Algebra & Number Theory

Volume 6
2012
No. 3

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 6:3(2012)

msp

The image of complex conjugation in l-adic representations associated to automorphic forms

Richard Taylor

If F^+ is a totally real field, if n is an odd integer and if Π is a regular, algebraic, essentially self-dual, cuspidal automorphic representation of $\mathrm{GL}_n(\mathbb{A}_{F^+})$, then we calculate the image of any complex conjugation under the l-adic representations $r_{l,\iota}(\Pi)$ associated to Π.

Introduction

Let F^+ denote a totally real number field and fix an isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. It is known that to a regular, algebraic, essentially self-dual, cuspidal automorphic representation Π of $\mathrm{GL}_n(\mathbb{A}_{F^+})$ one can associate a continuous semisimple Galois representation

$$r_{l,\iota}(\Pi) : \mathrm{Gal}(\overline{F}^+/F^+) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l).$$

(For the definition of "regular, algebraic, essentially self-dual, cuspidal" see the start of Section 1.) This representation is known to be de Rham and its Hodge–Tate numbers are known. (They can be simply calculated from the infinitesimal character of π_∞.) For all finite places v of F^+ not dividing l one can calculate the Frobenius semisimplification of the restriction of $r_{l,\iota}(\Pi)$ to a decomposition group above v in terms of π_v via the local Langlands correspondence. This uniquely (in fact, over) determines $r_{l,\iota}(\Pi)$. (See [Shin 2011; Clozel et al. 2011; Caraiani 2010; Chenevier and Harris 2011].) The representation $r_{l,\iota}(\Pi)$ is conjectured to be irreducible. This is known if Π is discrete series at some finite place [Taylor and Yoshida 2007]. Moreover $r_{l,\iota}(\Pi)^\vee \cong r_{l,\iota}(\Pi) \otimes \mu$ for some character μ of $\mathrm{Gal}(\overline{F}^+/F^+)$ which is either totally odd (takes the value -1 on all complex conjugations) or totally even (takes the value $+1$ on all complex conjugations).

Frank Calegari raised the question as to whether, for an infinite place v of F^+ one can calculate the conjugacy class of $r_{l,\iota}(\Pi)(c_v)$, where $c_v \in \mathrm{Gal}(\overline{F}^+/F^+)$ is a

The author is partially supported by NSF Grant DMS-0600716.
MSC2000: 11F80.
Keywords: Galois representations.

complex conjugation for v. This conjugacy class has order two, so it is semisimple with eigenvalues ± 1. The problem is to determine how many $+1$'s and how many -1's occur. Because Π was assumed to be regular, we expect that the number of $+1$'s and -1's differ by at most one:

$$|\mathrm{tr}\, r_{l,\iota}(\Pi)(c_v)| \leq 1.$$

As we know the determinant of $r_{l,\iota}(\Pi)$ this would completely determine the conjugacy class of $r_{l,\iota}(\Pi)(c_v)$.

If μ is totally odd then [Bellaïche and Chenevier 2011] shows that n is even and that $r_{l,\iota}(\Pi)$ preserves an alternating pairing up to multiplier μ. In this case, because $\mathrm{GSp}_n(\overline{\mathbb{Q}}_l)$ has a unique conjugacy class of elements of order two and multiplier -1, we see that $\mathrm{tr}\, r_{l,\iota}(\Pi)(c_v) = 0$ for all $v|\infty$. So the problem lies in the case that μ is totally even, i.e., that $r_{l,\iota}(\Pi)$ preserves an orthogonal pairing up to multiplier μ.

In this paper we will prove this conjecture in the case n is odd:

Proposition 1. *Suppose that F^+ is a totally real field, that n is an odd positive integer and that Π a regular, algebraic, essentially self-dual, cuspidal automorphic representation Π of $\mathrm{GL}_n(\mathbb{A}_{F^+})$. Suppose also that $r_{l,\iota}(\Pi)$ is irreducible. If*

$$c \in \mathrm{Gal}(\bar{F}^+/F^+)$$

is a complex conjugation (for some embedding $\bar{F}^+ \hookrightarrow \mathbb{C}$) then

$$|\mathrm{tr}\, r_{l,\iota}(\Pi)(c)| \leq 1.$$

We believe that essentially the same method works if n is even and Π is discrete series at a finite place, though we haven't taken the trouble to write the argument down in this case. (One would work with the construction of $r_{l,\iota}(\Pi)$ given in [Harris and Taylor 2001] rather than that given in [Shin 2011].) However we do not see how to treat the general case when n is even. When $r_{l,\iota}(\Pi)$ is reducible one can calculate the trace of $r(c)$ for some representation of r of $\mathrm{Gal}(\bar{F}^+/F^+)$ with the same restriction to $\mathrm{Gal}(\bar{F}^+/F)$, but this does not seem to be very helpful.

The construction of $r_{l,\iota}(\Pi)$ is via piecing together twists of representations of $\mathrm{Gal}(\bar{F}^+/F)$ which arise in the cohomology of unitary group Shimura varieties, as F runs over certain imaginary CM fields. For none of these twisted restrictions does complex conjugation make sense. For an infinite place of F one can assign a natural sign to the representations of $\mathrm{Gal}(\bar{F}^+/F)$ that arise in the cohomology of these Shimura varieties, because they are essentially conjugate self-dual. (See [Clozel et al. 2008] or [Bellaïche and Chenevier 2011].) As Calegari has stressed this sign is not related to the image of complex conjugation in our representation of $\mathrm{Gal}(\bar{F}^+/F^+)$. This latter image only makes sense for the Galois representations coming from certain automorphic forms on the unitary groups, namely those that arise from an automorphic form on $\mathrm{GL}_n(\mathbb{A}_{F^+})$ by some functoriality.

In the case that n is odd the unitary groups employed by Shin [2011] have rank n and we are able to use the moduli theoretic interpretation of its Shimura variety to write descent data to the maximal totally real subfield of F. This descent data does not commute with the action of the finite adelic points of the unitary group. However in the special case of an automorphic representation π which arises by functoriality from an automorphic form on GL_n over a totally real field we are able to show that, up to twist, this descent data preserves the π^∞ isotypical component of the cohomology, and hence gives a geometric realization of $r_{l,\iota}(\Pi)(c_v)$. Because of its geometric construction, $r_{l,\iota}(\Pi)(c_v)$ also makes sense in the world of variations of Hodge structures. Finally we can appeal to the fact that the Hodge structure corresponding to $r_{l,\iota}(\Pi)$ is regular (i.e., each $h^{p,q} \leq 1$) to show that $|\mathrm{tr}\, r_{l,\iota}(\Pi)(c_v)| \leq 1$.

In the case that n is even and Π is not discrete series at any finite place, [Shin 2011] realizes twists of $r_{l,\iota}(\Pi)|_{\mathrm{Gal}(\bar{F}^+/F)}$ in the cohomology of the Shimura varieties for unitary groups of rank $n+1$. One takes the π^∞ isotypic component of the cohomology for an unstable automorphic representation π of the unitary group, which one constructs from Π using the theory of endoscopy. In this case our descent data relates the π^∞ isotypic component of the cohomology, not to itself, but to a twist of the $(\pi')^\infty$ isotypic component for a second unstable automorphic representation π' of the unitary group also arising from Π. (This π' is not even nearly equivalent to a twist of π.) This does not seem to be helpful.

Notation. Let us establish some notation that we will use throughout the paper.

If ρ is a representation κ_ρ will denote its central character.

If F is a p-adic field with valuation v then F^{nr} will denote its maximal unramified extension and $\mathrm{Frob}_v \in \mathrm{Gal}(F^{\mathrm{nr}}/F)$ will denote geometric Frobenius. Moreover $\mathrm{Art}_F : F^\times \to \mathrm{Gal}(\bar{F}/F)^{\mathrm{ab}}$ will denote the Artin map (normalized to take uniformizers to geometric Frobenius elements). Suppose that $V/\overline{\mathbb{Q}}_l$ is a finite-dimensional vector space and that

$$r : \mathrm{Gal}(\bar{F}/F) \to \mathrm{GL}(V)$$

is a continuous homomorphism. If either $l \neq p$ or $l = p$ and V is de Rham (i.e., $\dim_{\overline{\mathbb{Q}}_l}(V \otimes_{\tau,F} B_{\mathrm{DR}})^{\mathrm{Gal}(\bar{F}/F)} = \dim_{\overline{\mathbb{Q}}_l} V$ for all continuous embeddings $\tau : F \hookrightarrow \overline{\mathbb{Q}}_l$) then we may associate to r a Weil–Deligne representation $\mathrm{WD}(r)$ of the Weil group W_K of K over $\overline{\mathbb{Q}}_l$. In the case $l \neq p$ the Weil–Deligne representation $\mathrm{WD}(r)$ determines r up to equivalence. (See for instance [Taylor and Yoshida 2007, Section 1] for details.) If (r, N) is a Weil–Deligne representation of W_K then we will let $(r, N)^{\text{F-ss}} = (r^{\mathrm{ss}}, N)$ denote the Frobenius semisimplification of (r, N). We will write rec_F for the local Langlands correspondence — a bijection from irreducible smooth representations of $\mathrm{GL}_n(F)$ over $\mathbb{C}$ to n-dimensional Frobenius semisimple Weil–Deligne representations of the Weil group W_F of F. (See the Introduction or

Section VII.2 of [Harris and Taylor 2001].) Recall that if χ is a character of $F^\times$ then $\mathrm{rec}(\chi) = \chi \circ \mathrm{Art}_F^{-1}$.)

If $F = \mathbb{R}$ or $\mathbb{C}$ we will write $\mathrm{Art}_F : F^\times \twoheadrightarrow \mathrm{Gal}(\bar{F}/F)$. If $F = \mathbb{R}$ then we will denote by c the nontrivial element of $\mathrm{Gal}(\bar{F}/F)$ and denote by sgn the unique surjection $F^\times \twoheadrightarrow \{\pm 1\}$.

If F is a number field then

$$\mathrm{Art}_F = \prod_v \mathrm{Art}_{F_v} : \mathbb{A}_F^\times / \overline{F^\times (F_\infty^\times)^0} \xrightarrow{\sim} \mathrm{Gal}(\bar{F}/F)^{\mathrm{ab}}$$

will denote the Artin map. If v is a real place of F then we will let c_v denote the image of $c \in \mathrm{Gal}(\bar{F}_v/F_v)$ in $\mathrm{Gal}(\bar{F}/F)$. Thus c_v is well defined up to conjugacy. Suppose that

$$\chi : \mathbb{A}_F^\times / F^\times \to \mathbb{C}^\times$$

is a continuous character for which there exists $a \in \mathbb{Z}^{\mathrm{Hom}(F,\mathbb{C})}$ such that

$$\chi|_{(F_\infty^\times)^0} : x \mapsto \prod_{\tau \in \mathrm{Hom}(F,\mathbb{C})} (\tau x)^{a_\tau}$$

(i.e., an algebraic grossencharacter). Suppose also that $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. Then we define

$$r_{l,\iota}(\chi) : \mathrm{Gal}(\bar{F}/F) \to \overline{\mathbb{Q}}_l^\times$$

to be the continuous character such that

$$\iota\left((r_{l,\iota}(\chi) \circ \mathrm{Art}_F)(x) \prod_{\tau \in \mathrm{Hom}(F,\mathbb{C})} (\iota^{-1}\tau)(x_l)^{-a_\tau}\right) = \chi(x) \prod_{\tau \in \mathrm{Hom}(F,\mathbb{C})} (\tau x)^{-a_\tau}.$$

1. Statement of the main result

Now let F^+ be a totally real field. By a *RAESDC* (regular, algebraic, essentially self dual, cuspidal) automorphic representation π of $\mathrm{GL}_n(\mathbb{A}_{F^+})$ we mean a cuspidal automorphic representation such that

- $\pi^\vee \cong \pi \otimes (\chi \circ \det)$ for some continuous character $\chi : \mathbb{A}_{F^+}^\times / (F^+)^\times \to \mathbb{C}^\times$ with $\chi_v(-1)$ independent of $v|\infty$, and
- π_∞ has the same infinitesimal character as some irreducible algebraic representation of the restriction of scalars from F^+ to $\mathbb{Q}$ of GL_n.

Note that χ is necessarily algebraic. Also, if n is odd and $\pi^\vee \cong \pi \otimes (\chi \circ \det)$, then $\chi_v(-1)$ is necessarily independent of $v|\infty$, in fact it is necessarily 1 for all such v.

If F^+ is totally real we will write $(\mathbb{Z}^n)^{\mathrm{Hom}(F^+,\mathbb{C}),+}$ for the set of $a = (a_{\tau,i}) \in (\mathbb{Z}^n)^{\mathrm{Hom}(F^+,\mathbb{C})}$ satisfying

$$a_{\tau,1} \geq \cdots \geq a_{\tau,n}.$$

If $F^{+\prime}/F^+$ is a finite totally real extension we define $a_{F^{+\prime}} \in (\mathbb{Z}^n)^{\mathrm{Hom}(F^{+\prime},\mathbb{C}),+}$ by

$$(a_{F^{+\prime}})_{\tau,i} = a_{\tau|_{F^+},i}.$$

If $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(F^+,\mathbb{C}),+}$, let Ξ_a denote the irreducible algebraic representation of $\mathrm{GL}_n^{\mathrm{Hom}(F^+,\mathbb{C})}$ which is the tensor product over τ of the irreducible representations of GL_n with highest weights a_τ. We will say that a RAESDC automorphic representation π of $\mathrm{GL}_n(\mathbb{A}_{F^+})$ has *weight* a if π_∞ has the same infinitesimal character as $\Xi_a^\vee$.

Fix once and for all an isomorphism $\iota : \overline{\mathbb{Q}}_l \xrightarrow{\sim} \mathbb{C}$. The following theorem is proved in [Shin 2011] (see also [Clozel et al. 2011]). (This is not explicitly stated in [Shin 2011], but see Remark 7.6 of that reference. For the last sentence see [Taylor and Yoshida 2007].)

Theorem 1.1. *Let F_0^+ be a totally real field and let n be an odd positive integer. Let $a \in (\mathbb{Z}^n)^{\mathrm{Hom}(F_0^+,\mathbb{C}),+}$. Suppose further that Π is a RAESDC automorphic representation of $\mathrm{GL}_n(\mathbb{A}_{F_0^+})$ of weight a. Specifically suppose that $\Pi^\vee \cong \Pi\chi$ where $\chi : \mathbb{A}_{F_0^+}^\times/(F_0^+)^\times \to \mathbb{C}^\times$ and $\chi_v(-1)$ is independent of $v|\infty$. Then there is a continuous semisimple representation*

$$r_{l,\iota}(\Pi) : \mathrm{Gal}(\bar{F}_0^+/F_0^+) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$$

with the following properties.

(1) *For every prime $v \nmid l$ of F_0^+ we have*

$$\mathrm{WD}(r_{l,\iota}(\Pi)|_{\mathrm{Gal}(\bar{F}_{0,v}^+/F_{0,v}^+)})^{\text{F-ss}} = r_l(\iota^{-1}\,\mathrm{rec}(\Pi_v \otimes |\det|_v^{(1-n)/2}).$$

(2) $r_{l,\iota}(\Pi)^\vee = r_{l,\iota}(\Pi)\epsilon^{n-1}r_{l,\iota}(\chi)$.

(3) $\det r_{l,\iota}(\Pi) = r_{l,\iota}(\kappa_\Pi)\epsilon_l^{n(1-n)/2}$.

(4) *If $v|l$ is a prime of F_0^+ then the restriction $r_{l,\iota}(\Pi)|_{\mathrm{Gal}(\bar{F}_{0,v}^+/F_{0,v}^+)}$ is de Rham. Moreover, if Π_v is unramified, if $(F_{0,v}^+)^0$ denotes the maximal unramified subextension of $F_{0,v}^+/\mathbb{Q}_l$ and if $\tau : (F_{0,v}^+)^0 \hookrightarrow \overline{\mathbb{Q}}_l$ then $r_{l,\iota}(\Pi)|_{\mathrm{Gal}(\bar{F}_{0,v}^+/F_{0,v}^+)}$ is crystalline and the characteristic polynomial of $\phi^{[(F_{0,v}^+)^0:\mathbb{Q}_l]}$ on*

$$(r_{l,\iota}(\Pi) \otimes_{\tau,(F_{0,v}^+)^0} B_{\mathrm{cris}})^{\mathrm{Gal}(\bar{F}_{0,v}^+/F_{0,v}^+)}$$

equals the characteristic polynomial of

$$\iota^{-1}\,\mathrm{rec}_{F_{0,v}^+}(\Pi_v \otimes |\det|_v^{(1-n)/2})(\mathrm{Frob}_v).$$

(5) *If $v|l$ is a prime of F_0^+ and if $\tau : F_0^+ \hookrightarrow \overline{\mathbb{Q}}_l$ lies above v then*

$$\dim_{\overline{\mathbb{Q}}_l} \mathrm{gr}^i(r_{l,\iota}(\Pi) \otimes_{\tau,F_{0,v}} B_{\mathrm{DR}})^{\mathrm{Gal}(\bar{F}_{0,v}^+/F_{0,v}^+)} = 0$$

unless $i = a_{\iota\tau,j} + n - j$ *for some* $j = 1, \dots, n$ *in which case*

$$\dim_{\overline{\mathbb{Q}}_l} \mathrm{gr}^i (r_{l,\iota}(\Pi) \otimes_{\tau, F_{0,v}^+} B_{\mathrm{DR}})^{\mathrm{Gal}(\bar{F}_{0,v}^+/F_{0,v}^+)} = 1.$$

(6) *If* Π *is discrete series at some finite place then* $r_{l,\iota}(\Pi)$ *is irreducible.*

The purpose of this paper is to calculate $r_{l,\iota}(\Pi)(c_v)$ for any infinite place v of F_0^+.

Proposition 1.2. *Keep the notation and assumptions of the above theorem and suppose that* $r_{l,\iota}(\Pi)$ *is irreducible.* (*In particular we are assuming that n is odd.*) *Let v denote an infinite place of* F_0^+. *Then*

$$r_{l,\iota}(\Pi)(c_v)$$

is semisimple with eigenvalues 1 *of multiplicity* $(n + \kappa_{\Pi,v}(-1))/2$ *and* -1 *with multiplicity* $(n - \kappa_{\Pi,v}(-1))/2$.

2. A geometric realization of complex conjugation

We must recall some of the construction of $r_{l,\iota}(\Pi)$ and explain how the action of complex conjugation can be constructed geometrically.

The basic set-up. There is a constant $\alpha \in \mathbb{Z}$ such that $a_{\tau,j} + a_{\tau,n+1-j} = \alpha$ for all $j = 1, \dots, n$ and all $\tau : F_0^+ \hookrightarrow \mathbb{C}$. Thus

$$\chi|_{((F_{0,\infty}^+)^\times)^0} = \mathbf{N}_{F_0^+/\mathbb{Q}}^{\alpha}.$$

Shin shows that one can choose

- a soluble Galois totally real extension F^+/F_0^+,
- an imaginary quadratic field E in which l splits,
- an embedding $\tau_0 : F = F^+E \hookrightarrow \mathbb{C}$,
- a continuous character

$$\phi : \mathbb{A}_F^\times / F^\times \to \mathbb{C}^\times,$$

- a continuous character

$$\psi : \mathbb{A}_E^\times / E^\times \to \mathbb{C}^\times,$$

with the following properties.

- $[F^+ : \mathbb{Q}]$ is even and > 2.
- If Ram denotes the set of (finite) rational primes above which any of F, Π, ϕ, or ψ ramifies, then every prime of F^+ above a prime of Ram splits in F.
- $r_{l,\iota}(\Pi)|_{\mathrm{Gal}(\bar{F}/F)}$ remains irreducible.

- $\phi\phi^c = \chi_F$ and $\phi|_{F_\infty^\times} = \prod_\tau \tau^{-\beta_\tau}$ where $\beta_\tau + \beta_{\tau c} = -\alpha$.
- $\psi^c/\psi = (\kappa_\Pi|_{\mathbb{A}^\times}^{[F^+:F_0^+]} \circ \mathbf{N}_{E/\mathbb{Q}})\phi|_{\mathbb{A}_E^\times}^n$.
- $\psi_\infty = \tau_0^{-\epsilon}(\tau_0 \circ c)^{-\epsilon'}$ with $\epsilon, \epsilon' \in \mathbb{Z}$.
- ψ is unramified at the prime of E above l corresponding to $\iota^{-1} \circ \tau_0$.

Let $V = F^n$ and let

$$\langle\ ,\ \rangle : V \times V \to \mathbb{Q}$$

be a nondegenerate alternating bilinear form such that

$$\langle xv, w\rangle = \langle v, {}^c x w\rangle$$

for all $x \in F$ and $v, w \in V$. Let G be the reductive subgroup of $\mathrm{GL}(V/F)$ consisting of elements which preserve $\langle\ ,\ \rangle$ up to a $\mathbb{G}_m$-multiple and let $\nu : G \to \mathbb{G}_m$ denote the multiplier character. We may, and do, suppose that V is chosen so that

- G is quasisplit at all finite places;
- if $\tau : F \hookrightarrow \mathbb{C}$ satisfies $\tau|_E = \tau_0|_E$ then the Hermitian form on $V \otimes_{F,\tau} \mathbb{C}$ defined by

$$(v, w) \mapsto \langle v, iw\rangle$$

has a maximal positive definite subspace of dimension 0 if $\tau \neq \tau_0$ and 1 if $\tau = \tau_0$.

(See [Shin 2011, Lemma 5.1].) There is an identification of $G \times_{\mathbb{Q}} E$ with the product of GL_1 and the restriction of scalars from F to E of GL_n. The map sends g to the product of its multiplier and its action on the direct summand $V \otimes_{E,1} E$ of $V \otimes_{\mathbb{Q}} E = V \otimes_{E,1} E \oplus V \otimes_{E,c} E$.

The group G. Letting $\ker^1(\mathbb{Q}, G)$ denote the kernel of

$$H^1(\mathbb{Q}, G) \to \prod_v H^1(\mathbb{Q}_v, G),$$

using the fact that n is odd, we see from [Kottwitz 1992, Section 8] that there is an identification

$$\ker^1(\mathbb{Q}, G) \cong ((F^+)^\times \cap (\mathbb{A}^\times \mathbf{N}_{F/F^+}\mathbb{A}_F^\times))/\mathbb{Q}^\times(\mathbf{N}_{F/F^+}F^\times).$$

As F/F^+ is unramified at all finite primes we see that $\mathbf{N}_{F/F^+}\mathbb{A}_F^\times \supset \widehat{\mathbb{Z}}^\times \mathbb{R}_{>0}^\times$ so that $\mathbb{A}^\times \mathbf{N}_{F/F^+}\mathbb{A}_F^\times = \mathbb{Q}^\times \mathbf{N}_{F/F^+}\mathbb{A}_F^\times$. Because $(F^+)^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times = \mathbf{N}_{F/F^+}F^\times$ we conclude that

$$\ker^1(\mathbb{Q}, G) \cong \mathbb{Q}^\times((F^+)^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times)/\mathbb{Q}^\times(\mathbf{N}_{F/F^+}F^\times) = \{1\}.$$

It follows from the proof of Lemma 3.1 of [Shin 2011] that the Tamagawa number $\tau(G) = 2$.

Let T denote the quotient of G by its derived subgroup. Then we may identify T by

$$T(R) = \{(x, y) \in R^\times \times (R \otimes_{\mathbb{Q}} F)^\times : \ x^n = y^c y\}$$

for any $\mathbb{Q}$-algebra R. The quotient map $d : G \to T$ sends g to $(\nu(g), \det g)$. Also let Z denote the centre of G so that

$$Z(R) = \{(x, y) \in R^\times \times (R \otimes_{\mathbb{Q}} F)^\times : \ x = y^c y\}$$

for any $\mathbb{Q}$-algebra R. The map $d|_Z$ sends (x, y) to (x, y^n) and the map $\nu|_Z$ sends (x, y) to x. Note that $Z \times E$ can be identified with the product of $\mathbb{G}_m$ with the restriction of scalars from F to E of $\mathbb{G}_m$ and the norm map sends (a, b) to $(a^c a, {}^c ab/{}^c b)$. Then

$$\nu : Z(\mathbb{A})/Z(\mathbb{Q})(\mathbf{N}_{E/\mathbb{Q}} Z(\mathbb{A}_E)) \xrightarrow{\sim} \mathbb{A}^\times/\mathbb{Q}^\times(\mathbf{N}_{E/\mathbb{Q}}\mathbb{A}_E^\times) \cong \mathrm{Gal}(E/\mathbb{Q}).$$

[To see this note that the left hand side is

$$\{y \in \mathbb{A}_F^\times : \ y^c y \in \mathbb{A}^\times\}/\mathbb{A}_E^\times\{y \in F^\times : \ y^c y \in \mathbb{Q}^\times\}\{y/{}^c y : \ y \in \mathbb{A}_F^\times\}.$$

As $\{y/{}^c y : \ y \in \mathbb{A}_F^\times\} = \mathbb{A}_F^{\mathbf{N}_{F/F^+}=1}$ we see that the group in the previous displayed equations maps isomorphically under $\nu = \mathbf{N}_{F/F^+}$ to

$$\begin{aligned} &(\mathbb{A}^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times)/(\mathbf{N}_{E/\mathbb{Q}}\mathbb{A}_E^\times)(\mathbb{Q}^\times \cap \mathbf{N}_{F/F^+}F^\times) \\ \cong\ &(\mathbb{A}^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times)/((\mathbf{N}_{E/\mathbb{Q}}\mathbb{A}_E^\times)\mathbb{Q}^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times). \end{aligned}$$

There is a natural injection from here to $\mathbb{A}^\times/(\mathbf{N}_{E/\mathbb{Q}}\mathbb{A}_E^\times)\mathbb{Q}^\times$. It only remains to see that this map is surjective, i.e., that

$$\mathbb{A}^\times/\mathbb{Q}^\times(\mathbf{N}_{E/\mathbb{Q}}\mathbb{A}_E^\times)(\mathbb{A}^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times) = \{1\}.$$

However as F/F^+ is everywhere unramified we have that

$$(\mathbb{A}^\times \cap \mathbf{N}_{F/F^+}\mathbb{A}_F^\times) \supset \widehat{\mathbb{Z}}^\times \times \mathbb{R}_{>0}^\times,$$

while $\mathbb{A}^\times = \mathbb{Q}^\times\widehat{\mathbb{Z}}^\times\mathbb{R}_{>0}^\times$.]

The involution I. We can choose a $\mathbb{Q}$-linear map $I : V \to V$ such that

- $I(xv) = {}^c x I(v)$ for all $x \in F$ and $v \in V$;
- $\langle Iv, Iw\rangle = -\langle v, w\rangle$ for all $v, w \in V$;
- $I^2 = 1$.

[To see this note that with respect to a suitable basis we have

$$\langle v, w\rangle = \mathrm{tr}_{F/\mathbb{Q}}({}^t v D^c w)$$

for some diagonal matrix D with ${}^c D = -D$. With respect to such a basis we can take I to simply be complex conjugation on coordinates.] The choice of I gives rise to an automorphism # of G of order two:

$$g^{\#} = I g I.$$

Note that

$$\nu \circ \# = \nu$$

and that

$$\det g^{\#} = {}^c \det g.$$

If we identify $G \times E$ with the product of $\mathbb{G}_m$ and the restriction of scalars from F to E of GL_n then # differs by composition with an inner automorphism from the automorphism:

$$(x, g) \mapsto (x, x^t g^{-1}).$$

Base change from $G(\mathbb{A}^\infty)$ to $(\mathbb{A}_E^\infty)^\times \times \mathbf{GL}_n(\mathbb{A}_F^\infty)$. As in [Harris and Taylor 2001, Section VI.2] we can define the base change $\mathrm{BC}(\tilde{\pi})$ of an irreducible admissible representation $\tilde{\pi}$ of $G(\mathbb{A}^\infty)$ which is unramified at a place v of $\mathbb{Q}$, unless all primes of F^+ above v split in F. The base change lift, $\mathrm{BC}(\tilde{\pi})$, is an irreducible admissible representation of $(\mathbb{A}_E^\infty)^\times \times \mathrm{GL}_n(\mathbb{A}_F^\infty)$. Note that if $\delta_{E/\mathbb{Q}}$ denotes the nontrivial character of $\mathbb{A}^\times/\mathbb{Q}^\times \mathbf{N}_{E/\mathbb{Q}}\mathbb{A}_E^\times$ then

$$\mathrm{BC}(\tilde{\pi}) = \mathrm{BC}(\tilde{\pi} \otimes (\delta_{E/\mathbb{Q}} \circ \nu)).$$

Also note that $\tilde{\pi}$ and $\tilde{\pi} \otimes (\delta_{E/\mathbb{Q}} \circ \nu)$ have different central characters and so can not be isomorphic. (Recall that

$$\nu : Z(\mathbb{A}^\infty) \twoheadrightarrow (\mathbb{A}^\infty)^\times \cap \mathbf{N}_{F/F^+}(\mathbb{A}_F^\infty)^\times \supset \widehat{\mathbb{Z}}^\times,$$

and that $\delta_{E/\mathbb{Q}}$ is ramified at some finite prime.) We have that

$$\kappa_{\mathrm{BC}(\tilde{\pi})} = \kappa_{\tilde{\pi}} \circ \mathbf{N},$$

where $\mathbf{N}$ denotes the norm map $Z(\mathbb{A}_E^\infty) \to Z(\mathbb{A}^\infty)$. If

$$\mathrm{BC}(\tilde{\pi}) = (\tilde{\phi}, \widetilde{\Pi})$$

then

$$\mathrm{BC}(\tilde{\pi}^{\#}) = (\tilde{\phi}\kappa_{\widetilde{\Pi}}|_{(\mathbb{A}_E^\infty)^\times}, \widetilde{\Pi}^\vee)$$

and

$$\kappa_{\tilde{\pi}^{\#}} = \kappa_{\tilde{\pi}} \kappa_{\widetilde{\Pi}}^c|_{Z(\mathbb{A}^\infty)},$$

where we think of $Z(\mathbb{A}^\infty) \subset (\mathbb{A}_F^\infty)^\times$.

Define

$$\begin{aligned}\omega : T(\mathbb{A})/T(\mathbb{Q}) &\to \mathbb{C}^\times \\ (x, y) &\mapsto \phi^c(y)^{-1}\kappa_{\Pi, F^+}(x)^{-1}.\end{aligned}$$

Note that

$$\omega^\# \omega = 1.$$

With the functorialities of the previous paragraph the next lemma is easy to verify.

Lemma 2.1. *Suppose that $\tilde{\pi}$ is as in the previous paragraph and that*

$$\mathrm{BC}(\tilde{\pi}) = (\psi^\infty, \Pi_F \phi).$$

Then

(1) $\kappa_{\tilde{\pi}^\# \otimes (\omega^\infty \circ d)} = \kappa_{\tilde{\pi}}$;

(2) $\mathrm{BC}(\tilde{\pi}^\# \otimes (\omega^\infty \circ d)) = \mathrm{BC}(\tilde{\pi})$;

(3) *and there exists an automorphism $A_{\tilde{\pi}}$ of the underlying space of $\tilde{\pi}$ such that*

$$A_{\tilde{\pi}} \tilde{\pi}(g) = \tilde{\pi}(g^\#)\omega(d(g))A_{\tilde{\pi}}$$

for all $g \in G(\mathbb{A}^\infty)$ and $A_{\tilde{\pi}}^2 = 1$. Moreover $A_{\tilde{\pi}}$ is unique up to sign.

Weights. We identify $G \times_{\mathbb{Q}} \mathbb{C}$ with

$$\mathbb{G}_m \times \prod_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C})} \mathrm{GL}(V \otimes_{F,\tau} \mathbb{C}),$$

where $\mathrm{Hom}_{E,\tau_0}(F, \mathbb{C})$ denotes the set of embeddings $\tau : F \hookrightarrow \mathbb{C}$ with $\tau|_E = \tau_0|_E$. The identification sends g to its multiplier and its push forward to each $\mathrm{GL}(V \otimes_{F,\tau} \mathbb{C})$. Let ξ denote the irreducible representations of $G \times_{\mathbb{Q}} \mathbb{C}$ with highest weights $(b_0; b_{\tau,i})_{\tau|_E = \tau_0|_E}$, where

- $b_0 = \epsilon$;
- $b_{\tau,i} = a_{\tau|_{F_0^+},i} + \beta_\tau$.

Then $\xi^\#$ has highest weights

$$\Big(b_0 + \sum_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C}),i} b_{\tau,i}; -b_{\tau,n+1-i}\Big)_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C});\ i=1,\ldots,n}.$$

Also let ζ be the irreducible representation with highest weights

$$\Big(-n\big([F^+ : \mathbb{Q}]\alpha/2 + \sum_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C})} \beta_\tau\big); \alpha + 2\beta_\tau\Big)_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C});\ i=1,\ldots,n}.$$

Then

- ζ is one-dimensional;
- $\xi^\# \otimes \zeta \cong \xi$;
- $\zeta^\# \cong \zeta^\vee$;
- and $\omega|_{T(\mathbb{R})} = \zeta^{-1}$.

Shimura varieties. Let U denote an open compact subgroup of $G(\mathbb{A}^\infty)$. Consider the functor $\mathfrak{X}_U$ from connected, locally noetherian F-schemes with a specified geometric point to sets, which sends a pair $(S, \bar{s})$ to the set of equivalence classes of 4-tuples

$$(A, i, \lambda, \bar{\eta})$$

where

(1) A/S is an abelian scheme of relative dimension n;

(2) $i : F \hookrightarrow \mathrm{End}^0(A/S)$ is such that for all $x \in F$ we have

$$\mathrm{tr}(x|_{\mathrm{Lie}\, A}) = x - {}^c x + n\, \mathrm{tr}_{F/E}\, {}^c x;$$

(3) $\lambda : A \to A^\vee$ is a polarization such that $i(x)^\vee \circ \lambda = \lambda \circ i({}^c x)$ for all $x \in F$;

(4) $\bar{\eta}$ is a $\pi_1(S, \bar{s})$-invariant U-orbit of $\mathbb{A}_F^\infty$-isomorphisms $\eta : V \otimes \mathbb{A}^\infty \xrightarrow{\sim} VA_{\bar{s}}$ such that for some isomorphism $\eta_0 : \mathbb{A}^\infty \xrightarrow{\sim} \mathbb{A}^\infty(1)$ and for all $v, w \in V \otimes \mathbb{A}^\infty$ we have

$$\langle \eta v, \eta w \rangle_\lambda = \eta_0 \langle v, w \rangle,$$

where $\langle\ ,\ \rangle_\lambda$ denotes the λ-Weil pairing.

Two 4-tuples $(A, i, \lambda, \bar{\eta})$ and $(A', i', \lambda', \bar{\eta}')$ are considered equivalent if there is an isogeny

$$\gamma : A \to A'$$

such that

(1) $\gamma i(x) = i'(x)\gamma$ for all $x \in F$,

(2) $\gamma^\vee \lambda' \gamma \in \mathbb{Q}^\times \lambda$,

(3) and $(V\gamma_{\bar{s}}) \circ \bar{\eta} = \bar{\eta}'$.

This functor is canonically independent of the choice of base point $\bar{s}$ and so can be considered as a functor from connected, locally noetherian F-schemes to sets. It can be extended to all locally noetherian F-schemes by setting

$$\mathfrak{X}_U(S_1 \amalg S_2) = \mathfrak{X}_U(S_1) \times \mathfrak{X}_U(S_2).$$

(See for instance [Harris and Taylor 2001, Section III.1] for more details. We are using $\mathrm{End}^0(A/S)$ to denote $\mathrm{End}(A/S) \otimes_{\mathbb{Z}} \mathbb{Q}$ and $VA_{\bar{s}}$ for $(\lim_{\leftarrow N} A[N](k(\bar{s}))) \otimes_{\mathbb{Z}} \mathbb{Q}$, where $k(\bar{s})$ denotes the residue field of $\bar{s}$.)

If U is sufficiently small then $\mathfrak{X}_U$ is represented by an abelian scheme

$$\mathscr{A}_U / X_U / \mathrm{Spec}\, F.$$

If $V \subset U$ is an open subgroup there is a natural map $X_V \to X_U$ such that $\mathscr{A}_U$ pulls back to $\mathscr{A}_V$. The inverse system of the X_U's carries a natural action of $G(\mathbb{A}^\infty)$, as does the inverse system of the $\mathscr{A}_U$'s. If V is a normal open subgroup of U

then U acts on X_V and induces an isomorphism between U/V and $\mathrm{Gal}(X_V/X_U)$. Thus $\iota^{-1}\xi$ gives a representation of U and hence a lisse $\overline{\mathbb{Q}}_l$-sheaf $\mathscr{L}_\xi$ on X_U. The $\overline{\mathbb{Q}}_l$-vector space

$$H^i(X, \mathscr{L}_\xi) = \lim_{\to U} H^i(X_U \times \bar{F}, \mathscr{L}_\xi)$$

has an action of $G(\mathbb{A}^\infty) \times \mathrm{Gal}(\bar{F}/F)$. It is admissible and semisimple as a $G(\mathbb{A}^\infty)$-module. If U is an open, compact subgroup of $G(\mathbb{A}^\infty)$ then

$$H^i(X, \mathscr{L}_\xi)^U = H^i(X_U \times \bar{F}, \mathscr{L}_\xi)$$

is a continuous representation of $\mathrm{Gal}(\bar{F}/F)$ on a finite-dimensional $\overline{\mathbb{Q}}_l$-vector space.

The pull back $X_U \times_{F,c} F$ represents the functor $\mathfrak{X}'_U$ defined exactly as $\mathfrak{X}_U$ except that the condition

$$\mathrm{tr}(x|_{\mathrm{Lie}\, A}) = x - {}^c x + n\, \mathrm{tr}_{F/E}\, {}^c x$$

is replaced by the condition

$$\mathrm{tr}(x|_{\mathrm{Lie}\, A}) = {}^c x - x + n\, \mathrm{tr}_{F/E}\, x.$$

There is a map of functors $\mathfrak{X}_U \to \mathfrak{X}'_U$ which sends $(A, i, \lambda, \bar{\eta})$ to $(A, i \circ c, \lambda, \overline{\eta \circ I})$. This induces an F-linear map $X_U \to X_U \times_{F,c} F$ and hence a c-linear map, which we will also denote I,

$$\begin{array}{ccc} X_U & \xrightarrow{I} & X_U \\ \downarrow & & \downarrow \\ \mathrm{Spec}\, F & \xrightarrow{c} & \mathrm{Spec}\, F. \end{array}$$

We have

- $I^2 = 1$;
- $IgI = g^\#$ for $g \in G(\mathbb{A}^\infty)$;
- and a natural isomorphism $I^*\mathscr{L}_\xi \otimes \mathscr{L}_\zeta \cong \mathscr{L}_\xi$, i.e.,

$$I^*\mathscr{L}_\xi \cong \mathscr{L}_{\xi^\#}. \tag{2-1}$$

Thus I provides a way to descend the system of the X_U to F^+; however this descended system of varieties no longer has an action of $G(\mathbb{A}^\infty)$ defined over F^+.

Complex points and connected components. We will need to consider the complex uniformization of $X_U \times_{F,\tau} \mathbb{C}$ for every homomorphism $\tau : F \hookrightarrow \mathbb{C}$. So suppose $\tau : F \hookrightarrow \mathbb{C}$. There is a nondegenerate alternating form

$$\langle\ ,\ \rangle_\tau : V \times V \to \mathbb{Q}$$

such that

$$\langle xv, w\rangle_\tau = \langle v, {}^c x w\rangle_\tau$$

for all $x \in F$ and $v, w \in V$ and such that

- there is an isomorphism $j_\tau : (V \otimes_{\mathbb{Q}} \mathbb{A}^\infty, \langle\ ,\ \rangle) \xrightarrow{\sim} (V \otimes_{\mathbb{Q}} \mathbb{A}^\infty, \langle\ ,\ \rangle_\tau)$ as $\mathbb{A}_F^\infty$-modules with alternating $\mathbb{A}^\infty$-bilinear pairing;
- if $\tau' : F \hookrightarrow \mathbb{C}$ satisfies $\tau'|_E = \tau|_E$ then the Hermitian form on $V \otimes_{F,\tau'} \mathbb{C}$ defined by
$$(v, w) \mapsto \langle v, iw\rangle_\tau$$
has a maximal positive definite subspace of dimension 0 if $\tau' \neq \tau$ and 1 if $\tau' = \tau$.

Let G_τ denote the group of symplectic F-linear similitudes for $(V, \langle\ ,\ \rangle_\tau)$ and $G_{\tau,1}$ the kernel of the multiplier character $G_\tau \to \mathbb{G}_m$. Note that $G_\tau \times_{\mathbb{Q}} \mathbb{A}^\infty \cong G \times_{\mathbb{Q}} \mathbb{A}^\infty$ and that $G_\tau / G_{\tau,1} \xrightarrow{\sim} T$. Choose a $\mathbb{Q}$-linear map $I_\tau : V \to V$ such that

- $I_\tau(xv) = {}^c x I_\tau(v)$ for all $x \in F$ and $v \in V$;
- $\langle I_\tau v, I_\tau w\rangle = -\langle v, w\rangle$ for all $v, w \in V$;
- $I_\tau^2 = 1$.

We may, and shall, take $\langle\ ,\ \rangle_{\tau_0} = \langle\ ,\ \rangle$ and $I_{\tau_0} = I$.

Let Ω_τ denote the set of homomorphisms
$$h : \mathbb{C} \to \mathrm{End}_{F \otimes_{\mathbb{Q}} \mathbb{R}}(V \otimes_{\mathbb{Q}} \mathbb{R})$$
such that

- $\langle h(z)v, w\rangle_\tau = \langle v, h({}^c z)w\rangle_\tau$ for all $z \in \mathbb{C}$ and $v, w \in V \otimes \mathbb{R}$,
- $\langle v, h(i)v\rangle_\tau \geq 0$ for all $v \in V$.

Then Ω_τ forms a single conjugacy class for $G_{\tau,1}(\mathbb{R})$ [Kottwitz 1992, Lemma 4.3]. This gives Ω_τ a topology (the quotient topology) and, as the group $G_{\tau,1}(\mathbb{R})$ is connected, we see that Ω_τ is connected. There are $G(\mathbb{A}^\infty)$-equivariant homeomorphisms (see [Kottwitz 1992, Section 8], for example)
$$G_\tau(\mathbb{Q}) \backslash (G(\mathbb{A}^\infty)/U \times \Omega_\tau) \xrightarrow{\sim} (X_U \times_{F,\tau} \mathbb{C})(\mathbb{C}).$$

Let Λ be a $\mathbb{Z}$-lattice in V. The map sends (g, h) to a the equivalence class of a four-tuple $(A, i, \lambda, \bar{\eta})$, which is determined as follows. The abelian variety A is characterized by the complex uniformization $A(\mathbb{C}) = (V \otimes_{\mathbb{Q}} \mathbb{R})/\Lambda$ with the complex structure coming from h. The map i arises from the natural action of F on $V \otimes_{\mathbb{Q}} \mathbb{R}$ and the (quasi)polarization λ corresponds to the Riemann form $\langle\ ,\ \rangle_\tau$. Note that VA is naturally identified with $V \otimes_{\mathbb{Q}} \mathbb{A}^\infty$. The level structure $\bar{\eta}$ is the class of $j_\tau \circ g$. Under $I \times c_\tau$ this is taken to $({}^c A, i \circ c, \lambda, \bar{\eta} \circ I)$, which has analytic uniformization as $(V \otimes_{\mathbb{Q}} \mathbb{R})/\Lambda$ but with the complex structure coming from $h \circ c$. The F action is the complex conjugate of the usual one. The Riemann form is sent

to its negative and the level structure is $j_\tau \circ g \circ I$. The map $I \otimes 1_{\mathbb{R}}$ shows that this is isomorphic to the abelian variety with additional structure corresponding to $((j_\tau^{-1} I_\tau j_\tau I) g^\#, I_\tau h I_\tau) \in G(\mathbb{A}^\infty) \times \Omega_\tau$. Set $s_\tau = j_\tau^{-1} I_\tau j_\tau I \in G(\mathbb{Q})$ and note that $s_\tau^\# s_\tau = 1$.

We conclude that there is a bijection ς_τ :

$$\pi_0(X_U \times_F \bar{F}) \cong \pi_0(X_U \times_{F,\tau} \mathbb{C})(\mathbb{C}) \cong G_\tau(\mathbb{Q}) \backslash G_\tau(\mathbb{A}^\infty)/U \xrightarrow{\sim} T(\mathbb{Q}) \backslash T(\mathbb{A}^\infty)/d(U).$$

(For the bijectivity of the third map, which is given by d, see [Milne 2005, Theorem 5.17] and the discussion following it.) Write ς for ς_{τ_0}. The map ς_τ is $G(\mathbb{A}^\infty)$-equivariant. It is also $I \times c_\tau$ equivariant if we let $I \times c_\tau$ act on $T(\mathbb{Q}) \backslash T(\mathbb{A}^\infty)/d(U)$ via $t \mapsto d(s_\tau) t^\#$. Note that because of the $G(\mathbb{A}^\infty)$ equivariance we must have $\varsigma_\tau = u_\tau \varsigma$ for some $u_\tau \in T(\mathbb{A})$. Thus we see that

- $\varsigma(Cg) = d(g)\varsigma(C)$ for all $C \in \pi_0(X_U \times_F \bar{F})$ and all $g \in G(\mathbb{A}^\infty)$,
- and for any infinite place v of $\bar{F}$ there is an $s_v \in T(\mathbb{A})$ such that $\varsigma((I \times c_v)x) = s_v \varsigma(x)^\#$ and $s_v s_v^\# = 1$.

(If $v|_F$ arises from $\tau : F \hookrightarrow \mathbb{C}$ then $s_v = d(s_\tau) u_\tau^\# u_\tau^{-1}$.)

We wish to also know the $\mathrm{Gal}(\bar{F}/F)$-equivariance of ς. Note that the X_U are the canonical models for the Shimura varieties $\mathrm{Sh}_U(G, [h^{-1}])$. (See [Kottwitz 1992, Section 8] and note that $\ker^1(\mathbb{Q}, G) = (0)$.) Define a map

$$r : \mathbb{A}_F^\times \to T(\mathbb{A}_E) \xrightarrow{\mathbf{N}_{E/\mathbb{Q}}} T(\mathbb{A})$$

where the first map sends

$$x \mapsto (\mathbf{N}_{F/E} x, x)^{-1}.$$

Note that $r \circ \mathrm{Art}_F^{-1}$ is a well defined map

$$(r \circ \mathrm{Art}_F^{-1}) : \mathrm{Gal}(\bar{F}/F) \to T(\mathbb{A})/T(\mathbb{Q})T(\mathbb{R}).$$

Then according to [Milne 2005, Section 13] we have

$$\varsigma(\sigma x) = (r \circ \mathrm{Art}_F^{-1})(\sigma)\varsigma(x)$$

for all $x \in \pi_0(X_U \times_F \bar{F})$ and all $\sigma \in \mathrm{Gal}(\bar{F}/F)$.

H^0 of sheaves on our Shimura varieties. Let $\tilde{\xi}$ be the irreducible representation of $G \times \mathbb{C}$ which has highest weight $(\tilde{b}_0, \tilde{b}_{\tau,i})_{\tau|_E = \tau_0|_E}$. The description of the previous section allows us to calculate $H^0(X_U \times \bar{F}, \mathscr{L}_{\tilde{\xi}})$. It will be (0) unless $\tilde{b}_{\tau,i} = \tilde{b}_\tau$ is independent of i. In this case $\tilde{\xi}$ factors through a map $T \times \mathbb{C} \to \mathbb{G}_m$ which we will also denote $\tilde{\xi}$. We can then identify $H^0(X_U \times \bar{F}, \mathscr{L}_{\tilde{\xi}})$ with the space of functions

$$f : T(\mathbb{A})/T(\mathbb{R})T(\mathbb{Q}) \to \overline{\mathbb{Q}}_l$$

such that

$$f(tu) = (\iota^{-1}\tilde{\xi})(u_l)^{-1} f(t)$$

for all $t \in T(\mathbb{A})$ and all $u \in d(U)$. The action of $G(\mathbb{A}^\infty)$ is via

$$(gf)(t) = (\iota^{-1}\tilde{\xi})(g_l) f(td(g))$$

and the action of $\mathrm{Gal}(\overline{F}/F)$ is via

$$(\sigma f)(t) = f((r \circ \mathrm{Art}_F^{-1})(\sigma)t).$$

The map that sends f to $\tilde{f}$ defined by

$$\tilde{f}(t) = (\iota^{-1} \circ \tilde{\xi})(t_\infty)^{-1}(\iota^{-1}\tilde{\xi})(t_l) f(t),$$

establishes an isomorphism between $H^0(X_U \times \overline{F}, \mathscr{L}_{\tilde{\xi}})$ and the space of functions $\tilde{f} : T(\mathbb{A})/T(\mathbb{Q})d(U) \to \overline{\mathbb{Q}}_l$ such that

$$\tilde{f}(tu_\infty) = (\iota^{-1} \circ \tilde{\xi})(u_\infty)^{-1} \tilde{f}(t)$$

for all $t \in T(\mathbb{A})$ and $u_\infty \in T(\mathbb{R})$. Now the action of $G(\mathbb{A}^\infty)$ is via right translation $((g\tilde{f})(t) = \tilde{f}(td(g)))$ and the action of $\mathrm{Gal}(\overline{F}/F)$ is via

$$(\sigma \tilde{f})(t) = (\iota^{-1} \circ \tilde{\xi})(s_\infty)(\iota^{-1}\tilde{\xi})(s_l)^{-1} \tilde{f}(st)$$

where s is a lift of $(r \circ \mathrm{Art}_F^{-1})(\sigma)$ to $T(\mathbb{A})$. From this it follows that we can write

$$H^0(X, \mathscr{L}_{\tilde{\xi}}) = \bigoplus_{\widetilde{\omega}} \overline{\mathbb{Q}}_l \upsilon_{\widetilde{\omega}}$$

where $\widetilde{\omega}$ runs over continuous characters

$$T(\mathbb{A})/T(\mathbb{Q}) \to \mathbb{C}^\times$$

such that $\widetilde{\omega}|_{T(\mathbb{R})} = \tilde{\xi}^{-1}$, and where:

- the action of $G(\mathbb{A}^\infty)$ on $\upsilon_{\widetilde{\omega}}$ is via $\iota^{-1} \circ \widetilde{\omega} \circ d$;
- the action of $\mathrm{Gal}(\overline{F}/F)$ on $\upsilon_{\widetilde{\omega}}$ is via $r_{l,\iota}(\widetilde{\omega} \circ r)$;
- and, if v is an infinite place of $\overline{F}$, then $(I \times c_v)\upsilon_{\widetilde{\omega}} \in \overline{\mathbb{Q}}_l \upsilon_{\widetilde{\omega}^\#}$.

