and Hecke's zeta-functions", pp. 305–347 in *Algebraic Number Theory* (Brighton, UT, 1965), Thompson, Washington, D.C., 1967. MR 36 #121

Communicated by Peter Sarnak
Received 2010-03-31 Revised 2011-07-04 Accepted 2011-08-01

sakellar@rutgers.edu

Department of Mathematics and Computer Science, Rutgers University, 101 Warren Street, Smith Hall 216, Newark, NJ 07102, United States

ALGEBRA AND NUMBER THEORY 6:4(2012)

Nonuniruledness results for spaces of rational curves in hypersurfaces

Roya Beheshti

We prove that the sweeping components of the space of smooth rational curves in a smooth hypersurface of degree d in $\mathbb{P}^n$ are not uniruled if $(n+1)/2 \le d \le n-3$. We also show that for any $e \ge 1$, the space of smooth rational curves of degree e in a general hypersurface of degree d in $\mathbb{P}^n$ is not uniruled roughly when $d \ge e\sqrt{n}$.

1. Introduction

Throughout this paper, we work over an algebraically closed field $\Bbbk$ of characteristic zero. Let X be a smooth hypersurface of degree d in $\mathbb{P}^n$, and for $e \ge 1$, let $R_e(X)$ denote the closure of the open subscheme of $\mathrm{Hilb}_{et+1}(X)$ parametrizing smooth rational curves of degree e in X. It is known that if $d < (n+1)/2$ and X is general, then $R_e(X)$ is an irreducible variety of dimension $e(n+1-d)+n-4$, and it is conjectured that the same holds for general Fano hypersurfaces; see [Harris et al. 2004; Coskun and Starr 2009]. If X is not general, $R_e(X)$ may be reducible. We call an irreducible component R of $R_e(X)$ a *sweeping component* if the curves parametrized by its points sweep out X, or equivalently, if for a general curve C parametrized by R the normal bundle of C in X is globally generated. If $d \le n-1$, or if $d = n$ and $e \ge 2$, then $R_e(X)$ has at least one sweeping component.

In this paper, we study the birational geometry of sweeping components of $R_e(X)$. Recall that a projective variety Y of dimension m is called uniruled if there is a variety Z of dimension $m-1$ and a dominant rational map $Z \times \mathbb{P}^1 \dashrightarrow Y$. We are interested in the following question: for which values of n, d, and e does $R_e(X)$ have nonuniruled sweeping components? Our original motivation for this study comes from the question of whether or not general Fano hypersurfaces of low indices are unirational.

We give a complete answer to the above question when $(n+1)/2 \le d \le n-3$:

Theorem 1.1. *Let X be any smooth hypersurface of degree d in $\mathbb{P}^n$, where $(n+1)/2 \le d \le n-3$. Then for all $e \ge 1$, no sweeping component of $R_e(X)$ is uniruled.*

MSC2010: primary 14J70; secondary 14J40, 14E05.
Keywords: rational curves on hypersurfaces.

We also consider the case $d = n-2$ and prove:

Theorem 1.2. *Let X be a smooth hypersurface of degree $n-2$ in $\mathbb{P}^n$, and let C be a smooth rational curve of degree e in X. Every irreducible sweeping component of $R_e(X)$ which contains C is nonuniruled provided that when we split the normal bundle of C in $\mathbb{P}^n$ as a sum of line bundles*

$$N_{C/\mathbb{P}^n} = \mathbb{O}_C(a_1) \oplus \cdots \oplus \mathbb{O}_C(a_{n-1}),$$

we have $a_i + a_j < 3e$ for every $1 \le i < j \le n-1$.

When $n = 5$ and $d = 3$, $R_e(X)$ is irreducible for any smooth X; see [Coskun and Starr 2009]. J. de Jong and J. Starr [2004] studied the birational geometry of $R_e(X)$ with regards to the question of rationality of general cubic fourfolds. Let $\overline{\mathcal{M}}_{0,0}(X, e)$ be the Kontsevich moduli stack of stable maps of degree e from curves of genus zero to X and $\overline{M}_{0,0}(X, e)$ the corresponding coarse moduli scheme. There is an open subscheme of $\overline{M}_{0,0}(X, e)$ parametrizing smooth rational curves of degree e in X. Presenting a general method to produce differential forms on desingularizations of $\overline{M}_{0,0}(X, e)$, de Jong and Starr prove that if X is a general cubic fourfold, then $R_e(X)$ is not uniruled when $e > 5$ is an odd integer, and the general fibers of the MRC fibration of a desingularization of $R_e(X)$ are at most 1-dimensional when $e > 4$ is an even integer.

If X is a general cubic fourfold, then for a general rational curve C of degree e in X, the normal bundle of C in $\mathbb{P}^5$ is isomorphic to $\mathbb{O}_C((3e-1)/2)^{\oplus 4}$ if $e \ge 5$ is odd and to $\mathbb{O}_C(3e/2)^{\oplus 2} \oplus \mathbb{O}_C((3e/2)-1)^{\oplus 2}$ if $e \ge 6$ is an even integer; see [de Jong and Starr 2004, Proposition 7.1]. Thus Theorem 1.2 gives a new proof of the result of de Jong and Starr when $e \ge 5$ is odd. In Section 4 we study the case when e is an even integer and show:

Theorem 1.3. *Let X be a smooth cubic fourfold, and let C be a general smooth rational curve of degree $e \ge 5$ in X.*

- *$R_e(X)$ is not uniruled if e is odd and $N_{C/\mathbb{P}^5} = \mathbb{O}_C((3e-1)/2)^{\oplus 4}$.*
- *If $\widetilde{R}$ is a desingularization of $R_e(X)$, then the general fibers of the MRC fibration of $\widetilde{R}$ are at most 1-dimensional if e is even and*

$$N_{C/\mathbb{P}^5} = \mathbb{O}_C(3e/2)^{\oplus 2} \oplus \mathbb{O}_C((3e/2)-1)^{\oplus 2}.$$

It is an interesting question whether or not the splitting type of $N_{C/\mathbb{P}^n}$ is always as above for a general rational curve C of degree ≥ 5 in an arbitrary smooth cubic fourfold.

Finally, we consider the case $d < (n+1)/2$. When $d^2 \le n$, $R_e(X)$ is uniruled. In fact, in this range a much stronger statement holds: for every $e \ge 2$, the space of based, 2-pointed rational curves of degree e in X is rationally connected in a suitable sense; see [de Jong and Starr 2006; Starr 2006]. By [Harris et al. 2004], when X is general and $d < (n+1)/2$, $\overline{M}_{0,0}(X,e)$ is irreducible and therefore it is birational to $R_e(X)$. Starr [2003] shows that if $d < \min(n-6, (n+1)/2)$ and $d^2+d \geq 2n+2$, then for every $e \geq 1$, the canonical divisor of $\overline{\mathcal{M}}_{0,0}(X,e)$ is big. This suggests that when $d^2+d \geq 2n+2$ and X is general, $R_e(X)$ may be nonuniruled. In Section 5, we show:

Theorem 1.4. *Let $X \subset \mathbb{P}^n$ $(n \geq 12)$ be a general hypersurface of degree d, and let $m \geq 1$ be an integer. If a general smooth rational curve C in X of degree e is m-normal (that is, if the global sections of $\mathbb{O}_{\mathbb{P}^n}(m)$ map surjectively to those of $\mathbb{O}_{\mathbb{P}^n}(m)|_C$), and if*

$$d^2+(2m+1)d \geq (m+1)(m+2)n+2,$$

then $R_e(X)$ is not uniruled.

In particular, since every smooth curve of degree $e \geq 3$ in $\mathbb{P}^n$ is $(e-2)$-normal, it follows that $R_e(X)$ is not uniruled when X is general and

$$d^2+(2e-3)d \geq e(e-1)n+2.$$

2. A consequence of uniruledness

In this section, we prove a proposition, analogous to the existence of free rational curves on nonsingular uniruled varieties, for varieties whose spaces of smooth rational curves are uniruled. We first fix notation and recall some definitions.

For a morphism $f: Y \to X$ between smooth varieties, by the *normal sheaf of f* we will mean the cokernel of the induced map on the tangent bundles $T_Y \to f^*T_X$.

If Y is an irreducible projective variety, and if $\widetilde{Y}$ is a desingularization of Y, then the maximal rationally connected (MRC) fibration of $\widetilde{Y}$ is a smooth morphism $\pi: Y^0 \to Z$ from an open subset $Y^0 \subset \widetilde{Y}$ such that the fibers of π are all rationally connected, and such that for a very general point $z \in Z$, any rational curve in $\widetilde{Y}$ intersecting $\pi^{-1}(z)$ is contained in $\pi^{-1}(z)$. The MRC fibration of any smooth variety exists and is unique up to birational equivalences [Kollár et al. 1992].

Let Y be an irreducible projective variety, and assume the fiber of the MRC fibration of $\widetilde{Y}$ at a general point is m-dimensional. Then it follows from the definition that there is an irreducible component Z of $\mathrm{Hom}(\mathbb{P}^1, Y)$ such that the map $\mu_1: Z \times \mathbb{P}^1 \to Y$ defined by $\mu_1([g], b) = g(b)$ is dominant and the image of the map $\mu_2: Z \times \mathbb{P}^1 \times \mathbb{P}^1 \to Y \times Y$ defined by $\mu_2([g], b_1, b_2) = (g(b_1), g(b_2))$ has dimension $\geq \dim Y + m$.

Proposition 2.1. *Let $X \subset \mathbb{P}^n$ be a nonsingular projective variety. If an irreducible sweeping component R of $R_e(X)$ is uniruled, then there exist a smooth rational surface S with a dominant morphism $\pi: S \to \mathbb{P}^1$ and a generically finite morphism $f: S \to X$ with the following two properties*:

(i) *If C is a general fiber of π, then $f|_C$ is a closed immersion onto a smooth curve parametrized by a general point of R.*

(ii) *If N_f denotes the normal sheaf of f, then π_*N_f is globally generated.*

Moreover, if the fiber of the MRC fibration of a desingularization of R at a general point is at least m-dimensional, then there are such S and f with the additional property that π_*N_f has an ample subsheaf of rank $= m-1$.

Proof. Let $U \subset R \times X$ be the universal family over R. Since R is uniruled, there exist a quasiprojective variety Z and a dominant morphism $\mu : Z \times \mathbb{P}^1 \to R$. Let $V \subset Z \times \mathbb{P}^1 \times X$ be the pullback of the universal family to $Z \times \mathbb{P}^1$, and denote by $q : V \to Z \times X$ and $p : V \to Z$ the projection maps.

Consider a desingularization $g : \widetilde{V} \to V$, and let $\tilde{q} = q \circ g$ and $\tilde{p} = p \circ g$. Let $z \in Z$ be a general point, and denote the fibers of p and $\tilde{p}$ over z by S and $\widetilde{S}$, respectively. Let $f : S \to X$ be the restriction of q to S, and let $\tilde{f} = f \circ g : \widetilde{S} \to X$. Since z is general, by generic smoothness $\widetilde{S}$ is a smooth surface whose general fiber over $\mathbb{P}^1$ is a smooth connected rational curve. We claim that $\widetilde{S}$ and $\tilde{f}$ satisfy the desired properties. The first property is clearly satisfied.

Since every coherent sheaf on $\mathbb{P}^1$ splits as a torsion sheaf and a direct sum of line bundles, to show that π_*N_f is globally generated it suffices to check that the restriction map $H^0(\mathbb{P}^1, \pi^*N_f) \to N_f|_b$ is surjective for a general point $b \in \mathbb{P}^1$, or equivalently, that the restriction map $H^0(S, N_f) \to H^0(C, N_f|_C)$ is surjective for a general fiber C. To show this, we consider the Kodaira–Spencer map associated to $\widetilde{V}$ at a general point $z \in Z$. Denote by $N_{\tilde{q}}$ the normal sheaf of the map $\tilde{q}$. We get a sequence of maps

$$T_{Z,z} \to H^0(\widetilde{S}, \tilde{p}^*T_Z|_{\widetilde{S}}) \to H^0(\widetilde{S}, \tilde{q}^*T_{X\times Z}|_{\widetilde{S}}) \to H^0(\widetilde{S}, N_{\tilde{q}}|_{\widetilde{S}}).$$

Let b be a general point of $\mathbb{P}^1$. Composing the above map with the projection map $T_{Z\times\mathbb{P}^1,(z,b)} \to T_{Z,z}$, we get a map $T_{Z\times\mathbb{P}^1,(z,b)} \to H^0(\widetilde{S}, N_{\tilde{q}}|_{\widetilde{S}})$. Note that if $N_{\tilde{f}}$ denotes the normal sheaf of $\tilde{f}$, then $N_{\tilde{q}}|_{\widetilde{S}}$ is naturally isomorphic to $N_{\tilde{f}}$. Also, if C is the fiber of $\pi : \widetilde{S} \to \mathbb{P}^1$ over b, then since b is general, C is smooth, and we have a short exact sequence

$$0 \to N_{C/\widetilde{S}} \to N_{\tilde{f}(C)/X} \to N_{\tilde{f}}|_C \to 0.$$

So we get a commutative diagram

$$\begin{tikzcd}
T_{Z\times\mathbb{P}^1,(z,b)} \arrow[r] \arrow[d, "d\mu_{(z,b)}"] & T_{Z,z} \arrow[r] & H^0(\widetilde{S}, N_{\tilde{f}}) \arrow[d] \\
T_{R,[\tilde{f}(C)]} = H^0(\tilde{f}(C), N_{\tilde{f}(C)/X}) \arrow[rr] & & H^0(C, N_{\tilde{f}}|_C)
\end{tikzcd}$$

Since μ is dominant, and since R is sweeping and therefore generically smooth, $d\mu_{(z,b)}$ is surjective. Since the bottom row is also surjective, the map $H^0(\widetilde{S}, N_{\tilde{f}}) \to H^0(C, N_{\tilde{f}}|_C)$ is surjective as well. Thus $\widetilde{\pi}_* N_{\tilde{f}}$ is globally generated.

Suppose now that R is uniruled and that the general fibers of the MRC fibration of R are at least m-dimensional. Let $\dim R = r$. Then there exists a morphism $\mu_1 : Z \times \mathbb{P}^1 \to R$ such that the image of

$$\mu_2 : Z \times \mathbb{P}^1 \times \mathbb{P}^1 \to R \times R, \quad \mu_2(z, b_1, b_2) = (\mu_1(z, b_1), \mu_1(z, b_2))$$

has dimension $\geq r + m$. Constructing $\widetilde{S}$ and $\tilde{f}$ as before, and if C_1 and C_2 denote the fibers of π over general points b_1 and b_2 of $\mathbb{P}^1$, then the image of the map

$$\begin{aligned} d\mu_2 : T_{Z\times\mathbb{P}^1\times\mathbb{P}^1,(z,b_1,b_2)} &\to T_{R\times R,([\tilde{f}(C_1)],[\tilde{f}(C_2)])} \\ &= H^0(C_1, N_{\tilde{f}(C_1)/X}) \oplus H^0(C_2, N_{\tilde{f}(C_2)/X}) \end{aligned}$$

is at least $(r+m)$-dimensional. The desired result now follows from the following commutative diagram and the observation that the kernel of the bottom row is 2-dimensional:

$$\begin{tikzcd} T_{Z\times\mathbb{P}^1\times\mathbb{P}^1,(z,b_1,b_2)} \arrow[r] \arrow[d, "(d\mu_2)_{(z,b_1,b_2)}"] & T_{Z,z} \arrow[r] & H^0(\widetilde{S}, N_{\tilde{f}}) \arrow[d] \\ T_{R\times R,([\tilde{f}(C_1)],[\tilde{f}(C_2)])} \arrow[rr] & & H^0(C_1, N_{\tilde{f}}|_{C_1}) \oplus H^0(C_2, N_{\tilde{f}}|_{C_2}) \end{tikzcd} \qquad \square$$

Proposition 2.1 will be enough for the proof of Theorem 1.1, but to prove Theorem 1.3 in the even case, we will need a slightly stronger variant. Let $f : Y \to X$ be a morphism between smooth varieties, and let N_f be the normal sheaf of f

$$0 \to T_Y \to f^* T_X \to N_f \to 0.$$

Suppose there is a dominant map $\pi : Y \to \mathbb{P}^1$, and let M be the image of the map induced by π on the tangent bundles $T_Y \to \pi^* T_{\mathbb{P}^1}$. Consider the push-out of the above sequence by the map $T_Y \to M$:

$$\begin{tikzcd} 0 \arrow[r] & T_Y \arrow[r] \arrow[d] & f^*T_X \arrow[r] \arrow[d] & N_f \arrow[r] \arrow[d, "="] & 0 \\ 0 \arrow[r] & M \arrow[r] \arrow[d] & N_{f,\pi} \arrow[r] \arrow[d] & N_f \arrow[r] & 0 \\ & 0 & 0 & & \end{tikzcd}$$

The sheaf $N_{f,\pi}$ in the above diagram will be referred to as the *normal sheaf of f relative to π*.

Property (ii) of Proposition 2.1 says that $H^0(S, N_f) \to H^0(C, N_f|_C)$ is surjective. An argument parallel to the proof of Proposition 2.1 shows the following:

Proposition 2.2. *Let X be as in Proposition 2.1. Then property (ii) can be strengthened as follows*:

(ii′) *If N_f denotes the normal sheaf of f and $N_{f,\pi}$ denotes the normal sheaf of f relative to π, then the composition of the maps*

$$H^0(S, N_{f,\pi}) \to H^0(C, N_{f,\pi}|_C) \to H^0(C, N_f|_C)$$

is surjective for a general fiber C of π.

Moreover, if the general fibers of the MRC fibration of a desingularization of R are at least m-dimensional, then there are S and f with properties (i) and (ii′) such that the image of the map

$$H^0(S, N_{f,\pi} \otimes I_C) \to H^0(C, (N_f \otimes I_C)|_C)$$

is at least $(m-1)$-dimensional.

3. The case when $\frac{n+1}{2} \le d$

Let X be a smooth hypersurface of degree d in $\mathbb{P}^n$. Assume that a sweeping component R of $R_e(X)$ is uniruled. The following result, along with Proposition 2.1, will prove Theorem 1.1.

Proposition 3.1. *Suppose $d \le n-3$, and let S and f be as in Proposition 2.1. If C is a general fiber of $\pi : S \to \mathbb{P}^1$ and I_C is the ideal sheaf of C in S, then the restriction map*

$$H^0(S, f^*\mathcal{O}_X(2d-n-1) \otimes I_C^\vee) \to H^0(C, f^*\mathcal{O}_X(2d-n-1) \otimes I_C^\vee|_C)$$

is zero.

Proof of Theorem 1.1. Granting Proposition 3.1, since

$$H^0(S, f^*\mathcal{O}_X(2d-n-1) \otimes I_C^\vee) \to H^0(C, f^*\mathcal{O}_X(2d-n-1) \otimes I_C^\vee|_C)$$

is the zero map, we have

$$H^0(S, f^*\mathcal{O}_X(2d-n-1)) = H^0(S, f^*\mathcal{O}_X(2d-n-1) \otimes I_C^\vee).$$

Thus,

$$\begin{aligned} H^0(\mathbb{P}^1, \pi_* f^*\mathcal{O}_X(2d-n-1)) &= H^0(\mathbb{P}^1, \pi_*(f^*\mathcal{O}_X(2d-n-1) \otimes I_C^\vee)) \\ &= H^0(\mathbb{P}^1, (\pi_* f^*\mathcal{O}_X(2d-n-1)) \otimes \mathcal{O}_{\mathbb{P}^1}(1)), \end{aligned}$$

which is only possible if $H^0(\mathbb{P}^1, \pi_* f^*\mathcal{O}_X(2d-n-1))$ vanishes. So we have $H^0(S, f^*\mathcal{O}_X(2d-n-1)) = 0$ and $d < (n+1)/2$. □

Proof of Proposition 3.1.

Let ω_S be the canonical sheaf of S. By Serre duality and the long exact sequence of cohomology, it suffices to show that if S and f satisfy the properties of Proposition 2.1, then the restriction map

$$H^1(S, f^*\mathcal{O}_X(n+1-2d)\otimes\omega_S) \to H^1(C, f^*\mathcal{O}_X(n+1-2d)\otimes\omega_S|_C)$$

is surjective. Let N be the normal sheaf of the map $f : S \to X$, and let N' be the normal sheaf of the map $S \to \mathbb{P}^n$.

There is a short exact sequence

$$0 \to N \to N' \to f^*\mathcal{O}_X(d) \to 0. \tag{1}$$

Taking the $(n-3)$-rd exterior power of this sequence, we get the exact sequence

$$0 \to \bigwedge^{n-3} N \otimes f^*\mathcal{O}_X(-d) \to \bigwedge^{n-3} N' \otimes f^*\mathcal{O}_X(-d) \to \bigwedge^{n-4} N \to 0.$$

For an exact sequence of sheaves of $\mathcal{O}_S$-modules $0 \to E \to F \to M \to 0$ with E and F locally free of ranks e and f, there is a natural map of sheaves

$$\bigwedge^{f-e-1} M \otimes \bigwedge^{e} E \otimes (\bigwedge^{f} F)^\vee \to M^\vee$$

which is defined locally at a point $s \in S$ as follows: assume $\gamma_1, \dots, \gamma_{f-e-1} \in M_s$, $\alpha_1, \dots, \alpha_e \in E_s$, and $\phi : \bigwedge^f F_s \to O_{S,s}$; then for $\gamma \in M_s$, we set $\gamma_{f-e} = \gamma$, and we define the map to be $\gamma \mapsto \phi(\tilde{\gamma}_1 \wedge \tilde{\gamma}_2 \wedge \cdots \wedge \tilde{\gamma}_{f-e} \wedge \alpha_1 \wedge \cdots \wedge \alpha_e)$, where $\tilde{\gamma}_i$ is any lifting of γ_i in F_s. Clearly, this map does not depend on the choice of the liftings, and thus it is defined globally. So from the short exact sequence $0 \to T_S \to f^*T_X \to N \to 0$, we get a map

$$\bigwedge^{n-4} N \to N^\vee \otimes f^*\mathcal{O}_X(n+1-d)\otimes\omega_S,$$

and from the short exact sequence $0 \to T_S \to f^*T_{\mathbb{P}^n} \to N' \to 0$, we get a map

$$\bigwedge^{n-3} N' \otimes f^*O_X(-d) \to (N')^\vee \otimes f^*O_X(n+1)\otimes\omega_S.$$

With the choices of the maps we have made, the following diagram, whose bottom row is obtained from dualizing sequence (1) and tensoring with

$$f^*\mathcal{O}_X(n+1-2d)\otimes\omega_S,$$

is commutative with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & \bigwedge^{n-3} N\otimes f^*\mathcal{O}_X(-d) \arrow[r] \arrow[d] & \bigwedge^{n-3} N'\otimes f^*O_X(-d) \arrow[r] \arrow[d] & \bigwedge^{n-4} N \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & f^*\mathcal{O}_X(n+1-2d)\otimes\omega_S \arrow[r] & (N')^\vee\otimes f^*\mathcal{O}_X(n+1-d)\otimes\omega_S \arrow[r] & N^\vee\otimes f^*\mathcal{O}_X(n+1-d)\otimes\omega_S \arrow[r] & 0
\end{tikzcd}$$

Since the cokernel of the first vertical map restricted to C is a torsion sheaf, to show the assertion it suffices to show that the map

$$H^1(S, \bigwedge^{n-3} N \otimes f^*\mathcal{O}_X(-d)) \to H^1(C, \bigwedge^{n-3} N \otimes f^*\mathcal{O}_X(-d)|_C)$$

is surjective. Applying the long exact sequence of cohomology to the top sequence, the surjectivity assertion follows if we show that

(1) $H^0(S, \bigwedge^{n-4} N) \to H^0(C, \bigwedge^{n-4} N|_C)$ is surjective,

(2) $H^1(C, \bigwedge^{n-3} N' \otimes f^*\mathcal{O}_X(-d)|_C) = 0$.

To prove (1), we consider the following commutative diagram:

$$\begin{array}{ccc} \bigwedge^{n-4} H^0(S, N) & \longrightarrow & \bigwedge^{n-4} H^0(C, N|_C) \\ \downarrow & & \downarrow \\ H^0(S, \bigwedge^{n-4} N) & \longrightarrow & H^0(C, \bigwedge^{n-4} N|_C) \end{array}$$

The top horizontal map is surjective since $H^0(S, N) \to H^0(C, N|_C)$ is surjective, and the right vertical map is surjective since $N|_C$ is a globally generated line bundle over $\mathbb{P}^1$. By commutativity of the diagram the bottom horizontal map is surjective.

To prove (2), note that there is a surjective map $f^*\mathcal{O}_{\mathbb{P}^n}(1)^{\oplus n+1} \to N'$. Taking the $(n-3)$-rd exterior power, and then tensoring with $f^*\mathcal{O}_X(-d)$, we get a surjective map

$$f^*\mathcal{O}_{\mathbb{P}^n}(n-3-d)^{\oplus \binom{n+1}{n-3}} \to \bigwedge^{n-3} N' \otimes f^*\mathcal{O}_X(-d).$$

Restricting to C, since $n-3-d \geq 0$, we have

$$H^1(C, \bigwedge^{n-3} N' \otimes f^*\mathcal{O}_X(-d)|_C) = 0. \qquad \square$$

Proof of Theorem 1.2. Suppose that X is a smooth hypersurface of degree $n-2$ in $\mathbb{P}^n$. Let C be a smooth rational curve of degree e in $\mathbb{P}^n$ whose normal bundle $N_{C/\mathbb{P}^n}$ is globally generated. If we write

$$N_{C/\mathbb{P}^n} = \mathcal{O}_C(a_1) \oplus \cdots \oplus \mathcal{O}_C(a_{n-1}),$$

then $\sum_{1 \leq i \leq n-1} a_i = e(n+1)-2$. Assume that $a_i + a_j < 3e$ for every $1 \leq i < j \leq n-1$.

Then $H^1(C, \bigwedge^{n-3} N_{C/\mathbb{P}^n} \otimes \mathcal{O}_{\mathbb{P}^n}(-d)|_C) = 0$, and so if N' is as in the proof of Theorem 1.1, then

$$H^1(C, \bigwedge^{n-3} N' \otimes f^*\mathcal{O}_X(-d)|_C) = 0.$$

The assertion now follows from the proof of Theorem 1.1. $\square$

We remark that when $d=n-1$ or n, the uniruledness of the sweeping subvarieties of $R_e(X)$ has been studied in [Beheshti and Starr 2008]. It is proved that if $e \leq n$, then a subvariety of $R_e(X)$ is nonuniruled if the curves parametrized by its points sweep out X or a divisor in X.

4. Cubic fourfolds

In this section we prove Theorem 1.3. When $e \geq 5$ is odd, the theorem follows from Theorem 1.2 and [de Jong and Starr 2004, Proposition 7.1].

So let $e \geq 6$ be an even integer, and assume to the contrary that the general fibers of the MRC fibration of $R_e(X)$ are at least 2-dimensional. Let S and f be as in Proposition 2.2, and let C be a general fiber of π. Set $N = N_f$ and $Q = N_{f,\pi}$. Then by Proposition 2.1 the following properties are satisfied:

- Property (i): The composition of the maps

  $$H^0(S, Q) \to H^0(S, Q|_C) \to H^0(C, N|_C)$$

  is surjective.
- Property (ii): The composition of the maps

  $$H^0(S, Q \otimes I_C) \to H^0(C, Q \otimes I_C|_C) \to H^0(C, N \otimes I_C|_C)$$

  is nonzero.

We show these lead to a contradiction. Note that $I_C|_C$ is isomorphic to the trivial bundle $\mathcal{O}_C$, but we write $I_C|_C$ instead of $\mathcal{O}_C$ to keep track of various maps and exact sequences involved in the proof.

Let Q' be the normal sheaf of the map $S \to \mathbb{P}^5$ relative to π. We have $Q|_C = N_{C/X}$ and $Q'|_C = N_{C/\mathbb{P}^5}$. Since $N_{X/\mathbb{P}^5} = \mathcal{O}_X(3)$, there is a short exact sequence

$$0 \to Q \to Q' \to f^*\mathcal{O}_X(3) \to 0. \tag{2}$$

Taking exterior powers, we obtain the short exact sequence

$$0 \to \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3) \to \bigwedge^2 Q' \otimes f^*\mathcal{O}_X(-3) \to Q \to 0. \tag{3}$$

Since this sequence splits locally, its restriction to C is also a short exact sequence

$$0 \to \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C \to \bigwedge^2 Q' \otimes f^*\mathcal{O}_X(-3)|_C \to Q|_C \to 0. \tag{4}$$

To get a contradiction, we show that the image of the boundary map

$$\gamma : H^0(C, Q|_C) \to H^1(C, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C)$$

is of codimension at least 2 in $H^1(C, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C)$. This is not possible since by our assumption $N_{C/\mathbb{P}^5} = \mathcal{O}_C(3e/2)^{\oplus 2} \oplus \mathcal{O}_C((3e/2)-1)^{\oplus 2}$, and so

$$\begin{aligned} H^1(C, \textstyle\bigwedge^2 Q' \otimes f^*\mathcal{O}_X(-3)|_C) &= H^1(C, \textstyle\bigwedge^2 N_{C/\mathbb{P}^5} \otimes f^*\mathcal{O}_X(-3)|_C) \\ &= H^1(C, \mathcal{O}_C(-2) \oplus \mathcal{O}_C(-1)^{\oplus 4} \oplus \mathcal{O}_C) \\ &= \mathbf{k}. \end{aligned}$$

Lemma 4.1. *The kernel of the map $f^*T_X \to Q$ is a line bundle which contains $\bigwedge^2 T_S \otimes \pi^*\omega_{\mathbb{P}^1}$ as a subsheaf.*

Proof. The kernel of $f^*T_X \to Q$ is equal to the kernel of the map induced by π on the tangent bundles $T_S \to \pi^*T_{\mathbb{P}^1}$, which we denote by F:

$$0 \to F \to T_S \to \pi^*T_{\mathbb{P}^1}.$$

Since F is reflexive, it is locally free on S, and it is clearly of rank 1. Also, the composition of the maps

$$\textstyle\bigwedge^2 T_S \otimes \pi^*\omega_{\mathbb{P}^1} \to \bigwedge^2 T_S \otimes \Omega_S = T_S \to \pi^*T_{\mathbb{P}^1}$$

is the zero map. So $\bigwedge^2 T_S \otimes \pi^*\omega_{\mathbb{P}^1}$ is a subsheaf of F. □

Given a section $r \in H^0(C, Q \otimes I_C|_C)$, we can define a map

$$\beta_r : H^1(C, \textstyle\bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C) \longrightarrow H^1(C, \omega_S|_C) = \mathbb{k}$$

as follows. Let F be the line bundle from the proof of Lemma 4.1. It follows from the proof of the lemma that there is an injection $\bigwedge^2 T_S \otimes \pi^*\omega_{\mathbb{P}^1} \to F$, and from the short exact sequence

$$0 \to F \to f^*T_X \to Q \to 0$$

we get a generically injective map of sheaves

$$\textstyle\bigwedge^3 Q \otimes F \to \bigwedge^4 f^*T_X.$$

Combining these, we get a morphism

$$\textstyle\bigwedge^3 Q \otimes (\omega_S \otimes \pi^*T_{\mathbb{P}^1})^\vee \to \bigwedge^4 f^*T_X.$$

Since $\bigwedge^4 f^*T_X = f^*\mathcal{O}_X(3)$, we get a generically injective map

$$\Psi : \textstyle\bigwedge^3 Q \otimes f^*\mathcal{O}_X(-3) \otimes I_C \to \omega_S \otimes \pi^*T_{\mathbb{P}^1} \otimes I_C,$$

and by restricting to C, we get a map

$$\Psi|_C : \big(\textstyle\bigwedge^3 Q \otimes f^*\mathcal{O}_X(-3) \otimes I_C\big)|_C \to \omega_S|_C.$$

Finally, r gives a map

$$\Phi_r : \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C \xrightarrow{\wedge r} \bigwedge^3 Q \otimes f^*\mathcal{O}_X(-3) \otimes I_C|_C,$$

and we define β_r to be the map induced by the composition $\Psi|_C \circ \Phi_r$. Note that β_r is nonzero if $r \neq 0$.

Lemma 4.2. *For $r, r' \in H^0(C, Q \otimes I_C|_C)$, $\ker(\beta_r) = \ker(\beta_{r'})$ if and only if r and r' are scalar multiples of each other.*

Proof. By Serre duality, it is enough to show that the images of the maps

$$H^0(C, I_C^\vee|_C) = H^0(C, \omega_S^\vee|_C \otimes \omega_C) \underset{\beta_{r'}^\vee}{\overset{\beta_r^\vee}{\rightrightarrows}} H^0\big(C, \big(\bigwedge^2 Q^\vee \otimes f^*\mathcal{O}_X(3)\big)|_C \otimes \omega_C\big)$$

are the same if and only if r and r' are scalar multiples of each other. Since $Q|_C = N_{C/X}$, we have $\bigwedge^3 Q|_C = \bigwedge^3 N_{C/X} = f^*\mathcal{O}_X(3) \otimes \omega_C$, so

$$\big(\bigwedge^2 Q^\vee \otimes f^*\mathcal{O}_X(3)\big)|_C \otimes \omega_C = Q|_C,$$

and the map

$$\beta_r^\vee : H^0(C, I_C^\vee|_C) \to H^0(C, Q|_C)$$

is simply given by r. Similarly, $\beta_{r'}^\vee$ is given by r', and the lemma follows. □

Recall that by definition, we have a short exact sequence

$$0 \to \pi^* T_{\mathbb{P}^1}|_C \to Q|_C \to N|_C \to 0,$$

and $\pi^* T_{\mathbb{P}^1}|_C = I_C^{-1}|_C$. If we tensor this sequence with $I_C|_C$, we get the short exact sequence

$$0 \to \mathcal{O}_C \to Q \otimes I_C|_C \to N \otimes I_C|_C \to 0.$$

Let i be a nonzero section in the image of $H^0(C, \mathcal{O}_C) \to H^0(C, Q \otimes I_C|_C)$. Then i induces a map

$$\beta_i : H^1(C, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C) \longrightarrow H^1(C, \omega_S|_C) = \Bbbk$$

as described before. Let

$$\gamma : H^0(C, Q|_C) \to H^1(C, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C)$$

be the connecting map in sequence (4).

Lemma 4.3. *We have* $\mathrm{image}(\gamma) \subset \ker \beta_i$.

Proof. Since the short exact sequence $0 \to N \to N' \to f^*\mathcal{O}_X(3) \to 0$ splits locally, there is an exact sequence

$$0 \to \bigwedge^2 N \otimes f^*\mathcal{O}_X(-3) \to \bigwedge^2 N' \otimes f^*\mathcal{O}_X(-3) \to N \to 0.$$

Applying the long exact sequence of cohomology to the restriction of this sequence to C, we get a map

$$H^0(C, N|_C) \to H^1(C, \bigwedge^2 N \otimes f^*\mathcal{O}_X(-3)|_C).$$

Also from the exact sequence $0 \to T_S \to f^*T_X \to N \to 0$, we get a map

$$\bigwedge^2 T_S \otimes \bigwedge^2 N \to \bigwedge^4 f^*T_X = f^*\mathcal{O}_X(3),$$

and hence a map

$$\bigwedge^2 N \otimes f^*\mathcal{O}_X(-3) \to \omega_S.$$

It follows from the definition of β_i that the map $\beta_i \circ \gamma$ factors through

$$H^0(C, Q|_C) \to H^0(C, N|_C) \to H^1(C, \bigwedge^2 N \otimes f^*\mathcal{O}_X(-3)|_C) \to H^1(C, \omega_S|_C),$$

so we have a commutative diagram

$$\begin{tikzcd}
 & H^0(S, N) \arrow[r] \arrow[d] & H^1(S, \bigwedge^2 N \otimes f^*\mathcal{O}_X(-3)) \arrow[r] & H^1(S, \omega_S) = 0 \arrow[d] \\
H^0(C, Q|_C) \arrow[r] & H^0(C, N|_C) \arrow[rr] & & H^1(C, \omega_S|_C)
\end{tikzcd}$$

Thus we can conclude the assertion by the fact that the restriction map $H^0(S, N) \to H^0(C, N|_C)$ is surjective, and so the image of the composition of the above maps is contained in the image of the restriction map $H^1(S, \omega_S) \to H^1(C, \omega_S|_C)$, which is zero. $\square$

In the following lemma we prove a similar result for the sections of $Q \otimes I_C|_C$ which are obtained by restricting the global sections of $Q \otimes I_C$ to C.

Lemma 4.4. *If* $\tilde{r} \in H^0(S, Q \otimes I_C)$, *and if* $r = \tilde{r}|_C$, *then* $\operatorname{image}(\gamma) \subset \ker(\beta_r)$.

Proof. We have a commutative diagram

$$\begin{tikzcd}
H^0(S, Q) \arrow[r] \arrow[d] & H^1(S, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)) \arrow[r] \arrow[d] & H^1(S, \omega_S) = 0 \arrow[d] \\
H^0(C, Q|_C) \arrow[r, "\gamma"] & H^1(C, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)) \arrow[r, "\beta_r"] & H^1(C, \omega_S|_C)
\end{tikzcd}$$

Therefore we have $\beta_r(\gamma(u)) = 0$ for any $u \in H^0(C, Q|_C)$ in the image of the restriction map $H^0(S, Q) \to H^0(C, Q|_C)$. Consider the exact sequence

$$0 \to I_C^{-1}|_C \to Q|_C \to N|_C \to 0.$$

From the hypothesis that the composition map

$$H^0(S, Q) \to H^0(C, Q|_C) \to H^0(C, N|_C)$$

is surjective, we see that to prove the statement it is enough to show that for any nonzero u in the image of $H^0(C, I_C^{-1}|_C) \to H^0(C, Q|_C)$, we have $\gamma(u) \in \ker \beta_r$.

Consider the following diagram, where λ is obtained from applying the long exact sequence of cohomology to the third wedge power of sequence (2), and ψ is induced by the map $\Psi|_C$:

$$\begin{tikzcd}
H^0(C, Q|_C) \arrow[r, "\gamma"] \arrow[d, bend right, "\wedge i"'] \arrow[d, bend left, "\wedge r"] & H^1(C, \bigwedge^2 Q \otimes f^*\mathcal{O}_X(-3)|_C) \arrow[d, bend right, "\wedge i"'] \arrow[d, bend left, "\wedge r"] \arrow[dr, "\beta_i"] \arrow[dr, bend left, "\beta_r"] & \\
H^0(C, \bigwedge^2 Q \otimes I_C|_C) \arrow[r, "\lambda"'] & H^1(C, \bigwedge^3 Q \otimes f^*\mathcal{O}_X(-3) \otimes I_C|_C) \arrow[r, "\psi"'] & H^1(C, \omega_S|_C)
\end{tikzcd}$$

Then we have

$$\begin{aligned}
\beta_r \circ \gamma(u) &= \psi \circ \lambda(u \wedge r) \\
&= \psi \circ \lambda(r \wedge i) \quad \text{(up to a scalar factor)} \\
&= \beta_i \circ \gamma(r) \\
&= 0,
\end{aligned}$$

where the last equality comes from $\gamma(H^0(C, Q|_C)) \subset \ker \beta_i$, by Lemma 4.3. □

Now, let $\tilde{r}_0 \in H^0(S, Q \otimes I_C)$ be so that its image in $H^0(C, N \otimes I_C|_C)$ is nonzero. Such an $\tilde{r}_0$ exists by Property (ii). Then $r_0 := \tilde{r}_0|_C$ defines a map β_{r_0}. Since the image of r_0 in $H^0(C, N \otimes I_C|_C)$ is nonzero, r_0 and i are not scalar multiples, so according to Lemma 4.2, $\ker \beta_{r_0} \neq \ker \beta_i$. Thus the codimension of $\ker \beta_i \cap \ker \beta_{r_0}$ is at least 2. On the other hand, by the previous lemmas, image$(\gamma) \subset \ker \beta_i \cap \ker \beta_{r_0}$. This is a contradiction since $\dim H^1(C, \bigwedge^2 Q' \otimes f^*\mathcal{O}_X(-3)|_C) = 1$.

5. The case when $d < \frac{n+1}{2}$

Throughout this section, $X \subset \mathbb{P}^n$ is a general hypersurface of degree $d < (n+1)/2$. By the main theorem of [Harris et al. 2004], $R_e(X)$ is irreducible for every $e \geq 1$. If $d^2 \leq n$ and $e \geq 2$, then by [de Jong and Starr 2006; Starr 2006], the space of rational curves of degree e in X passing through two general points of X is rationally connected. In particular, $R_e(X)$ is rationally connected for $e \geq 2$. If $e = 1$, then $R_1(X)$ is the Fano variety of lines in X which is rationally connected if and only if

$d^2+d \le 2n$ [Kollár 1996, V.4.7]. In this section, we will consider the case when $d^2+d > 2n$.

Assume that $R_e(X)$ is uniruled. Then there are S and f with the two properties given in Proposition 2.1. We can take the pair (S, f) to be minimal in the sense that a component of a fiber of π which is contracted by f cannot be blown down. Let N be the normal sheaf of f, and let C be a general fiber of π with ideal sheaf I_C in S. Denote by H the pullback of a hyperplane in $\mathbb{P}^n$ to S, and denote by K a canonical divisor on S. From the exact sequences $0 \to T_S \to f^*T_X \to N \to 0$ and $0 \to f^*T_X \to f^*T_{\mathbb{P}^n} \to f^*\mathbb{O}_{\mathbb{P}^n}(d) \to 0$, we get

$$\begin{aligned}
&\chi(N \otimes I_C)\\
&\quad = (n+1)\chi(f^*\mathbb{O}_{\mathbb{P}^n}(1) \otimes I_C) - \chi(f^*\mathbb{O}_{\mathbb{P}^n}(d) \otimes I_C) - \chi(I_C) - \chi(T_S \otimes I_C)\\
&\quad = (n+1)\Big(\frac{(H-C)\cdot(H-C-K)}{2}+1\Big) - \frac{(dH-C)\cdot(dH-C-K)}{2} - 1\\
&\qquad\qquad - \frac{-C\cdot(-C-K)}{2} - 1 - (2K^2-14)\\
&\quad = \frac{(n+1-d^2)}{2}H^2 - \frac{(n+1-d)}{2}H\cdot K - 2K^2 - (n+1-d)e + 14.
\end{aligned}$$

We claim that $2H+2C+K$ is base-point free and hence has a nonnegative self-intersection number. By the main theorem of [Reider 1988], if $2H+2C+K$ is not base-point free, then there exists an effective divisor E such that either

$$(2H+2C)\cdot E = 1,\, E^2 = 0 \quad \text{or} \quad (2H+2C)\cdot E = 0,\, E^2 = -1.$$

The first case is clearly not possible. In the second case, $H\cdot E = 0$, and $C \cdot E = 0$. So E is a component of one of the fibers of π which is contracted by f and which is a (-1)-curve. This contradicts the assumption that (S, f) is minimal. Thus $(2H+2C+K)^2 \ge 0$. Also, since $H^1(S, f^*\mathbb{O}_X(-1)) = 0$, we have

$$H\cdot(H+K) = 2\chi(f^*\mathbb{O}_X(-1)) - 2 \ge -2,$$

so we can write

$$\begin{aligned}
\chi(N \otimes I_C) &= \frac{2n+2-d^2-d}{2}H^2 - (n-d-15)(e-1) - 2 - 2(2H+2C+K)^2\\
&\qquad - \frac{n-d-15}{2}(H\cdot(H+K)+2)\\
&\le \frac{2n+2-d^2-d}{2}H^2 - (n-d-15)(e-1) - 2,
\end{aligned}$$

and therefore $\chi(N\otimes I_C)$ is negative when $d^2+d \ge 2n+2$ and $n \ge 30$.

The Leray spectral sequence gives a short exact sequence

$$0 \to H^1(\mathbb{P}^1, \pi_*(N\otimes I_C)) \to H^1(S, N\otimes I_C) \to H^0(\mathbb{P}^1, R^1\pi_*(N\otimes I_C)) \to 0,$$

and by our assumption on S and f, $H^1(\mathbb{P}^1, \pi_*(N \otimes I_C)) = 0$. If we could choose S such that $H^0(\mathbb{P}^1, R^1\pi_*(N \otimes I_C)) = 0$, then we could conclude that $\chi(N \otimes I_C) \geq 0$, and hence $R_e(X)$ could not be uniruled for $d^2 + d \geq 2n + 2$ and $n \geq 30$.

We cannot show that for a general X, a minimal pair (S, f) as in Proposition 2.1 can be chosen so that $H^0(\mathbb{P}^1, R^1\pi_*(N \otimes I_C)) = 0$. However, we prove that if X is general and (S, f) is minimal, then for every $t \geq 1$,

$$H^0(\mathbb{P}^1, R^1\pi_*(N \otimes I_C \otimes f^*\mathcal{O}_X(t))) = 0.$$

We also show that if $t \geq 0$ and $f(C)$ is t-normal, then

$$H^1(\mathbb{P}^1, \pi_*(N \otimes I_C \otimes f^*\mathcal{O}_X(t))) = 0.$$

These imply that $\chi(N \otimes I_C \otimes f^*\mathcal{O}_X(t))$ is nonnegative when X is general and $f(C)$ is t-normal. To finish the proof of Theorem 1.4, we compute $\chi(N \otimes I_C \otimes f^*\mathcal{O}_X(t))$ directly and show that it is negative when the inequality in the statement of the theorem holds.

Proof of Theorem 1.4. Let X be a general hypersurface of degree d in $\mathbb{P}^n$. If $R_e(X)$ is uniruled, then there are S and f as in Proposition 2.1. Assume the pair (S, f) is minimal. Let N be the normal sheaf of f, and let C be a general fiber of π. Then $H^0(S, N) \to H^0(C, N|_C)$ is surjective. The restriction map $H^0(S, f^*\mathcal{O}_X(m)) \to H^0(C, f^*\mathcal{O}_X(m)|_C)$ is also surjective since $f(C)$ is m-normal, so the restriction map $H^0(S, N \otimes f^*\mathcal{O}_X(m)) \to H^0(C, N \otimes f^*\mathcal{O}_X(m)|_C)$ is surjective as well. Therefore,

$$H^1(\mathbb{P}^1, \pi_*(N \otimes f^*\mathcal{O}_X(m) \otimes I_C)) = 0.$$

Now let C be an arbitrary fiber of π, and let C^0 be an irreducible component of C. Then by Proposition 5.2, $f^*(T_X(t))|_{C^0}$ is globally generated for every $t \geq 1$, and hence $N \otimes f^*\mathcal{O}_X(t)|_{C^0}$ is globally generated too. So Lemma 5.1 shows that for every $t \geq 1$,

$$H^0(\mathbb{P}^1, R^1\pi_*(N \otimes f^*\mathcal{O}_X(t) \otimes I_C)) = 0.$$

By the Leray spectral sequence,

$$\begin{aligned} &H^1(S, N \otimes f^*\mathcal{O}_X(m) \otimes I_C) \\ &\quad = H^1(\mathbb{P}^1, \pi_*(N \otimes f^*\mathcal{O}_X(m) \otimes I_C)) \oplus H^0(\mathbb{P}^1, R^1\pi_*(N \otimes f^*\mathcal{O}_X(m) \otimes I_C)) \\ &\quad = 0, \end{aligned}$$

and therefore, $\chi(N \otimes f^*\mathcal{O}_X(m) \otimes I_C) \geq 0$. We next compute $\chi(N \otimes f^*\mathcal{O}_X(m) \otimes I_C)$. For an integer $t \geq 0$, set

$$a_t = \chi(N \otimes I_C \otimes f^*\mathcal{O}_X(t)).$$

We have

$$a_t = \chi(N \otimes I_C) + \frac{2t(n+1-d)+t^2(n-3)}{2} H^2 - \frac{t(n-5)}{2} H \cdot K - t(n-3)e.$$

So

$$a_t = \frac{b_t}{2}\, H^2 + \frac{c_t}{2}\, H \cdot K - 2K^2 + d_t,$$

where

$$\begin{aligned} b_t &= (n+1-d^2) + 2t(n+1-d) + t^2(n-3),\\ c_t &= -(n+1-d) - t(n-5),\\ \text{and}\quad d_t &= -t(n-3)e - (n+1-d)e + 14. \end{aligned}$$

A computation similar to the computation in the beginning of this section shows that

$$\begin{aligned} a_t &= \frac{b_t - c_t}{2} H^2 - 2(2H + 2C + K)^2 + \frac{c_t + 16}{2}(H \cdot (H + K) + 2)\\ &\qquad\qquad + (d_t - c_t - 32 + 16e)\\ &\le \frac{b_t - c_t}{2} H^2 + (d_t - c_t - 32 + 16e). \end{aligned}$$

Since $d_t - c_t - 32 + 16e = -(e-1)(n - 15 - d + t(n-3)) - 2t - 2$, and since $n - 15 - d + t(n-3) \ge 2n - d - 18 \ge 0$ for $t \ge 1$ and $n \ge 12$, we get

$$a_t < \frac{b_t - c_t}{2} H^2.$$

When $d^2 + (2t+1)d \ge (t+1)(t+2)n + 2$, $b_t < c_t$, and so $a_t < 0$. If we let $t = m$, we get the desired result. □

Lemma 5.1. *If E is a locally free sheaf on S such that for every irreducible component C^0 of a fiber of π, $E|_{C^0}$ is globally generated, then $R^1\pi_* E = 0$.*

Proof. By cohomology and base change [Hartshorne 1977, Theorem III.12.11], it suffices to prove that for every fiber C of π, $H^1(C, E|_C) = 0$. We first show that if l is the number of irreducible components of C counted with multiplicity, then we can write $C = C_1 + \cdots + C_l$ such that each C_i is an irreducible component of C and for every $1 \le i \le l - 1$, $(C_1 + \cdots + C_i) \cdot C_{i+1} \le 1$. This is proven by induction on l. If $l = 1$, there is nothing to prove. Otherwise, there is at least one component C^0 of C which can be contracted. Let r be the multiplicity of C^0 in C. Blowing down C^0, we get a rational surface S' over $\mathbb{P}^1$. Denote by C' the blow-down of C. Then by the induction hypothesis, we can write

$$C' = C'_1 + \cdots + C'_{l-r}$$

such that $(C'_1+\cdots+C'_i)\cdot C'_{i+1}\le 1$ for every $1\le i\le l-r-1$. Let C_i be the proper transform of C'_i. Then if in the above sum we replace C'_i by C_i when C_i does not intersect C^0, and by C_i+C^0 when C_i intersects C^0, we get the desired result for C.

Since $E|_{C_{i+1}}$ is globally generated, $H^1(C_{i+1}, E(-C_1-\cdots-C_i)|_{C_{i+1}})=0$ for every $0\le i\le l-1$. On the other hand, for every $0\le i\le l-2$, we have a short exact sequence of $\mathbb{O}_S$-modules

$$0\to E(-C_1-\cdots-C_{i+1})|_{C_{i+2}+\cdots+C_l}\to E(-C_1-\cdots-C_i)|_{C_{i+1}+\cdots+C_l}\\ \to E(-C_1-\cdots-C_i)|_{C_{i+1}}\to 0.$$

So a decreasing induction on i shows that for every $0\le i\le l-2$, we have $H^1(S, E(-C_1-\cdots-C_i)|_{C_{i+1}+\cdots+C_l})=0$. Letting $i=0$, the statement follows. □

Proposition 5.2. *Let $X\subset\mathbb{P}^n$ be a general hypersurface of degree d.*

(i) *For any morphism $h:\mathbb{P}^1\to X$, $h^*(T_X(1))$ is globally generated.*

(ii) *If C is a smooth, rational, d-normal curve on X, then $H^1(C, T_X|_C)=0$.*

Proof. (i) This follows from [Voisin 1996, Proposition 1.1]. We give a proof here for the sake of completeness. Consider the short exact sequence

$$0\to h^*T_X\to h^*T_{\mathbb{P}^n}\to h^*\mathbb{O}_X(d)\to 0.$$

Since X is general, the image of the pullback map

$$H^0(X,\mathbb{O}_X(d))\to H^0(\mathbb{P}^1, h^*\mathbb{O}_X(d))$$

is contained in the image of the map $H^0(\mathbb{P}^1, h^*T_{\mathbb{P}^n})\to H^0(\mathbb{P}^1, h^*\mathbb{O}_X(d))$. Choose a homogeneous coordinate system for $\mathbb{P}^n$. Let p be a point in $\mathbb{P}^1$, and without loss of generality assume that $h(p)=(1:0:\cdots:0)$. We show that for any $r\in h^*(T_X(1))|_p$, there is $\tilde r\in H^0(\mathbb{P}^1, h^*(T_X(1)))$ such that $\tilde r|_p=r$.

Consider the exact sequence

$$0\longrightarrow H^0(\mathbb{P}^1, h^*T_X(1))\longrightarrow H^0(\mathbb{P}^1, h^*T_{\mathbb{P}^n}(1))\xrightarrow{\phi} H^0(\mathbb{P}^1, h^*\mathbb{O}_X(d+1)).$$

Denote by s the image of r in $h^*(T_{\mathbb{P}^n}(1))|_p$. There exists $S\in H^0(\mathbb{P}^n, T_{\mathbb{P}^n}(1))$ such that the restriction of $\tilde s:=h^*(S)$ to p is s. Denote by T the image of S in $H^0(\mathbb{P}^n,\mathbb{O}_{\mathbb{P}^n}(d+1))$, and let $\tilde t=h^*(T)$. Then T is a form of degree $d+1$ on $\mathbb{P}^n$, and since $\tilde t|_p=0$, we can write $T=x_1G_1+\cdots+x_nG_n$, where the G_i are forms of degree d. Our assumption implies that for every $1\le i\le n$, there is $\tilde s_i\in H^0(\mathbb{P}^1, h^*T_{\mathbb{P}^n})$ such that $\phi(\tilde s_i)=h^*G_i$. Then

$$\phi(\tilde s-h^*(x_1)\tilde s_1-\cdots-h^*(x_n)\tilde s_n)=\tilde t-h^*(x_1G_1)-\cdots-h^*(x_nG_n)=0,$$

and therefore there is some $\tilde r\in H^0(\mathbb{P}^1, h^*(T_X(1)))$ whose image is

$$\tilde s-h^*(x_1)\tilde s_1-\cdots-h^*(x_n)\tilde s_n.$$

Since $(\tilde{s}-h^*(x_1)\tilde{s}_1-\cdots-h^*(x_n)\tilde{s}_n)|_p=\tilde{s}|_p=s$, we have $\tilde{r}|_p=r$.

(ii) There is a short exact sequence

$$0\to T_X|_C\to T_{\mathbb{P}^n}|_C\to \mathbb{O}_C(d)\to 0.$$

The fact that X is general implies that any section of $\mathbb{O}_C(d))$ which is the restriction of a section of $\mathbb{O}_{\mathbb{P}^n}(d)$ can be lifted to a section of $T_{\mathbb{P}^n}|_C$. Since the first cohomology group of $T_{\mathbb{P}^n}|_C$ vanishes, the result follows. □

Although for every e and n with $e\ge n+1\ge 4$, there are smooth nondegenerate rational curves of degree e in $\mathbb{P}^n$ which are not $(e-n)$-normal [Gruson et al. 1983, Theorem 3.1], a general smooth rational curve of degree e in a general hypersurface of degree d has possibly a much smaller normality: if a maximal-rank type conjecture holds for rational curves contained in general hypersurfaces (at least when $d<(n+1)/2$), then it follows that if c is the smallest positive number such that $\binom{n+c}{n}-\binom{n+c-d}{n}\ge ce+1$, a general smooth rational curve of degree e in a general hypersurface of degree d in $\mathbb{P}^n$ is c-normal.

Acknowledgments

I am grateful to Izzet Coskun, N. Mohan Kumar, Mike Roth, and Jason Starr for many helpful conversations. I also thank the referee for a careful reading of the paper and several significant suggestions.

References

[Beheshti and Starr 2008] R. Beheshti and J. M. Starr, "Rational surfaces in index-one Fano hypersurfaces", *J. Algebraic Geom.* **17**:2 (2008), 255–274. MR 2008j:14075 Zbl 1141.14024

[Coskun and Starr 2009] I. Coskun and J. Starr, "Rational curves on smooth cubic hypersurfaces", *Int. Math. Res. Not.* **2009**:24 (2009), 4626–4641. MR 2010j:14056 Zbl 1200.14051

[Gruson et al. 1983] L. Gruson, R. Lazarsfeld, and C. Peskine, "On a theorem of Castelnuovo, and the equations defining space curves", *Invent. Math.* **72**:3 (1983), 491–506. MR 85g:14033 Zbl 0565.14014

[Harris et al. 2004] J. Harris, M. Roth, and J. Starr, "Rational curves on hypersurfaces of low degree", *J. Reine Angew. Math.* **571** (2004), 73–106. MR 2005e:14067 Zbl 1052.14027

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[de Jong and Starr 2004] A. J. de Jong and J. Starr, "Cubic fourfolds and spaces of rational curves", *Illinois J. Math.* **48**:2 (2004), 415–450. MR 2006e:14007 Zbl 1081.14007

[de Jong and Starr 2006] A. de Jong and J. Starr, "Low degree complete intersections are rationally simply connected", preprint, 2006, available at http://www.math.sunysb.edu/~jstarr/papers/nk1006g.pdf.

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

[Kollár et al. 1992] J. Kollár, Y. Miyaoka, and S. Mori, "Rational connectedness and boundedness of Fano manifolds", *J. Differential Geom.* **36**:3 (1992), 765–779. MR 94g:14021 Zbl 0759.14032

[Reider 1988] I. Reider, "Vector bundles of rank 2 and linear systems on algebraic surfaces", *Ann. of Math.* (2) **127**:2 (1988), 309–316. MR 89e:14038 Zbl 0663.14010

[Starr 2003] J. Starr, "The Kodaira dimension of spaces of rational curves on low degree hypersurfaces", preprint, 2003. arXiv math/0305432

[Starr 2006] J. Starr, "Hypersurfaces of low degree are rationally 1-connected", preprint, 2006. arXiv math/0602641

[Voisin 1996] C. Voisin, "On a conjecture of Clemens on rational curves on hypersurfaces", *J. Differential Geom.* **44**:1 (1996), 200–213. MR 97j:14047 Zbl 0883.14022

Communicated by David Eisenbud
Received 2010-09-20 Revised 2011-06-19 Accepted 2011-07-28

beheshti@math.wustl.edu *Department of Mathematics, Washington University, Campus Box 1146, Saint Louis, MO 63130, United States*

ALGEBRA AND NUMBER THEORY 6:4(2012)

Degeneracy of triality-symmetric morphisms

Dave Anderson

We define a new symmetry for morphisms of vector bundles, called *triality symmetry*, and compute Chern class formulas for the degeneracy loci of such morphisms. In an appendix, we show how to canonically associate an octonion algebra bundle to any rank-2 vector bundle.

1. Introduction

Let $\varphi : E \to F$ be a morphism of vector bundles on a smooth variety X, of respective ranks m and n. The r-th *degeneracy locus* of φ is the set of points of X defined by

$$D_r(\varphi) = \{x \in X \mid \operatorname{rk} \varphi(x) \leq r\},$$

where $\varphi(x) : E(x) \to F(x)$ is the corresponding linear map in the fibers over $x \in X$. Such loci are ubiquitous in algebraic geometry: many interesting varieties, from Veronese embeddings of projective spaces to Brill–Noether loci parametrizing special divisors in Jacobians, can be realized as degeneracy loci for appropriate maps of vector bundles. General geometric information about degeneracy loci is therefore often useful. In particular, one can ask for Chern class formulas for the cohomology class of $D_r(\varphi)$ in H^*X — what is $[D_r(\varphi)]$ as a polynomial in the Chern classes of E and F?

When φ is sufficiently general, so $D_r(\varphi)$ has *expected codimension* equal to $(m-r)(n-r)$, the answer is given by the Giambelli–Thom–Porteous determinantal formula. In two cases of particular interest, Chern class formulas are known for degeneracy loci where φ is not general in this sense. Taking $F = E^*$, one has the dual morphism $\varphi^* : E^{**} = E \to E^*$. Call φ *symmetric* if $\varphi^* = \varphi$, and *skew-symmetric* if $\varphi^* = -\varphi$. The codimension of $D_r(\varphi)$ is at most $\binom{m-r+1}{2}$ (in the symmetric case) or $\binom{m-r}{2}$ (in the skew-symmetric case), so such morphisms are never sufficiently general for the Giambelli–Thom–Porteous formula to apply. Formulas for these loci were given by Harris and Tu [1984] and Józefiak, Lascoux

This work was partially supported by NSF Grants DMS-0502170 and DMS-0902967.
MSC2010: primary 14M15; secondary 14F43, 14N15, 20G99, 17A75.
Keywords: degeneracy locus, triality, octonions, equivariant cohomology.

and Pragacz [Józefiak et al. 1981]. As explained in [Fehér et al. 2005], these formulas can also be found by computing the equivariant classes of appropriate orbit closures in the $\mathrm{GL}(E)$-representations $\operatorname{Sym}^2 E^*$ and $\bigwedge^2 E^*$, where E is a vector space. See [Fulton and Pragacz 1998, Chapter 6] for more detailed discussions of the formulas.

The primary goal of the present article is to give degeneracy locus formulas for a new class of morphisms, which we call *triality-symmetric* morphisms. Letting E be a rank-2 vector bundle, these are maps

$$\varphi : E \to \operatorname{End}(E) \oplus E^*$$

possessing a certain symmetry related to the S_3 symmetry of the D_4 Dynkin diagram. Specifically, we use the following definition:

Definition 1.1. Consider the canonical identification

$$\begin{aligned}\operatorname{Hom}(E, \operatorname{End}(E) \oplus E^*) &= (E^* \otimes E^* \otimes E) \oplus (E^* \otimes E^*)\\ &= (E^* \otimes E^* \otimes E^* \otimes \textstyle\bigwedge^2 E) \oplus (E^* \otimes E^*).\end{aligned}$$

A morphism $\varphi : E \to \operatorname{End}(E) \oplus E^*$ is *triality-symmetric* if the corresponding section of $\operatorname{Hom}(E, \operatorname{End}(E) \oplus E^*)$ lies in the subbundle

$$(\operatorname{Sym}^3 E^* \otimes \textstyle\bigwedge^2 E) \oplus \bigwedge^2 E^*.$$

That is, $\varphi = \varphi_1 \oplus \varphi_2$, with φ_1 defining a symmetric trilinear form $\operatorname{Sym}^3 E \to \bigwedge^2 E$ and φ_2 defining an alternating bilinear form $\bigwedge^2 E \to \mathbb{O}_X$.

We will sometimes write $\operatorname{tSym}(E^*) = (\operatorname{Sym}^3 E^* \otimes \bigwedge^2 E) \oplus \bigwedge^2 E^*$ for the subbundle of triality-symmetric morphisms.

A few words of motivation are in order concerning this definition. For simplicity, consider the case where X is a point, and take vector spaces E and F of respective dimensions m and n. The space of all linear maps $\operatorname{Hom}(E, F)$ is also the tangent space to the Grassmannian $\operatorname{Gr}(m, m+n) = \operatorname{Gr}(m, E \oplus F) = \mathrm{GL}_{m+n}/P$ (for an appropriate maximal parabolic subgroup P) at the point corresponding to E. When $F = E^*$, there is a canonical symplectic form ω on $E \oplus E^*$, defining the Lagrangian Grassmannian $\mathrm{LG}(m, 2m) \subseteq \operatorname{Gr}(m, 2m)$, and the space of symmetric morphisms $\operatorname{Sym}^2 E^*$ is naturally identified with the tangent space to $\mathrm{LG}(m, 2m) = \mathrm{Sp}_{2m}/P$ at the point $[E]$. Moreover, $\mathrm{LG}(m, 2m)$ is the fixed locus for the involution of $\operatorname{Gr}(m, 2m)$ which sends a subspace to its orthogonal complement under ω. The situation is similar for skew-symmetric morphisms, replacing the Lagrangian Grassmannian with the orthogonal Grassmannian $\mathrm{OG}(m, 2m) = \mathrm{SO}_{2m}/P$.

From this point of view, it is natural to expect nice degeneracy locus formulas corresponding to other finite symmetries of homogeneous spaces. A particularly interesting one is the *triality* action on $\mathrm{OG}(2, 8)$, which we identify as

$$\mathrm{OG}(2, E \oplus \mathrm{End}(E) \oplus E^*)$$

for a two-dimensional vector space E. A concise description of this S_3 action may be found in [Anderson 2009, Appendix B]; for more details, see [van der Blij and Springer 1960; Garibaldi 1999]. For our purposes, the relevant facts are that the fixed locus is the "G_2 Grassmannian" G_2/P (for P corresponding to the long root), and the tangent space to G_2/P is naturally identified with $\mathrm{tSym}(E^*)$ at the point $[E] \in G_2/P \subseteq \mathrm{OG}(2, 8)$. (In Section 3, we will explicitly exhibit the S_3 action on the tangent space $T_{[E]}\,\mathrm{OG}(2, 8) \cong \mathrm{Hom}(E, \mathrm{End}(E)) \oplus \bigwedge^2 E^*$ fixing $\mathrm{tSym}(E^*)$.) Further motivation comes from the fact that there is a canonical *octonion algebra* structure on $E \oplus \mathrm{End}(E) \oplus E^*$, when E is a rank-2 vector bundle, just as there is a canonical symplectic structure on $E \oplus E^*$. This is the content of Proposition A.1.

Since E is required to have rank-2, a triality-symmetric morphism may have rank 0, 1, or 2. Write $D_r(\varphi) \subseteq X$ for the locus of points where φ has rank at most r. For a triality-symmetric morphism φ, define the *expected codimension* of $D_r(\varphi)$ to be 5, 3, or 0 if $r = 0$, $r = 1$, or $r = 2$, respectively. With this understood, we may state our main theorem:

Theorem 1.2. *Let c_1, c_2 be the Chern classes of E^*, and let x_1, x_2 be Chern roots. Let $\varphi : E \to \mathrm{End}(E) \oplus E^*$ be a triality-symmetric morphism. If $D_r(\varphi)$ has expected codimension and X is Cohen–Macaulay, then we have $[D_r(\varphi)] = P_r(c_1, c_2)$ in H^*X, where*

$$\begin{aligned}
P_2 &= 1,\\
P_1 &= 3\,c_2\,c_1 = 3x_1x_2(x_1 + x_2),\\
P_0 &= c_2\,c_1\,(9\,c_2 - 2\,c_1^2) = x_1x_2(x_1 + x_2)(2x_1 - x_2)(-x_1 + 2x_2).
\end{aligned}$$

A secondary goal of this article is to illustrate two points of view on degeneracy loci. In this spirit, we will give two proofs of the main theorem, both involving the simple Lie group of type G_2, but using substantially different approaches. The first relates degeneracy loci for triality-symmetric morphisms to certain Schubert loci in a G_2 flag bundle, just as Fulton's generalization of the Harris–Tu formulas relates symmetric morphisms to type C flag bundles [Fulton 1996]. One then applies the formulas for G_2 Schubert loci developed in [Anderson 2011] to derive the formulas of Theorem 1.2.

The second proof uses equivariant cohomology in the spirit of [Fehér and Rimányi 2004; Fehér et al. 2005] (but see Remark 5.3). More precisely, when P is the maximal parabolic subgroup of G_2 which omits the long root and E is a two-dimensional vector space, we consider $(\mathrm{Sym}^3 E^* \otimes \bigwedge^2 E) \oplus \bigwedge^2 E^*$ as a P-module and compute the equivariant classes of the P-orbit closures in this vector space.

Certain of these orbit closures correspond to degeneracy loci, and one can deduce Theorem 1.2 from the equivariant formulas. Along the way, we explicitly identify the P-orbit closures in $(\mathrm{Sym}^3 E^* \otimes \bigwedge^2 E) \oplus \bigwedge^2 E^*$, and compute all their equivariant classes (Proposition 5.1 and Theorem 5.2).

Triality symmetry is the G_2 case of a general notion of symmetry for morphisms of vector bundles. In fact, two types of symmetry for morphisms can be naturally associated to any maximal parabolic subgroup P of a complex reductive group G, as described in [Anderson 2009, Appendix C]. The "orbit" approach used in the second proof of Theorem 1.2 generalizes to the following problem: *Compute the equivariant classes of P-orbit (or B-orbit) closures for the adjoint action on* $\mathfrak{g}/\mathfrak{p}$. Solutions to this problem account for many of the known degeneracy locus formulas; see, for example, [Fehér and Rimányi 2003; Knutson and Miller 2005].

A related problem is to classify situations where there are finitely many orbits. In the case of P acting on $\mathfrak{g}/\mathfrak{p}$, this problem was investigated by Popov and Röhrle [1997], and such parabolic actions have been classified [Bürgstein and Hesselink 1987; Hille and Röhrle 1999; Jürgens and Röhrle 2002]. The classification of Borel or Levi subgroup actions on $\mathfrak{g}/\mathfrak{p}$ with finitely many orbits appears to be unknown.

We have endeavored to make our perspective on triality and G_2 accessible to general algebraic geometers, and in this spirit, the ingredients of the first proof of Theorem 1.2 are spelled out quite explicitly. The second proof is more streamlined, but requires a little more specialized background; we hope that the reader versed in Lie theory will appreciate both points of view.

2. Preliminaries

All varieties are over $\mathbb{C}$. We will write X for the base variety. If E is a vector bundle on X, we write $E(x)$ for the fiber over $x \in X$. We often suppress notation for pullback of vector bundles.

2.1. *Octonions.* An *octonion algebra* over $\mathbb{C}$ is an 8-dimensional complex vector space C, equipped with

- a nondegenerate quadratic norm N, and
- a bilinear multiplication with unit e, written $u \otimes v \mapsto uv$,

such that $N(uv) = N(u)N(v)$ for all $u, v \in C$. Recall that any quadratic norm N corresponds to a symmetric bilinear form $\langle\cdot\,,\cdot\rangle$, by $N(v) = \frac{1}{2}\langle v, v\rangle$ and $\langle u, v\rangle = N(u+v) - N(u) - N(v)$, and a norm is called nondegenerate if the corresponding bilinear form is nondegenerate.

Up to isomorphism, there is only one octonion algebra over $\mathbb{C}$ (or over any algebraically closed field). The multiplication is only required to be bilinear, and indeed it is neither commutative nor associative.

The notion of an octonion algebra globalizes easily to *octonion bundles*, where C is a rank-8 vector bundle on a variety X, the multiplication is a vector bundle map $C \otimes C \to C$, and for simplicity we assume the norm takes values in $\mathbb{O}_X$. For more on octonions and octonion bundles, see [Springer and Veldkamp 2000, Sections 1–2; Petersson 1993; Anderson 2009, Section 2].

The group of algebra automorphisms of an octonion algebra (that is, linear automorphisms preserving multiplication) is the simple complex Lie group of type G_2 [Springer and Veldkamp 2000, Section 2]; abusing notation, we will write G_2 to denote this group.[1]

Let E be a rank-2 vector bundle on X. Then $C = E \oplus \operatorname{End}(E) \oplus E^*$ has a canonical structure of an octonion bundle, which is described in Proposition A.1. In the case where X is a point, so E is a 2-dimensional vector space, the same formulas (1) and (2) define an octonion algebra. It will be convenient to use a basis adapted to this construction. Let v_1, v_2 be a basis for E, with dual basis v_1^*, v_2^* for E^*, and extend to a basis for $C = E \oplus \operatorname{End}(E) \oplus E^*$ by setting

$$\begin{aligned} v_3 &= v_2^* \otimes v_1, \qquad & v_4 &= v_1^* \otimes v_1, \qquad & v_5 &= v_2^* \otimes v_2, \\ v_6 &= v_1^* \otimes v_2, \qquad & v_7 &= v_2^*, \qquad & v_8 &= v_1^*. \end{aligned} \tag{1}$$

One checks that the identity element of C is $e = v_4 + v_5$.

With respect to this basis, the symmetric bilinear form $\langle \cdot\,, \cdot \rangle$ is given by

$$\begin{aligned} \langle v_p, v_{9-q} \rangle &= -\delta_{pq}, \ \text{ for } \{p, q\} \neq \{4, 5\}; \\ \langle v_4, v_5 \rangle &= 1. \end{aligned} \tag{2}$$

Write $V = e^{\perp} \subset C$ for the orthogonal complement of the identity element with respect to $\langle \cdot\,, \cdot \rangle$. Thus V is defined by $v_4^* + v_5^* = 0$.

Let the torus $T = (\mathbb{C}^*)^2$ act on C in this basis via the matrix

$$\operatorname{diag}(z_1,\ z_2,\ z_1 z_2^{-1},\ 1,\ 1,\ z_1^{-1} z_2,\ z_2^{-1},\ z_1^{-1}), \tag{3}$$

with weights

$$\{t_1,\ t_2,\ t_1 - t_2,\ 0,\ 0,\ -t_1 + t_2,\ -t_2,\ -t_1\}. \tag{4}$$

This is induced from the standard action on $E = \operatorname{span}\{v_1, v_2\}$. The algebra structure of C is preserved by this action, so $T \subseteq G_2$; in fact, T is a maximal torus.

2.2. ***Roots and weights.*** For general Lie-theoretic notions, we refer to [Humphreys 1975]; here we explain the relevant facts for type G_2. Let G_2 be the automorphism group of an octonion algebra C, as above, so G_2 is presented as a subgroup of

[1] In fact, one can show that $u^2 = \langle u, e \rangle u - N(u)e$ for any element $u \in C$, so any algebra automorphism also preserves the norm.

$\mathrm{GL}(C) \cong \mathrm{GL}_8$. Let $T \subset B \subset G_2$ be a maximal torus and Borel subgroup, and let $\mathfrak{t} \subset \mathfrak{b} \subset \mathfrak{g}_2$ be the corresponding Lie algebras. Once a basis for C has been chosen as in (1), we will always take T to be the torus acting as in (3), and we may take B to be the intersection of the upper-triangular matrices in GL_8 with the subgroup G_2. Write α_1 and α_2 for the two *simple roots*, with α_2 the long root. In terms of the weights t_1, t_2 of (4), we have

$$\alpha_1 = t_1 - t_2, \qquad \alpha_2 = -t_1 + 2\,t_2. \tag{5}$$

The *positive roots* are $\alpha_1, \alpha_2, \alpha_1+\alpha_2, 2\,\alpha_1+\alpha_2, 3\alpha_1+\alpha_2, 3\alpha_1+2\,\alpha_2$; the *negative roots* are $-\alpha$, for α a positive root.

Let $P \subset G_2$ be the standard maximal parabolic subgroup omitting the long root, with Lie algebra $\mathfrak{p} \subset \mathfrak{g}_2$. Thus $\mathfrak{p} = \mathfrak{b} \oplus \mathfrak{g}_{-\alpha_1}$, where $\mathfrak{g}_{-\alpha_1} \subset \mathfrak{g}_2$ is the weight space for the negative root $-\alpha_1$.

The *Weyl group* is $W = N(T)/T$, where $N(T)$ is the normalizer of T in G_2. It is isomorphic to the dihedral group with 12 elements, and is generated by the *simple reflections* $s = s_{\alpha_1}$ and $t = s_{\alpha_2}$, and is defined by the relations $s^2 = t^2 = (st)^6 = 1$. There is an embedding $W \hookrightarrow S_7$ coming from the action of G_2 on $V \subset C$ given by

$$s \mapsto 2\,1\,5\,4\,3\,7\,6,$$
$$t \mapsto 1\,3\,2\,4\,6\,5\,7;$$

see [Anderson 2009, Section A.3]. We will sometimes treat elements of W as permutations via this embedding.

2.3. ***Flag bundles and Schubert loci.*** We refer to [Anderson 2009; 2011] for proofs of the following facts with more details. (There the term "γ-isotropic" is used instead of "G_2-isotropic" in reference to a trilinear form γ.)

Let C be an octonion algebra, and let $V = e^\perp \subset C$ be as before. A subspace $E \subseteq C$ is called G_2-*isotropic* if $E \subseteq V$ and $uv = 0$ for all $u, v \in E$. A maximal G_2-isotropic subspace has dimension 2, and a G_2-*isotropic flag* is a chain $E_1 \subset E_2 \subset V$ (with $\dim E_i = i$), where E_2 is G_2-isotropic. Such a flag can be canonically extended to a complete flag $E_1 \subset E_2 \subset E_3 \subset \cdots \subset E_7 = V$: When $E_1 \subset E_2$ is G_2-isotropic, with E_1 spanned by a vector u, then $E_u := \{v \in V \mid uv = 0\}$ is a three-dimensional subspace containing E_2. To get a complete flag, set $E_3 = E_u$, and then take orthogonal complements with respect to the norm N for the rest, so $E_4 = E_3^\perp$, etc. (See [Anderson 2011, Section 2.2] for this construction.)

The G_2 *flag variety* Fl_{G_2} parametrizes all G_2-isotropic flags in $V \subset C$. It is a six-dimensional projective homogeneous space, isomorphic to G_2/B for a Borel subgroup $B \subset G_2$. The G_2 *Grassmannian* Gr_{G_2} parametrizes two-dimensional G_2-isotropic subspaces of V; this is isomorphic to the five-dimensional homogeneous space G_2/P. The construction of a complete G_2-isotropic flag gives an embedding $\mathrm{Fl}_{G_2} \hookrightarrow \mathrm{Fl}(\mathbb{C}^7) = \mathrm{SL}_7/B$.

For an octonion bundle C on X with its rank-7 subbundle V, there is an associated *G_2-isotropic flag bundle* $\mathbf{Fl}_{G_2}(V) \to X$, as well as a *$G_2$-isotropic Grassmann bundle* $\mathbf{Gr}_{G_2}(V) \to X$. These are (étale-)locally trivial fiber bundles, with fibers Fl_{G_2} and Gr_{G_2}, respectively. The flag bundle $\mathbf{Fl}_{G_2}$ comes with a tautological flag of subbundles $\widetilde{E}_\bullet$ of V.

Given a complete G_2-isotropic flag of subbundles $F_1 \subset F_2 \subset \cdots \subset F_7 = V$ on X, the *Schubert loci* in $\mathbf{Fl}_{G_2}(V)$ are defined by

$$\Omega_w(F_\bullet) = \{x \in \mathbf{Fl}_{G_2} \mid \dim(\widetilde{E}_p(x) \cap F_q(x)) \geq r_w(q,p) \text{ for } 1 \leq p, q \leq 7\}, \tag{6}$$

where for $w \in W$, $r_w(q,p)$ is $\#\{i \leq q \mid w(8-i) \leq p\}$, and $\widetilde{E}_\bullet$ is the tautological flag on $\mathbf{Fl}_{G_2}$. Here we are using the embedding $W \hookrightarrow S_7$ discussed above.[2] The codimension of Ω_w is the *length* of w, i.e., the least number of simple transpositions needed to write w as a word in s and t.

If $E_\bullet$ is a second G_2-isotropic flag on X, it defines a section $s_{E_\bullet} : X \to \mathbf{Fl}_{G_2}$ such that $s_{E_\bullet}^* \widetilde{E}_\bullet = E_\bullet$. We define degeneracy loci in X as the scheme-theoretic inverse images of Schubert loci:

$$\Omega_w(E_\bullet, F_\bullet) = s_{E_\bullet}^{-1} \Omega_w(F_\bullet).$$

3. Triality symmetry

Triality symmetry is described in terms of coordinates as follows. Assume X is a point, so E is a two-dimensional vector space. Choose a basis $\{v_1, v_2\}$ for E, and let $\{v_3, \ldots, v_8\}$ be a basis for $\mathrm{End}(E) \oplus E^*$ as in (1). Suppose $\varphi : E \to \mathrm{End}(E) \oplus E^*$ is given by $\varphi = \varphi_1 \oplus \varphi_2$, with

$$\varphi_1(v_1) = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}, \quad \varphi_1(v_2) = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix},$$

and $\varphi_2(v_1) = z\, v_2^*$, $\varphi_2(v_2) = -z\, v_1^*$. In terms of the chosen bases for E and $\mathrm{End}(E) \oplus E^*$, φ has matrix A_φ^t, whose transpose is

$$A_\varphi = \begin{pmatrix} b_1 & a_1 & d_1 & c_1 & z & 0 \\ b_2 & a_2 & d_2 & c_2 & 0 & -z \end{pmatrix}. \tag{7}$$

[2]This definition of r_w differs slightly from that of [Anderson 2011]; there the assignment $(q,p) \mapsto \#\{i \leq q \mid w(i) \leq p\}$ is called r_w. The two are related by replacing w with $ww_\circ$, where $w_\circ$ is the longest element of W.

Identify $\mathrm{Hom}(E, \mathrm{End}(E)) = E^* \otimes E^* \otimes E$ with $E^* \otimes E^* \otimes E^*$ by mapping

$$v_i^* \otimes v_j^* \otimes v_1 \mapsto v_{ij2}^*,$$
$$v_i^* \otimes v_j^* \otimes v_2 \mapsto -v_{ij1}^*,$$

where $v_{ijk}^* = v_i^* \otimes v_j^* \otimes v_k^*$ for $1 \le i, j, k \le 2$. (The sign appears because of the canonical isomorphism $E^* \otimes E^* \otimes E \cong E^* \otimes E^* \otimes E^* \otimes \bigwedge^2 E$; we are using $v_1 \wedge v_2$ to identify $E \cong E^* \otimes \bigwedge^2 E$ with E^*.) Thus φ is triality-symmetric if and only if the corresponding coordinates of v_{ijk}^* are invariant under permutations of the indices. Explicitly, there is an S_3-action on $\mathrm{Hom}(E, \mathrm{End}(E)) \oplus \bigwedge^2 E^*$ generated by elements τ and σ whose action on matrices A_φ is given by

$$\tau\begin{pmatrix} b_1 & a_1 & d_1 & c_1 & z & 0 \\ b_2 & a_2 & d_2 & c_2 & 0 & -z \end{pmatrix} = \begin{pmatrix} -d_2 & -d_1 & c_2 & c_1 & z & 0 \\ b_2 & b_1 & -a_2 & -a_1 & 0 & -z \end{pmatrix}$$
$$\text{and} \quad \sigma\begin{pmatrix} b_1 & a_1 & d_1 & c_1 & z & 0 \\ b_2 & a_2 & d_2 & c_2 & 0 & -z \end{pmatrix} = \begin{pmatrix} a_2 & a_1 & c_2 & c_1 & z & 0 \\ b_2 & b_1 & d_2 & d_1 & 0 & -z \end{pmatrix}.$$

This means that the triality-symmetric maps are those whose (transposed) matrix is of the form

$$A_\varphi = \begin{pmatrix} a & -d & d & c & z & 0 \\ b & a & -a & d & 0 & -z \end{pmatrix}. \tag{8}$$

Here a is also the coordinate of v_{122}^*, b is the coordinate of v_{222}^*, $-c$ is the coordinate of v_{111}^*, and $-d$ is the coordinate of v_{112}^*. Note that the S_3-invariants coincide with the τ-invariants.

Remark 3.1. "Triality" usually refers to several phenomena related to the S_3 symmetry of the D_4 Dynkin diagram. It was first described by Cartan [1925]; see [Knus et al. 1998] for a thorough discussion. The connection with our context can be explained briefly as follows. Automorphisms of the D_4 Dynkin diagram correspond to outer automorphisms of the simply connected group Spin_8; these all fix a parabolic subgroup P, and therefore define automorphisms of $\mathrm{Spin}_8 / P \cong \mathrm{OG}(2, 8)$ and the tangent space $T_{eP} \mathrm{Spin}_8 / P$. The tangent space can be identified with matrices as in (7), and under this identification the automorphism group S_3 acts as described above.

4. Graphs

For any morphism $\varphi : E \to F$, let $E_\varphi \subset E \oplus F$ be its graph, that is, the subbundle whose fiber over x is $E_\varphi(x) = \{(v, \varphi(v)) \mid v \in E(x)\}$. If $\varphi : E \to E^*$ is symmetric, then its graph is isotropic for the *canonical skew-symmetric form* on $E \oplus E^*$ defined by $(v_1 \oplus f_1, v_2 \oplus f_2) = f_1(v_2) - f_2(v_1)$. Thus one obtains a map to the Lagrangian bundle of isotropic flags in $E \oplus E^*$, and formulas for the degeneracy loci of φ are deduced from formulas for Schubert loci; see [Fulton 1996; Fulton and Pragacz 1998].

In this section, we consider morphisms $\varphi : E \to \mathrm{End}(E) \oplus E^*$. There is, by Proposition A.1, a canonical octonion algebra structure on $E \oplus \mathrm{End}(E) \oplus E^*$. We give formulas for degeneracy loci of morphisms whose graphs are G_2-isotropic with respect to this structure. In general such morphisms are not triality-symmetric (nor vice versa). For rank-1 maps, however, the two notions agree.

After a suitable change of coordinates (including a switch to opposite Schubert cells), the parametrization of the open Schubert cell in G_2/P given in [Anderson 2009, Section D.1] becomes

$$\widetilde{\Omega}^\circ = \begin{pmatrix} 1 & 0 & a & -d & d & c & z & -X \\ 0 & 1 & b & a & -a & d & -Z & -Y \end{pmatrix}, \tag{9}$$

where $X = -ac - d^2$, $Y = z + ad - bc$, and $Z = -a^2 - bd$. Morphisms $E \to \mathrm{End}(E) \oplus E^*$ with G_2-isotropic graph are exactly those whose (transposed) matrix has the form of the last six columns of (9).

Lemma 4.1. *Suppose X is a point and $\varphi : E \to \mathrm{End}(E) \oplus E^*$ is a triality-symmetric map with matrix A_φ^t as in (8). Then the graph E_φ is contained in $V \subset C$, and is G_2-isotropic if and only if*

$$a^2 + bd = ac + d^2 = ad - bc = 0. \tag{10}$$

Conversely, suppose $\varphi : E \to \mathrm{End}(E) \oplus E^*$ has G_2-isotropic graph as in (9). Then φ is triality-symmetric if and only if the equations (10) hold.

Proof. This is a straightforward verification, using the basis $\{v_i\}$ as in (1). It is clear that the row span of (9) is always in $V \subset C$ since the fourth and fifth columns add to zero. The condition that the row span be the graph E_φ means $X = Z = 0$ and $Y = z$, which are precisely the equations (10). □

Corollary 4.2. *Let $\varphi : E \to \mathrm{End}(E) \oplus E^*$ be a morphism of rank at most* 1 *such that the component $\varphi_2 : E \to E^*$ is zero. Then φ is triality-symmetric if and only if $E_\varphi \subset C$ is G_2-isotropic.* (*This holds scheme-theoretically, that is, the equations locally defining these two subsets of* $\mathrm{Hom}(E, \mathrm{End}(E))$ *are the same.*)

Proof. This is a local statement, so we may assume X is a point and compute in coordinates. In this case, it follows from Lemma 4.1 by adding the equation $z = 0$. (In fact, for a morphism with G_2-isotropic graph, the rank condition is forced by $\varphi_2 \equiv 0$.) □

Corollary 4.2 implies that the formulas of Theorem 1.2 (for triality-symmetric morphisms) will agree with formulas for morphisms with G_2-isotropic graphs. Before proceeding with the proof of Theorem 1.2, we will describe the connection between triality-symmetry and G_2-isotropic graphs more precisely.

Let $\mathbf{Gr}_{G_2} \subseteq \mathbf{Gr}(2, C)$ be the G_2-Grassmannian bundle on X, and let $\mathbf{Gr}^\circ$ be the open subset parametrizing subbundles of $C = E \oplus \mathrm{End}(E) \oplus E^*$ whose projection onto E is an isomorphism; locally on X, coordinates for $\mathbf{Gr}^\circ$ are given as in (9). Identifying a morphism $E \to \mathrm{End}(E) \oplus E^*$ with its graph, note that $\mathrm{Hom}(E, \mathrm{End}(E) \oplus E^*)$ is identified with the corresponding open subset of $\mathbf{Gr}(2, C)$, so $\mathbf{Gr}^\circ = \mathbf{Gr}_{G_2} \cap \mathrm{Hom}(E, \mathrm{End}(E) \oplus E^*)$ parametrizes morphisms with G_2-isotropic graph.

When X is a point, we have remarked that the space of triality-symmetric maps $\mathrm{tSym}(E^*)$ is naturally isomorphic to the tangent space $T_{[E]}\,\mathrm{Gr}_{G_2}$. For general X, this globalizes to an identification of the vector bundle $\mathrm{tSym}(E^*)$ with the normal bundle $N_{X/\mathbf{Gr}} = N_{X/\mathbf{Gr}^\circ}$, where X is embedded in $\mathbf{Gr}^\circ \subset \mathbf{Gr}$ by the section corresponding to the subbundle $E \subset C$.

Now let $\mathbf{D}_1 \subseteq \mathrm{tSym}(E^*)$ be the locus of triality symmetric morphisms of rank at most 1, and let $\Omega_1^\circ \subset \mathbf{Gr}^\circ$ be the locus of morphisms with G_2-isotropic graph of rank at most 1 such that the component φ_2 is zero. The next lemma identifies $\mathbf{D}_1$ with the normal cone to X in Ω_1°.

Lemma 4.3. *Inside* $\mathrm{tSym}(E^*) = N_{X/\mathbf{Gr}^\circ}$, *we have* $\mathbf{D}_1 = C_X\Omega_1^\circ$.

Proof. This can be checked locally on X, so assume X is a point. Note that both $\mathrm{tSym}(E^*)$ and $\mathbf{Gr}^\circ$ are isomorphic to $\mathbb{A}^5$. By Corollary 4.2, $\mathbf{D}_1$ and Ω_1° are defined by the same equations, namely (10) together with $z = 0$; since these equations are already homogeneous, we have that Ω_1° is equal to its tangent cone at the origin. □

Corollary 4.4. *Let φ be any triality-symmetric morphism and let ψ be any morphism with G_2-isotropic graph. Let $s_\varphi : X \to \mathrm{tSym}(E^*)$ and $s_\psi : X \to \mathbf{Gr}^\circ$ be the sections determined by φ and ψ. Then $s_\varphi^*[\mathbf{D}_1] = s_\psi^*[\Omega_1]$ in H^*X.*

Proof. Let s_0 be the zero section of $\mathrm{tSym}(E^*)$, and let s_E be the section of $\mathbf{Gr}^\circ$ corresponding to $E \subset C$. By Lemma 4.3 and the basic construction of intersection theory (see [Fulton 1998, Section 6]), we have $s_0^*[\mathbf{D}_1] = s_E^*[\Omega_1]$. On the other hand, both $\mathrm{tSym}(E^*)$ and $\mathbf{Gr}^\circ$ are affine bundles on X, so every section determines the same pullback on cohomology. (In fact, $\mathbf{Gr}^\circ$ is isomorphic to $\mathrm{tSym}(E^*)$, as one sees from the parametrization in (9), although it is not a vector subbundle of $\mathrm{Hom}(E, \mathrm{End}(E) \oplus E^*)$.) □

Consequently, for morphisms φ and ψ as in Corollary 4.4, we have $[D_1(\varphi)] = [D_1(\psi)]$ whenever

$$s_\varphi^*[\mathbf{D}_1] = [s_\varphi^{-1}\mathbf{D}_1] \quad \text{and} \quad s_\psi^*[\Omega_1] = [s_\psi^{-1}\Omega_1]. \tag{11}$$

(Indeed, $D_1(\varphi) = s_\varphi^{-1}\mathbf{D}_1$ and $D_1(\psi) = s_\psi^{-1}\Omega_1$ by definition.) When X is Cohen–Macaulay, so are $\mathbf{D}_1$ and Ω_1; this can be seen directly from the equations, or by using the fact that Schubert varieties are Cohen–Macaulay. The conditions (11) are therefore equivalent to the condition that $D_1(\varphi)$ and $D_1(\psi)$ have expected codimension in X, by [Fulton and Pragacz 1998, Lemma, p. 108].

First proof of Theorem 1.2. Let $\varphi : E \to \mathrm{End}(E) \oplus E^*$ have G_2-isotropic graph E_φ. Suppose E has a rank-1 subbundle, so E_φ also does. (One can always arrange for this, by passing to a $\mathbb{P}^1$-bundle if necessary.) Write $E_1 \subset E_2 = E$ and $F_1 \subset F_2 = E_\varphi$, and extend these to complete G_2-isotropic flags $E_\bullet$ and $F_\bullet$. For $w \in W$, set $\Omega_w(\varphi) = \Omega_w(E_\bullet, F_\bullet)$. Since $E_\varphi \cong E$, the Chern classes are the same. Let $-x_1, -x_2$ be Chern roots of E, so x_1, x_2 are Chern roots of E^*. Then, by [Anderson 2011, Theorem 2.4 and Section 2.5], we have

$$[\Omega_w(\varphi)] = \mathfrak{G}_w(x_1, x_2; -x_1, -x_2) \tag{12}$$

in H^*X, where $\mathfrak{G}_w(x_1, x_2; y_1, y_2)$ is the "G_2 double Schubert polynomial" defined in the same reference.

It remains to determine the w for which $D_r(\varphi) = \Omega_w(\varphi)$. We have

$$D_r(\varphi) = \{x \in X \mid \dim(E(x) \cap E_\varphi(x)) \geq 2 - r\},$$

and it is easy to check that

$$D_2(\varphi) = \Omega_{\mathrm{id}}(\varphi) = X, \quad D_1(\varphi) = \Omega_{tst}(\varphi), \quad \text{and} \quad D_0(\varphi) = \Omega_{tstst}(\varphi). \tag{13}$$

Indeed, the element $tst \in W$ corresponds to the permutation 3 6 1 4 7 2 5 (see [Anderson 2011, Section A.3]), so the condition defining Ω_{tst} is $\dim(E_2 \cap F_2) \geq r_{tst}(2, 2) = 1$. The other two identities are clear. This also justifies our definition of expected codimension for triality-symmetric degeneracy loci: the expected codimension of $D_r(\varphi)$ is the length of the corresponding element of W.

Specializing the polynomials $\mathfrak{G}_w$ given in [Anderson 2009, Section D.2] for these three w, we obtain the desired formulas for P_r. □

Remark 4.5. The twelve polynomials $\mathfrak{G}_w(x_1, x_2; -x_1, -x_2)$ become the equivariant localizations of Schubert classes in G_2/B at the point eB after the substitution $x_i = -t_i$; see [Anderson 2009, Section D.3].

Remark 4.6. We defined the scheme structure on $D_1(\varphi)$ for a triality-symmetric morphism by taking the equations (10) together with $z = 0$. In fact, the ideal generated by 2×2 minors of the matrix (8) is the same as the one generated by minors of (9), but this ideal is not radical. (It is generated by (10) together with az, bz, cz, dz, z^2.) The requirement $\varphi_2 \equiv 0$ for rank-1 maps is transparent on the triality-symmetric side; for G_2-isotropic graphs, the scheme structure is defined by pullback from the Schubert locus Ω_{tst}, and one sees $\varphi_2 \equiv 0$ from a parametrization of Schubert cells [Anderson 2009, Section D.1].

5. Orbits

Another approach to the computation of triality-symmetric degeneracy loci is as follows. Inside the vector bundle

$$\left(\operatorname{Sym}^3 E^* \otimes \textstyle\bigwedge^2 E\right) \oplus \textstyle\bigwedge^2 E^* \subset \operatorname{Hom}\left(E, \operatorname{End}(E) \oplus E^*\right),$$

there is a locus $\mathbf{D}_r$ consisting of morphisms of rank at most r. By definition, a triality-symmetric morphism φ defines a section s_φ of $\left(\operatorname{Sym}^3 E^* \otimes \bigwedge^2 E\right) \oplus \bigwedge^2 E^*$, and $D_r(\varphi) = s_\varphi^{-1}\mathbf{D}_r$ is the scheme-theoretic preimage.

It suffices to solve this problem on the classifying space for the vector bundle E (or on algebraic approximations thereof), so let $X = \mathrm{BGL}_2$.[3] Replace E with the standard representation of GL_2, and write

$$U = \left(\operatorname{Sym}^3 E^* \otimes \textstyle\bigwedge^2 E\right) \oplus \textstyle\bigwedge^2 E^*.$$

The relevant vector bundle on BGL_2 is $U \times^{\mathrm{GL}_2} \mathrm{EGL}_2$, where $\mathrm{EGL}_2 \to \mathrm{BGL}_2$ is the universal principal GL_2-bundle. Letting $D_r \subseteq U \subset \operatorname{Hom}(E, \operatorname{End}(E) \oplus E^*)$ be the locus of maps of rank at most r, we have

$$\mathbf{D}_r = D_r \times^{\mathrm{GL}_2} \mathrm{EGL}_2 \subseteq U \times^{\mathrm{GL}_2} \mathrm{EGL}_2\,.$$

Therefore $[\mathbf{D}_r] = [D_r]^{\mathrm{GL}_2}$ in $H^*(U \times^{\mathrm{GL}_2} \mathrm{EGL}_2) = H^*_{\mathrm{GL}_2}(U)$, and the problem becomes a computation in the equivariant cohomology of the vector space U.

Moreover, as we shall see below, D_r is an orbit closure for the action of GL_2 on U. In fact, we will use a larger group action. As discussed in Section 1, U may be identified with the tangent space

$$T_{[E]}G_2/P \cong \mathfrak{g}_2/\mathfrak{p},$$

so P acts on U via the adjoint action on $\mathfrak{g}_2/\mathfrak{p}$. Let $P = L \cdot P_u$ be the Levi decomposition, with P_u the unipotent radical and L a Levi subgroup. We will be interested in P-orbit closures in $\mathfrak{g}_2/\mathfrak{p}$.

First observe that L is isomorphic to GL_2. Here is one way to see this. Since E defines a point in G_2/P, the parabolic P may be identified with the subgroup of G_2 stabilizing E. Every linear automorphism of E induces an algebra automorphism of $C = E \oplus \operatorname{End}(E) \oplus E^*$; therefore $\mathrm{GL}(E) \cong \mathrm{GL}_2$ is a (reductive) subgroup of P, and we have $\mathrm{GL}(E) \subseteq L$. On the other hand, L is connected (since P is) and four-dimensional (by the root decomposition), so this inclusion must be an equality.

[3] Topologically, we may assume E is pulled back from the tautological bundle on $\mathrm{Gr}(2, n)$, for $n \gg 0$, so one can take a Grassmannian for an approximation to BGL_2.

The L-action on $\mathfrak{g}_2/\mathfrak{p}$ is identified with the natural GL_2-action on U: as an L-module, we have

$$\mathfrak{g}_2/\mathfrak{p} \cong \big(\mathrm{Sym}^3 E^* \otimes \textstyle\bigwedge^2 E\big) \oplus \textstyle\bigwedge^2 E^*,$$

where $E \cong \mathbb{C}^2$ is the standard representation of $L \cong \mathrm{GL}_2$ (with weights $t_1 = 2\alpha_1 + \alpha_2$ and $t_2 = \alpha_1 + \alpha_2$). As a P-module, $\mathfrak{g}_2/\mathfrak{p}$ is indecomposable, but there is an exact sequence

$$0 \to \mathrm{Sym}^3 E^* \otimes \textstyle\bigwedge^2 E \to \mathfrak{g}_2/\mathfrak{p} \to \textstyle\bigwedge^2 E^* \to 0.$$

These identifications of L- and P-modules follow directly from the weight decomposition of $\mathfrak{g}_2/\mathfrak{p}$: the T-weights are

$$-\alpha_2, \quad -\alpha_1 - \alpha_2, \quad -2\alpha_1 - \alpha_2, \quad -3\alpha_1 - \alpha_2, \quad -3\alpha_1 - 2\alpha_2. \tag{14}$$

As a first step to computing the classes of P-orbits in $H_T^*(\mathfrak{g}/\mathfrak{p})$, we give explicit descriptions of these orbits.

By the classification given in [Jürgens and Röhrle 2002], there are finitely many P-orbits on $\mathfrak{g}/\mathfrak{p}$. In fact, there are five orbits. To describe them, let

$$U' = \mathrm{Sym}^3 E^* \otimes \textstyle\bigwedge^2 E \subset U = \mathfrak{g}_2/\mathfrak{p}.$$

Let b, a, d, c be coordinates on U' with weights $-\alpha_2, -\alpha_1-\alpha_2, -2\alpha_1-\alpha_2, -3\alpha_1-\alpha_2$, respectively. The five orbits are O_c, with $c = 0, 1, 2, 3, 5$ giving the codimension; their closures are nested and described by the following proposition:

Proposition 5.1. *The P-orbit closures in $U = \mathfrak{g}_2/\mathfrak{p}$ are as follows*:

- $\overline{O}_0 = U$.
- $\overline{O}_1 = U'$.
- $\overline{O}_2$ *is the discriminant locus in U' defined by the vanishing of the quartic polynomial* $a^2d^2 + 4a^3c + 4bd^3 - 27b^2c^2 + 18abcd$.
- $\overline{O}_3$ *is the* (*affine*) *cone over the twisted cubic curve in $\mathbb{P}^3 = \mathbb{P}U'$ defined by the condition that the matrix*

$$\begin{pmatrix} a & -d & c \\ b & a & d \end{pmatrix}$$

  have rank 1.
- $\overline{O}_5 = O_5 = \{0\}$.

The proof is straightforward, using the orbit classification of [Bürgstein and Hesselink 1987, Table 2]. See [Anderson 2009, Section 5.2] for details. Each of these orbit closures is Cohen–Macaulay, as may be checked easily from the equations.

From the description in terms of cubic polynomials, it is easy to find representatives for orbits in U'. Here we give representatives as weight vectors in $\mathfrak{g}/\mathfrak{p}$. Let $Y_\alpha \in \mathfrak{g}_2/\mathfrak{p}$ be a weight vector for α. We have

$$\begin{aligned}
O_0 &= P \cdot Y_{-3\alpha_1-2\alpha_2} = U \smallsetminus U',\\
O_1 &= P \cdot (Y_{-3\alpha_1-\alpha_2} + Y_{-\alpha_2}) \cong P/P_u \cong \mathrm{GL}_2,\\
O_2 &= P \cdot Y_{-\alpha_1-\alpha_2},\\
O_3 &= P \cdot Y_{-\alpha_2},\\
O_5 &= \{0\}.
\end{aligned}$$

Using Proposition 5.1, it is a simple matter to compute the equivariant classes.

Theorem 5.2. *In* $H_T^*(U) = \mathbb{Z}[\alpha_1, \alpha_2] = \mathbb{Z}[t_1, t_2]$, *we have*

$$\begin{aligned}
[\overline{O}_0] &= 1,\\
[\overline{O}_1] &= -3\alpha_1 - 2\alpha_2 = -t_1 - t_2,\\
[\overline{O}_2] &= 2(-3\alpha_1 - 2\alpha_2)^2 = 2(t_1+t_2)^2,\\
[\overline{O}_3] &= -3(\alpha_1+\alpha_2)(2\alpha_1+\alpha_2)(3\alpha_1+2\alpha_2)\\
&= -3t_1t_2(t_1+t_2),\\
[\overline{O}_5] &= -\alpha_2(\alpha_1+\alpha_2)(2\alpha_1+\alpha_2)(3\alpha_1+\alpha_2)(3\alpha_1+2\alpha_2)\\
&= t_1t_2(t_1+t_2)(2t_1-t_2)(t_1-2t_2).
\end{aligned}$$

Proof. The normal space to $U' = \overline{O}_1 \subset U$ has weight $-3\alpha_1 - 2\alpha_2$, so the formula for $[\overline{O}_1]$ is clear. Since the restriction $H_T^*(U) \to H_T^*(U')$ is an isomorphism, the Gysin pushforward $H_T^*(U') \to H_T^*(U)$ is multiplication by $[U']$. Therefore it suffices to compute the remaining classes in $H_T^*(U')$. The locus $\overline{O}_2$ is a hypersurface in U' defined by an equation of weight $-6\alpha_1 - 4\alpha_2$, so its class in $H_T^*(U)$ is $(-6\alpha_1 - 4\alpha_2) \cdot [U']$. The class of $[\overline{O}_3]$ in $H_T^*(U')$ is found by the classical Giambelli (or Salmon–Roberts) formula; see for example [Fulton and Pragacz 1998, Section 1.1]. Finally, the class of the origin is the product of all the T-weights on U. □

Remark 5.3. These classes cannot be computed using the "restriction equation" method of Fehér and Rimányi [2004] because the stabilizer of $O_1 = P/P_u$ is unipotent. This means the restriction map $H_P^*(U) \to H_P^*(O_1) \cong H_{P_u}^*(\mathrm{pt}) = H^*(\mathrm{pt})$ is zero in positive degrees, and all the restriction equations are of the form $0 = 0$. The problem persists for the other orbits.

Lemma 5.4. *The orbit* $O_3 \subset \mathfrak{g}/\mathfrak{p} \subset \mathrm{Hom}(E, \mathrm{End}(E) \oplus E^*)$ *consists of the triality-symmetric morphisms of rank* 1.

Proof. Any rank-1 map φ must correspond to an element $\varphi_1 \oplus \varphi_2 \in U = U' \oplus \bigwedge^2 E^*$ with $\varphi_2 = 0$, that is, φ lies in U'. (If $\varphi_2 \neq 0$, then φ surjects onto E^*.)

The action of P on U' is the same as that of its Levi subgroup GL_2. Let $P_{\hat{2}} \subset \mathrm{GL}_8$ be the parabolic which stabilizes E. The inclusion $P \hookrightarrow P_{\hat{2}} \subset \mathrm{GL}_8$ induces an inclusion of Levi subgroups $\mathrm{GL}_2 = \mathrm{GL}_2 \times 1 \hookrightarrow \mathrm{GL}_2 \times \mathrm{GL}_6$, and the latter group acts on $\mathrm{Hom}(E, \mathrm{End}(E) \oplus E^*)$ by left-right matrix multiplication,[4] so it preserves ranks of morphisms. Therefore it will suffice to check that a representative for O_2 has rank 2, and a representative from O_3 has rank 1.

For these, we use the coordinate description given in Section 3. Under the identification of U' with the space of cubic polynomials, the monomial xy^2 corresponds to the basis vector v^*_{122}. The orbit is O_2 (since xy^2 has two distinct zeroes), and the corresponding matrix A_φ has $b = c = d = 0$ and $a \neq 0$; it is easy to see this means φ has rank 2. Similarly, x^3 corresponds to v^*_{111}, and the corresponding A_φ has $a = b = d = 0$ and $c \neq 0$, so φ has rank 1. □

The formulas of Theorem 1.2 now follow from those of Theorem 5.2.

Second proof of Theorem 1.2. Let $f : X \to \mathrm{BGL}_2$ be the map defined (up to homotopy) by the given vector bundle E on X. The corresponding map

$$f^* : H^* \mathrm{BGL}_2 = H^*_{\mathrm{GL}_2}(\mathrm{pt}) = \mathbb{Z}[c_1, c_2] \to H^* X$$

is given by $c_i \mapsto c_i(E) = (-1)^i c_i(E^*)$. Equivalently, using the inclusion $H^*_{\mathrm{GL}_2}(\mathrm{pt}) \subset H^*_T(\mathrm{pt}) = \mathbb{Z}[t_1, t_2]$ and Chern roots x_1, x_2 for E^*, the map is given by $t_i \mapsto -x_i$.

Using Lemma 5.4, we have $f^{-1}\overline{O}_3 = D_1(\varphi)$, so by [Fulton and Pragacz 1998, p. 108] and the fact that $\overline{O}_3$ is Cohen–Macaulay, we obtain $f^*[\overline{O}_3] = [D_1(\varphi)]$ when $D_1(\varphi)$ has expected codimension. □

Remark 5.5. The proof of Theorem 1.2 given in Section 4 works verbatim for Chow cohomology. The proof in this section also works, though to apply equivariant techniques, one needs to take extra care to ensure that the bundle E is pulled back from an algebraic approximation to the classifying space. To achieve this, one can replace X with an appropriate composition of an affine bundle and a Chow envelope; see [Graham 1997, p. 486] for the argument.

Appendix: Octonion bundles

There is a G_2 analogue of the well-known fact that for any vector bundle E, the direct sum $E \oplus E^*$ carries canonical symplectic (type C) and symmetric (type D) forms; see for example [Fulton and Pragacz 1998, p. 71]. The intrinsic construction

[4] Identifying $\mathrm{Hom}(E, \mathrm{End}(E) \oplus E^*)$ with 6×2 matrices, the action is by $(g, h) \cdot A = hAg^{-1}$. This is the action induced by restricting the conjugation action of GL_8 on 8×8 matrices when the subspace of 6×2 matrices is placed in the lower-left corner.

presented here seems to appear first in [Landsberg and Manivel 2006, p. 151]; it is closely related to the Cayley–Dickson doubling construction [Petersson 1993].

We fix some notation. For any vector bundle E, let

$$\mathrm{Tr} : \mathrm{End}(E) = E^* \otimes E \to \mathcal{O}_X$$

be the canonical contraction map, and let

$$\mathrm{End}^0(E) = \ker(\mathrm{Tr}) \subset \mathrm{End}(E)$$

be the subbundle of trace-zero endomorphisms. Let $e : \mathcal{O}_X \to \mathrm{End}(E)$ be the identity section. Thus the composition $\mathrm{Tr} \circ e : \mathcal{O}_X \to \mathcal{O}_X$ is multiplication by $\mathrm{rk}(E)$. Also, when E has rank 2, the *conjugation* map $\mathrm{End}(E) \to \mathrm{End}(E)$ is given by $e \circ \mathrm{Tr} - \mathrm{id}$. (Here id is the identity morphism, as opposed to the identity section e.) Conjugation is an involution; locally, it is $\xi \mapsto \overline{\xi} := \mathrm{Tr}(\xi)e - \xi$.

The norm on an octonion bundle C corresponds to a nondegenerate symmetric bilinear form $\langle \cdot , \cdot \rangle$. Let $V \subset C$ be the orthogonal complement to the identity subbundle defined by e. A subbundle $E \subset C$ is *G_2-isotropic* if it is contained in V and the multiplication map $E \otimes E \to C$ is the zero map.

Proposition A.1 (cf. [Landsberg and Manivel 2006, p. 151]). *Let E be a rank-2 vector bundle on a variety X. Then $C = E \oplus \mathrm{End}(E) \oplus E^*$ has a canonical octonion bundle structure with identity section $e : \mathcal{O}_X \to \mathrm{End}(E) \subset C$. The subbundle $E = E \oplus 0 \oplus 0 \subset C$ is G_2-isotropic.*

More specifically, there is a quadratic norm $N : C \to \mathcal{O}_X$ and bilinear multiplication $m : C \otimes C \to C$ for $C = E \oplus \mathrm{End}(E) \oplus E^*$ which are compatible. The norm corresponds to the bilinear form $\langle \cdot , \cdot \rangle$ defined by

$$\langle x \oplus \xi \oplus f,\ y \oplus \eta \oplus g \rangle = \mathrm{Tr}(\xi)\,\mathrm{Tr}(\eta) - \mathrm{Tr}(\xi\eta) - f(y) - g(x). \tag{1}$$

The multiplication is given by

$$(x \oplus \xi \oplus f) \cdot (y \oplus \eta \oplus g) = (\eta x + \overline{\xi} y) \oplus (\overline{g \otimes x} + \xi\eta + f \otimes y) \oplus (g\xi + f\overline{\eta}). \tag{2}$$

One only needs to verify compatibility of the norm with multiplication; see [Anderson 2009, Section 2.4] for details.

Acknowledgments

This work is part of my Ph.D. thesis, and it is a pleasure to thank William Fulton for his encouragement in this project. Thanks also to Danny Gillam for conversations about triality, and the referees for helpful comments on the manuscript.

References

[Anderson 2009] D. Anderson, *Degeneracy loci and G_2 flags*, Ph.D. thesis, University of Michigan, 2009, Available at http://www.math.washington.edu/∼dandersn/papers/thesis.pdf.

[Anderson 2011] D. Anderson, "Chern class formulas for G_2 Schubert loci", *Trans. Amer. Math. Soc.* **363**:12 (2011), 6615–6646. MR 2012g:14101 Zbl 1234.14037

[van der Blij and Springer 1960] F. van der Blij and T. A. Springer, "Octaves and triality", *Nieuw Arch. Wisk.* (3) **8** (1960), 158–169. MR 23 #A947 Zbl 0127.11804

[Bürgstein and Hesselink 1987] H. Bürgstein and W. H. Hesselink, "Algorithmic orbit classification for some Borel group actions", *Compositio Math.* **61** (1987), 3–41. MR 88k:20069 Zbl 0612.17005

[Cartan 1925] E. Cartan, "Le principe de dualité et la théorie des groupes simples et semi-simples", *Bull. Sci. Math.* **49** (1925), 361–374. JFM 51.0322.02

[Fehér and Rimányi 2003] L. M. Fehér and R. Rimányi, "Schur and Schubert polynomials as Thom polynomials — cohomology of moduli spaces", *Cent. Eur. J. Math.* **1**:4 (2003), 418–434. MR 2005b:05219 Zbl 1038.57008

[Fehér and Rimányi 2004] L. M. Fehér and R. Rimányi, "Calculation of Thom polynomials and other cohomological obstructions for group actions", pp. 69–93 in *Real and complex singularities*, edited by T. Gaffney and M. A. S. Ruas, Contemp. Math. **354**, Amer. Math. Soc., Providence, RI, 2004. MR 2005j:58052 Zbl 1074.32008

[Fehér et al. 2005] L. M. Fehér, A. Némethi, and R. Rimányi, "Degeneracy of 2-forms and 3-forms", *Canad. Math. Bull.* **48**:4 (2005), 547–560. MR 2006h:53078 Zbl 1087.58012

[Fulton 1996] W. Fulton, "Determinantal formulas for orthogonal and symplectic degeneracy loci", *J. Differential Geom.* **43**:2 (1996), 276–290. MR 98d:14004 Zbl 0911.14001

[Fulton 1998] W. Fulton, *Intersection theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics] **2**, Springer, Berlin, 1998. MR 99d:14003 Zbl 0885.14002

[Fulton and Pragacz 1998] W. Fulton and P. Pragacz, *Schubert varieties and degeneracy loci*, Lecture Notes in Mathematics **1689**, Springer, Berlin, 1998. MR 99m:14092 Zbl 0913.14016

[Garibaldi 1999] R. S. Garibaldi, "Twisted flag varieties of trialitarian groups", *Comm. Algebra* **27**:2 (1999), 841–856. MR 2000a:20103 Zbl 0986.16006

[Graham 1997] W. Graham, "The class of the diagonal in flag bundles", *J. Differential Geom.* **45**:3 (1997), 471–487. MR 98j:14070 Zbl 0935.14015

[Harris and Tu 1984] J. Harris and L. W. Tu, "On symmetric and skew-symmetric determinantal varieties", *Topology* **23**:1 (1984), 71–84. MR 85c:14032 Zbl 0534.55010

[Hille and Röhrle 1999] L. Hille and G. Röhrle, "A classification of parabolic subgroups of classical groups with a finite number of orbits on the unipotent radical", *Transform. Groups* **4**:1 (1999), 35–52. MR 2000f:20072 Zbl 0924.20035

[Humphreys 1975] J. E. Humphreys, *Linear algebraic groups*, Graduate Texts in Mathematics **21**, Springer, New York, 1975. MR 53 #633 Zbl 0325.20039

[Józefiak et al. 1981] T. Józefiak, A. Lascoux, and P. Pragacz, "Classes of determinantal varieties associated with symmetric and skew-symmetric matrices", *Izv. Akad. Nauk SSSR Ser. Mat.* **45**:3 (1981), 662–673. In Russian; translated in *Mathematics of the USSR-Izvestiya* **18**:3 (1982), 575–586. MR 83h:14044 Zbl 0471.14028

[Jürgens and Röhrle 2002] U. Jürgens and G. Röhrle, "MOP—algorithmic modality analysis for parabolic group actions", *Experiment. Math.* **11**:1 (2002), 57–67. MR 2004a:20050

[Knus et al. 1998] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol, *The book of involutions*, American Mathematical Society Colloquium Publications **44**, American Mathematical Society, Providence, RI, 1998. MR 2000a:16031 Zbl 0955.16001

[Knutson and Miller 2005] A. Knutson and E. Miller, "Gröbner geometry of Schubert polynomials", *Ann. of Math.* (2) **161**:3 (2005), 1245–1318. MR 2006i:05177 Zbl 1089.14007

[Landsberg and Manivel 2006] J. M. Landsberg and L. Manivel, "The sextonions and $E_{7\frac{1}{2}}$", *Adv. Math.* **201**:1 (2006), 143–179. MR 2006k:17017 Zbl 1133.17007

[Petersson 1993] H. P. Petersson, "Composition algebras over algebraic curves of genus zero", *Trans. Amer. Math. Soc.* **337**:1 (1993), 473–493. MR 93g:17006 Zbl 0778.17001

[Popov and Röhrle 1997] V. Popov and G. Röhrle, "On the number of orbits of a parabolic subgroup on its unipotent radical", pp. 297–320 in *Algebraic groups and Lie groups*, edited by G. Lehrer, Austral. Math. Soc. Lect. Ser. **9**, Cambridge Univ. Press, 1997. MR 99f:14063 Zbl 0887.14020

[Springer and Veldkamp 2000] T. A. Springer and F. D. Veldkamp, *Octonions, Jordan algebras and exceptional groups*, Springer Monographs in Mathematics, Springer, Berlin, 2000. MR 2001f:17006 Zbl 1087.17001

Communicated by Ravi Vakil
Received 2010-09-27 Revised 2011-06-10 Accepted 2011-08-13

dandersn@math.washington.edu *Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, United States*

Department of Mathematics, University of Washington, Seattle, WA 98195, United States

ALGEBRA AND NUMBER THEORY 6:4(2012)

Multi-Frey $\mathbb{Q}$-curves and the Diophantine equation $a^2 + b^6 = c^n$

Michael A. Bennett and Imin Chen

We show that the equation $a^2+b^6=c^n$ has no nontrivial positive integer solutions with $(a, b) = 1$ via a combination of techniques based upon the modularity of Galois representations attached to certain $\mathbb{Q}$-curves, corresponding surjectivity results of Ellenberg for these representations, and extensions of multi-Frey curve arguments of Siksek.

1. Introduction

Following the proof of Fermat's last theorem [Wiles 1995], there has developed an extensive literature on connections between the arithmetic of modular abelian varieties and classical Diophantine problems, much of it devoted to solving *generalized Fermat equations* of the shape

$$a^p + b^q = c^r, \quad \frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1, \tag{1}$$

in coprime integers a, b, and c, and positive integers p, q, and r. That the number of such solutions (a, b, c) is finite, for a fixed triple (p, q, r), is a consequence of [Darmon and Granville 1995]. It has been conjectured that there are in fact at most finitely many such solutions, even when we allow the triples (p, q, r) to vary, provided we count solutions corresponding to $1^p + 2^3 = 3^2$ only once. Being extremely optimistic, one might even believe that the known solutions constitute a complete list, namely (a, b, c, p, q, r) corresponding to $1^p + 2^3 = 3^2$, for $p \geq 7$, and to nine other identities (see [Darmon and Granville 1995; Beukers 1998]):

$$2^5 + 7^2 = 3^4, \quad 7^3 + 13^2 = 2^9, \quad 2^7 + 17^3 = 71^2, \quad 3^5 + 11^4 = 122^2,$$
$$17^7 + 76271^3 = 21063928^2, \; 1414^3 + 2213459^2 = 65^7, \; 9262^3 + 15312283^2 = 113^7,$$
$$43^8 + 96222^3 = 30042907^2, \quad \text{and} \quad 33^8 + 1549034^2 = 15613^3.$$

Research supported by NSERC.
MSC2010: primary 11D41; secondary 11D61, 11G05, 14G05.
Keywords: Fermat equations, Galois representations, $\mathbb{Q}$-curves, multi-Frey techniques.

(For brevity, we omit listing the solutions which differ only by sign changes and permutation of coordinates: for instance, if p is even, $(-1)^p + 2^3 = 3^2$, etc.)

Since all known solutions have $\min\{p, q, r\} < 3$, a closely related formulation is that there are no nontrivial solutions in coprime integers once $\min\{p, q, r\} \geq 3$.

There are a variety of names associated to the above conjectures, including, alphabetically, Beal (see [Mauldin 1997]), Darmon and Granville [1995], van der Poorten, Tijdeman, and Zagier (see, for example, [Beukers 1998; Tijdeman 1989]), and apparently financial rewards have even been offered for their resolution, positive or negative.

Techniques based upon the modularity of Galois representations associated to putative solutions of (1) have, in many cases, provided a fruitful approach to these problems, though the limitations of such methods are still unclear. Each situation where finiteness results have been established for infinite families of triples (p, q, r) has followed along these lines. We summarize the results to date; in each case, no solutions outside those mentioned above have been discovered:

(p,q,r)	Reference(s)
(n,n,n), $n \geq 3$	[Wiles 1995; Taylor and Wiles 1995]
$(n,n,2)$, $n \geq 4$	[Darmon and Merel 1997; Poonen 1998]
$(n,n,3)$, $n \geq 3$	[Darmon and Merel 1997; Poonen 1998]
$(2n,2n,5)$, $n \geq 2$	[Bennett 2006]
$(2,4,n)$, $n \geq 4$	[Bruin 1999; Ellenberg 2004; Bennett et al. 2010]
$(2,n,4)$, $n \geq 4$	Immediate from [Bruin 2003; Bennett and Skinner 2004]
$(2,2n,k)$, $n \geq 2$, $k \in \{9,10,15\}$	[Bennett et al. $\geq$ 2012]
$(4,2n,3)$, $n \geq 2$	[Bennett et al. $\geq$ 2012]
$(2,n,6)$, $n \geq 3$	[Bennett et al. $\geq$ 2012]
$(3,3,n)$, $n \geq 3^*$	[Kraus 1998; Bruin 2000; Dahmen 2008; Chen and Siksek 2009]
$(3j,3k,n)$, $j,k,n \geq 2$	[Kraus 1998]
$(3,3,2n)$, $n \geq 2$	[Bennett et al. $\geq$ 2012]
$(3,6,n)$, $n \geq 2$	[Bennett et al. $\geq$ 2012]
$(2,2n,3)$, $n \geq 3^*$	[Bruin 1999; Chen 2008; Dahmen 2008; 2011; Siksek 2008]
$(2,2n,5)$, $n \geq 3^*$	[Chen 2010]
$(2,3,n)$, $6 \leq n \leq 10$	[Bruin 1999; 2003; 2005; Poonen et al. 2007; Siksek 2010; Brown 2012]

The $(*)$ here indicates that the result has been proven for a family of exponents of natural density one (but that there remain infinitely many cases of positive Dirichlet density untreated).

In this paper, we will prove the following theorem.

Theorem 1. *Let $n \geq 3$ be an integer. Then the equation*

$$a^2 + b^6 = c^n \tag{2}$$

has no solutions in positive integers a, b, and c, with a and b coprime.

Our motivations for considering this problem are two-fold. Firstly, the exponents $(2, 6, n)$ provide one of the final examples of an exponent family for which there is known to exist a corresponding family of Frey–Hellegouarch elliptic $\mathbb{Q}$-curves. A remarkable program for attacking the generalized Fermat equation of signature (n, n, m) (and perhaps others) is outlined in [Darmon 2000], relying upon the construction of Frey–Hellegouarch abelian varieties. Currently, however, it does not appear that the corresponding technology is suitably advanced to allow the application of such arguments to completely solve families of such equations for fixed $m \geq 5$.

In some sense, the signatures $(2, 6, n)$ and $(2, n, 6)$ (the latter equations are treated in [Bennett et al. $\geq$ 2012]) represent the final remaining families of generalized Fermat equations approachable by current techniques. More specifically, as discussed in [Darmon and Granville 1995], associated to a generalized Fermat equation $x^p + y^q = z^r$ is a triangle Fuchsian group with signature $(1/p, 1/q, 1/r)$. A reasonable precondition to applying the modular method using rational elliptic curves or $\mathbb{Q}$-curves is that this triangle group be commensurable with the full modular group. Such a classification has been performed in [Takeuchi 1977], where it is shown that the possible signatures containing ∞ are $(2, 3, \infty)$, $(2, 4, \infty)$, $(2, 6, \infty)$, $(2, \infty, \infty)$, $(3, 3, \infty)$, $(3, \infty, \infty)$, $(4, 4, \infty)$, $(6, 6, \infty)$, (∞, ∞, ∞). A related classification of Frey representations for prime exponents can be found in [Darmon and Granville 1995; Darmon 2000]. The above list does not, admittedly, explain all the possible families of generalized Fermat equations that have been amenable to the modular method. In all other known cases, however, the Frey curve utilized is derived from a descent step to one of the above "pure" Frey curve families. Concerning the applicability of using certain families of $\mathbb{Q}$-curves, see the conclusions section of [Chen 2010] for further remarks.

Our secondary motivation is as an illustration of the utility of the multi-Frey techniques of S. Siksek (see [Bugeaud et al. 2008a; 2008b]). A fundamental difference between the case of signature $(2, 4, n)$ considered in [Ellenberg 2004] and that of $(2, 6, n)$ is the existence, in this latter situation, of a potential obstruction to our arguments in the guise of a particular modular form *without* complex multiplication. To eliminate the possibility of a solution to the equation $x^2 + y^6 = z^n$ arising from this form requires fundamentally new techniques, based upon a generalization of the multi-Frey technique to $\mathbb{Q}$-curves (rather than just curves over $\mathbb{Q}$).

The computations in this paper were performed in MAGMA [Bosma et al. 1997]. The programs, data, and output files are posted in this paper's Electronic Supplement and at http://people.math.sfu.ca/ ichen/firstb3i-data. Throughout the text, we have included specific references to the MAGMA code employed, indicated as sample.txt.

2. Review of $\mathbb{Q}$-curves and their attached Galois representations

The exposition of $\mathbb{Q}$-curves and their attached Galois representations we provide in this section closely follows that of [Ribet 1992; Quer 2000; Ellenberg and Skinner 2001; Chen 2012]; we include it in the interest of keeping our exposition reasonably self-contained.

Let K be a number field and C/K be a non-CM elliptic curve such that there is an isogeny $\mu(\sigma) : {}^{\sigma}C \to C$ defined over K for each $\sigma \in G_{\mathbb{Q}}$. Such a curve C/K is called a $\mathbb{Q}$-curve defined over K. Let $\hat{\phi}_{C,p} : G_K \to \mathrm{GL}_2(\mathbb{Z}_p)$ be the representation of G_K on the Tate module $\hat{V}_p(C)$. One can attach a representation

$$\hat{\rho}_{C,\beta,p} : G_{\mathbb{Q}} \to \overline{\mathbb{Q}}_p^* \,\mathrm{GL}_2(\mathbb{Q}_p)$$

to C such that $\mathbb{P}\hat{\rho}_{C,\beta,p}|_{G_K} \cong \mathbb{P}\hat{\phi}_{C,p}$. The representation depends on the choice of splitting map β (in what follows, we will provide more details of this choice). Let π be a prime above p of the field M_β generated by the values of β. In practice, the representation $\hat{\rho}_{C,\beta,\pi}$ is constructed in a way so that its image lies in $M_{\beta,\pi}^* \,\mathrm{GL}_2(\mathbb{Q}_p)$, and we choose to use the notation $\hat{\rho}_{C,\beta,p} = \hat{\rho}_{C,\beta,\pi}$ to indicate the choice of π in this explicit construction.

Let

$$c_C(\sigma, \tau) = \mu_C(\sigma)\,{}^{\sigma}\mu_C(\tau)\mu_C(\sigma\tau)^{-1} \in (\mathrm{Hom}_K(C, C) \otimes_{\mathbb{Z}} \mathbb{Q})^* = \mathbb{Q}^*,$$

where $\mu_C^{-1} := (1/\deg \mu_C)\mu_C'$ and μ_C' is the dual of μ_C. Then $c_C(\sigma, \tau)$ determines a class in $H^2(G_{\mathbb{Q}}, \mathbb{Q}^*)$ which depends only on the $\overline{\mathbb{Q}}$-isogeny class of C. The class $c_C(\sigma, \tau)$ factors through $H^2(G_{K/\mathbb{Q}}, \mathbb{Q}^*)$, depending now only on the K-isogeny class of C. Alternatively,

$$c_C(\sigma, \tau) = \alpha(\sigma)\,{}^{\sigma}\alpha(\tau)\alpha(\sigma\tau)^{-1}$$

arises from a class $\alpha \in H^1(G_{\mathbb{Q}}, \overline{\mathbb{Q}}^*/\mathbb{Q}^*)$ through the map

$$H^1(G_{\mathbb{Q}}, \overline{\mathbb{Q}}^*/\mathbb{Q}^*) \to H^2(G_{\mathbb{Q}}, \mathbb{Q}^*),$$

resulting from the short exact sequence

$$1 \to \mathbb{Q}^* \to \overline{\mathbb{Q}}^* \to \overline{\mathbb{Q}}^*/\mathbb{Q}^* \to 1.$$

Explicitly, $\alpha(\sigma)$ is defined by $\sigma^*(\omega_C) = \alpha(\sigma)\omega_C$.

Tate showed that $H^2(G_{\mathbb{Q}}, \overline{\mathbb{Q}}^*)$ is trivial where the action of $G_{\mathbb{Q}}$ on $\overline{\mathbb{Q}}^*$ is trivial. Thus, there is a continuous map $\beta : G_{\mathbb{Q}} \to \overline{\mathbb{Q}}^*$ such that

$$c_C(\sigma, \tau) = \beta(\sigma)\beta(\tau)\beta(\sigma\tau)^{-1}$$

as cocycles, and we call β a splitting map for c_C. We define

$$\hat{\rho}_{C,\beta,\pi}(\sigma)(1 \otimes x) = \beta(\sigma)^{-1} \otimes \mu_C(\sigma)(\sigma(x)).$$

Given a splitting $c_C(\sigma, \tau) = \beta(\sigma)\beta(\tau)\beta(\sigma\tau)^{-1}$, Ribet attaches an abelian variety A_β defined over $\mathbb{Q}$, of GL_2-type and having C as a simple factor over $\overline{\mathbb{Q}}$.

In practice, what we do in this paper is find a continuous $\beta : G_{\mathbb{Q}} \to \overline{\mathbb{Q}}^*$, factoring over an extension of low degree, such that $c_C(\sigma, \tau) = \beta(\sigma)\beta(\tau)\beta(\sigma\tau)^{-1}$ as elements in $H^2(G_{\mathbb{Q}}, \overline{\mathbb{Q}}^*)$, using results in [Quer 2000]. Then we choose a suitable twist C_β/K_β of C, where K_β is the splitting field of β, such that $c_{C_\beta}(\sigma, \tau)$ is exactly the cocycle $c_\beta(\sigma, \tau) = \beta(\sigma)\beta(\tau)\beta(\sigma\tau)^{-1}$. In this situation, the abelian variety A_β is constructed as a quotient over $\mathbb{Q}$ of $\mathrm{Res}_{\mathbb{Q}}^{K_\beta} C_\beta$.

The endomorphism algebra of A_β is given by $M_\beta = \mathbb{Q}(\{\beta(\sigma)\})$ and the representation on the π^n-torsion points of A_β coincides with the representation $\hat{\rho}_{C,\beta,\pi}$ defined earlier.

Let $\epsilon : G_{\mathbb{Q}} \to \overline{\mathbb{Q}}^*$ be defined by

$$\epsilon(\sigma) = \beta(\sigma)^2 / \deg \mu(\sigma). \tag{3}$$

Then ϵ is a character and

$$\det \hat{\rho}_{C,\beta,\pi} = \epsilon^{-1} \cdot \chi_p, \tag{4}$$

where $\chi_p : G_{\mathbb{Q}} \to \mathbb{Z}_p^*$ is the p-th cyclotomic character.