In particular cupping with $\upsilon_{\delta_{E/\mathbb{Q}} \circ \nu} \in H^0(X, \overline{\mathbb{Q}}_l)$ we see that

$$\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\pi, H^i(X, \mathscr{L}_\xi)) \cong \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}(\pi \otimes (\delta_{E/\mathbb{Q}} \circ \nu)), H^i(X, \mathscr{L}_\xi)).$$

If v is a place of $\overline{F}$ above infinity then $I \times c_v$ defines a map $X_U \times_F \overline{F} \to X_U \times_F \overline{F}$, which in turn induces a map

$$H^i(X, \mathscr{L}_\xi) \to H^i(X, \mathscr{L}_{\xi^\#}).$$

Composing this with the cup product with $\omega(s_v)^{-1/2}\upsilon_\omega \in H^0(X, \mathscr{L}_\zeta)$, we get a map

$$I_v : H^i(X, \mathscr{L}_\xi) \to H^i(X, \mathscr{L}_\xi),$$

such that

- $I_v g I_v = g^{\#}(\iota^{-1} \circ \omega \circ d)(g)$ for $g \in G(\mathbb{A}^\infty)$;
- and $I_v \sigma I_v = (c_v \sigma c_v) r_{l,\iota}((\psi_F \phi)^c/(\psi_F \phi))(\sigma)$ for $\sigma \in \mathrm{Gal}(\bar{F}/F)$.

Galois representations. Shin shows that

- $\bigoplus_{\mathrm{BC}(\tilde{\pi})=(\psi^\infty, \Pi_F^\infty \otimes \phi^\infty)} \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\tilde{\pi}, H^i(X, \mathscr{L}_\xi)) \neq (0)$ if and only if $i = n-1$;
- $$\bigoplus_{\mathrm{BC}(\tilde{\pi})=(\psi^\infty, \Pi_F^\infty \otimes \phi^\infty)} \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\tilde{\pi}, H^{n-1}(X, \mathscr{L}_\xi))^{\mathrm{ss}} \cong r_{l,\iota}(\Pi)|^\vee_{\mathrm{Gal}(\bar{F}/F)} \otimes r_{l,\iota}((\psi_F^{-1}\phi^{-1})^2).$$

(See in particular Theorem 6.4, Corollary 6.5 and the proof of Lemma 3.1 of [Shin 2011]. The sums run over $\tilde{\pi}$ which only ramify above rational primes v, such that all places of F^+ above v split in F.) From the irreducibility of $r_{l,\iota}(\Pi)|_{\mathrm{Gal}(\bar{F}/F)}$ we see that at most two $\tilde{\pi}$'s can contribute to the latter sum. On the other hand if $\tilde{\pi}$ contributes so does $\tilde{\pi} \otimes (\delta_{E/\mathbb{Q}} \circ \nu)$, because one can cup with $\upsilon_{\delta_{E/\mathbb{Q}} \circ \nu}$. Thus exactly two $\tilde{\pi}$'s contribute. Choose one of them and from now on reserve the notation π for this one. Thus we have the following.

- Suppose that $\tilde{\pi}$ is an irreducible representation of $G(\mathbb{A}^\infty)$ and $j \in \mathbb{Z}_{\geq 0}$ such that
  - if $\tilde{\pi}$ is ramified above a rational prime v, then all places of F^+ above v split in F;
  - $\mathrm{BC}(\tilde{\pi}) = (\psi^\infty, \Pi_F^\infty \otimes \phi^\infty)$;
  - and $\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\tilde{\pi}, H^j(X, \mathscr{L}_\xi)) \neq (0)$.

  Then $j = n-1$ and $\tilde{\pi} \cong \pi$ or $\pi \otimes (\delta_{E/\mathbb{Q}} \circ \nu)$.
- $\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\pi, H^{n-1}(X, \mathscr{L}_\xi)) \otimes r_{l,\iota}(\psi_F \phi) \cong r_{l,\iota}(\Pi)|^\vee_{\mathrm{Gal}(\bar{F}/F)}$.
- $\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}(\pi \otimes (\delta_{E/\mathbb{Q}} \circ \nu)), H^{n-1}(X, \mathscr{L}_\xi)) \otimes r_{l,\iota}(\psi_F \phi) \cong r_{l,\iota}(\Pi)|^\vee_{\mathrm{Gal}(\bar{F}/F)}$.

If v is an infinite place of $\bar{F}$ then the map

$$f \mapsto I_v \circ f \circ A_\pi$$

induces a map $\tilde{c}_v$ on

$$\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\pi, H^{n-1}(X, \mathscr{L}_\xi)) \otimes r_{l,\iota}(\psi_F \phi)$$

such that

$$\tilde{c}_v \circ \sigma \circ \tilde{c}_v = (c_v \sigma c_v)$$

for all $\sigma \in \mathrm{Gal}(\bar{F}/F)$. Because $r_{l,\iota}(\Pi)|^{\vee}_{\mathrm{Gal}(\bar{F}/F)}$ is irreducible, we conclude that $\tilde{c}_v$ corresponds to a scalar multiple of $r_{l,\iota}(\Pi)^{\vee}(c_v)$. We can, and shall, replace $\tilde{c}_v$ by a scalar multiple so that $\tilde{c}_v^2 = 1$, so that $\tilde{c}_v = \pm r_{l,\iota}(\Pi)^{\vee}(c_v)$. We finally have our geometric realization of $r_{l,\iota}(\Pi)(c_v)$. To prove our proposition it suffices to check that the trace of $\tilde{c}_v$ on

$$\mathrm{Hom}_{G(\mathbb{A}^{\infty})}(\iota^{-1}\pi, H^{n-1}(X, \mathscr{L}_{\xi}))$$

is ± 1. This we will do in the next section by working with the variations of Hodge structure analogue of our l-adic sheaves.

3. Calculation of the trace of $\tilde{c}_v$

We must recall an alternative construction of the sheaves $\mathscr{L}_{\xi}$, $\mathscr{L}_{\xi^{\#}}$ and $\mathscr{L}_{\zeta}$, which will make sense also for variations of Hodge structures. First we recall the theory of Young symmetrizers.

Young symmetrizers. Let k denote a field of characteristic 0 and let $\mathscr{C}$ denote a Tannakian category over k in the terminology of [Deligne 1990]. Suppose that $e = (e_1, \ldots, e_n) \in \mathbb{Z}^n$ satisfies $e_1 \geq e_2 \geq \cdots \geq e_n \geq 0$. Let S_e denote the symmetric group on the set $\mathscr{T}_e$ of pairs of integers (i, j) with $1 \leq i \leq n$ and $1 \leq j \leq e_i$. Let S_e^+ denote the subgroup of S_e consisting of elements σ with $\sigma(i, j) = (i, j')$ some j' and let S_e^- denote the subgroup of S_e consisting of elements σ with $\sigma(i, j) = (i', j)$ for some i'. Further we set

$$A_e^{\pm} = \sum_{\sigma \in S_e^{\pm}} (\pm)^{\sigma} \sigma \in \mathbb{Q}[S_e],$$

where $(+)^{\sigma} = 1$ and $(-)^{\sigma}$ denotes the sign of σ. Note that $(A_e^{\pm})^2 = (\#S_e^{\pm})A_e^{\pm}$ and $(A_e^+ A_e^-)^2 = m(e)(A_e^+ A_e^-)$ and $(A_e^- A_e^+)^2 = m(e)(A_e^- A_e^+)$ for some nonzero integer $m(e)$ [Fulton and Harris 1991, Theorem 4.3]. If W is an object of $\mathscr{C}$ we define

$$\mathscr{S}_e(W) = W^{\otimes \mathscr{T}_e} A_e^+ A_e^-,$$

where S_e acts on $W^{\otimes \mathscr{T}_e}$ from the right by

$$(\otimes_{t \in \mathscr{T}_e} w_t)h = \otimes_{t \in \mathscr{T}_e} w_{ht}.$$

Then $\mathscr{S}_e$ is a functor from $\mathscr{C}$ to itself. Note that $\mathscr{S}_{(1,\ldots,1)}(W) = \bigwedge^n W$. Right multiplication by A_e^+ defines an isomorphism

$$\mathscr{S}_e(W) \xrightarrow{\sim} W^{\otimes \mathscr{T}_e} A_e^- A_e^+,$$

with inverse given by right multiplication by $m(e)^{-1}A_e^-$. Thus we get natural isomorphisms

$$\mathcal{S}_e(W)^\vee = (W^{\otimes\mathcal{T}_e}A_e^+A_e^-)^\vee \xrightarrow{\sim} (W^\vee)^{\otimes\mathcal{T}_e}A_e^-A_e^+ \xrightarrow{\sim} \mathcal{S}_e(W^\vee).$$

Let $e' = (e_1+1, \dots, e_n+1)$. Let

$$\iota : \mathcal{T}_{e'} \xrightarrow{\sim} \mathcal{T}_{(1,\dots,1)} \amalg \mathcal{T}_e$$

be the bijection which sends $(i,1)$ to $(i,1)$ in the first part and, if $j>1$, sends (i,j) to $(i,j-1)$ in the second part. Then ι induces an isomorphism

$$\iota^* : W^{\otimes n} \otimes W^{\otimes\mathcal{T}_e} \to W^{\otimes\mathcal{T}_{e'}}.$$

Note that

$$A_{e'}^+ \circ \iota^* \circ (A_{(1,\dots,1)}^- \otimes A_e^-A_e^+) = (\#S_e^+)(A_{e'}^-A_{e'}^+) \circ \iota^*$$

so that we get a natural surjection

$$\left(\textstyle\bigwedge^n W\right) \otimes \mathcal{S}_e(W) \xrightarrow{\sim} W^{\otimes n}A_{(1,\dots,1)}^- \otimes W^{\otimes\mathcal{T}_e}A_e^-A_e^+ \twoheadrightarrow W^{\otimes\mathcal{T}_{e'}}A_{e'}^-A_{e'}^+ \xrightarrow{\sim} \mathcal{S}'_{e'}(W),$$

where the middle map is $A_{e'}^+ \circ \iota^*$. If W has rank n then this map is an isomorphism. (This can be checked after applying a fibre functor where one can either count dimension, or use the fact that the map is $\mathrm{GL}(W)$ equivariant and $\left(\bigwedge^n W\right) \otimes \mathcal{S}_e(W)$ is an irreducible $\mathrm{GL}(W)$-module.) Thus for any $e = (e_1, \dots, e_n) \in (\mathbb{Z}^n)^+$ and any W of rank n we can define

$$\mathcal{S}_e(W) = \mathcal{S}_{e'}(W) \otimes \left(\textstyle\bigwedge^n W\right)^{\otimes -f}$$

where $f \in \mathbb{Z}$ satisfies $f \geq -e_n$ and where $e' = (e_1+f, \dots, e_n+f)$. We see that up to natural isomorphism this does not depend on the choice of f.

Lemma 3.1. *If $e \in (\mathbb{Z}^n)^+$ equals $(e_1, \dots, e_n)$ set $e^* = (-e_n, \dots, -e_1) \in (\mathbb{Z}^n)^+$. If W has rank n then there are natural isomorphisms*

$$\mathcal{S}_{e+(f,f,\dots,f)}(W) \cong \mathcal{S}_e(W) \otimes \mathcal{S}_{(f,f,\dots,f)}(W)$$

and

$$\mathcal{S}_e(W) \cong \mathcal{S}_{e^*}(W^\vee).$$

Proof. The first assertion has already been proved so we turn to the second. We may reduce to the case $e_n \geq 0$ and we may choose $f \in \mathbb{Z}_{\geq e_1}$. Set $e' = (f-e_n, \dots, f-e_1)$. Then it will suffice to show that

$$\mathcal{S}_e(W) \cong \mathcal{S}_{e'}(W)^\vee \otimes \left(\textstyle\bigwedge^n W\right)^{\otimes f}.$$

It even suffices to find a nontrivial natural map

$$\mathcal{S}_e(W) \otimes \mathcal{S}_{e'}(W) \to \left(\textstyle\bigwedge^n W\right)^{\otimes f} = (W^{\otimes\mathcal{T}_{(f,\dots,f)}})A_{(f,\dots,f)}^-.$$

(For this then gives a nontrivial natural map $\mathcal{S}_e(W) \to \mathcal{S}_{e'}(W)^\vee \otimes \left(\bigwedge^n W\right)^{\otimes f}$, which we can check is an isomorphism after applying a fibre functor, in which case the left and right hand sides become irreducible $\mathrm{GL}(W)$-modules.) To this end let ι denote the bijection

$$\iota : \mathcal{T}_{(f,\dots,f)} \xrightarrow{\sim} \mathcal{T}_e \amalg \mathcal{T}_{e'}$$

which sends (i, j) to (i, j) if $j \le e_i$ and to $(n+1-i, f+1-i)$ if $j > e_i$, and let ι^* denote the induced map

$$W^{\otimes \mathcal{T}_e} \otimes W^{\bigotimes \mathcal{T}_{e'}} \xrightarrow{\sim} W^{\bigotimes \mathcal{T}_{(f,\dots,f)}}.$$

Then we consider the map

$$A^-_{(f,\dots,f)} \circ \iota^* : \mathcal{S}_e(W) \otimes \mathcal{S}_{e'}(W) \to \mathcal{S}_{(f,\dots,f)}(W).$$

We must show that if W has rank n then this map is nontrivial. We can reduce this to the case of $\overline{\mathbb{Q}}$-vector spaces by applying a fibre functor. In this case let $w_1, \dots, w_n$ be a basis of W. Consider the element

$$x = (\otimes_{\mathcal{T}_e} u_t) A^-_e \otimes (\otimes_{\mathcal{T}_{e'}} v_t) A^-_{e'} \in W^{\otimes \mathcal{T}_e} \otimes W^{\otimes \mathcal{T}_{e'}}$$

where $u_{(i,j)} = w_i$ and $v_{(i,j)} = w_{n+1-i}$. Then

$$\begin{aligned}(\iota^* x) A^-_{(f,\dots,f)} &= \left(\textstyle\prod_{i=1}^f (\#\{j :\ e_j < i\})!(\#\{j :\ e_j \ge i\})!\right) (\otimes_{\mathcal{T}_{(f,\dots,f)}} x_t) A^-_{(f,\dots,f)} \\ &\neq 0,\end{aligned}$$

where $x_{(i,j)} = w_i$. The lemma follows. □

The relative cohomology of $\mathcal{A}/X_U$. If ϖ denotes the projection map from the universal abelian variety $\mathcal{A}$ to X_U then we decompose

$$R^1 \varpi_* \overline{\mathbb{Q}}_l = \bigoplus_{\tau \in \mathrm{Hom}(F, \mathbb{C})} \mathcal{L}_\tau$$

where $\mathcal{L}_\tau$ is the subsheaf of $R^1 \varpi_* \overline{\mathbb{Q}}_l$ where the action of F coming from the endomorphisms of the universal abelian variety is via $\iota^{-1}\tau$. The sheaves $\mathcal{L}_\tau$ on the inverse system of the X_U's carry a natural action of $G(\mathbb{A}^\infty)$ (coming from the action of $G(\mathbb{A}^\infty)$ on the inverse system of the $\mathcal{A}/X_U$. Let Std_τ denote the representation of $G \times_{\mathbb{Q}} \mathbb{C}$ on $V \otimes_{F,\tau} \mathbb{C}$, so that $\mathrm{Std}_{\tau c} \cong \nu\, \mathrm{Std}_\tau^\vee$. Then $\mathcal{L}_\tau \cong \mathcal{L}_{\mathrm{Std}_\tau^\vee}$ with the $G(\mathbb{A}^\infty)$-actions. We also define an action of $G(\mathbb{A}^\infty)$ on the sheaves $\overline{\mathbb{Q}}_l(1)$ by letting $g : g^* \overline{\mathbb{Q}}_l(1) \to \overline{\mathbb{Q}}_l(1)$ be $\nu(g_l)^{-1}$ times the canonical map. Then $\mathcal{L}_{\nu^m} \cong \overline{\mathbb{Q}}_l(m)$ with the $G(\mathbb{A}^\infty)$-actions. Moreover the Weil pairing gives $G(\mathbb{A}^\infty)$-equivariant isomorphisms

$$\mathcal{L}_\tau \cong \mathcal{L}_{\tau c}^\vee \otimes \overline{\mathbb{Q}}_l(-1)$$

corresponding to $\mathcal{L}_{\mathrm{Std}_\tau^\vee} \cong \mathcal{L}_{\mathrm{Std}_{\tau c}} \otimes \mathcal{L}_{\nu^{-1}}$.

Suppose that $\tilde{\xi}$ is an irreducible representation of $G \times_{\mathbb{Q}} \mathbb{C}$ with highest weight $(\tilde{b}_0, \tilde{b}_{\tau,i})_{\tau|_E=\tau_0|_E}$. Then we see that

$$\mathscr{L}_{\tilde{\xi}} \cong \left(\bigotimes_{\tau|_E=\tau_0|_E} \mathscr{S}_{(\tilde{b}_{\tau,1},\dots,\tilde{b}_{\tau,n})}(\mathscr{L}_\tau^\vee) \right) \otimes \overline{\mathbb{Q}}_l(\tilde{b}_0),$$

with their $G(\mathbb{A}^\infty)$-actions.

Note that there are natural isomorphisms $I^*\mathscr{L}_\tau \cong \mathscr{L}_{\tau c}$ and hence, by Lemma 3.1, natural isomorphisms

$$\begin{aligned} I^*\left(\bigotimes_{\tau|_E=\tau_0|_E} \mathscr{S}_{(\tilde{b}_{\tau,1},\dots,\tilde{b}_{\tau,n})}(\mathscr{L}_\tau^\vee) \right) \otimes \overline{\mathbb{Q}}_l(\tilde{b}_0) & \\ \cong \left(\bigotimes_{\tau|_E=\tau_0|_E} \mathscr{S}_{(\tilde{b}_{\tau,1},\dots,\tilde{b}_{\tau,n})}(\mathscr{L}_{\tau c}^\vee) \right) \otimes \overline{\mathbb{Q}}_l(\tilde{b}_0) & \\ \cong \left(\bigotimes_{\tau|_E=\tau_0|_E} \mathscr{S}_{(\tilde{b}_{\tau,1},\dots,\tilde{b}_{\tau,n})}(\mathscr{L}_\tau(1)) \right) \otimes \overline{\mathbb{Q}}_l(\tilde{b}_0) & \\ \cong \left(\bigotimes_{\tau|_E=\tau_0|_E} \mathscr{S}_{(-\tilde{b}_{\tau,n},\dots,-\tilde{b}_{\tau,1})}(\mathscr{L}_\tau^\vee) \right) \otimes \overline{\mathbb{Q}}_l\Big(\tilde{b}_0 + \sum_{\tau|_E=\tau_0|_E} \sum_i b_{\tau,i}\Big). & \end{aligned}$$

This isomorphism coincides up to scalar multiples with our previous isomorphism $I^*\mathscr{L}_{\tilde{\xi}} \cong \mathscr{L}_{\tilde{\xi}^\#}$ of (2-1).

Betti realizations. Fix $\sigma : \overline{F} \hookrightarrow \mathbb{C}$ which gives rise to our infinite place v of $\overline{F}$ and suppose that $\sigma|_E = \tau_0|_E$. Set $X_{U,\sigma}(\mathbb{C})$ to be the complex manifold $(X_U \times_{F,\sigma} \mathbb{C})(\mathbb{C})$. If $\tau : F \hookrightarrow \mathbb{C}$ let L_τ denote the maximal subsheaf of $R^1\varpi_*\mathbb{C}$ on $X_{U,\sigma}(\mathbb{C})$ where the action of F from endomorphisms of the universal abelian variety is via τ. The system of locally constant sheaves L_τ have a natural action of $G(\mathbb{A}^\infty)$. Also let $\mathbb{C}(1)$ denote the constant sheaf and endow the system of sheaves $\mathbb{C}(1)/X_{U,\sigma}(\mathbb{C})$ with an action of $G(\mathbb{A}^\infty)$ by letting $g : g^*\mathbb{C}(1) \to \mathbb{C}(1)$ be $|\nu(g)|^{-1}$ times the natural map. Then

$$L_\tau \cong L_{\tau c}^\vee \otimes \mathbb{C}(-1).$$

If $\tilde{\xi}$ is the irreducible representation of $G \times_{\mathbb{Q}} \mathbb{C}$ with highest weight $(\tilde{b}_0, \tilde{b}_{\tau,i})_{\tau|_E=\tau_0|_E}$, then we define a locally constant sheaf of finite-dimensional $\mathbb{C}$-vector spaces $L_{\tilde{\xi}}$ on $X_{U,\sigma}(\mathbb{C})$ as

$$\left(\bigotimes_{\tau|_E=\tau_0|_E} \mathscr{S}_{(\tilde{b}_{\tau,1},\dots,\tilde{b}_{\tau,n})}(L_\tau^\vee) \right) \otimes \mathbb{C}(\tilde{b}_0).$$

Then $L_{\tilde{\xi}}$ is the locally constant sheaf associated to the pull back of $\mathscr{L}_{\tilde{\xi}}$ to $X_U \times_{F,\sigma} \mathbb{C}$,

thought of as a sheaf of $\mathbb{C}$-vector spaces via ι^{-1}. This correspondence is $G(\mathbb{A}^\infty)$-equivariant. Note that by Lemma 3.1 if $\tilde{\xi}'$ is one-dimensional then

$$L_{\tilde{\xi}} \otimes L_{\tilde{\xi}'} \xrightarrow{\sim} L_{\tilde{\xi}\otimes\tilde{\xi}'}.$$

Let ${}^cX_{U,\sigma}(\mathbb{C})$ denote the complex conjugate complex manifold of $X_{U,\sigma}(\mathbb{C})$, that is, the same topological space but with complex conjugate charts. Then $I \times c$ induces an isomorphism

$$I \times c : X_{U,\sigma}(\mathbb{C}) \xrightarrow{\sim} {}^cX_{U,\sigma}(\mathbb{C}).$$

As we described above in the l-adic setting, Lemma 3.1 together with the isomorphisms $L_\tau \cong L_{\tau c}^\vee \otimes \mathbb{C}(-1)$ gives rise to an isomorphism

$$(I \times c)^* L_{\tilde{\xi}} \cong L_{\tilde{\xi}^\#}$$

compatible with the corresponding isomorphism in the l-adic setting ($I^*\mathscr{L}_{\tilde{\xi}} \cong \mathscr{L}_{\tilde{\xi}^\#}$).

We set

$$H^i(X_\sigma(\mathbb{C}), L_{\tilde{\xi}}) = \lim_{\to U} H^i(X_{U,\sigma}(\mathbb{C}), L_{\tilde{\xi}})$$

which is naturally a $G(\mathbb{A}^\infty)$-module and which satisfies

$$H^i(X_\sigma(\mathbb{C}), L_{\tilde{\xi}}) \cong H^i(X, \mathscr{L}_{\tilde{\xi}}) \otimes_{\overline{\mathbb{Q}}_l,\iota} \mathbb{C}$$

as $\mathbb{C}[G(\mathbb{A}^\infty)]$-modules. Again as in the l-adic setting we have a decomposition

$$H^0(X_\sigma(\mathbb{C}), L_\zeta) = \bigoplus_{\widetilde{\omega}} \mathbb{C}\upsilon_{\widetilde{\omega},B},$$

where $\widetilde{\omega}$ runs over continuous characters

$$T(\mathbb{A})/T(\mathbb{Q}) \to \mathbb{C}^\times$$

with $\widetilde{\omega}|_{T(\mathbb{R})} = \zeta^{-1}$, and where $G(\mathbb{A}^\infty)$ acts on $\upsilon_{\widetilde{\omega},B}$ via $\widetilde{\omega} \circ d$. If we define

$$I_{v,B} : H^i(X_\sigma(\mathbb{C}), L_\xi) \to H^i(X_\sigma(\mathbb{C}), L_\xi)$$

to be the composite

$$H^i(X_\sigma(\mathbb{C}), L_\xi) \xrightarrow{I\times c} H^i(X_\sigma(\mathbb{C}), L_{\xi^\#}) \xrightarrow{\cup\upsilon_{\omega,B}} H^i(X_\sigma(\mathbb{C}), L_\xi).$$

Then under the isomorphism $H^i(X_\sigma(\mathbb{C}), L_\xi) \cong H^i(X, \mathscr{L}_\xi) \otimes_{\overline{\mathbb{Q}}_l,\iota} \mathbb{C}$, this map $I_{v,B}$ corresponds to a scalar multiple of the previous map $I_v \otimes 1$.

Again we can define a map $\tilde{c}_{v,B}$ on

$$\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi, H^{n-1}(X_\sigma(\mathbb{C}), L_\xi)) \cong \mathbb{C}^n$$

to be the map which sends

$$f \mapsto I_{v,B} \circ f \circ A_\pi.$$

Then $\tilde{c}_{v,B}$ corresponds to a scalar multiple of the map $\tilde{c}_v$ previously defined on $\mathrm{Hom}_{G(\mathbb{A}^\infty)}(\iota^{-1}\pi, H^{n-1}(X, \mathscr{L}_\xi))$. Rescaling $\tilde{c}_{v,B}$ we may, and shall, suppose that $\tilde{c}_{v,B}^2 = 1$, in which case it corresponds to $\pm\tilde{c}_v$. Then it suffices to show that the trace of $\tilde{c}_{v,B}$ is ± 1.

Variation of Hodge structures I: generalities. We begin with a rather lengthy reminder about variations of pure Hodge structures on complex manifolds. We do this because we have not found a single clear reference for all the material we need, although it is all standard.

Recall that a (pure) $\mathbb{R}$-Hodge structure of weight w is a finite-dimensional $\mathbb{R}$-vector space H together with a decreasing, exhaustive and separated filtration Fil^i on the $\mathbb{C}$-vector space $H \otimes_{\mathbb{R}} \mathbb{C}$ such that

$$H \otimes_{\mathbb{R}} \mathbb{C} = \mathrm{Fil}^i(H \otimes_{\mathbb{R}} \mathbb{C}) \oplus (1 \otimes c)\,\mathrm{Fil}^{w-1-i}(H \otimes_{\mathbb{R}} \mathbb{C})$$

for all i. In this case $H \otimes_{\mathbb{R}} \mathbb{C} = \bigoplus_i H^{i,w-i}$, where

$$H^{i,w-i} = (\mathrm{Fil}^i H \otimes_{\mathbb{R}} \mathbb{C}) \cap (1 \otimes c)(\mathrm{Fil}^{w-i} H \otimes_{\mathbb{R}} \mathbb{C}).$$

By a polarization on $(H, \{\mathrm{Fil}^i\})$ we mean a perfect bilinear pairing

$$\langle\ ,\ \rangle : H \times H \to \mathbb{R}$$

such that the $\langle\ ,\ \rangle$-orthogonal complement of $\mathrm{Fil}^i H \otimes_{\mathbb{R}} \mathbb{C}$ is $\mathrm{Fil}^{w-1-i} H \otimes_{\mathbb{R}} \mathbb{C}$ and such that the following property holds. Define a sesquilinear pairing $(\ ,\ )$ on $H \otimes_{\mathbb{R}} \mathbb{C}$ by extending $\langle\ ,\ \rangle$ to a $\mathbb{C}$-bilinear pairing on $H \otimes \mathbb{C}$ and defining

$$(x, y) = \sqrt{-1}^{-w}\langle x, (1 \otimes c)y\rangle.$$

Note that $(\ ,\ )$ restricts to a perfect sesquilinear pairing on each $H^{i,w-i}$. We require that $(\ ,\ )$ is Hermitian (i.e., $(y, x) = c(x, y)$) and that the restriction of $(-1)^i(\ ,\ )$ to $H^{i,w-i}$ is positive definite. If $\phi : (H_1, \{\mathrm{Fil}_1^i\}) \to (H_2, \{\mathrm{Fil}_2^i\})$ is a map of $\mathbb{R}$-Hodge structures (i.e., a linear map $\phi : H_1 \to H_2$ such that $\phi \otimes 1$ maps $\mathrm{Fil}^i H_1 \otimes_{\mathbb{R}} \mathbb{C}$ to $\mathrm{Fil}^i H_2 \otimes_{\mathbb{R}} \mathbb{C}$ for all i) then

$$(\phi \otimes 1)(\mathrm{Fil}^i H_1 \otimes_{\mathbb{R}} \mathbb{C}) = (\mathrm{Fil}^i H_2 \otimes_{\mathbb{R}} \mathbb{C}) \cap (\phi(H_1) \otimes_{\mathbb{R}} \mathbb{C})$$

for all i. It follows that the category of $\mathbb{R}$-Hodge structures of weight w is an abelian category. The restriction of a polarization to a subobject is again a polarization and the orthogonal complement of the subobject is again a subobject. It follows that the full subcategory of polarizable pure Hodge structures is also (semisimple) abelian. The direct sums of over all integers w of the abelian category of $\mathbb{R}$-Hodge structures of weight w and of the abelian category of polarizable $\mathbb{R}$-Hodge structures of weight w are Tannakian. We will refer to them as the categories of pure $\mathbb{R}$-Hodge

structures and of pure polarizable $\mathbb{R}$-Hodge structures; although strictly speaking their objects are not pure, but direct sums of pure objects.

A (pure) $\mathbb{C}$-Hodge structure of weight w is a $\mathbb{C}$-vector space H together with two decreasing, exhaustive and separated filtrations Fil^i and $\overline{\mathrm{Fil}}^i$ on H such that $H = \mathrm{Fil}^i H \oplus \overline{\mathrm{Fil}}^{w-1-i} H$ for all i. If $\mathbb{H} = (H, \{\mathrm{Fil}^i\}, \{\overline{\mathrm{Fil}}^i\})$ is a $\mathbb{C}$-Hodge structure of weight w then we define the underlying $\mathbb{R}$-Hodge structure to be

$$(H, \{\mathrm{Fil}^i H \oplus \overline{\mathrm{Fil}}^i H\}),$$

where

$$H \otimes_{\mathbb{R}} \mathbb{C} \xrightarrow{\sim} H \oplus H \supset \mathrm{Fil}^i H \oplus \overline{\mathrm{Fil}}^i H$$

is given by $x \otimes a \mapsto (ax, ({}^c a)x)$. This establishes an equivalence of categories between $\mathbb{C}$-Hodge structures of weight w and $\mathbb{R}$-Hodge structures of weight w with an action of $\mathbb{C}$. If $\mathbb{H} = (H, \{\mathrm{Fil}^i\}, \{\overline{\mathrm{Fil}}^i\})$ is a $\mathbb{C}$-Hodge structure of weight $\mathbb{R}$ then $H = \bigoplus H^{i,w-i}$, where $H^{i,w-i} = \mathrm{Fil}^i H \cap \overline{\mathrm{Fil}}^{w-i} H$. By a polarization on $\mathbb{H}$ we mean a perfect Hermitian pairing

$$(\ ,\ ) : H \times H \to \mathbb{C},$$

such that for all i the orthogonal complement of $\mathrm{Fil}^i H$ is $\overline{\mathrm{Fil}}^{w-1-i} H$ and the restriction of $(-1)^i(\ ,\ )$ to $H^{i,w-i}$ is positive definite. This is equivalent to a polarization $\langle\ ,\ \rangle$ of the underlying $\mathbb{R}$-Hodge structure such that

$$\langle ax, y\rangle = \langle x, ({}^c a)y\rangle$$

for all $a \in \mathbb{C}$ and $x, y \in H$. The equivalence is given by

$$\langle x, y\rangle = \mathrm{Re}\ \sqrt{-1}^{w}(x, y).$$

The category of polarizable $\mathbb{C}$-Hodge structures of weight w is the full subcategory of the category of $\mathbb{C}$-Hodge structures of weight w whose objects are those that admit a polarization. It is closed under taking subobjects and quotients. By the category of (polarizable) pure $\mathbb{C}$-Hodge structures we mean the direct sum over w of the categories of (polarizable) $\mathbb{C}$-Hodge structures of weight w. They are Tannakian categories. (Again objects of these categories are not strictly speaking pure, but the direct sum of pure objects of different weights.)

If $(H, \{\mathrm{Fil}^i\})$ is an $\mathbb{R}$-Hodge structure of weight w then we define

$$(H, \{\mathrm{Fil}^i\}) \otimes \mathbb{C} = (H \otimes_{\mathbb{R}} \mathbb{C}, \{\mathrm{Fil}^i\}, \{(1 \otimes c)\,\mathrm{Fil}^i\}),$$

a $\mathbb{C}$-Hodge structure of weight w. If $(H, \{\mathrm{Fil}^i\})$ is polarizable, so is $(H, \{\mathrm{Fil}^i\}) \otimes \mathbb{C}$. (Define $(x \otimes a, y \otimes b) = \sqrt{-1}^{-w} a({}^c b)\langle x, y\rangle$.)

If $\mathbb{H} = (H, \{\mathrm{Fil}^i\}, \{\overline{\mathrm{Fil}}^i\})$ is a $\mathbb{C}$-Hodge structure we define its complex conjugate ${}^c\mathbb{H} = (H, \{\overline{\mathrm{Fil}}^i\}, \{\mathrm{Fil}^i\})$.

Recall also that a variation of $\mathbb{R}$-Hodge structures $\mathbb{H}$ of weight w on a complex manifold Y is a pair $(H, \{\mathrm{Fil}^i\})$, where H is a locally constant sheaf of finite-dimensional $\mathbb{R}$-vector spaces, where $\{\mathrm{Fil}^i\}$ is an exhaustive, separated, decreasing filtration of $H \otimes_{\mathbb{R}} \mathcal{O}_Y$ by local $\mathcal{O}_Y$-direct summands, such that

- the pull back of $\mathbb{H}$ to any point of Y is a pure $\mathbb{C}$-Hodge structure of weight w,
- and $1 \otimes d : \mathrm{Fil}^i(H \otimes_{\mathbb{R}} \mathcal{O}_Y) \to (\mathrm{Fil}^{i-1}(H \otimes_{\mathbb{R}} \mathcal{O}_Y)) \otimes_{\mathcal{O}_Y} \Omega^1_Y$.

If $\phi : \mathbb{H}_1 \to \mathbb{H}_2$ is a morphism of variation of $\mathbb{R}$-Hodge structures of weight w on Y then $(\phi \otimes 1)\,\mathrm{Fil}^i(H_1 \otimes_{\mathbb{R}} \mathcal{O}_Y) = ((\phi H_1) \otimes_{\mathbb{R}} \mathcal{O}_Y) \cap \mathrm{Fil}^i(H_2 \otimes_{\mathbb{R}} \mathcal{O}_Y)$. It follows that the category of variations of $\mathbb{R}$-Hodge structures of weight w on Y is abelian. By a polarization on $\mathbb{H}$ we mean a perfect bilinear pairing

$$\langle\ ,\ \rangle : H \times H \to \mathbb{R}$$

whose pull-back to any point of Y is a polarization. The full subcategory of the category of variations of $\mathbb{R}$-Hodge structures of weight w on Y consisting of polarizable objects is a semisimple abelian subcategory closed under taking subobjects and quotients. By the category of (polarizable) pure variations of $\mathbb{R}$-Hodge structures on Y we mean the direct sum over w of the categories of (polarizable) variations of $\mathbb{R}$-Hodge structures of weight w on Y. They are Tannakian categories. (Again objects of these categories are not strictly speaking pure, but the direct sum of pure objects of different weights.)

The pull back of a (polarizable) variation of $\mathbb{R}$-Hodge structures of weight w by any morphism is clearly again a (polarizable) variation of $\mathbb{R}$-Hodge structures of weight w. If Y is a compact Kähler manifold and $\mathbb{H}$ is a polarizable variation of $\mathbb{R}$-Hodge structures of weight w on Y then $H^i(Y, H)$ has a natural structure of a polarizable $\mathbb{R}$-Hodge structure of weight $i + w$ [Zucker 1979, Theorem (2.9)]. More precisely, we define $\Omega^\bullet(\mathbb{H})$ to be the complex

$$H \otimes_{\mathbb{R}} \mathcal{O}_Y \to H \otimes_{\mathbb{R}} \Omega^1_Y \to H \otimes_{\mathbb{R}} \Omega^2_Y \to \cdots,$$

and filter it by setting $\mathrm{Fil}^i\,\Omega^\bullet(\mathbb{H})$ to be the subcomplex

$$\mathrm{Fil}^i(H \otimes_{\mathbb{R}} \mathcal{O}_Y) \to \mathrm{Fil}^{i-1}(H \otimes_{\mathbb{R}} \mathcal{O}_Y) \otimes_{\mathcal{O}_Y} \Omega^1_Y \to \mathrm{Fil}^{i-2}(H \otimes_{\mathbb{R}} \mathcal{O}_Y) \otimes_{\mathcal{O}_Y} \Omega^2_Y \to \cdots.$$

Then the spectral sequence

$$E_1^{i,j} = \mathbb{H}^{i+j}(Y, \mathrm{gr}^i\,\Omega^\bullet(\mathbb{H})) \Rightarrow \mathbb{H}^{i+j}(Y, \Omega^\bullet(\mathbb{H})) \cong H^{i+j}(Y, H) \otimes_{\mathbb{R}} \mathbb{C}$$

degenerates at E_1 and defines the (Hodge) filtration on $H^i(Y, H) \otimes_{\mathbb{R}} \mathbb{C}$.

If $f : X \to Y$ is a smooth family of compact Kähler manifolds over a complex manifold Y then $R^i f_* \mathbb{R}$ is naturally a polarizable variation of $\mathbb{R}$-Hodge structures

of weight i. (See the Introduction and first two sections of [Zucker 1979].) More precisely, let $\Omega^{\bullet}_{X/Y}$ denote the complex

$$\mathcal{O}_X \to \Omega^1_{X/Y} \to \Omega^2_{X/Y} \to \cdots$$

and let $\mathrm{Fil}^i\,\Omega^{\bullet}_{X/Y}$ denote the subcomplex

$$\Omega^i_{X/Y} \to \Omega^{i+1}_{X/Y} \to \cdots$$

Then the filtration on $(R^i f_*\mathbb{R})\otimes\mathcal{O}_Y \cong \mathbb{R}^i f_*\Omega^{\bullet}_{X/Y}$ is the one induced by the spectral sequence

$$E_1^{i,j} = R^j f_*\Omega^i_{X/Y} \Rightarrow \mathbb{R}^{i+j} f_*\Omega^{\bullet}_{X/Y} \cong R^{i+j} f_*\mathbb{R}\otimes_{\mathbb{R}}\mathcal{O}_Y.$$

If moreover Y is a compact Kähler manifold then the Leray spectral sequence

$$E_2^{i,j} = H^i(Y, R^j f_*\mathbb{R}) \Rightarrow H^{i+j}(X, \mathbb{R})$$

degenerates at E_2 and the $\mathbb{R}$-Hodge structure on $H^i(Y, R^j f_*\mathbb{R})$ is compatible with the $\mathbb{R}$-Hodge structure on $H^{i+j}(X, \mathbb{R})$ [Zucker 1979, Proposition (2.16)].

By a variation of $\mathbb{C}$-Hodge structures $\mathbb{H}$ of weight w on a complex manifold Y we mean a triple $(H, \{\mathrm{Fil}^i\}, \{\overline{\mathrm{Fil}}^i\})$, where H is a locally constant sheaf of finite-dimensional $\mathbb{C}$-vector spaces, $\{\mathrm{Fil}^i\}$ is an exhaustive, separated, decreasing filtration of $H\otimes_{\mathbb{C}}\mathcal{O}_Y$ by local $\mathcal{O}_Y$-direct summands, and $\{\overline{\mathrm{Fil}}^i\}$ is an exhaustive, separated, decreasing filtration of $H\otimes_{\mathbb{C}}\mathcal{O}_{{}^cY}$ by local $\mathcal{O}_{{}^cY}$-direct summands such that

- the pull back of $\mathbb{H}$ to any point of Y is a pure $\mathbb{C}$-Hodge structure of weight w,
- $1\otimes d : \mathrm{Fil}^i(H\otimes_{\mathbb{C}}\mathcal{O}_Y) \to (\mathrm{Fil}^{i-1}(H\otimes_{\mathbb{C}}\mathcal{O}_Y))\otimes_{\mathcal{O}_Y}\Omega^1_Y$,
- and $1\otimes d : \overline{\mathrm{Fil}}^i(H\otimes_{\mathbb{C}}\mathcal{O}_{{}^cY}) \to (\overline{\mathrm{Fil}}^{i-1}(H\otimes_{\mathbb{C}}\mathcal{O}_{{}^cY}))\otimes_{\mathcal{O}_{{}^cY}}\Omega^1_{{}^cY}$.

(Recall that ${}^cY$ denote the same underlying topological space as Y but with complex conjugate charts.) If $\mathbb{H}$ is a variation of $\mathbb{C}$-Hodge structures of weight w on Y then $(H, \{\mathrm{Fil}^i \oplus (1\otimes c)\overline{\mathrm{Fil}}^i\})$ is a variation of $\mathbb{R}$-Hodge structures of weight w on Y, where we think of $\mathrm{Fil}^i \oplus (1\otimes c)\overline{\mathrm{Fil}}^i$ contained in

$$(H\otimes_{\mathbb{C}}\mathcal{O}_Y)\oplus(1\otimes c)(H\otimes_{\mathbb{C}}\mathcal{O}_{{}^cY}) = (H\otimes_{\mathbb{C}}\mathcal{O}_Y)\oplus(H\otimes_{\mathbb{C},c}\mathcal{O}_Y) = H\otimes_{\mathbb{R}}\mathcal{O}_Y.$$

This establishes an equivalence of categories between variations of $\mathbb{C}$-Hodge structures of weight w on Y and variations of $\mathbb{R}$-Hodge structures of weight w on Y together with an action of $\mathbb{C}$. Thus the category of variations of $\mathbb{C}$-Hodge structures of weight w on Y is abelian. By the category of pure variations of $\mathbb{C}$-Hodge structures of weight w on Y we mean the direct sum over w of the categories of variations of $\mathbb{C}$-Hodge structures of weight w. It is a Tannakian category. (Again the objects are not strictly speaking pure, but the direct sum of pure objects of different weights.)

By a polarization of a variation of $\mathbb{C}$-Hodge structures of weight w on Y we mean a perfect Hermitian pairing

$$(\ ,\ ) : H \times H \to \mathbb{C}$$

such that the pull back to any point of Y is a polarization. The category of polarizable $\mathbb{C}$-Hodge structures of weight w on Y is equivalent to the category of $\mathbb{R}$-Hodge structures of weight w on Y together with an action of $\mathbb{C}$, which admit a polarization for which the adjoint of any $a \in \mathbb{C}$ is ${}^c a$. Thus the category of polarizable variations of $\mathbb{C}$-Hodge structures of weight w on Y is a full abelian subcategory of the category of variations of $\mathbb{C}$-Hodge structures of weight w on Y and is closed under subobjects and quotients. By the category of pure polarizable variations of $\mathbb{C}$-Hodge structures of weight w on Y we mean the direct sum over w of the categories of variations of $\mathbb{C}$-Hodge structures of weight w. It is again a Tannakian category. (And again the objects are not strictly speaking pure, but the direct sum of pure objects of different weights.)

If $(H, \{\mathrm{Fil}^i\})$ is a variation $\mathbb{R}$-Hodge structures of weight w on Y then we define

$$(H, \{\mathrm{Fil}^i\}) \otimes \mathbb{C} = (H \otimes_{\mathbb{R}} \mathbb{C}, \{\mathrm{Fil}^i\}, \{(1 \otimes c)\,\mathrm{Fil}^i\}),$$

a variation of $\mathbb{C}$-Hodge structures of weight w on Y. If $(H, \{\mathrm{Fil}^i\})$ is polarizable then so is $(H, \{\mathrm{Fil}^i\}) \otimes \mathbb{C}$. (Define $(x \otimes a, y \otimes b) = \sqrt{-1}^{-w} a({}^c b)\langle x, y\rangle$.)

If $\mathbb{H} = (H, \{\mathrm{Fil}^i\}, \{\overline{\mathrm{Fil}}^i\})$ is a variation of $\mathbb{C}$-Hodge structures of weight w on Y we define its complex conjugate ${}^c\mathbb{H} = (H, \{\overline{\mathrm{Fil}}^i\}, \{\mathrm{Fil}^i\})$.

The pull back of a (polarizable) variation of $\mathbb{C}$-Hodge structures of weight w by any morphism is clearly again a (polarizable) variation of $\mathbb{C}$-Hodge structures of weight w. If Y is a compact Kähler manifold and $\mathbb{H}$ is a polarizable variation of $\mathbb{C}$-Hodge structures of weight w on Y then $H^i(Y, H)$ has a natural structure of a polarizable $\mathbb{C}$-Hodge structure of weight $i + w$). More precisely, define $\Omega^\bullet_Y(\mathbb{H})$ to be the complex

$$H \otimes_{\mathbb{C}} \mathbb{O}_Y \to H \otimes_{\mathbb{C}} \Omega^1_Y \to H \otimes_{\mathbb{C}} \Omega^2_Y \to \cdots$$

filtered by setting $\mathrm{Fil}^i\, \Omega^\bullet_Y(\mathbb{H})$ to be the subcomplex

$$\mathrm{Fil}^i(H \otimes_{\mathbb{C}} \mathbb{O}_Y) \to \mathrm{Fil}^{i-1}(H \otimes_{\mathbb{C}} \mathbb{O}_Y) \otimes_{\mathbb{O}_Y} \Omega^1_Y \to \mathrm{Fil}^{i-2}(H \otimes_{\mathbb{C}} \mathbb{O}_Y) \otimes_{\mathbb{O}_Y} \Omega^2_Y \to \cdots;$$

similarly $\Omega^\bullet_{{}^c Y}(\mathbb{H})$ is the complex

$$H \otimes_{\mathbb{C}} \mathbb{O}_{{}^c Y} \to H \otimes_{\mathbb{C}} \Omega^1_{{}^c Y} \to H \otimes_{\mathbb{C}} \Omega^2 Y \to \cdots$$

with $\mathrm{Fil}^i\, \Omega^\bullet_{{}^c Y}(\mathbb{H})$) the subcomplex

$$\overline{\mathrm{Fil}}^i(H \otimes_{\mathbb{C}} \mathbb{O}_{{}^c Y}) \to \overline{\mathrm{Fil}}^{i-1}(H \otimes_{\mathbb{C}} \mathbb{O}_{{}^c Y}) \otimes_{\mathbb{O}_{{}^c Y}} \Omega^1_{{}^c Y} \to \overline{\mathrm{Fil}}^{i-2}(H \otimes_{\mathbb{C}} \mathbb{O}_{{}^c Y}) \otimes_{\mathbb{O}_{{}^c Y}} \Omega^2_{{}^c Y} \ldots.$$

Then the spectral sequences

$$E_1^{i,j} = \mathbb{H}^{i+j}(Y, \operatorname{gr}^i \Omega_Y^\bullet(\mathbb{H})) \Rightarrow \mathbb{H}^{i+j}(Y, \Omega_Y^\bullet(\mathbb{H})) \cong H^{i+j}(Y, H)$$

and

$$\bar{E}_1^{i,j} = \mathbb{H}^{i+j}({}^cY, \operatorname{gr}^i \Omega_{{}^cY}^\bullet(\mathbb{H})) \Rightarrow \mathbb{H}^{i+j}(Y, \Omega_{{}^cY}^\bullet(\mathbb{H})) \cong H^{i+j}(Y, H)$$

degenerate at E_1 and define the (Hodge) filtrations on $H^i(Y, H)$. (This can be easily deduced from the corresponding facts for variations of $\mathbb{R}$-Hodge structures.)

If $f : X \to Y$ is a smooth family of compact Kähler manifolds over a complex manifold Y then $R^i f_*\mathbb{C}$ is naturally a polarizable variation of $\mathbb{C}$-Hodge structures of weight i. More precisely, the filtrations on $(R^i f_*\mathbb{C}) \otimes_{\mathbb{C}} \mathcal{O}_Y \cong \mathbb{R}^i f_* \Omega_{X/Y}^\bullet$ and $(R^i f_*\mathbb{C}) \otimes_{\mathbb{C}} \mathcal{O}_{{}^cY} \cong \mathbb{R}^i f_* \Omega_{{}^cX/{}^cY}^\bullet$ are the ones induced by the spectral sequences

$$E_1^{i,j} = R^j f_* \Omega_{X/Y}^i \Rightarrow \mathbb{R}^{i+j} f_* \Omega_{X/Y}^\bullet \cong R^{i+j} f_* \mathbb{C} \otimes_{\mathbb{C}} \mathcal{O}_Y$$

and

$$\bar{E}_1^{i,j} = R^j f_* \Omega_{{}^cX/{}^cY}^i \Rightarrow \mathbb{R}^{i+j} f_* \Omega_{{}^cX/{}^cY}^\bullet \cong R^{i+j} f_* \mathbb{C} \otimes_{\mathbb{C}} \mathcal{O}_{{}^cY}.$$

If moreover Y is a compact Kähler manifold then the Leray spectral sequence

$$E_2^{i,j} = H^i(Y, R^j f_*\mathbb{C}) \Rightarrow H^{i+j}(X, \mathbb{C})$$

degenerates at E_2 and the $\mathbb{C}$-Hodge structure on $H^i(Y, R^j f_*\mathbb{C})$ is compatible with the $\mathbb{C}$-Hodge structure on $H^{i+j}(X, \mathbb{C})$. (Again this is all easily deduced from the case of $\mathbb{R}$-Hodge structures.)