3. $\mathbb{Q}$-curves attached to $a^2 + b^6 = c^p$ and their Galois representations

Let $(a, b, c) \in \mathbb{Z}^3$ be a solution to $a^2 + b^6 = c^p$ where we suppose that p is a prime. We call (a, b, c) *proper* if $\gcd(a, b, c) = 1$ and *trivial* if $|c| = 1$. Note that a solution $(a, b, c) \in \mathbb{Z}^3$ is proper if and only if the integers a, b, and c are pairwise coprime. In what follows, we will always assume that the triple (a, b, c) is a proper, nontrivial solution. We consider the following associated (Frey or Frey–Hellegouarch) elliptic curve:

$$E : Y^2 = X^3 - 3(5b^3 + 4ai)bX + 2(11b^6 + 14ib^3a - 2a^2),$$

with j-invariant

$$j = 432i \frac{b^3(4a - 5ib^3)^3}{(a - ib^3)(a + ib^3)^3} \tag{5}$$

and discriminant $\Delta = -2^8 \cdot 3^3 \cdot (a - ib^3) \cdot (a + ib^3)^3$.

Lemma 2. *Suppose $a/b^3 \in \mathbb{P}^1(\mathbb{Q})$. Then the j-invariant of E does not lie in $\mathbb{Q}$ except when*

- $a/b^3 = 0$ *and* $j = 54000$, *or*
- $a/b^3 = \infty$ *and* $j = 0$.

Proof. This can be seen by solving the polynomial equation in $\mathbb{Q}[i][j, a/b^3]$ derived from (5) by clearing the denominators and collecting terms with respect to $\{1, i\}$, using the restriction that $j, a/b^3 \in \mathbb{P}^1(\mathbb{Q})$. □

Corollary 3. *E does not have complex multiplication unless*

- $a/b^3 = 0$, $j = 54000$, *and* $d(\mathbb{O}) = -12$, *or*
- $a/b^3 = \infty$, $j = 0$, *and* $d(\mathbb{O}) = -3$.

Proof. If E has complex multiplication by an order $\mathbb{O}$ in an imaginary quadratic field, then $j(E)$ has a real conjugate over $\mathbb{Q}$ (for instance, arising from $j(E_0)$, where E_0 is the elliptic curve associated to the lattice $\mathbb{O}$ itself). Hence, $j(E) \in \mathbb{Q}$ in fact. For a list of the j-invariants of elliptic curves with complex multiplication by an order of class number 1, see, for instance, [Cox 1989, p. 261]. □

Lemma 4. *If $(a, b, c) \in \mathbb{Z}^3$ with $\gcd(a, b, c) = 1$ and $a^2 + b^6 = c^p$, then either $c = 1$ or c is divisible by a prime not equal to 2 or 3.*

Proof. The condition $\gcd(a, b, c) = 1$ together with inspection of $a^2 + b^6$ modulo 3 shows that c is never divisible by 3. Similar reasoning allows us to conclude, since $p > 1$, that c is necessarily odd, whereby the lemma follows. □

From here on, let us suppose that E arises from a nontrivial proper solution to $a^2+b^6=c^p$ where p is an odd prime. Note that ab is even and, from the preceding discussion, that $a - b^3 i$ and $a + b^3 i$ are coprime p-th powers in $\mathbb{Z}[i]$.

The elliptic curve E is defined over $\mathbb{Q}(i)$. Its conjugate over $\mathbb{Q}(i)$ is 3-isogenous to E over $\mathbb{Q}(\sqrt{3}, i)$; see isogeny.txt. This is in contrast to the situation in [Ellenberg 2004], where the corresponding isogeny is defined over $\mathbb{Q}(i)$. We make a choice of isogenies $\mu(\sigma) : {}^{\sigma}E \to E$ such that $\mu(\sigma) = 1$ for $\sigma \in G_{\mathbb{Q}(i)}$ and $\mu(\sigma)$ is the 3-isogeny above when $\sigma \notin G_{\mathbb{Q}(i)}$.

Let $d(\sigma)$ denote the degree of $\mu(\sigma)$. We have $d(G_{\mathbb{Q}}) = \{1, 3\} \subseteq \mathbb{Q}^*/\mathbb{Q}^{*2}$. The fixed field K_d of the kernel of $d(\sigma)$ is $\mathbb{Q}(i)$ and so $(a, d) = (-1, 3)$ is a dual basis in the terminology of [Quer 2000]. The quaternion algebra $(-1, 3)$ is ramified at 2, 3 and so a choice of splitting character for $c_E(\sigma, \tau)$ is given by $\epsilon = \epsilon_2\epsilon_3$ where ϵ_2 is the nontrivial character of $\mathbb{Z}/4\mathbb{Z}^\times$ and ϵ_3 is the nontrivial character of $\mathbb{Z}/3\mathbb{Z}^\times$. The fixed field of ϵ is $K_\epsilon = \mathbb{Q}(\sqrt{3})$.

Let $G_{\mathbb{Q}(i)/\mathbb{Q}} = \{\sigma_1, \sigma_{-1}\}$. We have that

$$\alpha(\sigma_1) = 1 \quad \text{and} \quad \alpha(\sigma_{-1}) = i\sqrt{3}.$$

This can be checked by noting that the quotient of E by the 3-torsion point of E using Vélu multiplies the invariant differential by 1. The resulting quotient elliptic curve is then a twist over $\mathbb{Q}(\sqrt{3}, i)$ of the original E. This twisting multiplies the invariant differential by $i\sqrt{3}$.

So now we can express $c_E(\sigma, \tau) = \alpha(\sigma)^{\sigma}\alpha(\tau)\alpha(\sigma\tau)^{-1}$. Let $\beta(\sigma) = \sqrt{\epsilon(\sigma)}\sqrt{d(\sigma)}$ and $c_\beta(\sigma, \tau) = \beta(\sigma)\beta(\tau)\beta(\sigma\tau)^{-1} \in H^2(G_{\mathbb{Q}}, \mathbb{Q}^*)$. We know from [Quer 2000] that $c_\beta(\sigma, \tau)$ and $c_E(\sigma, \tau)$ represent the same class in $H^2(G_{\mathbb{Q}}, \mathbb{Q}^*)$. The fixed field of β is $K_\beta = K_\epsilon \cdot K_d = \mathbb{Q}(\sqrt{3}, i)$ and $M_\beta = \mathbb{Q}(\sqrt{3}, i)$.

Our goal is to find a $\gamma \in \overline{\mathbb{Q}}^*$ such that $c_\beta(\sigma, \tau) = \alpha_1(\sigma)^{\sigma}\alpha_1(\tau)\alpha_1(\sigma\tau)^{-1}$, where $\alpha_1(\sigma) = \alpha(\sigma)\,{}^{\sigma}(\sqrt{\gamma})/\sqrt{\gamma}$. Using a similar technique as for the equation $a^2 + b^{2p} = c^5$ (compare [Chen 2010], where the corresponding K_β is cyclic quartic), we can narrow down the possibilities for choices of γ and subsequently verify that a particular choice actually works.

In more detail, recall that $K_\beta = \mathbb{Q}(\sqrt{3}, i) = \mathbb{Q}(z)$, where $z = (i + \sqrt{3})/2$ is a primitive twelfth root of unity. Let $G_{\mathbb{Q}(\sqrt{3},i)/\mathbb{Q}} = \{\sigma_1, \sigma_{-1}, \sigma_3, \sigma_{-3}\}$ and assume that $\alpha_1(\sigma_{-3})^2/\alpha(\sigma_{-3})^2 = \alpha_1(\sigma_{-3})^2/-3$ is a unit, say 1. This implies that ${}^{\sigma_{-3}}\gamma/\gamma = 1$, whereby $\gamma \in \mathbb{Q}(\sqrt{-3})$. Furthermore, let us assume that ${}^{\sigma_{-1}}\gamma/\gamma$ is a square in K_β of a unit in $\mathbb{Q}(\sqrt{-3})$, say z^2 (the other choices produce isomorphic twists). Solving for γ, we obtain that $\gamma = (-3 + i\sqrt{3})/2$ is one possible choice.

The resulting values of $\alpha_2(\sigma) = \alpha(\sigma)\sqrt{{}^{\sigma}\gamma/\gamma}$ are

$$\alpha_2(\sigma_1) = 1, \quad \alpha_2(\sigma_{-1}) = i\sqrt{3}z, \quad \alpha_2(\sigma_3) = z, \quad \text{and} \quad \alpha_2(\sigma_{-3}) = i\sqrt{3},$$

where we have fixed a choice of square root for each $\sigma \in G_{K/\mathbb{Q}}$. It can be verified that $c_\beta(\sigma, \tau) = \alpha_2(\sigma)^{\sigma}\alpha_2(\tau)\alpha_2(\sigma\tau)^{-1}$.

Consider the twist E_β of E given by the equation

$$E_\beta : Y^2 = X^3 - 3(5b^3 + 4ai)b\gamma^2 X + 2(11b^6 + 14ib^3a - 2a^2)\gamma^3. \tag{6}$$

From the relationship between E_β and E, the initial $\mu(\sigma)$'s for E give rise to choices for $\mu_\beta(\sigma)$ for E_β which are, in general, locally constant on a smaller subgroup than G_K. For these choices of $\mu_\beta(\sigma)$ we have

$$\alpha_{E_\beta}(\sigma) = \alpha_1(\sigma) = \alpha(\sigma)\,{}^{\sigma}(\sqrt{\gamma})/\sqrt{\gamma}.$$

Now, $\sqrt{{}^{\sigma}\gamma/\gamma} = \xi(\sigma)\delta(\sigma)$ where $\delta(\sigma) = {}^{\sigma}(\sqrt{\gamma})/\sqrt{\gamma}$ and $\xi(\sigma) = \pm 1$. Since $\delta(\sigma)^{\sigma}\delta(\tau)\delta(\sigma\tau)^{-1} = 1$, it follows that $c_{E_\beta}(\sigma, \tau) = c_\beta(\sigma, \tau)\xi(\sigma)\xi(\tau)\xi(\sigma\tau)^{-1}$. Hence, by using the alternate set of isogenies $\mu'_\beta(\sigma) = \mu_\beta(\sigma)\xi(\sigma)$, which are now locally constant on G_K, the corresponding $\alpha_{E_\beta}(\sigma) = \alpha(\sigma)\sqrt{{}^{\sigma}\gamma/\gamma} = \alpha_2(\sigma)$, and hence $c_{E_\beta}(\sigma, \tau) = \alpha_2(\sigma)^{\sigma}\alpha_2(\tau)\alpha_2(\sigma\tau)^{-1} = c_\beta(\sigma, \tau)$ as cocycles, not just as classes in $H^2(G_{K/\mathbb{Q}}, \mathbb{Q}^*)$. The elliptic curve E_β/K_β is a $\mathbb{Q}$-curve defined over K_β; see isogenyp.txt.

Another way to motivate the preceding calculation is as follows. Without loss of generality, we may assume that γ is square-free in the ring of integers of K_β (if γ is a square, then the corresponding E_β is isomorphic over K_β to E). The field K_β has class number 1. If $\gamma = \lambda\gamma'$ where $\lambda \in \mathbb{Z}$, then using γ' instead of γ yields an E_β whose $c_{E_\beta}(\sigma, \tau)$ is the same cocycle in $H^2(G_{K/\mathbb{Q}}, \mathbb{Q}^*)$. The condition that $\sqrt{{}^\sigma\gamma/\gamma}$ be a square in K_β for all $\sigma \in G_{K/\mathbb{Q}}$ shows that only ramified primes divide γ and there are two such primes in $K_\beta = \mathbb{Q}(\sqrt{3}, i)$.

The discriminant of K_β is $d_{K_\beta/\mathbb{Q}} = 2^4 \cdot 3^2 = 144$. The prime factorizations of (2) and (3) in K_β are given by

$$(2) = \mathfrak{q}_2^2 \quad \text{and} \quad (3) = \mathfrak{q}_3^2.$$

Let ν_2 and ν_3 be uniformizers at $\mathfrak{q}_2$ and $\mathfrak{q}_3$ respectively with associated valuations v_2 and v_3. The units in K_β are generated by z of order 12 and a unit u_2 of infinite order. Thus, up to squares, γ is a product of a subset of the elements $\{z, u_2, \nu_2, \nu_3\}$.

The authors have subsequently learned that a similar technique for finding γ also appeared in [Dieulefait and Urroz 2009] (where K_β is polyquadratic).

It would be interesting to study the twists E_β which arise from other choices of splitting maps. We will not undertake this here.

Lemma 5. *Suppose that E and E' are elliptic curves defined by*

$$\begin{aligned} E{:}Y^2 + a_1XY + a_3Y &= X^3 + a_2X^2 + a_4X + a_6, \\ E'{:}Y^2 + a_1'XY + a_3'Y &= X^3 + a_2'X^2 + a_4'X + a_6', \end{aligned}$$

where the a_i and a_i' lie in a discrete valuation ring $\mathbb{O}$ with uniformizer ν.

(a) *Suppose the valuation at ν of the discriminants is, in each case, equal to* 12. *If E has reduction type II^* and $a_i' \equiv a_i \pmod{\nu^6}$, then E' also has reduction type II^*. If E has reduction type I_0 and $a_i' \equiv a_i \pmod{\nu^6}$, then E' also has reduction type I_0.*

(b) *Suppose the valuation at ν of the discriminants is, in each case, equal to* 16. *If E has reduction type II and $a_i' \equiv a_i \pmod{\nu^8}$, then E' also has reduction type II.*

(c) *Suppose the Weierstrass equation of E is in minimal form, and E has reduction type II or III. If $a_i' \equiv a_i \pmod{\nu^8}$, then E' has the same reduction type as E and is also in minimal form.*

Proof. We give a proof for case (a); the remaining cases are similar. Since the discriminants of E and E' have valuation 12, when E and E' are processed through Tate's algorithm [Silverman 1994], the algorithm terminates at one of Steps 1–10 or reaches Step 11 to loop back a second time at most once.

If E has reduction type II*, the algorithm applied to E terminates at Step 10. Since the transformations used in Steps 1–10 are translations, they preserve the congruence $a_i \equiv a_i' \pmod{\nu^6}$ as E and E' are processed through the algorithm, and since the conditions to exit at Steps 1–10 are congruence conditions modulo ν^6 on the coefficients of the Weierstrass equations, we see that if the algorithm applied to E terminates at Step 10, it must also terminate at Step 10 for E'.

If E has reduction type I_0, the algorithm applied to E reaches Step 11 to loop back a second time to terminate at Step 1 (because the valuation of the discriminant of the model for E is equal to 12). Again, since $a_i' \equiv a_i \pmod{\nu^6}$, it follows that the algorithm applied to E' also reaches Step 11 to loop back a second time and terminate at Step 1 (again because the valuation of the discriminant of the model for E' is equal to 12). □

Theorem 6. *The conductor of E_β is*

$$\mathfrak{m} = \mathfrak{q}_2^4\, \mathfrak{q}_3^\varepsilon \prod_{\substack{\mathfrak{q} \mid c \\ \mathfrak{q} \nmid 2,3}} \mathfrak{q},$$

where $\varepsilon = 0, 4$.

Proof. See tate2m.txt and tate3m.txt for the computations. Recall that E_β is given by

$$E_\beta : Y^2 = X^3 - 3(5b^3 + 4ai)b\gamma^2 X + 2(11b^6 + 14ib^3a - 2a^2)\gamma^3, \tag{7}$$

with

$$\Delta_{E_\beta} = -2^8 \cdot 3^3 \cdot (a - ib^3)(a + ib^3)^3 \cdot \gamma^6. \tag{8}$$

Then

$$\begin{aligned} c_4 &= 2^4 \cdot 3^2 \cdot b(4ia + 5b^3) \cdot \gamma^2 \\ c_6 &= 2^5 \cdot 3^3 \cdot (2a + (-7i - 6z^2 + 3)b^3)(2a + (-7i + 6z^2 - 3)b^3) \cdot \gamma^3. \end{aligned} \tag{9}$$

Let $\mathfrak{q}$ be a prime not dividing $2 \cdot 3$ but dividing Δ_{E_β}. The elliptic curve E_β has bad multiplicative reduction at $\mathfrak{q}$ if one of $c_4, c_6 \not\equiv 0 \pmod{\mathfrak{q}}$. Since γ is not divisible by $\mathfrak{q}$ and $\gcd(a, b) = 1$, we note that $c_4 \equiv c_6 \equiv 0 \pmod{\mathfrak{q}}$ happens if and only if

$$b^3 \equiv 0 \pmod{\mathfrak{q}} \quad \text{or} \quad 4ia + 5b^3 \equiv 0 \pmod{\mathfrak{q}},$$

and

$$2a + (-7i - 6z^2 + 3)b^3 \equiv 0 \pmod{\mathfrak{q}} \quad \text{or} \quad 2a + (-7i + 6z^2 - 3)b^3 \equiv 0 \pmod{\mathfrak{q}}.$$

The determinants of the resulting linear system in the variables a and b^3, in all four cases, are only divisible by primes above 2 and 3. It follows that E_β has bad multiplicative reduction at $\mathfrak{q}$.

By (8), since $\gcd(a,b)=1$, we have $v_3(\Delta_{E_\beta})=12$. We run through all possibilities for (a,b) modulo ν_3^6 and, in each case, we compute the reduction type of E_β at $\mathfrak{q}_3$ using MAGMA; in every case, said reduction type turns out to be of type II^* or I_0. By Lemma 5(a), this determines all the possible conductor exponents for E_β at $\mathfrak{q}_3$.

Since a and b are of opposite parity, (8) implies that $v_2(\Delta_{E_\beta})=16$. Checking all possibilities for (a,b) modulo ν_2^8, and in each case computing the reduction type of E_β at $\mathfrak{q}_2$, via MAGMA, we always arrive at reduction type II. By Lemma 5(b), this determines all the possible conductor exponents for E_β at $\mathfrak{q}_2$. ☐

Theorem 7. *The conductor of* $\mathrm{Res}_{\mathbb{Q}}^{K_\beta} E_\beta$ *is*

$$d_{K_\beta/\mathbb{Q}}{}^2 \cdot N_{K_\beta/\mathbb{Q}}(\mathfrak{m}) = 2^{16}\cdot 3^{4+2\varepsilon}\cdot \prod_{\substack{q\mid c\\ q\neq 2,3}} q^4,$$

where $\varepsilon = 0, 4$.

Proof. See [Milne 1972, Lemma, p. 178]. We also note that K_β is unramified outside $\{2,3\}$ so the product is of the form stated. ☐

Corollary 8. *The elliptic curve* E_β *has potentially good reduction at* $\mathfrak{q}_2$ *and* $\mathfrak{q}_3$. *In the latter case, the reduction is potentially supersingular.*

Let $A=\mathrm{Res}_{\mathbb{Q}}^{K_\beta} E_\beta$. By [Quer 2000, Theorem 5.4], A is an abelian variety of GL_2 type with $M_\beta=\mathbb{Q}(\sqrt{3},i)$. The conductor of the system of $M_{\beta,\pi}[G_{\mathbb{Q}}]$-modules $\{\hat{V}_\pi(A)\}$ is given by

$$2^4\cdot 3^{1+\varepsilon/2}\cdot \prod_{\substack{q\mid c\\ q\neq 2,3}} q, \tag{10}$$

using the conductor results explained in [Chen 2010].

For the next two theorems, it is useful to recall that $a-b^3i$ and $a+b^3i$ are coprime p-th powers in $\mathbb{Z}[i]$.

Theorem 9. *The representation* $\phi_{E,p}|_{I_p}$ *is finite flat for* $p\neq 2,3$.

Proof. This follows from the fact that E has good or bad multiplicative reduction at primes above p when $p\neq 2,3$, and in the case of bad multiplicative reduction, the exponent of a prime above p in the minimal discriminant of E is divisible by p. Also, p is unramified in K_β so that $I_p\subseteq G_{K_\beta}$. ☐

Theorem 10. *The representation* $\phi_{E,p}|_{I_\ell}$ *is trivial for* $\ell\neq 2,3,p$.

Proof. This follows from the fact that E has good or bad multiplicative reduction at primes above ℓ when $\ell\neq 2,3$, and, in the case of bad multiplicative reduction, the exponent of a prime above ℓ in the minimal discriminant of E is divisible by p. Also, ℓ is unramified in K_β so that $I_\ell\subseteq G_{K_\beta}$. ☐

Theorem 11. *Suppose $p \neq 2, 3$. The conductor of $\rho = \rho_{E,\beta,\pi}$ is one of* 48 *or* 432.

Proof. Since we are determining the Artin conductor of ρ, we consider only ramification at primes above $\ell \neq p$.

Suppose $\ell \neq 2, 3, p$. Since $\ell \neq 2, 3$, we see that K_β is unramified at ℓ and hence G_{K_β} contains I_ℓ. Now, in our case, $\rho|_{G_{K_\beta}}$ is isomorphic to $\phi_{E,p}$. Since $\phi_{E,p}|_{I_\ell}$ is trivial, $\rho|_{I_\ell}$ is trivial, so ρ is unramified outside $\{2, 3, p\}$.

Suppose $\ell = 2, 3$. The representation $\hat{\phi}_{E,p}|_{I_\ell}$ factors through a finite group of order only divisible by the primes 2 and 3. Now, in our case, $\hat{\rho}|_{G_{K_\beta}}$ is isomorphic to $\hat{\phi}_{E,p}$. Hence, the representation $\hat{\rho}|_{I_\ell}$ also factors through a finite group of order only divisible by the primes 2 and 3. It follows that the exponent of ℓ in the conductor of ρ is the same as in the conductor of $\hat{\rho}$ as $p \neq 2, 3$. □

Proposition 12. *Suppose $p \neq 2, 3$. Then the weight of $\rho = \rho_{E,\beta,\pi}$ is* 2.

Proof. The weight of ρ is determined by $\rho|_{I_p}$. Since $p \neq 2, 3$, we see that K_β is unramified at p and hence G_{K_β} contains I_p. Now, in our case, $\rho|_{G_{K_\beta}}$ is isomorphic to $\phi_{E,p}$. Since $\phi_{E,p}|_{I_p}$ is finite flat and its determinant is the p-th cyclotomic character, the weight of ρ is necessarily 2 [Serre 1987, Proposition 4]. □

Proposition 13. *The character of $\rho_{E,\beta,\pi}$ is ϵ.*

Proof. This follows from (4). □

Let $X_{0,B}^K(d, p)$, $X_{0,N}^K(d, p)$, and $X_{0,N'}^K(d, p)$ be the modular curves with level-p structure corresponding to a Borel subgroup B, the normalizer of a split Cartan subgroup N, the normalizer of a nonsplit Cartan subgroup N' of $\mathrm{GL}_2(\mathbb{F}_p)$, and level-$d$ structure consisting of a cyclic subgroup of order d, twisted by the quadratic character associated to K through the action of the Fricke involution w_d.

Lemma 14. *Let E be a $\mathbb{Q}$-curve defined over K', K a quadratic number field contained in K', and d a prime number such that*

(a) *the elliptic curve E is defined over K,*

(b) *the choices of $\mu_E(\sigma)$ are constant on G_K cosets, $\mu_E(\sigma) = 1$ when $\sigma \in G_K$, and* $\deg \mu_E(\sigma) = d$ *when $\sigma \notin G_K$, and*

(c) $\mu_E(\sigma)^\sigma\mu_E(\sigma) = \pm d$ *whenever $\sigma \notin G_K$.*

If $\rho_{E,\beta,\pi}$ has image lying in a Borel subgroup, normalizer of a split Cartan subgroup, or normalizer of a nonsplit Cartan subgroup of $\overline{\mathbb{F}}_p^\times\,\mathrm{GL}_2(\mathbb{F}_p)$, then E gives rise to a $\mathbb{Q}$-rational point on the corresponding modular curve above.

Proof. This proof is based on [Ellenberg 2004, Proposition 2.2]. Recall the action of $G_\mathbb{Q}$ on $\mathbb{P}E[d]$ is given by $x \mapsto \mu_E(\sigma)({}^\sigma x)$. Suppose $\mathbb{P}\rho_{E,\beta,p}$ has image lying in a Borel subgroup. Then we have that $\mu_E(\sigma)({}^\sigma C_p) = C_p$ for some cyclic subgroup C_p of order p in $E[p]$ and all $\sigma \in G_\mathbb{Q}$. Let C_d be the cyclic subgroup of order d in

$E[d]$ defined by $\mu_E(\sigma)({}^\sigma E[d])$ where σ is an element of $G_\mathbb{Q}$ which is nontrivial on K. This does not depend on the choice of σ. Suppose σ is an element of $G_\mathbb{Q}$ which is nontrivial on K. The kernel of $\mu_E(\sigma)$ is precisely ${}^\sigma C_d$ as $\mu_E(\sigma)({}^\sigma C_d) = \mu_E(\sigma)^\sigma\mu_E(\sigma)({}^{\sigma^2}E[d]) = [\pm d]({}^{\sigma^2}E[d]) = 0$. Hence, we see that

$$\begin{aligned} w_d{}^\sigma(E, C_d, C_p) &= w_d({}^\sigma E, {}^\sigma C_d, {}^\sigma C_p) \\ &= (\mu_E(\sigma)({}^\sigma E), \mu_E(\sigma)({}^\sigma E[d]), \mu_E(\sigma)({}^\sigma C_p)) \\ &= (E, C_d, C_p), \end{aligned}$$

so ${}^\sigma(E, C_d, C_p) = w_d(E, C_d, C_p)$, where w_d is the Fricke involution. Suppose σ is an element of $G_\mathbb{Q}$ which is trivial on K. In this case, we have ${}^\sigma(E, C_d, C_p) = (E, C_d, C_p)$. Thus, (E, C_d, C_p) gives rise to a $\mathbb{Q}$-rational point on $X_{0,B}(d, p)$.

The case when the image of $\rho_{E,\beta,\pi}$ lies in the normalizer of a Cartan subgroup is similar except now we have the existence of a set of distinct points $S_p = \{\alpha_p, \beta_p\}$ of $\mathbb{P}E[p] \otimes \mathbb{F}_{p^2}$ such that the action of $\sigma \in G_\mathbb{Q}$ by $x \mapsto \mu_E(\sigma)({}^\sigma x)$ fixes S_p as a set. □

Theorem 15. *Suppose the representation $\rho_{E,\beta,\pi}$ is reducible for $p \neq 2, 3, 5, 7, 13$. Then E has potentially good reduction at all primes above $\ell > 3$.*

Proof. See [Ellenberg 2004, Proposition 3.2]. E gives rise to a $\mathbb{Q}$-rational point on $X_{0,N}^K(3, p)$ by Lemma 14, even though the isogeny between E and its conjugate is only defined over $\mathbb{Q}(\sqrt{3}, i)$. □

Corollary 16. *The representation $\rho_{E,\beta,\pi}$ is irreducible for $p \neq 2, 3, 5, 7, 13$.*

Proof. Lemma 4 shows that E must have bad multiplicative reduction at some prime of K above $\ell > 3$. □

Proposition 17. *If $p = 13$, then $\rho_{E,\beta,\pi}$ is irreducible.*

Proof. By Lemma 14, if $\rho_{E,\beta,\pi}$ were reducible, then E would give rise to a noncuspidal K-rational point on $X_0(39)$ where $K = \mathbb{Q}(i)$ and a noncuspidal $\mathbb{Q}$-rational point on $X_0(39)/w_3$. We can now use [Kenku 1979] which says that $X_0(39)/w_3$ has four $\mathbb{Q}$-rational points. Two of them are cuspidal. Two of them arise from points in $X_0(39)$ defined over $\mathbb{Q}(\sqrt{-7})$. Hence, no such E can exist, since a K-rational point on $X_0(39)$ which is also $\mathbb{Q}(\sqrt{-7})$-rational must be $\mathbb{Q}$-rational (and again by [Kenku 1979], $X_0(39)$ has no noncuspidal $\mathbb{Q}$-rational points). □

Outline of proof of Theorem 1. Using the modularity of E, which now follows from Serre's conjecture [Serre 1987; Khare and Wintenberger 2009a; 2009b; Kisin 2009] plus the usual level-lowering arguments based on results in [Ribet 1990], we have $\rho_{E,\pi,\beta} \cong \rho_{g,\pi}$, where g is a newform in $S_2(\Gamma_0(M), \epsilon)$ where $M = 48$ or $M = 432$. This holds for $n = p \geq 11$.

There is one newform F_1 in $S_2(\Gamma_0(48), \epsilon)$ and this has CM by $\mathbb{Q}(-3)$; see inner-48.txt, cm-48.txt. At level 432, we find three newforms G_1, G_2, and G_3 in $S_2(\Gamma_0(432), \epsilon)$; inner-432.txt. As it transpires, both G_1 and G_2 have CM by $\mathbb{Q}(-3)$; cm-432.txt. The form G_3 is harder to eliminate as it does not have complex multiplication and its field of coefficients is $M_\beta = \mathbb{Q}(\sqrt{3}, i)$. Furthermore, the complex conjugate of G_3 is a twist of G_3 by ϵ^{-1}. In fact, G_3 arises from the near solution $1^2 + 1^6 = 2$ (this near solution gives rise to a form at level 432 and it is the unique non-CM form at that level) so it shares many of the same properties g should have as both arise from the same geometric construction. Note, however, that one cannot have $a \equiv b \equiv 1 \pmod 2$ in the equation $a^2 + b^6 = c^p$ as $p > 1$.

Unfortunately, it is not possible to eliminate the possibility of $g = G_3$ by considering the fields cut out by images of inertia at 2. Using [Kraus 1990, théorème 3] and its proof, it can be checked that these fields are the same regardless of whether or not $a \equiv b \equiv 1 \pmod 2$.

In the next two sections, we show that in each case $g = G_i$, for $i = 1, 2$ (CM case), and $i = 3$, we are led to a contradiction, if $n = p \geq 11$. Finally, in the last section, we deal with the cases $n = 3, 4, 5, 7$. This suffices to prove the theorem as any integer $n \geq 3$ is either divisible by an odd prime or by 4.

4. Eliminating the CM forms

When $g = G_i$ for $i = 1$ or 2, g has complex multiplication by $\mathbb{Q}(\sqrt{-3})$ so that $\rho_{E,\beta,\pi}$ has image lying in the normalizer of a Cartan subgroup for $p > 3$. However, this leads to a contradiction using the arguments below.

Proposition 18. *Let $p \geq 7$ be prime and suppose there exists either a p-newform in $S_2(\Gamma_0(3p^2))$ with $w_p f = f$, $w_3 f = -f$, or a p-newform in $S_2(\Gamma_0(p^2))$ with $w_p f = f$, such that $L(f \otimes \chi_{-4}, 1) \neq 0$, where χ_{-4} is the Dirichlet character associated to $K = \mathbb{Q}(i)$. Let E be an elliptic curve which gives rise to a noncuspidal $\mathbb{Q}$-rational point on $X^K_{0,N}(3, p)$ or $X^K_{0,N'}(3, p)$. Then E has potentially good reduction at all primes of K above $\ell > 3$.*

Proof. See [Ellenberg 2004] and comments in [Bennett et al. 2010, Proposition 6] about the applicability to the split case (see also the argument in [Ellenberg 2004, Lemma 3.5] which shows potentially good reduction at a prime of K above p in the split case). □

Proposition 19. *If $p \geq 11$ is prime, then there exists a p-newform $f \in S_2(\Gamma_0(p^2))$ with $w_p f = f$ and $L(f \otimes \chi_{-4}, 1) \neq 0$.*

Proof. For $p \geq 61$, this is, essentially, the content of the proof of [Bennett et al. 2010, Proposition 7] (the proof applies to $p \equiv 1 \pmod 8$, not just to $p \not\equiv 1 \pmod 8$)

as stated). Further, a relatively short Magma computation newform-twists.txt reveals the same to be true for smaller values of p with the following forms f (the number following the level indicates Magma's ordering of forms; one should note that this numbering is consistent neither with Stein's modular forms database nor with Cremona's tables):

p	f	dim f	p	f	dim f	p	f	dim f
11	121 (1)	1	29	841 (1)	2	47	2209 (9)	16
13	169 (2)	3	31	961 (1)	2	53	2809 (1)	1
17	289 (1)	1	37	1369 (1)	1	59	3481 (1)	2
19	361 (1)	1	41	1681 (1)	2			
23	529 (7)	4	43	1849 (1)	1			

□

Theorem 20. *Suppose the representation $\rho_{E,\beta,\pi}$ has image lying in the normalizer of a Cartan subgroup for $p \geq 11$. Then E has potentially good reduction at all primes of K above $\ell > 3$.*

Proof. We note that E still gives rise to a $\mathbb{Q}$-rational point on $X_{0,N}^K(3, p)$ or $X_{0,N'}^K(3, p)$ with $K = \mathbb{Q}(i)$, even though, as a consequence of Lemma 14, the isogeny between E and its conjugate is only defined over $\mathbb{Q}(\sqrt{3}, i)$. □

Theorem 21. *If $p \geq 11$ is prime, the representation $\rho_{E,\beta,\pi}$ does not have image lying in the normalizer of a Cartan subgroup.*

Proof. Lemma 4 immediately implies that E necessarily has bad multiplicative reduction at a prime of K lying above some $\ell > 3$. □

5. Eliminating the newform G_3

Recall that $E = E_{a,b}$ is given by

$$E : Y^2 = X^3 - 3(5b^3 + 4ai)bX + 2(11b^6 + 14ib^3a - 2a^2).$$

Let $E' = E'_{a,b}$ be the elliptic curve

$$E' : Y^2 = X^3 + 3b^2X + 2a,$$

which is a Frey–Hellegouarch elliptic curve over $\mathbb{Q}$ for the equation $a^2 + (b^2)^3 = c^p$. We will show how to eliminate the case of $g = G_3$ using a combination of congruences from the two Frey curves E and E'. This is an example of the multi-Frey technique [Bugeaud et al. 2008a; 2008b], as applied to the situation when one of the Frey curves is a $\mathbb{Q}$-curve. We are grateful to Siksek for suggesting a version of Lemma 24 which allows us to do this.

The discriminant of E' is given by

$$\Delta' = -2^6 \cdot 3^3 (a^2 + b^6). \tag{11}$$

For $a \not\equiv b \pmod 2$, $v_2(\Delta') = 6$, so E' is in minimal form at 2. Since a and b are not both multiples of 3, we have $v_3(\Delta') = 3$ and so E' is also minimal at 3. If q divides Δ' and is neither 2 nor 3, then E' has bad multiplicative reduction at q.

For each congruence class of (a, b) modulo 2^4 where $a \not\equiv b \pmod 2$, we compute the conductor exponent at 2 of E' using MAGMA. The conductor exponent at 2 of each test case was 5 (reduction type III) or 6 (reduction type II): tate2m-3.txt. By Lemma 5(c), the conductor exponent at 2 of E' is 5 or 6. In a similar way, the conductor exponent at 3 of E' is 2 (reduction type III) or 3 (reduction type II): tate3m-3.txt.

We are now almost in position to apply the modular method to E'. We need only show that the representation $\rho_{E',p}$ arising from the p-torsion points of E' is irreducible.

Lemma 22. *If $p \geq 11$ is prime, then $\rho_{E',p}$ is irreducible.*

Proof. If $p \neq 13$, the result follows essentially from [Mazur 1978] (see [Dahmen 2008, Theorem 22]), provided $j_{E'}$ is not one of

$$-2^{15},\ -11^2,\ -11\cdot 131^3,\ \frac{-17\cdot 373^3}{2^{17}},\ \frac{-17^2\cdot 101^3}{2},\ -2^{15}\cdot 3^3,\ -7\cdot 137^3\cdot 2083^3,$$
$$-7\cdot 11^3,\ -2^{18}\cdot 3^3\cdot 5^3,\ -2^{15}\cdot 3^3\cdot 5^3\cdot 11^3,\ -2^{18}\cdot 3^3\cdot 5^3\cdot 23^3\cdot 29^3.$$

Since

$$j_{E'} = \frac{1728 b^6}{a^2 + b^6} > 0,$$

we may thus suppose that $p = 13$. In this case, if $\rho_{E',p}$ were reducible, the representation would correspond to a rational point on the curve defined via the equation $j_{13}(t) = j_{E'}$, where $j_{13}(t)$ is the map from the modular curve $X_0(13)$ to $X(1)$, given by

$$\begin{aligned} j_{13}(t) &= \frac{(t^4+7t^3+20t^2+19t+1)^3(t^2+5t+13)}{t} \\ &= \frac{(t^6+10t^5+46t^4+108t^3+122t^2+38t-1)^2(t^2+6t+13)}{t} + 1728. \end{aligned}$$

Writing $s = a/b^3$, we thus have $1728/(s^2+1) = j_{13}(t)$, for some $t \in \mathbb{Q}$, and so

$$s^2 = \frac{1728 - j_{13}(t)}{j_{13}(t)} = -\frac{(t^6+10t^5+46t^4+108t^3+122t^2+38t-1)^2(t^2+6t+13)}{(t^4+7t^3+20t^2+19t+1)^3(t^2+5t+13)}.$$

It follows that there exist rational numbers x and y with

$$y^2 = -(x^2+6x+13)(x^2+5x+13)(x^4+7x^3+20x^2+19x+1),$$

and hence coprime, nonzero integers u and v, and an integer z for which

$$(u^2+6uv+13v^2)(u^2+5uv+13v^2)(u^4+7u^3v+20u^2v^2+19uv^3+v^4) = -z^2.$$

Note that, via a routine resultant calculation, if a prime p divides both $u^2+6uv+13v^2$ and the term $(u^2+5uv+13v^2)(u^4+7u^3v+20u^2v^2+19uv^3+v^4)$, then necessarily $p \in \{2, 3, 13\}$. Since $u^2+6uv+13v^2$ is positive-definite and u, and v are coprime (whereby $u^2+6uv+13v^2 \equiv \pm 1 \pmod 3$), we thus have

$$\begin{aligned} u^2+6uv+13v^2 &= 2^{\delta_1}13^{\delta_2}z_1^2, \\ (u^2+5uv+13v^2)(u^4+7u^3v+20u^2v^2+19uv^3+v^4) &= -2^{\delta_1}13^{\delta_2}z_2^2, \end{aligned}$$

for $z_1, z_2 \in \mathbb{Z}$ and $\delta_i \in \{0, 1\}$. The first equation, with $\delta_1 = 1$, implies that $u \equiv v \equiv 1 \pmod 2$, contradicting the second. We thus have $\delta_1 = 0$, whence

$$u^2+6uv+13v^2 \equiv u^2+v^2 \equiv z_1^2 \equiv 1 \pmod 3,$$

so that 3 divides one of u and v, again contradicting the second equation, this time modulo 3. □

Applying the modular method with E' as the Frey curve thus shows that $\rho_{E',p} \cong \rho_{g',\pi'}$ for some newform $g' \in S_2(\Gamma_0(M))$ where $M = 2^r 3^s$, $r \in \{5, 6\}$, and $s \in \{2, 3\}$ (here π' is a prime above p in the field of coefficients of g'). The possible forms g' were computed using b3i-modformagain.txt. The reason the multi-Frey method works is because when $a \not\equiv b \pmod 2$, we that $r \in \{5, 6\}$, whereas when $a \equiv b \equiv 1 \pmod 2$, we have $r = 7$. Thus, the 2-part of the conductor of $\rho_{E',\pi}$ separates the cases $a \not\equiv b \pmod 2$ and $a \equiv b \pmod 2$. Hence, the newform g' that the near solution $a = b = 1$ produces does not cause trouble from the point of view of the Frey curve E'. By linking the two Frey curves E and E', it is possible to pass this information from the Frey curve E' to the Frey curve E, by appealing to the multi-Frey technique.

The following lemma results from the condition $\rho_{E',p} \cong \rho_{g',\pi'}$ and standard modular method arguments.

Lemma 23. *Let $q \geq 5$ be prime and assume $q \neq p$, where $p \geq 11$ is a prime. Let*

$$C_{x,y}(q, g') = \begin{cases} a_q(E'_{x,y}) - a_q(g') & \text{if } x^2+y^6 \not\equiv 0 \pmod q, \\ (q+1)^2 - a_q(g')^2 & \text{if } x^2+y^6 \equiv 0 \pmod q. \end{cases}$$

If $(a, b) \equiv (x, y) \pmod q$, then $p \mid C_{x,y}(q, g')$.

For our choice of splitting map β, we attached a Galois representation $\rho_{E,\beta,\pi}$ to E such that $\rho_{E,\beta,\pi} \cong \rho_{g,\pi}$ for some newform $g \in S_2(\Gamma_0(M), \epsilon)$ where $M = 48, 432$. We wish to eliminate the case of $g = G_3$. The following is the analog of Lemma 23 for $E = E_{a,b}$.

Lemma 24. *Let $q \geq 5$ be prime and assume $q \neq p$, where $p \geq 11$ is prime. Let*

$$B_{x,y}(q,g) = \begin{cases} N(a_q(E_{x,y})^2 - \epsilon(q)a_q(g)^2) & \text{if } x^2+y^6 \not\equiv 0 \pmod{q} \text{ and } \left(\frac{-4}{q}\right) = 1, \\ N(a_q(g)^2 - a_{q^2}(E_{x,y}) - 2q\epsilon(q)) & \text{if } x^2+y^6 \not\equiv 0 \pmod{q} \text{ and } \left(\frac{-4}{q}\right) = -1, \\ N(\epsilon(q)(q+1)^2 - a_q(g)^2) & \text{if } x^2+y^6 \equiv 0 \pmod{q}, \end{cases}$$

where $a_{q^i}(E_{x,y})$ is the trace of Frob_q^i acting on the Tate module $T_p(E_{x,y})$.

If $(a,b) \equiv (x,y) \pmod{q}$, then $p \mid B_{x,y}(q,g)$.

Proof. Recall the setup in Sections 2 and 3. Let π be a prime of M_β above p. The mod π representation $\rho_{A_\beta,\pi}$ of $G_\mathbb{Q}$ attached to A_β is related to E_β by

$$\mathbb{P}\rho_{A_\beta,\pi}|_{G_K} \cong \mathbb{P}\phi_{E_\beta,p},$$

where $\phi_{E_\beta,p}$ is the representation of G_K on the p-adic Tate module $T_p(E_\beta)$ of E_β, and the $\mathbb{P}$ indicates that we are considering isomorphism up to scalars. The algebraic formula which describes $\rho_{E_\beta,\beta,\pi} = \rho_{A_\beta,\pi} \cong \rho_{f,\pi}$ is

$$\rho_{A_\beta,\pi}(\sigma)(1 \otimes x) = \beta(\sigma)^{-1} \otimes \mu'_\beta(\sigma)(\phi_{E_\beta,p}(\sigma)(x))$$

where $1 \otimes x \in M_{\beta,\pi} \otimes T_p(E_\beta)$. Here, $\mu'_\beta(\sigma)$ is the chosen isogeny from ${}^\sigma E_\beta \to E_\beta$ for each σ which is constant on G_K (see the paragraph after (6)).

If $x^2 + y^6 \equiv 0 \pmod{q}$, then $q \mid c$. Recall the conductor of A_β is given by

$$2^4 \cdot 3^{1+\varepsilon/2} \cdot \prod_{\substack{q \mid c \\ q \neq 2,3}} q,$$

so that q exactly divides the conductor of A_β. Using the condition $\rho_{f,\pi} \cong \rho_{g,\pi}$, we can deduce from [Carayol 1983, théorème 2.1], [Carayol 1986, théorème (A)], [Darmon et al. 1997, Theorem 3.1], and [Gross 1990, (0.1)] that

$$p \mid N\bigl(a_q(g)^2 - \epsilon^{-1}(q)(q+1)^2\bigr).$$

If $x^2 + y^6 \not\equiv 0 \pmod{q}$, then let $\mathfrak{q}$ be a prime of K_β over q. Let $\bar{E} = \bar{E}_{a,b}$ be the reduction modulo $\mathfrak{q}$ of E. Since $(a,b) \equiv (x,y) \pmod{q}$, we have $\bar{E} = E_{x,y}$. Furthermore, since $\mathfrak{q}$ is a prime of good reduction, $T_p(E) \cong T_p(\bar{E})$.

We now wish to relate the representation $\rho_{E_\beta,\beta,\pi} = \rho_{A_\beta,\pi} \cong \rho_{f,\pi}$ to the representation $\phi_{E,p}$ for the original E. We know that

$$c_{E_\beta}(\sigma,\tau) = \beta(\sigma)\beta(\tau)\beta(\sigma\tau)^{-1} \quad \text{and} \quad c_{E_\beta}(\sigma,\tau) = c_E(\sigma,\tau)\kappa(\sigma)\kappa(\tau)\kappa(\sigma\tau)^{-1},$$

where $\kappa(\sigma) = \dfrac{{}^{\sigma}(\sqrt{\gamma})}{\sqrt{\gamma}}$ and $\gamma = \dfrac{-3+i\sqrt{3}}{2}$. It follows that

$$c_E(\sigma,\tau) = \beta'(\sigma)\beta'(\tau)\beta'(\sigma\tau)^{-1},$$

where $\beta'(\sigma) = \beta(\sigma)\kappa(\sigma)$, so that β' is a splitting map for the original cocycle $c_E(\sigma,\tau)$. Also, recall that $\epsilon(\mathrm{Frob}_q) = \left(\frac{12}{q}\right)$.

Now we have

$$\rho_{A_{\beta'},\pi}(\sigma)(1\otimes x) = \beta'(\sigma)^{-1} \otimes \mu(\sigma)(\phi_{E,p}(\sigma)(x)),$$

where $1\otimes x \in M_{\beta,\pi} \otimes T_p(E)$. For this choice of $\beta'(\sigma)$,

$$\rho_{A_{\beta'},\pi} \cong \kappa(\sigma)\xi(\sigma) \otimes \rho_{A_\beta,\pi} \cong \kappa(\sigma)\xi(\sigma) \otimes \rho_{f,\pi}.$$

This can be seen by fixing the isomorphism $\iota : E \cong E_\beta$, using standard Weierstrass models and then appealing to the

$$\begin{tikzcd}
E_\beta \arrow[r, "\sigma"] & {}^{\sigma}E_\beta \arrow[r, "\mu_{E_\beta}(\sigma)"] & E_\beta \\
E \arrow[u, "\iota"] \arrow[r, "\sigma"] & {}^{\sigma}E \arrow[u, "{}^{\sigma}\iota"] \arrow[r, "\mu_E(\sigma)"] & E. \arrow[u, "\iota"]
\end{tikzcd}$$

Recall that $\beta(\sigma) = \sqrt{\epsilon(\sigma)}\sqrt{d(\sigma)}$, so that $\beta'(\sigma) = \sqrt{\epsilon(\sigma)}\sqrt{d(\sigma)}\kappa(\sigma)$. We note that $d(\sigma) = 1$ if $\sigma \in G_{\mathbb{Q}(\sqrt{-1})}$ and $d(\sigma) = 3$ if $\sigma \notin G_{\mathbb{Q}(\sqrt{-1})}$.

Now $\left(\frac{-4}{q}\right) = 1$ means $\sigma = \mathrm{Frob}_q \in G_{\mathbb{Q}(\sqrt{-1})}$. If $\sigma \in G_{\mathbb{Q}(\sqrt{-1})}$, then $\mu(\sigma) = \mathrm{id}$ and $d(\sigma) = 1$ so

$$\rho_{A_{\beta'},\pi}(\sigma)(1\otimes x) = \beta'(\sigma)^{-1} \otimes \mu(\sigma)(\phi_{E,p}(\sigma)(x)) = \sqrt{\epsilon(\sigma)}^{-1}\kappa(\sigma)^{-1} \otimes \phi_{E,p}(\sigma)(x),$$

so $\mathrm{tr}\,\rho_{A_{\beta'},\pi}(\sigma) = \sqrt{\epsilon(\sigma)}^{-1}\kappa(\sigma)^{-1} \cdot \mathrm{tr}\,\phi_{E,p}(\sigma)$ and $\epsilon(q)a_q(f)^2 = a_q(E)^2$. Also $a_q(f) \equiv a_q(g) \pmod{\pi}$, giving the assertion that $p \mid B_\alpha(q,g)$ in the case $\left(\frac{-4}{q}\right) = 1$.

If $\left(\frac{-4}{q}\right) = -1$, then $\sigma = \mathrm{Frob}_q \notin G_{\mathbb{Q}(\sqrt{-1})}$. But then $\sigma^2 \in G_{\mathbb{Q}(\sqrt{-1})}$, and in fact, $\sigma^2 \in G_{\mathbb{Q}(\sqrt{-1},\sqrt{3})}$, so by the argument above we get

$$\mathrm{tr}\,\rho_{A_{\beta'},\pi}(\sigma^2) = \sqrt{\epsilon(\sigma)}^{-1}\kappa(\sigma)^{-1} \cdot \mathrm{tr}\,\phi_{E,p}(\sigma^2) = \mathrm{tr}\,\phi_{E,p}(\sigma^2) = a_{q^2}(E).$$

Also, $\operatorname{tr} \rho_{A_{\beta'},\pi}(\sigma) = \kappa(\sigma)\xi(\sigma)a_q(f)$ so $\operatorname{tr} \rho_{A_{\beta'},\pi}(\sigma)^2 = a_q(f)^2$. We have

$$\frac{1}{\det(1 - \rho_{A_{\beta'},\pi}(\sigma)q^{-s})} = \exp \sum_{r=1}^{\infty} \operatorname{tr} \rho_{A_{\beta'},\pi}(\sigma^r)\frac{q^{-sr}}{r}$$
$$= \frac{1}{1 - \operatorname{tr} \rho_{A_{\beta'},\pi}(\sigma)q^{-s} + q\epsilon(q)q^{-2s}}.$$

The determinant and traces are of vector spaces over $M_{\beta,\pi}$. Computing the coefficient of q^{-2s} and equating, we find that $\operatorname{tr} \rho_{A_{\beta'},\pi}(\sigma^2) = \operatorname{tr} \rho_{A_{\beta'},\pi}(\sigma)^2 - 2q\epsilon(q)$ and hence conclude that $a_q(f)^2 - 2q\epsilon(q) = a_{q^2}(E)$. Since $a_q(f) \equiv a_q(g) \pmod{\pi}$, it follows that $p | B_\alpha(q, g)$ in the case $\left(\frac{-4}{q}\right) = -1$ as well. □

Let

$$A_q(g, g') := \prod_{\substack{(x,y)\in\mathbb{F}_q^2 \\ (x,y)\neq(0,0)}} \gcd(B_{x,y}(q, g), C_{x,y}(q, g')).$$

Then we must have $p | A_q(g, g')$. For a pair g, g', we can pick a prime q and compute $A_q(g, g')$. Whenever this $A_q(g, g') \neq 0$, we obtain a bound on p so that the pair g, g' cannot arise for p larger than this bound.

For $g = G_3$, and g' running through the newforms in $S_2(\Gamma_0(2^r 3^s))$ where $r \in \{5, 6\}$ and $s \in \{2, 3\}$, the above process eliminates all possible pairs $g = G_3$ and g'; see multi-frey.txt. In particular, using $q = 5$ or $q = 7$ for each pair shows that $p \in \{2, 3, 5\}$. Hence, if $p \notin \{2, 3, 5, 7\}$, then a solution to our original equation cannot arise from the newform $g = G_3$.

6. The cases $n = 3, 4, 5, 7$

It thus remains only to treat the equation $a^2 + b^6 = c^n$ for $n \in \{3, 4, 5, 7\}$. In each case, without loss of generality, we may suppose that we have a proper, nontrivial solution in positive integers a, b, and c. If $n = 4$ or 7, the desired result is immediate from [Bruin 1999] and [Poonen et al. 2007], respectively. In the case $n = 3$, a solution with $b \neq 0$ implies, via the equation

$$\left(\frac{a}{b^3}\right)^2 = \left(\frac{c}{b^2}\right)^3 - 1,$$

a rational point on the elliptic curve given by $E : y^2 = x^3 - 1$, Cremona's 144A1 of rank 0 over $\mathbb{Q}$ with $E(\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z}$. It follows that $c = b^2$ and hence $a = 0$.

Finally, we suppose that $a^2 + b^6 = c^5$, for coprime positive integers a, b, and c. From parametrizations for solutions to $x^2 + y^2 = z^5$ (see, for example, [Chen 2010, Lemma 2]), it is easy to show that there exist coprime integers u and v (and z) for which

$$v^4 - 10v^2u^2 + 5u^4 = 5^\delta z^3, \tag{12}$$

with either

(a) $v = \beta^3$, $\delta = 0$, β coprime to 5, or

(b) $v = 5^2\beta^3$, $\delta = 1$, for some integer β.

Let us begin by treating the latter case. From (12), we have

$$(u^2 - v^2)^2 - 4 \cdot 5^7 \cdot \beta^{12} = z^3;$$

and hence taking

$$x = \frac{z}{5^2\beta^4}, \qquad y = \frac{u^2 - v^2}{5^3\beta^6},$$

we have a rational point on $E : y^2 = x^3 + 20$, Cremona's 2700E1 of rank 0 and trivial torsion (with no corresponding solutions of interest to our original equation).

We may thus suppose that we are in situation (a), so that

$$\beta^{12} - 10\beta^6 u^2 + 5u^4 = z^3. \tag{13}$$

Since β and u are coprime, we may assume that they are of opposite parity (and hence that z is odd), since $\beta \equiv u \equiv 1 \pmod 2$ with (13) leads to an immediate contradiction modulo 8. Writing $T = \beta^6 - 5u^2$, (13) becomes $T^2 - 20u^4 = z^3$, where T is coprime to 10. Factoring over $\mathbb{Q}(\sqrt{5})$ (which has class number 1), we deduce the existence of integers m and n, of the same parity, such that

$$T + 2\sqrt{5}u^2 = \left(\frac{1+\sqrt{5}}{2}\right)^{\delta} \left(\frac{m+n\sqrt{5}}{2}\right)^3, \tag{14}$$

with $\delta \in \{0, 1, 2\}$.

Let us first suppose that $\delta = 1$. Then, expanding (14), we have

$$m^3 + 15m^2n + 15mn^2 + 25n^3 = 16T \quad \text{and} \quad m^3 + 3m^2n + 15mn^2 + 5n^3 = 32u^2.$$

It follows that

$$3m^2n + 5n^3 = 4T - 8u^2 \equiv 4 \pmod 8,$$

contradicting the fact that m and n have the same parity. Similarly, if $\delta = 2$, we find that

$$3m^3 + 15m^2n + 45mn^2 + 25n^3 = 16T \quad \text{and} \quad m^3 + 9m^2n + 15mn^2 + 15n^3 = 32u^2,$$

and so

$$3m^2n + 5n^3 = 24u^2 - 4T \equiv 4 \pmod 8,$$

again a contradiction.

We thus have $\delta = 0$, and so

$$m(m^2 + 15n^2) = 8T = 8(\beta^6 - 5u^2) \quad \text{and} \quad n(3m^2 + 5n^2) = 16u^2. \tag{15}$$

Combining these equations, we may write

$$16\beta^6 = (m+5n)(2m^2+5mn+5n^2). \tag{16}$$

Returning to the last equation of (15), since $\gcd(m,n)$ divides 2, we necessarily have $n = 2^{\delta_1}3^{\delta_2}r^2$ for some integers r and $\delta_i \in \{0,1\}$. Considering the equation $n(3m^2+5n^2) = 16u^2$ modulo 5 implies that $(\delta_1,\delta_2) = (1,0)$ or $(0,1)$. In case $(\delta_1,\delta_2) = (1,0)$, the two equations in (15), taken together, imply a contradiction modulo 9.

We may thus suppose that $(\delta_1,\delta_2) = (0,1)$ and, setting $y = (2\beta/r)^3$ and $x = 6m/n$ in (16), we find that

$$y^2 = (x+30)(x^2+15x+90).$$

This elliptic curve is Cremona's 3600G1, of rank 0 with nontrivial torsion corresponding to $x = -30$, $y = 0$.

It follows that there do not exist positive coprime integers a, b, and c for which $a^2+b^6 = c^5$, which completes the proof of Theorem 1.

Acknowledgements

The authors would like to thank Samir Siksek and Sander Dahmen for useful suggestions and correspondence pertaining to this paper, and the anonymous referees for numerous helpful comments.

References

[Bennett 2006] M. A. Bennett, "The equation $x^{2n} + y^{2n} = z^5$", *J. Théor. Nombres Bordeaux* **18**:2 (2006), 315–321. MR 2007i:11048 Zbl 1138.11009

[Bennett and Skinner 2004] M. A. Bennett and C. M. Skinner, "Ternary Diophantine equations via Galois representations and modular forms", *Canad. J. Math.* **56**:1 (2004), 23–54. MR 2005c:11035 Zbl 1053.11025

[Bennett et al. 2010] M. A. Bennett, J. S. Ellenberg, and N. C. Ng, "The Diophantine equation $A^4+2^\delta B^2 = C^n$", *Int. J. Number Theory* **6**:2 (2010), 311–338. MR 2011k:11045 Zbl 1218.11035

[Bennett et al. $\geq$ 2012] M. Bennett, I. Chen, S. Dahmen, and S. Yazdani, "Generalized Fermat equations: a miscellany", In preparation.

[Beukers 1998] F. Beukers, "The Diophantine equation $Ax^p + By^q = Cz^r$", *Duke Math. J.* **91**:1 (1998), 61–88. MR 99f:11039 Zbl 1038.11505

[Bosma et al. 1997] W. Bosma, J. Cannon, and C. Playoust, "The Magma algebra system, I: The user language", pp. 235–265 in *Computational algebra and number theory* (London, 1993), edited by J. Cannon and D. Holt, J. Symbolic Comput. **24**, 1997. MR 1484478

[Brown 2012] D. Brown, "Primitive Integral Solutions to $x^2+y^3=z^{10}$", *Int. Math. Res. Not.* **2012**:2 (2012), 423–436. MR 2876388 Zbl pre06013326

[Bruin 1999] N. Bruin, "The Diophantine equations $x^2 \pm y^4 = \pm z^6$ and $x^2+y^8 = z^3$", *Compositio Math.* **118**:3 (1999), 305–321. MR 2001d:11035 Zbl 0941.11013

[Bruin 2000] N. Bruin, "On powers as sums of two cubes", pp. 169–184 in *Algorithmic number theory* (Leiden, 2000), edited by W. Bosma, Lecture Notes in Comput. Sci. **1838**, Springer, Berlin, 2000. MR 2002f:11029 Zbl 0986.11021

[Bruin 2003] N. Bruin, "Chabauty methods using elliptic curves", *J. Reine Angew. Math.* **562** (2003), 27–49. MR 2004j:11051 Zbl 1135.11320

[Bruin 2005] N. Bruin, "The primitive solutions to $x^3 + y^9 = z^2$", *J. Number Theory* **111**:1 (2005), 179–189. MR 2006e:11040 Zbl 1081.11019

[Bugeaud et al. 2008a] Y. Bugeaud, F. Luca, M. Mignotte, and S. Siksek, "Almost powers in the Lucas sequence", *J. Théor. Nombres Bordeaux* **20**:3 (2008), 555–600. MR 2010d:11017 Zbl 1204.11030

[Bugeaud et al. 2008b] Y. Bugeaud, M. Mignotte, and S. Siksek, "A multi-Frey approach to some multi-parameter families of Diophantine equations", *Canad. J. Math.* **60**:3 (2008), 491–519. MR 2009b:11059 Zbl 1156.11014

[Carayol 1983] H. Carayol, "Sur les représentations l-adiques attachées aux formes modulaires de Hilbert", *C. R. Acad. Sci. Paris Sér. I Math.* **296**:15 (1983), 629–632. MR 85e:11039 Zbl 0537.10018

[Carayol 1986] H. Carayol, "Sur les représentations l-adiques associées aux formes modulaires de Hilbert", *Ann. Sci. École Norm. Sup.* (4) **19**:3 (1986), 409–468. MR 89c:11083 Zbl 0616.10025

[Chen 2008] I. Chen, "On the equation $s^2 + y^{2p} = \alpha^3$", *Math. Comp.* **77**:262 (2008), 1223–1227. MR 2009b:11099 Zbl 1183.11030

[Chen 2010] I. Chen, "On the equation $a^2 + b^{2p} = c^5$", *Acta Arith.* **143**:4 (2010), 345–375. MR x2012b:11051 Zbl 1215.11029

[Chen 2012] I. Chen, "On the equation $a^2 - 2b^6 = c^p$ and $a^2 - 2 = c^p$", *LMS J. Comput. Math.* **15** (2012), 158–171.

[Chen and Siksek 2009] I. Chen and S. Siksek, "Perfect powers expressible as sums of two cubes", *J. Algebra* **322**:3 (2009), 638–656. MR 2011d:11070 Zbl 1215.11026

[Cox 1989] D. A. Cox, *Primes of the form $x^2 + ny^2$: Fermat, class field theory and complex multiplication*, Wiley, New York, 1989. MR 90m:11016 Zbl 0701.11001

[Dahmen 2008] S. Dahmen, *Classical and modular methods applied to Diophantine equations*, Ph.D. thesis, University of Utrecht, 2008, Available at http://tinyurl.com/dahmen-thesis.

[Dahmen 2011] S. R. Dahmen, "A refined modular approach to the Diophantine equation $x^2 + y^{2n} = z^3$", *Int. J. Number Theory* **7**:5 (2011), 1303–1316. MR 2825973 Zbl 1226.11043

[Darmon 2000] H. Darmon, "Rigid local systems, Hilbert modular forms, and Fermat's last theorem", *Duke Math. J.* **102**:3 (2000), 413–449. MR 2001i:11071 Zbl 1008.11023

[Darmon and Granville 1995] H. Darmon and A. Granville, "On the equations $z^m = F(x, y)$ and $Ax^p + By^q = Cz^r$", *Bull. London Math. Soc.* **27**:6 (1995), 513–543. MR 96e:11042 Zbl 0838.11023

[Darmon and Merel 1997] H. Darmon and L. Merel, "Winding quotients and some variants of Fermat's last theorem", *J. Reine Angew. Math.* **490** (1997), 81–100. MR 98h:11076 Zbl 0976.11017

[Darmon et al. 1997] H. Darmon, F. Diamond, and R. Taylor, "Fermat's last theorem", pp. 2–140 in *Elliptic curves, modular forms & Fermat's last theorem* (Hong Kong, 1993), edited by J. Coates and S. T. Yau, International Press, Cambridge, MA, 1997. MR 99d:11067b Zbl 0877.11035

[Dieulefait and Urroz 2009] L. Dieulefait and J. J. Urroz, "Solving Fermat-type equations via modular $\mathbb{Q}$-curves over polyquadratic fields", *J. Reine Angew. Math.* **633** (2009), 183–195. MR 2011i:11042 Zbl 05640144

[Ellenberg 2004] J. S. Ellenberg, "Galois representations attached to $\mathbb{Q}$-curves and the generalized Fermat equation $A^4 + B^2 = C^p$", *Amer. J. Math.* **126**:4 (2004), 763–787. MR 2005g:11089 Zbl 1059.11041

[Ellenberg and Skinner 2001] J. S. Ellenberg and C. Skinner, "On the modularity of $\mathbb{Q}$-curves", *Duke Math. J.* **109**:1 (2001), 97–122. MR 2002i:11054 Zbl 1009.11038

[Gross 1990] B. H. Gross, "A tameness criterion for Galois representations associated to modular forms (mod p)", *Duke Math. J.* **61**:2 (1990), 445–517. MR 91i:11060 Zbl 0743.11030

[Kenku 1979] M. A. Kenku, "The modular curve $X_0(39)$ and rational isogeny", *Math. Proc. Cambridge Philos. Soc.* **85**:1 (1979), 21–23. MR 80g:14023 Zbl 0392.14011

[Khare and Wintenberger 2009a] C. Khare and J.-P. Wintenberger, "Serre's modularity conjecture, I", *Invent. Math.* **178**:3 (2009), 485–504. MR 2010k:11087 Zbl 05636295

[Khare and Wintenberger 2009b] C. Khare and J.-P. Wintenberger, "Serre's modularity conjecture, II", *Invent. Math.* **178**:3 (2009), 505–586. MR 2010k:11088 Zbl 05636296

[Kisin 2009] M. Kisin, "Modularity of 2-adic Barsotti–Tate representations", *Invent. Math.* **178**:3 (2009), 587–634. MR 2010k:11089 Zbl 05636297

[Kraus 1990] A. Kraus, "Sur le défaut de semi-stabilité des courbes elliptiques à réduction additive", *Manuscripta Math.* **69**:4 (1990), 353–385. MR 91j:11045 Zbl 0792.14014

[Kraus 1998] A. Kraus, "Sur l'équation $a^3 + b^3 = c^p$", *Experiment. Math.* **7**:1 (1998), 1–13. MR 99f:11040 Zbl 0923.11054

[Mauldin 1997] R. D. Mauldin, "A generalization of Fermat's last theorem: the Beal conjecture and prize problem", *Notices Amer. Math. Soc.* **44**:11 (1997), 1436–1437. MR 98j:11020 Zbl 924.11022

[Mazur 1978] B. Mazur, "Rational isogenies of prime degree", *Invent. Math.* **44**:2 (1978), 129–162. MR 80h:14022 Zbl 0386.14009

[Milne 1972] J. S. Milne, "On the arithmetic of Abelian varieties", *Invent. Math.* **17** (1972), 177–190. MR 48 #8512 Zbl 0249.14012

[Poonen 1998] B. Poonen, "Some Diophantine equations of the form $x^n + y^n = z^m$", *Acta Arith.* **86**:3 (1998), 193–205. MR 99h:11034 Zbl 0930.11017

[Poonen et al. 2007] B. Poonen, E. F. Schaefer, and M. Stoll, "Twists of $X(7)$ and primitive solutions to $x^2 + y^3 = z^7$", *Duke Math. J.* **137**:1 (2007), 103–158. MR 2008i:11085 Zbl 1124.11019

[Quer 2000] J. Quer, "$\mathbf{Q}$-curves and Abelian varieties of GL_2-type", *Proc. London Math. Soc.* (3) **81**:2 (2000), 285–317. MR 2001j:11040 Zbl 1035.11026

[Ribet 1990] K. A. Ribet, "On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms", *Invent. Math.* **100**:2 (1990), 431–476. MR 91g:11066 Zbl 0773.11039

[Ribet 1992] K. A. Ribet, "Abelian varieties over **Q** and modular forms", pp. 53–79 in *Algebra and topology 1992* (Taejŏn, 1992), edited by S. G. Hahn and D. Y. Suh, Korea Adv. Inst. Sci. Tech., Taejŏn, 1992. MR 94g:11042

[Serre 1987] J.-P. Serre, "Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$", *Duke Math. J.* **54**:1 (1987), 179–230. MR 88g:11022 Zbl 0641.10026

[Siksek 2008] S. Siksek, "personal communication", 2008.