For example $\mathbb{C}(m)$ is the variation of pure $\mathbb{C}$-Hodge structures of weight $-2m$ with underlying locally constant sheaf $\mathbb{C}$ and with $\operatorname{Fil}^i = (0)$ and $\overline{\operatorname{Fil}}^i = (0)$ for $i > -m$, but with Fil^i and $\overline{\operatorname{Fil}}^i$ everything for $i \leq m$.

If $\mathbb{H} = (H, \{\operatorname{Fil}^i\}, \{\overline{\operatorname{Fil}}^i\})$ is a variation of pure $\mathbb{C}$-Hodge structures of weight w on Y we define a variation pure $\mathbb{C}$-Hodge structures $\mathbb{H}\{j_1, j_2\}$ of weight $w + j_1 + j_2$ on Y by setting $H\{j_1, j_2\} = H$ and

$$\operatorname{Fil}^i H\{j_1, j_2\} \otimes_{\mathbb{C}} \mathcal{O}_Y = \operatorname{Fil}^{i-j_1} H \otimes_{\mathbb{C}} \mathcal{O}_Y,$$
$$\overline{\operatorname{Fil}}^i H\{j_1, j_2\} \otimes_{\mathbb{C}} \mathcal{O}_{{}^cY} = \overline{\operatorname{Fil}}^{i-j_2} H \otimes_{\mathbb{C}} \mathcal{O}_{{}^cY}.$$

Thus $\mathbb{C}(j) = \mathbb{C}(0)\{-j, -j\}$.

Variation of Hodge structures II. We will give $\mathbb{C}(j)$ (the constant variation of pure $\mathbb{C}$-Hodge structures of weight $-2j$ on $X_{U,\sigma}(\mathbb{C})$) an action of $G(\mathbb{A}^\infty)$ by letting $g : g^*\mathbb{C}(j) \to \mathbb{C}(j)$ be $|\nu(g)^{-j}|$ times the natural map. If $\mathbb{H}/X_{U,\sigma}(\mathbb{C})$ is a collection of variations of pure $\mathbb{C}$-Hodge structures with an action of $G(\mathbb{A}^\infty)$ we will give $\mathbb{H}\{j_1, j_2\}$ the action induced from the one on $\mathbb{H}$. Thus the actions of $G(\mathbb{A}^\infty)$ on $\mathbb{C}(j)$ and $\mathbb{C}(0)\{-j, -j\}$ are different.

$R^1\varpi_*\mathbb{C}$ is a variation of pure $\mathbb{C}$-Hodge structures of weight 1 on $X_{U,\sigma}(\mathbb{C})$ and we can decompose

$$R^1\varpi_*\mathbb{C} = \bigoplus_{\tau\in\mathrm{Hom}(F,\mathbb{C})} \mathbb{L}_\tau$$

where $\mathbb{L}_\tau$ is a variation of pure $\mathbb{C}$-Hodge structures of weight 1 extending L_τ. The projective system of variations of pure $\mathbb{C}$-Hodge structures $\mathbb{L}_\tau/X_{U,\sigma}(\mathbb{C})$ as U varies has an action of $G(\mathbb{A}^\infty)$. We have $G(\mathbb{A}^\infty)$-equivariant isomorphisms

$$\mathbb{L}_\tau \cong \mathbb{L}_{\tau c}^\vee \otimes \mathbb{C}(-1).$$

Also, if $\sigma, \tau \in \mathrm{Hom}_{E,\tau_0}(F, \mathbb{C})$ then

$$\left(\bigwedge^n \mathbb{L}_\tau\right)/X_{U,\sigma}(\mathbb{C})$$

is noncanonically isomorphic to $\mathbb{C}\{0, n\}$ if $\sigma \neq \tau$ and to $\mathbb{C}\{1, n-1\}$ if $\sigma = \tau$. This identification is not $G(\mathbb{A}^\infty)$-equivariant.

For $\tilde{\xi}$ an irreducible representation of $G\times_{\mathbb{Q}}\mathbb{C}$ with highest weight $(\tilde{b}_0, \tilde{b}_{\tau,i})$, we can then define a variation of pure $\mathbb{C}$-Hodge structures $\mathbb{L}_{\tilde{\xi}}$ of weight

$$-2\tilde{b}_0 - \sum_{\tau|_E=\tau_0|_E} \sum_i \tilde{b}_{\tau,i}$$

extending $L_{\tilde{\xi}}$ by

$$\mathbb{L}_{\tilde{\xi}} = \left(\bigotimes_{\tau|_E=\tau_0|_E} \mathcal{S}_{(\tilde{b}_{\tau,1},\ldots,\tilde{b}_{\tau,n})}(\mathbb{L}_\tau^\vee) \right) \otimes \mathbb{C}(\tilde{b}_0).$$

Again the system $\mathbb{L}_{\tilde{\xi}}/X_{U,\sigma}(\mathbb{C})$ has an action of $G(\mathbb{A}^\infty)$. Again by Lemma 3.1 we see that if $\tilde{\xi}'$ is one-dimensional then there is a natural isomorphism

$$\mathbb{L}_{\tilde{\xi}} \otimes \mathbb{L}_{\tilde{\xi}'} \xrightarrow{\sim} \mathbb{L}_{\tilde{\xi}\otimes\tilde{\xi}'}.$$

We set

$$H^i(X_\sigma(\mathbb{C}), \mathbb{L}_{\tilde{\xi}}) = \lim_{\to U} H^i(X_{U,\sigma}(\mathbb{C}), \mathbb{L}_{\tilde{\xi}}).$$

It is a direct limit of pure $\mathbb{C}$-Hodge structures with an action of $G(\mathbb{A}^\infty)$, such that the fixed subspace of any open subgroup of $G(\mathbb{A}^\infty)$ is a (finite-dimensional) pure $\mathbb{C}$-Hodge structure of weight $w = i - 2\tilde{b}_0 - \left(\sum_{\tau|_E=\tau_0|_E} \sum_j \tilde{b}_{\tau,j}\right)$.

If $\tilde{b}_{\tau,j} = \tilde{b}_\tau$ is independent of j for all $\tau \in \mathrm{Hom}_{E,\tau_0}(F, \mathbb{C})$ and if $\sigma|_E = \tau_0|_E$ then

$$\mathbb{L}_{\tilde{\xi}} \cong \mathbb{C}(0)\{-\tilde{b}_\sigma - \tilde{b}_0, \tilde{b}_\sigma - \tilde{b}_0 - n \sum_{\tau\in\mathrm{Hom}_{E,\tau_0}(E,\mathbb{C})} \tilde{b}_\tau\}$$

noncanonically on $X_{U,\sigma}(\mathbb{C})$. If

$$\widetilde{\omega} : T(\mathbb{A})/T(\mathbb{Q}) \longrightarrow \mathbb{C}^\times$$

is a continuous character with $\widetilde{\omega}|_{T(\mathbb{R})} = \tilde{\xi}^{-1}$ then $\upsilon_{\widetilde{\omega},B}$ spans a sub pure $\mathbb{C}$-Hodge structure of $H^0(X_\sigma(\mathbb{C}), \mathbb{L}_{\tilde{\xi}})$ isomorphic to

$$\mathbb{C}(0)\{-\tilde{b}_\sigma - \tilde{b}_0, \tilde{b}_\sigma - \tilde{b}_0 - n \sum_{\tau \in \mathrm{Hom}_{E,\tau_0}(E,\mathbb{C})} \tilde{b}_\tau\}.$$

The choice of $\widetilde{\omega}$ fixes an equivariant isomorphism

$$\mathbb{L}_{\tilde{\xi}} \cong \mathbb{C}(0)\{-\tilde{b}_\sigma - \tilde{b}_0, \tilde{b}_\sigma - \tilde{b}_0 - n \sum_{\tau \in \mathrm{Hom}_{E,\tau_0}(E,\mathbb{C})} \tilde{b}_\tau\}(\widetilde{\omega} \circ d).$$

The map $(I \times c) : X_{U,\sigma}(\mathbb{C}) \to {}^c X_{U,\sigma}(\mathbb{C})$ lifts to a map $\mathscr{A}_\sigma(\mathbb{C}) \to {}^c\mathscr{A}_\sigma(\mathbb{C})$. We deduce that there is a natural isomorphism

$$(I \times c)^* \mathbb{L}_\tau \cong {}^c\mathbb{L}_{\tau c},$$

and hence applying Lemma 3.1 and the isomorphism $\mathbb{L}_\tau \cong \mathbb{L}_{\tau c}^\vee \otimes \mathbb{C}(-1)$ we get natural isomorphisms

$$(I \times c)^* \mathbb{L}_{\tilde{\xi}} \cong {}^c\mathbb{L}_{\tilde{\xi}^\#}$$

extending our previous isomorphism $(I \times c)^* L_{\tilde{\xi}} \cong L_{\tilde{\xi}^\#}$. Thus we get maps

$$H^i(X_\sigma(\mathbb{C}), \mathbb{L}_{\tilde{\xi}}) \to H^i({}^c X_\sigma(\mathbb{C}), {}^c\mathbb{L}_{\tilde{\xi}^\#}) \cong {}^c H^i(X_\sigma(\mathbb{C}), \mathbb{L}_{\tilde{\xi}^\#}).$$

Now suppose that $\sigma|_E = \tau_0|_E$. The line $\mathbb{C}\upsilon_{\omega,B}$ is a subpure $\mathbb{C}$-Hodge structure of $H^0({}^c X_\sigma(\mathbb{C}), {}^c\mathbb{L}_\zeta)$ isomorphic to $\mathbb{C}\{\gamma, -\gamma\}$ with

$$\gamma = \alpha + 2\beta_\sigma - n \sum_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C})} (\beta_\tau + \alpha/2).$$

Thus the cup product map

$$\cup \upsilon_{\omega,B} : {}^c\mathbb{L}_{\xi^\#} \to ({}^c\mathbb{L}_\xi)\{-\gamma, \gamma\}$$

is a map of variations of pure $\mathbb{C}$-Hodge structures. Thus the map

$$I_{v,B} : H^i(X_\sigma(\mathbb{C}), L_\xi) \to H^i(X_\sigma(\mathbb{C}), L_\xi)$$

extends to a map of pure $\mathbb{C}$-Hodge structures

$$I_{v,B} : H^i(X_\sigma(\mathbb{C}), \mathbb{L}_\xi) \to ({}^c H^i(X_\sigma(\mathbb{C}), \mathbb{L}_\xi))\{-\gamma, \gamma\},$$

or to a map of pure $\mathbb{C}$-Hodge structures

$$I_{v,B} : H^i(X_\sigma(\mathbb{C}), \mathbb{L}_\xi)\{\epsilon + \beta_\sigma, \epsilon' - \alpha - \beta_\sigma\} \to {}^c(H^i(X_\sigma(\mathbb{C}), \mathbb{L}_\xi)\{\epsilon + \beta_\sigma, \epsilon' - \alpha - \beta_\sigma\}).$$

$\Big($Note that $\epsilon' - \alpha - \beta_\sigma - (\epsilon + \beta_\sigma) = -\alpha - 2\beta_\sigma + n \sum_{\tau \in \mathrm{Hom}_{E,\tau_0}(F,\mathbb{C})} (\beta_\tau + \alpha/2) = -\gamma.\Big)$

If we set

$$\mathbb{H} = \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi,\, H^{n-1}(X_\sigma(\mathbb{C}), \mathbb{L}_\xi))\{\epsilon + \beta_\sigma, \epsilon' - \alpha - \beta_\sigma\},$$

then $\mathbb{H}$ is a pure $\mathbb{C}$-Hodge structure of weight $w = n - 1 - \alpha \in 2\mathbb{Z}$. We see that $\tilde{c}_{v,B}$ extends to a map of pure $\mathbb{C}$-Hodge structures:

$$\tilde{c}_{v,B} : \mathbb{H} \to {}^c\mathbb{H}$$

with $\tilde{c}_{v,B}^2 = 1$. Moreover we see that $\tilde{c}_{v,B}$ interchanges $\mathrm{Fil}^{w/2-1}\mathbb{H}$ and $\overline{\mathrm{Fil}}^{w/2-1}\mathbb{H}$, and that these two spaces have trivial intersection. We deduce that

$$\begin{aligned}
|\operatorname{tr} \tilde{c}_{v,B}| &\le n - 2\dim_{\mathbb{C}} \mathrm{Fil}^{w/2-1}\mathbb{H} \\
&= \dim_{\mathbb{C}} \overline{\mathrm{Fil}}^{w/2}\mathbb{H} - \dim_{\mathbb{C}} \mathrm{Fil}^{w/2-1}\mathbb{H} \\
&= \dim_{\mathbb{C}} \mathrm{Fil}^{w/2}\mathbb{H} - \dim_{\mathbb{C}} \mathrm{Fil}^{w/2-1}\mathbb{H} \\
&= \dim_{\mathbb{C}} \mathrm{gr}^{w/2}\mathbb{H} = \mathrm{gr}^{w/2-\epsilon-\beta_\sigma} \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi,\, H^{n-1}(X_\sigma(\mathbb{C}), \mathbb{L}_\xi)).
\end{aligned}$$

Cupping with $\upsilon_{\delta_{E/\mathbb{Q}}\circ\nu,B}$ shows that

$$\begin{aligned}
&\dim_{\mathbb{C}} \mathrm{gr}^{w/2-\epsilon-\beta_\sigma} \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi,\, H^{n-1}(X_\sigma(\mathbb{C}), \mathbb{L}_\xi)) \\
&\qquad = \dim_{\mathbb{C}} \mathrm{gr}^{w/2-\epsilon-\beta_\sigma} \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi \otimes (\delta_{E/\mathbb{Q}} \circ \nu),\, H^{n-1}(X_\sigma(\mathbb{C}), \mathbb{L}_\xi)).
\end{aligned}$$

Thus it suffices to show that

$$\dim_{\mathbb{C}} \bigoplus_{\mathrm{BC}(\tilde{\pi}) = (\psi^\infty, \Pi_F^\infty \otimes \phi^\infty)} \mathrm{gr}^{w/2-\epsilon-\beta_\sigma} \mathrm{Hom}_{G(\mathbb{A}^\infty)}(\pi,\, H^{n-1}(X_\sigma(\mathbb{C}), \mathbb{L}_\xi)) \le 2.$$

However the proof of Corollary 6.7 of [Shin 2011] shows this. (Note that the constant $C_G = \tau(G)\#\ker^1(\mathbb{Q}, G)$ of [Shin 2011] in our case equals 2.) So we have finally completed the proof of Proposition 1.2.

Acknowledgements

The author would like to thank Frank Calegari for raising the question studied in this paper and for his persistence in urging the author to complete this project. The author would also like to thank Sug-Woo Shin for patiently answering his questions about [Shin 2011] and the referees for their helpful comments.

References

[Bellaïche and Chenevier 2011] J. Bellaïche and G. Chenevier, "The sign of Galois representations attached to automorphic forms for unitary groups", *Compos. Math.* **147**:5 (2011), 1337–1352. MR 2834723 Zbl 05961170

[Caraiani 2010] A. Caraiani, "Local-global compatibility and the action of monodromy on nearby cycles", preprint, 2010. arXiv 1010.2188v1

[Chenevier and Harris 2011] G. Chenevier and M. Harris, "Construction of automorphic Galois representations", preprint, 2011, Available at http://www.math.jussieu.fr/~cornut/Temp/CheHar.pdf.

[Clozel et al. 2008] L. Clozel, M. Harris, and R. Taylor, "Automorphy for some l-adic lifts of automorphic mod l Galois representations", *Publ. Math. Inst. Hautes Études Sci.* 108 (2008), 1–181. MR 2010j:11082 Zbl 1169.11020

[Clozel et al. 2011] L. Clozel, M. Harris, and J.-P. Labesse, "Construction of automorphic Galois representations, I", pp. 497–527 in *On the stabilization of the trace formula*, edited by L. Clozel, M. Harris, J.-P. Labesse and B.-C. Ngô, International Press, Somerville, MA, 2011. MR 2856383

[Deligne 1990] P. Deligne, "Catégories tannakiennes", pp. 111–195 in *The Grothendieck Festschrift*, vol. 2, edited by P. Cartier et al., Progr. Math. **87**, Birkhäuser, Boston, 1990. MR 92d:14002 Zbl 0727.14010

[Fulton and Harris 1991] W. Fulton and J. Harris, *Representation theory: a first course*, Graduate Texts in Mathematics **129**, Springer, New York, 1991. MR 93a:20069 Zbl 0744.22001

[Harris and Taylor 2001] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Math. Studies **151**, Princeton Univ. Press, 2001. MR 2002m:11050 Zbl 1036.11027

[Kottwitz 1992] R. E. Kottwitz, "Points on some Shimura varieties over finite fields", *J. Amer. Math. Soc.* **5**:2 (1992), 373–444. MR 93a:11053 Zbl 0796.14014

[Milne 2005] J. Milne, "Introduction to Shimura Varieties", pp. 265–378 in *Harmonic analysis, the trace formula, and Shimura varieties* (Toronto, 2003), edited by J. Arthur et al., Clay Math. Proc. **4**, Amer. Math. Soc., Providence, RI, 2005. MR 2006m:11087 Zbl 1148.14011

[Shin 2011] S. W. Shin, "Galois representations arising from some compact Shimura varieties", *Ann. of Math.* (2) **173**:3 (2011), 1645–1741. MR 2800722 Zbl 05960691

[Taylor and Yoshida 2007] R. Taylor and T. Yoshida, "Compatibility of local and global Langlands correspondences", *J. Amer. Math. Soc.* **20**:2 (2007), 467–493. MR 2007k:11193 Zbl 1210.11118

[Zucker 1979] S. Zucker, "Hodge theory with degenerating coefficients. L_2 cohomology in the Poincaré metric", *Ann. of Math.* (2) **109**:3 (1979), 415–476. MR 81a:14002 Zbl 0446.14002

Communicated by Bjorn Poonen
Received 2010-04-29 Revised 2011-01-02 Accepted 2011-02-20

rtaylor@math.harvard.edu *Department of Mathematics, Harvard University, One Oxford Street, Cambridge, MA 02138, United States*
http://www.math.harvard.edu/~rtaylor

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

Betti numbers of graded modules and the multiplicity conjecture in the non-Cohen–Macaulay case

Mats Boij and Jonas Söderberg

We use results of Eisenbud and Schreyer to prove that any Betti diagram of a graded module over a standard graded polynomial ring is a positive linear combination of Betti diagrams of modules with a pure resolution. This implies the multiplicity conjecture of Herzog, Huneke, and Srinivasan for modules that are not necessarily Cohen–Macaulay and also implies a generalized version of these inequalities. We also give a combinatorial proof of the convexity of the simplicial fan spanned by pure diagrams.

1. Introduction

The formula for the multiplicity of a standard graded algebra with a pure resolution found by C. Huneke and M. Miller [Huneke and Miller 1985], led to the formulation of the multiplicity conjecture by Huneke and H. Srinivasan, which was later generalized and published by J. Herzog and Srinivasan [Herzog and Srinivasan 1998]. J. Migliore, U. Nagel and T. Römer [Migliore et al. 2008a] extended the conjecture to modules and strengthened it.

In a series of papers [Srinivasan 1998; Gold 2003; Guardo and Van Tuyl 2005; Migliore et al. 2005; 2008b; Römer 2005; Francisco 2006; Herzog and Zheng 2006; Kubitzke and Welker 2006; Miró-Roig 2006; Francisco and Van Tuyl 2007; Gold et al. 2007; Seo and Srinivasan 2007; Zanello 2007; Goff 2008; Hibi and Singla 2008; Puthenpurakal 2008] different versions of the multiplicity conjecture have been proven in many special cases. An exposition of most of these results can be found in [Francisco and Srinivasan 2007].

Recently D. Eisenbud and F.-O. Schreyer [2009] proved the strongest version of the conjecture for Cohen–Macaulay modules by proving a set of conjectures formulated in [Boij and Söderberg 2008] on the set of possible Betti diagrams up to multiplication by positive rational numbers. Eisenbud and Schreyer introduced a set of linear functionals defined on the space of possible Betti diagrams. The

MSC2000: primary 13D02; secondary 13A02.
Keywords: graded modules, Betti numbers, multiplicity conjecture.

linear functionals are given by certain cohomology tables of vector bundles on $\mathbb{P}^{n-1}$, and they showed that the supporting hyperplanes of the exterior facets of the simplicial fan given by the pure Betti diagrams are given by the vanishing of these linear functionals, while the functionals are nonnegative on the Betti diagram of any minimal free resolution.

In this paper we generalize the construction given in [Boij and Söderberg 2008] to include Cohen–Macaulay pure Betti diagrams of various codimensions. We show that linear functionals similar to those introduced by Eisenbud and Schreyer define the supporting hyperplanes of the simplicial fan. Furthermore, these new linear functionals are limits of the Eisenbud–Schreyer functionals, which allows us to conclude that all Betti diagrams of graded modules can be uniquely written as positive linear combinations of pure diagrams in totally ordered chains. Together with the existence of modules with pure resolutions, proved by D. Eisenbud, G. Fløystad and J. Weyman [Eisenbud et al. 2011] in characteristic zero and by Eisenbud and Schreyer [Eisenbud and Schreyer 2009] in general, this gives a complete classification of the possible Betti diagrams up to multiplication by scalars.

As a consequence, we get the multiplicity conjecture for algebras and modules that are not necessarily Cohen–Macaulay. In fact, we get a stronger version of the inequalities of the multiplicity conjecture in terms of the Hilbert series of the module which is bounded from below by the Hilbert series corresponding to the lowest shifts in a minimal free resolution while it is bounded from above by the Hilbert series corresponding to the highest shifts in the first $s+1$ terms of the resolution, where s is the codimension of the module.

We also give a combinatorial proof of the convexity of the simplicial fan spanned by the pure diagrams, even though this convexity is an implicit consequence of the results involving the linear functionals.

Furthermore, we show that if we choose the basis of pure diagrams in a certain way, all the coefficients in the expansion of an actual Betti diagram into a chain of pure diagrams are nonnegative integers.

Remark 1.1. Before the submission of this paper, Eisenbud and Schreyer [2010] have gone further on the side of cohomology tables and extended their results from vector bundles to coherent sheaves. Apart from the addition of two remarks, this manuscript is in its original form and is not dependent on these newer developments.

2. The partially ordered set of pure Betti diagrams

Let $R = k[x_1, x_2, \ldots, x_n]$ be the polynomial ring with the standard grading. For any finitely generated graded module M, we have a minimal free resolution

$$0 \longleftarrow M \longleftarrow F_0 \longleftarrow \cdots \longleftarrow F_{n-1} \longleftarrow F_n \longleftarrow 0,$$

where

$$F_i = \bigoplus_{j\in\mathbb{Z}} R(-j)^{\beta_{i,j}}, \quad i = 0, 1, \ldots, n,$$

and we get that the Hilbert series of M can be recovered from the Betti numbers, $\beta_{i,j}$, by

$$H(M, t) = \frac{1}{(1-t)^n} \sum_{i=0}^{n} \sum_{j\in\mathbb{Z}} (-1)^i \beta_{i,j} t^j.$$

It was noted by Herzog and M. Kühl [Herzog and Kühl 1984] that this gives us s linearly independent equations

$$\sum_{i=0}^{n} \sum_{j\in\mathbb{Z}} (-1)^i \beta_{i,j} j^m = 0, \quad m = 0, 1, \ldots, s-1,$$

where s is the codimension of M.

Furthermore, they proved that in the case when M is Cohen–Macaulay and the resolution is *pure*, that is, if $F_i = R(-d_i)$ for some integers $d_0, d_1, \ldots, d_s$, we get a unique solution (up to scalar multiples) to these equations given by

$$\beta_{i,j} = \begin{cases} (-1)^i \displaystyle\prod_{\substack{j=0 \\ j\neq i}}^{s} \frac{1}{d_j - d_i}, & j = d_i, \\ 0, & j \neq d_i, \end{cases}$$

for $i = 0, 1, \ldots, s$.

Definition 2.1. For an increasing sequence of integers $\boldsymbol{d} = (d_0, d_1, \ldots, d_s)$, where $0 \leq s \leq n$, we denote by $\pi(\boldsymbol{d})$ the matrix in $\mathbb{Q}^{n-1} \times \mathbb{Q}^{\mathbb{Z}}$ given by

$$\pi(\boldsymbol{d})_{i,j} = (-1)^i \prod_{\substack{j=0 \\ j\neq i}}^{s} \frac{1}{d_j - d_i}, \quad \text{for } j = d_i$$

and zero elsewhere. We will call this the *pure diagram* given by the degree sequence $\boldsymbol{d} = (d_0, d_2, \ldots, d_s)$. We will use the notation $d_i(\pi)$ to denote the degree, d_i, when $\pi = \pi(\mathbf{d}) = \pi(d_0, d_1, \ldots, d_s)$.

For degree sequences with $d_0 = 0$, we will also use the *normalized pure diagram*

$$\bar{\pi}(\boldsymbol{d}) = d_1 d_2 \cdots d_s \pi(\boldsymbol{d})$$

so that normalized pure diagrams have $\bar{\pi}_{0,0} = 1$.

We define a partial ordering on the set of pure diagrams, extending the ordering used for Cohen–Macaulay diagrams of a fixed codimension.

Definition 2.2. We say that $\pi(d_0, d_1, \ldots, d_s) \leq \pi(d'_0, d'_1, \ldots, d'_t)$ if the $s \geq t$ and $d_i \leq d'_i$ for $i = 0, 1, \ldots, t$.

As in the case of Cohen–Macaulay pure diagrams, we get a simplicial structure given by the maximal chains of pure diagrams, which spans simplicial cones. In order to have maximal chains in this setting, we have to fix a bound on the region we are considering. We can do this by restricting the degrees to be in a given region $M + i \leq d_i \leq N + i$. We will denote the subspace generated by Betti diagrams with these restrictions by $B_{M,N} = \mathbb{Q}^{n+1} \times \mathbb{Q}^{N-M+1}$. Furthermore, we denote by $B^s_{M,N}$ the subspace of $B_{M,N}$ of diagrams satisfying the s first Herzog–Kühl equations.

Proposition 2.3. *For $s = 0, 1, \ldots, n$, we have that any maximal chain of pure diagrams of codimension at least s in $B_{M,N}$ form a basis for $B^s_{M,N}$.*

Proof. For any interval of length one $\pi < \pi'$ there is a unique nonzero entry in π which is zero in all pure diagrams above π'. Thus the pure diagrams in any maximal chain are linearly independent. The number of elements in any maximal chain of pure diagrams of codimension at least s is $(n+1)(N-M)+n-s+1$, since we have $n+1$ positions that has to be raised $N-M$ steps and then $n-s+1$ times when the codimension is lowered by one. On the other hand, we have that the dimension of $B^s_{M,N}$ is $(n+1)(N-M+1)-s$, since we have s independent equations on $B_{M,N}$. □

By looking at the order in which the different positions of a Betti diagram in $B_{M,N}$ disappear when going along a maximal chain of pure diagrams we get the following observation:

Proposition 2.4. *The maximal chains of pure diagrams in $B_{M,N}$ are in one to one correspondence with numberings of the entries of an $(N-M+1) \times (n+1)$-matrix which are increasing to the left and downwards.*

Example 2.5. For $n = 2$, $M = 0$, and $N = 1$, the numbering

5	3	1
6	4	2

corresponds to the maximal chain

$$\begin{array}{|ccc|}\hline * & * & * \\ - & - & - \\ \hline\end{array} < \begin{array}{|ccc|}\hline * & * & - \\ - & - & * \\ \hline\end{array} < \begin{array}{|ccc|}\hline * & * & - \\ - & - & - \\ \hline\end{array} < \begin{array}{|ccc|}\hline * & - & - \\ - & * & - \\ \hline\end{array} < \begin{array}{|ccc|}\hline * & - & - \\ - & - & - \\ \hline\end{array} < \begin{array}{|ccc|}\hline - & - & - \\ * & - & - \\ \hline\end{array}$$

and there are four other maximal chains corresponding to

$$\begin{array}{|ccc|}\hline 3 & 2 & 1 \\ 6 & 5 & 4 \\ \hline\end{array}, \quad \begin{array}{|ccc|}\hline 4 & 2 & 1 \\ 6 & 5 & 3 \\ \hline\end{array}, \quad \begin{array}{|ccc|}\hline 5 & 2 & 1 \\ 6 & 4 & 3 \\ \hline\end{array}, \quad \text{and} \quad \begin{array}{|ccc|}\hline 4 & 3 & 1 \\ 6 & 5 & 2 \\ \hline\end{array}.$$

3. Description of the boundary facets

We know that the partially ordered set of pure diagrams in $B_{M,N}$ give rise to a simplicial fan Δ, where the faces are the totally ordered subsets. The facets of this simplicial fan correspond to maximal chains in the partially ordered set. According to Proposition 2.3 each such set is a basis for the space $B_{M,N}$. If we look at normalized Betti diagrams, we get a simplicial complex which is the hyperplane section of the simplicial fan. This is the complex described in our previous paper, in the case of Cohen–Macaulay diagrams (see [Boij and Söderberg 2008, Proposition 2.9]).

Proposition 3.1. *The simplicial cones spanned by the totally ordered sets of pure diagrams in $B_{M,N}$ form a simplicial fan.*

Proof. We need to show that the cones meet only along faces. This is the same as to say that any element which can be written as a positive linear combination of pure diagrams in a chain can be written so in a unique way. If such a sum has only one term, the uniqueness is trivial. Thus, suppose that a diagram in $B_{M,N}$ can be written as a positive linear combination of totally ordered pure diagrams in two different ways and that this is the minimal number of terms in such an example. We then have

$$\beta = \sum_{i=1}^{m} \lambda_i \pi_i = \sum_{j=1}^{k} \mu_j \pi'_j,$$

where all the coefficients are positive. Look at the lowest degree in which β is nonzero for each column. These degrees have to be given by the degrees in π_1 and by the degrees in π'_1. Thus we must have $\pi_1 = \pi'_1$. If $\lambda_1 > \mu_1$, we can subtract $\mu_1\pi_1$ from β and from both sums to get

$$\beta - \lambda_1\pi_1 = \sum_{i=2}^{m} \lambda_i \pi_i = (\mu_1 - \lambda_1)\pi_1 + \sum_{j=2}^{k} \mu_j \pi'_j.$$

Since all the pure diagrams in the left expression are greater than π_1, the degrees in which $\beta - \lambda_1\pi_1$ is nonzero have to be given by π_2, but then the coefficient of π_1 in the right-hand expression has to be zero, that is, $\lambda_1 = \mu_1$. Since this was assumed to be a minimal example, we get that the expressions for $\beta - \lambda_1\pi_1$ are term wise equal and so were the original expressions for the diagram β. □

We will now show that the coefficients of the pure diagrams when a Betti diagram is expanded in the basis given by a maximal chain have nice expressions. In particular, we see that the coefficients are integers. Moreover, it will give us expressions for the inequalities that define the simplicial fan similar to the

inequalities used by Eisenbud and Schreyer to prove our conjectures in the Cohen–Macaulay case.

Proposition 3.2. *The coefficient of $\pi_1 = \pi(d_0, d_1, \dots, d_m)$ when a Betti diagram β in $B^s_{M,N}$ is expanded in a basis containing $\pi_0 < \pi_1 < \pi_2$ is given by the following formulas*:

(a) *When π_1 differs from π_0 in codimension and from π_2 in column k, it is given by*

$$\sum_{i=0}^{n} \sum_{d=M}^{d_i(\pi_0)} (-1)^i \prod_{\substack{j=0 \\ j\neq k}}^{m} (d_j - d)\beta_{i,d}.$$

(b) *When π_1 differs from π_0 in codimension and from π_2 in column k, it is given by*

$$\sum_{i=0}^{n} \sum_{d=M}^{d_i(\pi_0)} (-1)^i (d_k - d_m) \prod_{\substack{j=0 \\ j\neq k}}^{m-1} (d_j - d)\beta_{i,d}.$$

(c) *When π_1 differs from π_2 in column k and from π_0 in column $\ell \neq k$, it is given by*

$$\sum_{i=0}^{n} \sum_{d=M}^{d_i(\pi_0)} (-1)^i (d_\ell - d_k) \prod_{\substack{j=0 \\ j\notin\{k,\ell\}}}^{m} (d_j - d)\beta_{i,d}.$$

(d) *When π_1 differs from π_0 and π_2 in codimension, it is given by*

$$\sum_{i=0}^{n} \sum_{d=M}^{d_i(\pi_0)} (-1)^i \prod_{j=0}^{m-1} (d_j - d)\beta_{i,d}.$$

In particular, all the coefficients are integers.

Remark 3.3. For simplicity of notation, we use the convention that $d_i(\pi) = N$ if i is greater than the codimension of π.

Observe that when π_1 differs from π_0 and π_2 in column k, the coefficient is given by the individual Betti number $(-1)^k \prod_{j\neq k}(d_j - d_k)\beta_{k,d_k}$, since no other diagram than π_1 contributes in this position, (k, d_k).

Proof. In all four cases, (a)–(d), the coefficients of $\beta_{i,d}$ are zero for pure diagrams $\pi \geq \pi_2$ by construction since the sums are taken only up to degree $d_i(\pi_0)$ and the coefficients are zero in all positions where $d_i(\pi_0) = d_i(\pi_2)$.

Furthermore, the expressions are zero on all pure diagrams π, where $\pi \leq \pi_0$ by the Herzog–Kühl equations. Indeed, for such pure diagrams, we can extend the summation over d to be over all degrees from M to N. When we do this and

expand the products we get polynomials in d and, therefore, the expressions are linear combinations of the Herzog–Kühl equations

$$\sum_{i=0}^{n}\sum_{d=M}^{N}(-1)^i\beta_{i,d}d^j=0,$$

for $j=0,1,\ldots,m-1$ in cases (b) and (d), and for $j=0,1,\ldots,m$ in cases (a) and (d). Observe that all pure diagrams π with $\pi\leq\pi_0$, have codimension at least the codimension of π_0 and hence satisfy these Herzog–Kühl equations.

It remains to show that the value of the expressions are one on the pure diagram π_1. In case (a), the only nonzero term is in column k and equals

$$(-1)^k\prod_{\substack{j=0\\ j\neq k}}^{m}(d_j-d_k)\cdot(-1)^k\prod_{\substack{j=0\\ j\neq k}}^{m}\frac{1}{d_j-d_k}=1,$$

in case (b), the only nonzero term is in column m and equals

$$(-1)^m(d_k-d_m)\prod_{\substack{j=0\\ j\neq k}}^{m}(d_j-d_m)\cdot(-1)^m\prod_{j=0}^{m-1}\frac{1}{d_j-d_m}=1,$$

in case (c), the only nonzero term is in column k and equals

$$(-1)^k(d_\ell-d_k)\prod_{\substack{j=0\\ j\notin\{k,\ell\}}}^{m}(d_j-d_k)\cdot(-1)^k\prod_{\substack{j=0\\ j\neq \ell}}^{m}\frac{1}{d_j-d_k}=1,$$

and in case (d), the nonzero term is in column m and equals

$$(-1)^m\prod_{j=0}^{m-1}(d_j-d_m)\cdot(-1)^m\prod_{j=0}^{m-1}\frac{1}{d_j-d_m}=1.\qquad\square$$

Example 3.4. When $n=3$, $M=0$, and $N=2$ we can choose the basis of $B_{0,2}$ given by the numbering

10	4	3	1
11	6	5	2
12	9	8	7

.

When expanding a Betti diagram

	β_0	β_1	β_2	β_3
0 :	$\beta_{0,0}$	$\beta_{1,1}$	$\beta_{2,2}$	$\beta_{3,3}$
1 :	$\beta_{0,1}$	$\beta_{1,2}$	$\beta_{2,3}$	$\beta_{3,4}$
2 :	$\beta_{0,2}$	$\beta_{1,3}$	$\beta_{2,4}$	$\beta_{3,5}$

(where we use the Macaulay2 [Grayson and Stillman 1993] convention shifting the columns upwards in order for the diagrams to be more compact) into this chain we get the coefficients by applying the linear functionals corresponding to the following matrices:

$$\begin{bmatrix} 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
1

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 24 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
2

$$\begin{bmatrix} 0 & 0 & 6 & -18 \\ 0 & 0 & 0 & -36 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
3

$$\begin{bmatrix} 0 & 8 & -12 & 12 \\ 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
4

$$\begin{bmatrix} 0 & -4 & 6 & -6 \\ 0 & 0 & 6 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
5

$$\begin{bmatrix} 0 & 8 & -12 & 12 \\ 0 & 12 & -12 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
6

$$\begin{bmatrix} 0 & -6 & 8 & -6 \\ 0 & -8 & 6 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$
7

$$\begin{bmatrix} 0 & 2 & -2 & 0 \\ 0 & 2 & 0 & -4 \\ 0 & 0 & 4 & -10 \end{bmatrix}$$
8

$$\begin{bmatrix} 0 & 1 & -2 & 3 \\ 0 & 2 & -3 & 4 \\ 0 & 3 & -4 & 5 \end{bmatrix}$$
9

$$\begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & -1 & 1 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}$$
10

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
11

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$
12

Matrices 1, 2, 11, and 12 correspond to individual Betti numbers as we will see in the next proposition. Matrices 3, 4, 5, and 6 are as in Proposition 3.2 case (c), since they correspond to two consecutive degree changes in distinct columns. Matrix 7 is as in case (b), since it corresponds to a degree change followed by a change in codimension. Matrices 8 and 9 are as in case (d), since they correspond to two consecutive changes of codimension. Matrix 10 is as in case (a) since it corresponds to a change in codimension followed by a degree change.

In order to prove that the simplicial fan is convex and in order to prove that any Betti diagram of a module of codimension s is a positive linear combination of pure diagrams, we need to know the boundary of the simplicial fan. The description is very similar to the description given in our previous paper [Boij and Söderberg 2008, Proposition 2.12] and we only have to add one more kind of boundary facet.

Proposition 3.5. *A facet of the boundary of the simplicial fan given by the pure diagrams in $B^s_{M,N}$ is given by removing one element from a maximal chain such that there is a unique way to complete it to a maximal chain. There are four different cases*:

(i) *The removed element is maximal or minimal.*

(ii) *The removed element is in the middle of a chain of three degree sequences which differ in one single column, that is,*

$$\pi(\boldsymbol{d}) < \pi(\boldsymbol{d}') < \pi(\boldsymbol{d}''), \qquad d_i = d'_i - 1 = d''_i - 2.$$

(iii) *The removed element differs from the adjacent vertices in two adjacent degrees, that is,*

$$\pi(\boldsymbol{d}) < \pi(\boldsymbol{d}') < \pi(\boldsymbol{d}''), \qquad d_{i+1} = d'_{i+1} - 1, \quad d'_i = d''_i - 1.$$

(iv) *the removed element differs from the adjacent vertices in codimension, that is,*

$$\pi(\boldsymbol{d}) < \pi(\boldsymbol{d}') < \pi(\boldsymbol{d}''), \qquad \operatorname{codim}(\pi(\boldsymbol{d})) = \operatorname{codim}(\pi(\boldsymbol{d}'')) + 2.$$

Proof. A boundary facet is a codimension-one face of a simplicial cone and hence given by the removing of one vertex. Since it is on the boundary, it is contained in a unique cone of maximal dimension, so there has to be a unique extension of the chain into a maximal chain.

This clearly happens if we remove the maximal or the minimal element, since there are unique such elements in the partially ordered set. Suppose therefore that we remove $\pi(\boldsymbol{d}')$ and that the adjacent vertices in the chain are $\pi(\boldsymbol{d}) < \pi(\boldsymbol{d}')$ and $\pi(\boldsymbol{d}'') > \pi(\boldsymbol{d}')$. If the difference between $\pi(\boldsymbol{d})$ and $\pi(\boldsymbol{d}'')$ are in two columns that are not adjacent, we can find another element between $\pi(\mathbf{d})$ and $\pi(\boldsymbol{d}'')$ by exchanging the order in which the two degrees are increased. This can also be done if the columns are adjacent, but the degrees differ by more than one.

If one of the two differ from $\pi(\boldsymbol{d}')$ in codimension and the other by an increase of the degree in one column, we can alter the order and obtain another element between $\pi(\boldsymbol{d})$ and $\pi(\boldsymbol{d}'')$.

The remaining cases are those described by (ii), (iii), and (iv). □

Remark 3.6. In Example 3.4 matrices 1 and 12 correspond to boundary facets of type (i). Matrices 2 and 11 correspond to boundary facets of type (ii). Matrices 4 and 6 correspond to boundary facets of type (iii) and matrices 8 and 9 correspond

to boundary facets of type (iv). The remaining matrices correspond to inner faces of the fan.

Theorem 3.7. *The simplicial fan of pure diagrams in $B^s_{M,N}$ is convex.*

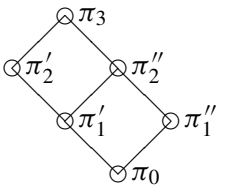
Proof. We will use the following observation which allows us to go from local convexity to global convexity: If the simplicial fan is not convex, then there will be one boundary facet whose supporting hyperplane passes through the interior of a neighboring simplicial cone. We can see this by looking at a line segment between two vertices which contains points outside the simplicial fan. If we take a two-dimensional plane through these two points and through a generic inner point of the fan, we get a two-dimensional picture where we can find two edges meeting at an inwards angle. The supporting hyperplane of the boundary facet meeting the two-dimensional plane in one of these edges meets the interior of the simplicial cone corresponding to the other edge.

Thus we will prove that any two boundary facets meet in a convex manner. If either of the two facets are of type (i) or (ii), described in Proposition 3.5, it is clear that any Betti diagram will lie on the correct side of the supporting hyperplane, since this hyperplane is given by the vanishing of a single Betti number. Thus we will assume that the facets are of types (iii) or (iv).

Let K denote the number of pure diagrams in a maximal chain in $B^s_{M,N}$. Any two boundary facets meeting along a codimension-one face gives us a chain with $K-2$ pure diagrams. If the two missing vertices are on levels differing by more than one, there is a unique way of completing the chain into a maximal chain and the two facets are faces of the same simplicial cone. Hence they meet in a convex way.

Thus we can assume that the two vertices missing are on adjacent levels and that there is a least element π_3 above these and a greatest element π_0 below them. Now we can see from the difference in codimension between π_0 and π_3 that facets of type (iii) described in Proposition 3.5 cannot meet facets of type (iv) in this way.

We must have that the two facets are given by removing π'_1 or π''_2 from the chains $\pi_0 < \pi'_1 < \pi'_2 < \pi_3$ and $\pi_0 < \pi''_1 < \pi''_2 < \pi_3$ in the following lattice:



We need to show that the coefficient of π'_1 is positive when π''_1 is expanded into the chain $\pi_0 < \pi'_1 < \pi'_2 < \pi_3$. Note that the codimension cannot differ by more than three between π_0 and π_3 and if it differs by three, there is only one chain between π_0 and π_3. Thus we can assume that the difference in codimension is zero, one, or two.

In order to do this we use Proposition 3.2 which gives us an expression of this coefficient involving only two terms.

Suppose that π_0 and π_3 differ in codimension by two. Then the coefficient is given by the fourth expression of Proposition 3.2 and the coefficient for $\pi_1' = \pi(d_0, d_1, \ldots, d_m)$ when $\pi_1'' = \pi(d_0, d_1, \ldots, d_{k-1}, d_k+1, d_{k+1}, \ldots, d_m)$ is expanded in the basis containing $\pi_0 < \pi_1' < \pi_2'$ is given by

$$\sum_{i=0}^{n} \sum_{d \le d_i(\pi_0)} (-1)^i \prod_{j=0}^{m-1} (d_j - d) \beta_{i,d}(\pi_1''),$$

which has only two nonzero terms in columns m and $m+1$ and equals

$$\begin{aligned}
(-1)^m \prod_{j=0}^{m-1} (d_j - d_m)(-1)^m \frac{d_k - d_m}{d_k + 1 - d_m} \frac{1}{d_{m+1} - d_m} \prod_{j=0}^{m-1} \frac{1}{d_j - d_m} \\
+ (-1)^{m+1} \prod_{j=0}^{m-1} (d_j - d_{m+1})(-1)^{m+1} \frac{d_k - d_{m+1}}{d_k + 1 - d_{m+1}} \frac{1}{d_m - d_{m+1}} \prod_{j=0}^{m-1} \frac{1}{d_j - d_{m+1}} \\
= \frac{1}{d_{m+1} - d_m} \left(\frac{d_k - d_m}{d_k + 1 - d_m} - \frac{d_k - d_{m+1}}{d_k + 1 - d_{m+1}} \right) \\
= \frac{1}{(d_k + 1 - d_{m+1})(d_k + 1 - d_m)} > 0,
\end{aligned}$$

since $d_k + 1 < d_m < d_{m+1}$.

If π_0 and π_3 differ in codimension by one we get that the coefficient of π_1' when $\pi_1'' = \pi(d_0, d_1, \ldots, d_{m-1})$ is expanded in the basis $\pi_0 < \pi_1' < \pi_2' < \pi_3$ is given by

$$\sum_{i=0}^{n} \sum_{d \le d_i(\pi_0)} (-1)^i (d_k + 2 - d_k) \prod_{\substack{j=0 \\ j \notin \{k, k+1\}}}^{m} (d_j - d) \beta_{i,d}(\pi_1''),$$

which has only two nonzero terms from columns k and $k+1$ and equals

$$\begin{aligned}
2(-1)^k \prod_{\substack{j=0 \\ j \notin \{k, k+1\}}}^{m} (d_j - d_k)(-1)^k \prod_{\substack{j=0 \\ j \ne k}}^{m-1} \frac{1}{d_j - d_k} \\
+ 2(-1)^{k+1} \prod_{\substack{j=0 \\ j \notin \{k, k+1\}}}^{m} (d_j - d_k - 1)(-1)^{k+1} \prod_{\substack{j=0 \\ j \ne k+1}}^{m-1} \frac{1}{d_j - d_k - 1} \\
= 2 \frac{d_m - d_k}{d_k + 1 - d_k} - 2 \frac{d_m - d_k - 1}{d_k - d_k - 1} = 2 > 0.
\end{aligned}$$

The last possibility is that the codimension of π_0 equals the codimension of π_3. In this case $\pi_0 < \pi_1' < \pi_2'$ corresponds to a facet of type (iii) and we get that the coefficient of $\pi_1' = \pi(d_0, d_1, \ldots, d_k, d_k+2, d_{k+2}, \ldots, d_m)$, when $\pi_1'' = \pi(d_0, d_1, \ldots, d_{\ell-1}, d_\ell + 1, d_{\ell+1}, \ldots, d_m)$ is expanded in the basis $\pi_0 < \pi_1' < \pi_2' < \pi_3$, is given by

$$\sum_{i=0}^{n} \sum_{d \le d_i(\pi_0)} (-1)^i (d_k + 2 - d_k) \prod_{\substack{j=0 \\ j \notin \{k,k+1\}}}^{m} (d_j - d)\beta_{i,d}(\pi_1'').$$

Again there are only two nonzero terms and the coefficient is equal to

$$2(-1)^k \prod_{\substack{j=0 \\ j \notin \{k,k+1\}}}^{m} (d_j - d_k)(-1)^k \frac{d_\ell - d_k}{d_\ell + 1 - d_k} \prod_{\substack{j=0 \\ j \neq k}}^{m} \frac{1}{d_j - d_k}$$

$$+ 2(-1)^{k+1} \prod_{\substack{j=0 \\ j \notin \{k,k+1\}}}^{m} (d_j - d_k - 1)(-1)^{k+1} \frac{d_\ell - d_k - 1}{d_\ell + 1 - d_k - 1} \prod_{\substack{j=0 \\ j \neq k+1}}^{m} \frac{1}{d_j - d_k - 1}$$

$$= 2\frac{d_\ell - d_k}{d_\ell + 1 - d_k} - 2\frac{d_\ell - d_k - 1}{d_\ell - d_k} = \frac{2}{(d_\ell - d_k)(d_\ell - d_k + 1)} > 0,$$

since $d_\ell - d_k$ and $d_\ell - d_k + 1$ are both negative or both positive. □

4. The expansion of any Betti diagram into sums of pure diagrams

We know from the work of Eisenbud and Schreyer that the inequalities given by the exterior facets of the cone are valid on all minimal free resolutions, not only on the resolutions of Cohen–Macaulay modules. The inequalities that we have to add because we now look at chains of pure diagrams of different codimensions can be seen to be limits of the inequalities already known.

As in the previous section, we look at Betti diagrams of graded modules under the restriction that $\beta_{i,j} = 0$ unless $M + i \le j \le N + i$, that is, diagrams in $B_{M,N}$.

Theorem 4.1. *Any Betti diagram of a finitely generated graded module P of codimension s can be uniquely written as a positive linear combination of a chain of pure diagrams in $B^s_{M,N}$, where M is the least degree of any generator of P and N is the regularity of P.*

Proof. By Proposition 3.5 we know what are the boundary facets of the simplicial fan given by the pure diagrams in $B^s_{M,N}$ and by Proposition 3.2 we know how to obtain the inequalities given by the boundary facets. In fact, the coefficient of the removed element of the chain is nonnegative on the half-space defined by the corresponding boundary facet. We need to show that any Betti diagram of a graded module satisfies these inequalities.