[Siksek 2010] S. Siksek, "Explicit Chabauty over number fields", preprint, 2010. arXiv 1010.2603v1

[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

[Takeuchi 1977] K. Takeuchi, "Commensurability classes of arithmetic triangle groups", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **24**:1 (1977), 201–212. MR 57 #3077 Zbl 0365.20055

[Taylor and Wiles 1995] R. Taylor and A. Wiles, "Ring-theoretic properties of certain Hecke algebras", *Ann. of Math.* (2) **141**:3 (1995), 553–572. MR 96d:11072 Zbl 0823.11030

[Tijdeman 1989] R. Tijdeman, "Diophantine equations and Diophantine approximations", pp. 215–243 in *Number theory and applications* (Banff, 1988), edited by R. A. Mollin, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **265**, Kluwer, Dordrecht, 1989. MR 92i:11072 Zbl 0719.11014

[Wiles 1995] A. Wiles, "Modular elliptic curves and Fermat's last theorem", *Ann. of Math.* (2) **141**:3 (1995), 443–551. MR 96d:11071 Zbl 0823.11029

Communicated by Richard Taylor
Received 2010-10-25 Revised 2011-02-23 Accepted 2011-04-01

bennett@math.ubc.ca *Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada*

ichen@math.sfu.ca *Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada*

ALGEBRA AND NUMBER THEORY 6:4(2012)

Detaching embedded points

Dawei Chen and Scott Nollet

Suppose that closed subschemes $X \subset Y \subset \mathbb{P}^N$ differ at finitely many points: when is Y a flat specialization of X union isolated points? Our main result says that this holds if X is a local complete intersection of codimension two and the multiplicity of each embedded point of Y is at most three. We show by example that no hypothesis can be weakened: the conclusion fails for embedded points of multiplicity greater than three, for local complete intersections X of codimension greater than two, and for nonlocal complete intersections of codimension two. As applications, we determine the irreducible components of Hilbert schemes of space curves with high arithmetic genus and show the smoothness of the Hilbert component whose general member is a plane curve union a point in $\mathbb{P}^3$.

1. Introduction

An attractive aspect of algebraic geometry is that moduli spaces for its objects tend themselves to be algebraic varieties. Ever since Grothendieck [1961] proved their existence, the Hilbert schemes $\mathrm{Hilb}^{p(z)}(\mathbb{P}^N)$ classifying flat families of subschemes in $\mathbb{P}^N$ with fixed Hilbert polynomial $p(z)$ have drawn great interest. One of the first major results was the connectedness of Hilbert schemes, proved in [Hartshorne 1966]. More recently Liaison theory [Peskine and Szpiro 1974; Martin-Deschamps and Perrin 1990; Migliore 1998] has focused attention on Hilbert schemes $H_{d,g}$ of degree d, arithmetic genus g, locally Cohen–Macaulay curves in $\mathbb{P}^3$. The connectedness of $H_{d,g}$ remains an open question [Nollet 1997; 2006; Hartshorne 2000; Nollet and Schlesinger 2003].

While Hilbert schemes can be quite complicated in general, Piene and Schlessinger [1985] gave a satisfying picture of $\mathrm{Hilb}^{3z+1}(\mathbb{P}^3)$: there are two smooth irreducible components of dimensions 12 and 15 which meet transversely along an 11-dimensional family. In [Chen 2008], Mori's program was applied to the 12-dimensional component of twisted cubics, working out the effective cone decomposition and the corresponding models, exhibiting it as a flip of the Kontsevich moduli space of stable maps over the Chow variety. Similarly the Hilbert scheme

MSC2010: primary 14B07; secondary 14H10, 14H50.

Keywords: Hilbert schemes, embedded points.

component of unions of a pair of codimension-two linear subspaces of $\mathbb{P}^N$ is a smooth Mori dream space [Chen et al. 2011].

In an effort to achieve a similar understanding of the geometry of the Hilbert scheme component H_1 of rational quartic curves in $\mathbb{P}^3$, the first obstacle is determining the other irreducible components of $\mathrm{Hilb}^{4z+1}(\mathbb{P}^3)$. There are three natural families whose general members consist of the disjoint union of a line and a plane cubic, the disjoint union of an elliptic quartic curve and a point, and the disjoint union of a plane quartic and three points, but what about a possible component whose general member has an embedded point? We show in Example 2.9 that such Hilbert scheme components exist for curves of degree four and sufficiently negative genus. This motivates the following question:

Question 1.1. If X is obtained from $Y \subset \mathbb{P}^N$ by removing the zero-dimensional components, under what conditions is Y in the Hilbert scheme closure of the family consisting of X union isolated points?

When this is the case, we say that Y is a *flat limit* of X union isolated points, or simply that one can *detach* the embedded points of Y.

Remarks 1.2. (a) From the Hilbert scheme perspective, we should allow X to vary in the flat family. On the other hand, it is clearly desirable to have results requiring no information on how X sits in its Hilbert scheme, for they will be easier to apply.

(b) Question 1.1 is already interesting when X is empty. Fogarty [1968] observed that $\mathrm{Hilb}^d(\mathbb{P}^2)$ is irreducible for all $d > 0$, but Iarrobino [1972] showed that $\mathrm{Hilb}^d(\mathbb{P}^3)$ is reducible for $d \gg 0$. The minimum such value of d is still unknown. Iarrobino and Emsalem [1978] showed that $\mathrm{Hilb}^8(\mathbb{P}^4)$ is reducible and [Mazzola 1980] showed that $\mathrm{Hilb}^d(\mathbb{P}^n)$ is irreducible for $d \leq 7$. Cartwright et al. [2009] extended this to prove that for $d \leq 8$, $\mathrm{Hilb}^d(\mathbb{P}^N)$ is reducible if and only if $d = 8$ and $N \geq 4$.

We are interested in the case $\dim X > 0$. The kernel K of the surjection $\mathcal{O}_Y \to \mathcal{O}_X$ has finite length and may be written $\bigoplus K_p$ with p in the support of K. For such p, we say that the *multiplicity* of p is length K_p. The following criterion tells when *all* subschemes obtained from X by adding an embedded point of multiplicity one at $p \in X$ are flat limits of X union an isolated point (see Theorem 2.3).

Theorem 1.3. *For $p \in X \subset \mathbb{P}^N$, the following are equivalent*:

(1) *All subschemes Y obtained from X by adding an embedded point of multiplicity one at p are flat limits of X union an isolated point.*

(2) *The ideal sheaf $\mathcal{I}_X$ has r minimal generators at p with $r \leq N$ and $\pi^{-1}(p) \cong \mathbb{P}^{r-1}$, where $\pi : \widetilde{\mathbb{P}}^N \to \mathbb{P}^N$ is the blow-up at X.*

In particular, *if X is a local complete intersection*, *then any embedded point of multiplicity one can be detached from X.*

Condition (2) makes it easy to recognize when there exist schemes Y obtained from X which are not flat limits of X union an isolated point (see Example 2.6). Sometimes an embedded point of multiplicity one cannot be detached even if X is allowed to move in the deformation (see Example 1.5). Our main result gives conditions under which embedded points of various multiplicities can be detached (see Theorem 3.9):

Theorem 1.4. *Let $X \subset \mathbb{P}^N$ be a local complete intersection of codimension two. If Y is obtained from X by adding embedded points of multiplicity at most three, then Y is a flat limit of X union isolated points.*

The hypotheses may seem restrictive, but Theorem 1.4 is sharp in all aspects, as the following examples show.

Example 1.5. For any $g \leq -15$, the Hilbert scheme $\mathrm{Hilb}^{4z+1-g}(\mathbb{P}^3)$ has an irreducible component H of dimension $9-2g$ whose general member is the union of a multiplicity 4-line containing the triple line of generic embedding dimension three and an embedded point of multiplicity one. Details are given in Example 2.9.

Example 1.6. There are local complete intersections $X \subset \mathbb{P}^N$ of codimension greater than two and Y obtained from X by adding an embedded point of multiplicity two which are not flat limits of X union two isolated points. Let X be the nonreduced curve in $\mathbb{P}^4$ with ideal $I_X = (x^2, y^2, z^2)$. The family of double point structures on X has dimension equal to eight, the same as the dimension of the family consisting of X union two isolated points, hence the former cannot lie in the closure of the latter. See Example 3.6 for details.

Example 1.7. There are local complete intersections $X \subset \mathbb{P}^N$ of codimension two and Y obtained from X by adding an embedded point of multiplicity four which are not flat limits of X union four isolated points. For X with ideal $I_X = (x^2, y^2)$ in $\mathbb{P}^N$, we give a family of such subschemes Y having dimension $5N-6$, hence the general member cannot be a flat limit of X union four isolated points for $N > 5$. See Example 3.10 for details.

Remark 1.8. (a) For Y and X as in Theorem 1.4, there is an exact sequence

$$0 \to \mathcal{I}_Y \to \mathcal{I}_X \xrightarrow{\varphi} K \to 0, \tag{1}$$

where K is a sheaf of finite length. It is clear that the sheaf K is uniquely determined by Y (it is the quotient $\mathcal{I}_X/\mathcal{I}_Y$) and that two surjections φ and φ' yield the same subscheme Y if and only if there exists an automorphism σ of K such that $\varphi' = \sigma \circ \varphi$. The technique of our proof deforms the pair (φ, K).

(b) It is *not* the case that the embedded points can be pulled away one at a time; see Example 3.5.

(c) If X is a *hypersurface* and Y is obtained from X by adding embedded points of *any* multiplicities, then Y is a flat limit of X union isolated multiple points. In particular, such Y is a flat limit of X union isolated reduced points if the multiplicities are less than eight (Proposition 2.4).

Applying Theorem 1.4 to plane curves in $\mathbb{P}^3$, we deduce the following:

Corollary 1.9. *For $d \geq 6$ and $(d-1)(d-2)/2 - 3 \leq g \leq (d-1)(d-2)/2$, the Hilbert scheme* $\mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3)$ *is irreducible.*

In Section 3 we give many other applications to space curves of low degree. Letting $g = (d-1)(d-2)/2$ be the genus of a degree-d plane curve, we give the following smoothness result:

Theorem 1.10. *Let $H_d \subset \mathrm{Hilb}^{dz+2-g}(\mathbb{P}^3)$ be the closure of the family of degree-d plane curves union an isolated point. Then H_d is smooth for all $d \geq 1$, and hence isomorphic to the blow-up of* $\mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3) \times \mathbb{P}^3$ *along the incidence correspondence.*

Remark 1.11. Similarly the Hilbert scheme of a hypersurface in $\mathbb{P}^N$ union an isolated point is smooth (Theorem 4.1), but the Hilbert scheme is not smooth at plane curves union certain double embedded points (Remark 4.4).

Regarding organization, we deal with the question of detaching embedded points of multiplicity one in Section 2, and with embedded points of multiplicities two or three in Section 3. Our applications to Hilbert schemes are found in Section 4.

Conventions. For a subscheme $Z \subset \mathbb{P}^N$, $\mathcal{I}_Z$ denotes its sheaf of ideals and I_Z denotes its homogeneous (saturated) ideal or sometimes the ideal of Z in an open affine chart. We often write $\mathbb{O}$ for the structure sheaf of the ambient projective space and S for the homogeneous coordinate ring. A curve is a (purely) one-dimensional scheme. We say that Y is a flat limit of X union isolated points if Y is in the Hilbert scheme closure of this family. This is equivalent to the existence of a one-parameter family $\{Y_t\}_{t \in T}$ in which Y_t is X union isolated points for t general and $Y = Y_0$, and this is typically how we exhibit such a flat limit. We sometimes speak of a flat limit of ideals (or ideal sheaves) when working with the corresponding ideals. If two schemes $X \subset Y$ differ at an embedded point supported at $p \in X$, the multiplicity of the embedded point is the length of $\mathcal{I}_{X,p}/\mathcal{I}_{Y,p}$. Throughout the paper we work over an algebraically closed field k of arbitrary characteristic, but will occasionally assume char $k \neq 2, 3$ to apply irreducibility results.

2. Detaching embedded points of multiplicity one

In this section we study embedded point structures of multiplicity one on a local complete intersection $X \subset \mathbb{P}^N$ of codimension two. We also give a global result for

ACM subschemes with 3-generated ideal (Proposition 2.7). We begin by determining when an embedded point of multiplicity one can be detached from a subscheme $X \subset \mathbb{P}^N$.

Proposition 2.1. *For a proper subscheme $X \subset \mathbb{P}^N$, let $V \subset \mathrm{Hilb}^{p(z)+1}(\mathbb{P}^N)$ be the closed subset of subschemes which may be obtained from X by adding a point p (embedded or isolated). Then there is a diagram*

$$\begin{tikzcd} \widetilde{\mathbb{P}}^N(X) \arrow[r, hook, "f"] \arrow[d, "\pi"] & V \arrow[d, "h"] \\ \mathbb{P}^N \arrow[r, phantom, "="] & \mathbb{P}^N, \end{tikzcd} \tag{2}$$

in which π is the blow-up at X, h sends a subscheme in V to the added point, and f extends the map $\mathbb{P}^N - X \to V$ given by $p \mapsto X \cup p$. Moreover, f is injective.

Proof. There is a uniform bound for the Castelnuovo–Mumford regularity of every ideal sheaf defining a closed subscheme with Hilbert polynomial $p(z)$, hence $h^0(\mathcal{I}_Y(m))$ is independent of $[Y] \in \mathrm{Hilb}^{p(z)+1}(\mathbb{P}^N)$ for sufficiently large m and the map

$$Y \mapsto \big(H^0(\mathbb{P}^N, \mathcal{I}_Y(m)) \subset H^0(\mathbb{P}^N, \mathcal{O}_{\mathbb{P}^N}(m))\big)$$

yields a closed immersion $F : \mathrm{Hilb}^{p(z)+1}(\mathbb{P}^N) \hookrightarrow \mathbb{G}$ to a suitable Grassmann variety [Harris and Morrison 1998]. Since $H^0(\mathbb{P}^N, \mathcal{I}_Y(m)) \subset H^0(\mathbb{P}^N, \mathcal{I}_X(m))$ has codimension one for $[Y] \in V$, the image $F(V)$ is contained in $\mathbb{P}(H^0(\mathbb{P}^N, \mathcal{I}_X(m)))^\vee \subset \mathbb{G}$. On the other hand, a standard construction [Peskine and Szpiro 1974, Proposition 4.1] yields a closed immersion $\widetilde{\mathbb{P}}^N(X) \overset{j}{\hookrightarrow} \mathbb{P}(H^0(\mathbb{P}^N, \mathcal{I}_X(m)))^\vee$ and for each $p \in \mathbb{P}^N - X$ we have $j(\pi^{-1}(p)) = F(h^{-1}(p))$. Since V is closed, we obtain an injective map $\widetilde{\mathbb{P}}^N(X) \hookrightarrow V$ and accompanying diagram (2). □

Proposition 2.2. *In diagram (2), the following are equivalent*:

(a) *V is irreducible.*

(b) *For each $p \in X$, $\dim_{k(p)} \mathcal{I}_{X,p} \otimes k(p) = r \le N$ and $\pi^{-1}(p) \cong \mathbb{P}^{r-1}$.*

(c) *The map $\widetilde{\mathbb{P}}^N(X) \xrightarrow{f} V$ is bijective.*

Proof. For each $[Y] \in V$, there is an exact sequence

$$0 \to \mathcal{I}_Y \to \mathcal{I}_X \to K_p \to 0,$$

where $K_p \cong \mathcal{O}_p$ is the skyscraper sheaf of length 1 supported at p. For fixed p, the set of all such Y is given by surjections

$$\phi \in \mathrm{Hom}(\mathcal{I}_X, \mathcal{O}_p) \cong \mathrm{Hom}(\mathcal{I}_{X,p}, k(p)) \cong \mathrm{Hom}(\mathcal{I}_X \otimes k(p), k(p)) \tag{3}$$

modulo scalar. In view of Nakayama's lemma, we see that $h^{-1}(p) \cong \mathbb{P}^{r-1}_{k(p)}$, where r is the minimal number of generators for $\mathcal{I}_X$ at p.

The equivalence of (a) and (c) is clear from Proposition 2.1. Condition (c) implies that $\pi^{-1}(p) \cong h^{-1}(p) \cong \mathbb{P}^{r-1}$ for each $p \in X$ and $r \le N$ because $\pi^{-1}(p) \subset \widetilde{\mathbb{P}}^N(X)$ is a proper subset, proving (b). Conversely if (b) holds, then for $p \in X$, we have injections $\mathbb{P}^{r-1} \cong \pi^{-1}(p) \hookrightarrow h^{-1}(p) \cong \mathbb{P}^{r-1}$ which must be surjective by reason of dimension, hence $f : \widetilde{\mathbb{P}}^N(X) \to V$ is bijective on the fibers over $\mathbb{P}^N$ and is therefore bijective. □

The next result follows from the argument above. It allows one to determine when *all* embedded structures of multiplicity one supported at a fixed point can be detached.

Theorem 2.3. *For $p \in X \subset \mathbb{P}^N$, the following are equivalent*:

(1) *Every subscheme Y obtained from X by adding an embedded point of multiplicity one at p is a flat limit of X union an isolated point.*

(2) *X satisfies condition* (*b*) *of Proposition 2.2 at p.*

In particular, *these conditions hold if X is a local complete intersection.*

Proof. In the setting of Proposition 2.1, let $U = h^{-1}(\mathbb{P}^N - X) \subset V$ correspond to the subschemes obtained from X by adding an isolated point. Note that $f(\widetilde{\mathbb{P}}^N(X)) = \overline{U} \subset V$, since it is a closed subset with dense open subset U; hence for fixed $p \in X$ we have an inclusion $f(\pi^{-1}(p)) \subset h^{-1}(p) \cong \mathbb{P}^{r-1}$. Now condition (b) holds if and only if $\pi^{-1}(p) \cong \mathbb{P}^{r-1}$, if and only if $f(\pi^{-1}(p)) = h^{-1}(p)$ by reason of dimension and irreducibility of $\mathbb{P}^{r-1}$; but this equality is equivalent to $h^{-1}(p) \subset f(\widetilde{\mathbb{P}}^N(X)) = \overline{U}$, which is equivalent to condition (a). If X is a local complete intersection of codimension r, then it is well-known that $\pi^{-1}(p) \cong \mathbb{P}^{r-1}$ [Hartshorne 1977, II, Theorem 8.24(b)]; hence condition (b) from Proposition 2.2 holds. □

We can make a stronger statement when X is a hypersurface.

Proposition 2.4. *If Y is obtained from a hypersurface $X \subset \mathbb{P}^N$ by adding embedded points of any multiplicities*, *then Y is a flat limit of X union isolated multiple points. In particular*, *Y is a flat limit of X union isolated reduced points if the multiplicities are at most seven. For $N \ge 4$*, *there exist embedded structures of multiplicity eight in Y such that Y is not a flat limit of X union eight reduced points.*

Proof. Suppose that Y is defined by the surjection $\mathcal{I}_X \to K$, where K is of finite length supported at the embedded points p. Then $K \cong \bigoplus_p \mathcal{O}_{Z_p}$ for finite length subschemes Z_p supported at p ($\mathcal{I}_X$ is principal) and $\mathcal{I}_Y = \mathcal{I}_X \cdot \mathcal{I}_Z$, where Z is the union of the zero-dimensional subschemes Z_p. Use automorphisms of $\mathbb{P}^N$ to deform Z to Z_t such that the support of Z_t does not intersect X for $t \ne 0$. Then $\mathcal{I}_{X \cup Z_t} = \mathcal{I}_X \cdot \mathcal{I}_{Z_t}$ for $t \ne 0$ and in considering the associated schemes it is clear that Y is a flat limit of $X \cup Z_t$. If the length of Z_t is ≤ 7, then Z_t is a flat limit of reduced points [Mazzola 1980; Cartwright et al. 2009]; hence Y is a flat limit of X union isolated reduced points.

For $N \geq 4$, there exists a nonsmoothable, length-8 subscheme $Z \subset \mathbb{P}^N$ [Iarrobino and Emsalem 1978; Cartwright et al. 2009]. Choose an open affine $U \cong \mathbb{A}^N$ on which $\mathcal{I}_X$ is trivial, apply an automorphism of $\mathbb{P}^N$ to translate Z so that the support of Z lies in $U \cap X$, and let Y be the subscheme defined by the surjection $\mathcal{I}_X \cong \mathcal{O} \to \mathcal{O}_Z$. If $\mathcal{I}_X = (f)$ locally, then $\mathcal{I}_Y = (f)\mathcal{I}_Z$ and Y cannot be a flat limit of X union eight isolated points, for then $\mathcal{I}_Y$ would be the flat limit of ideals $(f)\mathcal{I}_{Z_t}$, where Z_t consists of eight isolated points and from the expression of $\mathcal{I}_Y$ we would obtain $\mathcal{I}_Z$ as a flat limit of $\mathcal{I}_{Z_t}$, a contradiction. □

Example 2.5. We give two examples in which Theorem 2.3 applies.

(a) If $X \subset \mathbb{P}^N$ is a local complete intersection of codimension r at p, then $\mathcal{I}_{X,p} = (f_1, \ldots, f_r) \subset \mathcal{O}_{\mathbb{P}^N,p}$, where $f_1, \ldots, f_r$ cut out X at p. An embedded point is determined by a surjection $\varphi : \mathcal{I}_{X,p} \to k(p)$. After changing generators, we may assume $\varphi(f_1) = 1$ and $\varphi(f_i) = 0$ for $i > 1$ so that the ideal for the corresponding subscheme Y locally at p is $(m_p f_1, f_2, \ldots, f_r)$.

(b) Use $[x, y, z, w]$ to denote the coordinates of $\mathbb{P}^3$. Let $C \subset \mathbb{P}^3$ be the union of three coordinate axes with ideal $I_C = (xy, xz, yz)$. Away from the origin $[0, 0, 0, 1]$, C is a local complete intersection. Working on the affine patch $w \neq 0$, one computes that the blow-up at C has fiber $\mathbb{P}^2$ over the origin, so condition (b) of Proposition 2.2 holds at each point. It follows from Theorem 2.3 that any subscheme D obtained from C by adding an embedded point is a flat limit of C with an isolated point.

Example 2.6. We give two examples where Theorem 2.3 does not apply.

(a) Fix a line $L \subset \mathbb{P}^3$ and define X by $\mathcal{I}_X = \mathcal{I}_L^d$ with $d > 1$. Then $\mathcal{I}_X$ is generated by $d+1$ elements at each $p \in X$ ($I_X = I_L^d$), but $\pi^{-1}(p) \cong \mathbb{P}^1$ because the blow-ups of $\mathbb{P}^3$ at $\mathcal{I}_L$ and $\mathcal{I}_L^d$ are isomorphic [Hartshorne 1977, II, Example 7.11(a)], so condition (b) of Proposition 2.2 fails.

(b) The curve $X \subset \mathbb{P}^3$ with ideal (x^2, xy, y^3) is ACM with locally 3-generated ideal sheaf at each point $p \in X$; hence it is not possible that $\pi^{-1}(p) \cong \mathbb{P}^2$ for each $p \in C$, for then the exceptional divisor would have dimension 3. Therefore a general embedded point cannot be detached while leaving X fixed. Nevertheless, such an embedded point can be detached in the Hilbert scheme due to the following.

Proposition 2.7. *Let $X_0 \subset \mathbb{P}^N$ be ACM of codimension two with 3-generated homogeneous ideal I_{X_0}. Then each subscheme Y obtained from X_0 by adding an embedded point of multiplicity one is the flat limit of local complete intersection ACM subschemes union an isolated point.*

Proof. Write $\mathcal{O}$ for the structure sheaf of $\mathbb{P}^N$. Since X_0 is ACM and I_{X_0} is 3-generated, the ideal sheaf has minimal resolution

$$0 \to \bigoplus_{j=1}^{2} \mathcal{O}(-b_j) \xrightarrow{\psi_0} \bigoplus_{i=1}^{3} \mathcal{O}(-a_i) \xrightarrow{\pi_0} \mathcal{I}_{X_0} \to 0 \tag{4}$$

and $\mathcal{I}_Y$ is the kernel of a surjection $\varphi : \mathcal{I}_{X_0} \to \mathcal{O}_p$. Our strategy is to deform the exact sequence (4) along with φ to obtain subschemes X_t that are local complete intersections and maps $\varphi_t : \mathcal{I}_{X_t} \to \mathcal{O}_{p_t}$ to define the family Y_t. We carry this out in steps:

Claim 1. *There is a lift of* $\varphi \circ \pi_0 : \bigoplus_{i=1}^{3} \mathcal{O}(-a_i) \to \mathcal{O}_p$ *to* $\tilde{\varphi} : \bigoplus_{i=1}^{3} \mathcal{O}(-a_i) \to \mathcal{O}$ *such that the composite*

$$\tilde{\varphi} \circ \psi_0 : \bigoplus_{j=1}^{2} \mathcal{O}(-b_j) \to \mathcal{O}$$

is induced by multiplying (F, G), *where* F *and* G *are homogeneous polynomials of degrees* b_1 *and* b_2 *with no common factor.*

Claim 2. *There is a map*

$$\psi_1 : \bigoplus_{j=1}^{2} \mathcal{O}(-b_j) \to \bigoplus_{i=1}^{3} \mathcal{O}(-a_i)$$

whose cokernel is the ideal sheaf of a local complete intersection X_1.

Once we have established the claims, the rest is straightforward. Construct the linear deformation $\psi_t = t\psi_1 + (1-t)\psi_0$ for $t \in \mathbb{A}^1$ and write the composite maps $\tilde{\varphi} \circ \psi_t : \bigoplus_{j=1}^{2} \mathcal{O}(-b_j) \to \mathcal{O}$ as (F_t, G_t). Then the schemes S_t given by $F_t = G_t = 0$ are complete intersections in a neighborhood of $t = 0$ because this is true for S_0 by construction. If $S \subset \mathbb{P}^N \times \mathbb{A}^1$ is the total family, there is an integral curve T through $(p, 0)$ inside S which is not vertical at $(p, 0)$ and base extension by $T \to \mathbb{A}^1$ allows us to pick out a moving point $p_t \in S_t$ with $p_0 = p$. By abuse of notation we will use the same letter t for the parameter.

Let

$$\varphi_t : \bigoplus_{i=1}^{3} \mathcal{O}(-a_i) \xrightarrow{\tilde{\varphi}} \mathcal{O} \to \mathcal{O}_{p_t} = k(p_t)$$

be the composition. For general $t \neq 0$, $\operatorname{Coker} \psi_t$ is the ideal sheaf of an ACM local complete intersection X_t and $\varphi_t \circ \psi_t = 0$ by construction (since $p_t \in S_t$); hence we get induced maps $\mathcal{I}_{X_t} \to \mathcal{O}_{p_t}$. Since φ_0 is onto, so is φ_t for general t, the kernels giving a family of ideals $\mathcal{I}_{Y_t}$ for a family of local complete intersections X_t converging to X_0 along with a point p_t converging to $p = p_0$. If $p_t \notin X_t$, then we are done. If $p_t \in X_t$ for each t, then we have at least shown that Y is a flat limit of complete intersections having an embedded point. By Proposition 2.2, these are flat limits of local complete intersections with an isolated point and we again conclude the proof. It remains to establish the two claims above.

Proof of Claim 1. The composition $\varphi \circ \pi : \bigoplus_{i=1}^{3} \mathbb{O}(-a_i) \to \mathbb{O}_p$ lifts to

$$\widetilde{\varphi} : \bigoplus_{i=1}^{3} \mathbb{O}(-a_i) \to \mathbb{O}$$

because $H^0_*(\mathbb{O}_{\mathbb{P}^3}) \to H^0_*(\mathbb{O}_p)$ is surjective in positive degrees. Let us write this map as $\widetilde{\varphi} = (A_1, A_2, A_3) \in H^0(\bigoplus_{i=1}^{3} \mathbb{O}(a_i))$. Then the general such lift $\widetilde{\varphi}$ may be written $(A_1 + B_1, A_2 + B_2, A_3 + B_3)$ with $B_i \in I_p$. Writing

$$\psi_0 = \begin{pmatrix} f_1 & f_2 & f_3 \\ g_1 & g_2 & g_3 \end{pmatrix},$$

the desired composite map is given by $(F, G) = (\sum(A_i + B_i) f_i, \sum(A_i + B_i) g_i)$ and we need to show that F and G have no common factor. For this it suffices to show that the zero loci of F and G meet properly. Letting L be a line missing X_0 (and p), we will show that there are no common zeros along L, for general B_i. Restricting the resolution (4) to L and dualizing yields the exact sequence

$$0 \to \mathbb{O}_L \to \bigoplus_{i=1}^{3} \mathbb{O}_L(a_i) \xrightarrow{\psi_0^\vee \otimes \mathbb{O}_L} \bigoplus_{j=1}^{2} \mathbb{O}_L(b_j) \to 0.$$

Since $b_j > 0$, the rank-two bundle on the right has a nonvanishing section, which lifts to a section (r_1, r_2, r_3) of the rank-three bundle $\bigoplus_{i=1}^{3} \mathbb{O}_L(a_i)$. Since the equations in I_p of degree $d > 0$ cut out the complete linear system $H^0(\mathbb{O}_L(d))$, we can find $B_i \in I_p$ such that $(A_i + B_i)\big|_L = r_i$, for $i = 1, 2, 3$, and this choice proves our claim because the nonvanishing image of (r_1, r_2, r_3) in $\bigoplus_{j=1}^{2} \mathbb{O}_L(b_j)$ is given by the restrictions of the polynomials $\sum(A_i + B_i) f_i$ and $\sum(A_i + B_i) g_i$; hence these have no common zeros along L. □

Proof of Claim 2. It is well-known that the degeneracy locus X_1 of the general such map ψ_1 is codimension two and regular in codimension one [Chang 1989]. Here we show that X_1 is a local complete intersection as well. In the exact sequence (4) we may take $a_1 \le a_2 \le a_3$, $b_1 \le b_2$, and $b_1 > a_1$ (if $b_1 = a_1$, we can cancel off this summand and X_0 is a complete intersection, when Claim 2 is clear). Since $\sum b_j = \sum a_i$ (because $c_1(\mathcal{I}_{X_0}) = 0$), it follows that $d_1 = a_3 + a_2 - b_1 = b_2 - a_1 > 0$ and $d_2 = a_3 + a_2 - b_2 = b_1 - a_1 > 0$, so let Z be a complete intersection of two general hypersurfaces of degrees d_1 and d_2. It is easy to check $d_2 \le d_1 \le a_3$ and $d_2 \le a_2$, therefore we can link Z to X by a complete intersection $C = K_1 \cap K_2$ of hypersurfaces of degrees a_2 and a_3. The inclusion

$$\begin{tikzcd}
0 \arrow[r] & \mathbb{O}(b_1 + b_2 - 2a_2 - 2a_3) \arrow[r] & \mathbb{O}(b_2 - a_2 - a_3) \oplus \mathbb{O}(b_1 - a_2 - a_3) \arrow[r] & \mathcal{I}_Z \arrow[r] & 0 \\
0 \arrow[r] & \mathbb{O}(-a_2 - a_3) \arrow[r] \arrow[u] & \mathbb{O}(-a_2) \oplus \mathbb{O}(-a_3) \arrow[r] \arrow[u] & \mathcal{I}_C \arrow[r] \arrow[u] & 0,
\end{tikzcd}$$

and the cone construction from liaison theory [Migliore 1998, Proposition 5.2.10] yields the resolution

$$0 \to \mathcal{O}(-b_1) \oplus \mathcal{O}(-b_2) \to \mathcal{O}(-a_1) \oplus \mathcal{O}(-a_2) \oplus \mathcal{O}(-a_3) \to \mathcal{I}_X \to 0.$$

Hence X has the same type of resolution as X_0. By Bertini's theorem, the general hypersurface K_2 containing Z is smooth, so X is Cartier on K_2 and hence a local complete intersection. Now just take $X_1 = X$ and the claim is proved. □

Example 2.8. The easiest way to construct curves in $\mathbb{P}^3$ satisfying the hypotheses of Proposition 2.7 is by linking to a complete intersection, as in the proof of Claim 2.

(a) Any purely one-dimensional curve $C \subset \mathbb{P}^3$ of degree 3 and genus 0 is ACM [Piene and Schlessinger 1985] and has a resolution of the form

$$0 \to \mathcal{O}(-3)^2 \to \mathcal{O}(-2)^3 \to \mathcal{I}_C \to 0$$

as noted in [Ellingsrud 1975, Example 1], and links to a line by a complete intersection of two quadric surfaces. In particular, this holds for the triple line with ideal (x^2, xy, y^2).

(b) If $C \subset \mathbb{P}^3$ is any locally Cohen–Macaulay curve of degree 4 and genus 1, then C is nonplanar, so $h^1(\mathcal{I}_{C_0}(n)) \le (d-2)(d-3)/2 - g = 0$ for all n [Martin-Deschamps and Perrin 1993, Theorem 1.3] and therefore C is ACM. Now $\chi(\mathcal{I}_C(1)) = 0$, so $H^2(\mathcal{I}_C(1)) = 0$. Furthermore $H^1(\mathcal{I}_C(2)) = 0$ (C is ACM) and $H^3(\mathcal{I}_C(0)) = 0$ so $\mathcal{I}_C$ is Mumford 3-regular. In particular $\mathcal{I}_C(3)$ is generated by global sections, and we can link C by the complete intersection of a quadric and cubic to a curve D of degree 2 and genus 0. Since D is planar, it is a complete intersection, so using the method of the proof of Claim 2, above, we see that C has resolution

$$0 \to \mathcal{O}(-4) \oplus \mathcal{O}(-3) \to \mathcal{O}(-3) \oplus \mathcal{O}(-2)^2 \to \mathcal{I}_C \to 0$$

and again Proposition 2.7 applies to C. The quadruple line with ideal (x^2, xy, y^3) is such an example, explaining Example 2.6(b).

Sometimes a one-dimensional subscheme D with embedded points is not a flat limit of curves C union isolated points even if one allows C to deform. In other words, the Hilbert scheme can have irreducible components whose general member has an embedded point.

Example 2.9. We exhibit an irreducible component of $\mathrm{Hilb}^{4z+1-g}(\mathbb{P}^3)$ whose general member has an embedded point for any $g \le -15$. The irreducible components of the Hilbert schemes $H_{4,g}$ of locally Cohen–Macaulay curves of degree 4 and arithmetic genus g are known [Nollet and Schlesinger 2003, Table III]. We note two typographical errors in the table, namely the family G_5 of double conics has dimension $13 - 2g$ instead of $13 - 3g$ [Nollet and Schlesinger 2003, p. 189] and the general member of family $G_{7,a}$ should be $W \cup_{3p} L$ instead of just W. Now consider the irreducible component G_4 of dimension $9-3g$, consisting of thick quadruple lines. Each curve $[C] \in G_4$ has a supporting line L and there is an exact sequence

$$0 \to \mathscr{I}_C \to \mathscr{I}_W \xrightarrow{\phi} \mathbb{O}_L(-g-1) \to 0,$$

where W is the triple line given by $\mathscr{I}_W = \mathscr{I}_L^2$ [Nollet and Schlesinger 2003, Proposition 2.1]. The surjection ϕ factors through $\mathscr{I}_W \otimes \mathbb{O}_L \cong \mathbb{O}_L(-2)^2$, hence is given by $\phi(x^2) = a$, $\phi(xy) = b$, and $\phi(y^2) = c$ for three homogeneous polynomials a, b, and c of degree $-g+1$. Writing the ideal of C as

$$I_C = (x^3, x^2y, xy^2, y^3, axy - bx^2, by^2 - cxy),$$

we see that at general point $p \in L$, a, b, and c are units in the local ring $\mathbb{O}_{\mathbb{P}^3,p}$, therefore $I_{C,p} = (x^3, axy - bx^2, by^2 - cxy)$ and I_C is generically 3-generated for general ϕ.

Now consider the locus $V \subset \mathrm{Hilb}^{4z+2-g}(\mathbb{P}^3)$ obtained by adding an isolated or embedded point to C as above, as in Proposition 2.1. The closure of the component corresponding to C along with isolated points has dimension three. Since I_C is generically 3-generated, the set of embedded point structures at general $p \in C$ is parametrized by $\mathbb{P}^2$ and we obtain a second three-dimensional family. Thus V is reducible with at least these two three-dimensional components (conceivably the locus where I_C is generated by more elements could generate another family). Varying the curve $[C] \in G_4$, we obtain at least two corresponding families of dimension $12-3g$ (because $\dim G_4 = 9-3g$). Let F be the closure of the family whose general curve has an embedded point.

We claim that F is an irreducible component of $\mathrm{Hilb}^{4z+2-g}(\mathbb{P}^3)$. The general member $[D] \in F$ cannot be a flat limit of curves possessing more than one isolated or embedded point (counted with multiplicity). Since G_4 is an irreducible component of $H_{4,g}$ for $g \le -2$, D is not a flat limit of another family of curves with an isolated or embedded point of multiplicity one, because the underlying locally Cohen–Macaulay curve $C \subset D$ is not. Finally D cannot be a flat limit of locally Cohen–Macaulay curves of genus $g-1$ because the maximal dimension of such a family for $g \le -15$ is $12-3g = \dim F$.

3. Detaching embedded points of multiplicity two or three

In this section we prove that if Y has embedded points of multiplicity two (see Proposition 3.3) or three (see Proposition 3.7) and the underlying subscheme $X \subset \mathbb{P}^N$ is a local complete intersection of codimension two, then Y is a flat limit of X union isolated points. Along with Theorem 2.3, this shows that an embedded point of multiplicity at most three can be detached from X, from which we deduce our main result, Theorem 1.4.

We begin with several propositions that take care of the easier cases, leaving the more difficult cases to Proposition 3.7. We also show that these results may fail for local complete intersections of codimension greater than two (Example 3.6) and for embedded points of multiplicity greater than three (Example 3.10).

Proposition 3.1. *Let $X \subset \mathbb{P}^N$ be a local complete intersection of codimension two, Z a zero-dimensional subscheme of embedding dimension at most one and suppose that Y is defined by the exact sequence*

$$0 \to \mathcal{I}_Y \to \mathcal{I}_X \xrightarrow{\varphi} \mathcal{O}_Z \to 0.$$

Then Y is a flat limit of X union isolated points.

Proof. Since the result is local, we may assume that Z is supported at a point p and has length d. Since Z has embedding dimension ≤ 1, we can choose a smooth connected curve C_0 of high degree containing Z and not entirely in X. If $p \notin X$, the result is clear because Z is a flat limit of isolated points in C_0. In the interesting case $p \in X$, our idea is to take a deformation C_t of C_0 and use d isolated points in C_t to perform the detaching process.

Let C be a translation of C_0 by $\mathrm{PGL}(N+1)$ which misses X. Now for $m \gg 0$, the general pair $F, G \in H^0(\mathcal{I}_X(m))$ give hypersurfaces which cut out X in an open neighborhood of p. Write $\mathcal{O}$ for the structure sheaf of $\mathbb{P}^N$. For the purposes of this proof we may assume that X is equal to the complete intersection defined by F and G, giving the Koszul resolution

$$0 \to \mathcal{O}(-2m) \xrightarrow{\psi} \mathcal{O}(-m) \oplus \mathcal{O}(-m) \xrightarrow{\pi} \mathcal{I}_X \to 0. \tag{5}$$

Because the restriction map $H^0(\mathcal{O}(m)) \to H^0(\mathcal{O}_Z(m))$ is surjective for $m \gg 0$, we can lift the images of F, G to $\mathcal{O}$, hence the composition $\varphi \circ \pi : \mathcal{O}(-m)^2 \to \mathcal{O}_Z$ factors through $\mathcal{O}$ and we obtain $\widetilde{\varphi} : \mathcal{O}(-m)^2 \to \mathcal{O}$ inducing φ. The composition $\widetilde{\varphi} \circ \psi$ vanishes on a hypersurface S of degree $2m$ containing both X and Z.

By Bertini's theorem, we could have chosen the equations F and G cutting out X near p to be smooth away from X, meeting C in disjoint reduced sets of points, so the restrictions to C induce a sheaf surjection $\mathcal{O}_C^2 \to \mathcal{O}_C(m)$. If $\widetilde{\varphi}$ is given by $A_0, B_0 \in H^0(\mathcal{O}(m))$, then S has equation $FA_0 + GB_0 = 0$, but A_0 and B_0 are only determined up to elements of $H^0(\mathcal{I}_Z(m))$. Since the natural map $H^0(\mathcal{I}_Z(m))^2 \to H^0(\mathcal{O}_C(m))$ is surjective, given by $(A, B) \mapsto FA + GB$, we may choose A_0 and B_0 to assume that $S \cap C$ is a reduced set of $2m(\deg C)$ points.

Now consider a family of translations C_t from C to C_0, parametrized by $t \in \mathbb{A}^1$. Now $C_0 \cap S$ contains Z at p and $C_t \cap S$ consists of $2m(\deg C)$ reduced points for general $t \neq 0$. Possibly after a base extension, we may pick d distinct points

$p_{1,t}, p_{2,t}, \dots, p_{d,t}$ in $S \cap C_t$ near p. Letting $Z_t = \{p_{1,t}, p_{2,t}, \dots, p_{d,t}\}$, the flat limit of Z_t is exactly Z, because the ideal of the limit contains the equations of the curve C_0 by construction, and Z is the unique length-d subscheme of C_0 at p.

Letting φ_t be the composition

$$\mathcal{O}(-m)^2 \xrightarrow{\widetilde{\varphi}} \mathcal{O} \to \mathcal{O}_{Z_t},$$

we have $\varphi_t \circ \psi = 0$ by construction, so these maps factor through $\mathcal{I}_X$ and we obtain a family of maps $\varphi_t : \mathcal{I}_X \to \mathcal{O}_{Z_t}$. Since $\varphi_0 = \varphi$ is surjective, so are φ_t for t near 0 and the family $\mathcal{I}_{Y_t} = \ker \varphi_t$ gives the desired family. □

Proposition 3.2. *Let $X \subset \mathbb{P}^N$ be a local complete intersection, K a sheaf of finite length supported at p, and Y and Y^1 defined by the commutative diagram of short exact sequences*

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{I}_Y \arrow[r] \arrow[d] & \mathcal{I}_X \arrow[r, "\varphi=(\alpha,\beta)"] \arrow[d] & K \oplus \mathcal{O}_p \arrow[r] \arrow[d, "\pi_1"] & 0 \\
0 \arrow[r] & \mathcal{I}_{Y^1} \arrow[r] & \mathcal{I}_X \arrow[r, "\alpha"] & K \arrow[r] & 0.
\end{tikzcd}$$

Then Y is a flat limit of Y^1 union an isolated point. In particular, if Y^1 is a flat limit of X union isolated points, then so is Y.

Proof. The result is local at p. The direct sum allows us to write $\varphi = (\alpha, \beta)$, where α defines Y^1 as above. The surjection $\beta : \mathcal{I}_X \to \mathcal{O}_p$ defines an embedded point structure Y^2 on X. Since X is a local complete intersection, Y^2 is a flat limit of X union an isolated point by Theorem 2.3, meaning that there is a flat family Y_t^2 for $t \in T$ with $Y_0^2 = Y^2$ and $Y_t^2 = X \cup p_t$ with $p_t \notin X$ for $t \neq 0$. This gives a family of surjections $\beta_t : \mathcal{I}_X \to \mathcal{O}_{p_t}$ with $\mathcal{I}_{Y_t^2} = \ker \beta_t$ and $\beta_0 = \beta$.

Let $\gamma_t : \mathcal{I}_{Y^1} \subset \mathcal{I}_X \xrightarrow{\beta_t} \mathcal{O}_{p_t}$ be the composite map. Clearly γ_t is surjective for $t \neq 0$, because $p_t \notin \mathcal{I}_X$, so the inclusion $\mathcal{I}_{Y^1} \subset \mathcal{I}_X$ is an equality at these points. The map γ_0 is also a surjection, since, locally at p, if we choose $f \in \mathcal{I}_X$ such that $\varphi(f) = (0, 1)$, then $\alpha(f) = 0 \Rightarrow f \in \mathcal{I}_{Y^1}$ and $\gamma_0(f) = 1$. This family of maps gives a flat family Y_t, and for $t \neq 0$ Y_t consists of Y^1 union an isolated point. Finally, the kernel of $\gamma_0 : \mathcal{I}_{Y^1} \to \mathcal{O}_p$ is exactly $\mathcal{I}_Y$, for $g \in \mathcal{I}_{Y^1} \Rightarrow g \in \mathcal{I}_X$ and $\alpha(g) = 0$. Now $\gamma_0(g) = 0 \iff \beta(g) = 0 \iff \varphi(g) = 0 \iff g \in \mathcal{I}_Y$. □

Proposition 3.3. *Let $X \subset \mathbb{P}^N$ be a local complete intersection of codimension two and obtain Y by adding an embedded point of multiplicity two with associated exact sequence*

$$0 \to \mathcal{I}_Y \to \mathcal{I}_X \xrightarrow{\varphi} K_p \to 0,$$

where K_p is a sheaf of length 2 supported at p. Then either (a) *$K_p \cong \mathcal{O}_p \oplus \mathcal{O}_p$, or* (b) *$K_p \cong \mathcal{O}_Z$, where $Z \subset \mathbb{P}^N$ has length 2. In either case, Y is a flat limit of X union two isolated points.*

Proof. If $K_p \cong \mathcal{O}_p \oplus \mathcal{O}_p$, apply Proposition 3.2. Since Y^1 is obtained from X by adding an embedded point of multiplicity one, it is a flat limit of X union an isolated point, hence Y is a flat limit of X union two isolated points.

Now suppose $K_p \ncong \mathcal{O}_p \oplus \mathcal{O}_p$. Then the surjection $K_p \to K_p \otimes k(p)$ is not an isomorphism, thus $K_p \otimes \mathcal{O}_p$ is one-dimensional as an $\mathcal{O}_p = k(p)$ vector space. Therefore K_p is principal by Nakayama's lemma, so there is a surjection $\mathcal{O} \to K_p$ whose kernel is the ideal sheaf $\mathcal{I}_Z$ of a length-2 subscheme, which is contained in a unique line and has embedding dimension one. We apply Proposition 3.1 to see that Y is a flat limit of X union two isolated points. $\square$

Remark 3.4. We give the local equations of the embedded point structures for cases (a) and (b) of Proposition 3.3 for $X \subset \mathbb{P}^N$:

(a) If $I_{X,p} = (f, g)$, then $I_{Y,p} = m_p \cdot I_{X,p}$.

(b) Replacing generators so that $\varphi(f) = 1$ and $\varphi(g) = 0$, we obtain $I_{Y,p} = (g, f \cdot I_Z)$, Z being the length-2 subscheme.

Example 3.5. In case (b) of Proposition 3.3, there is a unique subscheme $X \subset E \subset Y$ with an embedded point of multiplicity one, because the unique length-1 quotient of $\mathcal{O}_Z$ is $\mathcal{O}_p$, obtained by modding out by the maximal ideal. Using such subschemes E, we explain why it was necessary to prove case (b) by pulling away two points simultaneously. For example, let $X \subset \mathbb{A}^3$ have ideal $I_X = (x^2, y^2)$ and let $p = (0, 0, 0)$, where $[x, y, z]$ denotes the coordinates of $\mathbb{A}^3$. Add an embedded point to X at p using the map $I_X \to k$ by $x^2 \mapsto 1$, $y^2 \mapsto 0$ to obtain E with $I_E = (y^2, x^3, x^2y, x^2z)$ being 4-generated. By Proposition 2.2, one can add a second point at p to obtain Y with an embedded point of multiplicity two, which is not a flat limit of E union an isolated point.

Example 3.6. Proposition 3.3 may fail for local complete intersections of codimension greater than two. For example, suppose that $C \subset \mathbb{P}^4$ is the complete intersection with ideal $I_C = (x^2, y^2, z^2)$, where $[x, y, z, u, w]$ denotes the projective coordinates. Consider the double point structures D on C given by surjections $\phi : (x^2, y^2, z^2) \to K = \mathcal{O}_Z$, where Z is the double point with ideal $I_Z = (x, y, z, u^2)$. An arbitrary map $\phi : I_C \to S/I_Z$ is given by

$$\phi(x^2) = a + bu, \quad \phi(y^2) = c + du, \quad \phi(z^2) = e + fu,$$

where S is the coordinate ring of $\mathbb{P}^4$, $a, b, c, d, e, f \in k$, and any tuple (a, b, c, d, e, f) is possible because $I_C \subset I_Z$. The automorphisms of $K = \mathcal{O}_Z$ are given by multiplication by $A + Bu$ with $A \neq 0$. Thus the maps for which $\phi(x^2)$ generates K (that is, $a \neq 0$) are uniquely determined up to automorphisms of K by the quotients $\phi(y^2)/\phi(x^2) = (c + du)/(a + bu)$ and $\phi(z^2)/\phi(x^2) = (e + fu)/(a + bu)$. By Remark 1.8(a), these quantities uniquely determine the corresponding subschemes D. In other words, if we compose the map above by the automorphism of $\mathcal{O}_Z$ given by multiplication by $(a+bu)^{-1} = (a-bu)/a^2$, we may assume that $\phi(x^2) = 1$ when the corresponding ideal of D is given by

$$(x^2(I_Z), y^2-(c+du)x^2, z^2-(e+fu)x^2) \\ = (x^3, x^2y, x^2z, x^2w^2, y^2-(c+du)x^2, z^2-(e+fu)x^2)$$

and each tuple $(c,d,e,f) \in k^4$ yields a distinct subscheme D, so we obtain a four-dimensional family of such double point structures D.

Finally, the same argument applies to any double point structure D on C. Since there is a choice of any point $p \in C$ for the support of $K = \mathcal{O}_Z$ and the structure of Z is uniquely determined by a line through p (parametrized by a hyperplane $\mathbb{P}^3$), the family of such double point structures has dimension $1+3+4=8$. The general such structure cannot be a flat limit of C union two isolated points, for this family also has dimension eight.

Now we turn to the case of multiplicity three.

Proposition 3.7. *Let $X \subset \mathbb{P}^N$ be a local complete intersection of codimension two. Let Y be the subscheme obtained from X by an exact sequence*

$$0 \to \mathcal{I}_Y \to \mathcal{I}_X \xrightarrow{\varphi} K \to 0,$$

*where K is a length-*3 *sheaf supported at p. Then one of the following holds*:

(a) *$K \cong \mathcal{O}_p \oplus \mathcal{O}_Z$, where $Z \subset \mathbb{P}^N$ is a double point on a line.*

(b) *$K \cong \mathcal{O}_Z$, where $Z \subset \mathbb{P}^N$ is a triple point on a line.*

(c) *$K \cong \mathcal{O}_Z$, where $Z \subset \mathbb{P}^N$ is a triple point on a smooth conic.*

(d) *$K \cong \mathcal{O}_Z$, where Z is contained in a plane H and $\mathcal{I}_{Z,H} = \mathcal{I}^2_{p,H}$.*

(e) *$K \cong \mathrm{Hom}_{\mathcal{O}_p}(\mathcal{O}_Z, \mathcal{O}_p)$ with Z as in case* (d).

In each case, *Y is a flat limit of X union three isolated points.*

Proof. If K is a direct summand, one summand is $\mathcal{O}_p$ and the other is $\mathcal{O}_p^2$ or $\mathcal{O}_Z$ for a double point Z. The former is not possible as a quotient of the locally 2-generated ideal $\mathcal{I}_C$, leading to case (a). If K is principal, then the surjection $\mathcal{O} \to K$ shows that $K \cong \mathcal{O}_Z$ for a length-3 subscheme supported at p. Since $h^0(\mathcal{O}_Z(1)) = 3$ and $h^0(\mathcal{O}(1)) = N+1$, Z is a planar triple point. It is easy to classify planar triple points, leading to cases (b), (c), and (d). If K is not principal and not a direct summand, then it is 2-generated as a quotient of $\mathcal{I}_X$ via φ. The two generators have a common nonzero multiple, otherwise they would express K as a direct sum of two principal modules. The common nonzero multiple is therefore a generator of the dual $\mathrm{Hom}(\mathcal{O}_Z, \mathcal{O}_p)$, where $\mathcal{O}_Z$ must be one of cases (b), (c), or (d). However, the duals to cases (b) and (c) are principal and we are left with the dual of case (d), which is case (e).

That Y is a flat limit of X union three isolated points follows from Propositions 3.2 and 3.3 in case (a) and from Proposition 3.1 in cases (b) and (c). Cases (d) and (e) require new ideas.

In case (d) we have $K_p \cong \mathcal{O}_Z$, where $Z \subset H$ is the planar triple point supported at p of embedding dimension two. As in Proposition 3.1, X is contained in hypersurfaces with equations F and G of degree $m \gg 0$, giving a Koszul resolution (5), $\varphi : \mathcal{I}_X \to \mathcal{O}_Z$ lifts to $\tilde{\varphi} : \mathcal{O}(-m)^2 \to \mathcal{O}$ and there is a hypersurface S of degree $2m$, where $\tilde{\varphi} \circ \psi = 0$. The intersection $H \cap S$ contains an integral curve T passing through p. Our idea is to realize this triple embedded structure as the flat limit of a fixed double embedded structure at p union a single point varying in T.

Let $\tilde{T} \xrightarrow{f} T \subset H \cong \mathbb{P}^2$ be the normalization of T and choose a point $0 \in T$ such that $f(0) = p$. For $t \neq 0$, let $L_t \subset H$ be the line through p and $f(t)$. As $t \to 0$, $f(t) \to p$ and the line L_t has a unique limit L_0 (complete the associated map $T - \{0\} \to (\mathbb{P}^2)^\vee$ to obtain this limiting line). Choose local coordinates x, y on $\mathbb{A}^2 \subset \mathbb{P}^2$ so that $p = (0, 0)$ and $L_0 = \{x = 0\}$. The double point W at p with ideal (x^2, y) is a closed subscheme of Z (which has ideal (x^2, xy, y^2)). We now show that $\lim_{t \to 0} f(t) \cup W = Z$ in the Hilbert scheme of length-3 subschemes of H. If $f(t) = (a(t), b(t))$ in the local coordinates above, then the ideal for $W \cup f(t)$ is

$$I_t = (x^2, y) \cap (x - a(t), y - b(t)),$$

which contains the product of the two ideals. Since $\lim_{t \to 0}(a(t), b(t)) = (0, 0)$, the limiting ideal contains (x^3, xy, y^2). If the line L_t has equation $l_t = 0$, then $l_t^2 \in I_t$ and by choice of coordinates we have $\lim_{t \to 0} l_t^2 = x^2$, so the limiting ideal also contains x^2 and hence (x^2, xy, y^2), which defines Z.

The rest is analogous to Proposition 3.1. The composite map

$$\mathcal{O}(-2m) \xrightarrow{\psi} \mathcal{O}(-m)^2 \xrightarrow{\tilde{\varphi}} \mathcal{O} \to \mathcal{O}_S \to \mathcal{O}_{W \cup f(t)}$$

is zero, inducing a family of maps $\varphi_t : \mathcal{I}_X \to \mathcal{O}_{W \cup f(t)}$. Since φ_0 is onto, so is φ_t for t near 0. Therefore the kernels $\mathcal{I}_{Y_t}$ give a flat family whose limit is Y as $t \to 0$. Using our earlier results, for $t \neq 0$ each Y_t is a flat limit of X union isolated points, and therefore so is Y.

Finally consider case (e), where $K_p \cong \mathrm{Hom}_{\mathcal{O}_p}(\mathcal{O}_Z, \mathcal{O}_p)$ with $\mathcal{O}_p = k(p)$ the residue field at p and $Z \subset H \subset \mathbb{P}^N$, with H a plane and $\mathcal{I}_{Z,H} = \mathcal{I}_{p,H}^2$. Choose affine coordinates $x, y, z_1, \ldots, z_{N-2}$ centered at p so that $I_H = (z_1, \ldots, z_{N-2})$ and x, y are coordinates for $H \cong \mathbb{A}^2$. Let f, g be the restrictions of F, G to this affine patch, so that $I_X = (f, g)$. If $g - uf = h \in I_H$ for some unit u in the local ring, replace g with h as a generator for I_X. In this way we may assume $g \in I_H$ or (f, g) is not principal modulo I_H locally around p. Now $\mathcal{O}_Z$ is generated by $1, x, y$ as an $\mathcal{O}_p$-vector space, so K_p is generated by dual basis $x^*, y^*, 1^*$ as a vector space and by x^*, y^* as an $\mathcal{O}_H$-module with structure given by $xx^* = 1^* = yy^*$, $xy^* = yx^* = 0$.

Since φ is surjective, $\varphi(f) = ax^* + by^* + c1^*$ and $\varphi(g) = dx^* + ey^* + f1^*$ are also module generators for K_p, and in particular $ae - bd \neq 0$. Now consider the new coordinates $X = ay - bx$, $Y = ex - dy$ for H. With these one sees that $\mathrm{Ann}(\varphi(f)) = (I_H, X)$, $\mathrm{Ann}(\varphi(g)) = (I_H, Y)$, and $Y\varphi(f) = (ae-bd)1^* = X\varphi(g)$. It follows that $I_Y = (I_H(f, g), Xf, Yg, Yf - Xg)$. So by replacing the coordinates, we can present the ideal of Y as

$$I_Y = (I_H(f, g), xf, yg, yf - xg).$$

We will directly deform this ideal to obtain the result. The locus

$$S = \{(A, B, C, D) : \\ f(A, B) = 0,\ g(C, D) = 0,\ (B - D)g(C, B) - (C - A)f(C, B) = 0\}$$

contains $(0, 0, 0, 0)$ and each component has dimension ≥ 1; hence S contains an integral curve T through the origin. Let $\sigma : T \to H \times H \cong \mathbb{A}^4$ be the inclusion with coordinate functions $\sigma(t) = (a(t), b(t), c(t), d(t))$ and $0 \in T$ chosen so that $\sigma(0) = (0, 0, 0, 0)$. We claim that T can be chosen with $(a(t), b(t)) \neq (c(t), d(t))$. This is clear if $g \in I_H$, for then the second equation $g(C, D) = 0$ puts no restriction on C and D, and S is defined by only two equations: on a surface there are many integral curves T through the origin. The other possibility by our assumption is that $g \neq uf$ modulo I_H for any invertible u in an affine neighborhood of the origin. Here the restrictions of f and g to H have a greatest common divisor h so that $f = hf_1$ and $g = hg_1$ with f_1 and g_1 vanishing at the origin and relatively prime modulo I_H locally around the origin. If we look at the sublocus of S defined as above with f_1 and g_1 in place of f and g, the condition of the claim holds and we obtain the desired integral curve T.

Now consider the family of ideals

$$I_t = \big(I_H(f, g),\ (x - c(t))f,\ (y - b(t))g,\ (y - d(t))f - (x - a(t))g\big).$$

We claim that the ideal I_t scheme-theoretically cuts out exactly X and the three points $(a(t), b(t))$, $(c(t), b(t))$, and $(c(t), d(t))$ (which may be isolated or embedded, two may coincide if $a(t) = c(t)$ or $b(t) = d(t)$) for generic t near 0.

The claim holds away from H via the generators $I_H(f, g)$. At points $(x, y) \in H$ away from (a, b), (c, b), and (c, d) (we suppress the variable t) the claim also holds. If $x \neq c$, then $x - c$ is a unit, $f \in I_t$ and there are two cases: if $x = a$, then $y \neq b$, hence $y - b$ is a unit and $g \in I_t$; otherwise $x \neq a$ and the last equation shows that $g \in I_t$. If $x = c$, then $y \neq b, d$, so $g \in I_t$ and $f \in I_t$ by the last equation.

Finally we consider $(x, y) \in \{(a, b), (c, b), (c, d)\}$. The claim is easily checked if these points are distinct ($a \neq c$ and $b \neq d$) by checking that length $I_X/I_t = 1$. For example, at $(x, y) = (a, b)$ we have $x \neq c$ so $f \in I_t$, when I_X/I_t is generated by g alone, and since $I_H g, (y-b)g, (x-a)g \in I_t$, we have $I_X/I_t \cong k$. The other points $(x, y) = (c, b), (c, d)$ are similar. In the degenerate case $a = c$, we need to show that length $I_X/I_t = 2$ at $(x, y) = (a, b) = (c, b)$. Here $y \neq d$ so $u = (y-d)$ is a unit and $uf - (x-a)g \in I_t$, showing that I_X/I_t is generated by g. Further I_t contains $I_H g$, $(y-b)g$, and $(x-c)^2 g$ (use $(x-c)f$ and $uf - (x-c)g$), so the quotient has length 2. The other degenerate case $b = d$ can be verified similarly. This proves the claim.

With the claim, the ideal I_t cuts out X and three other points (possibly embedded in X, but not all supported at the same point). Using our earlier results, these schemes are flat limits of X and isolated points. Since $\lim_{t\to 0}(a(t), b(t), c(t), d(t)) = (0, 0, 0, 0)$ by construction, we also have $\lim_{t\to 0} I_t = I_Y$, and we conclude. □

Remark 3.8. For $I_{X,p} = (f, g)$ locally at p in Proposition 3.7, we write local equations for the embedded point structure Y according to the various cases:

(a) If $K_p = \mathcal{O}_p \oplus \mathcal{O}_Z$ and $\varphi(f) = (1, 0)$, $\varphi(g) = (0, 1)$, then $I_{Y,p} = (f m_p, g I_Z)$ with $f \in I_Z$.

(b) If $K_p = \mathcal{O}_Z$ and $\varphi(f) = 1$, $\varphi(g) = 0$, then $I_{Y,p} = (f I_Z, g)$ with $g \in I_Z$.

(c) Similarly we have $I_{Y,p} = (f I_Z, g)$ with $g \in I_Z$.

(d) Again we have $I_{Y,p} = (f I_Z, g)$ with $g \in I_Z$.

(e) This is the most interesting structure. As shown in the proof, $I_{Y,p} = (xf - yg, yf, zf, xg, zg)$ for suitable coordinates x, y, z.

Putting these results together, we obtain our main theorem.

Theorem 3.9. *Let $X \subset \mathbb{P}^N$ be a local complete intersection of codimension two. If Y is obtained from X by adding embedded points of multiplicity at most three, then Y is a flat limit of X union isolated points.*

Proof. Suppose the embedded points are supported at $p_1, \dots, p_r$ with respective multiplicities $m_1, \dots, m_r \leq 3$. If Y_1 is the scheme which is isomorphic to Y near p_1 and equal to X away from p_1, it follows from Theorem 2.3 and Propositions 3.3 and 3.7 that Y_1 is in the Hilbert scheme closure of the family consisting of X union m_1 isolated points. Similarly if Y_2 is locally isomorphic to Y near p_2 and equal to X away from p_2, Y_2 is in the closure of the family of X union m_2 points. It follows that $Y_1 \cup Y_2$ is in the closure of the family of Y_1 union m_2 isolated points, the fixed embedded point at p_1 not affecting the relevant deformations. Since Y_1 union m_2 isolated points is in the closure of the family of X union $m_1 + m_2$ isolated points, we see that $Y_1 \cup Y_2$ is in this closure as well. Adding one point at a time in this way we find that Y is in the closure of the family of X union $m_1 + \cdots + m_r$ isolated points. □

Example 3.10. Here we show that it is not always possible to detach embedded points of multiplicity four. For linearly independent variables x, y, z, w, consider the $R = k[x, y, z, w]$-module K given by

$$K = \langle a, b\rangle/(za, wa, xb, yb, xa - zb, ya - wb).$$

In changing the choice of vector space basis for the linear forms x, y, z, w, we obtain a family of such modules on which the group GL(4) acts. It's easily checked that the R-module automorphisms of any fixed K have dimension five (one can write them down explicitly). For another K' determined by basis x', y', z', w' and an isomorphism $\psi : K \to K'$, the map ψ uniquely determines x', y', z', w' in terms of x, y, z, w, because the relations yield 16 equations in 16 unknowns. One can check that the family of candidate isomorphisms ψ has dimension 12 and a five-dimensional subspace corresponds to the identity coordinate change. Hence, we find that the isomorphism classes of such modules K has dimension $16 - (12 - 5) = 9$.

Now for $X \subset \mathbb{P}^N$ given by $I_X = (x^2, y^2)$, the family of embedded point structures on X given by such K has dimension $5N - 6$. The choice of the support of K has dimension equal to $\dim X = N - 2$; choosing the linear subspace $\langle x, y, z, w\rangle$ at p is given by $G(4, N)$ of dimension $4N - 16$; choosing the isomorphism class of K has dimension nine (see above); the choice of map $\varphi : \mathcal{I}_X \to K$ depends on eight parameters, but the resulting family of ideals $\mathcal{I}_Y$ given by the kernels has dimension three because the automorphisms of K have dimension five. All in all, the family has dimension $(N-2) + (4N-16) + 9 + (8-5) = 5N - 6$. For $N \geq 6$, we have $5N - 6 \geq 4N$, so the family cannot lie in the $4N$-dimensional closure of those obtained by unions of X with isolated points.

4. Applications to Hilbert schemes

In the previous section we proved various results about when a local complete intersection X with embedded points are flat limits of X union isolated points. In this section we apply these results to describe the irreducible components of certain Hilbert schemes. In view of Proposition 2.4, we deduce the following:

Theorem 4.1. *Let $p(z)$ be the Hilbert polynomial of a degree-d hypersurface in $\mathbb{P}^N$. Then:*

(a) *The Hilbert schemes* $\mathrm{Hilb}^{p(z)+e}(\mathbb{P}^N)$ *are irreducible for* $0 \leq e \leq 7$.

(b) *The Hilbert scheme* $\mathrm{Hilb}^{p(z)+1}(\mathbb{P}^N)$ *is smooth, isomorphic to* $\mathrm{Hilb}^{p(z)}(\mathbb{P}^N) \times \mathbb{P}^N$.

Proof. It follows from Proposition 2.4 that any (multiple) embedded point can be detached from a hypersurface, and for $e \leq 7$ we also know that any subscheme of length $e \leq 7$ is a flat limit of reduced points [Mazzola 1980; Cartwright et al. 2009]. Therefore $\mathrm{Hilb}^{p(z)+e}(\mathbb{P}^N)$ is the closure of the open subset formed by a degree-d

hypersurface and e isolated points and $\mathrm{Hilb}^{p(z)+e}(\mathbb{P}^N)$ is irreducible of dimension $\binom{d+N}{d} - 1 + Ne$.

Now take $e = 1$. It is easily checked that the Hilbert scheme is smooth at points corresponding to a hypersurface and an isolated point. Write $[x_0, x_1, \ldots, x_N]$ for the coordinates of $\mathbb{P}^N$. If $X \subset \mathbb{P}^N$ is a degree-d hypersurface and Y is obtained from X by adding an embedded point located at $x_1 = x_2 = \cdots = x_N = 0$, then the ideal of Y is simply $I_Y = (x_1, x_2, \ldots, x_N) \cdot I_X$, so I_Y is generated in degree $d+1$. Since the generator of $\mathcal{I}_X \to K$ is onto, $H^1(\mathcal{I}_Y(n)) = 0$ for $n \geq d$ and so the comparison theorem [Piene and Schlessinger 1985] applies (see also [Ellingsrud 1975; Kleppe 1979]). Now the argument of [Piene and Schlessinger 1985, Lemma 4, Case (iii)] goes through, which we include for self-containment: $H^0(\mathcal{N}_Y) = \mathrm{Hom}(I_Y, S/I_Y)_0$, where S is the coordinate ring of $\mathbb{P}^N$ and $\mathcal{N}_Y$ is the normal sheaf to Y. Given the dimension of $\mathrm{Hilb}^{p(z)+1}(\mathbb{P}^N)$, it suffices to prove that $\dim \mathrm{Hom}(I_Y, S/I_Y)_0 \leq \binom{d+N}{d} - 1 + N$. Setting $A = S/I_Y$ and $K = I_X/I_Y$, the S-module K has Koszul resolution of the form

$$0 \to S(-d-N) \to \cdots \to S(-d-2)^{N(N-1)/2} \to S(-d-1)^N \to S(-d) \to K \to 0.$$

Applying $\mathrm{Hom}(-, A)$ to this resolution shows $\mathrm{Hom}(K, A) = K(d)$ and $\mathrm{Ext}^1(K, A)$ is generated by the vectors $(f x_0)e_i$ with $1 \leq i \leq N$, where f is the defining equation of X. Applying $\mathrm{Hom}(-, A)$ to the short exact sequence $I_Y \to I_X \to K$ gives

$$0 \to \mathrm{Hom}(K, A) \to \mathrm{Hom}(I_X, A) \to \mathrm{Hom}(I_Y, A) \to \mathrm{Ext}^1(K, A) \to \cdots$$

but $\dim \mathrm{Hom}(K, A)_0 = \dim K(d)_0 = 1$ and $\mathrm{Hom}(I_X, A) \cong A(d)$, hence we have $\dim \mathrm{Hom}(I_X, A)_0 = \binom{d+N}{d}$. Since $\dim \mathrm{Ext}^1(K, A)_0 \leq N$ by the above, we conclude that the Hilbert scheme is smooth. The natural rational map $\mathrm{Hilb}^{p(z)}(\mathbb{P}^N) \times \mathbb{P}^N \to \mathrm{Hilb}^{p(z)+1}(\mathbb{P}^N)$ is actually a bijective morphism in view of the unique form of the ideal, and hence is an isomorphism by Zariski's main theorem. $\square$

We are also interested in Hilbert schemes of space curves and obtain the following irreducibility result for one-dimensional subschemes of high genus. Recall that if C is a space curve of degree d, then $g = p_a(C) \leq \binom{d-1}{2}$ with equality for plane curves.

Theorem 4.2. *The Hilbert scheme* $\mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3)$ *is irreducible for* (d, g) *satisfying* $d \geq 3$, $\binom{d-1}{2} - 4 < g \leq \binom{d-1}{2}$, *and* $g > \binom{d-2}{2}$, *with a general member consisting of a plane curve of degree d union isolated points.*

Proof. The Hilbert scheme is nonempty for all $g \leq \binom{d-1}{2}$ due to plane curves union isolated points. For $d \geq 3$, the genus of a nonplane curve satisfies $g \leq \binom{d-2}{2}$ [Hartshorne 1994], so if $[C] \in \mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3)$ and $C_0 \subset C$ is the curve remaining after removing embedded or isolated points, then C_0 is planar, hence a complete intersection. Since C is obtained by adding at most three embedded or isolated points, it is a flat limit of those with isolated points by Propositions 2.2, 3.3, and 3.7, and we conclude that the corresponding Hilbert scheme is irreducible. $\square$

When just one isolated or embedded point is added to a plane curve of degree d and genus $g=(d-1)(d-2)/2$, the resulting Hilbert scheme component is smooth:

Theorem 4.3. *For $g=(d-1)(d-2)/2$, the component $H_d \subset \mathrm{Hilb}^{dz+2-g}(\mathbb{P}^3)$ of the Hilbert scheme whose general member is a degree-d plane curve union an isolated point is smooth for all $d \geq 1$. Moreover, H_d is isomorphic to the blow-up of $\mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3)\times\mathbb{P}^3$ along the incidence correspondence.*

Proof. For $d=2$ and 3, this was proved in [Chen et al. 2011] and [Piene and Schlessinger 1985], respectively, even though H_d is not the full Hilbert scheme in these cases. For $d=1$ and $d\geq 4$, H_d is the full Hilbert scheme, and it suffices to compute the global sections $H^0(\mathcal{N}_D)$ of the normal sheaf associated to a point $[D]\in H_d$; so let D be the union of a plane curve C and the point $p=(0,0,0,1)$. If $p\notin C$, smoothness follows from $\mathcal{N}_D\cong\mathcal{N}_C\oplus\mathcal{N}_p$. If $p\in C$ is an embedded point, write $I_C=(z,f)$ with $f\in(x,y)$ and $z=0$ the equation of the plane H containing C. Consider the exact sequence (1):

$$0\to\mathcal{I}_D\to\mathcal{I}_C\xrightarrow{\varphi}\mathcal{O}_p\to 0.$$

If $\varphi(z)=0$, then $D\subset H$ and $h^0(\mathcal{N}_{D,H})=\binom{d+2}{2}+1$ from Theorem 4.1, so the exact sequence

$$0\to\mathcal{N}_{D,H}\to\mathcal{N}_{D,\mathbb{P}^3}\to\mathcal{O}_D(1)\to 0$$

yields $h^0(\mathcal{N}_{D,\mathbb{P}^3})\leq\binom{d+2}{2}+1+h^0(\mathcal{O}_D(1))$. If $d\geq 4$, then $h^0(\mathcal{O}_D(1))=4$ and we have $h^0(\mathcal{N}_{D,\mathbb{P}^3})\leq\dim H_d$, so H_d is smooth at $[D]$. Similarly, $h^0(\mathcal{O}_D(1))=3$ if $d=1$ and we obtain $h^0(\mathcal{N}_D)\leq 7=\dim H_1$.

Now suppose that $\varphi(z)\neq 0$ and $d\geq 4$, since the case $d=1$ is straightforward. Write $[x,y,z,w]$ for the coordinates of $\mathbb{P}^3$ and S for the coordinate ring. The exact sequence (1) shows that $h^1(\mathcal{I}_D(n))=0$ for all $n>0$; hence the map $(S/I_D)_n\to H^0(\mathcal{O}_D(n))$ is an isomorphism for all $n>0$. It follows that the comparison theorem [Piene and Schlessinger 1985] (see also [Ellingsrud 1975; Kleppe 1979]) applies to D so that $H^0(\mathcal{N}_D)\cong\mathrm{Hom}(I_D,S/I_D)_0$. Since $\varphi(f)=\lambda\varphi(zw^{d-1})$ for some $\lambda\in k$, $\varphi(f-\lambda zw^{d-1})=0$ and we may write $I_D=(xz,yz,z^2,f-\lambda zw^{d-1})$. For smoothness at $[D]$, it suffices to show this when $\lambda=0$, because the members of the family parametrized by λ are projectively equivalent for $\lambda\neq 0$. Thus we may assume $I_D=(xz,yz,z^2,f)$ with $f\in(x,y)$ and write $f=xg+yh$ for $g,h\in S_{d-1}$.

Now consider $\rho\in\mathrm{Hom}(I_D,S/I_D)_0$. Observe that a basis for $(S/I_D)_2$ consists of $\{x^2,xy,y^2,zw,w^2,wx,wy\}$ and there is a similar basis for $(S/I_D)_3$ consisting of eleven monomials because $\deg f>3$. In terms of these bases, the Koszul relations

$$z\rho(xz)=x\rho(z^2),\quad z\rho(xy)=y\rho(z^2),\quad x\rho(yz)=y\rho(xz)$$

require that

$$\rho(z^2) = a_1 wz, \quad \rho(xz) = a_2 wz + a_3 xz + a_4 x^2 + a_5 xy,$$
$$\rho(yz) = a_6 wz + a_3 zy + a_4 xy + a_5 y^2.$$

Modulo (xz, yz, z^2) we may write

$$\rho(f) = a_7 z w^{d-1} + G,$$

with $G \in k[x, y, w]_d$. Now $g\rho(zx) + h\rho(zy) = z\rho(f) = zG$ modulo I_D gives a linear relation between the coefficient of w^d in G and a_2 and a_6. Since $\rho(f)$ is only determined modulo f, there are $\binom{d+2}{2} - 2$ degrees of freedom in choosing $\rho(f)$, so that

$$\dim \operatorname{Hom}(I_D, S/I_D)_0 \le 7 + \binom{d+2}{2} - 2 = \binom{d+2}{2} + 5 = \dim H_d.$$

The second statement follows from Proposition 2.1 by varying C. Indeed, the rational map $M = \mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3) \times \mathbb{P}^3 \dashrightarrow H_d \subset \mathrm{Hilb}^{dz+2-g}(\mathbb{P}^3)$ given by $(C, p) \mapsto C \cup p$ has indeterminacy locus equal to the incidence correspondence $\Delta = \{(C, p) : p \in C\}$. For fixed $[C] \in \mathrm{Hilb}^{dz+1-g}(\mathbb{P}^3)$, the fiber is isomorphic to $\mathbb{P}^3$ and via this isomorphism the intersection with Δ is identified with $C \subset \mathbb{P}^3$. Thus when Δ is blown up, the fiber over C is identified with $\widetilde{\mathbb{P}}^3(C)$, which according to Proposition 2.1 is in bijective correspondence with $V \subset H_d$ (using the notation in Proposition 2.1). It follows that after blowing up the indeterminacy locus $\Delta \subset M$ we obtain a bijective morphism $\widetilde{M}(\Delta) \to H_d$, which is an isomorphism by Zariski's main theorem. $\square$

Remark 4.4. One can verify by similar tangent space calculations that the Hilbert scheme of plane curves with two isolated or embedded points is singular exactly along the plane curves with the double embedded points of type (a) in Proposition 3.3. It is interesting that the Hilbert scheme is smooth along curves with the double embedded points of type (b).

Example 4.5. The only locally Cohen–Macaulay curve of degree 1 is a line. By our results, any curve obtained from a line L by adding ≤ 3 embedded points is a flat limit of L union the right number of isolated points. It follows that $\mathrm{Hilb}^{z+1-g}(\mathbb{P}^3)$ is irreducible of dimension $4 - 3g$ for $-3 \le g \le 0$. On the other hand, it is reducible for $g \ll 0$ because the Hilbert scheme of sufficiently many points in $\mathbb{P}^3$ is not irreducible [Iarrobino 1972].

Example 4.6. For one-dimensional subschemes of degree 2 and high genus the irreducible components of $\mathrm{Hilb}^{2z+1-g}(\mathbb{P}^3)$ are as follows:

(a) If $g = 0$, the Hilbert scheme is irreducible, consisting of plane curves.

(b) If $g = -1$, there are two irreducible components. The first component H_1 has general member a pair of skew lines and has dimension eight. The second component H_2 has general member a plane conic union an isolated point and has dimension 11. There are also plane curves with embedded points, but these lie in H_2 by Proposition 2.2. Both components H_1 and H_2 are smooth [Chen et al. 2011].

(c) Similarly if $g = -2$, there are three irreducible components. There is the family H_1 of double lines of genus $g = -2$ with no embedded points of dimension nine, the family H_2 of two skew lines union an isolated point of dimension 11, and the family H_3 of conics union two isolated points of dimension 14. Because all the underlying locally Cohen–Macaulay curves in question are local complete intersections, we know from Proposition 2.2 and Proposition 3.3 that we have not missed any possibilities.

(d) For $g = -3$ we can write down four irreducible components following the same pattern as above and our results show that we have not missed any irreducible components. However when $g = -4$ we cannot be sure that there is not an irreducible component whose general member consists of a plane curve with some horrible quadruple point.

Example 4.7. For one-dimensional subschemes of degree 3 and high genus, we can make similar lists of the irreducible components of $\mathrm{Hilb}^{3z+1-g}(\mathbb{P}^3)$:

(a) If $g = 1$, the Hilbert scheme is irreducible and consists of plane curves.

(b) If $g = 0$, the Hilbert scheme has two irreducible components. The family H_1 has general member a twisted cubic and has dimension 12. The family H_2 has general member a plane cubic union an isolated point and has dimension 15. This example has been well-studied in [Piene and Schlessinger 1985].

(c) If $g = -1$, there are three irreducible components. The component H_1 whose general member is a line and a disjoint conic has dimension 12. The component H_2 whose general member is a twisted cubic union an isolated point has dimension 15. The component H_3 whose general member is a plane cubic union two isolated points has dimension 18. To see that these are all, we need to show that degenerations of a twisted cubic curve union an embedded point cannot form an irreducible component of their own, something which is not clear in view of Example 2.6(a). However all ACM curves of degree 3 and genus 0 have resolution

$$0 \to \mathcal{O}(-3)^2 \to \mathcal{O}(-2)^3 \to \mathcal{I}_C \to 0$$

[Ellingsrud 1975, Example 1] and we can apply Proposition 2.7.

Example 4.8. Consider the Hilbert schemes $\mathrm{Hilb}^{4z+1-g}(\mathbb{P}^3)$:

(a) If $g = 3$ or 2, the Hilbert scheme is irreducible by Theorem 4.2.

(b) If $g = 1$, the Hilbert scheme has two irreducible components. One component H_1 has general member a plane quartic union two isolated points and has dimension 23. Any subschemes not parametrized by H_1 have no isolated or embedded points (any nonplanar locally Cohen–Macaulay curve satisfies the genus bound $g \leq (d-2)(d-3)/2$ [Martin-Deschamps and Perrin 1993]), so we are looking at the Hilbert scheme $H_{4,1}$ of locally Cohen–Macaulay curves, which we described in Example 2.8.

This brings us to the last example, which might be known to experts, though we have not seen a rigorous proof in the literature.

Theorem 4.9. *The Hilbert scheme* $\mathrm{Hilb}^{4z+1}(\mathbb{P}^3)$ *has four irreducible components*:

H_1: *The closure of the family of rational quartic curves has dimension* 16.

H_2: *The family whose general member is a disjoint union of a plane cubic and a line has dimension* 16.

H_3: *The family whose general member is a disjoint union of an elliptic quartic curve and a point has dimension* 19.

H_4: *The family whose general member is a disjoint union of a plane quartic curve and three distinct points has dimension* 26.

Proof. The dimension counts are standard, so we only need to show that every subscheme parametrized by $\mathrm{Hilb}^{4z+1}(\mathbb{P}^3)$ is contained in one of these families and no family is contained in another. The second part is easy: the family H_4 has the largest dimension, but none of the others lie in its closure due to the three isolated or embedded points. Similarly H_3 has larger dimension than H_1 and H_2, but H_1 and H_2 are not in its closure due to the isolated or embedded point. Since families H_1 and H_2 have the same dimension, neither lies in the closure of the other.

To complete the proof, we show that each $[C] \in \mathrm{Hilb}^{4z+1}(\mathbb{P}^3)$ lies in one of the families H_i listed above. Fixing such $C \subset \mathbb{P}^3$, let $C_0 \subset C$ be the purely one-dimensional part. There is no such curve of genus $g = 2$ [Hartshorne 1994], leaving three cases. If $g(C_0) = 0$, then $C = C_0$ is locally Cohen–Macaulay, and it is known that the Hilbert scheme $H_{4,0}$ of locally Cohen–Macaulay curves has two irreducible components, described in H_1 and H_2 above [Nollet and Schlesinger 2003]. If $g(C_0) = 3$, then C_0 is a plane quartic and hence a complete intersection. It follows from Propositions 2.2, 3.3, and 3.7 that C is a flat limit of subschemes which are plane quartics union three isolated points, so $[C] \in H_4$. If $g(C_0) = 1$, then $[C] \in H_3$ by Example 2.8. □

It would be interesting to describe more precisely the intersection of the components H_1 through H_4 in $\mathrm{Hilb}^{4z+1}(\mathbb{P}^3)$ as done in [Piene and Schlessinger 1985] for $\mathrm{Hilb}^{3z+1}(\mathbb{P}^3)$, though this will require a classification of all curves of degree 4 and genus 0 up to projective equivalence. It would also be interesting to determine the birational geometry of the component H_1, as done in [Chen 2008] for $\mathrm{Hilb}^{3z+1}(P^3)$.

Acknowledgements

We thank John Brevik, Izzet Coskun, Daniel Erman, Robin Hartshorne, and Michael Stillman for useful conversations. Part of the work was modified during the workshop "Components of Hilbert Schemes" at the American Institute of Mathematics, July 2010. We thank the AIM and the organizers for their invitation and hospitality. We also would like to thank the referee for carefully reading the paper and for many helpful comments.

References

[Cartwright et al. 2009] D. A. Cartwright, D. Erman, M. Velasco, and B. Viray, "Hilbert schemes of 8 points", *Algebra Number Theory* **3**:7 (2009), 763–795. MR 2011c:14010 Zbl 1187.14005

[Chang 1989] M.-C. Chang, "A filtered Bertini-type theorem", *J. Reine Angew. Math.* **397** (1989), 214–219. MR 90i:14054 Zbl 0663.14008

[Chen 2008] D. Chen, "Mori's program for the Kontsevich moduli space $\overline{\mathcal{M}}_{0,0}(\mathbb{P}^3, 3)$", *Int. Math. Res. Not.* **2008** (2008), Art. ID rnn 067. MR 2009f:14019 Zbl 1147.14009

[Chen et al. 2011] D. Chen, I. Coskun, and S. Nollet, "Hilbert scheme of a pair of codimension two linear subspaces", *Comm. Algebra* **39**:8 (2011), 3021–3043. MR 2834144 Zbl 05973402

[Ellingsrud 1975] G. Ellingsrud, "Sur le schéma de Hilbert des variétés de codimension 2 dans $\mathbf{P}^e$ à cône de Cohen–Macaulay", *Ann. Sci. École Norm. Sup.* (4) **8**:4 (1975), 423–431. MR 52 #13831 Zbl 0325.14002

[Fogarty 1968] J. Fogarty, "Algebraic families on an algebraic surface", *Amer. J. Math* **90** (1968), 511–521. MR 38 #5778 Zbl 0176.18401

[Grothendieck 1961] A. Grothendieck, "Techniques de construction et théorèmes d'existence en géométrie algébrique, IV: Les schémas de Hilbert", exposé 221, pp. 249–276 in *Séminaire Bourbaki* 1960–1961, W. A. Benjamin, Amsterdam, 1961. Reprinted in *Séminaire Bourbaki* **6**, Soc. Math. France, Paris, 1995. Zbl 0236.14003

[Harris and Morrison 1998] J. Harris and I. Morrison, *Moduli of curves*, Graduate texts in mathematics **187**, Springer, New York, 1998. MR 99g:14031 Zbl 0913.14005

[Hartshorne 1966] R. Hartshorne, "Connectedness of the Hilbert scheme", *Inst. Hautes Études Sci. Publ. Math.* **29** (1966), 5–48. MR 35 #4232 Zbl 0171.41502

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate texts in mathematics **52**, Springer, New York, 1977. MR 57 #3116 Zbl 0367.14001

[Hartshorne 1994] R. Hartshorne, "The genus of space curves", *Ann. Univ. Ferrara Sez. VII* (*N.S.*) **40**:1 (1994), 207–223. MR 97d:14044 Zbl 0865.14013

[Hartshorne 2000] R. Hartshorne, "On the connectedness of the Hilbert scheme of curves in $\mathbb{P}^3$", *Comm. Algebra* **28**:12 (2000), 6059–6077. MR 2002d:14003 Zbl 0994.14002

[Iarrobino 1972] A. Iarrobino, "Reducibility of the families of 0-dimensional schemes on a variety", *Invent. Math.* **15** (1972), 72–77. MR 46 #170 Zbl 0227.14006

[Iarrobino and Emsalem 1978] A. Iarrobino and J. Emsalem, "Some zero-dimensional generic singularities; finite algebras having small tangent space", *Compositio Math.* **36**:2 (1978), 145–188. MR 81c:14004 Zbl 0393.14002

[Kleppe 1979] J. O. Kleppe, "Deformations of graded algebras", *Math. Scand.* **45**:2 (1979), 205–231. MR 82j:13017 Zbl 0436.14004

[Martin-Deschamps and Perrin 1990] M. Martin-Deschamps and D. Perrin, *Sur la classification des courbes gauches*, Astérisque **184-185**, 1990. MR 91h:14039 Zbl 0717.14017

[Martin-Deschamps and Perrin 1993] M. Martin-Deschamps and D. Perrin, "Sur les bornes du module de Rao", *C. R. Acad. Sci. Paris Sér. I Math.* **317**:12 (1993), 1159–1162. MR 95e:14023 Zbl 0796.14029

[Mazzola 1980] G. Mazzola, "Generic finite schemes and Hochschild cocycles", *Comment. Math. Helv.* **55**:2 (1980), 267–293. MR 82k:14010 Zbl 0463.14004

[Migliore 1998] J. C. Migliore, *Introduction to liaison theory and deficiency modules*, Progress in mathematics **165**, Birkhäuser, Boston, MA, 1998. MR 2000g:14058 Zbl 0921.14033

[Nollet 1997] S. Nollet, "The Hilbert schemes of degree three curves", *Ann. Sci. École Norm. Sup.* (4) **30**:3 (1997), 367–384. MR 99a:14003 Zbl 0892.14004

[Nollet 2006] S. Nollet, "Deformations of space curves: connectedness of Hilbert schemes", *Rend. Semin. Mat. Univ. Politec. Torino* **64**:4 (2006), 433–450. MR 2008d:14007 Zbl 1135.14023

[Nollet and Schlesinger 2003] S. Nollet and E. Schlesinger, "Hilbert schemes of degree four curves", *Compositio Math.* **139**:2 (2003), 169–196. MR 2005b:14010 Zbl 1053.14035

[Peskine and Szpiro 1974] C. Peskine and L. Szpiro, "Liaison des variétés algébriques, I", *Invent. Math.* **26** (1974), 271–302. MR 51 #526 Zbl 0298.14022

[Piene and Schlessinger 1985] R. Piene and M. Schlessinger, "On the Hilbert scheme compactification of the space of twisted cubics", *Amer. J. Math.* **107**:4 (1985), 761–774. MR 86m:14004 Zbl 0589.14009

Communicated by David Eisenbud
Received 2010-12-17 Revised 2011-05-10 Accepted 2011-06-30

dawei.chen@bc.edu — *Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, 851 S. Morgan St., Chicago, IL 60607, United States*

Current address: — *Department of Mathematics, Boston College, Chestnut Hill, MA 02467, United States*

s.nollet@tcu.edu — *Department of Mathematics, Texas Christian University, Box 298900, Fort Worth, TX 76129, United States*

ALGEBRA AND NUMBER THEORY 6:4(2012)

msp

Moduli of Galois p-covers in mixed characteristics

Dan Abramovich and Matthieu Romagny

We define a proper moduli stack classifying covers of curves of prime degree p. The objects of this stack are torsors $Y \to \mathscr{X}$ under a finite flat $\mathscr{X}$-group scheme, with $\mathscr{X}$ a twisted curve and Y a stable curve. We also discuss embeddings of finite flat group schemes of order p into affine smooth 1-dimensional group schemes.

1. Introduction

Fix a prime number p. The study of families of Galois p-cyclic covers of curves is well understood in characteristic 0, where there is a nice smooth proper stack classifying (generically étale) covers of stable curves, with a dense open substack composed of covers of smooth curves. The reduction of this stack at a prime $\ell \neq p$ is also well understood, but the question of the reduction at p is notably much harder. For the classical modular curves, namely the unramified genus-1 case, there has been in the last years renewed intense research on this topic; see, for example, [Edixhoven 1990; Bouw and Wewers 2004; McMurdy and Coleman 2010].

The aim of the present paper is to consider the case of arbitrary genus. More precisely, we define a complete moduli stack of degree-p covers $Y \to \mathscr{X}$, with Y a stable curve which is a $\mathscr{G}$-torsor over $\mathscr{X}$, for a suitable group scheme $\mathscr{G}/\mathscr{X}$. The curve $\mathscr{X}$ is a twisted curve in the sense of [Abramovich and Vistoli 2002;

Abramovich was supported in part by NSF grants DMS-0603284 and DMS-0901278.
MSC2010: primary 14H10; secondary 14L15.
Keywords: curves, group schemes of order p, moduli.

Abramovich et al. 2011] but in general not stable. This follows the same general approach as the characteristic-0 paper [Abramovich et al. 2003], but diverges from that of [Abramovich et al. 2011], where the curve $\mathscr{X}$ is stable, the group scheme $\mathscr{G}$ is assumed linearly reductive, but Y is in general much more singular. Here the approach is based on [Raynaud 1999, Proposition 1.2.1] of Raynaud, and the more general notion of *effective model* of a group-scheme action from [Romagny 2011]. The general strategy was outlined in [Abramovich 2012] in a somewhat special case.

The ideal goal is a moduli space where, on the one hand, the object parametrized are concrete and with minimal singularities — ideally nodes, and on the other hand the singularities of the moduli space are well understood. This would allow one to easily describe objects in characteristic p and to identify their liftings in characteristic 0. In this paper we have not given a description of the singularities of the moduli space, so we fall short of this goal.

1.1. ***Rigidified group schemes.*** The group scheme $\mathscr{G}$ in our covers comes with a supplementary structure which we call a *generator*. Before we define this notion, let us briefly recall from [Katz and Mazur 1985, §1.8] the concept of a *full set of sections*. Let $Z \to S$ be a finite locally free morphism of schemes of degree N. Then for all affine S-schemes $\mathrm{Spec}(R)$, the R-algebra $\Gamma(Z_R, \mathscr{O}_{Z_R})$ is locally free of rank N and has a canonical norm mapping. We say that a set of N sections $x_1, \dots, x_N \in Z(S)$ is a *full set of sections* if and only if for any affine S-scheme $\mathrm{Spec}(R)$ and any $f \in \Gamma(Z_R, \mathscr{O}_{Z_R})$, the norm of f is equal to the product $f(x_1) \dots f(x_N)$.

Definition 1.2. Let $G \to S$ be a finite locally free group scheme of order p. A *generator* is a morphism of S-group schemes $\gamma : (\mathbb{Z}/p\mathbb{Z})_S \to G$ such that the sections $x_i = \gamma(i)$, $0 \le i \le p-1$, are a full set of sections. A *rigidified group scheme* is a group scheme of order p with a generator.

The notion of generator is easily described in terms of the Tate–Oort classification of group schemes of order p. This is explained and complemented in Appendix A.

Remark 1.3. One can define the stack of rigidified group schemes a bit more directly: consider the Artin stack $\mathfrak{G}\mathfrak{S}_p$ of group schemes of order p, and let $\mathscr{G}^u \to \mathfrak{G}\mathfrak{S}_p$ be the universal group-scheme - an object of $\mathscr{G}^u$ over a scheme S consists of a group-scheme $\mathscr{G} \to S$ with a section $S \to \mathscr{G}$. It has a unique nonzero point over $\mathbb{Q}$ corresponding to $\mathbb{Z}/p\mathbb{Z}$ with the section 1. The stack of rigidified group schemes is canonically isomorphic to the closure of this point.

Of course describing a stack as a closure of a substack is not ideal from the moduli point of view, and we find the definition using Katz–Mazur generators more satisfying.

1.4. ***Stable p-torsors.*** Fix a prime number p and integers $g, h, n \geq 0$ with $2g - 2 + n > 0$.

Definition 1.5. A *stable n-marked p-torsor* of genus g (over some base scheme S) is a triple

$$(\mathscr{X}, \mathscr{G}, Y),$$

where

(1) $(\mathscr{X}, \{\Sigma_i\}_{i=1}^n)$ is an n-marked twisted curve of genus h,

(2) $(Y, \{P_i\}_{i=1}^n)$ is a nodal curve of genus g with étale marking divisors $P_i \to S$, which is stable in the sense of Deligne, Mumford, and Knudsen,

(3) $\mathscr{G} \to \mathscr{X}$ is a rigidified group scheme of order p,

(4) $Y \to \mathscr{X}$ is a $\mathscr{G}$-torsor and $P_i = \Sigma_i \times_{\mathscr{X}} Y$ for all i.

Note that as usual the markings Σ_i (resp. P_i) are required to lie in the smooth locus of $\mathscr{X}$ (resp. Y). They split into two groups. In the first group Σ_i is twisted and $[P_i : S] = 1$, while in the second group Σ_i is a section and $[P_i : S] = p$. The number m of twisted markings is determined by $(2g-2) = p(2h-2) + m(p-1)$ and it is equivalent to fix h or m.

The notion of stable marked p-torsor makes sense over an arbitrary base scheme S. Given stable n-marked p-torsors $(\mathscr{X}, \mathscr{G}, Y)$ over S and $(\mathscr{X}', \mathscr{G}', Y')$ over S', one defines as usual a morphism $(\mathscr{X}, \mathscr{G}, Y) \to (\mathscr{X}', \mathscr{G}', Y')$ over $S \to S'$ as a fiber diagram. This defines a category fibered over $\operatorname{Spec} \mathbb{Z}$ that we denote $ST_{p,g,h,n}$.

Our main result is:

Theorem 1.6. *The category $ST_{p,g,h,n}/\operatorname{Spec}\mathbb{Z}$ is a proper Deligne–Mumford stack with finite diagonal.*

Notice that $ST_{p,g,h,n}$ contains an open substack of étale $\mathbb{Z}/p\mathbb{Z}$-covers. Identifying the closure of this open locus remains an interesting question.

1.7. ***Organization.*** Section 2 is devoted to Proposition 2.1, in particular showing the algebraicity of $ST_{p,g,h,n}$. Section 3 completes the proof of Theorem 1.6 by showing properness. We give simple examples in Section 4. In Appendix A we discuss embeddings of group schemes of order p into smooth group schemes. In Appendix B we recall some facts about the Weil restriction of closed subschemes, and state the representability result in a form useful for us.

2. The stack $ST_{p,g,h,n}$

In this section, we review some basic facts on twisted curves and then we show:

Proposition 2.1. *The category $ST_{p,g,h,n}/\operatorname{Spec}\mathbb{Z}$ is an algebraic stack of finite type over $\mathbb{Z}$.*

2.2. ***Twisted curves and log twisted curves.*** We review some material from Olsson's treatment in [Abramovich et al. 2011, Appendix A], with some attention to properness of the procedure of "log twisting".

Recall that a *twisted curve* over a scheme S is a tame Artin stack $\mathscr{C} \to S$ (we refer to [Abramovich et al. 2008, Definition 3.1] for this notion) with a collection of gerbes $\Sigma_i \subset \mathscr{C}$ satisfying the following conditions:

(1) The coarse moduli space C of $\mathscr{C}$ is a prestable curve over S, and the images $\bar{\Sigma}_i$ of Σ_i in C are the images of disjoint sections $\sigma_i : S \to C$ of $C \to S$ landing in the smooth locus.

(2) Étale locally on S there are positive integers r_i such that, on a neighborhood of Σ_i we can identify $\mathscr{C}$ with the root stack $C(\sqrt[r_i]{\bar{\Sigma}_i})$.

(3) Near a node z of C write $C^{\mathrm{sh}} = \operatorname{Spec}(\mathbb{O}_S^{\mathrm{sh}}[x, y]/(xy - t))^{\mathrm{sh}}$. Then there exists a positive integer a_z and an element $s \in \mathbb{O}_S^{\mathrm{sh}}$ such that $s^{a_z} = t$ and
$$\mathscr{C}^{\mathrm{sh}} = [\operatorname{Spec} \mathbb{O}_S^{\mathrm{sh}}[u, v]/(uv - s))^{\mathrm{sh}}/\mu_{a_z}],$$
where μ_{a_z} acts via $(u, v) \mapsto (\zeta u, \zeta^{-1} v)$ and where $x = u^{a_z}$ and $y = v^{a_z}$.

The *index* of a geometric point z on a twisted curve is a measure of its automorphisms: it is the integer r_i for a twisted marking or the integer a_z for a twisted node.

The purpose of [Abramovich et al. 2011, Appendix A] was to show that twisted curves form an Artin stack which is locally of finite type over $\mathbb{Z}$. There are two steps involved.

The introduction of the stack structure over the markings is a straightforward step: the stack $\mathfrak{M}^{tw}_{g,n}$ of twisted curves with genus G and n markings is the infinite disjoint union $\mathfrak{M}^{tw}_{g,n} = \sqcup \mathfrak{M}^{r}_{g,n}$, where r runs over the possible *marking indices*, namely vectors of positive integers $r = (r_1, \dots, r_n)$, and the stacks $\mathfrak{M}^{r}_{g,n}$ are all isomorphic to each other - the universal family over $\mathfrak{M}^{r}_{g,n}$ is obtained form that over $\mathfrak{M}^{(1,\dots,1)}_{g,n}$ by taking the r_i-th root of $\bar{\Sigma}_i$.

The more subtle point is the introduction of twisting at nodes. Olsson achieves this using the canonical log structure of prestable curves, and provides an equivalence between twisted curves with $r = (1, \dots, 1)$ and log-twisted curves. A *log twisted curve* over a scheme S is the data of a prestable curve C/S along with a simple extension $\mathcal{M}^S_{C/S} \hookrightarrow \mathcal{N}$, see [Abramovich et al. 2011, Definition A.3]. Here $\mathcal{M}^S_{C/S}$ is F. Kato's canonical locally free log structure of the base S of the family of prestable curves C/S, and a *simple extension* is an injective morphism $\mathcal{M}^S_{C/S} \hookrightarrow \mathcal{N}$ of locally free log structures of equal rank where an irreducible element is sent to a multiple of an irreducible element up to units. See [Abramovich et al. 2011, Definition A.1].

We now describe an aspect of this equivalence which is relevant for our main results. Consider a family of prestable curves C/S and denote by $\iota : \operatorname{Sing} C/S \to C$

the embedding of the locus where $\pi : C \to S$ fails to be smooth. A *node function* is a section a of $\pi_* \iota_* \mathbb{N}_{\mathrm{Sing}\, C/S}$. In other words it gives a positive integer a_z for each singular point z of C/S in a continuous manner. Given a morphism $T \to S$, we say that a twisted curve $\mathscr{C}/T$ with coarse moduli space C_T is *a-twisted* over C/S if the index of a node of $\mathscr{C}$ over a node z of C is precisely a_z.

Proposition 2.3. *Fix a family of prestable curves C/S of genus g with n markings over a noetherian scheme S. Further fix marking indices $r = (r_1, \dots, r_n)$ and a node function a. Then the category of a-twisted curves over C/S with marking indices given by r is a proper and quasifinite tame stack over S.*

Proof. The problem is local on S, and further it is stable under base change in S. So it is enough to prove this when S is a versal deformation space of a prestable curve C_s of genus g with n markings, over a closed geometric point $s \in S$, in such a way that we have a chart $\mathbb{N}^k \to \mathcal{M}^S_{C/S}$ of the log structure, where k is the number of nodes of C_s. The image of the i-th generator of $\mathbb{N}^k$ in $\mathbb{O}_S$ is the defining equation of the smooth divisor D_i where the i-th node persists. Now consider an a-twisted curve over $\phi : T \to S$, corresponding to a simple extension $\phi^* \mathcal{M}^S_{C/S} \to \mathcal{N}$ where the image of the i-th generator m_i becomes an a_i-multiple up to units. This precisely means that $\mathbb{O}^*_{C_T} m_i$, the principal bundle associated to $\mathbb{O}_S(-D_i)$, is an a_i-th power. In other words, the stack of a-twisted curves over C/S is isomorphic to the stack

$$S(\sqrt[a_1]{D_1} \cdots \sqrt[a_n]{D_n}) \quad = \quad S(\sqrt[a_1]{D_1}) \times_S \cdots \times_S S(\sqrt[a_n]{D_n})$$

encoding a_i-th roots of $O_S(D_i)$. This is evidently a proper and quasifinite tame stack over S. $\square$

We now turn to the indices of twisted points in a stable p-torsor.

Lemma 2.4. *Let $(\mathscr{X}, \mathscr{G}, Y)$ be a stable p-torsor. Then the index of a point $x \in \mathscr{X}$ divides p.*

Proof. Let r be the index of x and d the local degree of $Y \to \mathscr{X}$ at a point y above x. Since $Y \to \mathscr{X}$ is finite flat of degree p and $\mathscr{G}$ acts transitively on the fibers, then $d \mid p$. Let $f : \mathscr{X} \to X$ be the coarse moduli space of $\mathscr{X}$. In order to compute d, we pass to strict henselizations on S, X and Y at the relevant points. Thus S is the spectrum of a strictly henselian local ring $(R, \mathfrak{m})$, and we have two cases to consider.

If x is a smooth point,

- $X \simeq \operatorname{Spec} R[a]^{\mathrm{sh}}$,
- $Y \simeq \operatorname{Spec} R[s]^{\mathrm{sh}}$,
- $\mathscr{X} \simeq [D/\mu_r]$ with $D = \operatorname{Spec} R[u]^{\mathrm{sh}}$ and $\zeta \in \mu_r$ acting by $u \mapsto \zeta u$.

Consider the fibered product $E = Y \times_{\mathscr{X}} D$. The map $E \to Y$ is a μ_r-torsor of the form $E \simeq \operatorname{Spec} \mathbb{O}_Y[w]/(w^r - f)$ for some invertible function $f \in \mathbb{O}_Y^\times$, and $E \to D$ is a μ_r-equivariant map given by $u \mapsto \varphi w$ for some function φ on Y. Let $\tilde{x} : \operatorname{Spec} k \to D$ be a point mapping to x in $\mathscr{X}$, i.e., corresponding to $u = \mathfrak{m} = 0$, and let $\bar{\varphi}$, $\bar{f}$ be the restrictions of φ, f to $Y_{\tilde{x}}$. The preimage of $\tilde{x}$ under $E \to D$ is a finite k-scheme with algebra $k[s][w]/(\bar{\varphi}, w^r - \bar{f})$. We see that $d = r \dim_k k[s]/(\bar{\varphi})$ and hence the index r divides p.

If x is a singular point, there exist λ, μ, ν in $\mathfrak{m}$ such that

- $X \simeq \operatorname{Spec}(R[a, b]/(ab - \lambda))^{\mathrm{sh}}$,
- $Y \simeq \operatorname{Spec}(R[s, t]/(st - \mu))^{\mathrm{sh}}$,
- $\mathscr{X} \simeq [D/\mu_r]$, where $D = \operatorname{Spec}(R[u, v]/(uv - \nu))^{\mathrm{sh}}$,

and $\zeta \in \mu_r$ acts by $u \mapsto \zeta u$ and $v \mapsto \zeta^{-1} v$. The scheme $E = Y \times_{\mathscr{X}} D$ is of the form $E \simeq \operatorname{Spec} \mathbb{O}_Y[w]/(w^r - f)$ for some invertible function $f \in \mathbb{O}_Y^\times$, and the map $E \to D$ is given by $u \mapsto \varphi w$, $v \mapsto \psi w^{-1}$ for some functions φ, ψ on Y satisfying $\varphi\psi = \nu$. Let $\tilde{x} : \operatorname{Spec} k \to D$ be a point mapping to x and let $\bar{\varphi}$, $\bar{\psi}$, $\bar{f}$ be the restrictions of φ, ψ, f to $Y_{\tilde{x}}$. The preimage of $\tilde{x}$ under $E \to D$ is a finite k-scheme with algebra $k[s, t][w]/(st, \bar{\varphi}, \bar{\psi}, w^r - \bar{f})$. We see that $d = r \dim_k k[s, t]/(st, \bar{\varphi}, \bar{\psi})$ and hence r divides p. □

Proof of Proposition 2.1. Let $\delta = (\delta_1, \dots, \delta_n)$ be the sequence of degrees of the markings P_i on the total space of stable p-torsors, with each δ_i equal to 1 or p. We build $ST_{p,g,h,n}$ from existing stacks: the stack $\overline{\mathcal{M}}_{g,\delta}$ of Deligne–Mumford–Knudsen stable marked curves (for the family of curves Y), the stack $\mathfrak{M}$ of twisted curves (for the family of marked twisted curves $\mathscr{X}$), and Hilbert schemes and Hom-stacks for construction of $Y \to \mathscr{X}$ and $\mathscr{G}$.

Bounding the twisted curves. We have an obvious forgetful functor $ST_{p,g,h,n} \to \overline{\mathcal{M}}_{g,\delta} \times \mathfrak{M}$. Note that the image of $ST_{p,g,h,n} \to \mathfrak{M}$ lies in an open substack $\mathfrak{M}'$ of finite type over $\mathbb{Z}$: the index of the twisted curve $\mathscr{X}$ divides p by Lemma 2.4, and its topological type is bounded by that of Y. The stack $\mathfrak{M}'$ parametrizing such twisted curves is of finite type over $\mathbb{Z}$ by [Abramovich et al. 2011, Corollary A.8].

Set $M_{Y,\mathscr{X}} = \overline{\mathcal{M}}_{g,\delta} \times \mathfrak{M}'$. This is an algebraic stack of finite type over $\mathbb{Z}$.

The map $Y \to \mathscr{X}$. Consider the universal family $Y \to M_{Y,\mathscr{X}}$ of stable curves of genus g and the universal family $\mathscr{X} \to M_{Y,\mathscr{X}}$ of twisted curves, with associated family of coarse curves $X \to M_{Y,\mathscr{X}}$. Since Hilbert schemes of fixed Hilbert polynomial are of finite type, there is an algebraic stack $\operatorname{Hom}^{\leq p}_{M_{Y,\mathscr{X}}}(Y, X)$, of finite type over $M_{Y,\mathscr{X}}$, parametrizing morphisms $Y_s \to X_s$ of degree $\leq p$ between the respective fibers. By [Abramovich et al. 2011, Corollary C.4] the stack $\operatorname{Hom}^{\leq p}_{M_{Y,\mathscr{X}}}(Y, \mathscr{X})$ corresponding to maps $Y_s \to \mathscr{X}_s$ with target the twisted curve is of finite type over $\operatorname{Hom}^{\leq p}_{M_{Y,\mathscr{X}}}(Y, X)$, hence over $M_{Y,\mathscr{X}}$. There is an open substack $M_{Y\to\mathscr{X}}$ parametrizing flat morphisms of degree precisely p. We have an obvious forgetful functor $ST_{p,g,h,n} \to M_{Y\to\mathscr{X}}$ lifting the functor $ST_{p,g,h,n} \to \bar{\mathcal{M}}_{g,\delta} \times \mathfrak{M}'$ above.

The rigidified group scheme $\mathscr{G}$. The scheme $Y_2 = Y \times_{\mathscr{X}} Y$ is flat of degree p over Y. Giving it the structure of a group scheme over Y with unit section equal to the diagonal $Y \to Y_2$ is tantamount to choosing structure Y-arrows $m : Y_2 \times_Y Y_2 \to Y_2$ and $i' : Y_2 \to Y_2$, which are parametrized by a Hom-scheme, and passing to the closed subscheme where these give a group-scheme structure (that this condition is closed follows from representability of the Weil restriction; see the discussion in Appendix B and in particular Corollary B.4). Giving a group scheme $\mathscr{G}$ over $\mathscr{X}$ with isomorphism $\mathscr{G} \times_{\mathscr{X}} Y \simeq Y_2$ is tantamount to giving descent data for Y_2 with its chosen group-scheme structure. This is again parametrized by a suitable Hom-scheme. Finally requiring that the projection $Y_2 \to Y$ correspond to an action of $\mathscr{G}$ on Y is a closed condition (again by Weil restriction, see Corollary B.4).

Passing to a suitable Hom-stack we can add a homomorphism $\mathbb{Z}/p\mathbb{Z} \to \mathscr{G}$, giving a section $\mathscr{X} \to \mathscr{G}$ (equivalently a morphism $\mathscr{X} \to \mathscr{G}^u$, see Remark 1.3). By [Katz and Mazur 1985, corollary 1.3.5], the locus of the base where this section is a generator is closed. Since $Y_2 \to Y$ and $Y \to \mathscr{X}$ are finite, all the necessary Hom-stacks are in fact of finite type.

The resulting stack is clearly isomorphic to $ST_{p,g,h,n}$. □

3. Properness

Since $ST_{p,g,h,n} \to \operatorname{Spec} \mathbb{Z}$ is of finite type, we need to prove the valuative criterion for properness.

We have the following situation:

(1) R is a discrete valuation ring with spectrum $S = \operatorname{Spec} R$, fraction field K with corresponding generic point $\eta = \operatorname{Spec} K$, and residue field κ with corresponding special point $s = \operatorname{Spec} \kappa$.

(2) $(\mathscr{X}_\eta, \mathscr{G}_\eta, Y_\eta)$ a stable marked p-torsor of genus g over η.

By an *extension* of $(\mathscr{X}_\eta, \mathscr{G}_\eta, Y_\eta)$ across s we mean

(1) a local extension $R \to R'$ with K'/K finite,

(2) a stable marked p-torsor $(\mathscr{X}', \mathscr{G}', Y')$ of genus g over $S' = \operatorname{Spec} R'$, and

(3) an isomorphism $(\mathscr{X}', \mathscr{G}', Y')'_\eta \simeq (\mathscr{X}_\eta, \mathscr{G}_\eta, Y_\eta) \times_\eta \eta'$.

Proposition 3.1. *An extension exists. When extension over S' exists, it is unique up to a unique isomorphism.*

Proof. We proceed in three steps.

Extension of Y_η. Since $\overline{\mathcal{M}}_{g,\delta}$ is proper, there is a stable marked curve Y' extending Y_η over some S', and this extension is unique up to a unique isomorphism. We replace S by S', and assume that there is Y over S with generic fiber Y_η.

Coarse extension of $\mathcal{X}_\eta$. By uniqueness, the action of $G = \mathbb{Z}/p\mathbb{Z}$ on Y_η induced by the map $G_{\mathcal{X}_\eta} \to \mathcal{G}_\eta$ extends to Y. There is a finite extension K'/K such that the intersection points of the orbits of geometric irreducible components of Y_η under the action of G are all K'-rational. We may and do replace S by the spectrum of the integral closure of R in K'. Let us call $Y_1, \dots, Y_m$ the orbits of irreducible components of Y and $\{y_{i,j}\}_{1\le i,j\le m}$ their intersections, which is a set of disjoint sections of Y. For each $i = 1, \dots, m$ we define a morphism $\pi_i : Y_i \to X_i$ as follows. If the action of G on Y_i is nontrivial we put $X_i := Y_i/G$ and π_i equal to the quotient morphism. If the action of G on Y_i is trivial, note that we must have $\mathrm{char}(K) = p$, since the map from Y_i to its image in $\mathcal{X}$ is a $\mathcal{G}$-torsor while $G_{\mathcal{X}} \to \mathcal{G}$ is an isomorphism in characteristic 0. Then we consider the Frobenius twist $X_i := Y_i^{(p)}$ and we define $\pi_i : Y_i \to X_i$ to be the relative Frobenius. Finally we let X be the scheme obtained by gluing the X_i along the sections $x_{i,j} = \pi_i(y_{i,j}) \in X_i$ and $x_{j,i} = \pi_j(y_{i,j}) \in X_j$. There are markings $\Sigma_i^X \subset X$ given by the closures in X of the generic markings $\Sigma_i^{X_\eta}$. It is clear that the morphisms π_i glue to a morphism $\pi : Y \to X$.

Extension of $\mathcal{X}_\eta$ and $Y_\eta \to \mathcal{X}_\eta$ along generic nodes and markings. In the following two lemmas we extend the stack structure of $\mathcal{X}_\eta$, and then the map $Y_\eta \to \mathcal{X}_\eta$, along the generic nodes and the markings:

Lemma 3.2. *There is a unique extension $\bar{\mathcal{X}}$ of the twisted curve $\mathcal{X}_\eta$ over X, such that $\bar{\mathcal{X}} \to X$ is an isomorphism away from the generic nodes and the markings.*

Proof. We follow [Abramovich et al. 2011, proof of Proposition 4.3]. First, let $\Sigma_{i,\eta}^{\mathcal{X}_\eta}$ be a marking on $\mathcal{X}_\eta$. There is an extension $\Sigma_i^X \subset X$. Let r be the index of $\mathcal{X}_\eta$ at $\Sigma_{i,\eta}^{\mathcal{X}_\eta}$. Then we define $\mathcal{X}$ to be the stack of r-th roots of Σ_i^X on X. This extension is unique by the separatedness of stacks of r-th roots.

Now let $x_\eta \in X_\eta$ be a node with index r and let $x \in X_s$ be its reduction. Locally in the étale topology, around x the curve X looks like the spectrum of $R[u, v]/(uv)$. Let B_u resp. B_v be the branches at x in X. The stacks of r-th roots of the divisor $u = 0$ in B_u and of the divisor $v = 0$ in B_v are isomorphic and glue to give a stack $\bar{\mathcal{X}}$. By definition of r we have $\bar{\mathcal{X}}_\eta \simeq \mathcal{X}_\eta$. This extension is unique by the separatedness of stacks of r-th roots, so the construction of $\bar{\mathcal{X}}$ descends to X. □

Lemma 3.3. *There is a unique lifting $Y \to \bar{\mathcal{X}}$.*

Proof. We need to check that there is a lifting at any point $y \in Y_s$ which either lies on a marking or is the reduction of a generic node. We can apply the purity lemma [Abramovich et al. 2011, Lemma 4.4] provided that the local fundamental group of Y at y is trivial and the local Picard group of Y at y is torsion-free. In order to see this, we replace R by its strict henselization and Y by the spectrum of the strict henselization of the local ring at y. We let $U = Y \setminus \{y\}$.

If y lies on a marking then Y is isomorphic to the spectrum of $R[a]^{\mathrm{sh}}$. Since this ring is local regular of dimension 2, the scheme U has trivial fundamental group by the Zariski–Nagata purity theorem, and trivial Picard group by Auslander–Buchsbaum. Hence the purity lemma applies.

If y is the reduction of a generic node, then Y is isomorphic to the strict henselization of $R[a,b]/(ab)$. Let $B_a = \mathrm{Spec}(R[a]^{\mathrm{sh}})$ resp. $B_b = \mathrm{Spec}(R[b]^{\mathrm{sh}})$ be the branches at y and $U_a = U \cap B_a$, $U_b = U \cap B_b$.

The schemes U_a and U_b have trivial fundamental group by Zariski–Nagata, and they intersect in Y in a single point of the generic fiber. Moreover the map $U_a \sqcup U_b \to U$, being finite surjective and finitely presented, is of effective descent for finite étale coverings [Grothendieck 1971, corollaire 4.12]. It then follows from the van Kampen theorem [ibid., théorème 5.1] that $\pi_1(U) = 1$.

For the computation of the local Picard group, first notice that since B_a, B_b are local regular of dimension 2 we have $\mathrm{Pic}(U_a) = \mathrm{Pic}(U_b) = 0$, and moreover it is easy to see that $H^0(U_a, \mathbb{O}_{U_a}^\times) = H^0(U_b, \mathbb{O}_{U_b}^\times) = R^\times$. Now we consider the long exact sequence in cohomology associated to the short exact sequence

$$0 \to \mathbb{O}_U^\times \to i_{a,*}\mathbb{O}_{U_a}^\times \oplus i_{b,*}\mathbb{O}_{U_b}^\times \to i_{ab,*}\mathbb{O}_{U_{ab}}^\times \to 0,$$

where the symbols $i_?$ stand for the obvious closed immersions. We obtain

$$\begin{aligned}\mathrm{Pic}(U) &= \mathrm{coker}\big(H^0(U_a, \mathbb{O}_{U_a}^\times) \oplus H^0(U_b, \mathbb{O}_{U_a}^\times) \to H^0(U_{ab}, \mathbb{O}_{U_{ab}}^\times)\big)\\ &= K^\times / R^\times = \mathbb{Z},\end{aligned}$$

which is torsion-free as desired. □

Note that we still need to introduce stack structure over special nodes of $\bar{\mathscr{X}}$.

Extension of $\mathscr{G}_\eta$ over generic points of $\bar{\mathscr{X}}_s$. Let ξ be the generic point of a component of $\bar{\mathscr{X}}_s$. Let U be the localization of $\bar{\mathscr{X}}$ at ξ and V be its inverse image in Y. Consider the closure $\mathscr{G}_\xi$ of $\mathscr{G}_\eta$ in $\mathrm{Aut}_U V$.

Proposition 3.4. *The scheme $\mathscr{G}_\xi \to U$ is a finite flat group scheme of order p, and $V \to U$ is a $\mathscr{G}_\xi$-torsor.*

Proof. This is a generalization of [Raynaud 1999, Proposition 1.2.1], see [Romagny 2011, Theorem 4.3.5]. □

Extension of $\mathscr{G}_\eta$ over the smooth locus of $\bar{\mathscr{X}}/S$. Quite generally, for a stable p-torsor $(\mathscr{X}, \mathscr{G}, Y)$ over a scheme T, by $\mathrm{Aut}_{\mathscr{X}} Y$ we denote the algebraic stack whose objects over an T-scheme U are pairs (u, f) with $u \in \mathscr{X}(U)$ and f a U-automorphism of

$Y \times_{\mathscr{X}} U$. Now consider $\bar{\mathscr{X}}^{\mathrm{sm}}$, the smooth locus of $\bar{\mathscr{X}}/S$, and its inverse image Y^{sm} in Y. Then $Y^{\mathrm{sm}} \to \bar{\mathscr{X}}^{\mathrm{sm}}$ is flat. Let $\mathscr{G}^{\mathrm{sm}}$ be the closure of $\mathscr{G}_\eta$ in $\mathrm{Aut}_{\bar{\mathscr{X}}^{\mathrm{sm}}} Y^{\mathrm{sm}}$.

Proposition 3.5. *The scheme $\mathscr{G}^{\mathrm{sm}} \to \bar{\mathscr{X}}^{\mathrm{sm}}$ is a finite flat group scheme of order p, and $Y^{\mathrm{sm}} \to \bar{\mathscr{X}}^{\mathrm{sm}}$ is a $\mathscr{G}^{\mathrm{sm}}$-torsor.*

Proof. Given Proposition 3.4, and since $\bar{\mathscr{X}}^{\mathrm{sm}}$ has local charts $U \to \bar{\mathscr{X}}^{\mathrm{sm}}$ with U regular 2-dimensional, this follows from [Abramovich 2012, Propositions 2.2.2 and 2.2.3]. □

Extension of $\mathscr{G}^{\mathrm{sm}}$ over generic nodes of $\mathscr{X}/S$. Consider the complement $\bar{\mathscr{X}}^0$ of the isolated nodes of $\bar{\mathscr{X}}_s$, and its inverse image Y^0 in Y.

Lemma 3.6. *The morphism $Y^0 \to \bar{\mathscr{X}}^0$ is flat.*

Proof. It is enough to verify the claim at the reduction x_s of an arbitrary generic node $x_\eta \in X_\eta$. Since generic nodes remain distinct in reduction, it is enough to prove that $Y \to \mathscr{X}$ is flat at a chosen point $y_s \in Y$ above x_s. Since the branches at y_s are not exchanged by $\mathscr{G}$, étale locally Y and $\mathscr{X}$ are the union of two branches which are flat over S and the restriction of $Y \to \mathscr{X}$ to each of the branches at x_s is flat. Since proper morphisms descend flatness [EGA IV$_3$ 1966, IV.11.5.3, p. 152], it follows that $Y \to \mathscr{X}$ is flat at y_s. □

Let $\mathscr{G}^0$ be the closure of $\mathscr{G}^{\mathrm{sm}}$ in $\mathrm{Aut}_{\bar{\mathscr{X}}^0} Y^0$.

Proposition 3.7. *The stack $\mathscr{G}^0 \to \bar{\mathscr{X}}^0$ is a finite flat group scheme of order p, and $Y^0 \to \bar{\mathscr{X}}^0$ is a $\mathscr{G}^0$ torsor.*

Proof. We only have to look around the closure of a generic node. Again since proper morphisms descend flatness, it is enough to prove the claim separately on the two branches. Then the result follows again from [Abramovich 2012, Propositions 2.2.2 and 2.2.3] by the same reason as in the proof of 3.5. □

Twisted structure at special nodes. Let P be a special node of X. By [Abramovich 2012, Section 3.2] there is a canonical twisted structure $\mathscr{X}$ at P determined by the local degree of Y/X. If near a given node Y_η/X_η is inseparable, then this degree is p. Otherwise Y/X has an action of $\mathbb{Z}/p\mathbb{Z}$ which is nontrivial near P, and therefore the local degree is either 1 or p. Then $\mathscr{X}$ is twisted with index p at P whenever this local degree is p. These twisted structures at the various nodes P glue to give a twisted curve $\mathscr{X}$.

We claim that this $\mathscr{X}$ is unique up to a unique isomorphism. This follows from Proposition 2.3 above. Indeed, let a be the node function which to a node P of X gives the local degree of Y/X at Y, and let r_i be the fixed indices at the sections. Then the stack of a-twisted curves over X/S with markings of indices r_i is proper over S, hence $\mathscr{X}$ is uniquely determined by $\mathscr{X}_\eta$ up to unique isomorphism.

By Lemma 3.2.1 of [Abramovich 2012], there is a unique lifting $Y \to \mathscr{X}$, and by Theorem 3.2.2 in the same reference the group scheme $\mathscr{G}^0$ extends uniquely to $\mathscr{G} \to \mathscr{X}$ such that Y is a $\mathscr{G}$-torsor. The rigidification extends immediately by taking the closure, since $\mathscr{G} \to \mathscr{X}$ is finite. $\square$

4. Examples

4.1. ***First, some nonexamples.*** Consider a smooth projective curve X of genus $h > 1$ in characteristic p and a p-torsion point in its Jacobian, corresponding to a μ_p-torsor $Y' \to X$. This is *not* a stable p-torsor in the sense of Definition 1.5: the curve Y' is necessarily unstable, with singularities which are not even nodal. In fact, $Y' \to X$ may be described by a locally logarithmic differential form ω on X, such that if locally $\omega = df/f$ for some $f \in \mathbb{O}_X^\times$ then Y' is given by an equation $z^p = f$. Since the genus $h > 1$, all differentials on X have zeroes, and each zero of ω (i.e., a zero of the derivative of f with respect to a coordinate) contributes to a unibranch singularity on Y'.

Now consider a ramified $\mathbb{Z}/p\mathbb{Z}$-cover $Y \to X$ of smooth projective curves over a field. Let $y \in Y$ be a fixed point for the action of $\mathbb{Z}/p\mathbb{Z}$ and let $x \in X$ be its image. In characteristic 0, since the stabilizer of y is a multiplicative group, the curve X may be twisted at x to yield a stable $\mathbb{Z}/p\mathbb{Z}$-torsor $Y \to \mathscr{X}$. However in characteristic p the stabilizer is additive and the result is not a $\mathbb{Z}/p\mathbb{Z}$-torsor. Hence ramified covers of smooth curves in characteristic p do not provide stable $\mathbb{Z}/p\mathbb{Z}$-torsors.

However something else does occur in both examples: the torsor $Y' \to X$ of the first example, and the branched cover $Y \to X$ in the second, lift to characteristic 0. The reduction back to characteristic p of the corresponding stable torsor "contains the original cover" in the following sense: there is a unique component $\mathscr{X}$ whose coarse moduli space is isomorphic to X. In particular that component $\mathscr{X}$ is necessarily a twisted curve, and the group scheme over it has to degenerate to α_p over the twisted points. We see a manifestation of this in the next example.

4.2. ***Limit of a p-isogeny of elliptic curves.*** Now consider the case where X is an elliptic curve, with a marked point x, over a discrete valuation ring R of characteristic 0 and residue characteristic p. For simplicity assume that R contains μ_p; let η be the generic point of $\operatorname{Spec} R$ and s the closed point of $\operatorname{Spec} R$. Given a p-torsion point on X with nontrivial reduction, we obtain a corresponding nontrivial μ_p-isogeny $Y' \to X$. Over the generic point η we can make Y'_η stable by marking the fiber P_η over x_η. But note that the reduction of P_η in Y' is not étale, hence something must modified. Since our stack is proper, a stable p-torsor $Y \to \mathscr{X}$ limiting $Y'_\eta \to X_\eta$ exists, at least over a base change of R. Here is how to describe it.

Consider the completed local ring $\hat{\mathcal{O}}_{Y',O} \simeq R[\![Z]\!]$ at the origin $O \in Y'_s$ and its spectrum $\mathbb{D}$. Then $\mathbb{D}_\eta$ is identified with an open p-adic disk modulo Galois action. Write $P_\eta = \{P_{\eta,1}, \dots, P_{\eta,p}\}$ as a sum of points permuted by the μ_p-action. Then the $P_{\eta,i}$ induce K-rational points of $\mathbb{D}_\eta$ which moreover are π-adically equidistant; i.e., the valuation $v = v_\pi(P_{\eta,i} - P_{\eta,j})$ is independent of i, j. It follows that after blowing-up the closed subscheme with ideal (π^v, Z) these points reduce to p distinct points in the exceptional divisor. Thus after twisting at the node, the fiber $Y_s \to \mathscr{X}_s$ over the special point s of R is described as follows:

$$\begin{tikzcd} Y_s \arrow[r, equal] \arrow[d] & Y'_s \cup \mathbb{P}^1 \arrow[d] & P \arrow[l, hook'] \arrow[d] \\ X'_s \arrow[r, equal] & E \cup Q & \{0\} \arrow[l, hook'] \end{tikzcd}$$

Here

- Y_s is a union of two components $Y'_s \cup \mathbb{P}^1$, attached at the origin of Y'_s,
- $\mathscr{X}_s$ is a twisted curve with two components $E \cup Q$,
- $E = X_s(\sqrt[p]{x})$ and $Q = \mathbb{P}^1(\sqrt[p]{\infty})$, with the twisted points attached,
- the map $Y_s \to \mathscr{X}_s$ decomposes into $Y'_s \to E$ and $\mathbb{P}^1 \to Q$,
- $\mathbb{P}^1 \to Q$ is an Artin–Schreier cover ramified at ∞,
- the curve is marked by the inverse image of $0 \in Q$ in $\mathbb{P}^1$, which is a $\mathbb{Z}/p\mathbb{Z}$-torsor $P \subset \mathbb{P}^1$,
- the map $Y'_s \to E$ is a lift of $Y'_s \to X_s$, and
- the group scheme $\mathscr{G} \to \mathscr{X}$ is generically étale on Q and generically μ_p on E, but the fiber over the node is α_p.

Notice that we can view $Y'_s \to E$, marked by the origin on Y'_s, as a twisted torsor as well, but this twisted torsor does not lift to characteristic 0 simply because the marked point on Y'_s can not be lifted to an invariant divisor. This is an example of the phenomenon described at the end of Section 4.1 above.

A very similar picture occurs when the cover $Y'_\eta \to X_\eta$ degenerates to an α_p-torsor. If, however, the reduction of the cover is a $\mathbb{Z}/p\mathbb{Z}$-torsor, then $Y' \to X$, marked by the fiber over the origin, is already stable and new components do not appear.

4.3. ***The double cover of $\mathbb{P}^1$ branched over 4 points.*** Consider an elliptic double cover Y over $\mathbb{P}^1$ in characteristic 0 given by the equation $y^2 = x(x-1)(x-\lambda)$. Marked by the four branched points, it becomes a stable μ_2-torsor over the twisted curve $Q = \mathbb{P}^1(\sqrt{0, 1, \infty, \lambda})$. What is its reduction in characteristic 2? We describe here one case, the others can be described in a similar way.

If the elliptic curve Y has good ordinary reduction E_s, the picture is as follows: Y_s has three components $\mathbb{P}^1 \cup E_s \cup \mathbb{P}^1$. The twisted curve $\mathscr{X}_s$ also has three rational components $Q_1 \cup Q_2 \cup Q_3$. The map splits as $\mathbb{P}^1 \to Q_1$, $E_s \to Q_2$ and $\mathbb{P}^1 \to Q_3$, where the first and last are generically μ_2-covers, and $E_s \to Q_2$ is a lift of the hyperelliptic cover $E_s \to \mathbb{P}^1$. The fibers of $\mathscr{G}$ at the nodes of X_s are both α_2. The points 0, 1, ∞, λ reduce to two pairs, one pair on each of the two $\mathbb{P}^1$ components, for instance:

$$\begin{tikzcd} & \mathbb{P}^1 \cup E_s \cup \mathbb{P}^1 \arrow[d] & \\ \{0, 1\} \arrow[r, hook] & Q_1 \cup Q_2 \cup Q_3 & \{\lambda, \infty\}. \arrow[l, hook'] \end{tikzcd}$$

Appendix A. Group schemes of order p

In this appendix, we give some complements on group schemes of order p. The main topic is the construction of an embedding of a given group scheme of order p into an affine smooth one-dimensional group scheme (an analogue of Kummer or Artin–Schreier theory). Although not strictly necessary in the paper, this result highlights the nature of our stable torsors in two respects: firstly because the original definition of generators in [Katz and Mazur 1985, §1.4] involves a smooth ambient group scheme, and secondly because the short exact sequence given by this embedding induces a long exact sequence in cohomology that may be useful for computations of torsors.