By Proposition 3.5 we have four types of boundary facets to consider, (i)–(iv). The inequalities corresponding to facets of type (i) and (ii) certainly hold for all Betti diagrams of modules.

The inequalities corresponding to facets of type (iii) are given by the expression of Proposition 3.2(c) and can be written as

$$\sum_{i=0}^{n}\sum_{d=M}^{d_i(\pi_0)}(-1)^i\prod_{\substack{j=0\\ j\notin\{k,k+1\}}}^{m}(d_j-d)\beta_{i,d}(P)\geq 0 \tag{4-1}$$

and the inequalities corresponding to facets of type (iv) are given by the expression of Proposition 3.2(d) and can be written as

$$\sum_{i=0}^{n}\sum_{d=M}^{d_i(\pi_0)}(-1)^i\prod_{j=0}^{m-1}(d_j-d)\beta_{i,d}(P)\geq 0. \tag{4-2}$$

As we can see from these expressions, they are very similar. Indeed, (4-2) is equal to (4-1) if we increase N and m by one and choose $k=m-1$. Thus it is sufficient to prove that (4-1) always holds for any $m\leq n$.

The case $m=n$ corresponds to the facets that occur in the cone of Cohen–Macaulay diagrams. In the Cohen–Macaulay case we have more Herzog–Kühl equations but fewer inequalities. Eisenbud and Schreyer [2009, Theorems 4.1 and 7.1] proved that the inequalities given by the boundary facets hold for Betti diagrams of all modules, not only Cohen–Macaulay modules. Hence (4-1) holds for Betti diagrams in $B_{M,N}$ when $m=n$.

In order to prove that (4-1) holds when $m<n$, we look at the inequality we get by exchanging $\pi_1=\pi(d_0,d_1,\dots,d_m)$ by $\pi_1^t=\pi(d_0,d_1,\dots,d_m,d_m+1+t,d_m+2+t,\dots,d_m+n-m+t)$, and similarly exchanging π_0 and π_2 by π_0^t and π_2^t. We then let t grow to infinity and get

$$\lim_{t\to\infty}\frac{1}{t^{n-m}}\sum_{i=0}^{n}\sum_{d=M}^{d_i(\pi_0^t)}(-1)^i\prod_{\substack{j=0\\ j\notin\{k,k+1\}}}^{m}(d_j-d)\prod_{j=m+1}^{n}(d_m+j-m+t-d)\beta_{i,d}$$
$$=\sum_{i=0}^{n}\sum_{d=M}^{d_i(\pi_0)}(-1)^i\prod_{\substack{j=0\\ j\notin\{k,k+1\}}}^{m}(d_j-d)\beta_{i,d}(P).$$

The limit is nonnegative since for each integer $t\geq 0$, the expression under the limit in the left-hand side is nonnegative. □

Remark 4.2. One of the questions raised by Eisenbud and Schreyer was what was the description of the convex cone cut out by all of their inequalities. The answer

to this question is that this convex cone in $B_{M,N}$ equals the convex cone of all Betti diagrams in $B_{M,N}$ of graded modules up to multiplication by nonnegative rational numbers. This follows from Theorem 4.1 and the fact that all the inequalities we use to define the simplicial fan, as seen in the proof, are limits of their inequalities, hence are consequences of the full set of those inequalities. Thus for a given diagram satisfying their inequalities, we find suitable M and N and use Theorem 4.1 to see that it can be written as a positive linear combination of pure diagrams and we can conclude that a multiple of the diagram is the Betti diagram of a module.

We can also see that the unique way of writing any Betti diagram of a module into a chain of pure diagrams leads to a way of writing the diagram as a linear combination of diagrams of Cohen–Macaulay modules, one of each codimension between the codimension of M and the projective dimension of M. As in the Cohen–Macaulay case, we get an algorithm for finding the expansion of a given Betti diagram by subtracting as much as possible of the pure diagram corresponding to the lowest shifts.

Remark 4.3. The method used in the proof of Theorem 4.1 is based only on what was known after the appearance of [Eisenbud and Schreyer 2009]. In view of the their more recent paper [Eisenbud and Schreyer 2010], the limiting process can be avoided using the functionals associated to coherent sheaves and, in fact, we only need to use vector bundles with supernatural cohomology supported on linear subspaces of $\mathbb{P}^{n-1}$, as pointed out by an anonymous referee.

Example 4.4. For $M = k[x, y, z]/(x^2, xy, xz^2)$ we get the Betti diagram

	1	3	3	1
0:	1	–	–	–
1:	–	2	1	–
2:	–	1	2	1

which can be expanded into

$$\begin{bmatrix} 1 & - & - & - \\ - & 2 & 1 & - \\ - & 1 & 2 & 1 \end{bmatrix} = 6 \cdot \begin{bmatrix} \frac{1}{30} & - & - & - \\ - & \frac{1}{6} & \frac{1}{6} & - \\ - & - & - & \frac{1}{30} \end{bmatrix} + 12 \cdot \begin{bmatrix} \frac{1}{40} & - & - & - \\ - & \frac{1}{12} & - & - \\ - & - & \frac{1}{8} & \frac{1}{15} \end{bmatrix}$$
$$+ 2 \cdot \begin{bmatrix} \frac{1}{12} & - & - & - \\ - & - & - & - \\ - & \frac{1}{3} & \frac{1}{4} & - \end{bmatrix} + 1 \cdot \begin{bmatrix} \frac{1}{3} & - & - & - \\ - & - & - & - \\ - & \frac{1}{3} & - & - \end{bmatrix}.$$

The coefficients can also be obtained by the functionals corresponding to the matrices 5, 6, 8, and 9 from Example 3.4.

We now will go on and prove a generalized version of the multiplicity conjecture in terms of the Hilbert series. In order to do this, we first prove that the Hilbert series behaves well with respect to the partial ordering on the normalized pure diagrams. We use the partial order on Hilbert series given by coefficientwise inequality in all degrees simultaneously.

Proposition 4.5. *The Hilbert series is strictly increasing on the partially ordered set of normalized pure diagrams generated in degree zero.*

Proof. First assume that $\bar{\pi} < \bar{\pi}'$ is a maximal chain of pure diagrams of codimension s. The Hilbert series can be recovered from the Betti diagrams and we can write

$$H(\bar{\pi}', t) - H(\bar{\pi}, t) = \frac{1}{(1-t)^n}(S(\bar{\pi}', t) - S(\bar{\pi}, t)),$$

where $S(\beta, t) = \sum_{i=0}^{n}(-1)^i \sum_j \beta_{i,j} t^j$. Since the polynomials $S(\bar{\pi}, t)$ and $S(\bar{\pi}', t)$ both have constant term 1, we get that the polynomial $S(\bar{\pi}', t) - S(\bar{\pi}, t)$, has only $s+1$ nonzero terms. Since it also satisfies the Herzog–Kühl equations, we have a unique solution to this up to a scalar multiple, and $H(\bar{\pi}, t) - H(\bar{\pi}', t)$ is λt^{d_1} times the Hilbert series of a pure module. Since we know that all pure diagrams have positive Hilbert series, we only have to check that this is a nonnegative multiple. The sign of λ can be obtained by looking at the sign of the term of $S(\bar{\pi}', t) - S(\bar{\pi}, t)$ which comes from $S(\bar{\pi}', t)$ and which is not present in $S(\bar{\pi}, t)$. This term comes with the sign $(-1)^i$, if it is in column i, which proves that λ has to be positive.

We now consider the case where $\bar{\pi} < \bar{\pi}'$ is a maximal chain such that the codimension of $\bar{\pi}$ is s and the codimension of $\bar{\pi}'$ is $s-1$. If we now look at the difference of the Hilbert series,

$$H(\bar{\pi}', t) - H(\bar{\pi}, t) = \frac{1}{(1-t)^n}(S(\bar{\pi}', t) - S(\bar{\pi}, t)),$$

where the polynomial $S(\bar{\pi}', t) - S(\bar{\pi}, t)$ has zero constant term and s nonzero terms. Since we know that it is divisible by $(1-t)^{s-1}$, there is again a unique possibility, which is λt^{d_1} times the Hilbert series of a module with pure resolution. This time, we can see that the sign of the last term is $-(-1)^s = (-1)^{s-1}$, which shows that λ is positive. □

Theorem 4.6. *For any finitely generated module M of projective dimension p and codimension s generated in degree* 0, *we have that*

$$H(\bar{\pi}(0, \underline{d}_1, \underline{d}_2, \ldots, \underline{d}_p), t) \leq \frac{H(M, t)}{\beta_0(M)} \leq H(\bar{\pi}(0, \bar{d}_1, \bar{d}_2, \ldots, \bar{d}_s), t),$$

where $\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_p$ are the minimal shifts and $\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_s$ are the maximal shifts in a minimal free resolution of M. Equality on either side implies that the resolution is pure.

In particular, the right-hand inequality implies the multiplicity conjecture, that is,

$$\frac{e(M)}{\beta_0(M)} \leq \frac{\bar{d}_1\bar{d}_2\cdots\bar{d}_s}{s!},$$

with equality if and only if M is Cohen–Macaulay with a pure resolution.

Proof. We can use Theorem 4.1 to write the Betti diagram of M as a positive linear combination of pure diagrams. Since the Hilbert series by Proposition 4.5 is increasing on along the chain, we get the first inequalities of the theorem.

The multiplicity of M is obtained from the leading coefficient of the Hilbert polynomial. The coefficients of the Hilbert series are eventually equal to the Hilbert polynomial, which shows that the multiplicity is increasing with the Hilbert series as long as the degree of the Hilbert polynomial is the same. Since this is the case for M and $\pi(0, \bar{d}_1, \bar{d}_2, \ldots, \bar{d}_s)$, we get the conclusion of the multiplicity conjecture. □

Remark 4.7. For a Cohen–Macaulay module M, the Hilbert coefficients are positive linear combinations of the entries of the h-vector. Since we can reduce M modulo a regular sequence and keep the same Betti diagram, we get that the Hilbert coefficients are increasing along chains of normalized pure diagrams. Thus the generalization of the multiplicity conjecture made by J. Herzog and X. Zheng [Herzog and Zheng 2009] is a consequence of Theorem 4.6.

Acknowledgement

This paper is heavily dependent on the beautiful proofs of our conjectures made by David Eisenbud and Frank-Olaf Schreyer and we are grateful for their efforts in this area. We would like to thank the participants at the workshop at Cornell University in March 2008 and the referee for useful comments and suggestions.

References

[Boij and Söderberg 2008] M. Boij and J. Söderberg, "Graded Betti numbers of Cohen–Macaulay modules and the multiplicity conjecture", *J. Lond. Math. Soc.,(2)* **78**:1 (2008), 85–106. MR 2009g:13018 Zbl 1189.13008

[Eisenbud and Schreyer 2009] D. Eisenbud and F.-O. Schreyer, "Betti numbers of graded modules and cohomology of vector bundles", *J. Amer. Math. Soc.* **22**:3 (2009), 859–888. MR 2011a:13024 Zbl 1213.13032

[Eisenbud and Schreyer 2010] D. Eisenbud and F.-O. Schreyer, "Cohomology of coherent sheaves and series of supernatural bundles", *J. Eur. Math. Soc.* **12**:3 (2010), 703–722. MR 2011e:14036 Zbl 1213.14035

[Eisenbud et al. 2011] D. Eisenbud, G. Fløystad, and J. Weyman, "The existence of equivariant pure free resolutions", *Ann. Inst. Fourier (Grenoble)* **61**:3 (2011), 905–926. MR 2918721 Zbl 139.13023

[Francisco 2006] C. A. Francisco, "New approaches to bounding the multiplicity of an ideal", *J. Algebra* **299**:1 (2006), 309–328. MR 2007b:13043 Zbl 1105.13003

[Francisco and Srinivasan 2007] C. A. Francisco and H. Srinivasan, "Multiplicity conjectures", pp. 145–178 in *Syzygies and Hilbert functions*, edited by I. Peeva, Lect. Notes Pure Appl. Math. **254**, Chapman & Hall/CRC, Boca Raton, FL, 2007. MR 2008f:13044 Zbl 1125.13011

[Francisco and Van Tuyl 2007] C. A. Francisco and A. Van Tuyl, "Some families of componentwise linear monomial ideals", *Nagoya Math. J.* **187** (2007), 115–156. MR 2008g:13017 Zbl 1140.13012

[Goff 2008] M. Goff, "On the multiplicity conjecture for non-Cohen–Macaulay simplicial complexes", preprint, 2008. arXiv 0802.1282v1

[Gold 2003] L. H. Gold, "A degree bound for codimension two lattice ideals", *J. Pure Appl. Algebra* **182**:2-3 (2003), 201–207. MR 2004i:13023 Zbl 1032.13006

[Gold et al. 2007] L. Gold, H. Schenck, and H. Srinivasan, "Betti numbers and degree bounds for some linked zero-schemes", *J. Pure Appl. Algebra* **210**:2 (2007), 481–491. MR 2008b:13020 Zbl 1123.14029

[Grayson and Stillman 1993] D. Grayson and M. Stillman, "Macaulay 2, a software system for research in algebraic geometry", 1993, Available at http://www.math.uiuc.edu/Macaulay2/.

[Guardo and Van Tuyl 2005] E. Guardo and A. Van Tuyl, "Powers of complete intersections: graded Betti numbers and applications", *Illinois J. Math.* **49**:1 (2005), 265–279. MR 2006k:13035 Zbl 1089.13008

[Herzog and Kühl 1984] J. Herzog and M. Kühl, "On the Betti numbers of finite pure and linear resolutions", *Comm. Algebra* **12**:13-14 (1984), 1627–1646. MR 85e:13021 Zbl 0543.13008

[Herzog and Srinivasan 1998] J. Herzog and H. Srinivasan, "Bounds for multiplicities", *Trans. Amer. Math. Soc.* **350**:7 (1998), 2879–2902. MR 99g:13033 Zbl 0899.13026

[Herzog and Zheng 2006] J. Herzog and X. Zheng, "Notes on the multiplicity conjecture", *Collect. Math.* **57**:2 (2006), 211–226. MR 2007a:13029 Zbl 1116.13022

[Herzog and Zheng 2009] J. Herzog and X. Zheng, "Bounds for Hilbert coefficients", *Proc. Amer. Math. Soc.* **137**:2 (2009), 487–494. MR 2009f:13024 Zbl 1162.13010

[Hibi and Singla 2008] T. Hibi and P. Singla, "Linear balls and the multiplicity conjecture", *J. Algebra* **319**:10 (2008), 4372–4390. MR 2009e:13026 Zbl 1166.13026

[Huneke and Miller 1985] C. Huneke and M. Miller, "A note on the multiplicity of Cohen–Macaulay algebras with pure resolutions", *Canad. J. Math.* **37**:6 (1985), 1149–1162. MR 87d:13024 Zbl 0579.13012

[Kubitzke and Welker 2006] M. Kubitzke and V. Welker, "The multiplicity conjecture for barycentric subdivisions", preprint, 2006. arXiv math/0606274v1

[Migliore et al. 2005] J. Migliore, U. Nagel, and T. Römer, "The multiplicity conjecture in low codimensions", *Math. Res. Lett.* **12**:5 (2005), 731–747. MR 2006i:13042 Zbl 1090.13020

[Migliore et al. 2008a] J. Migliore, U. Nagel, and T. Römer, "Extensions of the multiplicity conjecture", *Trans. Amer. Math. Soc.* **360**:6 (2008), 2965–2985. MR 2009a:13043 Zbl 1148.13016

[Migliore et al. 2008b] J. C. Migliore, U. Nagel, and F. Zanello, "An improved multiplicity conjecture for codimension 3 Gorenstein algebras", *Comm. Algebra* **36**:1 (2008), 112–119. MR 2008k:13033 Zbl 1135.13012

[Miró-Roig 2006] R. M. Miró-Roig, "A note on the multiplicity of determinantal ideals", *J. Algebra* **299**:2 (2006), 714–724. MR 2007a:13030 Zbl 1116.13008

[Puthenpurakal 2008] T. J. Puthenpurakal, "On the upper bound of the multiplicity conjecture", *Proc. Amer. Math. Soc.* **136**:10 (2008), 3429–3434. MR 2009e:13029 Zbl 1163.13018

[Römer 2005] T. Römer, "Note on bounds for multiplicities", *J. Pure Appl. Algebra* **195**:1 (2005), 113–123. MR 2005g:13041 Zbl 1073.13015

[Seo and Srinivasan 2007] S. Seo and H. Srinivasan, "Multiplicity of codimension three almost complete intersections", preprint, 2007. arXiv arXiv:0712.0818v1

[Srinivasan 1998] H. Srinivasan, "A note on the multiplicities of Gorenstein algebras", *J. Algebra* **208**:2 (1998), 425–443. MR 99m:13046 Zbl 0920.13018

[Zanello 2007] F. Zanello, "Improving the bounds of the multiplicity conjecture: the codimension 3 level case", *J. Pure Appl. Algebra* **209**:1 (2007), 79–89. MR 2008a:13034 Zbl 1116.13023

Communicated by David Eisenbud
Received 2010-07-02 Revised 2011-01-24 Accepted 2011-05-23

boij@kth.se *Department of Mathematics, KTH, SE-100 44 Stockholm, Sweden*

jonassod@gmail.com *Department of Mathematics, KTH, SE-100 44 Stockholm, Sweden*

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

$\mathscr{L}$-invariants and Shimura curves

Samit Dasgupta and Matthew Greenberg

In earlier work, the second named author described how to extract Darmon-style $\mathscr{L}$-invariants from modular forms on Shimura curves that are special at p. In this paper, we show that these $\mathscr{L}$-invariants are preserved by the Jacquet–Langlands correspondence. As a consequence, we prove the second named author's period conjecture in the case where the base field is $\mathbb{Q}$. As a further application of our methods, we use integrals of Hida families to describe Stark–Heegner points in terms of a certain Abel–Jacobi map.

1. Introduction

Let N and p be relatively prime positive integers with p prime and let

$$f = \sum_{n=1}^{\infty} a_n(f) q^n \in S_2(\Gamma_0(Np))^{p\text{-new}}$$

be a Hecke eigenform with $a_1(f) = 1$. In their study of p-adic L-functions associated to modular forms, Mazur, Tate and Teitelbaum [Mazur et al. 1986] introduce a p-adic invariant of f which they call its *$\mathscr{L}$-invariant*. Let $\mathscr{X}(f, p)$ be the set of primitive Dirichlet characters with conductor prime to p such that

MSC2000: primary 11F41; secondary 11G18, 11F67, 11F75.
Keywords: L-invariants, Shimura curves, Hida families, Stark–Heegner points.

$\chi(p) = a_p(f) = \pm 1$. If $\chi \in \mathscr{X}(f, p)$ then the interpolation property forces the p-adic L-function $L_p(f, \chi, s)$ of f twisted by χ to vanish at $s = 0$. This is called an *exceptional zero* phenomenon. In this case, it is conjectured in [Mazur et al. 1986] that there is a p-adic number $\mathscr{L}^{\mathrm{MTT}}(f)$ such that for all $\chi \in \mathscr{X}(f, p)$ of conductor c,

$$L_p'(f, \chi, 0) = \mathscr{L}^{\mathrm{MTT}}(f) \frac{c}{\tau(\overline{\chi})} \frac{L(f, \overline{\chi}, 1)}{\Omega_f^{\chi(-1)}}. \tag{1-1}$$

Here, $\tau(\overline{\chi})$ is the Gauss sum associated to $\overline{\chi}$ and $\Omega_f^{\chi(-1)}$ is the real or imaginary period of f, depending on the parity of χ. Note that (1-1) makes sense after fixing embeddings $\overline{\mathbb{Q}} \subset \mathbb{C}$, $\overline{\mathbb{Q}} \subset \mathbb{C}_p$, since $L(f, \overline{\chi}, 1)/\Omega_f^{\chi(-1)}$ is algebraic by a theorem of Shimura. It follows from nonvanishing results on critical L-values that $L(f, \overline{\chi}, 1) \neq 0$ for some $\chi \in \mathscr{X}(f, p)$, making (1-1) a nontrivial statement; see [Darmon 2001, Lemma 2.17] and the following remark.

The existence of $\mathscr{L}^{\mathrm{MTT}}(f)$ was proved in the influential paper [Greenberg and Stevens 1993]. Since f is p-ordinary, that is, $a_p(f)$ is a p-adic unit, f lives in a p-adic analytic family $\boldsymbol{f}$ of eigenforms by the work of Hida [1986]. More precisely, there is a p-adic disk $U \subset \mathbb{Z}_p \times \mathbb{Z}/(p-1)\mathbb{Z}$ containing 2 and a p-adic analytic function $\boldsymbol{a}_n(\boldsymbol{f}) : U \to \mathbb{C}_p$ for each $n \geq 1$, with $\boldsymbol{a}_1(\boldsymbol{f}) = 1$, such that

(1) for all integers $k \geq 2$ with $k \in U$, $\boldsymbol{a}_n(\boldsymbol{f}, k) \in \overline{\mathbb{Q}}$ and the image of

$$\boldsymbol{f}(k) := \sum_{n=1}^{\infty} \boldsymbol{a}_n(\boldsymbol{f}, k) q^n$$

in $\mathbb{C}[\![q]\!]$ is the q-expansion of an eigenform in $S_k(\Gamma_0(Np))$,

(2) $\boldsymbol{f}(2) = f$.

Moreover, up to shrinking U around 2, $\boldsymbol{f}$ is completely determined by f. Note that $1 - \boldsymbol{a}_p(\boldsymbol{f}, k)^2$ vanishes at $k = 2$ since $a_p(f) = \pm 1$. Thus, it is natural to consider the derivative of this quantity. Greenberg and Stevens show that (1-1) holds with

$$\mathscr{L}^{\mathrm{MTT}}(f) = \frac{d}{dk}\big(1 - \boldsymbol{a}_p(\boldsymbol{f}, k)^2\big)\Big|_{k=2} =: \mathscr{L}^{\mathrm{GS}}(f). \tag{1-2}$$

Also, (1-2) extends the definition of the $\mathscr{L}$-invariant from the case $a_p(f) = 1$ originally considered in [Mazur et al. 1986] to the case $a_p(f) = \pm 1$.

Mazur, Tate, and Teitelbaum further conjecture in the same work that $\mathscr{L}^{\mathrm{MTT}}(f)$ is of local type, that is, depends only on the two-dimensional p-adic representation $\sigma_p(f)$ of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ associated to f. Greenberg and Stevens [1993] proved this by showing that $\mathscr{L}^{\mathrm{GS}}(f)$ may be described in terms of the deformation theory of $\sigma_p(f)$.

Since the $\mathscr{L}$-invariant is a local-at-p invariant of f, it is natural to attempt to extract the $\mathscr{L}$-invariant of f from its Jacquet–Langlands lift g to another indefinite quaternion algebra B split at p, that is, with $B_p \cong M_2(\mathbb{Q}_p)$, since the corresponding automorphic representations have the same local components at p. (The case of definite quaternion algebras was resolved by Bertolini, Darmon and Iovita [Bertolini et al. 2010].) Following Darmon [2001], a conjectural method for doing this was proposed in [Greenberg 2009], as follows.

We first consider a certain p-arithmetic subgroup $\Theta \subset B^\times$ of level

$$N^+ := N/\operatorname{disc} B, \tag{1-3}$$

defined precisely in (6-1). We view Θ as a subgroup of $\mathrm{GL}_2(\mathbb{Q}_p)$ using the chosen isomorphism $B_p \cong M_2(\mathbb{Q}_p)$. Let $M^0(X)$ be the space of $\mathbb{C}_p$-valued measures on $X := \mathbb{P}^1(\mathbb{Q}_p)$ with total measure zero (see Section 4). The group Θ acts on X by linear fractional transformations. This induces an action of Θ on $M^0(X)$. A Mayer–Vietoris argument, together with multiplicity one, shows that for each choice of sign $\pm$ at infinity, $\dim_{\mathbb{C}_p} H^1(\Theta, M^0(X))^{g,\pm} = 1$. Here, the superscript g indicates the eigensubspace on which the Hecke operators act according to the Hecke eigenvalues of g. The superscript $\pm$ indicates the ± 1-eigenspace for the natural conjugation action of a matrix of determinant -1 that normalizes Θ. Let $\varphi_g^\pm$ be a nonzero element of $H^1(\Theta, M^0(X))^{g,\pm}$. Our definition of the $\mathscr{L}$-invariant of g will arise by considering the image of $\varphi_g^\pm$ under a certain integration pairing that we now define.

For each $\mathscr{L} \in \mathbb{C}_p$, there is a unique branch $\log_{\mathscr{L}}$ of the p-adic logarithm such that $\log_{\mathscr{L}}(p) = \mathscr{L}$. Let $\mathscr{H}_p = \mathbb{P}^1(\mathbb{C}_p) - \mathbb{P}^1(\mathbb{Q}_p)$ be the p-adic upper half-plane. Associated to each branch of the p-adic logarithm, there is a $\mathrm{PGL}_2(\mathbb{Q}_p)$-invariant integration pairing

$$\langle \cdot, \cdot \rangle_{\mathscr{L}} : M^0(X) \times \operatorname{Div}^0 \mathscr{H}_p \to \mathbb{C}_p$$

defined by

$$\langle \mu, \{\tau'\} - \{\tau\} \rangle_{\mathscr{L}} = \int_X \log_{\mathscr{L}}\Big(\frac{x-\tau'}{x-\tau}\Big)\mu(x),$$

which, in turn, induces a pairing $H^1(\Theta, M^0(X)) \times H_1(\Theta, \operatorname{Div}^0 \mathscr{H}_p) \to \mathbb{C}_p$. Let $\partial : H_2(\Theta, \mathbb{Z}) \to H_1(\Theta, \operatorname{Div}^0 \mathscr{H}_p)$ be the boundary map in the long exact sequence in Θ-cohomology associated to the short exact sequence defining $\operatorname{Div}^0 \mathscr{H}_p$:

$$0 \to \operatorname{Div}^0 \mathscr{H}_p \to \operatorname{Div} \mathscr{H}_p \xrightarrow{\deg} \mathbb{Z} \to 0.$$

Proposition 1 [Greenberg 2009, Prop. 30]. *There are unique constants $\mathscr{L}^D(\varphi_g^\pm)$ in $\mathbb{C}_p$ such that $\big\langle \varphi_g^\pm, \partial H_2(\Theta, \mathbb{Z}) \big\rangle_{-\mathscr{L}^D(\varphi_g^\pm)} = \{0\}$.*

We have chosen the notation $\mathscr{L}^{\mathrm{D}}(\varphi_g^{\pm})$ for these $\mathscr{L}$-invariants since they are defined following methods of Darmon [2001]. The goal of this paper is to relate these $\mathscr{L}$-invariants $\mathscr{L}^{\mathrm{D}}(\varphi_g^{\pm})$ arising from the cohomology of Shimura curves to those whose origins lie in the arithmetic of classical modular curves. The following is our main result:

Theorem 2. $$\mathscr{L}^{D}(\varphi_g^{\pm}) = \mathscr{L}^{GS}(f).$$

Using Theorem 2, we deduce Conjecture 2 of [Greenberg 2009] in the case where the base field is $\mathbb{Q}$; see Section 8 for details. The proof of Theorem 2 falls into two steps. Applying a result of Hida's theory, we deform the Jacquet–Langlands lift g of f into a cohomological Hida family $\Phi_g^{\pm}$. Let $\boldsymbol{a}_p = \boldsymbol{a}_p(k)$ be the eigenvalue of U_p acting on $\Phi_g^{\pm}$. Group cohomological calculations building upon those in [Dasgupta 2005] show that

$$\mathscr{L}^{\mathrm{D}}(\varphi_g^{\pm}) = \frac{d}{dk}\big(1 - \boldsymbol{a}_p(\boldsymbol{g}, k)^2\big)\Big|_{k=2} =: \mathscr{L}^{\mathrm{GS}}(g).$$

It remains to show that $\mathscr{L}^{\mathrm{GS}}(g) = \mathscr{L}^{\mathrm{GS}}(f)$. We prove this in Theorem 8, which asserts a compatibility between the Jacquet–Langlands correspondence with the formation of Hida families. This result is a weak analogue of results of Chenevier [2005] for definite quaternion algebras and may be of independent interest.

In the last section of this paper, we apply our computations to the theory of *Stark–Heegner points*. Let $E/\mathbb{Q}$ be an elliptic curve of conductor Np and suppose that $\mathbb{O}$ is a real quadratic order with fraction field K such that $(\operatorname{disc}\mathbb{O}, Np) = 1$. Assume further that the sign in the functional equation of $L(E/K, s)$ is -1. Then for each character $\chi : \mathrm{Cl}_{\mathbb{O}}^{+} \to \mathbb{C}^{\times}$ of the narrow ideal class group of $\mathbb{O}$, the sign in the functional equation of $L(E/K, \chi, s)$ is also -1. Thus, the conjecture of Birch and Swinnerton-Dyer leads one to expect that

$$\operatorname{rank} E(H_{\mathbb{O}}) = \operatorname{ord}_{s=1} L(E/H_{\mathbb{O}}, s) = \operatorname{ord}_{s=1} \prod_{\chi:\mathrm{Cl}_{\mathbb{O}}^{+}\to\mathbb{C}^{\times}} L(E/K, \chi, s) \geq |\mathrm{Cl}_{\mathbb{O}}^{+}|, \quad (1\text{-}4)$$

where $H_{\mathbb{O}}$ is the narrow ring class field associated to the order $\mathbb{O}$. In [Greenberg 2009], a p-adic analytic construction of local *Stark–Heegner points* on E was presented, generalizing a construction of Darmon [2001] applicable when p is inert in K and the primes dividing N split in K. The local definition of Stark–Heegner points given in [Greenberg 2009] is contingent upon Conjecture 2 [ibid.] over the base field $\mathbb{Q}$; this now follows from Theorem 2. The analytically defined Stark–Heegner points are conjectured to be defined over the field $H_{\mathbb{O}}$, and are expected to generate a finite index subgroup of $E(H_{\mathbb{O}})$ when the inequality in (1-4) is an equality.

The strongest theoretical evidence presented to date for the conjectures of [Darmon 2001] on the rationality of Stark–Heegner points is the main result of [Bertolini and Darmon 2009], which proves the rationality of certain linear combinations of Stark–Heegner points. A key tool in the proof of this result is a description of the formal group logarithms of Stark–Heegner points in terms of periods of Hida families. In Section 9, we prove such a formula for the Stark–Heegner points of [Greenberg 2009]. We intend to pursue the analogue of the rationality result of [Bertolini and Darmon 2009] in future work.

2. Modular forms on quaternion algebras and the cohomology of Shimura curves

Let f be as in the introduction with level Np, $p \nmid N$. In order to ensure that f admits a Jacquet–Langlands lift to an indefinite quaternion $\mathbb{Q}$-algebra, we suppose that the tame part N of the level of f admits a factorization

$$N = N^- N^+, \quad (N^-, N^+) = 1,$$

such that f is N^--new. We work under the additional simplifying assumption that N^- is squarefree.

Let B be the indefinite quaternion $\mathbb{Q}$-algebra with discriminant N^-. Let $R_{\max}$ be a maximal order in B. Let ℓ be a prime with $\ell \nmid N^-$. Since B is split at ℓ, we may choose an embedding

$$\iota_\ell : B \to M_2(\mathbb{Q}_\ell)$$

such that $\iota_\ell(R_{\max}) \subset M_2(\mathbb{Z}_\ell)$. Define

$$R = \left\{\alpha \in R_{\max} : \iota_\ell(\alpha) \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N^+\mathbb{Z}_\ell} \text{ for all } \ell \nmid N^-\right\}, \tag{2-1}$$

$$R_0 = \left\{\alpha \in R_{\max} : \iota_\ell(\alpha) \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{pN^+\mathbb{Z}_\ell} \text{ for all } \ell \nmid N^-\right\}. \tag{2-2}$$

The rings R and R_0 are Eichler orders in B of level N^+ and pN^+, respectively. Set

$$\Gamma = R_+^\times/\{\pm 1\}, \quad \Gamma_0 = R_{0,+}^\times/\{\pm 1\},$$

where the subscript $+$ indicates elements with positive reduced norm.

Since B is split at the infinite place of $\mathbb{Q}$, we may choose an embedding

$$\iota_\infty : B \to M_2(\mathbb{R}). \tag{2-3}$$

The groups Γ and Γ_0 may be viewed as discrete groups of transformations of the complex upper half-plane $\mathscr{H}$ by identifying them with subgroups of $\mathrm{PGL}_2(\mathbb{R})$ via ι_∞. The quotients

$$Y(\mathbb{C}) := \Gamma\backslash\mathscr{H}, \qquad Y_0(\mathbb{C}) := \Gamma_0\backslash\mathscr{H}$$

are Riemann surfaces, compact exactly when $N^- \neq 1$. Let $\mathcal{H}^* = \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q})$ be the extended complex upper half-plane and define

$$X(\mathbb{C}) = \begin{cases} Y(\mathbb{C}) & \text{if } N^- \neq 1, \\ \Gamma \backslash \mathcal{H}^* & \text{if } N^- = 1. \end{cases}$$

Define $X_0(\mathbb{C})$ analogously. The Riemann surfaces $X(\mathbb{C})$ and $X_0(\mathbb{C})$ are compact and may be identified with the loci of complex points of Shimura curves X and X_0 that admit canonical models over $\mathbb{Q}$. Of course, these are just the classical modular curves in the case $N^- = 1$. For the remainder of this section, we assume that $N^- \neq 1$.

Let $S_k(\Gamma)$ and $\overline{S_k(\Gamma)}$ be the spaces of holomorphic and, respectively, antiholomorphic weight k cusp forms on $X(\Gamma)$. The spaces $S_k(\Gamma_0)$ and $\overline{S_k(\Gamma_0)}$ are defined analogously. These spaces admit the action of a commutative algebra of Hecke operators, all commuting with complex conjugation (see Section 3).

Theorem 3 (Jacquet–Langlands correspondence). *Let $k \geq 2$ be an integer. There are isomorphisms*

$$S_k(\Gamma_0(N))^{N^-\text{-new}} \cong S_k(\Gamma) \quad \text{and} \quad S_k(\Gamma_0(Np))^{N^-\text{-new}} \cong S_k(\Gamma_0).$$

Both isomorphisms are equivariant with respect to the Hecke operators T_ℓ for $\ell \nmid Np$, U_ℓ for $\ell \mid N^+$ and W_ℓ for $\ell \mid N^-$. In addition, the first isomorphism equivariant with respect to T_p, and the second is equivariant with respect to U_p.

Therefore, there is a one-dimensional subspace of $S_2(\Gamma_0)$, independent of the choice of isomorphism in the Jacquet–Langlands correspondence, on which the Hecke operators act via the eigenvalues of f. Let g a nonzero element of this space. We call g a *Jacquet–Langlands lift of f*. Let $a_\ell(g) = a_\ell(f)$ be the eigenvalue of T_ℓ, U_ℓ, or $-W_\ell$ acting on g in the cases $\ell \nmid Np$, $\ell \mid pN^+$, and $\ell \mid pN^-$, respectively.

We are also interested in cohomological avatars of g. We have canonical isomorphisms of Betti and group cohomology

$$H^*(\Gamma, E) \cong H^*(X(\mathbb{C}), E), \quad H^*(\Gamma_0, E) \cong H^*(X_0(\mathbb{C}), E)$$

for any characteristic zero field E endowed with the trivial action of Γ. By the de Rham theorem and the Hodge decomposition,

$$\begin{aligned} H^1(\Gamma_0, \mathbb{C}) &= H^1(X_0(\mathbb{C}), \mathbb{C}) \\ &= H^{1,0}(X_0(\mathbb{C}), \mathbb{C}) \oplus H^{0,1}(X_0(\mathbb{C}), \mathbb{C}) \ \cong \ S_2(\Gamma_0) \oplus \overline{S_2(\Gamma_0)}. \end{aligned}$$

Therefore, if E is any field containing the Hecke eigenvalues of g, we have

$$\dim_E H^1(\Gamma_0, E)^g = 2,$$

where the superscript g indicates Hecke eigenspace corresponding to the system of Hecke eigenvalues of g:

$$H^1(\Gamma_0, E)^g = \{c \in H^1(\Gamma_0, E) : T_\ell(c) = a_\ell(g)c \text{ for } \ell \nmid N, \\ U_\ell(c) = a_\ell(g)c \text{ for } \ell \mid pN^+\}.$$

(See Section 3 for a detailed description of Hecke operators acting on group cohomology.) Note that this space is stable for the Atkin–Lehner involutions $-W_\ell$ for $\ell \mid pN^-$ with eigenvalues $a_\ell(g)$. Conjugation by an element of $R_0^\times$ of reduced norm -1 induces an automorphism of $H^1(\Gamma_0, E)$ under which the subspace $H^1(\Gamma_0, E)^g$ is stable. This action corresponds to complex conjugation of cusp forms and is denoted W_∞. Therefore, $H^1(\Gamma_0, E)^g$ decomposes into one-dimensional $\pm$-eigenspaces for this action:

$$H^1(\Gamma_0, E)^g = H^1(\Gamma_0, E)^{g,+} \oplus H^1(\Gamma_0, E)^{g,-}.$$

We denote by $g^\pm$ a nonzero element of $H^1(\Gamma_0, E)^{g,\pm}$. In Section 4 we construct a cohomological Hida family $\Phi_g^\pm$ that specializes to $g^\pm$ in weight 2, and in Section 6 we use $\Phi_g^\pm$ to define the Darmon $\mathscr{L}$-invariant $\mathscr{L}^{\mathrm{D}}(g^\pm)$.

3. Hecke operators and group cohomology

In anticipation of the delicate group cohomological calculations to follow, we carefully set up notation for describing the action of Hecke operators on various cohomology groups. Let $G \subset K$ be an inclusion of groups, x an element of K, M a G-module, and M' an xGx^{-1}-module. Suppose that $\xi : M \to M'$ is a group homomorphism such that

$$\xi(gm) = xgx^{-1}\xi(m). \tag{3-1}$$

for all $g \in G$ and $m \in M$. In our applications, $M \subset M''$ for a K-module M'', and ξ is the map $m \mapsto xm$ with $M' = xM \subset M''$. The map ξ induces a homomorphism

$$\xi_* : H^*(G, M) \to H^*(xGx^{-1}, M') \tag{3-2}$$

as follows: Let $F_\bullet \to \mathbb{Z}$ be a resolution of $\mathbb{Z}$ by free K-modules. Note that F_r is also a free G-module and a free xGx^{-1}-module. In what follows, we will often take $F_r = \mathbb{Z}[K^{r+1}]$. Formally, ξ induces a map of cochain complexes relative to this resolution,

$$\xi_* : \mathrm{Hom}_G(F_r, M) \to \mathrm{Hom}_{xGx^{-1}}(F_r, M'), \quad \xi_*(\varphi)(f_r) = \xi(\varphi(x^{-1}f_r)),$$

which induces (3-2). We now use this formalism to define the Hecke operators that play a role in this paper.

• Suppose that $\ell > 0$ is a prime divisor of N^-. Then there exists an element $\lambda \in R_0$ whose reduced norm is ℓ and such that λ generates the unique two-sided ideal of R_0 with norm ℓ. The element λ normalizes R_0 by [Vignéras 1980, chapitre II, corollaire 1.7]. Take $G = \Gamma_0$ or Γ, $K = B^\times/\mathbb{Q}^\times$, $x = \lambda$. Let M be a G-module such that $M = \lambda M$ (that is, this equality holds in a K-module M'' containing M). The formalism above then yields the *Atkin–Lehner involutions*

$$W_\ell : H^r(\Gamma_0, M) \to H^r(\Gamma_0, M), \quad W_\ell : H^r(\Gamma, M) \to H^r(\Gamma, M). \tag{3-3}$$

• Let $w_p \in R_0$ be an element of reduced norm p that generates the normalizer of Γ_0 in $R[1/p]_+^\times$ and define

$$\widetilde{\Theta} = R[1/p]_+^\times \,\big/\, \mathbb{Z}[1/p]^\times. \tag{3-4}$$

The groups Γ_0, Γ, and $\Gamma' := w_p \Gamma w_p^{-1}$ are all subgroups of $\widetilde{\Theta}$. Using the above formalism with $G = \Gamma_0$ or Γ, $K = \widetilde{\Theta}$, and $x = w_p$ yields *Atkin–Lehner maps*

$$W_p : H^r(\Gamma_0, M) \to H^r(\Gamma_0, M'), \quad W_p : H^r(\Gamma, M) \to H^r(\Gamma', M'), \tag{3-5}$$

with $M' = w_p M$. We note that these maps are isomorphisms, as applying the same formalism with w_p^{-1} instead of w_p yields inverse homomorphisms W_p^{-1}.

• Let $\ell > 0$ be a prime with $\ell \nmid N^-$. Choose an element $\lambda \in R_0$ of reduced norm ℓ. When $\ell \mid pN^+$, we insist that

$$\iota_\ell(\lambda)\mathfrak{I}_\ell \in \begin{pmatrix} 1 & 0 \\ 0 & \ell \end{pmatrix} \mathfrak{I}_\ell, \tag{3-6}$$

where $\mathfrak{I}_\ell$ is the Iwahori subgroup of $\mathrm{GL}_2(\mathbb{Z}_\ell)$ defined by

$$\mathfrak{I}_\ell = \left\{ \alpha \in \mathrm{GL}_2(\mathbb{Z}_\ell) : \alpha \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{\ell} \right\}.$$

Consider a double coset decomposition

$$\Gamma_0 \cdot \lambda \cdot \Gamma_0 = \bigcup_i \gamma_a \Gamma_0. \tag{3-7}$$

Let Σ be the subsemigroup of $\widetilde{\Theta}$ generated by Γ_0 together with λ, and let M be a Σ-module. Let $F_\bullet \to \mathbb{Z}$ be a resolution of $\mathbb{Z}$ by free $\widetilde{\Theta}$-modules, and define an endomorphism T_ℓ of the cochain complex $\mathrm{Hom}_{\Gamma_0}(F_\bullet, M)$ by

$$(T_\ell \varphi)(f_r) = \sum_i \gamma_i \varphi(\gamma_i^{-1} f_r), \quad f_r \in F_r. \tag{3-8}$$

It is routine to check that T_ℓ does not depend on the choice of coset representatives and descends to a well defined endomorphism T_ℓ of $H^*(\Gamma_0, M)$. When $\ell \mid pN^+$, we write U_ℓ instead of T_ℓ for this operator.

• Finally, let Π denote the matrix $\lambda \in R_0$ of reduced norm p chosen above to satisfy (3-6) when $\ell = p$. Let $\Pi' = w_p \Pi w_p^{-1}$. Then

$$\iota_p(\Pi')\mathfrak{I}_p = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \mathfrak{I}_p.$$

Let U_p' be the Hecke operator associated to the double coset $\Gamma_0 \Pi' \Gamma_0$. It is easy to check that

$$U_p' = W_p \circ U_p \circ W_p^{-1}. \tag{3-9}$$

Note that this holds on the level of cochains if we choose compatible double coset decompositions:

$$\Gamma_0 \Pi \Gamma_0 = \bigcup_i \gamma_a \Gamma_0, \qquad \Gamma_0 \Pi' \Gamma_0 = \bigcup_i (w_p \gamma_a w_p^{-1}) \Gamma_0.$$

4. *p*-adic measures, Hida families, and Greenberg–Stevens $\mathscr{L}$-invariants

Let Y be a compact topological space with a basis of compact-open subsets and let A be a subring of $\mathbb{C}_p$. Write $C^\infty(Y) = C^\infty(Y, A)$ for the group of locally constant, A-valued functions on Y, equipped with the sup-norm. An *A-valued measure on Y* is a bounded A-linear functional on $C^\infty(Y, A)$. We write $M(Y) = M(Y, A)$ for the space of such measures, which can be identified with the space of finitely additive, A-valued functions on the set of compact-open subsets of Y whose values are bounded. For details, see [Mazur and Swinnerton-Dyer 1974, §7.1].

Let

$$\mathbb{X} = (\mathbb{Z}_p^2)' := \mathbb{Z}_p^2 - p(\mathbb{Z}_p^2), \quad \mathbb{X}_\infty = \mathbb{Z}_p^\times \times p\mathbb{Z}_p \subset \mathbb{X}. \tag{4-1}$$

The spaces $M(\mathbb{X})$ and $M(\mathbb{X}_\infty)$ are naturally modules for the Iwasawa algebra $\Lambda := \mathbb{Z}_p[\![1 + p\mathbb{Z}_p]\!]$, where group-like elements act via the natural diagonal action of $1 + p\mathbb{Z}_p$ on $\mathbb{X}$; given $\ell \in 1 + p\mathbb{Z}_p$, we define $([\ell]\mu)(h(x, y)) := \mu(h(\ell x, \ell y))$.

Let

$$\epsilon : \Lambda \to \mathbb{Z}_p \tag{4-2}$$

be the augmentation map defined by $[\ell] \mapsto 1$ and let I_ϵ be the kernel of ϵ. Letting γ be a topological generator of $1 + p\mathbb{Z}_p$, it follows that I_ϵ is generated by

$$\varpi := [\gamma] - 1.$$

The group $\mathrm{GL}_2(\mathbb{Z}_p)$ acts on $\mathbb{X}$ from the left by viewing elements of $\mathbb{X}$ as column vectors. The group Γ acts on $\mathbb{X}$ via the embedding $\iota_p : R^\times \hookrightarrow \mathrm{GL}_2(\mathbb{Z}_p)$, and $\mathbb{X}_\infty$ is stable under Γ_0. Therefore, we may consider the cohomology groups $H^*(\Gamma, M(\mathbb{X}))$ and $H^*(\Gamma_0, M(\mathbb{X}_\infty))$. These cohomology groups are canonically isomorphic:

Lemma 4. *The map*

$$H^*(\Gamma, M(\mathbb{X})) \to H^*(\Gamma_0, M(\mathbb{X}_\infty))$$

induced by the Γ_0*-equivariant inclusion* $\mathbb{X}_\infty \hookrightarrow \mathbb{X}$ *is an isomorphism.*

Proof. The $p+1$ translates of $\mathbb{X}_\infty$ by Γ cover $\mathbb{X}$. It follows that

$$M(\mathbb{X}) = \text{Co-Ind}_{\Gamma_0}^{\Gamma} M(\mathbb{X}_\infty).$$

The lemma now follows from Shapiro's lemma. □

Let us assume that our measures take values in $\mathbb{Z}_p$ (so $M(\mathbb{X})$ denotes $M(\mathbb{X}, \mathbb{Z}_p)$, etc.). We set $\widetilde{\mathbb{W}} := H^1(\Gamma_0, M(\mathbb{X}_\infty)) \cong H^1(\Gamma, M(\mathbb{X}))$. View Λ as a $\mathbb{Z}_p[\![Z_p^\times]\!]$-algebra via the canonical projection

$$\mathbb{Z}_p^\times \to 1 + p\mathbb{Z}_p, \qquad \ell \mapsto \langle \ell \rangle := \ell/\omega(\ell),$$

where ω is the Teichmuller character. Define the Λ-algebra $\mathbb{W} := \widetilde{\mathbb{W}} \otimes_{\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]} \Lambda$.

As $\Pi\mathbb{X}_\infty \subset \mathbb{X}_\infty$, the semigroup Σ of Section 3 acts on $M(\mathbb{X}_\infty)$. Therefore, the formalism of Section 3 endows $\mathbb{W}$ with an action of the U_p-operator. In addition to the U_p-action, the group $\mathbb{W}$ enjoys an action of

- Hecke operators T_ℓ for primes $\ell \nmid pN$ and U_ℓ for $\ell \mid N^+$, and
- Atkin–Lehner involutions W_ℓ for $\ell \mid N^-$.

See Section 3 for the definitions of these operators. Let $\mathbb{T}$ be the commutative Λ-subalgebra of $\text{End}_\Lambda \mathbb{W}$ generated by these operators. Let $\rho : M(\mathbb{X}_\infty) \to \mathbb{Z}_p$ be the total measure map. It induces a corresponding map

$$\rho : \mathbb{W} \to H^1(\Gamma_0, \mathbb{Z}_p). \tag{4-3}$$

The map ρ respects the decomposition into $\pm$-eigenspaces:

$$\rho : \mathbb{W}^\pm \to H^1(\Gamma_0, \mathbb{Z}_p)^\pm.$$

Let $e = \lim_{n\to\infty} U_p^{n!}$ denote Hida's ordinary idempotent and, for any $\mathbb{T}$-module M, let $M^o = eM$. In particular, $\mathbb{T}^o = e\mathbb{T}$ is Hida's ordinary Hecke algebra.

Theorem 5 (Hida's control theorem). *There is an exact sequence*

$$0 \to \varpi\mathbb{W}^{\pm,o} \to \mathbb{W}^{\pm,o} \xrightarrow{\rho} H^1(\Gamma_0, \mathbb{Z}_p)^{\pm,o} \to 0. \tag{4-4}$$

The kernel of the Λ-algebra homomorphism $\mathbb{T}^o \to \overline{\mathbb{Q}}_p$ given by sending a Hecke operator to its eigenvalue on g is a prime ideal $\mathfrak{p} \subset \mathbb{T}^o$ lying above the augmentation ideal $I_\epsilon \subset \Lambda$. The following fundamental result is due to Hida in the case $N^- = 1$ (see [Greenberg and Stevens 1993]), and was extended in [Balasubramanyam and Longo 2011] to the case $N^- \neq 1$.

Theorem 6. *There is a unique minimal prime $\mathfrak{P} \subset \mathfrak{p}$, and the quotient $R := \mathbb{T}^o/\mathfrak{P}$ is a finite flat extension of Λ unramified above I_ϵ.*

Let R be as in the theorem, and let $R_\mathfrak{p}$ be the localization of R at $\mathfrak{p}$. Let E be the field of fractions of the integral closure of $\mathbb{Z}_p$ in R. It is a finite extension of $\mathbb{Q}_p$. We write $\epsilon : R_\mathfrak{p} \to R_\mathfrak{p}/\varpi R_\mathfrak{p} \cong E$ for the reduction map. This notation is justified as this map extends the augmentation $\epsilon : \Lambda \to \mathbb{Z}_p$.

Write $(\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,g}$ for the subspace of $(\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm}$ on which $\mathbb{T}$ acts via the canonical map $\mathbb{T} \to R_\mathfrak{p}$. Note that

$$(\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,g} \subset (\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,o} = \mathbb{W}^{\pm,o} \otimes_\Lambda R_\mathfrak{p}$$

and that

$$H^1(\Gamma_0, \mathbb{Z}_p) \otimes_\Lambda R_\mathfrak{p} = H^1(\Gamma_0, \mathbb{Z}_p) \otimes_{\mathbb{Z}_p} E = H^1(\Gamma_0, E). \tag{4-5}$$

On the left of (4-5), we view $H^1(\Gamma_0, \mathbb{Z}_p)$ as a Λ-module via the augmentation ϵ.

Corollary 7 [Balasubramanyam and Longo 2011, §3.6]. *The sequence*

$$0 \to \varpi(\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,g} \to (\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,g} \to H^1(\Gamma_0, E)^{\pm,g} \to 0$$

obtained by tensoring (4-4) *with $R_\mathfrak{p}$ over Λ and taking g-isotypic components is exact, and* $\operatorname{rank}_{R_\mathfrak{p}}(\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,g} = 1$.

We now view $g^\pm$ as an element of $H^1(\Gamma_0, E)^{\pm,g}$. By Corollary 7, we may choose a lift

$$\Phi_g^\pm \in (\mathbb{W} \otimes_\Lambda R_\mathfrak{p})^{\pm,g} \tag{4-6}$$

of $g^\pm$. The element $\Phi_g^\pm$ is well defined up to multiplication by an element of $1 + \varpi R_\mathfrak{p}$. We call $\Phi_g^\pm$ a *Hida family through* $g^\pm$. We denote its U_p-eigenvalue by $\boldsymbol{a}_p(\Phi_g^\pm) \in R_\mathfrak{p}$. Since $\epsilon(\boldsymbol{a}_p(\Phi_g^\pm)) = a_p(g^\pm) = a_p(g) = a_p(f) = \pm 1$, we see that $1 - \boldsymbol{a}_p(\Phi_g^\pm)^2$ lies in $\varpi R_\mathfrak{p}$. There is a "derivative map" $d_\epsilon : \varpi R_\mathfrak{p}/(\varpi R_\mathfrak{p})^2 \to E$ that extends the map $I_\epsilon/I_\epsilon^2 \to \mathbb{Z}_p$ given by the p-adic logarithm:

$$[\ell] - 1 \mapsto \log(\ell). \tag{4-7}$$

Since $\ell \in \mathbb{Z}_p^\times$, we need not specify a branch of the p-adic logarithm. We define the *Greenberg–Stevens $\mathscr{L}$-invariant of g* by

$$\mathscr{L}^{\mathrm{GS}}(\Phi_g^\pm) = d_\epsilon\big(1 - \boldsymbol{a}_p(\Phi_g^\pm)^2\big) \in E.$$

The derivative map d_ϵ is related to the usual notion of derivative in the following way. For $0 < r \le 1$, let $\mathscr{A}_r$ be the subring of $\overline{\mathbb{Q}}[\![x]\!]$ consisting of those powers series that converge on the closed disk centered at 0 with radius r. Evidently, if $r < s$, then there is a canonical inclusion $\mathscr{A}_s \subset \mathscr{A}_r$. Therefore, we may set $\mathscr{A} = \bigcup_r \mathscr{A}_r$. Define $i : \Lambda \to \mathscr{A}_1$ by sending a group-like element $[\ell]$, for $\ell \in 1 + p\mathbb{Z}_p$, to the

function $k \mapsto \ell^{k-2}$. Since R is unramified over I_ϵ and $\mathscr{A}$ is Henselian, there is a unique extension of i to a Λ-algebra homomorphism $i : R_\mathfrak{p} \to \mathscr{A}$. An element $\lambda \in R_\mathfrak{p}$ lies in $\varpi R_\mathfrak{p}$ if and only if the associated analytic function $i(\lambda)$ has a zero at $k = 2$. In this case, $d_\epsilon(\lambda) = i(\lambda)'(2)$.

Theorem 8. *We have the following equality of Greenberg–Stevens $\mathscr{L}$-invariants*:

$$\mathscr{L}^{GS}(\Phi_g^\pm) = \mathscr{L}^{GS}(f).$$

Proof. Suppose R' is a finitely generated R-subalgebra of $R_\mathfrak{p}$ such that $\Phi_g^\pm$ lies in $(\mathbb{W} \otimes_\Lambda R')^{g,\pm}$. With notation as above, there is some r_0 such that $i(R')$ is contained in $\mathscr{A}_{r_0}$.

Let $P_{k-2}(\overline{\mathbb{Q}})$ be the space of homogeneous polynomials of degree $k-2$ in indeterminates x and y, and let $V_{k-2}(\overline{\mathbb{Q}})$ be its $\overline{\mathbb{Q}}$-linear dual. Define a "specialization to weight k" map

$$\rho_k : M(\mathbb{X}_\infty) \to V_{k-2}(\overline{\mathbb{Q}})$$

by the rule

$$\rho_k(\Phi)(P) = \int_{\mathbb{X}_\infty} P(x, y)\Phi(x, y).$$

This map being Γ_0-equivariant, it induces a homomorphism

$$\rho_k : H^1(\Gamma_0, M(\mathbb{X}_\infty)) \to H^1(\Gamma_0, V_{k-2}(\overline{\mathbb{Q}})).$$

The map ρ defined in (4-3) coincides with ρ_2 in this more general notation.

If $|k-2|_p \le r$, we may extend ρ_k to a map

$$\rho_k : H^1(\Gamma_0, M(\mathbb{X}_\infty)) \otimes_\Lambda \mathscr{A}_r \to H^1(\Gamma_0, V_{k-2}(\overline{\mathbb{Q}}))$$

by setting

$$\rho_k\Big(\sum_i \varphi_i \otimes \alpha_i\Big) = \sum_i \alpha_i(k)\rho_k(\varphi_i).$$

One may verify formally that ρ_k is Hecke-equivariant.

Let $\boldsymbol{a}_\ell$ be the image in $\mathscr{A}_{r_0}$ of the eigenvalue of T_ℓ, $-\langle\ell\rangle^{(k-2)/2}W_\ell$, or U_ℓ acting on $\Phi_g^\pm$ in the cases $\ell \nmid Np$, $\ell \mid N^-$, and $\ell \mid N^+p$, respectively. Here $\langle\ell\rangle$ denotes the projection of ℓ onto $1 + p\mathbb{Z}_p$. Set $\boldsymbol{a}_1 = 1$ and define $\boldsymbol{a}_n$ in terms of the $\boldsymbol{a}_\ell$ with $\ell \mid n$ by the usual formulas for Hecke operators.

We may shrink r_0 if necessary to ensure that $\rho_k(\Phi_g^\pm)$ is a nonzero element of $H^1(\Gamma_0, V_{k-2}(\overline{\mathbb{Q}}_p))$ for all $k \ge 2$ with $|k-2|_p \le r_0$ and $k \equiv 2 \pmod{p-1}$. The class $\rho_k(\Phi_g^\pm)$ is an eigenvector for the ℓ-th Hecke operator with eigenvalue $\boldsymbol{a}_\ell(k)$. Thus, $\{\boldsymbol{a}_\ell(k)\}$ is a system of Hecke eigenvalues occurring in $H^1(\Gamma_0, V_{k-2}(\overline{\mathbb{Q}}_p))$. In particular, $\{\boldsymbol{a}_\ell(k)\} \subset \overline{\mathbb{Q}} \subset \overline{\mathbb{Q}}_p$. By the Eichler–Shimura isomorphism [Matsushima and Shimura 1963, §4], this system of Hecke eigenvalues also occurs in $S_k(\Gamma_0)$.

By the Jacquet–Langlands correspondence, it occurs in $S_k(\Gamma_0(pN))$ as well. Thus, if we set

$$\boldsymbol{h} := \sum_{n=1}^{\infty} \boldsymbol{a}_n q^n \in \mathscr{A}_{r_0}[\![q]\!],$$

then $\boldsymbol{h}(k) = \sum \boldsymbol{a}_n(k) q^n$ is in fact the q-expansion of a classical cusp form of weight k on $\Gamma_0(Np)$ for $k \geq 2$, $|k-2|_p \leq r_0$, $k \equiv 2 \pmod{p-1}$. Furthermore, it is clear that $\boldsymbol{h}(2) = f$. Therefore, by the uniqueness of the Hida family through f [Hida 1986, Corollary 1.3, pg. 554], it follows that $\boldsymbol{a}_n(k) = \boldsymbol{a}_n(\boldsymbol{f}, k)$ for $|k-2|_p \leq r_0$. In particular, this is true for $n = p$; Theorem 8 follows. □

Finally, we record a result that will be important later. Set

$$\mathbb{W}^0 = H^1(\Gamma, M^0(\mathbb{X})) \otimes_{\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]} \Lambda.$$

Lemma 9. *The canonical map*

$$(\mathbb{W}^0 \otimes_\Lambda R_{\mathfrak{p}})^{\pm, g} \to (\mathbb{W} \otimes_\Lambda R_{\mathfrak{p}})^{\pm, g} \tag{4-8}$$

is an isomorphism.

Proof. The map $\rho : M(\mathbb{X}) \to \mathbb{Z}_p$ gives rise to the short exact sequence

$$0 \to M^0(\mathbb{X}) \to M(\mathbb{X}) \xrightarrow{\rho} \mathbb{Z}_p \to 0.$$

Since R is Λ-flat, we may tensor $R_{\mathfrak{p}}$ with the associated long exact sequence in Γ-cohomology to obtain

$$\cdots \to H^0(\Gamma, E) \to \mathbb{W}^0 \otimes_\Lambda R_{\mathfrak{p}} \to \mathbb{W} \otimes_\Lambda R_{\mathfrak{p}} \to H^1(\Gamma, E) \to \cdots.$$

The space $H^0(\Gamma, E)$ is Eisenstein (that is, T_ℓ acts as $1+\ell$), so its g-isotypic component is trivial. Since the maps in the sequence above are Hecke-equivariant, it follows that the map (4-8) is injective. Similarly, if $\Phi \in (\mathbb{W} \otimes_\Lambda R_{\mathfrak{p}})^{\pm, g}$, then its image in $H^1(\Gamma, E)$ must be zero. This holds because g is p-new of level Γ_0, so the system of Hecke eigenvalues of g does not occur in $H^1(\Gamma, E)$. Therefore Φ is the image of an element $\widetilde{\Phi} \in \mathbb{W}^0 \otimes_\Lambda R_{\mathfrak{p}}$. Let ℓ be any prime such that the eigenvalue $a_\ell(g)$ of the Hecke operator T_ℓ is not equal to $\ell+1$. Let $\boldsymbol{a}_\ell(\Phi)$ denote the T_ℓ eigenvalue of Φ, that is, the image of T_ℓ in $R_{\mathfrak{p}}$. We claim that

$$\widetilde{\Phi}' := \frac{T_\ell - (\ell+1)}{\boldsymbol{a}_\ell(\Phi) - (\ell+1)} \widetilde{\Phi} \tag{4-9}$$

is a lift of Φ to $(\mathbb{W}^0 \otimes_\Lambda R_{\mathfrak{p}})^{\pm, g}$. First note that the division in (4-9) is allowed in the localization, since the image of $\boldsymbol{a}_\ell(\Phi) - (\ell+1)$ under reduction modulo $\mathfrak{p}$ is $a_\ell(g) - (\ell+1) \neq 0$. Next, it is clear that $\widetilde{\Phi}'$ maps to Φ under (4-8) since Φ has T_ℓ eigenvalue $\boldsymbol{a}_\ell(\Phi)$. Finally, let $\lambda \in \mathbb{T}^o$, and let $\boldsymbol{a}_\lambda(\Phi)$ be the corresponding eigenvalue of Φ. Then $(\lambda - \boldsymbol{a}_\lambda(\Phi))\widetilde{\Phi}$ maps to 0 in $\mathbb{W} \otimes_\Lambda R_{\mathfrak{p}}$ and hence arises from

$H^0(\Gamma, E)$. Since this module is Eisenstein, it is killed by $T_\ell - (\ell+1)$, and it follows that $(\lambda - \boldsymbol{a}_\lambda(\Phi))\widetilde{\Phi}' = 0$. This shows that $\widetilde{\Phi}'$ lies in $(\mathbb{W}^0 \otimes_\Lambda R_\mathfrak{p})^{\pm,g}$, and concludes the proof of the lemma. □

Using Lemma 9, we may view $\Phi_g^\pm$ an element of $(\mathbb{W}^0 \otimes_\Lambda R_\mathfrak{p})^{\pm,g}$.

5. Some commutative diagrams

In this section, we establish some commutative diagrams involving the operators U_p, U_p', and W_p acting on the group cohomology of various spaces of p-adic measures. In fact, these diagrams are so natural that they commute on the level of cochains; this fact will be used heavily in the calculations of Section 7. Recall the group $\widetilde{\Theta}$ defined in (3-4). We describe cohomology classes in terms of homogeneous cochains relative to the complex of projective $\widetilde{\Theta}$-modules

$$F_r := \mathbb{Z}[\widetilde{\Theta}^{r+1}]. \tag{5-1}$$

Thus, if G is a subgroup of $\widetilde{\Theta}$, our group of M-valued r-cochains is

$$C^r(G, M) := \mathrm{Hom}_G(F_r, M). \tag{5-2}$$

Coboundary maps $d : C^r(G, M) \to C^{r+1}(G, M)$ are defined by the usual formula

$$d\varphi(g_0, \ldots, g_{r+1}) = \sum_{i=0}^{r+1} (-1)^i \varphi(g_0, \ldots, \widehat{g_i}, \ldots, g_{r+1}).$$

We write

$$Z^r(G, M) = \mathrm{Ker}(d : C^r(G, M) \to C^{r+1}(G, M)),$$
$$B^r(G, M) = \mathrm{Image}(d : C^{r-1}(G, M) \to C^r(G, M)),$$

and have

$$H^r(G, M) = Z^r(G, M)/B^r(G, M).$$

Defining

$$\mathbb{X}_p = \mathbb{Z}_p \times \mathbb{Z}_p^\times = w_p^{-1}\mathbb{X}_\infty, \tag{5-3}$$

we obtain Atkin–Lehner maps as in (3-5) with $M = M(\mathbb{X}_\infty)$ and $M' = M(\mathbb{X}_p)$.

Proposition 10. *The following diagrams commute*:

$$\begin{array}{ccccc} & & C^r(\Gamma, M(\mathbb{X})) & & \\ & \swarrow^{\rho_{\mathbb{X}_\infty}} & & \searrow^{\rho_{\mathbb{X}_p}} & \\ C^r(\Gamma_0, M(\mathbb{X}_\infty)) & \xrightarrow[U_p]{} & C^r(\Gamma_0, M(\mathbb{X}_\infty)) & \xrightarrow[W_p^{-1}]{} & C^r(\Gamma_0, M(\mathbb{X}_p)) \end{array} \tag{5-4}$$

$$\begin{array}{ccccc} & & C^r(\Gamma', M(w_p\mathbb{X})) & & \\ & \swarrow^{\rho'_{p\mathbb{X}_p}} & & \searrow^{\rho'_{\mathbb{X}_\infty}} & \\ C^r(\Gamma_0, M(p\mathbb{X}_p)) & \xrightarrow[U'_p]{} & C^r(\Gamma_0, M(p\mathbb{X}_p)) & \xrightarrow[W_p^{-1}]{} & C^r(\Gamma_0, M(\mathbb{X}_\infty)) \end{array} \tag{5-5}$$

Here the maps ρ are the natural restriction maps.

Proof. Let $\varphi \in Z^r(\Gamma, M(\mathbb{X}))$. Let $g \in \widetilde{\Theta}^{r+1}$, and let h be a locally analytic function on $\mathbb{X}_p$. In the following, we will write $j_!$ for the extension-by-zero of a function j on $\mathbb{X}_\infty$ to a function on $\mathbb{X}$. We compute:

$$\begin{aligned} (W_p^{-1}U_p\rho_{\mathbb{X}_\infty}\varphi)(g)(h) &= (U_p\rho_{\mathbb{X}_\infty}\varphi)(w_pg)(h|w_p^{-1}) \\ &= \sum_{0\le i\le p-1} (\rho_{\mathbb{X}_\infty}\varphi)(\delta_i^{-1}w_pg)(h|w_p^{-1}\delta_i) \\ &= \sum_{0\le i\le p-1} \varphi(\delta_i^{-1}w_pg)((h|w_p^{-1}\delta_i)_!) \\ &= \sum_{0\le i\le p-1} \varphi(g)((h|w_p^{-1}\delta_i)_!|\delta_i^{-1}w_p) \\ &= \sum_{0\le i\le p-1} \varphi(g)(h_!\mathbf{1}_{\pi^{-1}(i+p\mathbb{Z}_p)}) \\ &= (\rho_{\mathbb{X}_p}\varphi)(g)(h). \end{aligned}$$

Essential in this calculation is that $w_p^{-1}\delta_i$ belongs to Γ and that

$$w_p^{-1}\delta_i(\mathbb{X}_\infty) = \gamma_i w_p^{-1}(\mathbb{X}_\infty) = \gamma_i(\mathbb{X}_p) = \pi^{-1}(i + p\mathbb{Z}_p).$$

The commutativity of (5-5) follows from applying the operator W_p to (5-4). □

Next, we will be interested in understanding the map

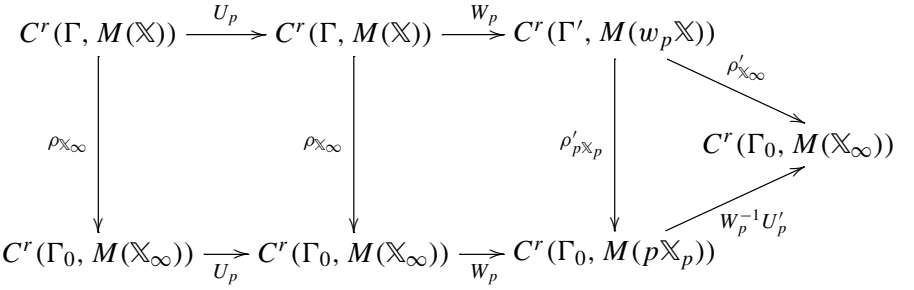
$$W_pU_p : H^r(\Gamma, M(\mathbb{X})) \to H^r(\Gamma', M(w_p\mathbb{X}))$$

with respect to the decomposition $w_p\mathbb{X} = \mathbb{X}_\infty \sqcup p\mathbb{X}_p$.

Proposition 11. *The following diagram commutes*:

$$\begin{array}{ccc} C^r(\Gamma, M(\mathbb{X})) & \xrightarrow{\rho_{\mathbb{X}_\infty}} & C^r(\Gamma_0, M(\mathbb{X}_\infty)) \\ \Big\downarrow{\scriptstyle W_pU_p} & & \Big\downarrow{\scriptstyle U_p^2} \\ C^r(\Gamma', M(w_p\mathbb{X})) & \xrightarrow[\rho'_{\mathbb{X}_\infty}]{} & C^r(\Gamma_0, M(\mathbb{X}_\infty)) \end{array}$$

Proof. The result follows from the following commutative diagram and (3-9). Note that the commutativity of the triangle on the right is given by that of (5-5).



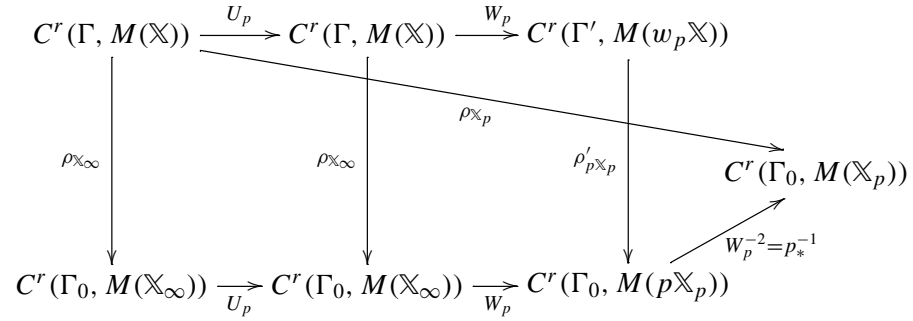
□

Proposition 12. *The following diagram commutes*:

$$\begin{array}{ccc} H^r(\Gamma, M(\mathbb{X})) & \xrightarrow{\rho_{\mathbb{X}_p}} & H^r(\Gamma_0, M(\mathbb{X}_p)) \\ {\scriptstyle W_pU_p}\downarrow & & \downarrow{\scriptstyle p_*} \\ H^r(\Gamma', M(w_p\mathbb{X})) & \xrightarrow[\rho'_{p\mathbb{X}_p}]{} & H^r(\Gamma_0, M(p\mathbb{X}_p)) \end{array}$$

Here the map $p_* : H^r(\Gamma_0, M(\mathbb{X}_p)) \to H^r(\Gamma_0, M(p\mathbb{X}_p))$ *is induced by* $p_*h(x, y) = h(px, py)$ *for a locally analytic function* h *on* $p\mathbb{X}_p$.

Proof. The result follows from the following commutative diagram.



The commutativity of the diagonal map $\rho_{\mathbb{X}_p}$ with the arrows that lie below it follows from that of (5-4). The fact that $W_p^2 = p_*$ follows from the fact that $w_p^2 \in p\Gamma_0$ and hence induces the same map on Γ_0-cohomology as multiplication by p. □

6. p-arithmetic cohomology classes and Darmon $\mathscr{L}$-invariants

Let

$$\Theta = \ker\left(\mathrm{ord}_p \circ \mathrm{nrd} : \widetilde{\Theta} \to \mathbb{Z}/2\mathbb{Z}\right), \tag{6-1}$$

where nrd : $B^\times \to \mathbb{Q}^\times$ is the reduced norm map. Thus, Θ is a normal subgroup of $\widetilde{\Theta}$ of index two and $\widetilde{\Theta}/\Theta$ is generated by the image of w_p. By analyzing its action on the Bruhat–Tits tree of $\mathrm{PGL}_2(\mathbb{Q}_p)$, the group Θ can be expressed as an amalgamation (free product) $\Theta \cong \Gamma *_{\Gamma_0} \Gamma'$ [Greenberg 2009]. Associated to such an amalgamation and a Θ-module M, there is a Mayer–Vietoris sequence

$$\cdots \to H^{r-1}(\Gamma_0, M) \xrightarrow{\delta} H^r(\Theta, M) \xrightarrow{(\mathrm{res}^\Theta_\Gamma, \mathrm{res}^\Theta_{\Gamma'})} H^r(\Gamma, M) \oplus H^r(\Gamma', M) \xrightarrow{(\mathrm{res}^\Gamma_{\Gamma_0} - \mathrm{res}^{\Gamma'}_{\Gamma_0})} H^r(\Gamma_0, M) \to \cdots. \quad (6\text{-}2)$$

Recall that we defined $X = \mathbb{P}^1(\mathbb{Q}_p)$. View $\mathbb{Q}_p$ as a subspace of $\mathbb{P}^1(\mathbb{Q}_p)$ via the inclusion $z \mapsto (z : 1)$. Thus, $(x : y)$ can be identified with the fraction x/y. Set $\infty = (1 : 0)$. We view $\mathbb{Z}_p \subset \mathbb{Q}_p$ as a subspace of X and set

$$X_\infty = X - \mathbb{Z}_p = w_p \mathbb{Z}_p.$$

Our first goal in this section is to use (6-2) in order to construct a cohomology class in $H^1(\Theta, M^0(X))^\pm$ associated to $g^\pm$. (Such a class is constructed in [Greenberg 2009] using different methods.) The map

$$\pi : \mathbb{X} \to X, \qquad \pi(x, y) = (x : y)$$

and the induced pushforward of measures $\pi_* : M(\mathbb{X}) \to M(X)$ can be described via the following isomorphism, a consequence of the fact that π is a $\mathbb{Z}_p^\times$-fibration:

$$M(X) \cong M(\mathbb{X}) \otimes_{\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]} \mathbb{Z}_p. \quad (6\text{-}3)$$

Here, $\mathbb{Z}_p$ is given the structure of a $\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]$-algebra via the augmentation map defined in (4-2). Recall that by Lemma 9, we may assume that the cohomological Hida family $\Phi_g^\pm$ associated to $g^\pm$ belongs to $H^1(\Gamma, M^0(\mathbb{X})) \otimes_{\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]} R_\mathfrak{p}$. For notational simplicity, we suppress the $\otimes_{\mathbb{Z}_p[\![\mathbb{Z}_p^\times]\!]} R_\mathfrak{p}$ in the sequel and write $g^\pm \in H^1(\Gamma, M^0(\mathbb{X}))$; this does not affect any subsequent arguments in a substantive way, though our measures now take values in E.

Proposition 13. *There is a unique cohomology class $\varphi_g^\pm \in H^1(\Theta, M^0(X))$ such that*

$$\mathrm{res}^\Theta_\Gamma \varphi_g^\pm = \pi_* \Phi_g^\pm, \quad \mathrm{res}^\Theta_{\Gamma'} \varphi_g^\pm = \pi_* W_p U_p \Phi_g^\pm.$$

Proof. The uniqueness follows from (6-2) as $H^0(\Gamma_0, M^0(X)) = 0$. We must show the existence of $\varphi_g^\pm$. To this end, let

$$\varphi_g^\pm = \pi_* \Phi_g^\pm \in H^1(\Gamma, M^0(X)), \quad \varphi_g'^\pm = \pi_* W_p U_p \Phi_g^\pm \in H^1(\Gamma', M^0(X)).$$

From (6-2), we must show that $\mathrm{res}^\Gamma_{\Gamma_0} \varphi_g^\pm = \mathrm{res}^\Gamma_{\Gamma_0} \varphi_g'^\pm$ in $H^1(\Gamma_0, M^0(X))$. Since the kernel of $H^1(\Gamma_0, M^0(X)) \to H^1(\Gamma_0, M(X))$ is Eisenstein, it suffices to prove this

equality after viewing $\varphi_g^{\pm}$ and $\varphi_g'^{\pm}$ as taking values in $M(X)$. Let

$$\begin{aligned}
\rho_{\mathbb{Z}_p} &: H^1(\Gamma, M(X)) \to H^1(\Gamma_0, M(\mathbb{Z}_p)) \\
\rho'_{X_\infty} &: H^1(\Gamma', M(X)) \to H^1(\Gamma_0, M(\mathbb{Z}_p)) \\
\rho_{X_\infty} &: H^1(\Gamma, M(X)) \to H^1(\Gamma_0, M(X_\infty)) \\
\rho'_{X_\infty} &: H^1(\Gamma', M(X)) \to H^1(\Gamma_0, M(X_\infty))
\end{aligned}$$

be the maps induced by the inclusions $\mathbb{Z}_p \hookrightarrow X$ and $X_\infty \hookrightarrow X$ and restriction of groups to Γ_0. From the decomposition

$$H^1(\Gamma_0, M(X)) = H^1(\Gamma_0, M(\mathbb{Z}_p)) \oplus H^1(\Gamma_0, M(X_\infty)),$$

we must show that $\rho_{\mathbb{Z}_p}\varphi_g^{\pm} = \rho_{\mathbb{Z}_p}\varphi_g'^{\pm}$ and $\rho_{X_\infty}\varphi_g^{\pm} = \rho'_{X_\infty}\varphi_g'^{\pm}$. By Propositions 11 and 12, the following diagrams commute:

$$\begin{array}{ccc}
H^1(\Gamma, M(X)) & & \\
{\scriptstyle W_pU_p}\downarrow & \searrow^{\rho_{\mathbb{Z}_p}} & \\
H^1(\Gamma', M(X) & \xrightarrow[\rho'_{\mathbb{Z}_p}]{} & H^1(\Gamma_0, M(\mathbb{Z}_p))
\end{array}
\qquad
\begin{array}{ccc}
H^1(\Gamma, M(X)) & \xrightarrow{\rho_{X_\infty}} & H^1(\Gamma_0, M(X_\infty)) \\
{\scriptstyle W_pU_p}\downarrow & & \downarrow{\scriptstyle U_p^2} \\
H^1(\Gamma', M(X)) & \xrightarrow[\rho'_{X_\infty}]{} & H^1(\Gamma_0, M(X_\infty))
\end{array}$$

The diagram on the left proves $\rho_{\mathbb{Z}_p}\varphi_g^{\pm} = \rho_{\mathbb{Z}_p}\varphi_g'^{\pm}$, one of the desired identities. The one on the right says $\rho'_{X_\infty}\varphi_g'^{\pm} = U_p^2\rho_{X_\infty}\varphi_g^{\pm}$. By (6-3),

$$U_p^2\rho_{X_\infty}\varphi_g^{\pm} = \epsilon(\boldsymbol{a}_p(\Phi_g^{\pm}))^2\rho_{X_\infty}\varphi_g^{\pm} = \rho_{X_\infty}\varphi_g^{\pm},$$

completing the proof. □

For each choice of $\mathscr{L} \in \mathbb{P}^1(E)$, we define an integration map

$$\kappa_{\mathscr{L}} : H^r(\Theta, M^0(X)) \to H^{r+1}(\Theta, E)$$

as follows: Let $C(X)$ denote the space of continuous E-valued functions on X. Choose a base-point $\tau \in \mathscr{H}_p(E) = \mathbb{P}^1(E) - \mathbb{P}^1(\mathbb{Q}_p)$ and define

$$\xi_{\mathscr{L},\tau} \in C^1(\widetilde{\Theta}, C(X)/E)$$

by

$$\xi_{\mathscr{L},\tau}(g_0, g_1) = \begin{cases} \log_{\mathscr{L}}\left(\dfrac{z-g_1\tau}{z-g_0\tau}\right) & \text{if } \mathscr{L} \in E, \\ \operatorname{ord}_p\left(\dfrac{z-g_1\tau}{z-g_0\tau}\right) & \text{if } \mathscr{L} = \infty. \end{cases}$$

It is easy to see that $d\xi_{\mathscr{L},\tau} = 0$ and that the cohomology class represented by $\xi_{\mathscr{L},\tau}$ does not depend on τ.

Let G be any subgroup of $\widetilde{\Theta}$, let $\varphi \in C^r(G, M^0(X))$, and consider the cup product

$$\xi_{\mathscr{L},\tau} \cup \varphi \in C^{r+1}(G, (C(X)/E) \otimes_E M^0(X)).$$

The $\widetilde{\Theta}$-invariant integration pairing $(C(X)/E) \otimes_E M^0(X) \to E$ induces a map

$$I : C^{r+1}(G, (C(X)/E) \otimes_E M^0(X)) \to C^{r+1}(G, E).$$

Set $\kappa_{\mathscr{L},\tau}(\varphi) = I(\xi_{\mathscr{L},\tau} \cup \varphi) \in C^{r+1}(G, E)$, i.e.,

$$\kappa_{\mathscr{L},\tau}(\varphi)(g_0, \ldots, g_{r+1}) = \int_X \log_{\mathscr{L}}\Big(\frac{z-g_1\tau}{z-g_0\tau}\Big)\varphi(g_1, \ldots, g_{r+1}). \tag{6-4}$$

One may compute directly that

$$d\kappa_{\mathscr{L},\tau}(\varphi) = \kappa_{\mathscr{L},\tau}(d\varphi). \tag{6-5}$$

Therefore, the correspondence $\varphi \mapsto \kappa_{\mathscr{L},\tau}(\varphi)$ induces a map

$$\kappa_{\mathscr{L}} : H^r(G, M^0(X)) \to H^{r+1}(G, E),$$

which, as our notation suggests, does not depend on the choice of τ. Define

$$H^1(\Gamma_0, E)_{p\text{-new}} := H^1(\Gamma_0, E) \Big/ \text{Image}\left(H^1(\Gamma, E) \oplus H^1(\Gamma', E) \to H^1(\Gamma_0, E)\right),$$

and let

$$\delta : H^1(\Gamma_0, E)_{p\text{-new}} \hookrightarrow H^2(\Theta, E) \tag{6-6}$$

be the injective map induced by the connecting homomorphism in the Mayer–Vietoris sequence (6-2).

Proposition 14. *The cohomology class $\varphi_g^{\pm}$ defined in Proposition 13 satisfies the following*:

(1) *The identity $\kappa_\infty(\varphi_g^{\pm}) = \delta(g^{\pm})$ holds in $H^2(\Theta, E)$.*

(2) *There is a unique $\mathscr{L} \in E$, denoted $-\mathscr{L}^D(g^{\pm})$, such that $\kappa_{\mathscr{L}}(\varphi_g^{\pm}) = 0$.*

Proof. The first statement is argued in the proof of [Greenberg 2009, Lemma 32]. By [ibid., Lemmas 32 and 33], the eigenspace of $H^2(\Theta, E)^{\pm}$ on which the Hecke operators away from p act via the eigenvalues of g is 1-dimensional and is spanned by $\kappa_\infty(\varphi_g^{\pm}) = \delta(g^{\pm})$, where δ is as in (6-6). The class $\delta(g^{\pm})$ is nonzero as $g^{\pm}$ is a nonzero p-new form and δ is injective on such classes. Since the map κ_0 (the one corresponding to $\mathscr{L} = 0$) is Hecke-equivariant, there is a unique constant $\mathscr{L}^{\mathrm{D}}(g^{\pm}) \in E$ such that $\kappa_0(\varphi_g^{\pm}) = \mathscr{L}^{\mathrm{D}}(g^{\pm})\kappa_\infty(\varphi_g^{\pm})$. But the identity $\log_{\mathscr{L}} = \log_0 + \mathscr{L}\,\mathrm{ord}_p$ implies that $\kappa_{\mathscr{L}} = \kappa_0 + \mathscr{L}\kappa_\infty$, and the second statement of the proposition follows with $\mathscr{L} = -\mathscr{L}^{\mathrm{D}}(g^{\pm})$. □

Definition 15. The quantity $\mathscr{L}^{\mathrm{D}}(g^{\pm})$ is called the *Darmon $\mathscr{L}$-invariant of $g^{\pm}$*.

7. Equality of the Greenberg–Stevens and Darmon $\mathscr{L}$-invariants

Let $\mathscr{L} \in E$. The goal of this section is to prove the following:

Theorem 16. *We have*

$$\kappa_{\mathscr{L}}(\varphi_g^{\pm}) = (\mathscr{L}^{GS}(g) + \mathscr{L})\delta(g^{\pm})$$

in $H^2(\Theta, E)$. *Therefore*, $\mathscr{L}^D(g^{\pm}) = \mathscr{L}^{GS}(g)$.

Since the Riemann surfaces $\Gamma\backslash\mathscr{H}$ and $\Gamma'\backslash\mathscr{H}$ are compact if and only if $N^- \neq 1$, we have

$$H^2(\Gamma, E) \cong \begin{cases} E & \text{if } N^- \neq 1, \\ \{0\} & \text{if } N^- = 1. \end{cases}$$

In either case, this space is Eisenstein for the Hecke operators. Since the restriction maps are Hecke-equivariant, $\mathrm{res}^{\Theta}_{\Gamma}\, \kappa_{\mathscr{L}}(\varphi_g^{\pm}) = 0$ and $\mathrm{res}^{\Theta}_{\Gamma'}\, \kappa_{\mathscr{L}}(\varphi_g^{\pm}) = 0$.

Fix a base point $\tau \in \mathscr{H}_p(E)$ and a representative $\varphi \in C^1(\Theta, M^0(X))$ for the cohomology class $\varphi_g^{\pm} \in H^1(\Theta, M^0(X))$. Let $\psi \in C^1(\Gamma, E)$ and $\psi' \in C^1(\Gamma', E)$ be 1-cochains such that $d\psi = \kappa_{\mathscr{L},\tau}(\varphi)|_{\Gamma}$ and $d\psi' = \kappa_{\mathscr{L},\tau}(\varphi)|_{\Gamma'}$. Then $\psi - \psi'$ is a 1-cocycle on $\Gamma_0 = \Gamma \cap \Gamma'$ and, tracing through the construction of the connecting homomorphism in the long exact sequence in cohomology associated to (6-6), one finds that

$$\delta([\psi - \psi']) = \kappa_{\mathscr{L}}(\varphi_g^{\pm}) \tag{7-1}$$

in $H^2(\Theta, E)$. Through a general cohomological calculation, we will find explicit formulas for ψ and ψ' and show that

$$[\psi - \psi'] = (\mathscr{L}^{\mathrm{GS}}(g) + \mathscr{L})g^{\pm}. \tag{7-2}$$

Equations (7-1) and (7-2) prove Theorem 16.

Let $\varphi \in C^1(\Theta, M^0(X))$ be a cocycle representing the class $\varphi_g^{\pm}$. Let

$$\Phi = \Phi_g^{\pm} \in H^1(\Gamma, M^0(\mathbb{X}))$$

denote the Hida family defined in (4-6) that lifts $\mathrm{res}^{\Theta}_{\Gamma}[\varphi]$ with respect to the push-forward map $\pi_* : M^0(\mathbb{X}) \to M^0(X)$. Let $\widetilde{\varphi}_0 \in C^1(\Gamma, M^0(\mathbb{X}))$ be a cocycle representing Φ. Then there exists a cochain $m \in Z^0(\Gamma, M^0(X))$ such that $\pi_*\widetilde{\varphi}_0 = \varphi + dm$. Since $F_0 = \mathbb{Z}[\widetilde{\Theta}]$ is Θ-projective and thus Γ-projective, we may lift m to a cochain $\widetilde{m} \in C^0(\Gamma, M^0(\mathbb{X}))$. Setting $\widetilde{\varphi} = \widetilde{\varphi}_0 - d\widetilde{m} \in C^1(\Gamma, M^0(\mathbb{X}))$, we obtain a cocycle representing Φ that satisfies

$$\pi_*\widetilde{\varphi} = \varphi. \tag{7-3}$$

For any $\sigma \in C^r(\Gamma, M^0(\mathbb{X}))$ and $\sigma' \in C^r(\Gamma', M^0(w_p\mathbb{X}))$, define $\lambda_{\mathscr{L}}(\sigma) \in C^r(\Gamma, E)$ and $\lambda'_{\mathscr{L}}(\sigma') \in C^r(\Gamma', E)$ by the formulas

$$\lambda_{\mathscr{L}}(\sigma)(g_0, g_1, \ldots, g_r) = \int_{\mathbb{X}} \log_{\mathscr{L}}(x - (g_0\tau)y)\, \sigma(g_0, g_1, \ldots, g_r)(x, y), \quad (7\text{-}4)$$

$$\lambda'_{\mathscr{L}}(\sigma')(g_0, g_1, \ldots, g_r) = \int_{w_p\mathbb{X}} \log_{\mathscr{L}}(x - (g_0\tau)y)\, \sigma'(g_0, g_1, \ldots, g_r)(x, y).$$

These maps are Γ and Γ'-invariant, respectively, because the values of σ and σ' have total measure zero.

Lemma 17. *For any $\sigma \in C^r(\Gamma, M^0(\mathbb{X}))$ and $\sigma' \in C^r(\Gamma', M^0(w_p\mathbb{X}))$, we have*

$$d\lambda_{\mathscr{L}}(\sigma) = \kappa_{\mathscr{L}}(\pi_*\sigma) + \lambda_{\mathscr{L}}(d\sigma), \qquad d\lambda'_{\mathscr{L}}(\sigma') = \kappa_{\mathscr{L}}(\pi_*\sigma') + \lambda'_{\mathscr{L}}(d\sigma').$$

Proof. Letting $h = (g_0, \ldots, g_{r+1})$ and $h_i = (g_0, \ldots, \widehat{g_i}, \ldots, g_r)$, we have

$$\begin{aligned} d\lambda(\sigma)(h) &= \int_{\mathbb{X}} \log_{\mathscr{L}}(x - (g_1\tau)y)\sigma(h_0)(x, y) \\ &\qquad + \sum_{i=1}^{r+1} (-1)^i \int_{\mathbb{X}} \log_{\mathscr{L}}(x - (g_0\tau)y)\sigma(h_i)(x, y) \\ &= \int_{\mathbb{X}} \log_{\mathscr{L}}\Big(\frac{x-(g_1\tau)y}{x-(g_0\tau)y}\Big)\sigma(h_0)(x, y) + \int_{\mathbb{X}} \log_{\mathscr{L}}(x - (g_0\tau)y)\, d\sigma(h)(x, y) \\ &= \int_X \log_{\mathscr{L}}\Big(\frac{z-g_1\tau}{z-g_0\tau}\Big)\pi_*\sigma(h_0)(z) + \lambda_{\mathscr{L}}(d\sigma)(h) \\ &= \kappa_{\mathscr{L}}(\pi_*\sigma)(h) + \lambda_{\mathscr{L}}(d\sigma)(h), \end{aligned}$$

as desired. The second equality is proved in a similar manner. □

Lemma 17 implies that if we define

$$\psi = \lambda_{\mathscr{L}}(\widetilde{\varphi}) \in C^1(\Gamma, E), \quad (7\text{-}5)$$

then $d\psi = \kappa_{\mathscr{L}}(\varphi)$. Similarly, define

$$\psi' = \lambda'_{\mathscr{L}}(W_pU_p\widetilde{\varphi}) \in C^1(\Gamma', E). \quad (7\text{-}6)$$

Then

$$\begin{aligned} d\psi' &= \kappa_{\mathscr{L}}(\pi_*W_pU_p\widetilde{\varphi}) + d\lambda'_{\mathscr{L}}(dW_pU_p\widetilde{\varphi}) \\ &= \kappa_{\mathscr{L}}(W_pU_p\varphi) + 0 \\ &= \kappa_{\mathscr{L}}(\varphi), \end{aligned}$$

where the last equality is justified by the following lemma:

Lemma 18. *We have the identity of Θ-cochains $W_pU_p\varphi = \varphi$.*

Proof. Consider the diagram

$$\begin{array}{ccccc} & & C^r(\Gamma, M(X)) & & \\ & \swarrow^{\rho_{X_\infty}} & & \searrow^{\rho_{\mathbb{Z}_p}} & \\ C^r(\Gamma_0, M(X_\infty)) & \xrightarrow[U_p]{} & C^r(\Gamma_0, M(X_\infty)) & \xrightarrow[W_p]{} & C^r(\Gamma_0, M(\mathbb{Z}_p)) \\ \downarrow{\scriptstyle \rho_{X_\infty}^{-1}} & & \downarrow{\scriptstyle \rho_{X_\infty}^{-1}} & & \downarrow{\scriptstyle \rho_{\mathbb{Z}_p}^{-1}} \\ C^r(\Gamma, M(X)) & \xrightarrow[U_p]{} & C^r(\Gamma, M(X)) & \xrightarrow[W_p]{} & C^r(\Gamma', M(X)) \end{array}$$

The maps ρ_{X_∞} and $\rho_{\mathbb{Z}_p}$ are isomorphisms by Shapiro's lemma. The bottom squares of the diagram commute by definition and the upper triangle commutes as it is the pushforward via π_* in (5-4). The lemma follows. □

Having found explicit formulas for ψ and ψ' in (7-5) and (7-6), respectively, we now turn towards proving (7-2). Recall that $\Phi = [\widetilde{\varphi}]$ is a U_p-eigenvector with eigenvalue $\boldsymbol{a}_p(\Phi)$ satisfying $\epsilon(\boldsymbol{a}_p(\Phi)) = \pm 1$. We defined

$$\mathscr{L}^{\mathrm{GS}}(\Phi) = d_\epsilon(1 - \boldsymbol{a}_p(\Phi)^2).$$

Proposition 19. *The class of the cocycle $\psi - \psi'$ in $H^1(\Gamma_0, E)$ is equal to*

$$(\mathscr{L}^{GS}(\Phi) + \mathscr{L})\rho_*[\varphi],$$

where $\rho_* : H^1(\Theta, M^0(X)) \to H^1(\Gamma_0, M(X_\infty)) \to H^1(\Gamma_0, E)$ is the composition of the canonical restriction map ρ_{X_∞} with the total measure on X_∞ map (as in (4-3)).

Proof. We use the decompositions $\mathbb{X} = \mathbb{X}_\infty \sqcup \mathbb{X}_p$ and $w_p\mathbb{X} = \mathbb{X}_\infty \sqcup p\mathbb{X}_p$ to study the integrals defining ψ and ψ' (see (4-1) and (5-3)). Writing $h = (g_0, g_1)$, we find:

$$\begin{aligned}(\psi - \psi')(h) = \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y)\widetilde{\varphi}(h) + \int_{\mathbb{X}_p} \log_{\mathscr{L}}(x - (g_0\tau)y)\widetilde{\varphi}(h) \\ - \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y) W_p U_p \widetilde{\varphi}(h) \\ - \int_{p\mathbb{X}_p} \log_{\mathscr{L}}(x - (g_0\tau)y) W_p U_p \widetilde{\varphi}(h). \quad (7\text{-}7)\end{aligned}$$

Propositions 11 and 12 allow us to rewrite these last two integrals as

$$\int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y) W_p U_p \widetilde{\varphi}(h) = \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y) U_p^2 \widetilde{\varphi}(h) \qquad (7\text{-}8)$$

and

$$\int_{p\mathbb{X}_p} \log_{\mathscr{L}}(x-(g_0\tau)y)W_pU_p\widetilde{\varphi}(h)$$
$$= \int_{p\mathbb{X}_p} \log_{\mathscr{L}}(x-(g_0\tau)y)p_*\widetilde{\varphi}(h) = \int_{\mathbb{X}_p} \log_{\mathscr{L}}(px-(g_0\tau)py)\widetilde{\varphi}(h)$$
$$= \int_{\mathbb{X}_p} \log_{\mathscr{L}}(x-(g_0\tau)y)\widetilde{\varphi}(h) + \mathscr{L}\widetilde{\varphi}(h)(\mathbb{X}_p). \tag{7-9}$$

Combining (7-7), (7-8), and (7-9), we obtain

$$(\psi-\psi')(h) = \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x-(g_0\tau)y)(1-U_p^2)\widetilde{\varphi}(h) - \mathscr{L}\widetilde{\varphi}(h)(\mathbb{X}_p). \tag{7-10}$$

We now view $\widetilde{\varphi}$ as an element of $Z^r(\Gamma_0, M^0(\mathbb{X}_\infty))$ and calculate the class in $H^r(\Gamma_0, E)$ represented by the right side of (7-10). We have that

$$\widetilde{\varphi}(h)(\mathbb{X}_p) = \varphi(h)(\mathbb{Z}_p) = -\varphi(h)(X_\infty),$$

and hence represents the class $-\rho_*[\varphi]$ in $H^r(\Gamma_0, E)$. Therefore the last term in (7-10) represents the class $\mathscr{L}\rho_*[\varphi]$.

It remains to prove that the first term in (7-10) represents the class $\mathscr{L}^{\mathrm{GS}}(\widetilde{\varphi})\rho_*[\varphi]$ in $H^1(\Gamma_0, E)$. Since $(1-U_p^2)\Phi = \alpha\Phi$ with $\alpha = 1-\boldsymbol{a}_p(\Phi)^2$, we may write

$$(1-U_p^2)\widetilde{\varphi} = \alpha\widetilde{\varphi} + d\nu \tag{7-11}$$

for some $\nu \in C^0(\Gamma_0, M(\mathbb{X}_\infty))$. Pushing forward via π_*, we obtain

$$(1-U_p^2)\varphi = 0 + \pi_*(d\nu).$$

Since the term on the left is zero, we obtain $d\pi_*(\nu) = 0$. Thus $\pi_*\nu$ represents a class in $H^0(\Gamma_0, M(X_\infty))$.

Lemma 20. *The cohomology group $H^0(\Gamma_0, M(X_\infty))$ is zero.*

Proof. It is easy to see that

$$\mathfrak{I}_p = \{g \in \mathrm{GL}_2(\mathbb{Z}_p) : g \text{ is upper-triangular modulo } p\}$$

acts transitively on the set of balls in X_∞ of radius p^{-n} for any $n \geq 1$. Since Γ_0 is p-adically dense in $\mathfrak{I}_p$, Γ_0 acts transitively on this set as well. It follows that if μ is a Γ_0-invariant measure on X_∞, then $\mu(B) = p^{-n+1}\mu(X_\infty)$ for all compact-open balls $B \subset X_\infty$ of radius p^{-n}. Since the values of μ are assumed to be p-adically bounded, it follows that $\mu = 0$. □

By the lemma, we conclude that $\pi_*\nu$ is a coboundary. Arguing above as in the definition of the cocycle $\widetilde{\varphi}$ satisfying (7-3), we may alter ν by a coboundary to assume that $\pi_*\nu = 0$.

We may now calculate the cohomology class represented by (7-10). Substituting (7-11) into (7-10), the term from $\alpha\widetilde{\varphi}$ yields

$$\int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y)\alpha\widetilde{\varphi}(h). \tag{7-12}$$

By Proposition 21 below, the expression in (7-12) represents the class $\mathscr{L}^{\mathrm{GS}}(\widetilde{\varphi})\rho_*[\varphi]$ in $H^1(\Gamma_0, E)$. It remains to prove that the term arising from $d\nu$ is trivial in cohomology, i.e., that

$$h \mapsto \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y)d\nu(h) \tag{7-13}$$

is a coboundary. Note that the right side of (7-13) is equal to

$$\int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x)d\nu(h) + \int_{X_\infty} \log_{\mathscr{L}}(1 - (g_0\tau)/z)\pi_* d\nu(h). \tag{7-14}$$

The last term of (7-14) is zero since $\pi_* d\nu = 0$. The first term of (7-14) is equal to the coboundary of the 0-cochain given by

$$g_0 \mapsto \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x)\nu(g_0). \tag{7-15}$$

We leave to the reader the exercise of using the equation $\pi_*\nu = 0$ to show that the 0-cochain in (7-15) is Γ_0-invariant. This proves that (7-13) is a coboundary and completes the proof of the proposition. □

The following proposition, applied with $\alpha = 1 - \boldsymbol{a}_p(\Phi)^2$, was used above to extract the invariant $\mathscr{L}^{\mathrm{GS}}(\Phi)$ from the cohomology class $[\Phi]$.