Anyway, let us now state the result.

Definition A.1. Let $G \to S$ be a finite locally free group scheme of order p.

(1) A *generator* is a morphism of S-group schemes $\gamma : (\mathbb{Z}/p\mathbb{Z})_S \to G$ such that the sections $x_i = \gamma(i)$, $0 \le i \le p-1$, are a full set of sections.

(2) A *cogenerator* is a morphism of S-group schemes $\kappa : G \to \mu_{p,S}$ such that the Cartier dual $(\mathbb{Z}/p\mathbb{Z})_S \to G^\vee$ is a generator.

We will prove the following.

Theorem A.2. *Let S be a scheme and let $G \to S$ be a finite locally free group scheme of order p. Let $\kappa : G \to \mu_{p,S}$ be a cogenerator. Then κ can be canonically inserted into a commutative diagram with exact rows*

$$\begin{tikzcd} 0 \arrow[r] & G \arrow[r] \arrow[d, "\kappa"] & \mathscr{G} \arrow[r, "\varphi_\kappa"] \arrow[d] & \mathscr{G}' \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & \mu_{p,S} \arrow[r] & \mathbb{G}_{m,S} \arrow[r, "p"] & \mathbb{G}_{m,S} \arrow[r] & 0 \end{tikzcd}$$

where $\varphi_\kappa : \mathscr{G} \to \mathscr{G}'$ is an isogeny between affine smooth one-dimensional S-group schemes with geometrically connected fibers.

In order to obtain this, we introduce two categories of invertible sheaves with sections: one related to groups with a cogenerator and one related to groups defined as kernels of isogenies, and we compare these categories.

Remark A.3. Not all group schemes of order p can be embedded into an affine smooth group scheme as in the theorem. For example, assume that there exists a closed immersion from $G = (\mathbb{Z}/p\mathbb{Z})_{\mathbb{Q}}$ to some affine smooth one-dimensional geometrically connected $\mathbb{Q}$-group scheme $\mathscr{G}$. Then $\mathscr{G}$ is a form of $\mathbb{G}_{m,\mathbb{Q}}$ and G is its p-torsion subgroup. Since $\mathscr{G}$ is trivialized by a quadratic field extension $K/\mathbb{Q}$, we obtain $G_K \simeq \mu_{p,K}$. This implies that K contains the p-th roots of unity, which is impossible for $p > 3$. Similar examples can be given for $\mathbb{Z}/p\mathbb{Z}$ over the Tate–Oort ring $\Lambda \otimes \mathbb{Q}$.

A.4. *Tate–Oort group schemes.* We recall the notations and results of the Tate–Oort classification of group schemes of order p over the ring Λ [Tate and Oort 1970, Section 2]. We introduce two fibered categories:

- a Λ-category TG of *triples* encoding *groups*, and
- a Λ-category TGC of *triples* encoding *groups with a cogenerator*.

Let $\chi : \mathbb{F}_p \to \mathbb{Z}_p$ be the unique multiplicative section of the reduction map, that is $\chi(0) = 0$ and if $m \in \mathbb{F}_p^\times$ then $\chi(m)$ is the $(p-1)$-st root of unity with residue equal to m. Set

$$\Lambda = \mathbb{Z}[\chi(\mathbb{F}_p), \frac{1}{p(p-1)}] \cap \mathbb{Z}_p.$$

There is in Λ a particular element w_p equal to p times a unit.

Definition A.5. The category TG is the category fibered over Spec Λ whose fiber categories over a Λ-scheme S are as follows.

- Its objects are the triples (L, a, b), where L is an invertible sheaf and $a \in \Gamma(S, L^{\otimes(p-1)})$, $b \in \Gamma(S, L^{\otimes(1-p)})$ satisfy $a \otimes b = w_p 1_{\mathbb{O}_S}$.
- Morphisms between (L, a, b) and (L', a', b') are the morphisms of invertible sheaves $f : L \to L'$, viewed as global sections of $L^{\otimes -1} \otimes L'$, such that $a \otimes f^{\otimes p} = f \otimes a'$ and $b' \otimes f^{\otimes p} = f \otimes b$.

The main result of [Tate and Oort 1970] is an explicit description of a covariant equivalence of fibered categories between TG and the category of finite locally free group schemes of order p. The group scheme associated to a triple (L, a, b) is denoted $G^L_{a,b}$. Its Cartier dual is isomorphic to $G^{L^{-1}}_{b,a}$.

Examples A.6. We have $(\mathbb{Z}/p\mathbb{Z})_S = G_{1,w_p}^{\mathbb{O}_S}$ and $\mu_{p,S} = G_{w_p,1}^{\mathbb{O}_S}$. Moreover if $G = G_{a,b}^L$ then a morphism $(\mathbb{Z}/p\mathbb{Z})_S \to G$ is given by a global section $u \in \Gamma(S, L)$ such that $u^{\otimes p} = u \otimes a$ and a morphism $G \to \mu_{p,S}$ is given by a global section $v \in \Gamma(S, L^{-1})$ such that $v^{\otimes p} = v \otimes b$.

Lemma A.7. *Let S be a Λ-scheme and let $G = G_{a,b}^L$ be a finite locally free group scheme of rank p over S. Then*:

(1) *Let $\gamma : (\mathbb{Z}/p\mathbb{Z})_S \to G$ be a morphism of S-group schemes given by a section $u \in \Gamma(S, L)$ such that $u^{\otimes p} = u \otimes a$. Then γ is a generator if and only if $u^{\otimes(p-1)} = a$.*

(2) *Let $\kappa : G \to \mu_{p,S}$ be a morphism of S-group schemes given by a section $v \in \Gamma(S, L^{-1})$ such that $v^{\otimes p} = v \otimes b$. Then κ is a cogenerator if and only if $v^{\otimes(p-1)} = b$.*

Proof. The proof of (2) follows from (1) by Cartier duality so we only deal with (1). The claim is local on S so we may assume that S is affine equal to $\mathrm{Spec}(R)$ and L is trivial. It follows from [Tate and Oort 1970] that $G = \mathrm{Spec}\, R[x]/(x^p - ax)$ and the section $\gamma(i) : \mathrm{Spec}(R) \xrightarrow{i} (\mathbb{Z}/p\mathbb{Z})_R \to G$ is given by the morphism of algebras $R[x]/(x^p - ax) \to R$, $x \mapsto \chi(i)u$. Thus γ is a generator if and only if $\mathrm{Norm}(f) = \prod f(\chi(i)u)$ for all functions $f = f(x)$. In particular for $f = 1+x$ one finds $\mathrm{Norm}(f) = (-1)^p a + 1$ and $\prod(1 + \chi(i)u) = (-1)^p u^{p-1} + 1$. Therefore if γ is a generator then $u^{p-1} = a$. Conversely, assuming that $u^{p-1} = a$ we want to prove that $\mathrm{Norm}(f) = \prod f(\chi(i)u)$ for all f. It is enough to prove this in the universal case where $R = \Lambda[a, b, u]/(ab - w_p, u^p - u)$. Since a is not a zerodivisor in R, it is in turn enough to prove the equality after base change to $K = R[1/a]$. Then G_K is étale and the morphism

$$K[x]/(x^p - ax) = K[x]/\prod(x - \chi(i)u) \to K^p$$

taking f to the tuple $(f(\chi(i)u))_{0 \le i \le p-1}$ is an isomorphism of algebras. Since the norm in K^p is the product of the coordinates, the result follows. □

Definition A.8. The category TGC is the category fibered over $\mathrm{Spec}\,\Lambda$ whose fibers over a Λ-scheme S are as follows.

- Its objects are the triples (L, a, v), where L is an invertible sheaf and $a \in \Gamma(S, L^{\otimes(p-1)})$, $v \in \Gamma(S, L^{\otimes -1})$ satisfy $a \otimes v^{\otimes(p-1)} = w_p 1_{\mathbb{O}_S}$.
- Morphisms between (L, a, v) and (L', a', v') are the morphisms of invertible sheaves $f : L \to L'$, viewed as global sections of $L^{\otimes -1} \otimes L'$, such that $a \otimes f^{\otimes p} = f \otimes a'$ and $v' \otimes f = v$.

By Lemma A.7, the category TGC is equivalent to the category of group schemes with a cogenerator. The functor from group schemes with a cogenerator to group schemes that forgets the cogenerator is described in terms of categories of invertible sheaves by the functor $\omega : TGC \to TG$ given by $\omega(L, a, v) = (L, a, v^{\otimes(p-1)})$.

Note also that Lemma A.7 tells us that for any locally free group scheme G over a Λ-scheme S, there exists a finite locally free morphism $S' \to S$ of degree $p-1$ such that $G \times_S S'$ admits a generator or a cogenerator.

A.9. ***Congruence group schemes.*** Here, we introduce and describe a $\mathbb{Z}$-category TCG of *triples* encoding *congruence groups*.

Let R be ring with a discrete valuation v and let $\lambda \in R$ be such that $(p-1)v(\lambda) \le v(p)$. In [Sekiguchi et al. 1989] are introduced some group schemes $H_\lambda = \operatorname{Spec} R[x]/(((1+\lambda x)^p - 1)/\lambda^p)$ with multiplication $x_1 \star x_2 = x_1 + x_2 + \lambda x_1 x_2$. (The notation in *loc. cit.* is $\mathcal{N}$.) Later Raynaud called them *congruence groups of level* λ and we will follow his terminology. We now define the analogues of these group schemes over a general base. The objects that are the input of the construction constitute the following category.

Definition A.10. The category TCG is the category fibered over $\operatorname{Spec} \mathbb{Z}$ whose fibers over a scheme S are as follows.

- Its objects are the triples (M, λ, μ), where M is an invertible sheaf over S and the global sections $\lambda \in \Gamma(S, M^{-1})$ and $\mu \in \Gamma(S, M^{p-1})$ are subject to the condition $\lambda^{\otimes(p-1)} \otimes \mu = p1_{\mathcal{O}_S}$.
- Morphisms between (M, λ, μ) and (M', λ', μ') are morphisms of invertible sheaves $f : M \to M'$ viewed as sections of $M^{-1} \otimes M'$ such that $f \otimes \lambda' = \lambda$ and $f^{\otimes(p-1)} \otimes \mu = \mu'$.

We will exhibit a functor $(M, \lambda, \mu) \rightsquigarrow H^M_{\lambda,\mu}$ from TCG to the category of group schemes, with $H^M_{\lambda,\mu}$ defined as the kernel of a suitable isogeny.

First, starting from (M, λ) we construct a smooth affine one-dimensional group scheme denoted $\mathcal{G}^{(M,\lambda)}$, or simply $\mathcal{G}^{(\lambda)}$. We see λ as a morphism $\lambda : \mathbb{V}(M) \to \mathbb{G}_{a,S}$ of (geometric) line bundles over S, where $\mathbb{V}(M) = \operatorname{Spec} \operatorname{Sym}(M^{-1})$ is the (geometric) line bundle associated to M. We define $\mathcal{G}^{(\lambda)}$ as a scheme by the fibered product

$$\begin{tikzcd} \mathcal{G}^{(\lambda)} \arrow[r, "1+\lambda"] \arrow[d] & \mathbb{G}_{m,S} \arrow[d] \\ \mathbb{V}(M) \arrow[r, "1+\lambda"] & \mathbb{G}_{a,S} \end{tikzcd} .$$

The points of $\mathcal{G}^{(\lambda)}$ with values in an S-scheme T are the global sections $u \in \Gamma(T, M \otimes \mathcal{O}_T)$ such that $1 + \lambda \otimes u$ is invertible. We endow $\mathcal{G}^{(\lambda)}$ with a multiplication given on the T-points by

$$u_1 \star u_2 = u_1 + u_2 + \lambda \otimes u_1 \otimes u_2 .$$

The zero section of $\mathbb{V}(M)$ sits in $\mathcal{G}^{(\lambda)}$ and is the unit section for the law just defined. The formula

$$(1+\lambda\otimes u_1)(1+\lambda\otimes u_2)=1+\lambda\otimes(u_1\star u_2)$$

shows that $1+\lambda:\mathcal{G}^{(\lambda)}\to\mathbb{G}_{m,S}$ is a morphism of group schemes. Moreover, if the locus where $\lambda:\mathbb{V}(M)\to\mathbb{G}_{a,S}$ is an isomorphism is scheme-theoretically dense, then $\star$ is the unique group law on $\mathcal{G}^{(\lambda)}$ for which this holds. This construction is functorial in (M,λ): given a morphism of invertible sheaves $f:M\to M'$, in other words a global section of $M^{-1}\otimes M'$, such that $f\otimes\lambda'=\lambda$, there is a morphism $f:\mathcal{G}^{(\lambda)}\to\mathcal{G}^{(\lambda')}$ making the diagram

$$\begin{tikzcd}
\mathcal{G}^{(\lambda)} \arrow[r, "1+\lambda"] \arrow[d, "f"'] & \mathbb{G}_{m,S} \\
\mathcal{G}^{(\lambda')} \arrow[ur, "1+\lambda'"'] &
\end{tikzcd}$$

commutative. The notation is coherent since that morphism is indeed induced by the extension of f to the sheaves of symmetric algebras.

Then, we use the section $\mu\in\Gamma(S,M^{p-1})$ and the relation $\lambda^{\otimes(p-1)}\otimes\mu=p1_{\mathcal{O}_S}$ to define an isogeny φ fitting into a commutative diagram

$$\begin{tikzcd}
\mathcal{G}^{(\lambda)} \arrow[r, "\varphi"] \arrow[d, "1+\lambda"'] & \mathcal{G}^{(\lambda^{\otimes p})} \arrow[d, "1+\lambda^{\otimes p}"] \\
\mathbb{G}_{m,S} \arrow[r, "^\wedge p"] & \mathbb{G}_{m,S}\ .
\end{tikzcd}$$

The formula for φ is given on the T-points $u\in\Gamma(T,M\otimes\mathcal{O}_T)$ by

$$\varphi(u)=u^{\otimes p}+\sum_{i=1}^{p-1}\left\{\begin{smallmatrix}p\\ i\end{smallmatrix}\right\}\lambda^{\otimes(i-1)}\otimes\mu\otimes u^{\otimes i},$$

where $\left\{\begin{smallmatrix}p\\ i\end{smallmatrix}\right\}=\frac{1}{p}\binom{p}{i}$ is the binomial coefficient divided by p. In order to check that the diagram is commutative and that φ is an isogeny, we may work locally on S hence we may assume that S is affine and that $M=\mathcal{O}_S$. In this case, the two claims follow from the universal case; i.e., from points (1) and (2) in the following lemma.

Lemma A.11. *Let $\mathcal{O}=\mathbb{Z}[E,F]/(E^{p-1}F-p)$ and let $\lambda,\mu\in\mathcal{O}$ be the images of the indeterminates E, F. Then, the polynomial*

$$P(X)=X^p+\sum_{i=1}^{p-1}\left\{\begin{smallmatrix}p\\ i\end{smallmatrix}\right\}\lambda^{i-1}\mu X^i\ \in\mathcal{O}[X]$$

satisfies:

(1) $1+\lambda^p P(X) = (1+\lambda X)^p$, *and*

(2) $P(X+Y+\lambda XY) = P(X)+P(Y)+\lambda^p P(X)P(Y)$.

Proof. Point (1) follows by expanding $(1+\lambda X)^p$ and using the fact that $p=\lambda^{p-1}\mu$ in $\mathbb{O}$. Then we compute:

$$\begin{aligned}1+\lambda^p P(X+Y+\lambda XY) &= (1+\lambda(X+Y+\lambda XY))^p \\ &= (1+\lambda X)^p(1+\lambda Y)^p \\ &= (1+\lambda^p P(X))(1+\lambda^p P(Y)) \\ &= 1+\lambda^p(P(X)+P(Y)+\lambda^p P(X)P(Y)).\end{aligned}$$

Since λ is a nonzerodivisor in $\mathbb{O}$, point (2) follows. □

Definition A.12. We denote by $H^M_{\lambda,\mu}$ the kernel of φ, and call it the *congruence group scheme* associated to (M,λ,μ).

This construction is functorial in (M,λ,μ). Precisely, consider two triples (M,λ,μ) and (M',λ',μ') and a morphism of invertible sheaves $f: M\to M'$ viewed as a section of $M^{-1}\otimes M'$ such that $f\otimes\lambda'=\lambda$ and $f^{\otimes(p-1)}\otimes\mu=\mu'$. Then we have morphisms $f:\mathcal{G}^{(\lambda)}\to\mathcal{G}^{(\lambda')}$ and $f^{\otimes p}:\mathcal{G}^{(\lambda^{\otimes p})}\to\mathcal{G}^{(\lambda'^{\otimes p})}$ compatible with the isogenies φ and φ', and f induces a morphism $H^M_{\lambda,\mu}\to H^{M'}_{\lambda',\mu'}$. Note also that the image of $H^M_{\lambda,\mu}$ under $1+\lambda:\mathcal{G}^{(\lambda)}\to\mathbb{G}_{m,S}$ factors through $\mu_{p,S}$, so that by construction $H^M_{\lambda,\mu}$ comes embedded into a diagram

$$\begin{tikzcd} 0 \arrow[r] & H^M_{\lambda,\mu} \arrow[r] \arrow[d, "\kappa"] & \mathcal{G}^{(\lambda)} \arrow[r] \arrow[d, "1+\lambda"] & \mathcal{G}^{(\lambda^{\otimes p})} \arrow[r] \arrow[d, "1+\lambda^{\otimes p}"] & 0 \\ 0 \arrow[r] & \mu_{p,S} \arrow[r] & \mathbb{G}_{m,S} \arrow[r] & \mathbb{G}_{m,S} \arrow[r] & 0. \end{tikzcd}$$

The formation of this diagram is also functorial.

Lemma A.13. *The morphism $\kappa: H^M_{\lambda,\mu}\to\mu_{p,S}$ is a cogenerator.*

Proof. We have to show that the dual map $(\mathbb{Z}/p\mathbb{Z})_S\to(H^M_{\lambda,\mu})^\vee$ is a generator. This means verifying locally on S certain equalities of norms. Hence we may assume that S is affine and that M is trivial, then reduce to the universal case where S is the spectrum of the ring $\mathbb{O}$ with elements λ,μ satisfying $\lambda^{p-1}\mu=p$ as in Lemma A.11, and finally restrict to the schematically dense open subscheme $S'=D(\lambda)\subset S$. Since $\mathcal{G}^{(\lambda)}\times_S S'\to\mathbb{G}_{m,S'}$ is an isomorphism, then $H^M_{\lambda,\mu}\times_S S'\to\mu_{p,S'}$ and the dual morphism also are. The claim follows immediately. □

A.14. ***Equivalence between TGC and $TCG \otimes_{\mathbb{Z}} \Lambda$.*** The results of the previous subsection imply that for a Λ-scheme S, a triple $(M, \lambda, \mu) \in TCG(S)$ gives rise in a functorial way to a finite locally free group scheme with cogenerator $\kappa : H^M_{\lambda,\mu} \to \mu_{p,S}$, that is, an object of $TGC(S)$.

Theorem A.15. *The functor*

$$F : TCG \otimes_{\mathbb{Z}} \Lambda \to TGC$$

defined above is an equivalence of fibered categories over Λ. If (M, λ, μ) has image (L, a, v) then $H^M_{\lambda,\mu} \simeq G^L_{a,v^{\otimes(p-1)}}$.

Proof. The main point is to describe F in detail using the Tate–Oort classification, and to see that it is essentially surjective. The description of the action of F on morphisms and the verification that it is fully faithful offers no difficulty and will be omitted.

Let (M, λ, μ) be a triple in $TCG(S)$ and let $G = H^M_{\lambda,\mu}$. We use the notations of Section 2 of [Tate and Oort 1970], in particular the structure of the group μ_p is described by a function z, the sheaf of χ-eigensections $J = y\mathbb{O}_S \subset \mathbb{O}_{\mu_p}$ with distinguished generator $y = (p-1)e_1(1-z)$, and constants

$$w_1 = 1,\ w_2,\ \dots,\ w_{p-1},\ w_p = pw_{p-1} \in \Lambda.$$

The augmentation ideal of the algebra $\mathbb{O}_G$ is the sheaf I generated by M^{-1}, and by Tate and Oort's results the subsheaf of χ-eigensections is the sheaf $I_1 = e_1(I)$, where e_1 is the $\mathbb{O}_S$-linear map defined in [Tate and Oort 1970]. It is an invertible sheaf and L is (by definition) its inverse.

We claim that in fact $I_1 = e_1(M^{-1})$. In order to see this, we may work locally. Let x be a local generator for M^{-1} and let

$$t := (p-1)e_1(-x) \in I_1.$$

Let us write $\lambda = \lambda_0 x$ for some local function λ_0. We first prove that

$$x = \frac{1}{1-p}\left(t + \frac{\lambda_0 t^2}{w_2} + \cdots + \frac{\lambda_0^{p-2} t^{p-1}}{w_{p-1}}\right). \tag{$\star$}$$

In fact, by construction the map $\mathbb{O}_{\mu_p} \to \mathbb{O}_G$ is given by $z = 1 + \lambda_0 x$, so we get $y = (p-1)e_1(1-z) = \lambda_0 t$. In order to check the expression for x in terms of t, we can reduce to the universal case (Lemma A.11). Then λ_0 is not a zerodivisor and we can harmlessly multiply both sides by λ_0. In this form, the equality to be proven is nothing else than the identity (16) in [Tate and Oort 1970]. Now write $t = \alpha t^*$ with t^* a local generator for I_1 and α a local function. Using ($\star$) we find that $x = \alpha x^*$ for some $x^* \in \mathbb{O}_G$. Since x generates M^{-1} in the fibers over S, this proves that α is invertible. Finally t is a local generator for I_1 and this finishes the proof that $I_1 = e_1(M^{-1})$.

Let $x^\vee$ be the local generator for M dual to x and write $\mu = \mu_0 (x^\vee)^{\otimes(p-1)}$ for some local function μ_0 such that $(\lambda_0)^{p-1}\mu_0 = p$. Let $t^\vee$ be the local generator for L dual to t. We define a local section a of $L^{\otimes(p-1)}$ by

$$a = w_{p-1}\mu_0(t^\vee)^{\otimes(p-1)}$$

and a local section v of L^{-1} by

$$v = \lambda_0 t.$$

These sections are independent of the choice of the local generator x, because if $x' = \alpha x$ then

$$(x')^\vee = \alpha^{-1}x^\vee \ ; \ t' = \alpha t \ ; \ (t')^\vee = \alpha^{-1}t^\vee \ ; \ \lambda_0' = \alpha^{-1}\lambda_0 \ ; \ \mu_0' = \alpha^{p-1}\mu_0$$

so that

$$a' = w_{p-1}\mu_0'(t'^\vee)^{\otimes(p-1)} = w_{p-1}\alpha^{p-1}\mu_0\alpha^{1-p}(t^\vee)^{\otimes(p-1)} = a$$

and

$$v' = \lambda_0' t' = \alpha^{-1}\lambda_0\alpha t = v.$$

They glue to global sections a and v satisfying

$$a \otimes v^{\otimes(p-1)} = w_p 1_{\mathbb{O}_S}.$$

Let us prove that a and v are indeed the sections defining G and the cogenerator in the Tate–Oort classification. The verification for a amounts to checking that the relation

$$t^p = w_{p-1}\mu_0 t$$

holds in the algebra $\mathbb{O}_G$. This may be seen in the universal case where λ_0 is not a zerodivisor, hence after multiplying by $(\lambda_0)^p$ this follows from the equality $y^p = w_p y$ from [Tate and Oort 1970]. The verification for v amounts to noting that the cogenerator $G \to \mu_{p,S}$ is indeed given by $y \mapsto v$.

This completes the description of F on objects. Finally we prove that F is essentially surjective. Assume given (L, a, v) and let t be a local generator for $I_1 = L^{-1}$. Write $a = w_{p-1}\mu_0(t^\vee)^{\otimes(p-1)}$, $v = \lambda_0 t$ and define an element $x \in \mathbb{O}_G$ by the expression $(\star)$ above. If we change the generator t to another $t' = \alpha t$, then $\lambda_0' = \alpha^{-1}\lambda_0$ and $x' = \alpha x$. It follows that the subsheaf of $\mathbb{O}_G$ generated by x does not depend on the choice of the generator for I_1, call it N. Reducing to the universal case as before, we prove that $t = (p-1)e_1(-x)$. This shows that in fact N is an invertible sheaf and we take M to be its inverse. Finally we define sections $\lambda \in \Gamma(S, M^{-1})$ and $\mu \in \Gamma(S, M^{\otimes(p-1)})$ by the local expressions $\lambda = \lambda_0 x$ and

$\mu = \mu_0(x^\vee)^{\otimes(p-1)}$. It is verified like in the case of a, v before that they do not depend on the choice of t and hence are well-defined global sections. The equality $\lambda^{\otimes(p-1)} \otimes \mu = p1_{\mathcal{O}_S}$ holds true and the proof is now complete. □

Proof of Theorem A.2. We keep the notation of the theorem. Since the construction of the isogeny φ_κ and the whole commutative diagram is canonical, if we perform it after fppf base change $S' \to S$ then it will descend to S. We choose $S' = S_1 \amalg S_2$, where $S_1 = S \otimes_{\mathbb{Z}} \mathbb{Z}[1/p]$ and $S_2 = S \otimes_{\mathbb{Z}} \Lambda$. Over S_1 the group scheme G is étale and the cogenerator is an isomorphism by [Katz and Mazur 1985, Lemma 1.8.3]. We take $\mathcal{G} = \mathcal{G}' = \mathbb{G}_{m,S}$ and φ_κ is the p-th power map. Over S_2 we use Theorem A.15 which provides a canonical isomorphism between κ and $H^M_{\lambda,\mu}$ with its canonical cogenerator, embedded into a diagram of the desired form. This completes the proof. □

Appendix B. Weil restriction of closed subschemes

Let $Z \to X$ be a morphism of S-schemes (or algebraic spaces) and denote by $h : X \to S$ the structure map. The Weil restriction h_*Z of Z along h is the functor on S-schemes defined by $(h_*Z)(T) = \mathrm{Hom}_X(X \times_S T, Z)$. It may be seen as a left adjoint to the pullback along h, or as the functor of sections of $Z \to X$.

If $Z \to X$ is a closed immersion of schemes (or algebraic spaces) of finite presentation over S, there are two main cases where h_*Z is known to be representable by a closed subscheme of S. As is well-known, this has applications to representability of various equalizers, kernels, centralizers, normalizers, etc. These two cases are:

(i) if $X \to S$ is proper flat and $Z \to S$ is separated, by the Grothendieck–Artin theory of the Hilbert scheme,

(ii) if $X \to S$ is essentially free, by [Grothendieck 1970, théorème 6.4].

In this appendix, we want to prove that h_*Z is representable by a closed subscheme of S in a case that includes both situations and is often easier to check in practice, namely the case where $X \to S$ is flat and pure.

B.1. ***Essentially free and pure morphisms.*** We recall the notions of essentially free and pure morphisms and check that essentially free morphisms and proper morphisms are pure.

In [Grothendieck 1970, §6], a morphism $X \to S$ is called *essentially free* if and only if there exists a covering of S by open affine subschemes S_i, and for each i an affine faithfully flat morphism $S'_i \to S_i$ and a covering of $X'_i = X \times_S S'_i$ by open affine subschemes $X'_{i,j}$ such that the function ring of $X'_{i,j}$ is free as a module over the function ring of S'_i.

In fact, the proof of [Grothendieck 1970, théorème 6.4] works just as well with a slightly weaker notion than freeness of modules. Namely, for a module M over a ring A, let us say that M is *quasireflexive* if the canonical map $M \to M^{\vee\vee}$ from M to its linear bidual is injective after any change of base ring $A \to A'$. It is a simple exercise to see that this is equivalent to M being a submodule of a product module A^I for some set I, over A and after any base change $A \to A'$. For instance, free modules, projective modules, product modules are quasireflexive. This gives rise to a notion of *essentially quasireflexive* morphism, and in particular *essentially projective* morphism. Then inspection of the proof of [Grothendieck 1970, théorème 6.4] shows that it remains valid for these morphisms.

In [Raynaud and Gruson 1971, 3.3.3], a morphism locally of finite type $X \to S$ is called *pure* if and only if for all points $s \in S$, with henselization $(\tilde{S}, \tilde{s})$, and all points $\tilde{x} \in \tilde{X}$ where $\tilde{X} = X \times_S \tilde{S}$, if $\tilde{x}$ is an associated point in its fiber then its closure in $\tilde{X}$ meets the special fiber. Examples of pure morphisms include proper morphisms (by the valuative criterion for properness) and morphisms locally of finite type and flat, with geometrically irreducible fibers without embedded components [ibid., 3.3.4].

Finally if $X \to S$ is locally of finite presentation and essentially free, then it is pure. Indeed, with the notations above for an essentially free morphism, one sees using [ibid., 3.3.7] that it is enough to see that for each i, j the scheme $X'_{i,j}$ is pure over S'_i. But since the function ring of $X'_{i,j}$ is free over the function ring of S'_i, this follows from [ibid., 3.3.5].

B.2. *Representability of h_*Z.*

Proposition B.3. *Let $h : X \to S$ be a morphism of finite presentation, flat and pure, and let $Z \to X$ be a closed immersion. Then the Weil restriction h_*Z is representable by a closed subscheme of S.*

Proof. The question is local for the étale topology on S. Let $s \in S$ be a point and let $\mathbb{O}^h$ be the henselization of the local ring at s. By [Raynaud and Gruson 1971, 3.3.13], for each $x \in X$ lying over s, there exists an open affine subscheme U^h_x of $X \times_S \operatorname{Spec}(\mathbb{O}^h)$ containing x and whose function ring is free as an $\mathbb{O}^h$-module. Since X_s is quasi-compact, there is a finite number of points $x_1, \dots, x_n$ such that the open affines $U^h_i = U^h_{x_i}$ cover it. Since X is locally of finite presentation, after restricting to an étale neighborhood $S' \to S$ of s, there exist affine open subschemes U_i of X inducing the U^h_i. According to [ibid., 3.3.8], the locus of the base scheme S where $U_i \to S$ is pure is open, so after shrinking S we may assume that for each i the affine U_i is flat and pure. This means that its function ring is projective by [ibid., 3.3.5]. In other words, the union $U = U_1 \cup \cdots \cup U_n$ is essentially projective over S in the terms of the comments in B.1. If $k : U \to X$ denotes the structure map, it follows from [Grothendieck 1970, théorème 6.4] that $k_*(Z \cap U)$ is representable by a closed subscheme of S. On the other hand, according to [Romagny 2011, 3.1.8], replacing S again by a smaller neighborhood of s, the open immersion $U \to X$

is S-universally schematically dense. One deduces immediately that the natural morphism $h_*Z \to k_*(Z \cap U)$ is an isomorphism. This finishes the proof. $\square$

This proposition has a long list of corollaries and applications, listed in [Grothendieck 1970, §6]. In particular:

Corollary B.4. *Let $X \to S$ be a morphism of finite presentation, flat and pure and $Y \to S$ a separated morphism. Consider two morphisms $f, g : X \to Y$. Then the condition $f = g$ is represented by a closed subscheme of S.*

Proof. Apply the previous proposition to the pullback of the diagonal of Y along $(f, g) : X \to Y \times_S Y$. $\square$

Acknowledgements

We thank Sylvain Maugeais for helping us clarify a point in this paper.

References

[Abramovich 2012] D. Abramovich, "Raynaud's group-scheme and reduction of coverings", pp. 1–18 in *Number theory, analysis and geometry*, edited by D. Goldfeld et al., Springer, New York, 2012. MR 2867909

[Abramovich and Vistoli 2002] D. Abramovich and A. Vistoli, "Compactifying the space of stable maps", *J. Amer. Math. Soc.* **15**:1 (2002), 27–75. MR 2002i:14030 Zbl 0991.14007

[Abramovich et al. 2003] D. Abramovich, A. Corti, and A. Vistoli, "Twisted bundles and admissible covers", *Comm. Algebra* **31**:8 (2003), 3547–3618. MR 2005b:14049 Zbl 1077.14034

[Abramovich et al. 2008] D. Abramovich, M. Olsson, and A. Vistoli, "Tame stacks in positive characteristic", *Ann. Inst. Fourier* (*Grenoble*) **58**:4 (2008), 1057–1091. MR 2009c:14002 Zbl 1222.14004

[Abramovich et al. 2011] D. Abramovich, M. Olsson, and A. Vistoli, "Twisted stable maps to tame Artin stacks", *J. Algebraic Geom.* **20**:3 (2011), 399–477. MR 2012c:14024 Zbl 1225.14020

[Bouw and Wewers 2004] I. I. Bouw and S. Wewers, "Stable reduction of modular curves", pp. 1–22 in *Modular curves and Abelian varieties*, edited by J. Cremona et al., Progr. Math. **224**, Birkhäuser, Basel, 2004. MR 2005b:11079 Zbl 1147.11316

[Edixhoven 1990] B. Edixhoven, "Minimal resolution and stable reduction of $X_0(N)$", *Ann. Inst. Fourier* (*Grenoble*) **40**:1 (1990), 31–67. MR 92f:11080 Zbl 0679.14009

[EGA IV$_3$ 1966] A. Grothendieck, *Eléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, III*, Publications Mathématiques de l'I.H.E.S. **28**, Inst. Hautes Études Sci., Paris, 1966. MR 0217086 Zbl 0144.19904

[Grothendieck 1970] A. Grothendieck, "Groupes diagonalisables", exposé VIII in *Schémas en groupes, II: Groupes de type multiplicatif, et structure des schémas en groupes généraux* (SGA 3), Lecture Notes in Mathematics **152**, Springer, Berlin, 1970. MR 43 #223b Zbl 0209.24201

[Grothendieck 1971] A. Grothendieck, "Descente des morphismes étales; application au groupe fondamental", exposé IX in *Revêtements étales et groupe fondamental* (SGA 1), Lecture Notes in Mathematics **224**, Springer-Verlag, Berlin, 1971. Reprinted Soc. Math. de France, Paris, 2003. MR 0354651 (50 #7129)

[Katz and Mazur 1985] N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies **108**, Princeton University Press, 1985. MR 86i:11024 Zbl 0576.14026

[McMurdy and Coleman 2010] K. McMurdy and R. Coleman, "Stable reduction of $X_0(p^3)$", *Algebra Number Theory* **4**:4 (2010), 357–431. MR 2011k:14022 Zbl 1215.11060

[Raynaud 1999] M. Raynaud, "Spécialisation des revêtements en caractéristique $p > 0$", *Ann. Sci. École Norm. Sup.* (4) **32**:1 (1999), 87–126. MR 2000e:14016 Zbl 0999.14004

[Raynaud and Gruson 1971] M. Raynaud and L. Gruson, "Critères de platitude et de projectivité: Techniques de "platification" d'un module", *Invent. Math.* **13** (1971), 1–89. MR 46 #7219

[Romagny 2011] M. Romagny, "Effective models of group schemes", preprint, 2011. To appear in *J. Algebraic Geom.*

[Sekiguchi et al. 1989] T. Sekiguchi, F. Oort, and N. Suwa, "On the deformation of Artin–Schreier to Kummer", *Ann. Sci. École Norm. Sup.* (4) **22**:3 (1989), 345–375. MR 91g:14041 Zbl 0714.14024

[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 2011-02-26 Revised 2011-06-11 Accepted 2011-07-17

abrmovic@math.brown.edu *Department of Mathematics, Brown University, Box 1917, 151 Thayer Street, Providence, RI 02912, United States*

romagny@math.jussieu.fr *Université Pierre et Marie Curie (Paris 6), 4 place Jussieu, F-75252 Paris, France*

ALGEBRA AND NUMBER THEORY 6:4(2012)

msp

Block components of the Lie module for the symmetric group

Roger M. Bryant and Karin Erdmann

Let F be a field of prime characteristic p and let B be a nonprincipal block of the group algebra FS_r of the symmetric group S_r. The block component $\mathrm{Lie}(r)_B$ of the Lie module $\mathrm{Lie}(r)$ is projective, by a result of Erdmann and Tan, although $\mathrm{Lie}(r)$ itself is projective only when $p \nmid r$. Write $r = p^m k$, where $p \nmid k$, and let S_k^* be the diagonal of a Young subgroup of S_r isomorphic to $S_k \times \cdots \times S_k$. We show that $p^m \,\mathrm{Lie}(r)_B \cong (\mathrm{Lie}(k){\uparrow}_{S_k^*}^{S_r})_B$. Hence we obtain a formula for the multiplicities of the projective indecomposable modules in a direct sum decomposition of $\mathrm{Lie}(r)_B$. Corresponding results are obtained, when F is infinite, for the r-th Lie power $L^r(E)$ of the natural module E for the general linear group $\mathrm{GL}_n(F)$.

1. Introduction and summary of results

Let r be a positive integer and let S_r denote the symmetric group of degree r. For any field F the Lie module $\mathrm{Lie}(r)$ is the FS_r-module given by the right ideal $\omega_r FS_r$ of the group algebra FS_r where ω_r is the *Dynkin–Specht–Wever element*, defined by $\omega_1 = 1$ and, for $r \geqslant 2$,

$$\omega_r = (1 - c_r)(1 - c_{r-1}) \cdots (1 - c_2), \tag{1-1}$$

where c_i is the i-cycle $(1\ i\ i{-}1\ \ldots\ 2)$. It is known that $\mathrm{Lie}(r)$ has dimension $(r-1)!$ (see Section 2A).

If F has prime characteristic p and $p \nmid r$ then $\mathrm{Lie}(r)$ is a direct summand of FS_r because, as is well known, $\omega_r^2 = r\omega_r$ (see, for example, [Bryant 2009, Section 3]); so in this case $\mathrm{Lie}(r)$ is projective. However, if $\mathrm{char}\, F = p$ and $p \mid r$ then $\mathrm{Lie}(r)$ is not projective (because its dimension is not then divisible by the order of a Sylow p-subgroup of S_r), but it was shown recently that every nonprincipal block component of $\mathrm{Lie}(r)$ *is* projective (see [Erdmann and Tan 2011]). Here we show that each such component can be described in a surprisingly simple way in terms of $\mathrm{Lie}(k)$, where k is the p'-part of r.

Supported by EPSRC Standard Research Grant EP/G025487/1.

MSC2010: primary 20C30; secondary 20G43, 20C20.

Keywords: Lie module, symmetric group, Lie power, Schur algebra, block.

The Lie module occurs naturally in a number of contexts in algebra, algebraic topology and elsewhere (see [Erdmann and Tan 2011] for a fuller discussion). Here we shall only be concerned with the connection with free Lie algebras, where our results on the Lie module give new insight into the module structure of the homogeneous components.

Let G be a group and V an FG-module. Let $L^r(V)$ denote the homogeneous component of degree r in the free Lie algebra $L(V)$ freely generated by any basis of V. (Here $L(V)$ may be regarded as the Lie subalgebra generated by V in the tensor algebra or free associative algebra on V: see Section 2A.) The vector space $L^r(V)$ is called the *r-th Lie power* of V and it inherits the structure of an FG-module.

Suppose that F is infinite, let n be a positive integer, and let E denote the natural n-dimensional module over F for the general linear group $\mathrm{GL}_n(F)$. Then $L^r(E)$, as a module for $\mathrm{GL}_n(F)$, is a homogeneous polynomial module of degree r. In other words it is a module for the Schur algebra $S_F(n,r)$ (see [Green 1980]). In the case where $n \geqslant r$ the Schur functor f_r maps $S_F(n,r)$-modules to FS_r-modules and we have $f_r(L^r(E)) \cong \mathrm{Lie}(r)$ (see Section 2D).

Recall that if $\mathcal{B}$ is the set of blocks of an algebra Γ and V is a Γ-module then we may write V as a direct sum of block components: $V = \bigoplus_{B\in\mathcal{B}} V_B$, where $V_B \in B$ for all B. Our main results concern the block components of $\mathrm{Lie}(r)$ and $L^r(E)$ when F has prime characteristic p. The basic results are for $\mathrm{Lie}(r)$, and the results for $L^r(E)$ are obtained from these by means of the Schur functor. To state the results we write $r = p^m k$ where $m \geqslant 0$, $k \geqslant 1$, and $p \nmid k$.

Let S_k^* be a subgroup of S_r such that $S_k^* \cong S_k$ and S_k^* is the diagonal of a Young subgroup of S_r isomorphic to $S_k \times \cdots \times S_k$ (with p^m factors). (See Section 2D for more details.) Since $S_k^* \cong S_k$ we may regard $\mathrm{Lie}(k)$ as an FS_k^*-module and, since $p \nmid k$, this module is projective. Thus the induced module $\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r}$ is also projective. It was proved in [Erdmann and Tan 2011, Theorem 3.1] that if B is a nonprincipal block of FS_r then $\mathrm{Lie}(r)_B$ is projective. Here we shall prove (Theorem 3.1) that

$$\mathrm{Lie}(r)_B \cong \frac{1}{p^m}\,(\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r})_B \tag{1-2}$$

when B satisfies the condition $\tilde{B} \neq \varnothing$ (see Section 2B): this condition is satisfied when B is nonprincipal. (The notation $U \cong (1/q)\,V$ used in (1-2) means that $q\,U \cong V$, where $q\,U$ denotes $U \oplus \cdots \oplus U$ with q summands.) The special case where $k = 1$ is of particular interest: it yields

$$\mathrm{Lie}(p^m)_B \cong \frac{1}{p^m}\,(FS_{p^m})_B.$$

Since (1-2) holds for each nonprincipal block B, a comparison of dimensions gives a weaker result (Corollary 3.4) for the principal block B_0:

$$\dim \mathrm{Lie}(r)_{B_0} = \frac{1}{p^m} \dim (\mathrm{Lie}(k){\uparrow}_{S_k^*}^{S_r})_{B_0}.$$

The projective indecomposable FS_r-modules may be labelled P^λ where λ ranges over the p-regular partitions of r (see Section 2B). For any FS_r-block B we write $\lambda \in B$ when $P^\lambda \in B$. Since $\mathrm{Lie}(r)_B$ is projective when B is nonprincipal, there are nonnegative integers m_λ such that

$$\mathrm{Lie}(r)_B \cong \bigoplus_{\lambda \in B} m_\lambda \, P^\lambda.$$

In Theorem 3.5 we prove that

$$m_\lambda = \frac{1}{r} \sum_{d \mid k} \mu(d) \, \beta^\lambda(\tau^{k/d}), \tag{1-3}$$

where μ is the Möbius function, τ is an element of S_r of cycle type $(k, k, \dots, k)$, and β^λ denotes the Brauer character of D^λ, the irreducible FS_r-module isomorphic to the head of P^λ.

Now suppose that F is infinite and n is a positive integer. We have observed that $L^r(E)$ is an $S_F(n,r)$-module. Similarly, $L^k(E^{\otimes p^m})$ is an $S_F(n,r)$-module, and (by the argument in [Donkin and Erdmann 1998, Section 3.1]) it is isomorphic to a direct summand of $E^{\otimes r}$. It is a consequence of [Erdmann and Tan 2011, Theorem 3.2] that if B is a block of $S_F(n,r)$ satisfying the condition $\tilde{B} \neq \varnothing$ (see Section 2C) then $L^r(E)_B$ is isomorphic to a direct sum of summands of $E^{\otimes r}$. Here we shall prove (Theorem 3.6) that

$$L^r(E)_B \cong \frac{1}{p^m} L^k(E^{\otimes p^m})_B.$$

The indecomposable summands of $E^{\otimes r}$ are tilting modules $T(\lambda)$, where λ is a p-regular partition of r with at most n parts (see Section 2C). For any $S_F(n,r)$-block B we write $\lambda \in B$ when $T(\lambda) \in B$. In Theorem 3.7 we prove that if $\tilde{B} \neq \varnothing$ then

$$L^r(E)_B \cong \bigoplus_{\lambda \in B} m_\lambda \, T(\lambda),$$

where the multiplicities m_λ are given by (1-3). This extends [Donkin and Erdmann 1998, Section 3.3, Theorem], which gives the same result in the case where $p \nmid r$.

All modules over fields in this paper will be assumed to be finite-dimensional, and modules for algebras are right modules unless otherwise specified.

2. Preliminaries

2A. ***The Lie module.*** Let r and n be positive integers where $n \geqslant r$. Let Δ be the free associative ring ($\mathbb{Z}$-algebra) on free generators $x_1, \dots, x_n$ and let L be the

Lie subring of Δ generated by $x_1, \ldots, x_n$. By [Bourbaki 1972, chapitre II, §3, théorème 1], L is free on $x_1, \ldots, x_n$. Let Δ_r denote the homogeneous component of Δ of degree r. Then S_r has a left action by "place permutations" on Δ_r, given by $\alpha(y_1 \cdots y_r) = y_{1\alpha} \cdots y_{r\alpha}$ for all $\alpha \in S_r$ and all $y_1, \ldots, y_r \in \{x_1, \ldots, x_n\}$. (Note that we write multiplication in S_r from left to right.) Hence Δ_r is a left $\mathbb{Z}S_r$-module. Let ω_r be the element of $\mathbb{Z}S_r$ given by (1-1). Then it is well known and easily verified that $\omega_r(y_1 \cdots y_r) = [y_1, \ldots, y_r]$ where $[y_1, \ldots, y_r]$ denotes the left-normed Lie product $[\cdots[[y_1, y_2], y_3], \ldots, y_r]$.

The group S_r also has a right action on Δ by automorphisms, where $x_i\alpha = x_{i\alpha}$ for $i = 1, \ldots, r$ and $x_i\alpha = x_i$ for $i > r$. Thus Δ_r becomes a $(\mathbb{Z}S_r, \mathbb{Z}S_r)$-bimodule. Let Δ_r^0 be the $\mathbb{Z}$-subspace of Δ_r spanned by the monomials $x_{1\alpha} \cdots x_{r\alpha}$ with $\alpha \in S_r$. Thus Δ_r^0 is a subbimodule of Δ_r and the map $\xi : \mathbb{Z}S_r \to \Delta_r^0$ defined by $\xi(\alpha) = x_{1\alpha} \cdots x_{r\alpha}$ is an isomorphism of bimodules.

Let $L_r^0 = L \cap \Delta_r^0$. Then L_r^0 is spanned over $\mathbb{Z}$ by all elements of the form $[x_{1\alpha}, \ldots, x_{r\alpha}]$. Also, by [Bourbaki 1972, chapitre II, §3, théorème 2], L_r^0 is free of rank $(r-1)!$ as a $\mathbb{Z}$-module. We have $\omega_r \Delta_r^0 = L_r^0$. Thus the isomorphism ξ maps $\omega_r\mathbb{Z}S_r$ to L_r^0, and so $\omega_r\mathbb{Z}S_r$ is isomorphic to L_r^0 as a right $\mathbb{Z}S_r$-module.

All of the above still applies if $\mathbb{Z}$ is replaced by R, where R is an arbitrary commutative ring with unity. Also (using subscripts to show coefficient rings) we have $R \otimes_{\mathbb{Z}} L_{r,\mathbb{Z}}^0 \cong L_{r,R}^0$. The Lie module $\mathrm{Lie}_R(r)$ is the RS_r-module defined by $\mathrm{Lie}_R(r) = \omega_r RS_r$. Thus $\mathrm{Lie}_R(r) \cong L_{r,R}^0$. It follows that $\mathrm{Lie}_R(r) \cong R \otimes_{\mathbb{Z}} \mathrm{Lie}_{\mathbb{Z}}(r)$ and $\mathrm{Lie}_R(r)$ is free of rank $(r-1)!$ as an R-module. If F is a field we have

$$\dim \mathrm{Lie}_F(r) = (r-1)! \tag{2-1}$$

and, if F is understood, we write $\mathrm{Lie}_F(r)$ as $\mathrm{Lie}(r)$.

When K is a field of characteristic zero there is a formula for the character ψ_r of $\mathrm{Lie}_K(r)$. Let μ denote the Möbius function and let σ be an r-cycle of S_r. For each divisor d of r let C_d denote the conjugacy class of $\sigma^{r/d}$ in S_r. Then, for $g \in S_r$,

$$\psi_r(g) = \begin{cases} \mu(d)(r-1)!/|C_d| & \text{if } g \in C_d, \\ 0 & \text{if } g \notin C_d \text{ for all } d. \end{cases} \tag{2-2}$$

(See, for example, [Donkin and Erdmann 1998, Section 3.2].) Hence, if θ is any class function on S_r with values in K and we write

$$(\theta, \psi_r)_{S_r} = \frac{1}{|S_r|} \sum_{g \in S_r} \theta(g)\psi_r(g^{-1}), \tag{2-3}$$

we have

$$(\theta, \psi_r)_{S_r} = \frac{1}{r} \sum_{d|r} \mu(d)\theta(\sigma^{r/d}). \tag{2-4}$$

2B. ***Representations of S_r.*** By a partition of a nonnegative integer r we mean, as usual, a finite sequence $\lambda = (\lambda_1, \ldots, \lambda_s)$ of integers satisfying $\lambda_1 \geqslant \cdots \geqslant \lambda_s > 0$ and $\lambda_1 + \cdots + \lambda_s = r$. We call $\lambda_1, \ldots, \lambda_s$ the *parts* of λ. We write $\Lambda^+(r)$ for the set of all partitions of r. If $r = 0$ then $\Lambda^+(r)$ contains only the empty partition, which we denote by $\varnothing$. Let p be a prime number. A partition λ is *p-regular* if λ does not have p or more equal parts, and we write $\Lambda_p^+(r)$ for the set of all p-regular partitions of r. The *p-core* of a partition λ of r is the partition $\tilde{\lambda}$ of r', for some $r' \leqslant r$, obtained from (the diagram of) λ by the removal of as many "rim p-hooks" as possible: see [James and Kerber 1981, Section 2.7]. We write $\mathscr{C}_r$ for the set of all p-cores of elements of $\Lambda^+(r)$.

Let r be a positive integer and let F be a field of prime characteristic p. The irreducible FS_r-modules may be labelled (up to isomorphism) as D^λ with $\lambda \in \Lambda_p^+(r)$, where D^λ is a quotient module of the Specht module S^λ (see [James and Kerber 1981, 7.1.14]). Here $D^{(r)}$ is isomorphic to the trivial FS_r-module F because $S^{(r)} \cong F$. For each λ we write P^λ for the projective cover of D^λ. Thus the projective indecomposable FS_r-modules are the P^λ with $\lambda \in \Lambda_p^+(r)$. If F' is an extension field of F then (using subscripts to show coefficient fields) we have $D^\lambda_{F'} \cong F' \otimes_F D^\lambda_F$ and $P^\lambda_{F'} \cong F' \otimes_F P^\lambda_F$.

We recall a few general facts about blocks. If Γ is a finite-dimensional F-algebra we may write Γ uniquely as a finite direct sum of indecomposable two-sided ideals, $\Gamma = \bigoplus_{B \in \mathscr{B}} \Gamma_B$. These ideals are the blocks of Γ, but it is convenient also to refer to the labels B as the blocks. The identity element of Γ may be written as $\sum_{B \in \mathscr{B}} e_B$, with $e_B \in \Gamma_B$ for all B. The elements e_B are the block idempotents: they are primitive central idempotents of Γ (see, for example, [Benson 1995]). Any Γ-module V satisfying $Ve_B = V$ is said to belong to B, and we write $V \in B$. Every Γ-module V may be written uniquely in the form $V = \bigoplus_{B \in \mathscr{B}} V_B$, where $V_B \in B$ for all B (indeed, $V_B = Ve_B$). We call V_B the block component of V corresponding to B.

By the Nakayama conjecture (see [James and Kerber 1981, 6.1.21]), the blocks of FS_r may be labelled $B(\nu)$ with $\nu \in \mathscr{C}_r$ in such a way that $S^\lambda \in B(\tilde{\lambda})$ for all $\lambda \in \Lambda_p^+(r)$. Since D^λ is a quotient of S^λ and P^λ is indecomposable with D^λ as a quotient, we have $D^\lambda, P^\lambda \in B(\tilde{\lambda})$. We use the same notation for the blocks of FS_r for every field F of characteristic p. By consideration of composition factors we see that, for any FS_r-module V, any extension field F' of F, and any ν, we have

$$(F' \otimes_F V)_{B(\nu)} \cong F' \otimes_F V_{B(\nu)}. \tag{2-5}$$

If B is a block and $B = B(\nu)$ we write $\tilde{B} = \nu$. Also, for $\lambda \in \Lambda_p^+(r)$, we write $\lambda \in B$ if $D^\lambda \in B$ (or equivalently $P^\lambda \in B$). The *principal block* is the block B_0 containing the trivial irreducible $D^{(r)}$. Thus $\tilde{B}_0 = (\bar{r})$, where $\bar{r}$ denotes the remainder on dividing r by p. If $\tilde{B} = \varnothing$ then $p \mid r$ and $\bar{r} = 0$ so that $B = B_0$. Hence if B is nonprincipal we have $\tilde{B} \neq \varnothing$.

If $p \nmid r$ then $\mathrm{Lie}(r)$ is projective (see Section 1). But if $p \mid r$ and $\tilde{B} \neq \varnothing$ then $B \neq B_0$ and so $\mathrm{Lie}(r)_B$ is projective by [Erdmann and Tan 2011, Theorem 3.1]. Hence we have the following result.

Theorem 2.1 [Erdmann and Tan 2011]. *If B is a block of FS_r such that $\tilde{B} \neq \varnothing$ then $\mathrm{Lie}(r)_B$ is projective.*

As is well known, Brauer characters of FS_r-modules have integer values: this follows, for example, from [Nagao and Tsushima 1989, Chapter 3, Lemma 6.13]. (Consequently Brauer characters of FS_r-modules are uniquely defined and do not depend upon choices of roots of unity.) We regard Brauer characters as maps from S_r to $\mathbb{Z}$ by assigning the value zero to p-singular elements of S_r. For each $\lambda \in \Lambda_p^+(r)$ we write β^λ and ζ^λ for the Brauer characters of D^λ and P^λ, respectively. By the orthogonality relations for Brauer characters (see [Nagao and Tsushima 1989, Chapter 3, Theorem 6.10]) we have

$$(\beta^\lambda, \zeta^\rho)_{S_r} = \begin{cases} 1 & \text{if } \lambda = \rho, \\ 0 & \text{if } \lambda \neq \rho, \end{cases} \tag{2-6}$$

where $(\beta^\lambda, \zeta^\rho)_{S_r}$ is defined as in (2-3).

2C. ***Polynomial representations of*** $\mathbf{GL}_n(F)$**.** Suppose now that F is an infinite field of prime characteristic p and let n and r be positive integers. We refer to [Green 1980] and [Donkin and Erdmann 1998] for background concerning polynomial $\mathrm{GL}_n(F)$-modules and the Schur algebra $S_F(n,r)$. Let E denote the natural $\mathrm{GL}_n(F)$-module. Thus $E^{\otimes r}$ is an $S_F(n,r)$-module. If k and t are positive integers such that $r = kt$ and if V is an $S_F(n,t)$-module then $V^{\otimes k}$ and $L^k(V)$ are $S_F(n,r)$-modules.

Let $\Lambda^+(n,r)$ denote the set of all partitions of r with at most n parts and let $\Lambda_p^+(n,r)$ denote the set of all p-regular partitions in $\Lambda^+(n,r)$. The irreducible $S_F(n,r)$-modules may be labelled $L(\lambda)$ with $\lambda \in \Lambda^+(n,r)$. For each such λ there is also an indecomposable $S_F(n,r)$-module $T(\lambda)$ called a "tilting module", and (see [Donkin and Erdmann 1998, Section 1.3]) there are nonnegative integers n_λ such that

$$E^{\otimes r} \cong \bigoplus_{\lambda \in \Lambda_p^+(n,r)} n_\lambda T(\lambda). \tag{2-7}$$

The main facts about the blocks of $S_F(n,r)$ were obtained in [Donkin 1994] and summarised in [Erdmann and Tan 2011]. When $n \geqslant r$ the blocks may be labelled $B(\nu)$ with $\nu \in \mathscr{C}_r$ in such a way that $L(\lambda) \in B(\tilde{\lambda})$. If B is a block and $B = B(\nu)$ we write $\tilde{B} = \nu$. When $n < r$, p-cores do not necessarily label unique blocks, but if $L(\lambda)$ and $L(\rho)$ are in the same block then $\tilde{\lambda} = \tilde{\rho}$. Thus, for each block B, there is an element $\tilde{B}$ of $\mathscr{C}_r$ (where $\tilde{B}$ has at most n parts) with the property that $\tilde{\lambda} = \tilde{B}$ whenever $L(\lambda) \in B$. For each $\nu \in \mathscr{C}_r$ we write $B(\nu)$ for the set of blocks B such that $\tilde{B} = \nu$. (Thus $B(\nu)$ is empty if ν has more than n parts.) If V is an $S_F(n,r)$-module we write $V_{B(\nu)}$ for the direct sum of the block components V_B of V corresponding to blocks B in $B(\nu)$. For all n and all $\lambda \in \Lambda^+(n,r)$, $T(\lambda)$ is indecomposable and has $L(\lambda)$ as a composition factor (see [Erdmann 1994, Section 1.3]); thus $T(\lambda)$ and $L(\lambda)$ belong to the same block. For a block B and $\lambda \in \Lambda^+(n,r)$ we write $\lambda \in B$ if $L(\lambda) \in B$ (or equivalently $T(\lambda) \in B$). We define the principal block to be the block B_0 containing $L(\lambda)$ where $\lambda = (r)$. Thus $\tilde{B}_0 = (\bar{r})$, with $\bar{r}$ as before. As in the case of FS_r, if $n \geqslant r$ and B is nonprincipal then $\tilde{B} \neq \varnothing$.

Let $\mathscr{T}$ denote the class of all $S_F(n,r)$-modules that are isomorphic to direct sums of tilting modules $T(\lambda)$ where $\lambda \in \Lambda_p^+(n,r)$. Thus $E^{\otimes r} \in \mathscr{T}$ by (2-7). If $p \nmid r$ then $L^r(E)$ is isomorphic to a direct summand of $E^{\otimes r}$ (see Section 1) and so $L^r(E) \in \mathscr{T}$. But if $p \mid r$ and $\tilde{B} \neq \varnothing$ then $L^r(E)_B \in \mathscr{T}$ by [Erdmann and Tan 2011, Theorem 3.2]. Hence we have the following result.

Theorem 2.2 [Erdmann and Tan 2011]. *If B is a block of $S_F(n,r)$ such that $\tilde{B} \neq \varnothing$ then $L^r(E)_B \in \mathscr{T}$.*

Suppose now that n_1 and n_2 are positive integers with $n_1 \geqslant n_2$ and let d_{n_1,n_2} denote the functor from the category of $S_F(n_1,r)$-modules to the category of $S_F(n_2,r)$-modules described in [Green 1980, Section 6.5]. This functor is exact (in particular it preserves direct sums) and we call it *truncation*. Note that $\Lambda^+(n_2,r) \subseteq \Lambda^+(n_1,r)$. We temporarily use subscripts to distinguish between modules for $S_F(n_1,r)$ and $S_F(n_2,r)$. Then, if $\lambda \in \Lambda^+(n_1,r)$ and $M(\lambda)$ denotes either $L(\lambda)$ or $T(\lambda)$, we have

$$d_{n_1,n_2}(M_{n_1}(\lambda)) \cong \begin{cases} M_{n_2}(\lambda) & \text{if } \lambda \in \Lambda^+(n_2,r), \\ 0 & \text{otherwise.} \end{cases} \tag{2-8}$$

(For the case of $L(\lambda)$ see [Green 1980, Section 6.5] and for $T(\lambda)$ see [Erdmann 1994, Section 1.7].)

Write $d = d_{n_1,n_2}$ and use the same notation for arbitrary r. Then, if k and t are positive integers and V is an $S_F(n_1,t)$-module, it is easy to check that $d(V^{\otimes k}) \cong d(V)^{\otimes k}$ and $d(L^k(V)) \cong L^k(d(V))$. Furthermore $d(E_{n_1}^{\otimes t}) \cong E_{n_2}^{\otimes t}$. Also, if V is an $S_F(n_1,r)$-module and $\nu \in \mathscr{C}_r$, it follows from (2-8) that

$$d(V_{B(\nu)}) \cong d(V)_{B(\nu)}. \tag{2-9}$$

2D. *The Schur functor.* We continue with the notation of the previous subsection but now assume that $n \geqslant r$. The Schur functor f_r is an exact functor from the category of $S_F(n,r)$-modules to the category of FS_r-modules (see [Green 1980, Chapter 6]). If U is an $S_F(n,r)$-module then $f_r(U)$ may be thought of as the weight space of U corresponding to the weight $(1,\dots,1,0,\dots,0)$, with r coordinates equal to 1, and the action of S_r on $f_r(U)$ comes by taking S_r as a group of permutation matrices in $\mathrm{GL}_n(F)$ (see, for example, [Donkin and Erdmann 1998, Section 1.2]). It is easily seen that

$$f_r(E^{\otimes r}) \cong FS_r. \tag{2-10}$$

Let $\{e_1,\dots,e_n\}$ be the standard basis of E. Then $f_r(L^r(E))$ is the subspace of $L^r(E)$ spanned by the left-normed Lie products $[e_{1\alpha},\dots,e_{r\alpha}]$ with $\alpha\in S_r$. In the notation of Section 2A, $f_r(L^r(E))\cong L^0_{r,F}$. Thus, since $L^0_{r,F}\cong \mathrm{Lie}(r)$, we obtain

$$f_r(L^r(E)) \cong \mathrm{Lie}(r). \tag{2-11}$$

For all $\lambda\in\Lambda^+_p(n,r)=\Lambda^+_p(r)$, we have (see [Donkin and Erdmann 1998, Section 1.3])

$$f_r(T(\lambda)) \cong P^\lambda. \tag{2-12}$$

As observed in [Erdmann and Tan 2011], f_r sends modules in the $S_F(n,r)$-block $B(\nu)$ to modules in the FS_r-block $B(\nu)$ labelled by the same p-core. Thus, if V is any $S_F(n,r)$-module, we have

$$f_r(V_{B(\nu)}) \cong f_r(V)_{B(\nu)}. \tag{2-13}$$

Let k be a divisor of r, and write $t=r/k$. (We do not at present assume that $p\nmid k$.) For each $\alpha\in S_k$ we may define $\alpha^*\in S_r$ by $((i-1)t+j)\alpha^*=(i\alpha-1)t+j$ for $i=1,\dots,k$ and $j=1,\dots,t$. The set $\{\alpha^*:\alpha\in S_k\}$ is a subgroup S^*_k of S_r isomorphic to S_k. The subgroup of S_r consisting of all permutations fixing $\{(i-1)t+j : i=1,\dots,k\}$ setwise for $j=1,\dots,t$ is a Young subgroup of S_r isomorphic to $S_k\times\cdots\times S_k$, and S^*_k may be thought of as the diagonal of this subgroup. The diagonal of any other Young subgroup isomorphic to $S_k\times\cdots\times S_k$ is a conjugate of S^*_k in S_r. Note that if σ is the r-cycle $(1\ 2\ \dots\ r)$ of S_r and σ_k is the k-cycle $(1\ 2\ \dots\ k)$ of S_k then $\sigma^t=\sigma^*_k\in S^*_k$. For $i=1,\dots,k$, write $\Omega_i=\{(i-1)t+j : j=1,\dots,t\}$. The subgroup $S^{(k)}_t$ of S_r consisting of all permutations fixing each Ω_i setwise is a Young subgroup isomorphic to $S_t\times\cdots\times S_t$. For each $\alpha\in S_k$ we have $\Omega_i\alpha^*=\Omega_{i\alpha}$ for $i=1,\dots,k$. The subgroup $S^{(k)}_t S^*_k$ of S_r is isomorphic to the wreath product $S_t \mathbin{\mathrm{wr}} S_k$.

Let V be an $S_F(n,t)$-module. Then $f_t(V)$ is an FS_t-module, so $f_t(V)^{\otimes k}$ is an $FS^{(k)}_t$-module. Indeed, $f_t(V)^{\otimes k}$ may be regarded as an $FS^{(k)}_t S^*_k$-module, where the action of S^*_k is to permute the tensor factors. We regard $\mathrm{Lie}(k)$ as an FS^*_k-module by means of the isomorphism $\alpha\mapsto\alpha^*$ from S_k to S^*_k. Then $\mathrm{Lie}(k)$ may also be regarded as an $FS^{(k)}_t S^*_k$-module, by taking trivial action of $S^{(k)}_t$. The following result is part of [Lim and Tan 2012, Corollary 3.2].

Lemma 2.3 [Lim and Tan 2012]. *In the above notation,*

$$f_r(L^k(V)) \cong (f_t(V)^{\otimes k} \otimes \mathrm{Lie}(k))\uparrow_{S_t^{(k)} S_k^*}^{S_r}.$$

Corollary 2.4. *In the above notation,*

$$f_r(L^k(E^{\otimes t})) \cong \mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r}.$$

Proof. By (2-10) and Lemma 2.3,

$$f_r(L^k(E^{\otimes t})) \cong ((FS_t)^{\otimes k} \otimes \mathrm{Lie}(k))\uparrow_{S_t^{(k)} S_k^*}^{S_r}.$$

Clearly $(FS_t)^{\otimes k}$ is a transitive permutation module under the action of $S_t^{(k)} S_k^*$ and the stabiliser of the basis element $1 \otimes \cdots \otimes 1$ is S_k^*. Thus $(FS_t)^{\otimes k}$ is induced from a one-dimensional trivial module for S_k^* and (by [Benson 1995, Proposition 3.3.3(i)]) we have

$$(FS_t)^{\otimes k} \otimes \mathrm{Lie}(k) \cong \mathrm{Lie}(k)\uparrow_{S_k^*}^{S_t^{(k)} S_k^*}.$$

The result follows. □

3. Main results

Recall from Section 2B that if B is a nonprincipal block of FS_r then $\tilde{B} \neq \varnothing$. Our main result on the Lie module is as follows. We use the notation of Section 2D, regarding $\mathrm{Lie}(k)$ as an FS_k^*-module.