Proposition 21. *Let $\sigma \in Z^r(\Gamma_0, M(\mathbb{X}_\infty))$, let $\alpha \in I_\epsilon \subset \Lambda$ and define*

$$\eta(g_0, \ldots, g_r) = \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y)\alpha\sigma(g_0, \ldots, g_r).$$

Then $\eta \in Z^r(\Gamma_0, E)$ and represents the class

$$[\eta] = d_\epsilon(\alpha)\rho_*[\sigma] \in H^r(\Gamma_0, E).$$

Proof. Since $\alpha \in I_\epsilon$, we have $\pi_*(\alpha\sigma) = 0$; in particular, $\alpha\sigma$ has total measure 0. It follows from this fact and a routine calculation that η is a cochain. That η is a cocycle follows from the equations $d(\alpha\sigma) = \alpha\, d\sigma = 0$.

To evaluate the class $[\eta] \in H^r(\Gamma_0, E)$, we consider α of the form $[\ell] - 1$ for $\ell \in 1 + p\mathbb{Z}_p$. Writing $h = (g_0, \ldots, g_r)$, we have

$$\begin{aligned}\eta(h) &= \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(x - (g_0\tau)y)([\ell]\sigma - \sigma)(h) \\ &= \int_{\mathbb{X}_\infty} \big(\log_{\mathscr{L}}(\ell x - (g_0\tau)\ell y) - \log_{\mathscr{L}}(x - (g_0\tau)y)\big)\sigma(h) \\ &= \int_{\mathbb{X}_\infty} \log_{\mathscr{L}}(\ell)\sigma(h) = \log(\ell) \cdot \sigma(h)(\mathbb{X}_\infty) = d_\epsilon([\ell] - 1)\rho_*\sigma(h).\end{aligned}$$

This proves the result for $\alpha = [\ell] - 1$, and hence gives the result for general $\alpha \in I_\epsilon$ as the ideal I_ϵ is generated over Λ by such elements. □

This concludes the proof of Proposition 19, and since $\rho_*\varphi_g^\pm = g^\pm$, we deduce (7-2) and hence Theorem 16. Combining with Theorem 8, we also complete the proof of Theorem 2.

8. Multiplicative integrals and period lattices

In this section, we suppose that the Hecke eigenvalues of g belong to $\mathbb{Z}$. In this case, it is shown in [Greenberg 2009, §8] that we may take

$$\varphi_g^\pm \in H^1(\Theta, M^0(X, \mathbb{Z}))^{g,\pm}.$$

That is, we may find an element $\varphi_g^\pm \in H^1(\Theta, M^0(X, \mathbb{Z}))^{g,\pm}$ whose image in $H^1(\Theta, M^0(X, E))$ is a basis for $H^1(\Theta, M^0(X, E))^{g,\pm}$. Using this integral cohomology class, we may define multiplicative versions of many of the objects considered in previous sections.

Following Darmon [2001], we consider the *multiplicative integration pairing*

$$C(X)^\times/E^\times \times M^0(X, \mathbb{Z}) \to E^\times, \quad (f, \mu) \mapsto \oint_X f\mu \tag{8-1}$$

defined by

$$\oint_X f\mu = \lim_{||\mathcal{U}||\to 0} \prod_{U\in\mathcal{U}} f(z_U)^{\mu(U)}.$$

Here, $\mathcal{U}$ is a finite cover of X by compact open sets and z_U is an arbitrary point of U. The limit is taken over uniformly finer covers $\mathcal{U}$. It is clear that

$$\log_{\mathscr{L}} \oint f\mu = \int \log_{\mathscr{L}}(f)\mu \quad \text{for any } \mathscr{L}.$$

The pairing (8-1) is easily seen to be $\mathrm{GL}_2(\mathbb{Q}_p)$-equivariant and thus induces a corresponding pairing

$$\langle \cdot, \cdot \rangle^\times : H_1(\Theta, C(X)^\times/E^\times) \times H^1(\Theta, M^0(X, \mathbb{Z})) \to E^\times. \tag{8-2}$$

Let $\Delta = \mathrm{Div}\,\mathscr{H}_p$ and let $\Delta^0 = \mathrm{Div}^0\,\mathscr{H}_p$. From the long exact sequence associated to the short exact sequence of $\mathrm{GL}_2(\mathbb{Q}_p)$-modules $0 \to \Delta^0 \to \Delta \to \mathbb{Z} \to 0$, we extract a connecting homomorphism $\partial : H_2(\Theta, \mathbb{Z}) \to H_1(\Theta, \Delta^0)$. Let $j : \Delta^0 \to C(X)^\times / E^\times$ be the map sending a divisor D to a rational function on X with divisor D. (Note that such a function is only well-defined up to multiplication by a nonzero scalar.) The map j being $\mathrm{GL}_2(\mathbb{Q}_p)$-equivariant, it induces a corresponding map

$$j_* : H_1(\Theta, \Delta^0) \to H_1(\Theta, C(X)^\times / E^\times).$$

We may also define multiplicative refinements of the cocycles $\kappa_{\mathscr{L},\tau}(\varphi)$ as follows. Let $\tau \in \mathscr{H}_p$, let $\varphi \in C^r(\Theta, M^0(X, \mathbb{Z}))$, and define $\kappa_\tau(\varphi) \in C^{r+1}(\Theta, E^\times)$ by the rule

$$\kappa_\tau(\varphi)(g_0, \ldots, g_{r+1}) = \int\!\!\!\!\!\!\diagup_X \left(\frac{z - g_1\tau}{z - g_0\tau}\right) \varphi(g_1, \ldots, g_{r+1}) \in E^\times.$$

As with $\kappa_{\mathscr{L},\tau}$, the homomorphism κ_τ induces a map

$$\kappa : H^r(\Theta, M^0(X, \mathbb{Z})) \to H^{r+1}(\Theta, E^\times)$$

that does not depend on τ.

By the universal coefficients theorem, there is a natural surjective map

$$H^{r+1}(\Theta, E^\times) \to \mathrm{Hom}(H_{r+1}(\Theta, \mathbb{Z}), E^\times).$$

Lemma 22. *The image of* $\kappa(\varphi_g^\pm)$ *in* $\mathrm{Hom}(H_2(\Theta, \mathbb{Z}), E^\times)$ *is given by*

$$\xi \mapsto \langle -j_*\partial\xi, \varphi_g^\pm \rangle^\times.$$

Proof. Suppose

$$\xi = \sum_i 1 \otimes (\gamma_i, \delta_i, \epsilon_i) \in Z_2(\Theta, \mathbb{Z}) = \mathbb{Z} \otimes_\Theta \mathbb{Z}[\Theta^3]$$

is a 2-cycle on Θ with values in $\mathbb{Z}$. Tracing through the construction of the connecting homomorphism, one computes that $\partial[\xi]$ is represented by the cycle

$$\sum_i (\gamma_i\tau - \delta_i\tau) \otimes (\delta_i, \epsilon_i).$$

Therefore,

$$\langle j_*\partial\xi, \varphi_g^\pm \rangle^\times = \prod_i \int\!\!\!\!\!\!\diagup_X \left(\frac{z - \gamma_i\tau}{z - \delta_i\tau}\right) \varphi_g^\pm(\delta_i, \epsilon_i).$$

By the definition of the map in the universal coefficients theorem, the image of $\kappa(\varphi_g^\pm)$ in $\mathrm{Hom}(H_2(\Theta, \mathbb{Z}), E^\times)$ sends ξ to

$$\prod_i \kappa(\varphi_g^\pm)(\gamma_i, \delta_i, \epsilon_i) = \prod_i \int\!\!\!\!\!\!\diagup_X \left(\frac{z - \delta_i\tau}{z - \gamma_i\tau}\right) \varphi_g^\pm(\delta_i, \epsilon_i).$$

The result follows. □

In view of Lemma 22, we set

$$L_g^{\pm} = \langle j_* \partial H_2(\Theta, \mathbb{Z}), \varphi_g^{\pm} \rangle^{\times} = \langle H_2(\Theta, \mathbb{Z}), \kappa(\varphi_g^{\pm}) \rangle \subset E^{\times}.$$

Proposition 23 [Greenberg 2009, Proposition 30]. *$L_g^{\pm}$ is a lattice in $E^{\times}$.*

Therefore, there is a unique $\mathscr{L} \in E$ such that $\log_{\mathscr{L}}(L_g^{\pm}) = 0$. We define the *$\mathscr{L}$-invariant of the lattice $L_g^{\pm}$*, denoted $\mathscr{L}(L_g^{\pm})$, to be the negative of this constant $\mathscr{L}$.

Proposition 24. *The $\mathscr{L}$-invariant of the lattice $L_g^{\pm}$ is equal to $\mathscr{L}^D(\varphi_g^{\pm})$.*

Proof. By the universal coefficients theorem,

$$\begin{aligned} \log_{\mathscr{L}}(L_g^{\pm}) &= \log_{\mathscr{L}} \langle H_2(\Theta, \mathbb{Z}), \kappa(\varphi_g^{\pm}) \rangle \\ &= \langle H_2(\Theta, \mathbb{Z}), \kappa_{\mathscr{L}}(\varphi_g^{\pm}) \rangle \end{aligned}$$

is equal to 0 if and only if $\kappa_{\mathscr{L}}(\varphi_g^{\pm}) = 0$. By definition, this occurs if and only if $\mathscr{L} = -\mathscr{L}^{\mathrm{D}}(\varphi_g^{\pm})$. □

Corollary 25 [Greenberg 2009, Conjecture 2]. *Let q be the Tate period of the elliptic curve $\mathscr{E}/\mathbb{Q}$ associated to f. Then*

$$\mathscr{L}(L_g^{\pm}) = \log_p(q)/\operatorname{ord}_p(q).$$

Proof. By Proposition 23 and Theorem 2, $\mathscr{L}(L_g^{\pm}) = \mathscr{L}^{\mathrm{D}}(\varphi_g^{\pm}) = \mathscr{L}^{\mathrm{GS}}(f)$. By the Galois-theoretic portion of the proof of the Greenberg–Stevens theorem [Greenberg and Stevens 1993, Theorem 3.18], we have $\mathscr{L}^{\mathrm{GS}}(f) = \log_p(q)/\operatorname{ord}_p(q)$. □

In [Greenberg 2009], a construction was given for local *Stark–Heegner points* on $E^{\times}/L_g^{\pm}$. We conjectured that the elliptic curve $E^{\times}/L_g^{\pm}$ is isogenous to $\mathscr{E}/E$, yielding a construction of local points on $\mathscr{E}$. Corollary 25 proves this conjecture and makes the construction unconditional. In the following section, we apply the above techniques further to obtain a formula for the formal group logarithms of these Stark–Heegner points in terms of Hida families.

9. Abel–Jacobi maps and Stark–Heegner points

In this section we recall the definition of Stark–Heegner points and give a formula for the formal group logarithms of these points in terms of Hida families. This formula will be used in [Greenberg and Shahabi ≥ 2012] to prove partial results towards the rationality of the Stark–Heegner points following the methods of [Bertolini and Darmon 2009].

Let $\mathscr{H}_{p,\mathrm{ur}}$ denote the *unramified p-adic upper half-plane*:

$$\mathscr{H}_{p,\mathrm{ur}} = \mathbb{P}^1(\mathbb{C}_p) - r^{-1}(\mathbb{P}^1(\mathbb{F}_p)) \subset \mathscr{H}_p,$$

where $r : \mathbb{P}^1(\mathbb{C}_p) \to \mathbb{P}^1(\overline{\mathbb{F}}_p)$ is the reduction map. The action of $\mathrm{GL}_2(\mathbb{Z}_p)$ on $\mathscr{H}_p$ preserves $\mathscr{H}_{p,\mathrm{ur}}$. We set $\Delta_{\mathrm{ur}} = \mathrm{Div}\,\mathscr{H}_{p,\mathrm{ur}}$ and $\Delta^0_{\mathrm{ur}} = \mathrm{Div}^0\,\mathscr{H}_{p,\mathrm{ur}}$. If $\tau_1, \tau_2 \in \mathscr{H}_{p,\mathrm{ur}}$, $z \in X$ and $(x, y) \in \mathbb{X}$, then the quantities

$$\log_{\mathscr{L}}\left(\frac{z-\tau}{z-\tau'}\right), \quad \log_{\mathscr{L}}(x - y\tau)$$

do not depend on $\mathscr{L}$ because the arguments are p-adic units. For this reason, we do not specify a branch of the p-adic logarithm and simply write log. The natural $\mathrm{GL}_2(\mathbb{Q}_p)$-equivariant pairing

$$\langle \cdot, \cdot \rangle : M^0(X) \times C(X)/E \to \mathbb{C}_p$$

induces a pairing

$$\langle \cdot, \cdot \rangle : H^1(\Gamma, M^0(X)) \times H_1(\Gamma, C(X)/E) \to \mathbb{C}_p. \tag{9-1}$$

Define $j : \Delta^0_{\mathrm{ur}} \to C(X)/E$ by

$$j(\{\tau_2\} - \{\tau_1\})(z) = \log\left(\frac{z-\tau_2}{z-\tau_1}\right).$$

Since it is Γ-equivariant, j induces a homomorphism

$$j_* : H_1(\Gamma, \Delta^0_{\mathrm{ur}}) \to H_1(\Gamma, C(X)/E).$$

We define one more pairing

$$\langle \cdot, \cdot \rangle : H^1(\Gamma, M^0(X)) \times H_1(\Gamma, \Delta^0_{\mathrm{ur}}) \to \mathbb{C}_p$$

by $\langle \varphi, \xi \rangle = \langle \varphi, j_*\xi \rangle$.

Let $\mathbb{T}^{(p)}$ be the Hecke generated by the operators away from p, that is, the operators T_ℓ for $\ell \nmid pN$, U_ℓ for $\ell \mid N^+$, and the involutions W_ℓ for $\ell \mid N^-$ (see §3). There is a natural action of $\mathbb{T}^{(p)}$ on $H_1(\Gamma, \Delta^0_{\mathrm{ur}})$ described by double cosets such that, endowing $\mathrm{Hom}(H_1(\Gamma, \Delta^0_{\mathrm{ur}}), E)$ with the corresponding dual action, the map

$$\begin{aligned} A : H^1(\Gamma, M^0(X)) &\to \mathrm{Hom}(H_1(\Gamma, \Delta^0_{\mathrm{ur}}), E), \\ \varphi &\mapsto \big(\xi \mapsto \langle \varphi, \xi \rangle\big) \end{aligned}$$

induced by the pairing (9-1) is $\mathbb{T}^{(p)}$-equivariant. For g as in the previous sections, define

$$A^{\pm}_g = A(\mathrm{Res}^{\Theta}_{\Gamma}\, \varphi^{\pm}_g).$$

We have $A^{\pm}_g \in \mathrm{Hom}(H_1(\Gamma, \Delta^0_{\mathrm{ur}}), E)^{g,\pm}$, where $\mathrm{Hom}(H_1(\Gamma, \Delta^0_{\mathrm{ur}}), E)^{g,\pm}$ is the eigenspace on which $\mathbb{T}^{(p)}$ acts via the Hecke eigenvalues of g and W_∞ acts as ± 1.

Proposition 26. *There is a unique homomorphism* $\mathrm{AJ}_g^{\pm} \in \mathrm{Hom}(H_1(\Gamma, \Delta_{ur}), E)^{g,\pm}$ *such that the diagram*

$$\begin{array}{ccc} H_1(\Gamma, \Delta_{ur}^0) & \longrightarrow & H_1(\Gamma, \Delta_{ur}) \\ & {\scriptstyle A_g^{\pm}} \searrow \quad \swarrow {\scriptstyle \mathrm{AJ}_g^{\pm}} & \\ & E & \end{array}$$

commutes, where the horizontal map is induced by the inclusion $\Delta_{ur}^0 \hookrightarrow \Delta_{ur}$.

The proof of Proposition 26 is given in [Greenberg 2009, §10] and is very similar to the first half of the proof of Lemma 9.

Remark 27. We have chosen the notation $\mathrm{AJ}_g^{\pm}$ for this map because it formally resembles an Abel–Jacobi map.

Define $J : \Delta_{\mathrm{ur}} \to C(\mathbb{X})/E$ by $J(\{\tau\})(x, y) = \log(x - y\tau)$. Since it is Γ-equivariant, J induces a homomorphism $J_* : H_1(\Gamma, \Delta) \to H_1(\Gamma, C(\mathbb{X})/E)$. The natural Γ-equivariant pairing $M^0(\mathbb{X}) \times C(\mathbb{X})/E \to E$ induces a corresponding pairing $H^1(\Gamma, M^0(\mathbb{X})) \times H_1(\Gamma, C(\mathbb{X})/E) \to E$.

Corollary 28. *The map* $\mathrm{AJ}_g^{\pm} : H_1(\Gamma, \Delta_{ur}) \to E$ *is given by* $\mathrm{AJ}_g^{\pm}(\xi) = \langle \Phi_g^{\pm}, J_*\xi \rangle$.

Proof. It is easy to see that the element $\widetilde{\mathrm{AJ}}_g^{\pm}$ of $\mathrm{Hom}(H_1(\Gamma, \Delta_{\mathrm{ur}}), E)$ defined by $\xi \mapsto \langle \Phi_g^{\pm}, J_*\xi \rangle$ belongs to the $(g, \pm)$-eigenspace. Since $\pi_* \Phi_g^{\pm} = \mathrm{Res}_{\Gamma}^{\Theta} \varphi_g^{\pm}$, the diagram

$$\begin{array}{ccccc} H_1(\Gamma, \Delta_{\mathrm{ur}}^0) & \xrightarrow{j_*} & H_1(\Gamma, C(X)/E) & & \\ \downarrow & & \downarrow {\scriptstyle \pi^*} & \searrow {\scriptstyle \langle \mathrm{Res}_{\Gamma}^{\Theta} \varphi_g^{\pm}, \cdot \rangle} & \\ & & & & E \\ H_1(\Gamma, \Delta_{\mathrm{ur}}) & \xrightarrow[J_*]{} & H_1(\Gamma, C(\mathbb{X})/E) & \nearrow {\scriptstyle \langle \Phi_g^{\pm}, \cdot \rangle} & \end{array}$$

commutes, implying that

$$\begin{array}{ccc} H_1(\Gamma, \Delta_{\mathrm{ur}}^0) & \longrightarrow & H_1(\Gamma, \Delta_{\mathrm{ur}}) \\ & {\scriptstyle A_g^{\pm}} \searrow \quad \swarrow {\scriptstyle \widetilde{\mathrm{AJ}}_g^{\pm}} & \\ & E & \end{array}$$

commutes as well. Therefore, by Proposition 26, $\mathrm{AJ}_g^{\pm} = \widetilde{\mathrm{AJ}}_g^{\pm}$. □

Let K be a real quadratic field and let $\mathcal{O} \subset K$ be an order such that $\operatorname{disc} \mathcal{O}$ is relatively prime to N and p. There is an embedding

$$\psi : K \to B$$

such that $\psi(\mathcal{O}) = \psi(K) \cap R$. For details regarding this point, see [Vignéras 1980, chapitre III, 5C]. Suppose further that p is inert in K. Then $\psi(K^\times)$ acts on $\mathbb{P}^1(E)$ via ι_p with two fixed points τ_ψ and $\bar{\tau}_\psi$ in $\mathcal{H}_{p,\mathrm{ur}}$, conjugate under the action of $\mathrm{Gal}(K_p/\mathbb{Q}_p)$. Let ϵ be a generator of the unit group of $\mathcal{O}$. Then since $\psi(\epsilon)\tau_\psi = \tau_\psi$, we have

$$\{\tau_\psi\} \otimes (1, \psi(\epsilon)) \in Z_1(\Gamma, \Delta_{\mathrm{ur}}).$$

Let $C_{[\psi]}$ be the corresponding class in $H_1(\Gamma, \Delta_{\mathrm{ur}})$. The brackets around ψ indicate that $C_{[\psi]}$ depends only on the Γ-conjugacy class of the embedding ψ. Assuming that the Hecke eigenvalues of g lie in $\mathbb{Z}$, we may associate an elliptic curve $\mathscr{E}/\mathbb{Q}$ to g by the Eichler–Shimura construction. Let $\log_\omega$ be the logarithm of the formal group law on $\mathscr{E}$ associated to the differential dq/q on $E^\times/q^{\mathbb{Z}}$. Note that $\log_\omega$ factorizes as

$$\mathscr{E}(E) \to E^\times/q^{\mathbb{Z}} \to E,$$

where the left arrow is the inverse of the Tate uniformization of $\mathscr{E}$ and the right arrow is $\log_{\mathscr{L}}$ with

$$-\mathscr{L} = \mathscr{L}^{\mathrm{GS}}(g) = \mathscr{L}^{\mathrm{D}}(\varphi_g^\pm) = \mathscr{L}^{\mathrm{MTT}}(g) = \frac{\log_p(q)}{\operatorname{ord}_p(q)}.$$

The points

$$\mathrm{AJ}_g^\pm(C_{[\psi]}) \in E = \log_\omega \mathscr{E}(E)$$

are called *Stark–Heegner points* on $\mathscr{E}$. We conjecture in [Greenberg 2009, §10] that the locally defined points $\mathrm{AJ}_g^\pm(C_{[\psi]})$ in fact belong to $\log_{\mathscr{E}}(\mathscr{E}(H_{\mathcal{O}}))$, where $H_{\mathcal{O}}$ is the ring class field of K associated to the order $\mathcal{O}$. By the results of this section, we have the following formula for $\mathrm{AJ}_g^\pm(C_{[\psi]})$ in terms of the Hida family $\Phi_g^\pm$:

Corollary 29. $$\mathrm{AJ}_g^\pm(C_{[\psi]}) = \langle \Phi_g^\pm, J_* C_{[\psi]} \rangle.$$

In [Greenberg and Shahabi ≥ 2012], we apply this formula with the methods of [Bertolini and Darmon 2009] to prove partial results towards the rationality of the Stark–Heegner points $\mathrm{AJ}_g^\pm(C_{[\psi]})$ over $H_{\mathcal{O}}$.

References

[Balasubramanyam and Longo 2011] B. Balasubramanyam and M. Longo, "Λ-adic modular symbols over totally real fields", *Comment. Math. Helv.* **86**:4 (2011), 841–865. MR 2851871 Zbl 0596 5407

[Bertolini and Darmon 2009] M. Bertolini and H. Darmon, "The rationality of Stark–Heegner points over genus fields of real quadratic fields", *Ann. of Math.* (2) **170**:1 (2009), 343–370. MR 2010m:11072 Zbl 1203.11045

[Bertolini et al. 2010] M. Bertolini, H. Darmon, and A. Iovita, "Families of automorphic forms on definite quaternion algebras and Teitelbaum's conjecture", pp. 29–64 Astérisque **331**, Société Mathématique de France, Paris, 2010. MR 2012c:11109 Zbl 05765039

[Chenevier 2005] G. Chenevier, "Une correspondance de Jacquet–Langlands p-adique", *Duke Math. J.* **126**:1 (2005), 161–194. MR 2006f:11144 Zbl 1070.11016

[Darmon 2001] H. Darmon, "Integration on $\mathscr{H}_p \times \mathscr{H}$ and arithmetic applications", *Ann. of Math.* (2) **154**:3 (2001), 589–639. MR 2003j:11067 Zbl 1035.11027

[Dasgupta 2005] S. Dasgupta, "Stark–Heegner points on modular Jacobians", *Ann. Sci. École Norm. Sup.* (4) **38**:3 (2005), 427–469. MR 2006e:11080 Zbl 1173.11334

[Greenberg 2009] M. Greenberg, "Stark–Heegner points and the cohomology of quaternionic Shimura varieties", *Duke Math. J.* **147**:3 (2009), 541–575. MR 2010f:11097 Zbl 1183.14030

[Greenberg and Shahabi $\geq$ 2012] M. Greenberg and S. Shahabi, "p-adic L-functions associated to families of modular forms on Shimura curves", in preparation.

[Greenberg and Stevens 1993] R. Greenberg and G. Stevens, "p-adic L-functions and p-adic periods of modular forms", *Invent. Math.* **111**:2 (1993), 407–447. MR 93m:11054 Zbl 0778.11034

[Hida 1986] H. Hida, "Galois representations into $\mathrm{GL}_2(\mathbf{Z}_p[\![X]\!])$ attached to ordinary cusp forms", *Invent. Math.* **85**:3 (1986), 545–613. MR 87k:11049 Zbl 0612.10021

[Matsushima and Shimura 1963] Y. Matsushima and G. Shimura, "On the cohomology groups attached to certain vector valued differential forms on the product of the upper half planes", *Ann. of Math.* (2) **78** (1963), 417–449. MR 27 #5274

[Mazur and Swinnerton-Dyer 1974] B. Mazur and P. Swinnerton-Dyer, "Arithmetic of Weil curves", *Invent. Math.* **25** (1974), 1–61. MR 50 #7152 Zbl 0281.14016

[Mazur et al. 1986] B. Mazur, J. Tate, and J. Teitelbaum, "On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer", *Invent. Math.* **84**:1 (1986), 1–48. MR 87e:11076 Zbl 0699.14028

[Vignéras 1980] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics **800**, Springer, Berlin, 1980. MR 82i:12016 Zbl 0422.12008

Communicated by Richard Taylor
Received 2010-07-15 Revised 2011-04-08 Accepted 2011-05-23

sdasgup2@ucsc.edu *Department of Mathematics, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA 95064, United States*
http://people.ucsc.edu/~sdasgup2/

mgreenbe@math.ucalgary.ca *Department of Mathematics and Statistics, University of Calgary, Calgary, AL T2N 1N4, Canada*

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

On the weak Lefschetz property for powers of linear forms

Juan C. Migliore, Rosa M. Miró-Roig and Uwe Nagel

Ideals generated by prescribed powers of linear forms have attracted a great deal of attention recently. In this paper we study properties that hold when the linear forms are general, in a sense that we make precise. Analogously, one could study so-called "general forms" of the same prescribed degrees. One goal of this paper is to highlight how the differences between these two settings are related to the weak Lefschetz property (WLP) and the strong Lefschetz property (SLP). Our main focus is the case of powers of $r+1$ general linear forms in r variables. For four variables, our results allow the exponents to all be different, and we determine when the WLP holds and when it does not in a broad range of cases. For five variables, we solve this problem in the case where all the exponents are equal (uniform powers), and in the case where one is allowed to be greater than the others. For evenly many variables (≥ 6) we solve the case of uniform powers, and in particular we prove half of a recent conjecture by Harbourne, Schenck and Seceleanu by showing that for evenly many variables, an ideal generated by d-th powers of $r+1$ general linear forms fails the WLP if and only if $d > 1$. For uniform powers of an odd number of variables, we also give a result for seven variables, missing only the case $d = 3$. Our approach in this paper is via the connection (thanks to Macaulay duality) to fat point ideals, together with a reduction to a smaller projective space, and the use of Cremona transformations.

Part of the work for this paper was done while Migliore was sponsored by the National Security Agency under Grant Number H98230-09-1-0031.
Miró-Roig was partially supported by MTM2010-15256.
Part of the work for this paper was done while Nagel was sponsored by the National Security Agency under Grant Number H98230-09-1-0032.

MSC2010: primary 14C20, 13D40; secondary 13E10, 13C13, 13D02.
Keywords: weak Lefschetz property, artinian algebra, powers of linear forms, fat points.

1. Introduction

Ideals generated by powers of linear forms have attracted a great deal of attention recently. For example, their Hilbert functions have been the focus of the papers [Ardila and Postnikov 2010; Sturmfels and Xu 2010; Harbourne et al. 2011], among others. In this paper we obtain further results in this direction, and relate them to the presence or failure of the weak Lefschetz property, which we now recall.

Let $A = R/I$ be a standard graded artinian algebra, where $R = k[x_1, \ldots, x_r]$ and k is a field. If ℓ is a linear form, then multiplication by ℓ induces a homomorphism from any component $[A]_i$ to the next. Such linear forms are parametrized by $[R]_1$. A natural question is whether there is a Zariski-open subset U of $[R]_1$ such that if ℓ corresponds to any point of U, then for each i, multiplication by ℓ induces a homomorphism of maximal rank. When this property holds, the algebra is said to have the *weak Lefschetz property* (*WLP*), and we say that multiplication by a general linear form has maximal rank from each degree to the next. One would naively expect this property to hold, and so it is interesting to find classes of algebras where it fails and to understand what is it about the algebra that prevents this property from holding. There has been a long series of papers by many authors studying different aspects of this problem. Even the characteristic of k plays an interesting role; see for instance [Migliore et al. 2011; Li and Zanello 2010; Brenner and Kaid 2011; Cook and Nagel 2009; 2011].

The first result in this direction is due to R. Stanley [1980], J. Watanabe [1987], and L. Reid, L. Roberts and M. Roitman [Reid et al. 1991], who showed that, in characteristic 0, the WLP holds for an artinian complete intersection generated by powers of variables. In fact, they showed that there is a Zariski-open subset U of $[R]_1$ such that if ℓ corresponds to a point of U, then for each i, multiplication by any power ℓ^d induces a homomorphism of maximal rank from $[A]_i$ to $[A]_{i+d}$; that is, the *strong Lefschetz property* (*SLP*) holds. Since both the WLP and the SLP are preserved after a change of variables, their result shows that both properties hold for any complete intersection whose generators are powers of linear forms. By semicontinuity, it holds for a complete intersection whose generators (of arbitrary degree) are chosen generically.

There are (at least) three natural directions suggested by this theorem. First, we can ask whether the WLP holds for arbitrary complete intersections. It was shown by T. Harima, J. Watanabe and the first and third authors in [Harima et al. 2003] that in two variables, *all* artinian algebras have the WLP. In the same paper,

it was shown that it also holds for arbitrary artinian complete intersections in three variables. It remains open whether it also holds for arbitrary complete intersections in arbitrarily many variables.

Second, a natural question arising from the theorem of [Stanley 1980; Watanabe 1987; Reid et al. 1991] is to ask for which monomial ideals does the WLP hold or not hold. F. Zanello [2006] and H. Brenner and A. Kaid [2007] gave very simple examples to show that even level monomial ideals need not have this property, and the latter gave an example that was even an almost complete intersection (the ideal was in a ring with three variables and had four minimal generators). This latter fact gave a negative answer to a question in [Migliore and Miró-Roig 2003]. In [Migliore et al. 2011], we gave a much more extensive study of monomial almost complete intersections and when they fail to have the WLP. This work was extended in [Cook and Nagel 2009; 2011]. In [Boij et al. 2012], we showed that the only other situation where level monomial ideals have to have the WLP is 3 variables, type 2.

A third interesting problem suggested by the result of [Stanley 1980; Watanabe 1987; Reid et al. 1991] is to ask when the WLP holds for powers of $\geq r+1$ linear forms, since up to a change of variables their result says that any complete intersection of powers of linear forms has the WLP. In [Migliore et al. 2011], we showed by example that in four variables, for $d = 3, \ldots, 12$, an ideal generated by the d-th powers of five general linear forms does not have the WLP. On the other hand, H. Schenck and A. Seceleanu [2010] then gave the surprising result that in three variables, *any* ideal generated by powers of linear forms has the WLP. (We give a new proof of this result in Section 2.) In contrast, Harbourne, Schenck and Seceleanu [Harbourne et al. 2011] have recently shown the following: Let

$$I = \langle \ell_1^t, \ldots, \ell_n^t \rangle \subset k[x_1, \ldots, x_4]$$

with ℓ_i general linear forms. If $n \in \{5, 6, 7, 8\}$ then the WLP fails, respectively, for $t \geq \{3, 27, 140, 704\}$.

A famous conjecture of Fröberg [1985] gives the expected Hilbert function for an ideal of s general forms of prescribed degrees $d_1, \ldots, d_s$. The result of [Stanley 1980; Watanabe 1987; Reid et al. 1991] shows that when $s = r+1$, the same Hilbert function is obtained by the same powers $d_1, \ldots, d_{r+1}$ of general linear forms. Many authors have studied the question of when an ideal of powers of general linear forms has the Hilbert function predicted by Fröberg, and it is known that often it fails even when $s = r+2$; see for instance [Iarrobino 1997]. It is also known (and strongly used in this paper) that there is a strong connection between the Hilbert function of powers of general linear forms and the Hilbert function of a related set of fat points in projective space. The connections between these topics, and a geometric study of the Hilbert function of a set of fat points via Bézout methods, can be found in [Iarrobino 1997] and in [2005; Chandler 2007].

In this paper, we study the WLP for quotients $k[x_1, \ldots, x_r]/I$ where I is an almost complete intersection ideal generated by powers of general linear forms. By the main result of [Schenck and Seceleanu 2010], the first interesting case is $r = 4$, and by the result of [Stanley 1980; Watanabe 1987; Reid et al. 1991], the first interesting situation is that of $r + 1$ forms. This is the focus of this paper. As a main tool, we first use the inverse system dictionary to relate an ideal $I \subset k[x_1, \ldots, x_r]$ generated by powers of linear forms to an ideal of fat points in $\mathbb{P}^{r-1}$, and then we show that the WLP problem of an ideal generated by powers of linear forms is closely connected to the geometry of the linear system of hypersurfaces in $\mathbb{P}^{r-2}$ of fixed degree with preassigned multiple points.

Let us briefly explain how this paper is organized. In Section 2 we give the connection between the WLP problem and Fröberg's conjecture. As a consequence, the results of this paper (about the failure of WLP for explicit ideals generated by $r + 1$ powers of general linear forms in r variables) can be interpreted as explicit results about when an ideal generated by $r + 1$ powers of general linear forms in $r - 1$ variables fails to have the Hilbert function predicted by Fröberg.

We begin Section 3 by explaining the tools that are applied throughout the paper. First, we recall a result of Emsalem and Iarrobino [1995] which gives a duality between powers of linear forms and ideals of fat points in $\mathbb{P}^{r-1}$. Then, we reduce our WLP problem to one of computing the Hilbert function of n general fat points in $\mathbb{P}^{r-2}$ or, equivalently, to computing the dimension of the linear system of hypersurfaces in $\mathbb{P}^{r-2}$ of degree d having some points of fixed multiplicity. Moreover, using Cremona transformations, one can relate two different linear systems to reduce the problem; see [Laface and Ugaglia 2006] or [Dumnicki 2009, Theorem 3].

In Section 4, we consider the case of 4 variables and we give a fairly complete answer about the failure of the WLP for

$$I = \langle L_1^{a_1}, L_2^{a_2}, L_3^{a_3}, L_4^{a_4}, L_5^{a_5} \rangle \subset k[x_1, x_2, x_3, x_4],$$

where the L_i are general linear forms and $2 \leq a_1 \leq a_2 \leq a_3 \leq a_4 \leq a_5$. In Section 5, we deal with 5 variables and we completely determine when an ideal generated by uniform or almost uniform powers of six linear forms fails the WLP. We add some examples to illustrate that our methods extend beyond the mentioned results. The main result of Section 6 is Theorem 6.1, where we give a complete answer to the uniform case when the number of variables is even; in particular, we solve [Harbourne et al. 2011, Conjecture 5.5.2] when the number of variables is even. The case of an odd number of variables is left as an open conjecture and we present some evidence for this conjecture, including a proof for the case of seven variables.

Finally, it is worthwhile to point out that the approach used in this work can be applied to many other situations, in particular when the generators do not all

have the same degree, but the calculations quickly become overwhelming. The key steps of our approach involve identifying the consecutive degrees where the WLP fails, determining the appropriate inequality (or equality) in the dimensions, and then after making the translation to fat points, showing that the dimension of the corresponding linear system has the predicted failure. As we will see, all of these steps become very intricate.

2. Lefschetz properties and Fröberg's conjecture

This section discusses briefly the relations between Fröberg's conjecture and the Lefschetz properties. For all references to Fröberg's work, see [Fröberg 1985].

Let $R = k[x_1, \ldots, x_r]$ be a polynomial ring over a field k with its standard grading. Assume that k is infinite. Denote by h_A the Hilbert function

$$h_A(t) = \dim_k [A]_t$$

of a standard graded k-algebra A. Fix a positive integer s, and fix positive integers $d_1, \ldots, d_s$. Consider the product $P = [R]_{d_1} \times \cdots \times [R]_{d_s}$, with its Zariski topology. Fröberg's idea is that there should be a dense open subset U of P such that for all points $(F_1, \ldots, F_s)$ of U, the coordinate rings of the ideals $\langle F_1, \ldots, F_s \rangle$ should have the same Hilbert function, which was precisely described by Fröberg; we call this *the expected Hilbert function* for the given values of $s, d_1, \ldots, d_s$. (In particular, U will avoid any instances where $F_i = F_j$.) Throughout this paper, when we say that *an ideal of general forms* has some property $(*)$, either the values of $s, d_1, \ldots, d_s$ will be understood from the context, or the statement will be valid for any choices of $s, d_1, \ldots, d_s$, and we will mean that there is a Zariski-open subset U as above whose points correspond to generators of an ideal satisfying $(*)$. Beginning in the next section, usually we will have $s = r + 1$.

Fröberg's conjecture is equivalent to the following structural statement: If $I \subset R$ is an ideal generated by s general forms of degrees $d_1, \ldots, d_s$, and $f \in R$ is another general form of degree d, then the multiplication map

$$\times f : [R/I]_{i-d} \to [R/I]_i$$

has maximal rank for every integer i; that is, it is injective or surjective. This conjecture is known if the number of variables is at most three, due to Anick [1986], and open if the number of variables is greater.

Recall that an artinian graded algebra $A = R/I$ has the *weak Lefschetz property* (*WLP*) if there is a linear form ℓ such that the multiplication $\times \ell : [A]_{i-1} \to [A]_i$ has maximal rank for each integer i. Furthermore, A has the *strong Lefschetz property* (*SLP*) if for each integer $d \geq 1$ the multiplication

$$\times \ell^d : [A]_{i-d} \to [A]_i$$

has maximal rank for each i. It is known that the WLP does not imply the SLP.

Fröberg's conjecture implies, in particular, that all ideals generated by general forms have the WLP. In fact, these concepts are even more closely related, as the following result and its corollaries show.

Proposition 2.1. (a) *If Fröberg's conjecture is true for all ideals generated by general forms in r variables, then all ideals generated by general forms in $r+1$ variables have the WLP.*

(b) *Let $R = k[x_1, \dots, x_{r+1}]$, $\ell \in R$ be a general linear form, and $S = R/\langle \ell \rangle \cong k[x_1, \dots, x_r]$. Fix positive integers $s, d_1, \dots, d_{s+1}$. Let $L_1, L_2, \dots, L_{s+1} \in R$ be linear forms. Denote by $\bar{}$ the restriction from R to $S \cong R/\langle \ell \rangle$. Make the following assumptions*:

(i) *The ideal $I = \langle L_1^{d_1}, \dots, L_s^{d_s} \rangle$ has the WLP.*

(ii) *The multiplication $\times \bar{L}_{s+1}^{d_{s+1}} : [S/\bar{I}]_{j-d_{s+1}} \to [S/\bar{I}]_j$ has maximal rank.*

Then $R/\langle L_1^{d_1}, \dots, L_{s+1}^{d_{s+1}} \rangle$ has the WLP.

Proof. As pointed out in the paragraph preceding [Migliore and Miró-Roig 2003, Proposition 4.3], the proof of [ibid., Proposition 4.3] implies Claim (a). Essentially, it also proves Claim (b). For the reader's convenience, we give the proof of (b) here.

Let $A = R/I$. Let $f = L_{s+1}^{d_{s+1}}$. We have to show that $A/fA \cong R/(I, f)$ has the WLP. To this end, for simplicity let $d = d_{s+1}$ and consider the following commutative diagram with exact rows and columns,

$$\begin{array}{ccccccc}
[A]_{j-d-1} & \xrightarrow[\rho]{} & [A]_{j-1} & \longrightarrow & [A/fA]_{j-1} & \longrightarrow & 0 \\
\downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\
[A]_{j-d} & \xrightarrow[\psi]{} & [A]_j & \longrightarrow & [A/fA]_j & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \\
[A/\ell A]_{j-d} & \xrightarrow[\varphi]{} & [A/\ell A]_j & \longrightarrow & [A/(f,\ell)A]_j & \longrightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \\
0 & & 0 & & 0, & &
\end{array}$$

where α, β, γ are multiplications by ℓ and ρ, ψ, φ are multiplications by f.

By (i), α and β have maximal rank. We have to show that the same is true for γ. If β is surjective, then so is γ. Thus, assume that β is injective. Since the algebras in the bottom row are quotients of $\overline{R} = R/\ell R$ by ideals generated by powers of general linear forms in $\overline{R}$, the assumption implies that also φ has maximal rank. If φ is surjective, then so is γ, and we are done. If φ is injective, then a routine diagram chase shows that γ is injective as well, which completes the argument. □

The last result provides a new short proof of the main result of [Schenck and Seceleanu 2010].

Corollary 2.2 [Schenck and Seceleanu 2010]. *If k has characteristic zero, then every ideal generated by powers of linear forms in 3 variables has the WLP.*

Proof. Let $J \subset R = k[x, y]$ be an ideal generated by powers of linear forms in two variables. By [Harima et al. 2003, Proposition 4.4], every ideal in two variables over a field of characteristic zero has the SLP. Hence, if ℓ in R is a general linear form, then multiplication by any power of ℓ, say ℓ^e, on R/J has maximal rank. However, since the statement is about ideals generated by powers of arbitrary linear forms, we need to show that the multiplication still has maximal rank whenever ℓ is not one of the linear forms whose powers generate J. To this end, it is enough to argue that the Hilbert function of the cokernel of multiplication by ℓ^e, namely $R/(J, \ell^e)$, does not change. However, this is clear because the latter is determined by the Hilbert function of fat points in projective 1-space (see Theorem 3.3 below), which only depends on the degree of the scheme since it is defined by a principal ideal.

Now we can employ the argument used in the proof of Proposition 2.1. Indeed, we just showed that the map φ in the commutative diagram above has maximal rank. □

Corollary 2.3. *Assume the characteristic is zero. Let $R = k[x_1, \dots, x_{r+1}]$, $\ell \in R$ be a general linear form, and $S = R/\langle \ell \rangle \cong k[x_1, \dots, x_r]$. For integers $d_1, \dots, d_{r+2}$, if an ideal of powers of general linear forms $\langle L_1^{d_1}, \dots, L_{r+2}^{d_{r+2}} \rangle \subset R$ fails to have the WLP, then an ideal of powers of general linear forms $\langle \bar{L}_1^{d_1}, \dots, \bar{L}_{r+2}^{d_{r+2}} \rangle \subset S$ fails to have the Hilbert function predicted by Fröberg's conjecture.*

Proof. Taking $s = r + 1$, condition (i) of Proposition 2.1(b) is satisfied by the result of [Stanley 1980; Watanabe 1987; Reid et al. 1991]. Thus (ii) must fail. We also know that $[S/\bar{I}]_{j-d_{r+1}}$ and $[S/\bar{I}]_j$ have the expected dimensions. The failure of $\times \bar{L}_{r+2}^{d_{r+2}}$ to have maximal rank then immediately gives the result. □

If all ideals in $R = k[x_1, \dots, x_r]$ that are generated by powers of general linear forms were to have the SLP, then these ideals would have the expected Hilbert function of ideals generated by general forms of the corresponding degrees, and thus Fröberg's conjecture would be true in R. Since complete intersections generated by powers of linear forms have the SLP, it follows that ideals generated by $r + 1$ powers of general linear forms in r variables have the expected Hilbert functions, so Fröberg's conjecture is true for ideals generated by $r + 1$ general forms of R.

It has already been shown by Iarrobino in [1997] that ideals generated by $r + 2$ powers of general linear forms in r variables do not necessarily have the expected Hilbert function, leading to almost complete intersections generated by powers of

general linear forms in $r+1$ variables that fail the WLP. By Corollary 2.2, this forces $r \geq 4$. Since complete intersections *do* have the WLP, these considerations suggest as a crucial test case the question of whether ideals generated by $r+1$ powers of general linear forms in $r \geq 4$ variables have the WLP.

In the following sections we study this phenomenon much more systematically. We give many explicit ideals of powers of $r+1$ general linear forms that fail the WLP, and so by Corollary 2.3 these all give examples in $r-1$ variables of ideals generated by powers of $r+1$ linear forms that fail to have the Hilbert function predicted by Fröberg. We will not repeat this remark after this section, but it is an important motivation for our work.

Example 2.4. For some specific examples, see [Migliore et al. 2011; Schenck and Seceleanu 2010]. However, we give two illustrations here.

(a) Taking $a_1 = \cdots = a_5 = 5$, in Theorem 4.2 we have $\lambda = 8$, so $k[x_1, \ldots, x_4]/\langle x_1^5, \ldots, x_4^5, L^5\rangle$ fails the WLP from degree 7 to degree 8. Thus the ideal generated by the fifth powers of five general linear forms in $k[x, y, z]$ fails to have the Hilbert function predicted by Fröberg.

(b) If r is odd and $d > 1$, then an ideal of powers of general linear forms of the form $\langle L_1^d, \ldots, L_r^d, L_1^d, L_2^d\rangle$ fails to have the Hilbert function predicted by Fröberg. This follows from Theorem 6.1 and the above discussion.

3. General approach

Let $R = k[x_1, \ldots, x_r]$ be a polynomial ring, where k is a field of characteristic zero.

Notation 3.1. Throughout this paper, when m is any integer, we will denote

$$[m]_+ = \max\{m, 0\}.$$

For any artinian ideal $I \subset R$ and a general linear form $\ell \in R$, the exact sequence

$$\cdots \to [R/I]_{m-1} \xrightarrow{\times \ell} [R/I]_m \to [R/(I, \ell)]_m \to 0$$

gives, in particular, that the multiplication by ℓ will fail to have maximal rank exactly when

$$\dim_k [R/(I, \ell)]_m \neq \max\{\dim_k [R/I]_m - \dim_k [R/I]_{m-1}, 0\}; \tag{3-1}$$

in that case, we will say that R/I fails the WLP in degree m.

Remark 3.2. Notice that to show that the multiplication by ℓ fails to have maximal rank from degree $m-1$ to degree m, it is enough to check the failure of injectivity if $\dim[R/I]_{m-1} \leq \dim[R/I]_m$, and it is enough to check the failure of surjectivity if $\dim[R/I]_{m-1} \geq \dim[R/I]_m$. We will then say that R/I *fails injectivity* or *fails*

surjectivity, respectively. An important part of the argument is the verification of the inequality of the dimensions.

In several of the papers mentioned above, this failure was studied via an examination of the splitting type of the first syzygy bundle of I. For powers of linear forms, we give an alternative approach, which we will implement in the subsequent sections.

We first recall a result of Emsalem and Iarrobino giving a duality between powers of linear forms and ideals of fat points in $\mathbb{P}^{n-1}$. We quote it in the form that we need.

Theorem 3.3 [Emsalem and Iarrobino 1995, Theorem I]. *Let $\langle L_1^{a_1}, \dots, L_n^{a_n}\rangle \subset R$ be an ideal generated by powers of n linear forms. Let $\wp_1, \dots, \wp_n$ be the ideals of n points in $\mathbb{P}^{r-1}$. (Each point is actually obtained explicitly from the corresponding linear form by duality.) Choose positive integers $a_1, \dots, a_n$. Then, for any integer $j \geq \max\{a_i\}$,*

$$\dim_k\big[R/\langle L_1^{a_1}, \dots, L_n^{a_n}\rangle\big]_j = \dim_k\big[\wp_1^{j-a_1+1} \cap \cdots \cap \wp_n^{j-a_n+1}\big]_j.$$

Now, we observe that the ideal (I, ℓ) is also an ideal generated by powers of linear forms! We conclude that if $\wp$ is the ideal of the point dual to ℓ then

$$\dim_k\big[R/\langle L_1^{a_1}, \dots, L_n^{a_n}, \ell\rangle\big]_j = \dim_k\big[\wp_1^{j-a_1+1} \cap \cdots \cap \wp_n^{j-a_n+1} \cap \wp^j\big]_j.$$

Consider the points $P_1, \dots, P_n, P$ in $\mathbb{P}^{r-1}$ defined by the ideals $\wp_1, \dots, \wp_n, \wp$, respectively. Let λ_i be the line joining P to P_i, and let $H = \mathbb{P}^{r-2}$ be a general hyperplane defined by a linear form L_H. Let Q_i be the point of intersection of λ_i with H. For any positive integer m, we will denote by λ_i^m the curve with defining ideal $I_{\lambda_i}^m$, and notice that λ_i^m is arithmetically Cohen–Macaulay. Thus the hyperplane section of λ_i^m by H has saturated ideal $\mathfrak{q}_i^m = (I_{\lambda_i}^m, L_H)/(L_H)$ in the coordinate ring $R/(L_H)$ of H, and $\mathfrak{q}_i^m$ defines a fat point Q_i^m in $H = \mathbb{P}^{r-2}$. The curve

$$Y = \lambda_1^{j-a_1+1} \cup \cdots \cup \lambda_n^{j-a_n+1}$$

is the cone over $Q_1^{a_1} \cup \cdots \cup Q_n^{a_n}$, and thus is also arithmetically Cohen–Macaulay.