Theorem 3.1. *Let F be a field of prime characteristic p. Let r be a positive integer and write $r = p^m k$ where $m \geqslant 0$, $k \geqslant 1$, and $p \nmid k$. Let B be a block of FS_r such that $\tilde{B} \neq \varnothing$ and let S_k^* be the diagonal of a Young subgroup $S_k \times \cdots \times S_k$ of S_r. Then*

$$\mathrm{Lie}(r)_B \cong \frac{1}{p^m}\,(\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r})_B.$$

Note that $\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r}$ is projective since $\mathrm{Lie}(k)$ is projective (see Section 1).

We commence the proof of Theorem 3.1. If F' is an extension field of F then, by the description of the Lie module in Section 2A, $\mathrm{Lie}_{F'}(r) \cong F' \otimes \mathrm{Lie}_F(r)$ and $\mathrm{Lie}_{F'}(k) \cong F' \otimes \mathrm{Lie}_F(k)$. Thus, if B is any block of FS_r, we have $\mathrm{Lie}_{F'}(r)_B \cong F' \otimes \mathrm{Lie}_F(r)_B$ and $(\mathrm{Lie}_{F'}(k)\uparrow_{S_k^*}^{S_r})_B \cong F' \otimes (\mathrm{Lie}_F(k)\uparrow_{S_k^*}^{S_r})_B$ by (2-5). Hence it suffices to prove Theorem 3.1 for the prime field $\mathbb{F}_p$ and then, by the Noether–Deuring theorem, it suffices to prove the theorem for any chosen field F of characteristic p. We choose F so that there is a p-modular system (K, R, F) with the properties specified in [Nagao and Tsushima 1989, Chapter 3, Section 6]. Note, in particular, that K has characteristic 0 and contains sufficient roots of unity, K is the field of fractions of R, and $F = R/(\pi)$ where (π) is the maximal ideal of R.

We state some standard facts associated with p-modular systems in order to establish terminology and notation. If G is any finite group then the natural epimorphism $R \to F$ yields an epimorphism $RG \to FG$. If this epimorphism maps u to v, where $u \in RG$ and $v \in FG$, we say that v *lifts* to u. By an *RG-lattice* we mean an RG-module that is free of finite rank as an R-module. If M is an RG-lattice we write $\overline{M} = M/\pi M$. Thus $\overline{M} \cong F \otimes_R M$ and $\overline{M}$ has the structure of an FG-module. An FG-module V is said to be *liftable* if there exists M such that $\overline{M} \cong V$, in which case we say that V lifts to M. If M is an RG-lattice then $K \otimes_R M$ is a KG-module. If U is any KG-module then there is an RG-lattice M such that $U \cong K \otimes_R M$ (see [Benson 1995, Lemma 1.9.1]) and we say that $\overline{M}$ is obtained from U by *modular reduction*.

By a standard result (see [ibid., Theorem 1.9.4]), each block idempotent e_B of FS_r can be lifted to an element $\widehat{e}_B$ of RS_r to obtain pairwise-orthogonal primitive central idempotents of RS_r summing to the identity. If M is an RS_r-lattice such that $M\widehat{e}_B = M$ we write $M \in B$. Every RS_r-lattice M may be written uniquely in the form $M = \bigoplus_B M_B$ where, for each B, M_B is an RS_r-lattice belonging to B. Similar facts and notation apply to KS_r-modules, using the same idempotents $\widehat{e}_B$. If U is a KS_r-module then, since K has characteristic zero, U_B is a direct sum of irreducible KS_r-modules belonging to B.

It is easily verified that if M is an RS_r-lattice and B is a block then

$$(K \otimes_R M)_B \cong K \otimes_R M_B \quad \text{and} \quad \overline{M_B} \cong \overline{M}_B. \tag{3-1}$$

We let σ be an r-cycle of S_r chosen as in Section 2D (with $t = p^m$) so that $\sigma^{p^m} = \sigma_k^* \in S_k^*$, where σ_k is a k-cycle of S_k.

Lemma 3.2. *If g is an element of the cyclic subgroup $\langle \sigma \rangle$ such that g has order divisible by p and if χ is the character of an irreducible KS_r-module U belonging to a block B such that $\tilde{B} \neq \varnothing$ then $\chi(g) = 0$.*

Proof. Let M be an RS_r-lattice such that $U \cong K \otimes_R M$. Since U belongs to B it follows from (3-1) that M belongs to B. Let D be the defect group of the FS_r-block B (see [Benson 1995, Section 6.1]). (Thus D is a p-group, determined up to conjugacy in S_r.) By [ibid., Corollary 6.1.3], D is also the defect group of B regarded as a block of RS_r. Thus, by [ibid., Proposition 6.1.2], M is projective relative to D.

Let $B = B(\nu)$ where $\nu \in \mathscr{C}_r$. Thus $\nu \neq \varnothing$ and so ν is a partition of r' for some r' satisfying $0 < r' \leqslant r$. It follows from [James and Kerber 1981, 6.2.45] that D can be taken to be a Sylow p-subgroup of a subgroup $S_{r-r'}$ of S_r fixing r' points of $\{1, \ldots, r\}$. Hence every element of D fixes some point of $\{1, \ldots, r\}$.

Let g be as in the statement of the lemma. The p-part of g is a nontrivial element of $\langle \sigma \rangle$ and hence has no fixed points in $\{1, \ldots, r\}$. It follows that the p-part of g is not conjugate in S_r to an element of D. Therefore, by [Nagao and Tsushima 1989, Chapter 4, Theorem 7.4], we have $\chi(g) = 0$, as required. $\square$

Lemma 3.3. *If B is a block such that $\tilde{B} \neq \varnothing$ then $p^m \operatorname{Lie}_K(r)_B \cong (\operatorname{Lie}_K(k)\uparrow_{S_k^*}^{S_r})_B$.*

Proof. The result is trivial if $r = k$. Thus we may assume that $p \mid r$. Let ψ_r denote the character of the KS_r-module $\operatorname{Lie}_K(r)$ and let ψ_k denote the character of the KS_k^*-module $\operatorname{Lie}_K(k)$. In order to prove the lemma it suffices to show that the multiplicity of each irreducible KS_r-module U belonging to B is the same in $p^m \operatorname{Lie}_K(r)$ as in $\operatorname{Lie}_K(k)\uparrow_{S_k^*}^{S_r}$. Let χ be the character of U. By the orthogonality relations and Frobenius reciprocity for ordinary characters, it suffices to prove

$$p^m\,(\chi, \psi_r)_{S_r} = (\chi\downarrow_{S_k^*}^{S_r}, \psi_k)_{S_k^*}. \tag{3-2}$$

By (2-4) we have

$$\begin{aligned} r\,(\chi, \psi_r)_{S_r} &= \sum_{d \mid r} \mu(d)\chi(\sigma^{r/d}) \\ &= \sum_{d \mid k} \mu(d)\chi(\sigma^{r/d}) - \sum_{d \mid k} \mu(d)\chi(\sigma^{r/pd}). \end{aligned}$$

However, for $d \mid k$, we have $\chi(\sigma^{r/pd}) = 0$ by Lemma 3.2. Thus

$$r\,(\chi, \psi_r)_{S_r} = \sum_{d \mid k} \mu(d)\chi((\sigma^{p^m})^{k/d}).$$

Recall that $\sigma^{p^m} = \sigma_k^* \in S_k^*$ where σ_k is a k-cycle of S_k. Hence, by (2-4) applied to S_k^*,

$$k\,(\chi\downarrow_{S_k^*}^{S_r}, \psi_k)_{S_k^*} = \sum_{d \mid k} \mu(d)\chi((\sigma^{p^m})^{k/d}).$$

This gives (3-2). $\square$

We can now prove Theorem 3.1. Let B be a block of FS_r such that $\tilde{B} \neq \varnothing$. By the description of the Lie module in Section 2A, $\operatorname{Lie}(r)$ lifts to the RS_r-lattice $\operatorname{Lie}_R(r)$ and $\operatorname{Lie}(k)$ lifts to the RS_k^*-lattice $\operatorname{Lie}_R(k)$. Thus $p^m \operatorname{Lie}(r)$ and $\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r}$ lift to $p^m \operatorname{Lie}_R(r)$ and $\operatorname{Lie}_R(k)\uparrow_{S_k^*}^{S_r}$, respectively. Also, $K \otimes p^m \operatorname{Lie}_R(r) \cong p^m \operatorname{Lie}_K(r)$ and $K \otimes \operatorname{Lie}_R(k)\uparrow_{S_k^*}^{S_r} \cong \operatorname{Lie}_K(k)\uparrow_{S_k^*}^{S_r}$. Hence $p^m \operatorname{Lie}(r)$ and $\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r}$ are modular reductions of $p^m \operatorname{Lie}_K(r)$ and $\operatorname{Lie}_K(k)\uparrow_{S_k^*}^{S_r}$, respectively. It follows by (3-1) that $p^m \operatorname{Lie}(r)_B$ and $(\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r})_B$ are modular reductions of $p^m \operatorname{Lie}_K(r)_B$ and $(\operatorname{Lie}_K(k)\uparrow_{S_k^*}^{S_r})_B$, respectively. However, by Lemma 3.3, these two last modules are isomorphic. Therefore, by [Nagao and Tsushima 1989, Chapter 3, Lemma 6.4], $p^m \operatorname{Lie}(r)_B$ and $(\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r})_B$ have the same Brauer character.

By Theorem 2.1, $\operatorname{Lie}(r)_B$ is projective. Since $\operatorname{Lie}(k)$ is a projective FS_k^*-module, $\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r}$ is a projective FS_r-module and so $(\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r})_B$ is projective. Thus

$p^m \operatorname{Lie}(r)_B$ and $(\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r})_B$ are projective modules with the same Brauer characters. Therefore, by [Benson 1995, Corollary 5.3.6], these modules are isomorphic. This proves Theorem 3.1.

Corollary 3.4. *If B_0 is the principal block of FS_r then*

$$\dim \operatorname{Lie}(r)_{B_0} = \frac{1}{p^m} \dim\,(\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r})_{B_0}\,.$$

Proof. For each nonprincipal block B of FS_r we have

$$\dim \operatorname{Lie}(r)_{B} = \frac{1}{p^m} \dim\,(\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r})_{B}\,,$$

by Theorem 3.1. However, by (2-1),

$$\dim \operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r} = (k-1)!\,r!/k! = p^m(r-1)! = p^m \dim \operatorname{Lie}(r).$$

The result follows. □

Theorem 3.5. *In the notation of Theorem 3.1, we have*

$$\operatorname{Lie}(r)_B \cong \bigoplus_{\substack{\lambda\in\Lambda_p^+(r)\\ \lambda\in B}} m_\lambda\, P^\lambda, \tag{3-3}$$

where, for each λ,

$$m_\lambda = \frac{1}{r}\sum_{d\mid k} \mu(d)\,\beta^\lambda(\tau^{k/d}), \tag{3-4}$$

where τ is an element of S_r of cycle type $(k, k, \dots, k)$ and β^λ denotes the Brauer character of D^λ.

Proof. By Theorem 2.1, $\operatorname{Lie}(r)_B$ is projective. Thus it satisfies (3-3) for suitable nonnegative integers m_λ. It remains to prove (3-4). If F' is an extension field of F then $\operatorname{Lie}_{F'}(r) \cong F' \otimes \operatorname{Lie}_F(r)$ and $P^\lambda_{F'} \cong F' \otimes P^\lambda_F$. Also, block components are preserved under field extensions, by (2-5). Hence it suffices to prove the result for the field $\mathbb{F}_p$ and then, by a similar argument, it suffices to prove the result for any chosen field F of characteristic p. We take F from the p-modular system (K, R, F) used in the proof of Theorem 3.1.

Since $\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r}$ is projective we have

$$\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r} \cong \bigoplus_{\rho\in\Lambda_p^+(r)} m'_\rho\, P^\rho$$

for suitable nonnegative integers m'_ρ. Let $\lambda \in \Lambda_p^+(r)$ where $\lambda \in B$. By Theorem 3.1 we have $m_\lambda = (1/p^m)m'_\lambda$. Let ϕ denote the Brauer character of $\operatorname{Lie}(k)\uparrow_{S_k^*}^{S_r}$. By the orthogonality relation (2-6) we have

$$m_\lambda = \frac{1}{p^m} m'_\lambda = \frac{1}{p^m} (\beta^\lambda, \phi)_{S_r}.$$

As observed in the proof of Theorem 3.1, $\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r}$ is a modular reduction of $\mathrm{Lie}_K(k)\uparrow_{S_k^*}^{S_r}$. The character of $\mathrm{Lie}_K(k)\uparrow_{S_k^*}^{S_r}$ is $\psi_k\uparrow_{S_k^*}^{S_r}$, where ψ_k denotes the character of $\mathrm{Lie}_K(k)$ as a KS_k^*-module. By [Nagao and Tsushima 1989, Chapter 3, Lemma 6.4], ϕ and $\psi_k\uparrow_{S_k^*}^{S_r}$ take the same value on p'-elements of S_r. Thus, by Frobenius reciprocity,

$$m_\lambda = \frac{1}{p^m} (\beta^\lambda, \psi_k\uparrow_{S_k^*}^{S_r})_{S_r} = \frac{1}{p^m} (\beta^\lambda\downarrow_{S_k^*}^{S_r}, \psi_k)_{S_k^*}.$$

Let τ be as in the statement of the theorem. Then τ is conjugate to, and therefore can be taken to be, an element σ_k^* of S_k^* corresponding to a k-cycle σ_k of S_k. Thus, by (2-4), we have

$$\frac{1}{p^m} (\beta^\lambda\downarrow_{S_k^*}^{S_r}, \psi_k)_{S_k^*} = \frac{1}{p^m k} \sum_{d \mid k} \mu(d)\, \beta^\lambda(\tau^{k/d}).$$

The result follows. □

We now turn to Lie powers and, for the rest of this section, we assume that F is infinite. As before, let n be a positive integer and let E be the natural $\mathrm{GL}_n(F)$-module.

Theorem 3.6. *Let F be an infinite field of prime characteristic p. Let r be a positive integer and write $r = p^m k$ where $m \geqslant 0$, $k \geqslant 1$, and $p \nmid k$. Let B be a block of $S_F(n, r)$ such that $\tilde{B} \neq \varnothing$. Then*

$$L^r(E)_B \cong \frac{1}{p^m}\, L^k(E^{\otimes p^m})_B.$$

Proof. Let $\mathcal{T}$ be as defined in Section 2C. Thus, by Theorem 2.2, $L^r(E)_B \in \mathcal{T}$. Also, since $L^k(E^{\otimes p^m})$ is a direct summand of $E^{\otimes r}$, we have $L^k(E^{\otimes p^m}) \in \mathcal{T}$, by (2-7).

Suppose first that $n \geqslant r$. Then we may write $B = B(\nu)$ where $\nu \neq \varnothing$. By (2-11) and (2-13),

$$f_r(p^m\, L^r(E)_{B(\nu)}) \cong p^m\, \mathrm{Lie}(r)_{B(\nu)}.$$

Similarly, by Corollary 2.4 and (2-13),

$$f_r(L^k(E^{\otimes p^m})_{B(\nu)}) \cong (\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r})_{B(\nu)}.$$

Also, by Theorem 3.1, $p^m\, \mathrm{Lie}(r)_{B(\nu)} \cong (\mathrm{Lie}(k)\uparrow_{S_k^*}^{S_r})_{B(\nu)}$. It follows from (2-12) that if $U, V \in \mathcal{T}$ and $f_r(U) \cong f_r(V)$ then $U \cong V$. Hence the isomorphism in Theorem 3.6 holds when $n \geqslant r$.

Now suppose that $n < r$ and let $\tilde{B} = \nu$. Thus $B \in B(\nu)$. Consider the $S_F(r,r)$-block $B(\nu)$. By the first case, there is an isomorphism of $S_F(r,r)$-modules,

$$L^r(E)_{B(\nu)} \cong \frac{1}{p^m} L^k(E^{\otimes p^m})_{B(\nu)}. \tag{3-5}$$

We apply truncation $d_{r,n}$ to (3-5). By (2-9) and the other properties of truncation given in Section 2C, we obtain (3-5) for $S_F(n,r)$-modules. Hence the corresponding block components are isomorphic for all $S_F(n,r)$-blocks in $B(\nu)$ and we obtain the isomorphism of Theorem 3.6. □

Theorem 3.7. *In the notation of Theorem 3.6, we have*

$$L^r(E)_B \cong \bigoplus_{\substack{\lambda \in \Lambda_p^+(n,r) \\ \lambda \in B}} m_\lambda\, T(\lambda),$$

where m_λ is given by (3-4).

Proof. By Theorem 2.2, $L^r(E)_B \in \mathcal{T}$. Thus $L^r(E)_B$ is isomorphic to a direct sum of tilting modules $T(\lambda)$ with $\lambda \in \Lambda_p^+(n,r)$ and $\lambda \in B$. Let $\tilde{B} = \nu$. Then, for $n \geqslant r$, we have $f_r(L^r(E)_{B(\nu)}) \cong \mathrm{Lie}(r)_{B(\nu)}$, by (2-11) and (2-13), and $f_r(T(\lambda)) \cong P^\lambda$ for all $\lambda \in \Lambda_p^+(n,r)$, by (2-12). Thus, for $n \geqslant r$, the result is given by Theorem 3.5. For $n < r$ the result follows by truncation, as in the proof of Theorem 3.6. (Note that the effect of truncation on $T(\lambda)$ is given by (2-8).) □

References

[Benson 1995] D. J. Benson, *Representations and cohomology, vol. 1: Basic representation theory of finite groups and associative algebras*, Cambridge Studies in Advanced Mathematics **30**, Cambridge University Press, 1995. MR 1644252 Zbl 0908.20001

[Bourbaki 1972] N. Bourbaki, *Groupes et algèbres de Lie: Chapitre II: Algèbres de Lie libres; Chapitre III: Groupes de Lie*, Actualités Scientifiques et Industrielles **1349**, Hermann, Paris, 1972. MR 58 #28083a Zbl 0244.22007

[Bryant 2009] R. M. Bryant, "Lie powers of infinite-dimensional modules", *Beiträge Algebra Geom.* **50**:1 (2009), 179–193. MR 2010f:20004 Zbl 1185.17003

[Donkin 1994] S. Donkin, "On Schur algebras and related algebras, IV: The blocks of the Schur algebras", *J. Algebra* **168**:2 (1994), 400–429. MR 95j:20037 Zbl 0832.20013

[Donkin and Erdmann 1998] S. Donkin and K. Erdmann, "Tilting modules, symmetric functions, and the module structure of the free Lie algebra", *J. Algebra* **203**:1 (1998), 69–90. MR 99e:20056 Zbl 0984.20029

[Erdmann 1994] K. Erdmann, "Symmetric groups and quasi-hereditary algebras", pp. 123–161 in *Finite-dimensional algebras and related topics* (Ottawa, 1992), edited by V. Dlab and L. L. Scott, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **424**, Kluwer, Dordrecht, 1994. MR 95k:20012 Zbl 0831.20013

[Erdmann and Tan 2011] K. Erdmann and K. M. Tan, "The non-projective part of the Lie module for the symmetric group", *Arch. Math.* (*Basel*) **96**:6 (2011), 513–518. MR 2012h:20025 Zbl 1234.20013

[Green 1980] J. A. Green, *Polynomial representations of* GL_n, Lecture Notes in Mathematics **830**, Springer-Verlag, Berlin, 1980. MR 606556 (83j:20003) Zbl 0451.20037

[James and Kerber 1981] G. James and A. Kerber, *The representation theory of the symmetric group*, Encyclopedia of Mathematics and its Applications **16**, Addison-Wesley, Reading, MA, 1981. MR 83k:20003 Zbl 0491.20010

[Lim and Tan 2012] K. J. Lim and K. M. Tan, "The Schur functor on tensor powers", *Arch. Math.* (*Basel*) **98**:2 (2012), 99–104. Zbl 06018039

[Nagao and Tsushima 1989] H. Nagao and Y. Tsushima, *Representations of finite groups*, Academic Press, Boston, MA, 1989. MR 90h:20008 Zbl 0673.20002

Communicated by David Benson
Received 2011-03-10 Revised 2011-06-08 Accepted 2011-07-06

roger.bryant@manchester.ac.uk
School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

erdmann@maths.ox.ac.uk *Mathematical Institute, University of Oxford, 24-29 St. Giles, Oxford, OX1 3LB, United Kingdom*
http://www.maths.ox.ac.uk/~erdmann

ALGEBRA AND NUMBER THEORY 6:4(2012)

Basepoint-free theorems: saturation, b-divisors, and canonical bundle formula

Osamu Fujino

We reformulate basepoint-free theorems using notions introduced by Shokurov, such as b-divisors and saturation of linear systems. Our formulation is flexible and has some important applications. One of the main purposes of this paper is to prove a generalization of the basepoint-free theorem in Fukuda's paper "On numerically effective log canonical divisors".

1. Introduction

In this paper, we reformulate basepoint-free theorems by using Shokurov's ideas [2003] of *b-divisors* and *saturation of linear systems*. Combining the refined Kawamata–Shokurov basepoint-free theorem (quoted here as Theorem 2.1) or its generalization (Theorem 6.1) with Ambro's formulation of Kodaira's canonical bundle formula, we obtain new basepoint-free theorems (Theorems 4.4 and 6.2), which are flexible and have some important applications (Theorem 7.11). One of the main purposes of this paper is to prove the following generalization of the basepoint-free theorem given in [Fukuda 2002, Proposition 3.3]:

Theorem 1.1. *Let (X, B) be an lc pair and let $\pi : X \to S$ be a proper morphism onto a variety S. Assume the following conditions*:

MSC2010: primary 14C20; secondary 14N30, 14E30.

Keywords: basepoint-free theorem, canonical bundle formula, b-divisor, saturation.

(A) *H is a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X.*

(B) *$H-(K_X+B)$ is π-nef and π-abundant.*

(C) *$\kappa(X_\eta, (aH-(K_X+B))_\eta) \geq 0$ and*

$$\nu(X_\eta, (aH-(K_X+B))_\eta) = \nu(X_\eta, (H-(K_X+B))_\eta)$$

for some $a \in \mathbb{Q}$ with $a > 1$, where η is the generic point of S.

(D) *There is a positive integer c such that cH is Cartier and*

$$\mathcal{O}_T(cH) := \mathcal{O}_X(cH)|_T$$

is π-generated, where $T = \mathrm{Nklt}(X, B)$ is the non-klt locus of (X, B).

Then H is π-semiample.

This will be proved on page 816. As an application of Theorem 1.1, we have:

Theorem 1.2 [Fujino and Gongyo 2011, Theorem 4.12]. *Let $\pi : X \to S$ be a projective morphism between projective varieties. Let (X, B) be an lc pair such that $K_X + B$ is nef and log abundant over S. Then $K_X + B$ is f-semiample.*

We also used Theorem 1.1 to prove the finite generation of the log canonical ring for log canonical 4-folds in [Fujino 2010]; see Remark 3.4 of that paper. As we explain elsewhere [Fujino 2007b, Remark 3.10.3; 2011d, 5.1], the proof of Theorem 4.3 of [Kawamata 1985] contains a gap. Because of that gap, Theorem 5.1 of [Kawamata 1985] was also not rigorously proved, and since Proposition 3.3 of [Fukuda 2002] depends on it, our proof of Theorem 1.1 is the first rigorous proof of this important result of Fukuda.

Another purpose of this paper is to show how to use Shokurov's ideas, such as *b-divisors*, *saturation of linear systems*, various kinds of *adjunction*, and so on, by reproving some known results in our formulation. Thus one can regard this paper as Chapter $8\frac{1}{2}$ of the book [Corti et al. 2007]. It is also a complement of the paper [Fujino 2011d]. We do not use the powerful new method developed in [Ambro 2003; Fujino 2009a; 2009b; 2009c; 2011a; 2011b; 2011c]. For related topics and applications, see [Fujino 2010; Gongyo 2010, Section 6; Cacciola 2011; Fujino and Gongyo 2011].

Remark 1.3. Professor Yujiro Kawamata [2011a] has announced a correction to the error in the proof of [Kawamata 1985, Theorem 4.3]. The new proof seems to depend heavily on arguments in his preprints [2011b; 2010]. If we accept his correction, then Theorem 1.1 holds under the assumption that (X, B) is dlt and S is a point, by [Fukuda 2002, Proposition 3.3] (see Remark 6.7 (ii)). As stated in the introduction of [Kawamata 2011a], our arguments are simpler. We note that our approach is completely different from Kawamata's original one. Anyway, Theorem 1.1 plays a crucial role in our study of the log abundance conjecture for log canonical pairs; see [Fujino and Gongyo 2011, Section 4]. Therefore, this paper is very relevant for the minimal model program for log canonical pairs.

Let us explain the motivation for our formulation.

1.4. ***Motivation.*** Let (X, B) be a projective klt pair and let D be a nef Cartier divisor on X such that $D-(K_X+B)$ is nef and big. Then the Kawamata–Shokurov basepoint-free theorem means that $|mD|$ is free for every $m \gg 0$. Let $f : Y \to X$ be a projective birational morphism from a normal projective variety Y such that $K_Y + B_Y = f^*(K_X + B)$. We note that f^*D is a nef Cartier divisor on Y and that $f^*D-(K_Y+B_Y)$ is nef and big. It is obvious that $|mf^*D|$ is free for every $m \gg 0$ because $|mD|$ is free for every $m \gg 0$. In general, we cannot directly apply the Kawamata–Shokurov basepoint-free theorem to f^*D and (Y, B_Y). This is because (Y, B_Y) is sub-klt but is not always klt. Note that a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor L on X is nef, big, or semiample if and only if so is f^*L. However, the notion of klt is not stable under birational pull-backs. By adding a *saturation condition*, which is trivially satisfied for klt pairs, we can apply the Kawamata–Shokurov basepoint-free theorem for sub-klt pairs (see Theorem 2.1). By this new formulation, the basepoint-free theorem becomes more flexible and has some important applications.

1.5. ***Background.*** A key result we need is [Ambro 2004, Theorem 0.2], which is a generalization of [Fujino 2003, Section 4: Pull-back of $L^{ss}_{X/Y}$]. It originates from *Kawamata's positivity theorem* [1998] and Shokurov's idea of *adjunction*. For details, see [Ambro 2004, Introduction]. The formulation and calculation we borrow from [Ambro 2005b; 2007] grew out from Shokurov's *saturation of linear systems* [2003, 4.32].

1.6. ***Outline of the paper.*** In Section 2, we reformulate the Kawamata–Shokurov basepoint-free theorem for sub-klt pairs with a saturation condition. To state our theorem, we use the notion of b-divisors. It is very useful to discuss linear systems with some base conditions. In Section 3, we collect basic properties of b-divisors and prove some elementary properties. In Section 4, we discuss a slight generalization of the main theorem of [Kawamata 1985]. We need this generalization in Section 7. The main ingredient of our proof is Ambro's formulation of Kodaira's canonical bundle formula. By this formula and the refined Kawamata–Shokurov basepoint-free theorem obtained in Section 2, we can quickly prove Kawamata's theorem in [Kawamata 1985] and its generalization without appealing to the notion of generalized normal crossing varieties. In Section 5, we treat the basepoint-free theorem of Reid–Fukuda type. In this case, the saturation condition behaves very well for inductive arguments. It helps us understand the saturation condition of linear systems. In Section 6, we prove some variants of basepoint-free theorems, mainly due to Fukuda [2002]. We reformulate them by using b-divisors and saturation conditions. Then we use Ambro's canonical bundle formula to reduce them to the easier case instead of proving them directly by the X-method. In Section 7, we generalize the Kawamata–Shokurov basepoint-free theorem and Kawamata's main theorem in [Kawamata 1985] for *pseudo-klt pairs*. Theorem 7.11, which is new, is the main theorem of this section. It will be useful for the study of lc centers (Theorem 7.13).

Notation. Let $B = \sum b_i B_i$ be a $\mathbb{Q}$-divisor on a normal variety X such that B_i is prime for every i and that $B_i \neq B_j$ for $i \neq j$. We denote by

$$\lceil B \rceil = \sum \lceil b_i \rceil B_i, \quad \lfloor B \rfloor = \sum \lfloor b_i \rfloor B_i, \quad \text{and} \quad \{B\} = B - \lfloor B \rfloor$$

the *round-up*, the *round-down*, and the *fractional part* of B. Note that we do not use $\mathbb{R}$-divisors in this paper. We make one general remark here. Since the freeness (or semiampleness) of a Cartier divisor D on a variety X depends only on the linear equivalence class of D, we can freely replace D by a linearly equivalent divisor to prove the freeness (or semiampleness) of D.

We will work over an algebraically closed field k of characteristic zero throughout this paper.

2. Kawamata–Shokurov basepoint-free theorem revisited

Kawamata and Shokurov claimed the following theorem for klt pairs, that is, they assumed that B is effective, which implies that condition (2) is trivially satisfied. We think that our formulation is useful for some applications. Readers not familiar with the notion of b-divisors are referred to Section 3.

Theorem 2.1 (Basepoint-free theorem). *Let (X, B) be a sub-klt pair, let $\pi : X \to S$ be a proper surjective morphism onto a variety S and let D be a π-nef Cartier divisor on X. Assume the following conditions*:

(1) *$rD - (K_X + B)$ is nef and big over S for some positive integer r.*

(2) (Saturation condition.) *There exists a positive integer j_0 such that*

$$\pi_* \mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil + j\overline{D}) \subseteq \pi_* \mathcal{O}_X(jD)$$

for every integer $j \geq j_0$.

Then mD is π-generated for every $m \gg 0$, that is, there exists a positive integer m_0 such that for every $m \geq m_0$ the natural homomorphism $\pi^* \pi_* \mathcal{O}_X(mD) \to \mathcal{O}_X(mD)$ is surjective.

Proof. The usual proof of the basepoint-free theorem, that is, the X-method, works without any changes if we note Lemma 3.10. For the details, see, for example, [Kawamata et al. 1987, Section 3-1]. See also Remarks 3.14–3.17. □

The assumptions in Theorem 2.1 are birational in nature. This point is indispensable in Section 4. We note that we can assume that X is nonsingular and Supp B is a simple normal crossing divisor because conditions (1) and (2) are invariant for birational pull-backs. So, it is easy to see that Theorem 2.1 is equivalent to the following theorem.

Theorem 2.2. *Let X be a nonsingular variety and let B be a $\mathbb{Q}$-divisor on X such that $\lfloor B \rfloor \leq 0$ and* Supp B *is a simple normal crossing divisor. Let $\pi : X \to S$ be a projective morphism onto a variety S and let D be a π-nef Cartier divisor on X. Assume the following conditions*:

(1) *$rD - (K_X + B)$ is nef and big over S for some positive integer r.*

(2) (Saturation condition.) *There exists a positive integer j_0 such that*
$$\pi_* \mathcal{O}_X(\lceil -B \rceil + jD) \simeq \pi_* \mathcal{O}_X(jD)$$
for every integer $j \geq j_0$.

Then mD is π-generated for every $m \gg 0$.

The following example says that the original Kawamata–Shokurov basepoint-free theorem does not necessarily hold for *sub*-klt pairs.

Example 2.3. Let $X = E$ be an elliptic curve. We take a Cartier divisor H such that deg $H = 0$ and $lH \not\sim 0$ for every $l \in \mathbb{Z} \setminus \{0\}$. In particular, H is nef. We put $B = -P$, where P is a closed point of X. Then (X, B) is sub-klt and $H - (K_X + B)$ is ample. However, H is not semiample. In this case, $H^0(X, \mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil + j\overline{H})) \simeq H^0(X, \mathcal{O}_X(P + jH)) \simeq k$ for every j. However, $H^0(X, \mathcal{O}_X(jH)) = 0$ for all j. Therefore, the saturation condition in Theorem 2.1 does not hold.

We note that Kollár's effective basepoint-freeness holds under the same assumption as in Theorem 2.1.

Theorem 2.4 (Effective freeness). *We use the same notation and assumption as in Theorem 2.1. Then there exists a positive integer l, which depends only on* dim X *and* $\max\{r, j_0\}$, *such that lD is π-generated, that is, $\pi^* \pi_* \mathcal{O}_X(lD) \to \mathcal{O}_X(lD)$ is surjective.*

Sketch of the proof. We need no new ideas. So, we just explain how to modify the arguments in [Kollár 1993, Section 2]. From now on, we use the notation in [Kollár 1993]. In that reference, (X, Δ) is assumed to be klt, that is, (X, Δ) is sub-klt and Δ is effective. The effectivity of Δ implies that H' is f-exceptional in [ibid., (2.1.4.3)]. We need this to prove $H^0(Y, \mathcal{O}_Y(f^*N + H')) = H^0(X, \mathcal{O}_X(N))$ in [ibid., (2.1.6)].

It is not difficult to see that $0 \le H' \le \lceil \mathbf{A}(X, \Delta)_Y \rceil$ in our notation. Therefore, it is sufficient to assume the saturation condition Theorem 2.1(2) in the proof of Kollár's effective freeness (see [ibid., Section 2]). We make one more remark. Applying the argument in the first part of [ibid., 2.4] to $\mathcal{O}_X(j\overline{D}+\lceil \mathbf{A}(X, B)\rceil)$ on the generic fiber of $\pi : X \to S$ with the saturation condition (2) in Theorem 2.1, we obtain a positive integer l_0 that depends only on $\dim X$ and $\max\{r, j_0\}$ such that $\pi_*\mathcal{O}_X(l_0 D) \neq 0$. As explained above, the arguments in Section 2 in [ibid.] work with only minor modifications in our setting. We leave the details as an exercise for the reader. □

3. b-divisors

Let us recall the notion of singularities of pairs, referring the reader to [Fujino 2007b] for a more extended treatment.

Definition 3.1 (Singularities of pairs). Let X be a normal variety and let B be a $\mathbb{Q}$-divisor on X such that $K_X + B$ is $\mathbb{Q}$-Cartier. Let $f : Y \to X$ be a resolution of singularities such that $\mathrm{Exc}(f) \cup f_*^{-1}B$ has a simple normal crossing support, where $\mathrm{Exc}(f)$ is the exceptional locus of f. We write

$$K_Y = f^*(K_X + B) + \sum a_i A_i.$$

We note that a_i is called the *discrepancy* of A_i. Then the pair (X, B) is *sub-klt* (resp. *sub-lc*) if $a_i > -1$ (resp. $a_i \ge -1$) for every i. The pair (X, B) is *klt* (resp. *lc*) if (X, B) is sub-klt (resp. sub-lc) and B is effective. (In some literature, sub-klt and sub-lc are sometimes called klt and lc.)

Let (X, B) be an lc pair. If there exists a resolution $f : Y \to X$ such that $\mathrm{Exc}(f)$ and $\mathrm{Exc}(f) \cup f_*^{-1}B$ are simple normal crossing divisors on Y and

$$K_Y = f^*(K_X + B) + \sum a_i A_i$$

with $a_i > -1$ for all f-exceptional A_i's, then (X, B) is called *dlt*.

Remark 3.2. Let (X, B) be a klt (resp. lc) pair and let $f : Y \to X$ be a proper birational morphism of normal varieties. We put $K_Y + B_Y = f^*(K_X + B)$. Then (Y, B_Y) is not necessarily klt (resp. lc) but it is sub-klt (resp. sub-lc).

Let us recall the definition of *log canonical centers*.

Definition 3.3 (Log canonical center). Let (X, B) be a sub-lc pair. A subvariety $W \subset X$ is called a *log canonical center* or an *lc center* of (X, B) if there is a resolution $f : Y \to X$ such that $\mathrm{Exc}(f) \cup \mathrm{Supp}\, f_*^{-1}B$ is a simple normal crossing divisor on Y and a divisor E with discrepancy -1 such that $f(E) = W$. A log canonical center $W \subset X$ of (X, B) is called *exceptional* if there is a unique divisor E_W on Y with discrepancy -1 such that $f(E_W) = W$ and $f(E) \cap W = \varnothing$ for every other divisor $E \neq E_W$ on Y with discrepancy -1; see [Kollár 2007, 8.1].

3.4. ***b-divisors.*** The notion of *b-divisors*, introduced by Shokurov, plays a central role in this paper, and we now recall its definition. For details, we refer to [Ambro 2005b, 1-B] and [Corti 2007, 2.3.2]. The reader can find various examples of b-divisors in [Iskovskikh 2003].

Definition 3.5 (b-divisor). Let X be a normal variety and let $\mathrm{Div}(X)$ be the free abelian group generated by Weil divisors on X. A *b-divisor* on X is an element

$$\mathbf{D} \in \mathbf{Div}(X) = \mathrm{projlim}_{Y\to X}\, \mathrm{Div}(Y),$$

where the projective limit is taken over all proper birational morphisms $f : Y \to X$ of normal varieties, under the push forward homomorphism $f_* : \mathrm{Div}(Y) \to \mathrm{Div}(X)$. A $\mathbb{Q}$-*b-divisor* on X is an element of $\mathbf{Div}_{\mathbb{Q}}(X) = \mathbf{Div}(X) \otimes_{\mathbb{Z}} \mathbb{Q}$.

Definition 3.6 (Discrepancy $\mathbb{Q}$-b-divisor). Let X be a normal variety and let B be a $\mathbb{Q}$-divisor on X such that $K_X + B$ is $\mathbb{Q}$-Cartier. Then the *discrepancy* $\mathbb{Q}$*-b-divisor* of the pair (X, B) is the $\mathbb{Q}$-b-divisor $\mathbf{A} = \mathbf{A}(X, B)$ with the trace $\mathbf{A}_Y$ defined by the formula

$$K_Y = f^*(K_X + B) + \mathbf{A}_Y,$$

where $f : Y \to X$ is a proper birational morphism of normal varieties.

Definition 3.7 (Cartier closure). Let D be a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on a normal variety X. Then the $\mathbb{Q}$-b-divisor $\overline{D}$ denotes the *Cartier closure* of D, whose trace on Y is $\overline{D}_Y = f^*D$, where $f : Y \to X$ is a proper birational morphism of normal varieties.

Definition 3.8. Let $\mathbf{D}$ be a $\mathbb{Q}$-b-divisor on X. The round up $\lceil \mathbf{D} \rceil \in \mathbf{Div}(X)$ is defined componentwise. The restriction of $\mathbf{D}$ to an open subset $U \subset X$ is a well-defined $\mathbb{Q}$-b-divisor on U, denoted by $\mathbf{D}|_U$. Then $\mathcal{O}_X(\mathbf{D})$ is an $\mathcal{O}_X$-module whose sections on an open subset $U \subset X$ are given by

$$H^0(U, \mathcal{O}_X(\mathbf{D})) = \{a \in k(X)^\times;\ (\overline{(a)} + \mathbf{D})|_U \geq 0\} \cup \{0\},$$

where $k(X)$ is the function field of X. Note that $\mathcal{O}_X(\mathbf{D})$ is not necessarily coherent.

3.9. ***Basic properties.*** We recall the first basic property of discrepancy $\mathbb{Q}$-b-divisors. We will treat a generalization of Lemma 3.10 for sub-lc pairs below.

Lemma 3.10. *Let (X, B) be a sub-klt pair and let D be a Cartier divisor on X. Let $f : Y \to X$ be a proper surjective morphism from a nonsingular variety Y. We write $K_Y = f^*(K_X + B) + \sum a_i A_i$. We assume that $\sum A_i$ is a simple normal crossing divisor. Then, for every integer j,*

$$\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil + j\overline{D}) = f_*\mathcal{O}_Y\Big(\sum \lceil a_i \rceil A_i\Big) \otimes \mathcal{O}_X(jD)$$

Let E be an effective divisor on Y such that $E \leq \sum \lceil a_i \rceil A_i$. Then

$$\pi_* f_* \mathcal{O}_Y(E + f^* jD) \simeq \pi_* \mathcal{O}_X(jD)$$

if

$$\pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil + j\bar{D}) \subseteq \pi_*\mathcal{O}_X(jD),$$

where $\pi : X \to S$ is a proper surjective morphism onto a variety S.

Proof. For the first equality, see [Corti 2007, Lemmas 2.3.14 and 2.3.15] or their generalizations: Lemmas 3.19 and 3.20 below. Since E is effective,

$$\pi_*\mathcal{O}_X(jD) \subseteq \pi_* f_*\mathcal{O}_Y(E + f^* jD) \simeq \pi_*(f_*\mathcal{O}_Y(E) \otimes \mathcal{O}_X(jD)).$$

By the assumption and $E \le \sum \lceil a_i\rceil A_i$,

$$\begin{aligned}\pi_*(f_*\mathcal{O}_Y(E) \otimes \mathcal{O}_X(jD)) &\subseteq \pi_*\big(f_*\mathcal{O}_Y\big(\textstyle\sum \lceil a_i\rceil A_i\big)\otimes\mathcal{O}_X(jD)\big)\\ &= \pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil + j\bar{D})\\ &\subseteq \pi_*\mathcal{O}_X(jD).\end{aligned}$$

Therefore, we obtain $\pi_* f_*\mathcal{O}_Y(E + f^* jD) \simeq \pi_*\mathcal{O}_X(jD)$. □

We will use Lemma 3.11 in Section 4. The vanishing theorem in Lemma 3.11 is nothing but the Kawamata–Viehweg–Nadel vanishing theorem.

Lemma 3.11. *Let X be a normal variety and let B be a $\mathbb{Q}$-divisor on X such that $K_X + B$ is $\mathbb{Q}$-Cartier. Let $f : Y \to X$ be a proper birational morphism from a normal variety Y. We put $K_Y + B_Y = f^*(K_X + B)$. Then*

$$f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil) = \mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil)$$

and

$$R^i f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil) = 0$$

for every $i > 0$.

Proof. Let $g : Z \to Y$ be a resolution such that $\mathrm{Exc}(g) \cup g_*^{-1} B_Y$ has a simple normal crossing support. We put $K_Z + B_Z = g^*(K_Y + B_Y)$. Then $K_Z + B_Z = h^*(K_X + B)$, where $h = f \circ g : Z \to X$. By Lemma 3.10,

$$\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil) = g_*\mathcal{O}_Z(\lceil -B_Z\rceil)$$

and

$$\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil) = h_*\mathcal{O}_Z(\lceil -B_Z\rceil).$$

Therefore, $f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil) = \mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil)$. Since, $-B_Z = K_Z - h^*(K_X + B)$, we have

$$\lceil -B_Z\rceil = K_Z + \{B_Z\} - h^*(K_X + B).$$

Therefore, $R^i g_*\mathcal{O}_Z(\lceil -B_Z\rceil) = 0$ and $R^i h_*\mathcal{O}_Z(\lceil -B_Z\rceil) = 0$ for every $i > 0$ by the Kawamata–Viehweg vanishing theorem. Thus, $R^i f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil) = 0$ for every $i > 0$ by Leray's spectral sequence. □

Remark 3.12. We use the same notation as in Remark 3.2. Let (X, B) be a klt pair. Let D be a Cartier divisor on X and let $\pi : X \to S$ be a proper morphism onto a variety S. We put $p = \pi \circ f : Y \to S$. Then

$$p_*\mathcal{O}_Y(jf^*D) \simeq \pi_*\mathcal{O}_X(jD) \simeq p_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil + j\overline{f^*D})$$

for every integer j. This is because $f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil) = \mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil) \simeq \mathcal{O}_X$ by Lemma 3.11.

Remark 3.13 (Multiplier ideal sheaf). Let D be an effective $\mathbb{Q}$-divisor on a nonsingular variety X. Then $\mathcal{O}_X(\lceil \mathbf{A}(X, D)\rceil)$ is nothing but the *multiplier ideal sheaf* $\mathcal{J}(X, D) \subseteq \mathcal{O}_X$ of D on X. See [Lazarsfeld 2004, Definition 9.2.1]. More generally, let X be a normal variety and let Δ be a $\mathbb{Q}$-divisor on X such that $K_X+\Delta$ is $\mathbb{Q}$-Cartier. Let D be a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X. Then $\mathcal{O}_X(\lceil \mathbf{A}(X, \Delta + D)\rceil) = \mathcal{J}((X, \Delta); D)$, where the right hand side is the *multiplier ideal sheaf* defined (but not investigated) in [Lazarsfeld 2004, Definition 9.3.56]. In general, $\mathcal{O}_X(\lceil \mathbf{A}(X, \Delta+D)\rceil)$ is a fractional ideal of $k(X)$.

The next four remarks help us understand Theorem 2.1.

Remark 3.14 (Nonvanishing theorem). By Shokurov's nonvanishing theorem (see [Kawamata et al. 1987, Theorem 2-1-1]), we have $\pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil + j\overline{D}) \neq 0$ for every $j \gg 0$. Thus $\pi_*\mathcal{O}_X(jD) \neq 0$ for every $j \gg 0$ by condition (2) in Theorem 2.1.

Remark 3.15. We know that $\lceil \mathbf{A}(X, B)\rceil \geq 0$ since (X, B) is sub-klt. Therefore, $\pi_*\mathcal{O}_X(jD) \subseteq \pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil + j\overline{D})$. This implies that

$$\pi_*\mathcal{O}_X(jD) \simeq \pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil + j\overline{D})$$

for $j \geq j_0$, by condition (2) in Theorem 2.1.

Remark 3.16. If the pair (X, B) is klt, then $\lceil \mathbf{A}(X, B)\rceil$ is effective and exceptional over X. In this case, it is obvious that $\pi_*\mathcal{O}_X(jD) = \pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B)\rceil + j\overline{D})$.

Remark 3.17. Condition (2) in Theorem 2.1 is a very elementary case of *saturation of linear systems*. See [Corti 2007, 2.3.3] and [Ambro 2005b, 1-D].

We next introduce the notion of *non-klt $\mathbb{Q}$-b-divisor*, which is trivial for sub-klt pairs. We will use this in Section 5.

Definition 3.18 (Non-klt $\mathbb{Q}$-b-divisor). Let X be a normal variety and let B be a $\mathbb{Q}$-divisor on X such that $K_X + B$ is $\mathbb{Q}$-Cartier. Then the *non-klt $\mathbb{Q}$-b-divisor* of the pair (X, B) is the $\mathbb{Q}$-b-divisor $\mathbf{N} = \mathbf{N}(X, B)$ with the trace $\mathbf{N}_Y = \sum_{a_i \leq -1} a_i A_i$ for

$$K_Y = f^*(K_X + B) + \sum a_i A_i,$$

where $f : Y \to X$ is a proper birational morphism of normal varieties. It is easy to see that $\mathbf{N}(X, B)$ is a well-defined $\mathbb{Q}$-b-divisor. We put $\mathbf{A}^*(X, B) = \mathbf{A}(X, B) - \mathbf{N}(X, B)$.

Of course, $\mathbf{A}^*(X, B)$ is a well-defined $\mathbb{Q}$-b-divisor and $\lceil \mathbf{A}^*(X, B)\rceil \geq 0$. If (X, B) is sub-klt, then $\mathbf{N}(X, B) = 0$ and $\mathbf{A}(X, B) = \mathbf{A}^*(X, B)$.

The next lemma is a generalization of Lemma 3.10.

Lemma 3.19. *Let (X, B) be a sub-lc pair and let $f : Y \to X$ be a resolution such that $\mathrm{Exc}(f) \cup \mathrm{Supp}\, f_*^{-1}B$ is a simple normal crossing divisor on Y. We write $K_Y = f^*(K_X + B) + \sum a_i A_i$. Then*

$$\mathcal{O}_X(\lceil \mathbf{A}^*(X, B)\rceil) = f_*\mathcal{O}_Y\left(\sum_{a_i \neq -1} \lceil a_i \rceil A_i\right).$$

In particular, $\mathcal{O}_X(\lceil \mathbf{A}^*(X, B)\rceil)$ is a coherent $\mathcal{O}_X$-module. If (X, B) is lc, then $\mathcal{O}_X(\lceil \mathbf{A}^*(X, B)\rceil) \simeq \mathcal{O}_X$.

Let D be a Cartier divisor on X and let E be an effective divisor on Y such that $E \leq \sum_{a_i \neq -1} \lceil a_i \rceil A_i$. Then

$$\pi_* f_* \mathcal{O}_Y(E + f^* jD) \simeq \pi_* \mathcal{O}_X(jD)$$

if

$$\pi_* \mathcal{O}_X(\lceil \mathbf{A}^*(X, B)\rceil + j\bar{D}) \subseteq \pi_* \mathcal{O}_X(jD),$$

where $\pi : X \to S$ is a proper morphism onto a variety S.

Proof. By definition, $\mathbf{A}^*(X, B)_Y = \sum_{a_i \neq -1} a_i A_i$. If $g : Y' \to Y$ is a proper birational morphism from a normal variety Y', then

$$\lceil \mathbf{A}^*(X, B)_{Y'}\rceil = g^*\lceil \mathbf{A}^*(X, B)_Y\rceil + F,$$

where F is a g-exceptional effective divisor, by Lemma 3.20 below. This implies $f_*\mathcal{O}_Y(\lceil \mathbf{A}^*(X, B)_Y\rceil) = f'_*\mathcal{O}_{Y'}(\lceil \mathbf{A}^*(X, B)_{Y'}\rceil)$, where $f' = f \circ g$, from which it follows that $\mathcal{O}_X(\lceil \mathbf{A}^*(X, B)\rceil) = f_*\mathcal{O}_Y(\sum_{a_i \neq -1} \lceil a_i \rceil A_i)$ is a coherent $\mathcal{O}_X$-module. The last statement is easy to check. $\square$

Lemma 3.20. *Let (X, B) be a sub-lc pair and let $f : Y \to X$ be a resolution as in Lemma 3.19. We consider the $\mathbb{Q}$-b-divisor $\mathbf{A}^* = \mathbf{A}^*(X, B) = \mathbf{A}(X, B) - \mathbf{N}(X, B)$. If Y' is a normal variety and $g : Y' \to Y$ is a proper birational morphism, then*

$$\lceil \mathbf{A}^*_{Y'}\rceil = g^*\lceil \mathbf{A}^*_Y\rceil + F,$$

where F is a g-exceptional effective divisor.

Proof. By definition, we have $K_Y = f^*(K_X + B) + \mathbf{A}_Y$. Therefore we may write

$$\begin{aligned} K_{Y'} &= g^* f^*(K_X + B) + \mathbf{A}_{Y'} = g^*(K_Y - \mathbf{A}_Y) + \mathbf{A}_{Y'} \\ &= g^*(K_Y + \{-\mathbf{A}^*_Y\} - \mathbf{N}_Y + \lfloor -\mathbf{A}^*_Y \rfloor) + \mathbf{A}_{Y'} \\ &= g^*(K_Y + \{-\mathbf{A}^*_Y\} - \mathbf{N}_Y) + \mathbf{A}_{Y'} - g^*\lceil \mathbf{A}^*_Y\rceil. \end{aligned}$$

We note that $(Y, \{-\mathbf{A}^*_Y\} - \mathbf{N}_Y)$ is lc and that the set of lc centers of $(Y, \{-\mathbf{A}^*_Y\} - \mathbf{N}_Y)$ coincides with that of $(Y, -\mathbf{A}^*_Y - \mathbf{N}_Y) = (Y, -\mathbf{A}_Y)$. Therefore, the round-up of $\mathbf{A}_{Y'} - g^*\lceil \mathbf{A}^*_Y \rceil - \mathbf{N}_{Y'}$ is effective and g-exceptional. Thus, we can write $\lceil \mathbf{A}^*_{Y'} \rceil = g^*\lceil \mathbf{A}^*_Y \rceil + F$, where F is a g-exceptional effective divisor. □

The next lemma is obvious by Lemma 3.19.

Lemma 3.21. *Let (X, B) be a sub-lc pair and let $f : Y \to X$ be a proper birational morphism from a normal variety Y. We put $K_Y + B_Y = f^*(K_X + B)$. Then $f_*\mathcal{O}_Y(\lceil \mathbf{A}^*(Y, B_Y) \rceil) = \mathcal{O}_X(\lceil \mathbf{A}^*(X, B) \rceil)$.*

4. Basepoint-free theorem; nef and abundant case

We recall the definition of *abundant* divisors, which are called *good* divisors in [Kawamata 1985]. See [Kawamata et al. 1987, Section 6-1].

Definition 4.1 (Abundant divisor). Let X be a complete normal variety and let D be a $\mathbb{Q}$-Cartier nef $\mathbb{Q}$-divisor on X. We define the *numerical Iitaka dimension* to be

$$\nu(X, D) = \max\{e;\ D^e \not\equiv 0\}.$$

This means that $D^{e'} \cdot S = 0$ for any e'-dimensional subvarieties S of X with $e' > e$ and there exists an e-dimensional subvariety T of X such that $D^e \cdot T > 0$. Then it is easy to see that $\kappa(X, D) \leq \nu(X, D)$, where $\kappa(X, D)$ denotes *Iitaka's D-dimension*. A nef $\mathbb{Q}$-divisor D is said to be *abundant* if the equality $\kappa(X, D) = \nu(X, D)$ holds. Let $\pi : X \to S$ be a proper surjective morphism of normal varieties and let D be a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X. Then D is said to be π-abundant if $D|_{X_\eta}$ is abundant, where X_η is the generic fiber of π.

The next theorem is the main theorem of [Kawamata 1985]. For the relative statement, see [Nakayama 1986, Theorem 5]. We reduced Theorem 4.2 to Theorem 2.1 by using Ambro's results in [Ambro 2004] and [Ambro 2007], which is the main theme of [Fujino 2011d]. For the details, see [Fujino 2011d, Section 2].

Theorem 4.2 cf. [Kawamata et al. 1987, Theorem 6-1-11]. *Let (X, B) be a klt pair and let $\pi : X \to S$ be a proper morphism onto a variety S. Assume the following conditions*:

(a) *H is a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X.*

(b) *$H - (K_X + B)$ is π-nef and π-abundant.*

(c) *$\kappa(X_\eta, (aH - (K_X + B))_\eta) \geq 0$ and*

$$\nu(X_\eta, (aH - (K_X + B))_\eta) = \nu(X_\eta, (H - (K_X + B))_\eta)$$

for some $a \in \mathbb{Q}$ with $a > 1$, where η is the generic point of S.

Then H is π-semiample.

Definition 4.3 (Iitaka fibration). Let $\pi : X \to S$ be a proper surjective morphism of normal varieties. Let D be a $\mathbb{Q}$-Cartier $\mathbb{Q}$-Weil divisor on X such that $\kappa(X_\eta, D_\eta) \geq 0$, where η is the generic point of S. Let $X \dashrightarrow W$ be the rational map over S induced by $\pi^*\pi_*\mathcal{O}_X(mD) \to \mathcal{O}_X(mD)$ for a sufficiently large and divisible integer m. We consider a projective surjective morphism $f : Y \to Z$ of nonsingular varieties that is birational to $X \dashrightarrow W$. We call $f : Y \to Z$ the *Iitaka fibration* with respect to D over S.

We now state the main result of this section, which will be used in the proof of Theorem 7.11. It is a slight generalization of Theorem 4.2.

Theorem 4.4. *Let (X, B) be a sub-klt pair and let $\pi : X \to S$ be a proper morphism onto a variety S. Assume the following conditions*:

(a) *H is a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X.*

(b) *$H - (K_X + B)$ is π-nef and π-abundant.*

(c) *$\kappa(X_\eta, (aH - (K_X + B))_\eta) \geq 0$ and*

$$\nu(X_\eta, (aH - (K_X + B))_\eta) = \nu(X_\eta, (H - (K_X + B))_\eta)$$

for some $a \in \mathbb{Q}$ with $a > 1$, where η is the generic point of S.

(d) *Let $f : Y \to Z$ be the Iitaka fibration with respect to $H - (K_X + B)$ over S. We assume that there exists a proper birational morphism $\mu : Y \to X$ and put $K_Y + B_Y = \mu^*(K_X + B)$. In this setting, we assume* rank $f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y) \rceil) = 1$.

(e) (Saturation condition.) *There exist positive integers b and j_0 such that bH is Cartier and $\pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil + jb\overline{H}) \subseteq \pi_*\mathcal{O}_X(jbH)$ for every positive integer $j \geq j_0$.*

Then H is π-semiample.

Proof. The proof of Theorem 4.2 given in [Fujino 2011d, Section 2] works without any changes. We note that condition (d) implies [ibid., Lemma 2.3] and that we can use condition (e) in the proof of [ibid., Lemma 2.4]. □

Remark 4.5. The rank of $f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y) \rceil)$ is a birational invariant for $f : Y \to Z$ by Lemma 3.11.

Remark 4.6. If (X, B) is klt and bH is Cartier, it is obvious that

$$\pi_*\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil + jb\overline{H}) \simeq \pi_*\mathcal{O}_X(jbH)$$

for every positive integer j (see Remark 3.16).

Remark 4.7. We can easily generalize Theorem 4.4 to varieties in class $\mathcal{C}$ by suitable modifications. For details, see [Fujino 2011d, Section 4].

The following examples help us understand condition (d).

Example 4.8. Let $X = E$ be an elliptic curve and let $P \in X$ be a closed point. Take a general member $P_1 + P_2 + P_3 \in |3P|$. We put $B = \frac{1}{3}(P_1 + P_2 + P_3) - P$. Then (X, B) is sub-klt and $K_X + B \sim_{\mathbb{Q}} 0$. In this case, $\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil) \simeq \mathcal{O}_X(P)$ and $H^0(X, \mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil)) \simeq k$.

Example 4.9. Let $f : X = \mathbb{P}_{\mathbb{P}^1}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(1)) \to Z = \mathbb{P}^1$ be the Hirzebruch surface and let C (resp. E) be the positive (resp. negative) section of f. We take a general member $B_0 \in |5C|$. Note that $|5C|$ is a free linear system on X. We put $B = -\frac{1}{2}E + \frac{1}{2}B_0$ and consider the pair (X, B). Then (X, B) is sub-klt. We put $H = 0$. Then H is a nef Cartier divisor on X and $aH - (K_X + B) \sim_{\mathbb{Q}} \frac{1}{2}F$ for every rational number a, where F is a fiber of f. Therefore, $aH - (K_X + B)$ is nef and abundant for every rational number a. In this case, $\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil) \simeq \mathcal{O}_X(E)$. Thus

$$\begin{aligned} H^0(X, \mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil + j\overline{H})) &\simeq H^0(X, \mathcal{O}_X(E)) \simeq k \\ &\simeq H^0(X, \mathcal{O}_X) \simeq H^0(X, \mathcal{O}_X(jH)) \end{aligned}$$

for every integer j. Therefore, $\pi : X \to \operatorname{Spec} k$, H, and (X, B) satisfy conditions (a), (b), (c), and (e) in Theorem 4.4. However, (d) is not satisfied. In our case, it is easy to see that $f : X \to Z$ is the Iitaka fibration with respect to $H - (K_X + B)$. Since $f_*\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil) \simeq f_*\mathcal{O}_X(E)$, we have $\operatorname{rank} f_*\mathcal{O}_X(\lceil \mathbf{A}(X, B) \rceil) = 2$.

Remark 4.10. In Theorem 4.4, assumptions (a)–(c) are the same as in Theorem 4.2. Condition (e) is indispensable by Example 2.3 for sub-klt pairs. By using the nonvanishing theorem for generalized normal crossing varieties in [Kawamata 1985, Theorem 5.1], which is the hardest part to prove in [Kawamata 1985], the semiampleness of H seems to follow from conditions (a), (b), (c), and (e). However, we need (d) to apply Ambro's canonical bundle formula to the Iitaka fibration $f : Y \to Z$. See, for example, [Fujino 2011d, Section 3]. Unfortunately, as we saw in Example 4.9, condition (d) does not follow from the other assumptions. Anyway, condition (d) is automatically satisfied if (X, B) is klt; see [Fujino 2011d, Lemma 2.3].

The following two examples show that the effective version of Theorem 4.2 does not necessarily hold. The first one is an obvious example.

Example 4.11. Let $X = E$ be an elliptic curve and let m be an arbitrary positive integer. Then there is a Cartier divisor H on X such that $mH \sim 0$ and $lH \not\sim 0$ for $0 < l < m$. Therefore, the effective version of Theorem 4.2 does not necessarily hold.

The next one shows the reason why Theorem 2.4 does not imply the effective version of Theorem 4.2.

Example 4.12. Let E be an elliptic curve and $G = \mathbb{Z}/m\mathbb{Z} = \langle \zeta \rangle$, where ζ is a primitive m-th root of unity. We take an m-torsion point $a \in E$. The cyclic group

G acts on $E \times \mathbb{P}^1$ as follows:

$$E \times \mathbb{P}^1 \ni (x, [X_0 : X_1]) \mapsto (x + a, [\zeta X_0 : X_1]) \in E \times \mathbb{P}^1.$$

We put $X = (E \times \mathbb{P}^1)/G$. Then X has a structure of elliptic surface $p : X \to \mathbb{P}^1$. In this setting,

$$K_X = p^* \left(K_{\mathbb{P}^1} + \frac{m-1}{m}[0] + \frac{m-1}{m}[\infty] \right).$$

We put $H = p^{-1}(0)_{\mathrm{red}}$. Then H is a Cartier divisor on X. It is easy to see that H is nef and $H - K_X$ is nef and abundant. Moreover, $\kappa(X, aH - K_X) = \nu(X, aH - K_X) = 1$ for every rational number $a > 0$. It is obvious that $|mH|$ is free. However, $|lH|$ is not free for $0 < l < m$. Thus, the effective version of Theorem 4.2 does not hold.

5. Basepoint-free theorem of Reid–Fukuda type

The following result is a reformulation of the main theorem of [Fujino 2000].

Theorem 5.1 (Basepoint-free theorem of Reid–Fukuda type). *Let X be a nonsingular variety and let B be a $\mathbb{Q}$-divisor on X such that* Supp B *is a simple normal crossing divisor and (X, B) is sub-lc. Let $\pi : X \to S$ be a proper morphism onto a variety S and let D be a π-nef Cartier divisor on X. Assume the following conditions*:

(1) *$rD - (K_X + B)$ is nef and log big over S for some positive integer r.*

(2) (Saturation condition.) *There exists a positive integer j_0 such that*

$$\pi_* \mathcal{O}_X(\lceil \mathbf{A}^*(X, B) \rceil + j\overline{D}) \subseteq \pi_* \mathcal{O}_X(jD)$$

for every integer $j \geq j_0$.

Then mD is π-generated for every $m \gg 0$, that is, there exists a positive integer m_0 such that for every $m \geq m_0$ the natural homomorphism $\pi^* \pi_* \mathcal{O}_X(mD) \to \mathcal{O}_X(mD)$ is surjective.

Definition 5.2. Let (X, B) be a sub-lc pair and let $\pi : X \to S$ be a proper morphism onto a variety S. Let $\mathcal{L}$ be a line bundle on X. We say that $\mathcal{L}$ is *nef and log big* over S if and only if $\mathcal{L}$ is π-nef and π-big and the restriction $\mathcal{L}|_W$ is big over $\pi(W)$ for every lc center W of the pair (X, B). A $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor H on X is said to be nef and log big over S if and only if so is $\mathcal{O}_X(cH)$, where c is a positive integer such that cH is Cartier.

Proof of Theorem 5.1. We write $B = T + B_+ - B_-$ such that T, B_+, and B_- are effective divisors, they have no common irreducible components, $\lfloor B_+ \rfloor = 0$, and

$\lfloor T \rfloor = T$. If $T = 0$, then (X, B) is sub-klt. So, theorem follows from Theorem 2.1. Thus, we assume $T \neq 0$. Let T_0 be an irreducible component of T. If $m \geq r$, then

$$mD + \lceil B_- \rceil - T_0 - (K_X + B + \lceil B_- \rceil - T_0) = mD - (K_X + B)$$

is nef and log big over S for the pair $(X, B + \lceil B_- \rceil - T_0)$. We note that $B + \lceil B_- \rceil - T_0$ is effective. Therefore, $R^1\pi_*\mathcal{O}_X(\lceil B_- \rceil - T_0 + mD) = 0$ for $m \geq r$ by the vanishing theorem: Lemma 5.3. Thus, we obtain the following commutative diagram for $m \geq \max\{r, j_0\}$:

$$\begin{tikzcd}
\pi_*\mathcal{O}_X(\lceil B_- \rceil + mD) \arrow[r] & \pi_*\mathcal{O}_{T_0}(\lceil B_-|_{T_0} \rceil + mD|_{T_0}) \arrow[r] & 0 \\
\pi_*\mathcal{O}_X(mD) \arrow[u, "\cong"] \arrow[r, "\alpha"] & \pi_*\mathcal{O}_{T_0}(mD|_{T_0}). \arrow[u, "\iota"] &
\end{tikzcd}$$

Here, we used

$$\begin{aligned}
\pi_*\mathcal{O}_X(mD) &\subseteq \pi_*\mathcal{O}_X(\lceil B_- \rceil + mD) \\
&\simeq \pi_*\mathcal{O}_X(\lceil \mathbf{A}^*(X, B) \rceil + m\overline{D}) \\
&\subseteq \pi_*\mathcal{O}_X(mD)
\end{aligned}$$

for $m \geq j_0$ (see Lemma 3.19). We put $K_{T_0} + B_{T_0} = (K_X + B)|_{T_0}$ and $D_{T_0} = D|_{T_0}$. Then (T_0, B_{T_0}) is sub-lc and it is easy to see that $rD_{T_0} - (K_{T_0} + B_{T_0})$ is nef and log big over $\pi(T_0)$. It is obvious that T_0 is nonsingular and Supp B_{T_0} is a simple normal crossing divisor. We note that $\pi_*\mathcal{O}_{T_0}(\lceil \mathbf{A}^*(T_0, B_{T_0}) \rceil + j\overline{D_{T_0}}) \simeq \pi_*\mathcal{O}_{T_0}(jD_{T_0})$ for every $j \geq \max\{r, j_0\}$ follows from the above diagram, that is, the natural inclusion ι is isomorphism for $m \geq \max\{r, j_0\}$. Thus, α is surjective for $m \geq \max\{r, j_0\}$. By induction, mD_{T_0} is π-generated for every $m \gg 0$. We can apply the same argument to every irreducible component of T. Therefore, the relative base locus of mD is disjoint from T for every $m \gg 0$ since the restriction map $\alpha : \pi_*\mathcal{O}_X(mD) \to \pi_*\mathcal{O}_{T_0}(mD_{T_0})$ is surjective for every irreducible component T_0 of T. The arguments in [Fukuda 1996, Proof of Theorem 3], which is a variant of the X-method, work without any changes (cf. Theorem 6.1). So, we obtain that mD is π-generated for every $m \gg 0$. □

The following vanishing theorem was already used in the proof of Theorem 5.1. The proof is an easy exercise by induction on dim X and on the number of the irreducible components of $\lfloor \Delta \rfloor$.