Proposition 3.4. *Let $\langle L_1^{a_1}, \dots, L_n^{a_n}\rangle \subset R$ be an ideal generated by powers of n linear forms, and let ℓ be a general linear form. For $j \geq \max\{a_i\}$, we have*

$$\begin{aligned}\dim_k\big[R/\langle L_1^{a_1}, \dots, L_n^{a_n}, \ell\rangle\big]_j &= \dim_k\big[\wp_1^{j-a_1+1} \cap \cdots \cap \wp_n^{j-a_n+1} \cap \wp^j\big]_j \quad (\text{in } k[x_1, \dots, x_r])\\ &= \dim_k\big[\mathfrak{q}_1^{j-a_1+1} \cap \cdots \cap \mathfrak{q}_n^{j-a_n+1}\big]_j \quad (\text{in } k[x_1, \dots, x_{r-1}]).\end{aligned}$$

Proof. The first equality is Theorem 3.3. Without loss of generality let $P = [0, \dots, 0, 1]$, with defining ideal $\wp = \langle x_1, \dots, x_{r-1}\rangle$, and assume that H is defined

by $x_r = 0$. Any form $F \in [\wp_1^{j-a_1+1} \cap \cdots \cap \wp_n^{j-a_n+1} \cap \wp^j]_j$ involves only the variables $x_1, \ldots, x_{r-1}$ since it is in $[\wp^j]_j$. Thus $F \in [I_Y]_j$, so viewing F in $k[x_1, \ldots, x_{r-1}]$, we see that F vanishes on the hyperplane section of Y; that is,

$$F \in \left[\mathfrak{q}_1^{j-a_1+1} \cap \cdots \cap \mathfrak{q}_n^{j-a_n+1}\right]_j. \tag{$*$}$$

Now suppose F satisfies $(*)$. Viewing F in $k[x_1, \ldots, x_r]$ we see that F vanishes on Y, hence also on the subscheme of Y defined by $\wp_1^{j-a_1+1} \cap \cdots \cap \wp_n^{j-a_n+1}$. But we also have that $F \in \wp^j$, since it involves only $x_1, \ldots, x_{r-1}$. Moreover F has degree j. Thus $F \in \wp_1^{j-a_1+1} \cap \cdots \cap \wp_n^{j-a_n+1} \cap \wp^j]_j$ as desired. □

Using Proposition 3.4, we reduce our WLP problem to one of computing the Hilbert function of n general fat points in $\mathbb{P}^{r-2}$. From now on, we will denote by

$$\mathscr{L}_{r-2}(j;\, j-a_1+1,\, j-a_2+1,\, \ldots,\, j-a_n+1)$$

the linear system $[\mathfrak{q}_1^{j-a_1+1} \cap \cdots \cap \mathfrak{q}_n^{j-a_n+1}]_j \subset [k[x_1, \ldots, x_{r-1}]]_j$. In order to simplify notation, we use superscripts to indicate repeated entries. For example, $\mathscr{L}_3(j; 5^2, 2^3) = \mathscr{L}_3(j; 5, 5, 2, 2, 2)$.

Notice that, for every linear system $\mathscr{L}_r(j; a_1, \ldots, a_n)$, one has

$$\dim_k \mathscr{L}_r(j; a_1, \ldots, a_n) \geq \max\left\{0, \binom{j+r}{r} - \sum_{i=1}^{n} \binom{a_i+r-1}{r}\right\},$$

where the right-hand side is called the *expected dimension* of the linear system. If the inequality is strict, then the linear system $\mathscr{L}_r(j; a_1, \ldots, a_n)$ is called *special*. It is a difficult problem to classify the special linear systems.

Using Cremona transformations, one can relate two different linear systems; see [Nagata 1960; Laface and Ugaglia 2006; Dumnicki 2009, Theorem 3].

Lemma 3.5. *Let $n > r \geq 2$ and $j, a_1, \ldots, a_n$ be nonnegative integers, and set $m = (r-1)j - (a_1 + \cdots + a_{r+1})$. If $a_i + m \geq 0$ for all $i = 1, \ldots, r+1$, then*

$$\dim_k \mathscr{L}_r(j; a_1, \ldots, a_n) = \dim_k \mathscr{L}_r(j+m; a_1+m, \ldots, a_{r+1}+m, a_{r+2}, \ldots, a_n).$$

Following [De Volder and Laface 2007], the linear system $\mathscr{L}_r(j; a_1, \ldots, a_n)$ is said to be in *standard form* if $(r-1)j \geq a_1 + \cdots + a_{r+1}$ and $a_1 \geq \cdots \geq a_n \geq 0$. If $r = 2$, then every linear system in standard form is nonspecial. This is no longer true if $r \geq 3$. However, De Volder and Laface were able to compute the speciality in the case of at most 8 fat points in $\mathbb{P}^3$. We state their result only in the form we need it.

Theorem 3.6 [De Volder and Laface 2007, Theorem 5.3]. *If the linear system* $\mathcal{L}_3(j; a_1, \ldots, a_6)$ *is in standard form, then*

$$\dim_k \mathcal{L}_3(j; a_1, \ldots, a_6) = \max\left\{0, \binom{j+r}{r} - \sum_{i=1}^{6}\binom{a_i+r-1}{r}\right\} + \sum_{i=2}^{6}\binom{t_i+1}{3},$$

where $t_i = a_1 + a_i - j$.

Notice that we always use the vector space dimension of the linear system rather than the dimension of its projectivization and that we adjusted the formula for the expected dimension. Furthermore, we always use the convention that a binomial coefficient $\binom{a}{r}$ is zero if $a < r$.

In this note, we are interested in certain almost complete intersections. Then one can compute the right-hand side of the inequality (3-1).

Lemma 3.7. *Let* $I = \langle L_1^{a_1}, \ldots, L_{r+1}^{a_{r+1}}\rangle \subset R$ *be an almost complete intersection generated by powers of* $r+1$ *general linear forms. Then, for each integer* j,

$$\begin{aligned}\dim_k[R/I]_j - \dim_k[R/I]_{j-1} &= \big[h_A(j) - h_A(j-a_{r+1})\big]_+ - \big[h_A(j-1) - h_A(j-1-a_{r+1})\big]_+,\end{aligned}$$

where $A = R/\langle L_1^{a_1}, \ldots, L_r^{a_r}\rangle$. *Furthermore, if* $j \le \frac{1}{2}a_{r+1} + \frac{1}{2}\sum_{i=1}^{r}(a_i - 1)$, *then the formula simplifies to*

$$\begin{aligned}\dim_k[R/I]_j - \dim_k[R/I]_{j-1} &= \big[h_A(j) - h_A(j-1)\big] - \big[h_A(j-a_{r+1}) - h_A(j-1-a_{r+1})\big].\end{aligned}$$

Proof. Considering multiplication by $l_{r+1}^{a_{r+1}}$ on A, the first equation follows because the complete intersection A has the SLP according to [Stanley 1980; Watanabe 1987; Reid et al. 1991]. The latter also implies that the Hilbert function of A is unimodal. Its midpoint is $\frac{1}{2}\sum_{i=1}^{r}(a_i - 1)$. Thus, the differences in brackets in the first formula are not negative if $j \le \frac{1}{2}a_{r+1} + \frac{1}{2}\sum_{i=1}^{r}(a_i - 1)$, proving the second formula. □

Notice that the Hilbert function of the complete intersection A can be computed using the Koszul complex that provides its minimal free resolution.

4. Powers of linear forms in four variables

In this section we let $R = k[x_1, x_2, x_3, x_4]$, where k is a field of characteristic zero. Our main result will be to determine, in almost all cases, when an ideal generated by powers of five general linear forms has the WLP. To this end, without loss of generality we set $I = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}\rangle$, where L is a general linear form and $a_1 \le a_2 \le a_3 \le a_4 \le a_5$. Part of the argument involves an application of Bézout's

theorem to remove one-dimensional components of the relevant linear systems; this approach was also used in [Chandler 2005].

Lemma 4.1. *Let $P_1, \dots, P_6$, be points in $\mathbb{P}^2$ in linear general position. Assign multiplicities $m_1, \dots, m_6$ respectively to the points, with $0 \le m_1 \le \cdots \le m_6$. (In particular, taking some of the $m_i = 0$ allows us to consider fewer than six points.) Assume that $d \ge m_5 + m_6$ and that $2d \ge \sum_{i=2}^{6} m_i$. Then the fat point scheme $Z = m_1P_1 + \cdots + m_6P_6$ imposes independent conditions on curves of degree d.*

Proof. Let X be the rational surface obtained by blowing up $\mathbb{P}^2$ at the points P_i. Let $L, E_1, \dots, E_6$ be the standard basis of the divisor class group of X, that is, L is the pullback of the class of a line in $\mathbb{P}^2$ and $E_1, \dots, E_6$ are the exceptional divisors. Under the stated assumptions $d \ge m_5 + m_6$ and $2d \ge \sum_{i=2}^{6} m_i$, the divisor $dL - m_1E_1 - \cdots - m_6E_6$ is numerically effective (nef), by [Di Rocco 1996, Theorem 3.4]. Then [Geramita et al. 2009, Theorem 2.3 and Remark 2.4] show that the Castelnuovo–Mumford regularity of I_Z is $\le d + 1$. This implies the claimed result. □

The following theorem represents "half" of our results (the "even" case) concerning when an ideal of five powers of general linear forms has the WLP. It is not a complete list, however, because of the second condition in (iii). See Remark 4.4 for further discussion.

Theorem 4.2. *Let L be a general linear form and let $I = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5} \rangle$. Assume that $a_1 + a_2 + a_3 + a_4$ is even. Let $\lambda = (a_1 + a_2 + a_3 + a_4)/2 - 2$.*

(i) *If $a_5 \ge \lambda$ then the ring R/I has the WLP.*

(ii) *If $a_5 < \lambda$ and $a_1 + a_4 \ge a_2 + a_3$ then R/I fails the WLP from degree $\lambda - 1$ to degree λ.*

(iii) *If $a_5 < \lambda$, $a_1 + a_4 < a_2 + a_3$ and $2a_5 + a_1 - a_2 - a_3 - a_4 \ge 0$ then R/I fails the WLP from degree $\lambda - 1$ to degree λ.*

The hypotheses of (ii) and of (iii) both force $a_1 \ge 3$.

Proof. Let ℓ be a general linear form. Letting $J = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4} \rangle$, consider the commutative diagram

$$\begin{array}{ccc} [R/J]_t & \xrightarrow{\times \ell} & [R/J]_{t+1} \\ \downarrow & & \downarrow \\ [R/I]_t & \xrightarrow{\times \ell} & [R/I]_{t+1} \end{array}$$

(where the vertical arrows are the natural restriction). If we know that the multiplication on the top row is surjective, then we immediately conclude surjectivity on the bottom row. Notice that 2λ is the socle degree of the artinian complete intersection $R/(x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4})$, so λ is the midpoint.

First assume that $a_5 > \lambda$. Then clearly R/I and R/J coincide in degrees $\leq \lambda$, so we have injectivity (by the result of Stanley, Watanabe, Reid, Roberts and Roitman) for $(\times \ell) : [R/I]_t \to [R/I]_{t+1}$ for all $t \leq \lambda - 1$. When $t \geq \lambda$ we have surjectivity for R/J, so by the above result we also have it for R/I.

Now assume that $a_5 = \lambda$. We wish to show that R/I has the WLP. Again surjectivity is immediate for $t \geq \lambda$, and injectivity is immediate for $t \leq \lambda - 2$. When $t = \lambda - 1$ we consider two cases. If $\dim[R/J]_{\lambda-1} = \dim[R/J]_\lambda$ then by Stanley–Watanabe we have surjectivity for $(\times \ell) : [R/J]_{\lambda-1} \to [R/J]_\lambda$, so the same holds for R/I. If $\dim[R/J]_{\lambda-1} < \dim[R/J]_\lambda$, the image of $[R/J]_{\lambda-1}$ in $[R/J]_\lambda$ under multiplication by ℓ is not surjective. Hence the vector space $\left[\langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, \ell\rangle\right]_\lambda$ is not all of R_λ. But in characteristic zero the λ-th powers of linear forms span $[R]_\lambda$. Thus for a general L, the image of L^λ in $[R/J]_\lambda$ is outside the image of $[R/J]_{\lambda-1}$ in $[R/J]_\lambda$. Thus $(\times \ell) : [R/I]_{\lambda-1} \to [R/I]_\lambda$ is injective since it is for R/J in that degree. This completes the proof of (i).

We now prove (ii). Our assumptions now are that $a_1 + a_2 + a_3 + a_4$ is even, $a_5 < \lambda$, and $a_1 + a_4 \geq a_2 + a_3$. We first note that the hypotheses force $a_1 \geq 3$, since if $a_1 = 2$ then $a_5 < \lambda = (a_2 + a_3 + a_4 - 2)/2 \leq (2 + 2a_4 - 2)/2 = a_4$.

We will show that the multiplication by ℓ fails to have maximal rank from degree $\lambda - 1$ to degree λ. If $\dim[R/I]_{\lambda-1} \leq \dim[R/I]_\lambda$ it is enough to check the failure of injectivity, and if $\dim[R/I]_{\lambda-1} \geq \dim[R/I]_\lambda$ it is enough to check the failure of surjectivity. (See Example 4.3.)

According to (3-1), the task is to show that

$$\dim[R/(I, \ell)]_\lambda > [\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}]_+. \tag{4-1}$$

We will compute the left-hand side using Proposition 3.4 and the right-hand side using Lemma 3.7.

By Proposition 3.4 and Theorem 3.3, we have

$$\begin{aligned} &\dim_k[R/(I, \ell)]_\lambda \\ &= \dim_k[\mathfrak{q}_1^{\lambda-a_1+1} \cap \mathfrak{q}_2^{\lambda-a_2+1} \cap \mathfrak{q}_3^{\lambda-a_3+1} \cap \mathfrak{q}_4^{\lambda-a_4+1} \cap \mathfrak{q}_5^{\lambda-a_5+1}]_\lambda \\ &= \dim_k \mathscr{L}_2(\lambda; \lambda - a_1 + 1, \lambda - a_2 + 1, \lambda - a_3 + 1, \lambda - a_4 + 1, \lambda - a_5 + 1). \end{aligned} \tag{4-2}$$

Notice that $\lambda - a_1 + 1 \geq \cdots \geq \lambda - a_5 + 1 \geq 2$. The vector space defines a linear system, and we want to find its dimension. The first step is to understand the one-dimensional base locus, which has components of degree 2 and of degree 1. We will use Bézout's theorem to formally reduce the degree of the polynomials and the order of vanishing at the points, without changing the dimension of the linear system. If the result has dimension zero then (4-2) is also zero.

Let $F_\lambda \in \mathscr{L}_2(\lambda; \lambda - a_1 + 1, \lambda - a_2 + 1, \lambda - a_3 + 1, \lambda - a_4 + 1, \lambda - a_5 + 1)$. There is a unique quadratic polynomial F_2 vanishing at the five general points. By Bézout's

theorem, if F_2 is not a factor of F_λ then it intersects F_λ with multiplicity 2λ. On the other hand, considering the multiplicities at the five points, we get that the curves meet with multiplicity at least $5\lambda - \sum_{i=1}^{5} a_i + 5$. But

$$5\lambda - \sum_{i=1}^{5} a_i + 5 \le 2\lambda \iff a_5 \ge \frac{a_1+a_2+a_3+a_4}{2} - 1 = \lambda + 1,$$

a contradiction. Hence F_2 is a factor of F_λ. We next want to know what power of F_2 divides F_λ. Thus we want to know the smallest i for which

$$2(\lambda - 2i) \ge (\lambda - a_1 + 1 - i) + \cdots + (\lambda - a_5 + 1 - i),$$

and this is equivalent to $i \ge \lambda - a_5 + 1$. Thus we have $F_\lambda = F_2^{\lambda - a_5 + 1} \cdot F_{2a_5 - \lambda - 2}$, where $F_{2a_5-\lambda-2} \in \mathscr{L}_2(2a_5 - \lambda - 2; a_5 - a_1, a_5 - a_2, a_5 - a_3, a_5 - a_4)$. (Notice that now there are only four points, and some of these multiplicities might even be zero.)

Now we consider linear factors coming from the lines joining two of these four points. There are six such lines; we denote by L_{ij}, $1 \le i < j \le 4$, the line (as well as the linear form) passing through the points with multiplicities $a_5 - a_i$ and $a_5 - a_j$. Notice that if $a_5 - a_i > 2a_5 - \lambda - 2$ then there are no such forms $F_{2a_5-\lambda-2}$, so the desired dimension is zero.

Arguing in a similar manner as above, we obtain that if $(a_5 - a_i) + (a_5 - a_j) > 2a_5 - \lambda - 2$, then L_{ij} appears as a factor

$$(a_5 - a_i) + (a_5 - a_j) - (2a_5 - \lambda - 2) = \lambda + 2 - a_i - a_j$$

times. Thus, letting $A_{ij} = [\lambda + 2 - a_i - a_j]_+ = \max\{\lambda + 2 - a_i - a_j, 0\}$, we see that (formally)

$$F_\lambda = F_2^{\lambda - a_5 + 1} \cdot \prod_{1 \le i < j \le 4} L_{ij}^{A_{ij}} \cdot G, \quad \text{where } \deg G = 2a_5 - \lambda - 2 - \sum_{1 \le i < j \le 4} A_{ij}. \tag{4-3}$$

Notice that

$$\text{if } 2a_5 - \lambda - 2 < \sum_{1 \le i < j \le 4} A_{ij}, \text{ then } \dim[\langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}, \ell \rangle]_\lambda = 0. \tag{4-4}$$

Now, since we have assumed $a_1 + a_4 \ge a_2 + a_3$, observe that

$$\begin{array}{lll} a_1 + a_4 - a_2 - a_3 \ge 0 & a_2 + a_3 - a_1 - a_4 \le 0 & a_2 + a_4 - a_1 - a_3 \ge 0 \\ a_1 + a_3 - a_2 - a_4 \le 0 & a_3 + a_4 - a_1 - a_2 \ge 0 & a_1 + a_2 - a_3 - a_4 \le 0. \end{array} \tag{4-5}$$

We claim that after removing the one-dimensional base locus, we obtain a set of at most 4 fat points of uniform multiplicities (possibly 0). We have already seen that after removing the powers of F_2 we are left with the problem of finding the

dimension of the forms of degree $2a_5 - \lambda - 2$ passing through four general fat points with multiplicity $a_5 - a_1, \ldots, a_5 - a_4$. To compute this, we have to determine precisely what is left when we remove the lines. At each of the four points we compute the multiplicity of the fat point that remains after we remove the (multiple) lines passing through it in the base locus (which do not contribute to the desired dimension):

$$\begin{aligned} &a_5 - a_4 - \sum_{i=1,2,3} [\lambda + 2 - a_i - a_4]_+ \\ &\quad = a_5 - a_4 - \left[\frac{-a_1+a_2+a_3-a_4}{2}\right]_+ - \left[\frac{a_1-a_2+a_3-a_4}{2}\right]_+ - \left[\frac{a_1+a_2-a_3-a_4}{2}\right]_+ \\ &\quad = a_5 - a_4. \end{aligned}$$

Similarly,

$$\begin{aligned} a_5 - a_3 - \textstyle\sum_{i=1,2,4} [\lambda + 2 - a_i - a_3]_+ &= a_5 - a_4, \\ a_5 - a_2 - \textstyle\sum_{i=1,3,4} [\lambda + 2 - a_i - a_2]_+ &= a_5 - a_4, \\ a_5 - a_1 - \textstyle\sum_{i=2,3,4} [\lambda + 2 - a_i - a_1]_+ &= a_5 - a_4. \end{aligned}$$

Concluding, we want to find the dimension of the vector space of forms of degree

$$2a_5 - \lambda - 2 - \sum_{1 \le i < j \le 4} A_{ij} = 2(a_5 - a_4)$$

passing through four points with multiplicity $a_5 - a_4$. By Lemma 4.1, this dimension is $a_5 - a_4 + 1$ (which in particular is at least 1). That is,

$$\dim[\langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}, \ell\rangle]_\lambda = a_5 - a_4 + 1 \ge 1. \tag{4-6}$$

Using Lemma 3.7, we now compute the "expected" dimension, that is, the right-hand side of (4-1). Let $A = R/J$, where $J = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}\rangle$. In the following, recall also that λ is the midpoint of the h-vector of R/J, so $0 \le h_{R/I}(\lambda) = h_A(\lambda) - h_A(\lambda - a_5)$. Observe that $a_4 < a_1 + a_2 + a_3$ since otherwise

$$a_5 - \lambda \ge a_4 - \lambda = \frac{a_4 - a_1 - a_2 - a_3}{2} + 2 > 0,$$

contradicting our assumption that $a_5 < \lambda$. This implies, in particular, that for determining $h_A(\lambda)$ by using the Koszul resolution of A, we only need to consider the free modules up to $\bigoplus_{1 \le i < l \le 4} R(-a_i - a_j)$.

From (4-5) we see that only $\lambda - a_2 - a_3 + 2$, $\lambda - a_1 - a_3 + 2$ and $\lambda - a_1 - a_2 + 2$ can be positive. Similarly, we have $\lambda - a_5 - a_i + 2 \le 0$ since $a_1 + a_4 \ge a_2 + a_3$.

Then the "expected dimension" is

$$\begin{aligned}
&\big[\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}\big]_+ \\
&\quad = \big[h_A(\lambda) - h_A(\lambda - a_5) - h_A(\lambda - 1) + h_A(\lambda - a_5 - 1)\big]_+ \\
&\quad = \left[\binom{\lambda+2}{2} - \sum_{i=1}^{4}\binom{\lambda-a_i+2}{2} + \sum_{1\le i<j\le 3}\binom{\lambda-a_i-a_j+2}{2} - \binom{\lambda-a_5+2}{2}\right]_+.
\end{aligned}$$

If this is zero then clearly the actual dimension exceeds the expected one and we are done. If not, one verifies (we used CoCoA version 4.7.5 [CoCoA 2009]) that

$$\begin{aligned}
&\dim_k[R/(I,\ell)]_\lambda - \big[\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}\big]_+ \\
&\quad = (a_5 - a_4 + 1) \\
&\qquad - \left(\binom{\lambda+2}{2} - \sum_{i=1}^{4}\binom{\lambda-a_i+2}{2} + \sum_{1\le i<j\le 3}\binom{\lambda-a_i-a_j+2}{2} - \binom{\lambda-a_5+2}{2}\right) \\
&\quad = \binom{\lambda+1-a_5}{2},
\end{aligned}$$

and this last binomial coefficient is at least 1. Thus, in either case we have shown the inequality (4-1), which completes the proof of (ii).

Finally we prove (iii). Our assumptions now are that $a_1+a_2+a_3+a_4$ is even, $a_5 < \lambda$, $a_1+a_4 < a_2+a_3$ and $2a_5+a_1-a_2-a_3-a_4 \ge 0$. The calculations from (ii) continue to be valid up to (4-3) and (4-4). We first note that the hypotheses again force $a_1 \ge 3$ since if $a_1 = 2$ then $a_5 < \lambda = (a_2+a_3+a_4-2)/2 \le (2+2a_5-2)/2 = a_5$.

In our current situation, though, observe that

$$\begin{array}{lll}
a_3+a_4-a_1-a_2 > 0 & a_1+a_2-a_3-a_4 < 0 & a_2+a_4-a_1-a_3 > 0 \\
a_1+a_3-a_2-a_4 < 0 & a_2+a_3-a_1-a_4 > 0 & a_1+a_4-a_2-a_3 < 0.
\end{array} \tag{4-7}$$

Now we examine the linear system remaining after we remove the one-dimensional base locus. As in (ii), we obtain a linear system of curves of degree

$$2a_5 - \lambda - 2 - \sum_{1\le i<j\le 4} A_{ij}.$$

Now we compute the order of vanishing at the four points:

$$\begin{aligned}
&a_5 - a_4 - \sum_{i=1,2,3}[\lambda + 2 - a_i - a_4]_+ \\
&\quad = a_5 - a_4 - \left[\frac{-a_1+a_2+a_3-a_4}{2}\right]_+ - \left[\frac{a_1-a_2+a_3-a_4}{2}\right]_+ - \left[\frac{a_1+a_2-a_3-a_4}{2}\right]_+ \\
&\quad = a_5 - \frac{-a_1+a_2+a_3+a_4}{2}.
\end{aligned}$$

By the additional hypothesis $2a_5 + a_1 - a_2 - a_3 - a_4 \geq 0$, this order of vanishing is ≥ 0. A similar computation gives that the order of vanishing at the other three points is the same. Thus Lemma 4.1 shows that

$$\dim[\langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}, \ell\rangle]_\lambda = a_5 - \frac{-a_1+a_2+a_3+a_4}{2} + 1 \geq 1.$$

The computation of the "expected" dimension is very similar to what we did above. From (4-7) we see that only $\lambda - a_1 - a_2 + 2$, $\lambda - a_1 - a_3 + 2$ and $\lambda - a_1 - a_4 + 2$ can be positive. We again have $a_4 < a_1 + a_2 + a_3$. Thus, Lemma 3.7 provides

$$\begin{aligned}
&\big[\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}\big]_+ \\
&\quad = \big[h_A(\lambda) - h_A(\lambda - a_5) - h_A(\lambda - 1) + h_A(\lambda - a_5 - 1)\big]_+ \\
&\quad = \left[\binom{\lambda+2}{2} - \sum_{i=1}^{4}\binom{\lambda-a_i+2}{2} + \sum_{i=2}^{4}\binom{\lambda-a_1-a_i+2}{2} - \binom{\lambda-a_5+2}{2}\right]_+.
\end{aligned}$$

As above, if this is zero then we are done. If not, one verifies (for example with CoCoA) that

$$\begin{aligned}
&\dim_k[R/(I,\ell)]_\lambda - \big[\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}\big]_+ \\
&\quad = a_5 - \frac{-a_1+a_2+a_3+a_4}{2} + 1 \\
&\qquad - \left(\binom{\lambda+2}{2} - \sum_{i=1}^{4}\binom{\lambda-a_i+2}{2} + \sum_{i=2}^{4}\binom{\lambda-a_1-a_i+2}{2} - \binom{\lambda-a_5+2}{2}\right) \\
&\quad = \binom{\lambda+1-a_5}{2},
\end{aligned}$$

and this last binomial coefficient is at least 1, establishing the inequality (4-1). This completes the proof of (iii). □

Example 4.3. To illustrate that sometimes it is injectivity that fails and sometimes it is surjectivity, consider the following (produced using CoCoA).

When $a_1 = 5$, $a_2 = 7$, $a_3 = 8$, $a_4 = 10$, $a_5 = 10$ we get $\dim[R/I]_{12} = 225$ and $\dim[R/I]_{13} = 220$, so we expect surjectivity, but the image under multiplication by a general linear form has dimension 219.

When $a_1 = 5$, $a_2 = 7$, $a_3 = 8$, $a_4 = 10$, $a_5 = 12$ we get $\dim[R/I]_{12} = 234$, $\dim[R/I]_{13} = 236$, so we expect injectivity, but the image under multiplication by a general linear form has dimension 233.

Let us now discuss cases that are not covered by Theorem 4.2, still assuming that $a_1 + a_2 + a_3 + a_4$ is even.

Remark 4.4. As above, let $I = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}\rangle$. Assume $2a_5 + a_1 - a_2 - a_3 - a_4 < 0$ and $a_1 \geq 3$. Then:

(i) If $a_1 = 2$ then R/I has the WLP (see Theorem 4.6).

(ii) If $a_1 = 3$, then in the following cases, R/I fails to have the WLP, and the failure is in degree $\lambda - 1$ and it fails by 1.

- $(3, 9, m, m, m)$ for $m \geq 9$
- $(3, 10, m, m+1, m+1)$ for $m \geq 10$
- $(3, 11, m, m, m+1)$ for $m \geq 11$
- $(3, 11, m, m+2, m+2)$ for $m \geq 11$
- $(3, 12, m, m+1, m+2)$ for $m \geq 12$
- $(3, 12, m, m+3, m+3)$ for $m \geq 12$
- $(3, 13, m, m, m+2)$ for $m \geq 13$
- $(3, 13, m, m+2, m+3)$ for $m \geq 13$
- $(3, 13, m, m+4, m+4)$ for $m \geq 13$

(This is shown using arguments as in the proof of Theorem 4.2. We omit the details.)

Furthermore, the reader can easily construct examples using CoCoA to support the following statements, although we do not have proofs (see also the beginning of the proof of Theorem 6.5):

(iii) When $a_1 = 3$ and $a_2 \leq 13$, all cases apart from the ones above have the WLP.

(iv) When $a_1 = 3, 4$, some examples have the WLP and others do not.

(v) When $a_1 \geq 5$, R/I fails the WLP.

We now consider the case where $a_1 + a_2 + a_3 + a_4$ is odd. There are some interesting differences to Theorem 4.2. As before, Theorem 4.5 is not a complete classification because of the extra condition in (iii).

Theorem 4.5. *Let L be a general linear form and let $I = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}\rangle$. Assume that $a_1 + a_2 + a_3 + a_4$ is odd. Let $\lambda = (a_1 + a_2 + a_3 + a_4 - 5)/2$.*

(i) *If $a_5 \geq \lambda - 1$ then the ring R/I has the WLP.*

(ii) *If $a_5 < \lambda - 1$ and $a_1 + a_4 \geq a_2 + a_3$ then R/I fails the WLP from degree $\lambda - 1$ to degree λ. These hypotheses force $a_1 \geq 5$.*

(iii) *If $a_5 < \lambda - 1$, $a_1 + a_4 < a_2 + a_3$ and $2a_5 + 3 - a_4 - a_3 - a_2 + a_1 \geq 0$ then R/I fails the WLP from degree $\lambda - 1$ to degree λ. These hypotheses force $a_1 \geq 3$.*

Proof. The outline of part of the proof is the same as that for Theorem 4.2, and we only highlight the differences. First note that with the hypotheses of (ii), if $a_1 \leq 4$ then we have $a_5 < \lambda - 1 = (a_1 + a_2 + a_3 + a_4 - 7)/2 \leq (2a_1 + 2a_4 - 7)/2 \leq a_4 + (1)/2$, a contradiction. Similarly, with the hypotheses of (iii), if $a_1 = 2$ then we have $a_2 + a_3 + a_4 - 5 \leq 2a_5 < 2\lambda - 2 = a_2 + a_3 + a_4 - 5$, again a contradiction.

Now, for all three parts of the theorem, we show that the one-dimensional part of the base locus corresponding to the quadratic polynomial F_2 has equation $F_2^{\lambda-5}$. Thus, in the first step we want to compute

$$\dim \mathscr{L}_2(2a_5-\lambda; a_5-a_1+1, a_5-a_2+1, a_5-a_3+1, a_5-a_4+1, 1). \quad (4\text{-}8)$$

In the second step, we obtain that if $(a_5-a_i+1)+(a_5-a_j+1) > 2a_5-\lambda$ then L_{ij} appears as a factor

$$(a_5-a_i+1)+(a_5-a_j+1)-(2a_5-\lambda) = \lambda-a_i-a_j+2$$

times. Thus we let $A_{ij} = [\lambda+2-a_i-a_j]_+$ as before, and formally we have

$$F_\lambda = F_2^{\lambda-a_5} \cdot \prod_{1\le i<j\le 4} L_{ij}^{A_{ij}} \cdot G, \quad \text{where} \quad \deg G = 2a_5-\lambda-\sum_{1\le i<j\le 4} A_{ij}. \quad (4\text{-}9)$$

Notice that

$$\text{if} \quad 2a_5-\lambda < \sum_{1\le i<j\le 4} A_{ij} \quad \text{then} \quad \dim[\langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}, \ell\rangle]_\lambda = 0. \quad (4\text{-}10)$$

For (i), we want to show that whenever $a_5 \ge \lambda-1$, the multiplication $(\times\ell)\colon [R/I]_t \to [R/I]_{t+1}$ has maximal rank. Let $J = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}\rangle$. Notice that we have $h_{R/J}(\lambda) = h_{R/J}(\lambda+1)$. We consider several cases.

- If $t \ge \lambda$, we know that the multiplication on R/J from degree t to degree $t+1$ is surjective, by the result of Watanabe, Stanley, Reid, Roberts and Roitman. Since R/I is a quotient, the same holds for R/I. This holds no matter what a_5 is.
- If $t \le a_5-2$, then $[R/I]_t = [R/J]_t$ and $[R/I]_{t+1} = [R/J]_{t+1}$, so again the result follows trivially. Notice that as a result of these first two cases, we are done if $a_5 \ge \lambda+1$.
- If $(t, a_5) = (\lambda-1, \lambda)$ or $(t, a_5) = (\lambda-2, \lambda-1)$, we know that the multiplication for R/J is injective in either of these cases, and that $\dim[R/I]_t = \dim[R/J]_t$ and $\dim[R/I]_{t+1} = \dim[R/J]_{t+1}-1$. Then we argue exactly as in the case $a_5 = \lambda$ at the beginning of the proof of Theorem 4.2, using that the λ-th and $(\lambda-1)$-st powers of linear forms span $[R]_\lambda$ and $[R]_{\lambda-1}$, respectively. In particular, this completes the argument if $a_5 = \lambda$.
- The case where $t = \lambda-1$, $a_5 = \lambda-1$ is the most subtle, and we now give the argument for this case.

We claim that the multiplication $(\times\ell)\colon [R/I]_{\lambda-1} \to [R/I]_\lambda$ is injective. To see this, we will show that $\dim[R/I]_{\lambda-1} < \dim[R/I]_\lambda$ and that $\dim[R/(I,\ell)]_\lambda = \dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}$.

First we compute $\dim[R/(I,\ell)]_\lambda$. The first step (4-8) now becomes

$$\dim[R/(I,\ell)]_\lambda = \dim \mathscr{L}_2(\lambda-2;\, \lambda-a_1, \lambda-a_2, \lambda-a_3, \lambda-a_4, 1).$$

We now consider three cases. First, if $a_1 = a_2 = a_3 = a_4 - 1$ then (4-8) becomes

$$\begin{aligned}\dim[R/(I,\ell)]_\lambda &= \dim \mathscr{L}_2(2a_4-6;\, a_4-3, a_4-3, a_4-3, a_4-4, 1)\\ &= 2a_4-6 = 2a_1-4\end{aligned} \tag{4-11}$$

thanks to Lemma 4.1.

We now assume that we are not in the first case. We obtain

$$A_{1,2} = \left[\frac{-a_1-a_2+a_3+a_4-1}{2}\right]_+ \geq 0 \qquad A_{1,3} = \left[\frac{-a_1+a_2-a_3+a_4-1}{2}\right]_+ \geq 0$$
$$A_{1,4} = \left[\frac{-a_1+a_2+a_3-a_4-1}{2}\right]_+ \qquad A_{2,3} = \left[\frac{a_1-a_2-a_2+a_4-1}{2}\right]_+$$
$$A_{2,4} = \left[\frac{a_1-a_2+a_3-a_4-1}{2}\right]_+ = 0 \qquad A_{3,4} = \left[\frac{a_1+a_2-a_3-a_4-1}{2}\right]_+ = 0.$$

Our second case is $a_2+a_3 > a_1+a_4$. Then $A_{1,2}$, $A_{1,3}$ and $A_{1,4}$ are possibly nonzero. A calculation shows that we then must have

$$\begin{aligned}\dim[R/(I,\lambda)]_\lambda &= \dim \mathscr{L}_2(2a_1-3;\, a_1-1, a_1-2, a_1-2, a_1-2, 1)\\ &= 2a_1-3.\end{aligned} \tag{4-12}$$

Our third case is $a_1+a_4 > a_2+a_3$. Then $A_{1,2}$, $A_{1,3}$ and $A_{2,3}$ are possibly nonzero. Another calculation shows that we must have

$$\begin{aligned}\dim[R/(I,\ell)]_\lambda &= \dim \mathscr{L}_2(a_1+a_2+a_3-a_4-3;\, (\lambda-a_4+1)^3, \lambda-a_4, 1)\\ &= a_1+a_2+a_3-a_4-3.\end{aligned} \tag{4-13}$$

Now we compute the expected dimension $\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}$ (recall $a_5 = \lambda - 1$) by using Lemma 3.7.

In the first case ($a_1 = a_2 = a_3 = a_4 - 1$) we use the Koszul complex and easily compute that the expected dimension is $2a_4 - 6 = 2a_1 - 4$, agreeing with (4-11).

In the second case, using the observations about which $A_{i,j}$ are positive, the Koszul resolution gives the expected dimension $2a_1 - 3$, agreeing with (4-12).

In the third case, again using the observations about the $A_{i,j}$ and the Koszul complex, we obtain the expected dimension $a_1 + a_2 + a_3 - a_4 - 3$, agreeing with (4-13). Thus when $a_5 = \lambda - 1$, R/I has the WLP, concluding the proof of (i).

Now, we prove (ii). Hence, we are assuming that $a_1 + a_4 \geq a_2 + a_3$. In fact, since $a_1 + a_2 + a_3 + a_4$ is odd, we actually have $a_1 + a_4 > a_2 + a_3$. We obtain

$$\begin{aligned}&a_1+a_4-a_2-a_3>0, \quad a_2+a_3-a_1-a_4<0, \quad a_2+a_4-a_1-a_3>0,\\ &a_1+a_3-a_2-a_4<0, \quad a_3+a_4-a_1-a_2>0, \quad a_1+a_2-a_3-a_4<0.\end{aligned} \tag{4-14}$$

Therefore, we get

$$\begin{aligned}
a_5 - a_4 + 1 - \textstyle\sum_{i=1,2,3}[\lambda + 2 - a_i - a_4]_+ &= a_5 - a_4 + 1,\\
a_5 - a_3 + 1 - \textstyle\sum_{i=1,2,4}[\lambda + 2 - a_i - a_3]_+ &= a_5 - a_4 + 2,\\
a_5 - a_2 + 1 - \textstyle\sum_{i=1,3,4}[\lambda + 2 - a_i - a_2]_+ &= a_5 - a_4 + 2,\\
a_5 - a_1 + 1 - \textstyle\sum_{i=2,3,4}[\lambda + 2 - a_i - a_1]_+ &= a_5 - a_4 + 2,\\
2a_5 - \lambda - \textstyle\sum_{1\le i<j\le 4} A_{ij} &= 2(a_5 - a_4 + 2).
\end{aligned}$$

Concluding, we want the dimension of the linear system

$$\mathscr{L}_2(2a_5 - 2a_4 + 4;\, a_5 - a_4 + 2,\, a_5 - a_4 + 2,\, a_5 - a_4 + 2,\, a_5 - a_4 + 1,\, 1).$$

By Lemma 4.1 we have

$$\begin{aligned}
&\dim_k \mathscr{L}_2(2a_5 - 2a_4 + 4;\, (a_5 - a_4 + 2)^3,\, a_5 - a_4 + 1,\, 1)\\
&\qquad \ge \dim_k \mathscr{L}_2(2a_5 - 2a_4 + 4;\, (a_5 - a_4 + 2)^3,\, a_5 - a_4 + 1) - 1\\
&\qquad = \dim_k \mathscr{L}_2(2a_5 - 2a_4 + 4;\, (a_5 - a_4 + 2)^4) + a_5 - a_4 + 1\\
&\qquad = 2a_5 - 2a_4 + 4.
\end{aligned}$$

This is clearly positive. Arguing as in the proof of Theorem 4.2, and using the inequalities (4-14), we compute the expected dimension and we get

$$\begin{aligned}
&[\dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}]_+\\
&\quad = \left[\binom{\lambda+2}{2} - \sum_{i=1}^{4}\binom{\lambda-a_i+2}{2} + \sum_{1\le i<j\le 3}\binom{\lambda-a_j-a_i+2}{2} - \binom{\lambda-a_5+2}{2}\right]_+.
\end{aligned}$$

If the part inside the brackets is negative, the actual dimension clearly exceeds the expected one, and we are done. If not, a straightforward computation shows that

$$\begin{aligned}
&\dim_k[R/(I,\ell)]_\lambda - [\dim_k[R/I]_\lambda - \dim_k[R/I]_{\lambda-1}]\\
&\quad \ge (2a_5 - 2a_4 + 4)\\
&\qquad - \left(\binom{\lambda+2}{2} - \sum_{i=1}^{4}\binom{\lambda-a_i+2}{2} + \sum_{1\le i<j\le 3}\binom{\lambda-a_j-a_i+2}{2} - \binom{\lambda-a_5+2}{2}\right)\\
&\quad = \binom{\lambda-a_5}{2} > 0.
\end{aligned}$$

Thus the actual dimension exceeds the expected dimension, and this completes the proof of (ii).

(iii) We break this into two cases: first we will assume that $a_1 + a_4 + 3 \le a_2 + a_3$, and then we will handle the case $a_1 + a_4 + 1 = a_2 + a_3$.

So to begin, our assumptions now are that $a_1+a_2+a_3+a_4$ is odd, $a_5<\lambda-1$, $a_1+a_4+3\le a_2+a_3$ and $2a_5+3-a_4-a_3-a_2+a_1\ge 0$. Hence, we have

$$\begin{aligned} &a_3+a_4-a_1-a_2>0, \quad a_1+a_2-a_3-a_4<0, \quad a_2+a_4-a_1-a_3>0,\\ &a_1+a_3-a_2-a_4<0, \quad a_2+a_3-a_1-a_4>0, \quad a_1+a_4-a_2-a_3<0. \end{aligned} \tag{4-15}$$

Now we examine the linear system remaining after we remove the one-dimensional base locus. Observe that

$$\begin{aligned} &a_5-a_4+1-\textstyle\sum_{i=1,2,3}[\lambda+2-a_i-a_4]_+ = a_5+1-(a_2+a_3+a_4-a_1-1)/2,\\ &a_5-a_3+1-\textstyle\sum_{i=1,2,4}[\lambda+2-a_i-a_3]_+ = a_5+1-(a_2+a_3+a_4-a_1-1)/2,\\ &a_5-a_2+1-\textstyle\sum_{i=1,3,4}[\lambda+2-a_i-a_2]_+ = a_5+1-(a_2+a_3+a_4-a_1-1)/2,\\ &a_5-a_1+1-\textstyle\sum_{i=2,3,4}[\lambda+2-a_i-a_1]_+ = a_5+2-(a_2+a_3+a_4-a_1-1)/2,\\ &2a_5-\lambda-\textstyle\sum_{1\le i<j\le 4}A_{ij} = 2a_5+4-a_4-a_3-a_2+a_1=:2b. \end{aligned}$$

The additional hypothesis $2a_5+3-a_4-a_3-a_2+a_1\ge 0$ guarantees that the orders of vanishing are positive. Therefore, applying Lemma 4.1, we obtain

$$\begin{aligned} \dim_k[R/(I,\ell)]_\lambda &= \dim_k \mathscr{L}_2(2b;\, b,\, (a_5+1-(a_2+a_3+a_4-a_1-1)/2)^3,\, 1)\\ &\ge 2a_5+5-(a_2+a_3+a_4-a_1)-1. \end{aligned}$$

The hypothesis also guarantees that this value is positive. The computation of the "expected" dimension is very similar to what we did above. The extra hypothesis $a_1+a_4+3\le a_2+a_3$ implies that only $\lambda-a_1-a_2+2$, $\lambda-a_1-a_3+2$ and $\lambda-a_1-a_4+2$ are >0. We again have $a_4<a_1+a_2+a_3$. Thus, Lemma 3.7 provides

$$\begin{aligned} &\big[\dim_k[R/I]_\lambda-\dim_k[R/I]_{\lambda-1}\big]_+\\ &\quad=\big[h_A(\lambda)-h_A(\lambda-a_5)-h_A(\lambda-1)+h_A(\lambda-a_5-1)\big]_+\\ &\quad=\left[\binom{\lambda+2}{2}-\sum_{i=1}^4\binom{\lambda-a_i+2}{2}+\sum_{i=2}^4\binom{\lambda-a_1-a_i+2}{2}-\binom{\lambda-a_5+2}{2}\right]_+. \end{aligned}$$

If this is zero, we are done. Otherwise, a straightforward computation shows that $\dim_k[R/(I,\ell)]_\lambda-[\dim_k[R/I]_\lambda-\dim_k[R/I]_{\lambda-1}]$ is given by

$$\begin{aligned} &2a_5+4-(a_2+a_3+a_4-a_1)\\ &\quad-\left(\binom{\lambda+2}{2}-\sum_{i=1}^4\binom{\lambda-a_i+2}{2}+\sum_{i=2}^4\binom{\lambda-a_1-a_i+2}{2}-\binom{\lambda-a_5+2}{2}\right)\\ &\qquad=\binom{\lambda-a_5}{2}>0. \end{aligned}$$

Thus in either case the actual value of $\dim[R/(I,\ell)]_\lambda$ exceeds the expected dimension. This completes the proof of the case $a_1+a_4+3 \le a_2+a_3$.

Finally, we assume that $a_1+a_4+1=a_2+a_3$. We note first that this assumption actually forces the stronger condition $a_1 \ge 4$ since $a_4 \le a_5 < \lambda - 1$ implies

$$7 < a_1+a_2+a_3-a_4 = 2a_1+1,$$

and hence $a_1 > 3$. The condition $a_1+a_4+1=a_2+a_3$ implies that $\lambda-a_1-a_4+2=0$, so the computation of the expected dimension above can only get smaller, while the computation of $\dim[R/(I,\ell)]_\lambda$ remains unchanged. Thus the difference can only grow, and we again have shown the failure of the WLP. This completes the proof of (iii). □

In the previous results we excluded the case $a_1 = 2$ for the most part. The reason is that then the algebra does have the WLP.

Theorem 4.6. *Let L be a general linear form in $R := k[x_1,\dots,x_4]$, and let I be the ideal $\langle x_1^2, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}\rangle$ of R. Then the algebra R/I has the WLP.*

The proof will be based on a result about almost complete intersections in three variables generated by powers of four general linear forms. According to [Schenck and Seceleanu 2010] such an algebra has the WLP, that is, multiplication by a general linear form has maximal rank. We show that this is also true when one multiplies by the square of such a form.

Proposition 4.7. *Let $\ell_2,\dots,\ell_5,\ell$ be five general linear forms of $S := k[x,y,z]$, and let $\mathfrak{a} \subset S$ be the ideal $\langle \ell_2^{a_2},\dots,\ell_5^{a_5}\rangle$. Set $A := S/\mathfrak{a}$. Then, for each integer j, the multiplication map $\times\ell^2 : [A]_{j-2} \to [A]_j$ has maximal rank.*

Proof. If any of the numbers $a_2,\dots,a_5$ equals one, then A has the SLP by [Harima et al. 2003], so the claim is true. Thus, we may assume $2 \le a_2 \le \cdots \le a_5$.

First, assume also that

$$a_5 \le \frac{a_2+a_3+a_4-3}{2}.$$

Define integers p and b by

$$p := \left\lfloor \frac{a_2+a_3+a_4+a_5-4}{3} \right\rfloor \quad \text{and} \quad a_2+a_3+a_4+a_5 = 3(p+1)+b,$$

thus $1 \le b \le 3$.

According to [Schenck and Seceleanu 2010], A has the WLP and $A/\ell A$ has socle degree p. This implies that multiplication by ℓ on A is injective until degree p and surjective in larger degrees. Symbolically, this reads as

$$[A]_0 \hookrightarrow \cdots \hookrightarrow [A]_{p-1} \hookrightarrow [A]_p \twoheadrightarrow [A]_{p+1} \twoheadrightarrow \cdots. \tag{4-16}$$

Hence, to show our claim it suffices to prove that the multiplication

$$\times \ell^2 : [A]_{p-1} \to [A]_{p+1}$$

has maximal rank, which is equivalent to

$$\dim_k [A/\ell^2 A]_{p+1} = \max\{0, h_A(p+1) - h_A(p-1)\}. \tag{4-17}$$

In order to see this, we first compute the left-hand side and then the right-hand side. Using Theorem 3.3, we get

$$\begin{aligned}\dim_k [A/\ell^2 A]_{p+1} &= \dim_k [S/\langle \ell^2, \ell_2^{a_2}, \ldots, \ell_5^{a_5}\rangle]_{p+1} \\ &= \dim_k [\mathfrak{p}^p \cap \mathfrak{p}_2^{p-a_2+2} \cap \ldots \cap \mathfrak{p}_5^{p-a_5+2}]_{p+1} \\ &= \dim_k \mathscr{L}_2(p+1;\ p, p-a_2+2, \ldots, p-a_5+2),\end{aligned}$$

where $\mathfrak{p}, \mathfrak{p}_2, \ldots, \mathfrak{p}_5$ are the homogenous ideals of five general points in $\mathbb{P}^2$. Let $Q \in S$ be the unique quadric that vanishes at these five points. Again, we use Bézout's theorem to estimate the multiplicity of Q in the base locus of the linear system

$$\mathscr{L}_2(p+1;\ p, p-a_2+2, \ldots, p-a_5+2).$$

For an integer j, the condition

$$\begin{aligned}2(p+1-2j) &\geq p - j + \textstyle\sum_{i=2}^5 [p - a_i + 2 - j] \\ &= 5p + 8 - 5j - \textstyle\sum_{i=2}^5 a_i \\ &= 5p + 8 - 5j - [3(p+1) + b] \\ &= 2p + 5 - 5j - b\end{aligned}$$

is equivalent to $j \geq 3 - b$. It follows that Q appears with multiplicity at least $3 - b$ in the base locus. Thus, we get

$$\begin{aligned}\dim_k [A/\ell^2 A]_{p+1} &= \dim_k \mathscr{L}_2(p+1;\ p, p-a_2+2, \ldots, p-a_5+2) \\ &= \dim_k \mathscr{L}_2(p+2b-5;\ p+b-3, p+b-a_2-1, \\ &\qquad \ldots, p+b-a_5-1).\end{aligned}$$

The latter linear system is clearly empty if $b = 1$. To compute its dimension if $2 \leq b \leq 3$, we consider the lines L_i, $2 \leq i \leq 5$, passing through the points $\mathfrak{p}$ and $\mathfrak{p}_i$. By Bézout's theorem, the line L_i appears with multiplicity at least B_i in the base locus, where

$$B_i := [p + b - 3 + p + b - a_i - 1 - \{p + 2b - 5\}]_+ = [p + 1 - a_i]_+.$$

Our assumption $a_5 \leq (a_2 + a_3 + a_4 - 3)/2$ implies $a_5 \leq p + 1$. Thus, we get $B_i = p + 1 - a_i$, so

$$B_2 + \cdots + B_5 = 4p + 4 - [a_2 + a_3 + a_4 + a_5] = 4p + 4 - [3(p+1) + b] = p + 1 - b.$$

Removing the lines from the base locus we obtain

$$\begin{aligned}\dim_k [A/\ell^2 A]_{p+1} &= \dim_k \mathscr{L}_2(p + 2b - 5;\ p + b - 3, p + b - a_2 - 1, \\ &\qquad \ldots, p + b - a_5 - 1) \\ &= \dim_k \mathscr{L}_2(3b - 6;\ 2b - 4, b - 2, \ldots, b - 2).\end{aligned}$$

It follows that $\dim_k [A/\ell^2 A]_{p+1} = 1$ if $b = 2$.