Lemma 5.3. *Let $\pi : X \to S$ be a proper morphism from a nonsingular variety X. Let $\Delta = \sum d_i \Delta_i$ be a sum of distinct prime divisors such that* Supp Δ *is a simple normal crossing divisor and d_i is a rational number with $0 \leq d_i \leq 1$ for every i. Let D be a Cartier divisor on X. Assume that $D - (K_X + \Delta)$ is nef and log big over S for the pair (X, Δ). Then $R^i\pi_*\mathcal{O}_X(D) = 0$ for every $i > 0$.*

As in Theorem 2.4, effective freeness holds under the same assumption as in Theorem 5.1.

Theorem 5.4 (Effective freeness). *We use the same notation and assumption as in Theorem 5.1. Then there exists a positive integer l, which depends only on* $\dim X$ *and* $\max\{r, j_0\}$, *such that lD is π-generated, that is, $\pi^*\pi_*\mathcal{O}_X(lD) \to \mathcal{O}_X(lD)$ is surjective.*

Sketch of the proof. If (X, B) is sub-klt, then this theorem is nothing but Theorem 2.4. So, we can assume that (X, B) is not sub-klt. In this case, the arguments in [Fukuda 1996, Section 4] work with only minor modifications. From now on, we use the notation in [Fukuda 1996, Section 4]. By minor modifications, the proof in [Fukuda 1996, Section 4] works under the following weaker assumptions: X is nonsingular and Δ is a $\mathbb{Q}$-divisor on X such that $\operatorname{Supp}\Delta$ is a simple normal crossing divisor and (X, Δ) is sub-lc. In [Fukuda 1996, Claim 5], E_i is f-exceptional. In our setting, this is not true. However, the bound

$$0 \le \sum_{cb_i - e_i + p_i < 0} \lceil -(cb_i - e_i + p_i)\rceil E_i \le \lceil \mathbf{A}^*(X, \Delta)_Y \rceil,$$

which always holds even when Δ is not effective, is sufficient for us. It is because we can use the saturation condition (2) in Theorem 5.1. We leave the details as an exercise for the reader since all we have to do is to repeat the arguments in [Kollár 1993, Section 2] and [Fukuda 1996, Section 4]. □

The final statement in this section is the (effective) basepoint-free theorem of Reid–Fukuda type for dlt pairs.

Corollary 5.5. *Let (X, B) be a dlt pair and let $\pi : X \to S$ be a proper morphism onto a variety S. Let D be a π-nef Cartier divisor on X. Assume that $rD-(K_X+B)$ is nef and log big over S for some positive integer r. Then there exists a positive integer m_0 such that mD is π-generated for every $m \ge m_0$ and we can find a positive integer l, which depends only on* $\dim X$ *and r, such that lD is π-generated.*

Proof. Let $f : Y \to X$ be a resolution such that $\operatorname{Exc}(f)$ and $\operatorname{Exc}(f) \cup \operatorname{Supp} f_*^{-1}B$ are simple normal crossing divisors, $K_Y + B_Y = f^*(K_X + B)$, and f is an isomorphism over all the generic points of lc centers of the pair (X, B). Then (Y, B_Y) is sub-lc, and $rD_Y-(K_Y+B_Y)$ is nef and log big over S, where $D_Y = f^*D$. Since $\lceil \mathbf{A}^*(X, B)\rceil$ is effective and exceptional over X, $p_*\mathcal{O}_Y(\lceil \mathbf{A}^*(Y, B_Y)\rceil + j\overline{D_Y}) \simeq p_*\mathcal{O}_Y(jD_Y)$ for every j, where $p = \pi \circ f$. So, we can apply Theorems 5.1 and 5.4 to D_Y and (Y, B_Y). This concludes the proof. □

For the (effective) basepoint-freeness for lc pairs, see [Fujino 2009a; Fujino 2009b, 3.3.1 Base Point Free Theorem; Fujino 2011a, Theorem 13.1; Fujino 2011c, Theorem 1.2].

6. Variants of basepoint-free theorems due to Fukuda

The starting point of this section is a slight generalization of Theorem 2.1. It is essentially the same as [Fukuda 1996, Theorem 3].

Theorem 6.1. *Let X be a nonsingular variety and let B be a $\mathbb{Q}$-divisor on X such that (X, B) is sub-lc and* Supp B *is a simple normal crossing divisor. Let $\pi : X \to S$ be a proper morphism onto a variety S and let H be a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X. Assume the following conditions*:

(1) $H - (K_X + B)$ *is nef and big over S.*

(2) (Saturation condition.) *There exist positive integers b and j_0 such that*

$$\pi_* \mathcal{O}_X(\lceil \mathbf{A}^*(X, B) \rceil + jb\overline{H}) \subseteq \pi_* \mathcal{O}_X(jbH)$$

for every integer $j \geq j_0$.

(3) *There is a positive integer c such that cH is Cartier and*

$$\mathcal{O}_T(cH) := \mathcal{O}_X(cH)|_T$$

is π-generated, where $T = -\mathbf{N}(X, B)_X$.

Then H is π-semiample.

Proof. If (X, B) is sub-klt, then this follows from Theorem 2.1. By replacing H by a multiple, we can assume that $b = 1$, $j_0 = 1$, and $c = 1$. Since

$$lH + \lceil \mathbf{A}^*_X \rceil - T - (K_X + \{B\}) = lH - (K_X + B)$$

is nef and big over S for every positive integer l, we have the following commutative diagram by the Kawamata–Viehweg vanishing theorem:

$$\begin{array}{ccccc}
\pi_* \mathcal{O}_X(lH + \lceil \mathbf{A}^*_X \rceil) & \longrightarrow & \pi_*(\mathcal{O}_T(lH) \otimes \mathcal{O}_T(\lceil \mathbf{A}^*_X|_T \rceil)) & \longrightarrow & 0 \\
\uparrow{\cong} & & \uparrow{\iota} & & \\
\pi_* \mathcal{O}_X(lH) & \xrightarrow[\alpha]{} & \pi_* \mathcal{O}_T(lH). & &
\end{array}$$

Thus, the natural inclusion ι is an isomorphism and α is surjective for every $l \geq 1$. In particular, $\pi_* \mathcal{O}_X(lH) \neq 0$ for every $l \geq 1$. The same arguments as in [Fukuda 1996, Proof of Theorem 3] show that H is π-semiample. $\square$

The main purpose of this section is to prove Theorem 6.2 below, which is a generalization of Theorem 4.4 and Theorem 6.1. The basic strategy of the proof is the same as that of Theorem 4.4. That is, by using Ambro's canonical bundle formula, we reduce it to the case when $H - (K_X + B)$ is nef and big. This is nothing but Theorem 6.1.

Theorem 6.2. *Let X be a nonsingular variety and let B be a $\mathbb{Q}$-divisor on X such that (X, B) is sub-lc and* Supp B *is a simple normal crossing divisor. Let $\pi : X \to S$ be a proper morphism onto a variety S. Assume the following conditions*:

(a) *H is a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on X.*

(b) *$H - (K_X + B)$ is π-nef and π-abundant.*

(c) *$\kappa(X_\eta, (aH - (K_X + B))_\eta) \geq 0$ and $\nu(X_\eta, (aH - (K_X + B))_\eta) = \nu(X_\eta, (H - (K_X + B))_\eta)$ for some $a \in \mathbb{Q}$ with $a > 1$, where η is the generic point of S.*

(d) *Let $f : Y \to Z$ be the Iitaka fibration with respect to $H - (K_X + B)$ over S. We assume that there exists a proper birational morphism $\mu : Y \to X$ and put $K_Y + B_Y = \mu^*(K_X + B)$. We also assume* rank $f_*\mathcal{O}_Y(\lceil \mathbf{A}^*(Y, B_Y)\rceil) = 1$.

(e) (Saturation condition.) *There exist positive integers b and j_0 such that bH is Cartier and $\pi_*\mathcal{O}_X(\lceil \mathbf{A}^*(X, B)\rceil + jb\overline{H}) \subseteq \pi_*\mathcal{O}_X(jbH)$ for every positive integer $j \geq j_0$.*

(f) *There is a positive integer c such that cH is Cartier and*

$$\mathcal{O}_T(cH) := \mathcal{O}_X(cH)|_T$$

is π-generated, where $T = -\mathbf{N}(X, B)_X$.

Then H is π-semiample.

Proof. If $H - (K_X + B)$ is big, this follows from Theorem 6.1. So, we can assume that $H - (K_X + B)$ is not big. Form now on, we use the notation from the proof of Theorem 4.2, which is given in [Fujino 2011d, Section 2]. We just explain how to modify that proof. Let us recall the commutative diagram

$$\begin{array}{ccc} Y & \xrightarrow{f} & Z \\ {\scriptstyle\mu}\downarrow & & \downarrow{\scriptstyle\varphi} \\ X & \xrightarrow[\pi]{} & S \end{array}$$

from the proof of [Fujino 2011d, Theorem 1.1], where $f : Y \to Z$ is the Iitaka fibration with respect to $H - (K_X + B)$ over S. For the details, see [Fujino 2011d, Section 2]. We note that $\mu^*H = H_Y$ and $H_Y \sim f^*D$. Here, we replaced H with a multiple and assumed that H and D are Cartier (see [Fujino 2011d, page 307]). We can also assume that $b = j_0 = 1$ in (e) and $c = 1$ in (f) by replacing H with a multiple. We start with the following obvious lemma.

Lemma 6.3. *We put $T' = -\mathbf{N}(X, B)_Y$. Then $\mu(T') \subset T$. Therefore, $\mathcal{O}_{T'}(H_Y) := \mathcal{O}_Y(H_Y)|_{T'}$ is p-generated, where $p = \pi \circ \mu$.*

Lemma 6.4. *If $f(T') = Z$, then H_Y is p-semiample. In particular, H is π-semiample.*

Proof. There exists an irreducible component T'_0 of T' such that $f(T'_0) = Z$. Since $(H_Y)|_{T'_0} \sim (f^*D)|_{T'_0}$ is p-semiample, D is φ-semiample. This implies that H_Y is p-semiample and H is π-semiample. □

Therefore, we can assume that T' is not dominant onto Z. Thus $\mathbf{A}(Y, B_Y) = \mathbf{A}^*(Y, B_Y)$ over the generic point of Z. Equivalently, (Y, B_Y) is sub-klt over the generic point of Z. As in [Fujino 2011d, Proof of Theorem 1.1], we have these properties:

(1) $K_Y + B_Y \sim_{\mathbb{Q}} f^*(K_Z + B_Z + M)$, where B_Z is the *discriminant* $\mathbb{Q}$-divisor of (Y, B_Y) on Z and M is the *moduli* $\mathbb{Q}$-divisor on Z.

(2′) (Z, B_Z) is sub-lc.

(3) M is a φ-nef $\mathbb{Q}$-divisor on Z.

(4′) $\varphi_*\mathcal{O}_Z(\lceil \mathbf{A}^*(Z, B_Z)\rceil + j\overline{D}) \subseteq \varphi_*\mathcal{O}_Z(jD)$ for every positive integer j.

(5) $D - (K_Z + B_Z)$ is φ-nef and φ-big.

(6) Y and Z are nonsingular and $\operatorname{Supp} B_Y$ and $\operatorname{Supp} B_Z$ are simple normal crossing divisors.

(7) $\mathcal{O}_{T''}(D) := \mathcal{O}_Z(D)|_{T''}$ is φ-generated where $T'' = -\mathbf{N}(Z, B_Z)_Z$.

Once the conditions above were satisfied, D is φ-semiample by Theorem 6.1. Therefore, H is π-semiample. So, all we have to do is check the conditions. Conditions (1), (2′), (3), (5), (6) are satisfied by a result of Ambro; see [Fujino 2011d, Proof of Theorem 1.1]. We note that $f_*\mathcal{O}_Y(\lceil \mathbf{A}(Y, B_Y)\rceil)$ and $f_*\mathcal{O}_Y(\lceil \mathbf{A}^*(Y, B_Y)\rceil)$ both have rank 1. By the same computation as in [Ambro 2007, Lemma 9.2.2 and Proposition 9.2.3], we have the following lemma.

Lemma 6.5. *$\mathcal{O}_Z(\lceil \mathbf{A}^*(Z, B_Z)\rceil + j\overline{D}) \subseteq f_*\mathcal{O}_Y(\lceil \mathbf{A}^*(Y, B_Y)\rceil + j\overline{H_Y})$ for every integer j.*

Thus, we have (4′) by the saturation condition (e) (for details, see [Fujino 2011d, Proof of Theorem 1.1], and Lemma 3.21). By definition, we have

$$lH_Y + \lceil \mathbf{A}^*_Y \rceil - T' - (K_Y + \{B_Y\}) \sim_{\mathbb{Q}} f^*((l-1)D + M_0),$$

where

$$H_Y - (K_Y + B_Y) = \mu^*(H - (K_X + B)) \sim_{\mathbb{Q}} f^*M_0.$$

Note that $(l-1)D + M_0$ is φ-nef and φ-big for $l \geq 1$. By the Kollár type injectivity theorem,

$$R^1 p_*\mathcal{O}_Y(lH_Y + \lceil \mathbf{A}^*_Y \rceil - T') \to R^1 p_*\mathcal{O}_Y(lH_Y + \lceil \mathbf{A}^*_Y \rceil)$$

is injective for $l \geq 1$. Note that the above injectivity can be checked easily by [Fujino 2007a, Theorem 1.1]. Here, we used the fact that $f(T') \subsetneq Z$. So, we have the commutative diagram

$$\begin{tikzcd} p_*\mathcal{O}_Y(lH_Y+\lceil \mathbf{A}^*_Y\rceil) \arrow[r] & p_*(\mathcal{O}_{T'}(lH_Y)\otimes \mathcal{O}_{T'}(\lceil \mathbf{A}^*_Y|_{T'}\rceil)) \arrow[r] & 0 \\ p_*\mathcal{O}_Y(lH_Y) \arrow[u, "\cong"] \arrow[r, "\alpha"'] & p_*\mathcal{O}_{T'}(lH_Y) \arrow[u, "\iota"] & \end{tikzcd}$$

The isomorphism of the left vertical arrow follows from the saturation condition (e). Thus, the natural inclusion ι is an isomorphism and α is surjective for $l \geq 1$. In particular, the relative base locus of lH_Y is disjoint from T' since $\mathcal{O}_{T'}(lH_Y)$ is p-generated (cf. Lemma 6.3). On the other hand, $H_Y \sim f^*D$. Therefore, $\mathcal{O}_{T''}(D)$ is φ-generated since $T'' \subset f(T')$. So, we obtain condition (7). This completes the proof of Theorem 6.2. □

As a corollary of Theorem 6.2, we obtain the generalization of [Fukuda 2002, Proposition 3.3] stated in the introduction (Theorem 1.1). Before we derive it, we recall the definition of *non-klt loci*.

Definition 6.6 (Non-klt locus). Let (X, B) be an lc pair. We consider the closed subset

$$\mathrm{Nklt}(X, B) = \{x \in X \mid (X, B) \text{ is not klt at } x\}$$

of X. We call $\mathrm{Nklt}(X, B)$ the *non-klt locus* of (X, B).

Proof of Theorem 1.1. Let $h : X' \to X$ be a resolution such that $\mathrm{Exc}(h)\cup \mathrm{Supp}\, h^{-1}_*B$ is a simple normal crossing divisor and $K_{X'}+B_{X'} = h^*(K_X+B)$. Then $H_{X'} = h^*H$, $(X', B_{X'})$, and $\pi' = \pi\circ h : X' \to S$ satisfy assumptions (a), (b), and (c) in Theorem 6.2. By the same argument as in the proof of [Fujino 2011d, Lemma 2.3], we obtain $\operatorname{rank} f_*\mathcal{O}_Y(\lceil \mathbf{A}^*(Y, B_Y)\rceil) = 1$, where $f : Y \to Z$ is the Iitaka fibration as in (d) in Theorem 6.2. Note that $\lceil \mathbf{A}^*(Y, B_Y)\rceil$ is effective and exceptional over X. Since B is effective, $\lceil \mathbf{A}^*(X, B)\rceil$ is effective and exceptional over X,

$$\pi'_*\mathcal{O}_{X'}(\lceil \mathbf{A}^*(X', B_{X'})\rceil + jb\overline{H_{X'}}) \subseteq \pi'_*\mathcal{O}_{X'}(jbH_{X'})$$

for every integer j, where b is a positive integer such that bH is Cartier. So, the saturation condition (e) in Theorem 6.2 is satisfied. Finally, $\mathcal{O}_{T'}(cH_{X'}) := \mathcal{O}_{X'}(cH_{X'})|_{T'}$ is π'-generated, where $T' = -\mathbf{N}(X, B)_{X'}$, by assumption (D) and the fact that $h(T') \subset T$. So, condition (f) in Theorem 6.2 for $H_{X'}$ and $(X', B_{X'})$ is satisfied. Therefore, $H_{X'}$ is π'-semiample by Theorem 6.2. Thus, H is π-semiample. □

Remark 6.7. (i) It is obvious that $\mathrm{Supp}(-\mathbf{N}(X, B)_X) \subseteq \mathrm{Nklt}(X, B)$. In general, $\mathrm{Supp}(-\mathbf{N}(X, B)_X) \subsetneq \mathrm{Nklt}(X, B)$. In particular, $\mathrm{Nklt}(X, B)$ is not necessarily of pure codimension one in X.

(ii) If (X, B) is dlt, then $\mathrm{Nklt}(X, B) = \mathrm{Supp}(-\mathbf{N}(X, B)_X) = \lfloor B \rfloor$. Therefore, if (X, B) is dlt and S is a point, then Theorem 1.1 is nothing but Fukuda's result [Fukuda 2002, Proposition 3.3].

(iii) The reader can find applications of this corollary in [Fukuda 2002; Fujino 2010; Fujino and Gongyo 2011].

By combining Theorem 1.1 with [Gongyo 2010, Theorem 1.5], we obtain the following result.

Corollary 6.8. *Let (X, B) be a projective dlt pair such that $\nu(K_X+B)=\kappa(K_X+B)$ and that $(K_X + B)|_{\lfloor B \rfloor}$ is numerically trivial. Then $K_X + B$ is semiample.*

Remark 6.9. We can easily generalize Theorem 6.2 and Theorem 1.1 to varieties in class $\mathscr{C}$ by suitable modifications. For details, see [Fujino 2011d, Section 4].

7. Basepoint-free theorems for pseudo-klt pairs

In this section, we generalize the Kawamata–Shokurov base point free theorem and Kawamata's theorem: Theorem 4.2 for *klt pairs* to *pseudo-klt pairs*. We think that our formulation is useful when we study lc centers (see Proposition 7.8). First, we introduce the notion of *pseudo-klt pairs*.

Definition 7.1 (Pseudo-klt pair). Let W be a normal variety. Assume the following conditions:

(1) there exist a sub-klt pair (V, B) and a projective surjective morphism $f : V \to W$ with connected fibers.

(2) $f_*\mathscr{O}_V(\lceil \mathbf{A}(V, B) \rceil) \simeq \mathscr{O}_W$.

(3) There exists a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor $\mathscr{H}$ on W such that $K_V + B \sim_{\mathbb{Q}} f^*\mathscr{H}$.

Then the pair $[W, \mathscr{H}]$ is called a *pseudo-klt pair*.

Although it is the first time that we use the name of *pseudo-klt pair*, the notion of pseudo-klt pair appeared in [Fujino 1999], where we proved the cone and contraction theorem for *pseudo-klt* pairs (cf. [Fujino 1999, Section 4]). We note that all the fundamental theorems for the log minimal model program for pseudo-klt pairs can be proved by the theory of quasilog varieties (cf. [Ambro 2003; Fujino 2009b; 2011b]).

Remark 7.2. In Definition 7.1, we assume that W is normal. However, the normality of W follows from condition (2) and the normality of V. Note that $\lceil \mathbf{A}(V, B) \rceil$ is effective.

Remark 7.3. In the definition of pseudo-klt pairs, if (V, B) is klt, the condition $f_*\mathscr{O}_V(\lceil \mathbf{A}(V, B) \rceil) \simeq \mathscr{O}_W$ is automatically satisfied. This is because $\lceil \mathbf{A}(V, B) \rceil$ is effective and exceptional over V.

We note that a pseudo-klt pair is a very special example of Ambro's quasilog varieties (see [Ambro 2003, Definition 4.1]). More precisely, if $[V, \mathcal{K}]$ is a pseudo-klt pair, then we can easily check that $[V, \mathcal{K}]$ is a *qlc pair*. See, for example, [Fujino 2011b, Definition 3.1]. For the details of the theory of quasilog varieties, see [Fujino 2009b].

Theorem 7.4. *Let $[W, \mathcal{K}]$ be a pseudo-klt pair. Assume that (V, B) is klt and W is projective or that W is affine. Then we can find an effective $\mathbb{Q}$-divisor B_W on W such that (W, B_W) is klt and that $\mathcal{K} \sim_{\mathbb{Q}} K_W + B_W$.*

Proof. When (X, B) is klt and W is projective, we can find B_W by [Ambro 2005a, Theorem 4.1]. When W is affine, this theorem follows from [Fujino 1999, Theorem 1.2]. □

Remark 7.5. It is conjectured that one can always find an effective $\mathbb{Q}$-divisor B_W on W such that (W, B_W) is klt and $\mathcal{K} \sim_{\mathbb{Q}} K_W + B_W$.

We now collect basic examples of pseudo-klt pairs.

Example 7.6. A klt pair is a pseudo-klt pair.

Example 7.7. Let $f : X \to W$ be a Mori fiber space. Then we can find a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor $\mathcal{K}$ on W such that $[W, \mathcal{K}]$ is a pseudo-klt pair. It is because we can find an effective $\mathbb{Q}$-divisor B on X such that $K_X + B \sim_{\mathbb{Q}, f} 0$ and (X, B) is klt.

Proposition 7.8. *An exceptional lc center W of an lc pair (X, B) is a pseudo-klt pair for some $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor $\mathcal{K}$ on W.*

Proof. We take a resolution $g : Y \to X$ such that $\mathrm{Exc}(g) \cup g_*^{-1}B$ has a simple normal crossing support. We put $K_Y + B_Y = g^*(K_X + B)$. Then $-B_Y = \mathbf{A}(X, B)_Y = \mathbf{A}_Y = \mathbf{A}^*_Y + \mathbf{N}_Y$, where $\mathbf{N}_Y = -\sum_{i=0}^{k} E_i$. Without loss of generality, we can assume that $f(E) = W$ and $E = E_0$. By shrinking X around W, we can assume that $\mathbf{N}_Y = -E$. Note that $R^1 g_*\mathcal{O}_Y(\lceil \mathbf{A}^*_Y \rceil - E) = 0$ by the Kawamata–Viehweg vanishing theorem since $\lceil \mathbf{A}^*_Y \rceil - E = K_Y + \{-\mathbf{A}^*_Y\} - g^*(K_X + B)$. Therefore, $g_*\mathcal{O}_Y(\lceil \mathbf{A}^*_Y \rceil) \simeq \mathcal{O}_X \to g_*\mathcal{O}_E(\lceil \mathbf{A}^*_Y|_E \rceil)$ is surjective. This implies that $g_*\mathcal{O}_E(\lceil \mathbf{A}^*_Y|_E \rceil) \simeq \mathcal{O}_W$. In particular, W is normal. If we put $K_E + B_E = (K_Y + B_Y)|_E$, then (E, B_E) is sub-klt and $\mathbf{A}^*_Y|_E = \mathbf{A}(E, B_E)_E = -B_E$. So, $g_*\mathcal{O}_E(\lceil \mathbf{A}(E, B_E) \rceil) = g_*\mathcal{O}_E(\lceil -B_E \rceil) \simeq \mathcal{O}_W$. Since $K_E + B_E = (K_Y + B_Y)|_E$ and $K_Y + B_Y = g^*(K_X + B)$, we can find a $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor $\mathcal{K}$ on W such that $K_E + B_E \sim_{\mathbb{Q}} g^*\mathcal{K}$. Therefore, W is a pseudo-klt pair. □

We make an important remark on *minimal* lc centers.

Remark 7.9 (Subadjunction for minimal lc center). Let (X, B) be a projective or affine lc pair and let W be a minimal lc center of the pair (X, B). Then we can find an effective $\mathbb{Q}$-divisor B_W on W such that (W, B_W) is klt and $K_W + B_W \sim_{\mathbb{Q}} (K_X + B)|_W$. For the details, see [Fujino and Gongyo 2012, Theorems 4.1, 7.1].

The following theorem is the Kawamata–Shokurov basepoint-free theorem for pseudo-klt pairs. We give a simple proof depending on Kawamata's positivity theorem. Although Theorem 7.10 seems to be contained in [Ambro 2003, Theorem 7.2], no proof is given there.

Theorem 7.10. *Let $[W, \mathcal{H}]$ be a pseudo-klt pair, let $\pi : W \to S$ be a proper morphism onto a variety S and let D be a π-nef Cartier divisor on W. Assume that $rD - \mathcal{H}$ is π-nef and π-big for some positive integer r. Then mD is π-generated for every $m \gg 0$.*

Proof. Without loss of generality, we can assume that S is affine. By the usual technique (see [Kawamata 1998, Theorem 1] and [Fujino 1999, Theorem 1.2]), we have

$$\mathcal{H} + \varepsilon(rD - \mathcal{H}) \sim_{\mathbb{Q}} K_W + \Delta_W$$

such that (W, Δ_W) is klt for some sufficiently small rational number $0 < \varepsilon \ll 1$ (see also [Kollár 2007, Theorem 8.6.1]). Then $rD - (K_W + \Delta_W) \sim_{\mathbb{Q}} (1 - \varepsilon)(rD - \mathcal{H})$, which is π-nef and π-big. Therefore, mD is π-generated for every $m \gg 0$ by the usual Kawamata–Shokurov basepoint-free theorem. □

The next theorem is the main theorem of this section. It is a generalization of Kawamata's theorem in [Kawamata 1985] (cf. Theorem 4.2) for pseudo-klt pairs.

Theorem 7.11. *Let $[W, \mathcal{H}]$ be a pseudo-klt pair and let $\pi : W \to S$ be a proper morphism onto a variety S. Assume the following conditions*:

(i) *H is a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on W.*

(ii) *$H - \mathcal{H}$ is π-nef and π-abundant.*

(iii) *$\kappa(W_\eta, (aH - \mathcal{H})_\eta) \geq 0$ and $\nu(W_\eta, (aH - \mathcal{H})_\eta) = \nu(W_\eta, (H - \mathcal{H})_\eta)$ for some $a \in \mathbb{Q}$ with $a > 1$, where η is the generic point of S.*

Then H is π-semiample.

Proof. By definition, there exists a proper surjective morphism $f : V \to W$ from a sub-klt pair (V, B). Without loss of generality, we can assume that V is nonsingular and $\operatorname{Supp} B$ is a simple normal crossing divisor. By definition, $f_*\mathcal{O}_V(\lceil -B \rceil) \simeq \mathcal{O}_W$. From now on, we assume that H is Cartier by replacing it with a multiple. Then $f_*\mathcal{O}_V(\lceil -B \rceil + jH_V) \simeq \mathcal{O}_W(jH)$ by the projection formula for every integer j, where $H_V = f^*H$. Pushing forward by π, we have

$$\begin{aligned} p_*\mathcal{O}_V(\lceil \mathbf{A}(V, B) \rceil + j\overline{H_V}) &= p_*\mathcal{O}_V(\lceil -B \rceil + jH_V) \\ &\simeq \pi_*\mathcal{O}_W(jH) \\ &\simeq p_*\mathcal{O}_V(jH_V) \end{aligned}$$

for every integer j, where $p=\pi\circ f$. This is nothing but the saturation condition Theorem 4.4(e). We put $L=H-\mathcal{K}$. We consider the Iitaka fibration with respect to L over S as in [Fujino 2011d, Proof of Theorem 1.1]. Then we obtain the following commutative diagram:

$$\begin{tikzcd}
V \arrow[r, equal] \arrow[d, "f"'] & V \arrow[d] \\
W \arrow[d, "\pi"'] & U \arrow[l, "\mu"'] \arrow[d, "g"] \\
S & Z \arrow[l, "\varphi"]
\end{tikzcd}$$

where $g: U\to Z$ is the Iitaka fibration over S and $\mu: U\to W$ is a birational morphism. Note that we can assume that $f: V\to W$ factors through U by blowing up V.

Lemma 7.12. $\operatorname{rank} h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)=1$, *where* $h: V\to U\to Z$.

Proof. This proof is essentially the same as that of [Fujino 2011d, Lemma 2.3]. First, we can assume that S is affine. Let A be an ample divisor on Z such that $h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)\otimes\mathcal{O}_Z(A)$ is φ-generated. We note that we can assume that $\mu^*L\sim_{\mathbb{Q}} g^*M$ since L is π-nef and π-abundant, where M is a φ-nef and φ-big $\mathbb{Q}$-divisor on Z. If we choose a large and divisible integer m, then $\mathcal{O}_Z(A)\subset\mathcal{O}_Z(mM)$. Thus

$$\begin{aligned}
&\varphi_*(h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)\otimes\mathcal{O}_Z(A))\\
&\subseteq \varphi_*(h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)\otimes\mathcal{O}_Z(mM))\\
&\simeq p_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil+m\overline{f^*L})\\
&\simeq \pi_*\mathcal{O}_W(mL)\\
&\simeq \varphi_*\mathcal{O}_Z(mM).
\end{aligned}$$

Therefore, $\operatorname{rank} h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)\le 1$. Since $\mathcal{O}_Z\subset h_*\mathcal{O}_V\subset h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)$, we obtain $\operatorname{rank} h_*\mathcal{O}_V(\lceil \mathbf{A}(V,B)\rceil)=1$ □

Note that $h: V\to Z$ is the Iitaka fibration with respect to f^*L over S. Assumption (c) in Theorem 4.4 easily follows from (iii). Thus, by Theorem 4.4, we have that H_V is p-semiample. Equivalently, H is π-semiample. □

The final theorem of this paper is a basepoint-free theorem for minimal lc centers.

Theorem 7.13. *Let* (X,B) *be a quasi-projective lc pair and let* W *be a minimal lc center of* (X,B). *Let* $\pi: W\to S$ *be a proper morphism onto a variety* S. *Assume the following conditions*:

(i) *H is a π-nef $\mathbb{Q}$-Cartier $\mathbb{Q}$-divisor on W.*

(ii) *$H-(K_X+B)|_W$ is π-nef and π-abundant.*

(iii) *$\kappa(W_\eta, (aH-(K_X+B))|_{W_\eta}) \geq 0$ and*

$$\nu(W_\eta, (aH-(K_X+B))|_{W_\eta}) = \nu(W_\eta, (H-(K_X+B))|_{W_\eta})$$

for some $a \in \mathbb{Q}$ with $a > 1$, where η is the generic point of S.

Then H is π-semiample.

Proof. Let $f : Y \to X$ be a dlt blow-up such that $K_Y + B_Y = f^*(K_X + B)$ (see, for example, [Fujino 2011a, Theorem 10.4] or [Fujino 2011e, Section 4]). Then we can take a minimal lc center Z of (Y, B_Y) such that $f(Z) = W$. Note that $K_Z + B_Z = (K_Y + B_Y)|_Z$ is klt. We also note that W is normal (see, for example, [Fujino 2011c, Theorem 2.4 (4)] or [Fujino 2011a, Theorem 9.1 (4)]). Let

$$f : Z \xrightarrow{g} V \xrightarrow{h} W$$

be the Stein factorization of $f : Z \to W$. Then $[V, h^*((K_X+B)|_W)]$ is a pseudo-klt pair by $g : (Z, B_Z) \to V$. We note that H is π-semiample if and only if h^*H is $\pi \circ h$-semiample. By Theorem 7.11, h^*H is semiample over S. This concludes the proof. $\square$

Acknowledgement

The first version of this paper was written in Nagoya in 2005 and posted on arXiv [Fujino 2005]. At the time, the author was partially supported by The Sumitomo Foundation and by the Grant-in-Aid for Young Scientists (A) 17684001 from JSPS. He revised this paper in Kyoto in 2011, where he was partially supported by the Inamori Foundation and by the Grant-in-Aid for Young Scientists (A) 20684001 from JSPS.

The author thanks the referee for comments.

References

[Ambro 2003] F. Ambro, "Quasi-log varieties", *Tr. Mat. Inst. Steklova* **240** (2003), 220–239. Reprinted in *Proc. Steklov Inst. Math.* **240**: (2003), 214-233. MR 2004f:14027 Zbl 1081.14021

[Ambro 2004] F. Ambro, "Shokurov's boundary property", *J. Differential Geom.* **67**:2 (2004), 229–255. In Russian; translated in *Proc. Steklov Inst. Math.* **240**: (2003), 214-233. MR 2006d:14033 Zbl 1097.14029

[Ambro 2005a] F. Ambro, "The moduli *b*-divisor of an lc-trivial fibration", *Compos. Math.* **141**:2 (2005), 385–403. MR 2006d:14015 Zbl 1094.14025

[Ambro 2005b] F. Ambro, "A semiampleness criterion", *J. Math. Sci. Univ. Tokyo* **12**:3 (2005), 445–466. MR 2007i:14018 Zbl 1121.14005

[Ambro 2007] F. Ambro, "Non-klt techniques", pp. 163–170 in *Flips for* 3*-folds and* 4*-folds*, edited by A. Corti, Oxford Lecture Ser. Math. Appl. **35**, Oxford Univ. Press, 2007. MR 2359347

[Cacciola 2011] S. Cacciola, "On the semiampleness of the positive part of CKM Zariski decompositions", preprint, 2011. arXiv 1011.6564v2

[Corti 2007] A. Corti, "3-fold flips after Shokurov", pp. 18–48 in *Flips for* 3*-folds and* 4*-folds*, edited by A. Corti, Oxford Lecture Ser. Math. Appl. **35**, Oxford Univ. Press, 2007. MR 2359339

[Corti et al. 2007] A. Corti (editor), *Flips for* 3*-folds and* 4*-folds*, Oxford Lecture Ser. Math. Appl. **35**, Oxford Univ. Press, 2007. MR 2359340

[Fujino 1999] O. Fujino, "Applications of Kawamata's positivity theorem", *Proc. Japan Acad. Ser. A Math. Sci.* **75**:6 (1999), 75–79. MR 2000f:14089 Zbl 0967.14012

[Fujino 2000] O. Fujino, "Base point free theorem of Reid–Fukuda type", *J. Math. Sci. Univ. Tokyo* **7**:1 (2000), 1–5. MR 2001c:14013 Zbl 0971.14009

[Fujino 2003] O. Fujino, "A canonical bundle formula for certain algebraic fiber spaces and its applications", *Nagoya Math. J.* **172** (2003), 129–171. MR 2005b:14024 Zbl 1072.14040

[Fujino 2005] O. Fujino, "A remark on the basepoint-free theorem", preprint, 2005. arXiv 0508.554v1

[Fujino 2007a] O. Fujino, "A transcendental approach to Kollár's injectivity theorem II", preprint, 2007. To appear in *J. reine angew. math.* arXiv 0705.1199v1

[Fujino 2007b] O. Fujino, "What is log terminal?", pp. 49–62 in *Flips for* 3*-folds and* 4*-folds*, edited by A. Corti, Oxford Lecture Ser. Math. Appl. **35**, Oxford Univ. Press, 2007. MR 2359341

[Fujino 2009a] O. Fujino, "Effective base point free theorem for log canonical pairs—Kollár type theorem", *Tohoku Math. J., (2)* **61**:4 (2009), 475–481. MR 2011c:14014 Zbl 1189.14025

[Fujino 2009b] O. Fujino, "Introduction to the log minimal model program for log canonical pairs", preprint, 2009. arXiv 0907.1506v1

[Fujino 2009c] O. Fujino, "On injectivity, vanishing and torsion-free theorems for algebraic varieties", *Proc. Japan Acad. Ser. A Math. Sci.* **85**:8 (2009), 95–100. MR 2011c:14053 Zbl 1189.14024

[Fujino 2010] O. Fujino, "Finite generation of the log canonical ring in dimension four", *Kyoto J. Math.* **50**:4 (2010), 671–684. MR 2012c:14032 Zbl 1210.14020

[Fujino 2011a] O. Fujino, "Fundamental theorems for the log minimal model program", *Publ. Res. Inst. Math. Sci.* **47**:3 (2011), 727–789. MR 2832805 Zbl 1234.14013

[Fujino 2011b] O. Fujino, "Introduction to the theory of quasi-log varieties", pp. 289–303 in *Classification of algebraic varieties*, edited by C. Faber et al., Eur. Math. Soc., Zürich, 2011. MR 2012f:14025 Zbl 1213.14030

[Fujino 2011c] O. Fujino, "Non-vanishing theorem for log canonical pairs", *J. Algebraic Geom.* **20**:4 (2011), 771–783. MR 2819675 Zbl 05955440

[Fujino 2011d] O. Fujino, "On Kawamata's theorem", pp. 305–315 in *Classification of algebraic varieties*, edited by C. Faber et al., Eur. Math. Soc., Zürich, 2011. MR 2012c:14011 Zbl 1213.14015

[Fujino 2011e] O. Fujino, "Semi-stable minimal model program for varieties with trivial canonical divisor", *Proc. Japan Acad. Ser. A Math. Sci.* **87**:3 (2011), 25–30. MR 2802603 Zbl 1230.14016

[Fujino and Gongyo 2011] O. Fujino and Y. Gongyo, "Log pluricanonical representations and abundance conjecture", preprint, 2011. arXiv 1104.0361

[Fujino and Gongyo 2012] O. Fujino and Y. Gongyo, "On canonical bundle formulas and subadjunctions", *Michigan Math. J.* **61**:2 (2012), 255–264.

[Fukuda 1996] S. Fukuda, "On base point free theorem", *Kodai Math. J.* **19**:2 (1996), 191–199. MR 97f:14004 Zbl 0860.14005

[Fukuda 2002] S. Fukuda, "On numerically effective log canonical divisors", *Int. J. Math. Math. Sci.* **30**:9 (2002), 521–531. MR 2003e:14010 Zbl 1058.14027

[Gongyo 2010] Y. Gongyo, "Abundance theorem for numerically trivial log canonical divisors of semi-log canonical pairs", preprint, 2010. To appear in *J. Algebraic Geom.* arXiv 1005.2796v2

[Iskovskikh 2003] V. A. Iskovskikh, "b-divisors and Shokurov functional algebras", *Tr. Mat. Inst. Steklova* **240** (2003), 8–20. In Russian; translated in *Proc. Steklov Inst. Math.* **240**:1 (2003), 4–15. MR 2004g:14019 Zbl 1081.14022

[Kawamata 1985] Y. Kawamata, "Pluricanonical systems on minimal algebraic varieties", *Invent. Math.* **79**:3 (1985), 567–588. MR 87h:14005 Zbl 0593.14010

[Kawamata 1998] Y. Kawamata, "Subadjunction of log canonical divisors. II", *Amer. J. Math.* **120**:5 (1998), 893–899. MR 2000d:14020 Zbl 0919.14003

[Kawamata 2010] Y. Kawamata, "Variation of mixed Hodge structures and the positivity for algebraic fiber spaces", preprint, 2010. arXiv 1008.1489v1

[Kawamata 2011a] Y. Kawamata, "Hodge theory on generalized normal crossing varieties", preprint, 2011. arXiv 1104.0524v1

[Kawamata 2011b] Y. Kawamata, "Semipositivity theorem for reducible algebraic fiber spaces", *Pure Appl. Math. Q.* **7**:4 (2011), 1427–1447.

[Kawamata et al. 1987] Y. Kawamata, K. Matsuda, and K. Matsuki, "Introduction to the minimal model problem", pp. 283–360 in *Algebraic geometry* (Sendai, 1985), edited by T. Oda, Adv. Stud. Pure Math. **10**, North-Holland, Amsterdam, 1987. MR 89e:14015 Zbl 0672.14006

[Kollár 1993] J. Kollár, "Effective base point freeness", *Math. Ann.* **296**:4 (1993), 595–605. MR 94f:14004 Zbl 0818.14002

[Kollár 2007] J. Kollár, "Kodaira's canonical bundle formula and adjunction", pp. 134–162 in *Flips for* 3*-folds and* 4*-folds*, edited by A. Corti, Oxford Lecture Ser. Math. Appl. **35**, Oxford Univ. Press, 2007. MR 2359346

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, II: Positivity for vector bundles, and multiplier ideals*, Ergebnisse der Math. (3) **49**, Springer, Berlin, 2004. MR 2005k:14001b Zbl 1093.14500

[Nakayama 1986] N. Nakayama, "Invariance of the plurigenera of algebraic varieties under minimal model conjectures", *Topology* **25**:2 (1986), 237–251. MR 87g:14034 Zbl 0596.14026

[Shokurov 2003] V. V. Shokurov, "Prelimiting flips", *Tr. Mat. Inst. Steklova* **240** (2003), 82–219. Reprinted in *Proc. Steklov Inst. Math.* **240**: (2003), 75–213. MR 2004k:14024 Zbl 1082.14019

Communicated by Shigefumi Mori
Received 2011-04-30 Revised 2011-05-18 Accepted 2011-06-20

fujino@math.kyoto-u.ac.jp *Department of Mathematics, Faculty of Science, Kyoto University, Sakyo-ku Kitashirakawa, Oiwake-cho, Kyoto 606-8502, Japan*

ALGEBRA AND NUMBER THEORY 6:4(2012)

msp

Realizing large gaps in cohomology for symmetric group modules

David J. Hemmer

Using results of the author with Cohen and Nakano, we find examples of Young modules Y^λ for the symmetric group Σ_d for which the Tate cohomology $\widehat{\mathrm{H}}^i(\Sigma_d, Y^\lambda)$ does not vanish identically, but vanishes for approximately $\frac{1}{3}d^{3/2}$ consecutive degrees. We conjecture these vanishing ranges are maximal among all Σ_d-modules with nonvanishing cohomology. The best known upper bound on such vanishing ranges stands at $(d-1)^2$, due to work of Benson, Carlson and Robinson. Particularly striking, and perhaps counterintuitive, is that these Young modules have maximum possible complexity.

1. Introduction

Let G be a finite group and k an algebraically closed field of characteristic p. If G contains an element x of order p such that the centralizer $C_G(x)$ is not p-nilpotent, then a result of Benson [1995] guarantees the existence of a nonprojective kG-module M in the principal block such that the cohomology $\mathrm{H}^*(G, M)$ is identically zero. For the remaining principal block modules, those with nonvanishing cohomology, one might ask for the smallest degree that is nonzero, or the number of consecutive degrees in which the cohomology vanishes. In [Benson et al. 1990], Benson, Carlson and Robinson gave an upper bound $r = r(G)$ on the number of consecutive i for which the cohomology $\mathrm{H}^i(G, M)$ can vanish, without being identically zero:

Theorem 1.1 [Benson et al. 1990, Theorem 2.4]. *Given a finite group G, there exists a positive integer r such that for any commutative ring R of coefficients and any RG-module M, if $\hat{\mathrm{H}}^i(G, M) = 0$ for $r + 1$ consecutive values of i then $\hat{\mathrm{H}}^i(G, M) = 0$ for all i positive and negative.*

The $\hat{\mathrm{H}}$ above denotes Tate cohomology, which agrees with the ordinary cohomology in positive degrees. The proof of Theorem 1.1 expresses r in terms of the

Research of the author was supported in part by NSF grant DMS-0808968.
MSC2010: 20C30.
Keywords: symmetric group, cohomology, Young module.

degrees of a set of homogenous generators for the cohomology ring of G. However, there is no expectation that this r should be the best possible bound.

There do not appear to be any examples in the literature demonstrating large gaps in cohomology, or determining the smallest possible value of r for particular groups. This is not surprising, as calculating $\mathrm{H}^*(G, M)$ is generally difficult. In [Cohen et al. 2010] the author, with Cohen and Nakano, obtained some very general results when M is a Young module Y^λ for the symmetric group Σ_d. The goal of this paper is to use these results to find very large gaps in Young module cohomology. For certain partitions $\lambda \vdash d$ in characteristic two, we find the minimal $i \geq 0$ with $\mathrm{H}^i(\Sigma_d, Y^\lambda) \neq 0$. These gaps turn out to be the largest possible among all Young modules, and come "close" to realizing the value of r arising from Theorem 1.1. Remarkably the Young modules with the largest vanishing ranges also have maximum possible complexity. That is the dimensions in a minimal projective resolution grow as quickly as possible. See [Benson 1998, p.153] for the precise definition of complexity.

2. Computing Young module cohomology

In this section we recall results from [Cohen et al. 2010] on computing Young module cohomology. Let $V \cong k^d$ be the natural module for the general linear group $G := \mathrm{GL}_d(k)$. For a partition $\lambda \vdash d$, let $L(\lambda)$ denote the simple G-module of highest weight λ, and let Y^λ denote the Young module for Σ_d. We denote by $\trianglerighteq$ the usual dominance order on partitions of d, and by λ' the transpose or conjugate partition. Definitions and information on all these modules can be found in [Martin 1993].

The commuting actions of G and Σ_d on $V^{\otimes d}$ give the homology $\mathrm{H}_i(\Sigma_d, V^{\otimes d})$ the structure of a G-module. The composition multiplicities of this G-module are related to the dimensions of Young module cohomology in the following way. Let $[M : S]$ denotes the multiplicity of a simple module S in a composition series of M.

Theorem 2.1 [Doty et al. 2004, Proposition 2.6B].

$$\dim_k \mathrm{H}^i(\Sigma_d, Y^\lambda) = [\mathrm{H}_i(\Sigma_d, V^{\otimes d}) : L(\lambda)], \ \ i \geq 0.$$

Theorem 2.1 indicates that determining the simple constituents of $\mathrm{H}_*(\Sigma_d, V^{\otimes d})$ as a graded G-module allows one to calculate Young module cohomology in all degrees. It turned out to be easier to study this for all d simultaneously, using methods from algebraic topology. In [Cohen et al. 2010, Theorem 8.1.4] the algebra $\oplus_{d\geq 0} \mathrm{H}_*(\Sigma_d, V^{\otimes d})$ is described as a G-module. It is a polynomial algebra, tensored with an exterior algebra if p is odd. Each generator belongs to a certain G-module direct summand, and this summand belongs to $\mathrm{H}_i(\Sigma_d, V^{\otimes d})$ for a particular i and d.

The G-modules that occur are described below. For a G-module M let $M^{(a)}$ denote the a-th Frobenius twist of M (see [Jantzen 2003, p. 132]), and let $S^a(M)$ and $\Lambda^a(M)$ denote respectively the a-th symmetric and exterior power of M.

Theorem 2.2 [Cohen et al. 2010, Corollary 8.2.1]. *In characteristic two, the G-module $\mathrm{H}_*(\Sigma_d, V^{\otimes d})$ is a direct sum of modules of the form*

$$S^{a_0}(V) \otimes S^{a_1}(V^{(c_1)}) \otimes \cdots \otimes S^{a_s}(V^{(c_s)}) \tag{2-1}$$

where each $a_i \geq 0$, $c_i > 0$ and $d = a_0 + \sum_{j=1}^{s} a_j 2^{c_j}$.

In odd characteristic the G-module $\mathrm{H}_*(\Sigma_d, V^{\otimes d})$ is a direct sum of modules of the form

$$S^{a_0}(V) \otimes S^{a_1}(V^{(c_1)}) \otimes \cdots \otimes S^{a_s}(V^{(c_s)}) \otimes \Lambda^{d_1}(V^{(e_1)}) \otimes \cdots \otimes \Lambda^{d_t}(V^{(e_t)}) \tag{2-2}$$

where each $a_i \geq 0$, each c_i, d_i, $e_i > 0$ and where $d = a_0 + \sum_{j=1}^{s} a_j p^{c_j} + \sum_{j=1}^{t} d_j p^{e_j}$.

Each summand in (2-1) or (2-2) occurs in $\mathrm{H}_i(\Sigma_d, V^{\otimes d})$ for a single value of d but for infinitely many different degrees i, for a description see Theorem 8.1.4 in [Cohen et al. 2010] or the special cases below, which are all we will use. To compute a particular $\mathrm{H}^i(\Sigma_d, Y^\lambda)$ one must first determine the (finitely many) summands which contribute to this d and i, and then compute the multiplicities of $L(\lambda)$ in each summand. In the next section we will let $p = 2$ and make a strategic choice for λ. For these λ we can determine precisely the summand (2-1) of smallest degree which contains $L(\lambda)$ as a composition factor, and thus determine the initial vanishing range. In Section 4 we use these computations to produce Young modules with very large gaps in cohomology. In the final section we discuss the situation in odd characteristic, and present a few open problems.

3. Initial vanishing ranges in characteristic two

In this section assume $p = 2$. Notation such as $(2^3, 1^2)$ will be shorthand for the partition $(2, 2, 2, 1, 1)$, not the partition $(8, 1)$. It is clear from (2-1) that understanding the composition factors of $S^a(V)$ is necessary for computing Young module cohomology (but not sufficient, as one must also decompose the tensor products).

Fortunately, Doty [1985] has determined the entire submodule structure for $S^a(V)$. The composition factors all occur with multiplicity at most one, and have a particularly nice form in characteristic two:

Proposition 3.1 [Doty 1985]. (See also [Cohen et al. 2010, Proposition 12.2.1].) *Let $\lambda \vdash s$ have a 2-adic expansion*

$$\lambda = \sum_{i=0}^{m} 2^i \lambda_{(i)}$$

where each $\lambda_{(i)}$ is 2-restricted. Then $L(\lambda)$ is a constituent of $S^s(V)$ if and only if each $\lambda_{(i)}$ is of the form (1^{a_i}) for $a_i \geq 0$.

Let $\mu = (\mu_1, \mu_2, \ldots, \mu_r) \vdash d$ be 2-restricted. Set $\mu' = ((\mu')_1, (\mu')_2, \ldots, (\mu')_{\mu_1})$. We will compute the first i such that $\mathrm{H}^i(\Sigma_{2d}, Y^{2\mu})$ is nonzero, and see that a particular such μ will maximize the initial vanishing range.

Since μ is 2-restricted, the 2-adic expansion of 2μ is just 2μ. So Steinberg's tensor product theorem (STPT) [Jantzen 2003, II.3.17] implies the summands from (2-1) with any $c_i > 1$ do not have $L(2\mu)$ as a composition factor. So to compute $\mathrm{H}^i(\Sigma_{2d}, Y^{2\mu})$ we must determine the multiplicity of $L(2\mu)$ in summands of the form

$$S^a(V) \otimes S^{a_1}(V^{(1)}) \otimes \cdots \otimes S^{a_s}(V^{(1)}) \cong S^a(V) \otimes S^{\tau}(V^{(1)}) \tag{3-1}$$

where we can assume without loss that $a_i \geq a_{i+1}$, so $\tau = (a_1, a_2, \ldots, a_s) \vdash d - \frac{a}{2}$.

Analysis just as in Section 10 of [Cohen et al. 2010] shows that a summand of the form (3-1) corresponds to monomials in the polynomial algebra of the form

$$v^a \cdot Q_{i_1}^{a_1}(v) \cdots Q_{i_s}^{a_s}(v),$$

for distinct i_t. By [Cohen et al. 2010, Theorem 8.1.4(a)], such a summand contributes to the cohomology in degree $a_1 i_1 + a_2 i_2 + \cdots + a_s i_s$. To determine the smallest i with $\mathrm{H}^i(\Sigma_{2d}, Y^{2\mu}) \neq 0$ we must first determine which modules (3-1) contain $L(2\mu)$ as a composition factor. Then for each we must determine the smallest possible corresponding degree where the summand can occur. Our assumption on μ limits how $L(2\mu)$ can arise as a composition factor in (3-1):

Proposition 3.2. *Let $2\mu \vdash 2d$ where μ is 2-restricted. Then $\mathrm{H}^i(\Sigma_{2d}, Y^{2\mu}) \neq 0$ if and only if there exists an integer $a \geq 0$, a partition $\tau = (a_1, a_2, \ldots, a_s) \vdash d - a$ and integers $\{i_t > 0\}$ such that*

(i) $i = a_1 i_1 + a_2 i_2 + \cdots + a_s i_s$,

(ii)
$$[L(2^a) \otimes L(2^{a_1}) \otimes \cdots \otimes L(2^{a_s}) : L(2\mu)] \neq 0.$$

Proof. By Proposition 3.1 and the STPT, the composition factors of $S^m(V)$ are all of the form

$$L(1^{c_0}) \otimes L(2^{c_1}) \otimes L(4^{c_2}) \otimes \cdots .$$

But 2μ is its own 2-adic expansion, so any $L(2\mu)$ occurring in (3-1) must arise as in part (2) by the STPT. The corresponding degree i follows from [Cohen et al. 2010, Theorem 8.1.4(a)]. □

Notice that the "if" part of the preceding result did not require μ be 2-restricted, a fact we will need later.

Now we want to find the smallest degree i where the cohomology $\mathrm{H}^i(\Sigma_{2d}, Y^{2\mu})$is nonzero. Since $a_1 \geq a_2 \geq \cdots$, it is clear from Proposition 3.2(1) that we should choose $i_t = t$ to minimize the degree i. The smallest nonzero degree is given in terms of the following function on partitions. Let $\rho = (\rho_1, \rho_2, \dots \rho_s) \vdash d$. Define

$$x(\rho) = \sum_{l=1}^{s} (l-1)\rho_l.$$

The following easy lemma is left to the reader:

Lemma 3.3. *Suppose $\lambda \trianglerighteq \mu$. Then $x(\lambda) \le x(\mu)$. If $\lambda \ne \mu$ the inequality is strict.*

We can now determine the first nonvanishing degree for $\mathrm{H}^*(\Sigma_{2d}, Y^{2\mu})$.

Theorem 3.4. *Let $\mu \vdash d$ be arbitrary. Then*:

(i) $\mathrm{H}^{x(\mu')}(\Sigma_{2d}, Y^{2\mu}) \ne 0$.

(ii) *If μ is 2-restricted, then*

$$\dim \mathrm{H}^i(\Sigma_{2d}, Y^{2\mu}) = \begin{cases} 0 & \text{for } 0 \le i < x(\mu'), \\ 1 & \text{for } i = x(\mu'). \end{cases}$$

Proof. For convenience let $\tau = \mu'$. Observe that

$$\mu = (1^{\tau_1}) + (1^{\tau_2}) + \cdots + (1^{\tau_{\mu_1}}).$$

Then $L(2^{\tau_1}) \otimes L(2^{\tau_2}) \otimes \cdots \otimes L(2^{\tau_{\mu_1}})$ has highest weight 2μ with multiplicity one, so

$$[L(2^{\tau_1}) \otimes L(2^{\tau_2}) \otimes \cdots \otimes L(2^{\tau_{\mu_1}}) : L(2\mu)] = 1. \tag{3-2}$$

Thus

$$[S^{2\tau_1}(V) \otimes S^{\tau_2}(V^{(1)}) \otimes S^{\tau_3}(V^{(1)}) \cdots \otimes S^{\tau_{\mu_1}}(V^{(1)}) : L(2\mu)] \ge 1. \tag{3-3}$$

Choosing $a = \tau_1$ and $i_t = t$, the proof of Proposition 3.2 tells us that

$$\mathrm{H}^{x(\mu')}(\Sigma_{2d}, Y^{2\mu}) \ne 0.$$

(The “if” part did not require μ be 2-restricted.)

Now suppose further that μ is 2-restricted, and consider Proposition 3.2. Suppose

$$[L(2^a) \otimes L(2^{a_1}) \otimes \cdots \otimes L(2^{a_s}) : L(2\mu)] \ne 0.$$

In order to minimize the degree i it is clear from Proposition 3.2(2) that we may assume $a \ge a_1 \ge a_2 \ge \cdots \ge a_s$. Then $\rho := (a, a_1, a_2, \dots, a_s) \vdash d$, and by Proposition 3.2(1), the corresponding cohomological degree is $x(\rho)$. Since $L(2^a) \otimes L(2^{a_1}) \otimes \cdots \otimes L(2^{a_s})$ has highest weight $2\rho'$ then $\rho' \trianglerighteq \mu$, and thus $\mu' \trianglerighteq \rho$. When $\rho = \mu'$ we get a single copy of $L(2\mu)$ as above, contributing to degree $x(\mu')$. Otherwise $\mu' \triangleright \rho$. Then Lemma 3.3 implies $x(\mu') < x(\rho)$, so $x(\mu')$ is the smallest degree with nonzero cohomology. So the cohomology is one-dimensional in degree $x(\mu')$ and zero in smaller degrees. □

4. Mind the gap

In this section we apply Theorem 3.4 to find large gaps in cohomology. For comparison we first compute the smallest currently known $r(\Sigma_d)$ which satisfies Theorem 1.1.

A faithful complex representation of a group G gives rise to an embedding into a compact unitary group $G \hookrightarrow U(n)$. The cohomology of the classifying space $BU(n)$ is a polynomial ring on generators in degrees $2, 4, 6, \ldots, 2n$ (see [Benson 1998, Section 2.6]). The value of r coming from these generators by the construction in [Benson et al. 1990] is $1+3+5+\cdots+(2n-1)=n^2$. Thus if G has a faithful representation of dimension n, one can take $r=r(G)=n^2$ in Theorem 1.1, see [Benson 1998, Sections 5.14–15] for details.

The smallest faithful irreducible $\mathbb{C}\Sigma_d$ module is $d-1$ dimensional, so one can take $r(\Sigma_d)=(d-1)^2$, and this is the smallest known bound. To find Young modules with large vanishing ranges, Theorem 3.4(2) suggests finding p-restricted μ with $x(\mu')$ as large as possible. In this section we show careful choice of Young module can realize gaps on the order of $\frac{1}{3}d^{3/2}$.

Fix $n \geq 1$ and define $\rho_n=(n, n-1, n-2, \ldots, 2, 1) \vdash \frac{1}{2}(n^2+n)$. Notice that

$$\rho_n=(\rho_n)'=(1^n)+(1^{n-1})+\cdots+(1^2)+(1).$$

One easily computes that

$$x(\rho_n)=\frac{n^3-n}{6}. \tag{4-1}$$

Proposition 4.1. *Let $p=2$ and $\rho_n \vdash (n^2+n)/2$ be as above. Then*:

$$\dim \hat{\mathrm{H}}^i(\Sigma_{n^2+n}, Y^{2\rho_n})=\begin{cases}0 & \text{for } -\frac{1}{6}(n^3-n)<i<\frac{1}{6}(n^3-n),\\ 1 & \text{for } i=\pm\frac{1}{6}(n^3-n).\end{cases}$$

Proof. Since ρ_n is 2-restricted, we can apply Theorem 3.4(2) and (4-1). The extension to negative degrees comes from Tate duality, using the fact that Young modules are self-dual. □

Proposition 4.1 shows that for $d=n^2+n$, the best possible $r(\Sigma_d)$ is at least $\frac{1}{3}(n^3-n)$.

Example 4.2. Let $\lambda=(28, 26, 24, \ldots, 6, 4, 2) \vdash 210$. Then

$$\mathrm{H}^i(\Sigma_{210}, Y^\lambda)=\begin{cases}0 & \text{if } -455<i<455,\\ k & \text{if } i=455.\end{cases}$$

It follows from [Hemmer and Nakano 2002, Theorem 3.3.2] that the Y^λ in Example 4.2 has complexity 105, the maximum possible among Σ_{210}-modules. This means the dimension of the module P_i in the minimal projective resolution

$P_* \to k$ of the trivial module grows like a polynomial of degree 104 in i. However it is not until P_{455} that the projective cover $P(k)$ makes its first appearance!

Remark 4.3. Proposition 4.1 applies to Σ_{2d} where d is a triangular number $T(n) = \frac{n^2+n}{2}$. For arbitrary d one can still choose a 2-restricted μ maximizing $x(\mu')$ in a similar way. Write $d = T(n) + a$ for $0 \le a < n+1$ and choose

$$\mu = (n, n-1, \ldots, a+1, a, a, a-1, a-2, \ldots, 2, 1). \tag{4-2}$$

One still has $x((2\mu)')$ asymptotic to a constant times n^3.

So there is a constant c so that for arbitrary d we can obtain Young modules in characteristic two with cohomology vanishing for the first $cd^{3/2}$ degrees.

5. Odd primes and further directions

Since Theorem 1.1 gives a bound $r(G)$ independent of the characteristic, we have focused on $p = 2$ which gives the cleanest results. For an arbitrary prime one can still achieve gaps that are a constant times $d^{3/2}$ in length, using $\mu = p(p-1)\rho$, although the answer is messier, and involves polynomials in p. For example the nice compact form for $x(\rho_n)$ in (4-1) becomes replaced by

$$(p-1)[n(2p-3) + (n-1)(4p-5) + \cdots + 1(2n(p-1)-1)].$$

The corresponding result, which we state without proof, is this:

Proposition 5.1. *Let $d = \frac{1}{2}p(p-1)(n^2+n)$. Let $\mu = p(p-1)\rho_n \vdash d$. Then there is a constant $c(p)$ and a polynomial $p(n) = c(p)n^3 + an^2 + bn$ such that*

$$\mathrm{H}^i(\Sigma_d, Y^\mu) = 0 \text{ if } -p(n) < i < p(n)$$

So once again we have find an $r(\Sigma_d)$ asymptotic to a constant times $d^{3/2}$. The function $c(p)$ is decreasing, so the best estimates for $r(\Sigma_d)$ come from the $p = 2$ case. This might lead one to make a wild conjecture:

Conjecture 5.2. Let $d = \frac{1}{2}p(p-1)(n^2+n)$. Let $\mu = p(p-1)\rho_n \vdash d$. Among all Σ_d modules in the principal block with nonvanishing cohomology, the Young module Y^μ has the largest gap in cohomology, and thus determines the best possible r in Theorem 1.1. For d not of this form, a similar choice, in the spirit of (4-2), for μ achieves the maximal gap.

There are many problems which remain, although it isn't clear one should expect nice answers to any of them. For example one might find the smallest positive i with $\mathrm{H}^i(\Sigma_d, Y^\lambda) \ne 0$. The corresponding problem for simple modules is a subject of active research, for example for groups of Lie type. A first step would be to generalize Doty's work from the $\mu = (d)$ case to something more general:

Problem 5.3. Given $\lambda \vdash d$, find the maximal $\mu \vdash d$ such that $[S^{\mu}(V) : L(\lambda)] \neq 0$.

Determining the λ for which $\mu = (d)$ is just Doty's result on the composition factors of $S^d(V)$. At the opposite extreme, such a μ always exists, because $S^{(1^d)}(V) \cong V^{\otimes d}$ and each $L(\mu)$ occurs as a composition factor of $V^{\otimes d}$.

Finally we observe that the partition μ appearing in Proposition 5.1 is just the twist of the Steinberg weight (see [Jantzen 2003, p. 199]), but there seems to be no representation-theoretic interpretation of this fact.

References

[Benson 1995] D. J. Benson, "Cohomology of modules in the principal block of a finite group", *New York J. Math.* **1** (1995), 196–205. MR 96h:20095 Zbl 0879.20004

[Benson 1998] D. J. Benson, *Representations and cohomology, II: Cohomology of groups and modules*, 2nd ed., Cambridge Studies in Advanced Mathematics **31**, Cambridge University Press, 1998. MR 99f:20001b Zbl 0908.20002

[Benson et al. 1990] D. J. Benson, J. F. Carlson, and G. R. Robinson, "On the vanishing of group cohomology", *J. Algebra* **131**:1 (1990), 40–73. MR 91c:20073 Zbl 0697.20043

[Cohen et al. 2010] F. R. Cohen, D. J. Hemmer, and D. K. Nakano, "On the cohomology of Young modules for the symmetric group", *Adv. Math.* **224**:4 (2010), 1419–1461. MR 2011f:20118 Zbl 1206.20057

[Doty 1985] S. R. Doty, "The submodule structure of certain Weyl modules for groups of type A_n", *J. Algebra* **95**:2 (1985), 373–383. MR 86j:20035 Zbl 0577.20031

[Doty et al. 2004] S. R. Doty, K. Erdmann, and D. K. Nakano, "Extensions of modules over Schur algebras, symmetric groups and Hecke algebras", *Algebr. Represent. Theory* **7**:1 (2004), 67–100. MR 2005e:20004 Zbl 1084.20004

[Hemmer and Nakano 2002] D. J. Hemmer and D. K. Nakano, "Support varieties for modules over symmetric groups", *J. Algebra* **254**:2 (2002), 422–440. MR 2003k:20013 Zbl 1028.20009

[Jantzen 2003] J. C. Jantzen, *Representations of algebraic groups*, 2nd ed., Mathematical Surveys and Monographs **107**, American Mathematical Society, Providence, RI, 2003. MR 2004h:20061 Zbl 1034.20041

[Martin 1993] S. Martin, *Schur algebras and representation theory*, Cambridge Tracts in Mathematics **112**, Cambridge University Press, 1993. MR 95f:20071 Zbl 0802.20011

Communicated by David Benson
Received 2011-06-20 Accepted 2011-08-13

dhemmer@math.buffalo.edu *Department of Mathematics, University at Buffalo, SUNY, College of Arts and Sciences, The State University of New York, Buffalo, NY 14260-2900, United States*
http://math.buffalo.edu/~dhemmer/

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. 4 2012