If $b = 3$, then we get, using Theorem 3.3 again,

$$\begin{aligned}\dim_k [A/\ell^2 A]_{p+1} &= \dim_k \mathscr{L}_2(3;\ 2, 1, \ldots, 1) \\ &= \dim_k [S/\langle \ell^2, \ell_2^3, \ldots, \ell_5^3 \rangle]_3 \\ &= \dim_k [S/\langle \ell^2 \rangle]_3 - 4 \\ &= 3,\end{aligned}$$

because the linear forms are general. Summarizing, we have shown so far that

$$\dim_k [A/\ell^2 A]_{p+1} = \begin{cases} 0 & \text{if } b = 1, \\ 1 & \text{if } b = 2, \\ 3 & \text{if } b = 3. \end{cases} \tag{4-18}$$

Now we compute the right-hand side of (4-17). To this end consider the ring $B := S/\langle \ell_3^{a_3}, \ell_4^{a_4}, \ell_5^{a_5} \rangle$. Observe that

$$\begin{aligned}p + 1 - a_3 - a_4 &= \left\lfloor \frac{a_2 + a_3 + a_4 + a_5 - 1}{3} \right\rfloor - a_3 - a_4 \\ &\leq \frac{a_2 + a_5 - 1 - 2a_3 - 2a_4}{3} \\ &\leq \frac{3a_2 - 5 - 3a_3 - 3a_4}{6} \\ &< 0,\end{aligned}$$

where we used again $a_5 \leq (a_2 + a_3 + a_4 - 3)/2$. Hence, the Koszul resolution of the complete intersection B provides for its Hilbert function if $j \leq p + 1$ that

$$h_B(j) = \binom{j+2}{2} - \sum_{i=3}^{5} \binom{j+2-a_i}{2}, \tag{4-19}$$

where, as above, we define a binomial coefficient $\binom{c}{2}$ to be zero if $c < 0$. Since the complete intersection B has the SLP and since $A \cong B/\ell_2^{a_2}B$, we get

$$h_A(p+1) - h_A(p-1) = \big[h_B(p+1) - h_B(p+1-a_2)\big]_+ - \big[h_B(p-1) - h_B(p-1-a_2)\big]_+.$$

One easily checks that our assumptions provide $p+1 \le \lceil (a_3+a_4+a_5-3)/2 \rceil$. Since $a_3+a_4+a_5-3$ is the socle degree of B, this implies that $h_B(j-1) \le h_B(j)$ if $j \le p+1$. Hence the last formula simplifies to

$$h_A(p+1) - h_A(p-1) = h_B(p+1) - h_B(p+1-a_2) - h_B(p-1) + h_B(p-1-a_2).$$

The socle degree of $A/\ell A$ is at most the socle degree of $S/\langle \ell, \ell_2^{a_2}, \ell_3^{a_3}\rangle$, thus $p \le a_2 + a_3 - 2$. Combining with (4-19) and using that

$$\binom{j+1}{2} - \binom{j-1}{2} = 2j-1$$

if $j \ge 1$, this provides

$$\begin{aligned} h_A(p+1) - h_A(p-1) &= \binom{p+3}{2} - \binom{p+1}{2} - \sum_{i=3}^{5} \left[\binom{p+3-a_a}{2} - \binom{p+1-a_i}{2}\right] \\ &\qquad - \left[\binom{p+3-a_2}{2} - \binom{p+1-a_2}{2}\right] \\ &= -6p - 9 + 2\sum_{i=2}^{4} a_i \\ &= 2b - 3. \end{aligned}$$

Hence, we get

$$[h_A(p+1) - h_A(p-1)]_+ = [2b-3]_+ = \begin{cases} 0 & \text{if } b = 1, \\ 1 & \text{if } b = 2, \\ 3 & \text{if } b = 3. \end{cases}$$

Comparing with (4-18), this establishes the desired equality (4-17).

It remains to consider the case $a_5 > (a_2 + a_3 + a_4 - 3)/2$. Let us call $\mathfrak{b} := \langle \ell_2^{a_2}, \ell_3^{a_3}, \ell_4^{a_4}\rangle$ and set $B := S/\mathfrak{b}$. Note that the socle degree of B is $a_2+a_3+a_4-3$. Stanley [1980], Watanabe [1987] and Reid, Roberts and Roitman [1991] showed that, in characteristic 0, SLP holds for an artinian complete intersection generated by powers of linear forms. In particular, for each integer j, the multiplication map $\times \ell^2 : [B]_{j-2} \to [B]_j$ has maximal rank. We want to prove that, for each integer j, the multiplication map $\times \ell^2 : [A]_{j-2} \to [A]_j$ has maximal rank. To this end, we will examine several cases:

(a) Assume $a_2 + a_3 + a_4$ odd.

(a1) For any $j < a_5$, we have $[A]_j \cong [B]_j$ and hence $\times \ell^2 : [A]_{j-2} \to [A]_j$ has maximal rank.

(a2) For any $j \geq a_5 + 1$ or $j = a_5$ and $a_5 > (a_2 + a_3 + a_4 - 3)/2 + 1$, we have $j - 2 \geq (a_2 + a_3 + a_4 - 3)/2$. Hence, the surjectivity $\times \ell^2 : [B]_{j-2} \twoheadrightarrow [B]_j$ together with the commutative diagram

$$\begin{array}{ccc} [B]_{j-2} & \twoheadrightarrow & [A]_{j-2} \\ \downarrow & & \downarrow \\ [B]_j & \twoheadrightarrow & [A]_j \end{array}$$

allows us to conclude that $\times \ell^2 : [A]_{j-2} \to [A]_j$ has maximal rank.

(a3) For $j = a_5 = (a_2 + a_3 + a_4 - 3)/2 + 1$ we have $[A]_{j-2} \cong [B]_{j-2} \cong [B]_j \twoheadrightarrow [A]_j$. Therefore, we also conclude that $\times \ell^2 : [A]_{j-2} \to [A]_j$ has maximal rank.

(b) Assume $a_2 + a_3 + a_4$ even.

(b1) For any $j < a_5$, we have $[A]_j \cong [B]_j$ and hence $\times \ell^2 : [A]_{j-2} \to [A]_j$ has maximal rank.

(b2) It is identical to (a2).

(b3) For $j = a_5 = (a_2 + a_3 + a_4 - 2)/2$ we have to prove that $\times \ell^2 : [A]_{j-2} \to [A]_j$ is injective. Since $\dim[B]_{j-2} < \dim[B]_j$, the image of $[B]_{j-2}$ in $[B]_j$ under multiplication by ℓ^2 is not surjective. So the vector space $[\langle \ell_2^{a_2}, \ell_3^{a_3}, \ell_4^{a_4}, \ell^2 \rangle]_j$ is not all of S_j. But in characteristic zero the j-th powers of linear forms span $[S]_j$. Thus for a general L, the image of $L^j = L^{a_5}$ in $[B]_j$ is outside the image of $[B]_{j-2}$ in $[B]_j$. Thus $(\times \ell^2) : [A]_{j-2} \to [A]_j$ is injective, since it is for B in that degree. This completes the proof of the proposition. □

Remark 4.8. Observe that extensions of Proposition 4.7 to multiplication by higher powers of ℓ fail in general. There are many such examples. The smallest is when $a_2 = \cdots = a_5 = 3$, for which multiplication by ℓ^3 fails to have maximal rank from degree 1 to degree 4. It is easy to see that this extends to the case $a_2 = \cdots = a_5 = d$, for which multiplication by ℓ^d fails to have maximal rank from degree $d - 2$ to degree $2d - 2$. Many more complicated examples (produced by CoCoA) exist as well.

Proof of Theorem 4.6. If one of the numbers $a_2, \ldots, a_5$ is one, then the result follows by [Schenck and Seceleanu 2010]. Thus, we may assume $2 \leq a_2 \leq \cdots \leq a_5$.

If $a_5 > (a_2 + a_3 + a_4 - 3)/2$, then R/I has the WLP by Theorems 4.2(i) and 4.5(i). Thus, we may assume $a_5 \leq (a_2 + a_3 + a_4 - 3)/2$ for the remainder of the proof. Let $\ell \in R$ be another general linear form. We have to show that for all integers j,

$$\dim_k [R/(I, \ell)]_j = \max\{0, h_{R/I}(j) - h_{R/I}(j-1)\}. \tag{4-20}$$

To this end, consider the ideal $J := \langle x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, L^{a_5}\rangle \subset R$. The complete intersection R/J has the SLP, which implies

$$\begin{aligned}&\big[h_{R/I}(j) - h_{R/I}(j-1)\big]_+ \\ &\qquad = \Big[\big[h_{R/J}(j) - h_{R/J}(j-2)\big]_+ - \big[h_{R/J}(j-1) - h_{R/J}(j-3)\big]_+\Big]_+.\end{aligned}$$

The complete intersection R/J has socle degree $a_2 + \cdots + a_5 - 4 \geq 4$, thus the multiplication $\times \ell : [R/J]_{j-1} \to [R/J]_j$ is injective if $j \leq \lceil (a_2+a_3+a_4+a_5-4)/2 \rceil$. One checks that

$$\left\lceil \frac{a_2+a_3+a_4+a_5-4}{2} \right\rceil \geq \left\lfloor \frac{a_2+a_3+a_4+a_5-1}{3} \right\rfloor =: p+1.$$

It follows that $h_{R/J}(j-1) \leq h_{R/J}(j)$ if $j \leq p+1$. Thus, if $j \leq p+1$, then we get that

$$\begin{aligned}&h_{R/I}(j) - h_{R/I}(j-1) \\ &\qquad = h_{R/J}(j) - h_{R/J}(j-2) - h_{R/J}(j-1) + h_{R/J}(j-3) \\ &\qquad = \big[h_{R/J}(j) - h_{R/J}(j-1)\big]_+ - \big[h_{R/J}(j-2) - h_{R/J}(j-3)\big]_+ \\ &\qquad = h_{R/(J,\ell)}(j) - h_{R/(J,\ell)}(j-2),\end{aligned}$$

where we used again that R/J has the SLP. Invoking Proposition 4.7, we obtain

$$\big[h_{R/I}(j) - h_{R/I}(j-1)\big]_+ = h_{R/(I,\ell)}(j)$$

if $j \leq p+1$. This equality is also true (using the same computations) if $h_{R/J}(p+1) \leq h_{R/J}(p+2)$. Otherwise, we get $\big[h_{R/I}(p+2) - h_{R/I}(p+1)\big]_+ = 0$. However, using (4-16), we get $[R/(I,\ell)]_{p+2} = 0$. Hence, we have in any case

$$\big[h_{R/I}(p+2) - h_{R/I}(p+1)\big]_+ = h_{R/(I,\ell)}(p+2) = 0,$$

which completes the argument. □

5. Almost uniform powers of linear forms in 5 variables

In this section, we let $R = k[x_1, x_2, x_3, x_4, x_5]$, where k is a field of characteristic zero, and we will apply the approach described in Section 3 and results on fat points in $\mathbb{P}^3$ to determine exactly when an ideal generated by uniform powers of six general linear forms in R fails the WLP. Some nonuniform cases are also discussed. To this end, without loss of generality we set $I = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, x_5^{a_5}, L^{a_6}\rangle$, where L is a general linear form and $a_1 \leq a_2 \leq a_3 \leq a_4 \leq a_5 \leq a_6$.

Theorem 5.1. *Let L be a general linear form and let $I = \langle x_1^d, x_2^d, x_3^d, x_4^d, x_5^d, L^d\rangle$. Then the ring R/I fails the WLP if and only if $d > 3$.*

Proof. Using CoCoA we check that if $d < 4$ then R/I has the WLP. Assume $d \geq 4$ and we will show that R/I fails the WLP in degree $2d-1$. To this end, we take $\ell \in R$ a general linear form. According to (3-1), it is enough to show that

$$\dim[R/(I,\ell)]_{2d-1} > \big[\dim[R/I]_{2d-1} - \dim[R/I]_{2d-2}\big]_+. \tag{5-1}$$

We will compute the left-hand side using Proposition 3.4 and the right-hand side using the fact that R/J has the SLP.

Now let $S = k[x_1, x_2, x_3, x_4]$. We want to compute the vector space dimension

$$\dim_k\big[\mathfrak{q}_1^d \cap \mathfrak{q}_2^d \cap \mathfrak{q}_3^d \cap \mathfrak{q}_4^d \cap \mathfrak{q}_5^d \cap \mathfrak{q}_6^d\big]_{2d-1} = \dim_k \mathscr{L}_3(2d-1; d^6). \tag{5-2}$$

Applying a sequence of cubo-cubic Cremona transformations

$$\begin{array}{ccc} \mathbb{P}^3 & \dashrightarrow & \mathbb{P}^3 \\ (x_1 : x_2 : x_3 : x_4) & \dashrightarrow & (x_1^{-1}, x_2^{-1}, x_3^{-1}, x_4^{-1}) = (x_2x_3x_4 : x_1x_3x_4 : x_1x_2x_4 : x_1x_2x_3) \end{array}$$

we will transform the last linear system to another one which has the same dimension, but it will be nonspecial and hence we will be able to compute its dimension. In fact, we apply Lemma 3.5 and we get

$$\begin{aligned} \dim \mathscr{L}_3(2d-1; d^6) &= \dim \mathscr{L}_3(2d-3; d^2, (d-2)^4) \\ &= \dim \mathscr{L}_3(2d-5; (d-2)^4, (d-4)^2) \\ &= \dim \mathscr{L}_3(2d-7; (d-4)^6). \end{aligned} \tag{5-3}$$

Since $2(2d-7) \geq 4(d-4)$, the linear system $\mathscr{L}_3(2d-7; (d-4)^6)$ is in standard form. Therefore, Theorem 3.6 provides that it is nonspecial and its dimension is given by

$$\dim \mathscr{L}_3\big(2d-7; (d-4)^6\big) = \binom{2d-4}{3} - 6\binom{d-2}{3}. \tag{5-4}$$

Now we have to compute the right-hand side of (5-1). Let $A = R/J$, where $J = \langle x_1^d, x_2^d, x_3^d, x_4^d, x_5^d\rangle$. Since R/J has the SLP, we have

$$0 \leq h_{R/I}(2d-1) = h_A(2d-1) - h_A(d-1).$$

Therefore,

$$\begin{aligned} &[\dim[R/I]_{2d-1} - \dim[R/I]_{2d-2}] \\ &\qquad = h_A(2d-1) - h_A(d-1) - h_A(2d-2) + h_A(d-2) \\ &\qquad = \binom{2d+2}{3} - 6\binom{d+2}{3}. \end{aligned}$$

We easily verify that

$$\dim[R/(I,\ell)]_{2d-1} - [\dim[R/I]_{2d-1} - \dim[R/I]_{2d-2}] = 4$$

and this shows that R/I fails the WLP in degree $2d-1$, which is what we wanted to prove. □

Remark 5.2. Note that (5-4) could also have been proven using Proposition 3.5 of [Catalisano et al. 1999]. However, we use our approach because it also applies to setups below where the hypotheses of the latter proposition are not satisfied.

There are several possible extensions of the above theorem. First, we can ask whether the WLP property holds for the case of nonuniform powers and, in particular, we can ask what happens in the almost uniform case. We have

Theorem 5.3. *Let L be a general linear form and let $I = \langle x_1^d, x_2^d, x_3^d, x_4^d, x_5^d, L^{d+e}\rangle$ with $e \geq 1$. Then*:

(i) *If d is odd, then R/I has the WLP if and only if $e \geq (3d-5)/2$.*

(ii) *If d is even, R/I has the WLP if and only if $e \geq (3d-8)/2$.*

Proof. Set $\lambda := \lfloor (5d-5)/2 \rfloor$. We will actually prove the following sequence of statements:

(i′) If $d+e \geq \lambda$, then R/I has the WLP.

(ii′) If d is odd and $e \leq d-2$, then R/I fails the WLP.

(iii′) If d is even and $e \leq d-3$, then R/I fails the WLP.

(iv′) If d is odd and $d-1 \leq e \leq (3d-7)/2$ then R/I fails the WLP.

(v′) If d is even and $d-2 \leq e \leq (3d-10)/2$ then R/I fails the WLP.

(vi′) If d is even and $e = (3d-8)/2$ (that is $d+e = \lambda - 1$) then R/I has the WLP.

Throughout this proof we will denote $A = R/J$, where $J = \langle x_1^d, x_2^d, x_3^d, x_4^d, x_5^d\rangle$.

(i′) The proof is the same as in Theorems 4.2(i) and 4.5(i).

(ii′) Since $d+e \leq 2d-2$, we can write $e = 2j+\epsilon$ with $0 \leq \epsilon \leq 1$ and $j \leq (d-3)/2$. We will show that R/I fails the WLP in degree $2d-1+j$. To this end, we take $\ell \in R$, a general linear form. According to (3-1), it is enough to show that

$$\dim[R/(I,\ell)]_{2d-1+j} > [\dim[R/I]_{2d-1+j} - \dim[R/I]_{2d-2+j}]_+. \tag{5-5}$$

We will compute the left-hand side using Proposition 3.4 and the right-hand side using Lemma 3.7.

We begin by computing the vector space dimension

$$\dim_k[\mathfrak{q}_1^{d+j} \cap \mathfrak{q}_2^{d+j} \cap \mathfrak{q}_3^{d+j} \cap \mathfrak{q}_4^{d+j} \cap \mathfrak{q}_5^{d+j} \cap \mathfrak{q}_6^{d+j-e}]_{2d-1+j} = \dim_k \mathscr{L}_3(2d-1+j; (d+j)^5, d+j-e).$$

Applying a sequence of cubo-cubic Cremona transformations (see Lemma 3.5), we get

$$\begin{aligned}
&\dim_k \mathscr{L}_3(2d-1+j;\ (d+j)^5, d+j-e)\\
&\quad = \dim_k \mathscr{L}_3(2d-3-j;\ d+j, d+j-e), (d-j-2)^4\\
&\quad = \dim_k \mathscr{L}_3(2d-5-3j+e;\ d-j-2+e, (d-j-2)^3, (d-3j-4+e)^2)\\
&\quad = \dim_k \mathscr{L}_3(2d-7-5j+2e;\ d-3j-4+2e, (d-3j-4+e)^5).
\end{aligned}$$

(Here we use the hypothesis $d+e \le 2d-2$ to guarantee that $d-3j-4+2e \ge 0$.)

Since $2(2d-7-5j+2e) \ge 3(d-3j-4+e)+d-3j-4+2e$, the linear system

$$\mathscr{L}_3(2d-7-5j+2e; d-3j-4+2e, (d-3j-4+e)^5)$$

is in standard form. Therefore, Theorem 3.6 provides

$$\begin{aligned}
&\dim_k \mathscr{L}_3(2d-7-5j+2e; (d-3j-4+e)^5, d-3j-4+2e)\\
&= \left[\binom{2d-4-5j+2e}{3} - 5\binom{d-2-3j+e}{3} - \binom{d-2-3j+2e}{3}\right]_+ + 5\cdot\binom{e-j}{3}.
\end{aligned}$$

We claim that

$$\begin{aligned}
&\binom{2d-4-5j+2e}{3} - 5\binom{d-2-3j+e}{3} - \binom{d-2-3j+2e}{3} + 5\cdot\binom{e-j}{3}\\
&\quad = \frac{1}{6}(d-2-2j)\left[2d^2-3e^2+12e(1+j)-4(3+8j+4j^2)+d(3e-2-2j)\right]
\end{aligned}$$

is positive. In fact, using $d \ge 2j+3$ and $e \ge 2e$, one gets

$$\begin{aligned}
&2d^2-3e^2+12e(1+j)-4(3+8j+4j^2)+d(3e-2-2j)\\
&\ge 2(2j+3)^2-3e^2+12e(1+j)-4(3+8j+4j^2)+(2j+3)(3e-2-2j) > 0.
\end{aligned}$$

We conclude that $\dim[R/(I,\ell)]_{2d-1+j} > 0$.

Now we compute the right-hand side of (5-5). Lemma 3.7 provides

$$\begin{aligned}
&\dim[R/I]_{2d-1+j} - \dim[R/I]_{2d-2+j}\\
&\quad = h_A(2d-1+j) - h_A(d-1+j-e) - h_A(2d-2+j) + h_A(d-2+j-e)\\
&\quad = \binom{2d+j+2}{3} - 5\binom{d+j+2}{3} - \binom{d+j+2-e}{3} + 10\binom{j+2}{3}.
\end{aligned}$$

If $\dim[R/I]_{2d-1+j} - \dim[R/I]_{2d-2+j} \le 0$, then the WLP fails because we have seen that $\dim[R/(I,\ell)]_{2d-1+j} > 0$.

If $\dim[R/I]_{2d-1+j} - \dim[R/I]_{2d-2+j} > 0$, then

$$\begin{aligned}
&\dim[R/(I,\ell)]_{2d-1+j} - [\dim[R/I]_{2d-1+j} - \dim[R/I]_{2d-2+j}] \\
&\quad \geq \binom{2d-4-5j+2e}{3} - 5\binom{d-2-3j+e}{3} - \binom{d-2-3j+2e}{3} + 5\cdot\binom{e-j}{3} \\
&\qquad - \Big[\binom{2d+j+2}{3} - 5\binom{d+j+2}{3} - \binom{d+j+2-e}{3} + 10\binom{j+2}{3}\Big] \\
&\quad = \binom{3j-e+4}{3} > 0.
\end{aligned}$$

Hence, we conclude in every case that R/I fails the WLP in degree $2d-1+j$, which is what we wanted to prove.

(iii′) The proof is completely analogous.

(iv′) Suppose that d is odd and $d-1 \leq e \leq (3d-7)/2$. We claim that R/I fails the WLP (usually by failing injectivity) from degree $(5d-7)/2$ to degree $(5d-5)/2$. We first consider (by applying a sequence of cubo-cubic Cremona transformations; see Lemma 3.5)

$$\begin{aligned}
\dim[R/(I,\ell)]_{(5d-5)/2} &= \dim_k \mathscr{L}\Big(\frac{5d-5}{2};\ \Big(\frac{3d-3}{2}\Big)^5,\ \frac{3d-3}{2}-e\Big) \\
&= \dim_k \mathscr{L}_3\Big(\frac{3d-3}{2};\ \Big(\frac{d-1}{2}\Big)^4,\ \frac{3d-3}{2},\ \frac{3d-3}{2}-e\Big) \\
&= \dim_k \mathscr{L}_3\Big(\frac{3d-3}{2};\ \frac{3d-3}{2},\ \Big(\frac{d-1}{2}\Big)^4,\ \frac{3d-3}{2}-e\Big) \\
&= \dim_k \mathscr{L}_2\Big(\frac{3d-3}{2};\ \Big(\frac{d-1}{2}\Big)^4,\ \frac{3d-3}{2}-e\Big),
\end{aligned}$$

the last step being a consequence of Proposition 3.4. Notice that since $d-1 \leq e$, we have $(3d-3)/2 - e \leq (d-1)/2$. Thanks to Lemma 4.1, these fat points impose independent conditions, so we have

$$\begin{aligned}
\dim[R/(I,\ell)]_{(5d-5)/2} &= \binom{\frac{3d-3}{2}+2}{2} - 4\cdot\binom{\frac{d-1}{2}+1}{2} - \binom{\frac{3d-3}{2}-e+1}{2} \\
&= \frac{9d^2-1}{8} - 4\cdot\frac{d^2-1}{8} - \frac{1}{2}\Big[\frac{9d^2-12d+3}{4} - e(3d-2) + e^2\Big] \\
&= \frac{e}{2}(3d-2-e) - \binom{d-1}{2} + 1
\end{aligned}$$

(The fact that

$$\dim_k \mathscr{L}_3\Big(\frac{3d-3}{2};\ \frac{3d-3}{2},\ \frac{d-1}{2},\ \frac{d-1}{2},\ \frac{d-1}{2},\ \frac{d-1}{2},\ \frac{3d-3}{2}-e\Big) = \frac{e}{2}(3d-2-e) - \binom{d-1}{2} + 1$$

could also be obtained via Theorem 3.6.)

Combined with Lemma 3.7, we obtain

$$\dim[R/(I,\ell)]_{(5d-5)/2} - \big(\dim[R/I]_{(5d-5)/2} - \dim[R/I]_{((5d-5)/2)-1}\big) = \binom{\frac{3d-7}{2}-e+3}{3}.$$

If the part in parentheses is nonnegative then we expect injectivity, but the positivity of the part on the right (since we assumed $e \le (3d-7)/2$) implies that the WLP fails. Now suppose that the part in parentheses is negative, so that we expect surjectivity. Then one checks that

$$\begin{aligned}\dim[R/(I,\ell)]_{(5d-5)/2} &= \frac{e}{2}(3d-2-e) - \binom{d-1}{2} + 1\\ &= \frac{(e-d)(2d-e-2)}{2} + \binom{d+1}{2}\\ &\ge \frac{(e-d)(d+3)}{4} + \binom{d+1}{2} \qquad (\text{since } e \le \frac{3d-7}{2})\\ &> 0.\end{aligned}$$

(the last inequality is for $d-1 \le e \le (3d-7)/2$), and so surjectivity, and hence WLP, fails. This completes the proof of (iv′).

(v′) We show that R/I fails the WLP from degree $(5d-8)/2$ to degree $(5d-6)/2$. Since we will use part of the computations also for (vi′), we consider all even d such that $d-2 \le e \le (3d-8)/2$. In the same way as above we get

$$\begin{aligned}\dim[R/(I,\ell)]_{(5d-6)/2} &= \dim_k \mathscr{L}_3\Big(\frac{5d-6}{2};\ \Big(\frac{3d-4}{2}\Big)^5, \frac{3d-4}{2} - e\Big)\\ &= \dim_k \mathscr{L}_3\Big(\frac{3d-2}{2};\ \Big(\frac{d}{2}\Big)^4, \frac{3d-4}{2}, \frac{3d-4}{2} - e\Big).\end{aligned}$$

Using Theorem 3.6, we obtain

$$\dim[R/(I,\ell)]_{(5d-6)/2} = -d^2 + 3de - e^2 + 6d - 4e - 4 = \frac{5d^2}{4} - \Big(\frac{3d-4}{2} - e\Big)^2,$$

which is at least d^2 in the given range for e. In particular, we are finished whenever $\dim[R/I]_{(5d-6)/2} \le \dim[R/I]_{((5d-6)/2)-1}$.

If $\dim[R/I]_{(5d-6)/2} \ge \dim[R/I]_{((5d-6)/2)-1}$ (so we expect injectivity), then Lemma 3.7 provides, minding the bound on e,

$$\dim[R/(I,\ell)]_{(5d-6)/2} - \big(\dim[R/I]_{(5d-6)/2} - \dim[R/I]_{((5d-6)/2)-1}\big) = \binom{\frac{3d-10}{2}-e+3}{3}. \quad (5\text{-}6)$$

Since we are assuming that the part in parentheses is nonnegative, the positivity of the part on the right implies that WLP fails if $e \le (3d-10)/2$, as claimed.

(vi′) Arguing as in Theorem 4.5(iii), we see that if $t \leq \lambda - 2$ or $t \geq \lambda$ then $(\times \ell) : [R/I]_t \to [R/I]_{t+1}$ has maximal rank. So, it only remains to study the case $t = \lambda - 1$. We are going to prove that the multiplication $(\times \ell) : [R/I]_{\lambda-1} \to [R/I]_\lambda$ is injective, that is $\dim[R/(I, \ell)]_\lambda = \dim[R/I]_\lambda - \dim[R/I]_{\lambda-1}$.

Notice that the assumptions force $d \geq 4$, so we may apply the computations of (v′). Thus, the desired follows by (5-6). □

The same argument gives us the following result.

Proposition 5.4. *Let L be a general linear form and let*

$$I = \langle x_1^{a_1}, x_2^{a_2}, x_3^{a_3}, x_4^{a_4}, x_5^{a_5}, L^{a_6} \rangle.$$

Assume that $5 \leq a_1 \leq a_2 \leq a_3 \leq a_4 \leq a_5 \leq a_6 \leq a_1 + 2$. Then the ring R/I fails the WLP.

Our methods extend beyond the results mentioned above.

Example 5.5. Using the above notation, one shows:

(i) If $d \geq 4$ and $(a_1, a_2, a_3, a_4, a_5, a_6) = (d, d+1, d+2, d+3, d+4, d+5)$, then the ring R/I fails the WLP in degree $2d + 4$.

(ii) If $d \geq 4$ and $(a_1, a_2, a_3, a_4, a_5, a_6) = (d, d+3, d+4, d+7, d+7, d+10)$, then the ring R/I fails the WLP in degree $2d + 9$.

6. Uniform powers of linear forms

In this section, we consider the case of an almost complete intersection of general linear forms whose generators all have the same degree d. We give a complete answer, in the case of an even number of variables, to the question of when the WLP holds. Interestingly, the case of an odd number of variables is more delicate, and we are only able to give a partial result, concluding with a conjecture.

We first consider the case where there are an even number of variables. Set $R = k[x_1, \ldots, x_{2n}]$, where k is a field of characteristic zero. We determine when an ideal generated by uniform powers of $2n + 1$ general linear forms in R fails the WLP. If $n = 1$, then R/I always has the WLP due to [Harima et al. 2003]. If $n = 2$, then R/I fails the WLP if and only if $d \geq 3$ by Theorem 4.2.

Theorem 6.1. *Let $L \in R$ be a general linear form, and let $I = \langle x_1^d, \ldots, x_{2n}^d, L^d \rangle$, where $n \geq 3$. Then the ring R/I fails the WLP if and only if $d > 1$.*

Proof. It is clear that for $d = 1$, R/I has WLP.

Assume $d \geq 2$. We will show that R/I fails WLP in degree $nd - n$. To this end, we take $\ell \in R$, a general linear form. According to (3-1), it is enough to show that

$$\dim_k[R/(I, \ell)]_{nd-n} > \left[\dim_k[R/I]_{nd-n} - \dim_k[R/I]_{nd-n-1}\right]_+. \tag{6-1}$$

First, we compute the left-hand side of (6-1).

Claim 1. $\dim[R/(I,\ell)]_{nd-n} = 1$.

Proof. By Proposition 3.4 and Theorem 3.3, we have

$$\begin{aligned}&\dim_k[R/(I,\ell)]_{nd-n}\\ &\quad= \dim_k[\mathfrak{q}_1^{(n-1)d-(n-1)} \cap \mathfrak{q}_2^{(n-1)d-(n-1)} \cap \cdots \cap \mathfrak{q}_{2n}^{(n-1)d-(n-1)} \cap \mathfrak{q}_{2n+1}^{(n-1)d-(n-1)}]_{nd-n}\\ &\quad= \dim_k \mathscr{L}_{2n-2}(nd-n; ((n-1)d-(n-1))^{2n+1}).\end{aligned}$$

Applying Lemma 3.5, we get

$$\begin{aligned}&\dim_k \mathscr{L}_{2n-2}(nd-n; ((n-1)d-(n-1))^{2n+1})\\ &= \dim_k \mathscr{L}_{2n-2}((n-1)d-(n-1); ((n-1)d-(n-1))^2, ((n-2)d-(n-2))^{2n-1}).\end{aligned}$$

Using Proposition 3.4 twice, it follows that

$$\begin{aligned}&\dim_k \mathscr{L}_{2n-2}(nd-n; ((n-1)d-(n-1))^{2n+1})\\ &\quad= \dim_k \mathscr{L}_{2n-3}((n-1)d-(n-1); (n-1)d-(n-1), ((n-2)d-(n-2))^{2n-1})\\ &\quad= \dim_k \mathscr{L}_{2n-4}((n-1)d-(n-1); ((n-2)d-(n-2))^{2n-1}). \qquad (6\text{-}2)\end{aligned}$$

If $n = 3$, then we get by applying again Lemma 3.5

$$\dim_k \mathscr{L}_2(2d-2; (d-1)^5) = \dim_k \mathscr{L}_2(2d-2; (d-1)^2, 0^3) = 1,$$

as desired.

If $n \geq 4$, then we conclude by induction using (6-2). Thus, Claim 1 is shown.

Next we consider the right-hand side of inequality (6-1). Taking into account Claim 1, we see that R/I fails the WLP, once we have shown the following:

Claim 2. $\dim_k[R/I]_{nd-n} \leq \dim_k[R/I]_{nd-n-1}$.

Proof. We use induction on $n \geq 3$. Let $A = R/\langle x_1^d, \ldots, x_{2n}^d\rangle$. Assume $n = 3$. Lemma 3.7 provides

$$\begin{aligned}&\dim_k[R/I]_{3d-3} - \dim_k[R/I]_{3d-4}\\ &\qquad= h_A(3d-3) - h_A(2d-3) - h_A(3d-4) + h_A(2d-4)\\ &\qquad= \binom{3d+1}{4} - 7\binom{2d+1}{4} + 21\binom{d+1}{4}\\ &\qquad= -\frac{1}{12}d(d-2)(5d^2+2d+5)\\ &\qquad\leq 0,\end{aligned}$$

as desired.

Let $n \geq 4$. By Lemma 3.7, Claim 2 can be rewritten as

$$h_A(nd-n) - h_A(nd-n-1) \leq h_A(nd-n-d) - h_A(nd-n-d-1). \qquad (6\text{-}3)$$

Consider now the ring $B = k[x_1, \dots, x_{2n-1}]/\langle x_1^d, \dots, x_{2n-1}^d \rangle$. Then $A \cong B \otimes_k k[x]/(x^d)$, which implies

$$h_A(j) = h_B(j) + h_B(j-1) + \cdots + h_B(j-(d-1)). \tag{6-4}$$

Thus, the last inequality becomes

$$h_B(nd-n) - h_B(nd-n-d) \le h_B(nd-n-d) - h_B(nd-n-2d).$$

The Hilbert function of B is symmetric about $\frac{1}{2}(2n-1)(d-1)$, so $h_B(nd-n) = h_B(nd-n-d+1)$. Thus, we have to show

$$h_B(nd-n-d+1) - h_B(nd-n-d) \le h_B(nd-n-d) - h_B(nd-n-2d).$$

To this end, put $C = k[x_1, \dots, x_{2n-2}]/\langle x_1^d, \dots, x_{2n-2}^d \rangle$. Then $B \cong C \otimes_k k[x]/(x^d)$. Hence, using a relation similar to (6-4), Claim 2 follows, once we have shown

$$\begin{aligned} &h_C(nd-n-d+1) - h_C(nd-n-2d+1) \\ &\quad \le h_C(nd-n-d) + h_C(nd-n-d) + \cdots + h_C(nd-n-2d+1) \\ &- \big[h_C(nd-n-2d) + h_C(nd-n-2d-1) + \cdots + h_C(nd-n-3d+1)\big] \end{aligned} \tag{6-5}$$

Our induction hypothesis (see inequality (6-3)) provides

$$h_C(nd-n-d+1) - h_C(nd-n-d) \le h_C(nd-n-2d+1) - h_C(nd-n-2d).$$

Since the Hilbert function of C is unimodal with peak in degree $(n-1)(d-1)$, we have the estimates:

$$\begin{aligned} h_C(nd-n-d) - h_C(nd-n-2d+1) & \\ &\le h_C(nd-n-d) - h_C(nd-n-2d-1), \\ 0 &\le h_C(nd-n-d-1) - h_C(nd-n-2d-2), \\ &\vdots \\ 0 &\le h_C(nd-n-2d) - h_C(nd-n-3d-1). \end{aligned}$$

Adding the last inequalities, we get the desired inequality (6-5), which completes the proof of Claim 2, and we are done. □

Remark 6.2. Theorem 6.1 proves half of Conjecture 5.5.2 in [Harbourne et al. 2011] — namely, the case when the number of variables is even.

Remark 6.3. In [Harbourne et al. 2011, Theorem 5.2.2], Harbourne, Schenck and Seceleanu have recently given an alternative proof of the above theorem for $d \gg 0$.

Claim 2 can be restated as a result about the growth of the coefficients of a certain univariate polynomial.

Proposition 6.4. *Let $n \geq 3$ and $d \geq 1$ be integers. For the univariate polynomial*

$$a_0 + a_1 z + \cdots + a_{2nd} z^{2nd} := \left(1 + z + \cdots + z^d\right)^{2n},$$

we have $a_{nd} - a_{nd-1} \leq a_{nd-d-1} - a_{nd-d-2}$.

Proof. Note that a_i is the Hilbert function of $k[x_1, \ldots, x_{2n}]/\langle x_1^{d+1}, \ldots, x_{2n}^{d+1}\rangle$ in degree i. Hence inequality (6-3) establishes the claim. □

We now turn to an odd number of variables. We are not able to give a result as comprehensive as that of an even number of variables, and we only consider the case of seven variables.

Theorem 6.5. *Let $L \in R = k[x_1, \ldots, x_7]$ be a general linear form, and let $I = \langle x_1^d, \ldots, x_7^d, L^d\rangle$. If $d = 2$ then the ring R/I has the WLP. If $d \geq 4$ then R/I fails the WLP.*

Proof. If $d = 2$, we have verified on CoCoA that R/I has the WLP. In fact, CoCoA has also given the result that when $d = 3$, R/I fails to have the WLP because of the failure of injectivity. However, a computer verification is not enough to show the failure of the WLP, since it is impossible to justify that the linear forms are "general enough". We conjecture that WLP also fails for $d = 3$.

We now assume that $d \geq 4$. We will show the failure of surjectivity in a suitable degree. Let $\ell \in R$ be a general linear form. Let $j = \left\lfloor \frac{17}{5}(d-1) \right\rfloor$. We want to compute $\dim_k [R/(I, \ell)]_j = \dim_k \mathscr{L}_5(j; (j+1-d)^8)$, and in particular, to show that this dimension is nonzero. Using Lemma 3.5 four times we get

$$\begin{aligned}
&\dim_k [R/(I, \ell)]_j \\
&\qquad = \dim_k \mathscr{L}_5(-j + 6(d-1); (j+1-d)^2, (-j+5(d-1))^6) \\
&\qquad = \dim_k \mathscr{L}_5(-3j + 12(d-1); (-j+5(d-1))^4, (-3j+11(d-1))^4,) \\
&\qquad = \dim_k \mathscr{L}_5(-5j + 18(d-1); (-3j+11(d-1))^6, (-5j+17(d-1))^2) \\
&\qquad = \dim_k \mathscr{L}_5(-7j + 24(d-1); (-5j+17(d-1))^8). \qquad (6\text{-}6)
\end{aligned}$$

This computation is correct and has a chance of resulting in a nonempty linear system if $0 \leq -5j + 17(d-1) < -7j + 24(d-1)$, which is true since $j \leq \frac{17}{5}(d-1)$.

Thus, we distinguish five cases, where e is an integer:

Case 1: $d - 1 = 5e$, thus $j = 17e$.

Case 2: $d - 1 = 5e + 1$, thus $j = 17e + 3$.

Case 3: $d - 1 = 5e + 2$, thus $j = 17e + 6$.

Case 4: $d - 1 = 5e + 3$, thus $j = 17e + 10$.

Case 5: $d - 1 = 5e + 4$, thus $j = 17e + 13$.

The computation (6-6) then shows that $\dim_k[R/(I,\ell)]_j$ equals $\dim_k \mathscr{L}_5(e;0^8)$ in Case 1, $\dim_k \mathscr{L}_5(e+3;2^8)$ in Case 2, $\dim_k \mathscr{L}_5(e+6;4^8)$ in Case 3, $\dim_k \mathscr{L}_5(e+2;1^8)$ in Case 4, and $\dim_k \mathscr{L}_5(e+5;3^8)$ in Case 5. It is clear that these linear systems are not empty if $e \geq 0$, thus $d \geq 1$.

To prove failure of the WLP in degree j it remains to check that the expected dimension is zero. Using Lemma 3.7, we obtain

$$\dim_k[R/I]_j - \dim_k[R/I]_{j-1} = \binom{j+5}{5} - 8\binom{j+5-d}{5} + 28\binom{j+5-2d}{5} - 56\binom{j+5-3d}{5}.$$

Notice that the last binomial coefficient is zero if $d \leq 10$, while the third one is zero for $d \leq 2$. However, the computations of the polynomials below are not affected. Distinguishing the five cases above, this dimension times 5! equals

$$\text{Case 1:} -101995\,e^5 - 69925\,e^4 - 15975\,e^3 + 565\,e^2 + 730\,e + 120$$

$$\text{Case 2:} -101995\,e^5 - 139850\,e^4 - 60225\,e^3 - 1330\,e^2 + 5080\,e + 960$$

$$\text{Case 3:} -101995\,e^5 - 209775\,e^4 - 133975\,e^3 - 8145\,e^2 + 19730\,e + 5040$$

$$\text{Case 4:} -101995\,e^5 - 359875\,e^4 - 499175\,e^3 - 336365\,e^2 - 107910\,e - 12600$$

$$\text{Case 5:} -5\,(e+1)\,(20399\,e^4 + 65561\,e^3 + 74044\,e^2 + 32716\,e + 3840)$$

Clearly the first three polynomials are negative if $e \geq 1$, and the last two whenever $e \geq 0$. Thus the expected dimension is zero whenever $d \geq 4$. In particular, we have shown the failure of surjectivity (in particular the failure of the WLP) for $d \geq 4$. □

In applying the approach of Theorem 6.5 to the general case, we were able to mimic the choice of j, as well as the main details of the proof that $\dim[R/(I,\ell)]_j > 0$. However, a proof of the required inequality to verify that it is surjectivity rather than injectivity that fails eluded us. Based on experiments with CoCoA, we end with the following conjecture (also to complete the case of seven variables). Notice that the case $d = 2$ has the WLP in seven variables (as noted above).

Conjecture 6.6. Let $R = k[x_1, \ldots, x_{2n+1}]$, where $n \geq 4$. Let $L \in R$ be a general linear form, and let $I = \langle x_1^d, \ldots, x_{2n+1}^d, L^d \rangle$. Then the ring R/I fails the WLP if and only if $d > 1$. Furthermore, if $n = 3$ then R/I fails the WLP when $d = 3$.

Acknowledgments

Shortly before the end of the writing of this paper we were informed that B. Harbourne, H. Schenck and A. Seceleanu were able to prove Theorem 6.1 for $d \gg 0$. We are very grateful to H. Schenk for sending their paper to us, and for the helpful discussions we have had. We also thank M. Boij for interesting and useful discussions, and the referees for helpful comments.

References

[Anick 1986] D. J. Anick, "Thin algebras of embedding dimension three", *J. Algebra* **100**:1 (1986), 235–259. MR 88d:13016a Zbl 0588.13013

[Ardila and Postnikov 2010] F. Ardila and A. Postnikov, "Combinatorics and geometry of power ideals", *Trans. Amer. Math. Soc.* **362**:8 (2010), 4357–4384. MR 2011g:05322 Zbl 1226.05019

[Boij et al. 2012] M. Boij, J. Migliore, R. Miró-Roig, U. Nagel, and F. Zanello, *On the shape of a pure O-sequence*, Mem. Amer. Math. Soc. **1024**, Amer. Math. Soc., Providence, RI, 2012.

[Brenner and Kaid 2007] H. Brenner and A. Kaid, "Syzygy bundles on $\mathbb{P}^2$ and the weak Lefschetz property", *Illinois J. Math.* **51**:4 (2007), 1299–1308. MR 2009j:13012 Zbl 1148.13007

[Brenner and Kaid 2011] H. Brenner and A. Kaid, "A note on the weak Lefschetz property of monomial complete intersections in positive characteristic", *Collect. Math.* **62**:1 (2011), 85–93. MR 2012c:13031 Zbl 1221.13019

[Catalisano et al. 1999] M. V. Catalisano, P. Ellia, and A. Gimigliano, "Fat points on rational normal curves", *J. Algebra* **216**:2 (1999), 600–619. MR 2000d:13025 Zbl 0940.14035

[Chandler 2005] K. A. Chandler, "The geometric interpretation of Fröberg–Iarrobino conjectures on infinitesimal neighbourhoods of points in projective space", *J. Algebra* **286**:2 (2005), 421–455. MR 2005k:14008 Zbl 1067.14047

[Chandler 2007] K. A. Chandler, "Examples and counterexamples on the conjectured Hilbert function of multiple points", pp. 13–31 in *Algebra, geometry and their interactions* (Notre Dame, IN, 2005), edited by A. Corso et al., Contemp. Math. **448**, Amer. Math. Soc., Providence, RI, 2007. MR 2009c:13036 Zbl 1067.14047

[CoCoA 2009] CoCoA, "CoCoA: a system for doing computations in commutative algebra", 2009, Available at http://cocoa.dima.unige.it.

[Cook and Nagel 2009] D. Cook II and U. Nagel, "The weak Lefschetz property, monomial ideals, and lozenges", preprint, 2009. To appear in *Illinois J. Math.* arXiv 0909.3509

[Cook and Nagel 2011] D. Cook II and U. Nagel, "Enumerations deciding the weak Lefschetz property", preprint, 2011. arXiv 1105.6062

[De Volder and Laface 2007] C. De Volder and A. Laface, "On linear systems of $\mathbb{P}^3$ through multiple points", *J. Algebra* **310**:1 (2007), 207–217. MR 2008b:14009 Zbl 1113.14036

[Di Rocco 1996] S. Di Rocco, "k-very ample line bundles on del Pezzo surfaces", *Math. Nachr.* **179** (1996), 47–56. MR 97h:14007 Zbl 0870.14031

[Dumnicki 2009] M. Dumnicki, "An algorithm to bound the regularity and nonemptiness of linear systems in $\mathbb{P}^n$", *J. Symbolic Comput.* **44**:10 (2009), 1448–1462. MR 2010i:14108 Zbl 1205.14006

[Emsalem and Iarrobino 1995] J. Emsalem and A. Iarrobino, "Inverse system of a symbolic power. I", *J. Algebra* **174**:3 (1995), 1080–1090. MR 96i:13017 Zbl 0842.14002

[Fröberg 1985] R. Fröberg, "An inequality for Hilbert series of graded algebras", *Math. Scand.* **56**:2 (1985), 117–144. MR 87f:13022 Zbl 0582.13007

[Geramita et al. 2009] A. V. Geramita, B. Harbourne, and J. Migliore, "Classifying Hilbert functions of fat point subschemes in $\mathbb{P}^2$", *Collect. Math.* **60**:2 (2009), 159–192. MR 2010g:13025 Zbl 1186.14008

[Harbourne et al. 2011] B. Harbourne, H. Schenck, and A. Seceleanu, "Inverse systems, Gelfand–Tsetlin patterns and the weak Lefschetz property", *J. Lond. Math. Soc.* (2) **84**:3 (2011), 712–730. MR 2855798 Zbl 05987713

[Harima et al. 2003] T. Harima, J. Migliore, U. Nagel, and J. Watanabe, "The weak and strong Lefschetz properties for Artinian K-algebras", *J. Algebra* **262**:1 (2003), 99–126. MR 2004b:13001 Zbl 1018.13001

[Iarrobino 1997] A. Iarrobino, "Inverse system of a symbolic power, III: Thin algebras and fat points", *Compositio Math.* **108**:3 (1997), 319–356. MR 98k:13017 Zbl 0899.13016

[Laface and Ugaglia 2006] A. Laface and L. Ugaglia, "On a class of special linear systems of $\mathbb{P}^3$", *Trans. Amer. Math. Soc.* **358**:12 (2006), 5485–5500. MR 2007e:14009 Zbl 1160.14003

[Li and Zanello 2010] J. Li and F. Zanello, "Monomial complete intersections, the weak Lefschetz property and plane partitions", *Discrete Mathematics* **310**:24 (2010), 3558–3570. MR 2012a:13032 Zbl 1202.13013

[Migliore and Miró-Roig 2003] J. Migliore and R. M. Miró-Roig, "Ideals of general forms and the ubiquity of the weak Lefschetz property", *J. Pure Appl. Algebra* **182**:1 (2003), 79–107. MR 2004c:13027 Zbl 1041.13011

[Migliore et al. 2011] J. C. Migliore, R. M. Miró-Roig, and U. Nagel, "Monomial ideals, almost complete intersections and the weak Lefschetz property", *Trans. Amer. Math. Soc.* **363**:1 (2011), 229–257. MR 2011i:13020

[Nagata 1960] M. Nagata, "On the fourteenth problem of Hilbert", pp. 459–462 in *Proc. Internat. Congress Math. 1958*, Cambridge Univ. Press, New York, 1960. MR 22 #6851 Zbl 0127.26302

[Reid et al. 1991] L. Reid, L. G. Roberts, and M. Roitman, "On complete intersections and their Hilbert functions", *Canad. Math. Bull.* **34**:4 (1991), 525–535. MR 93b:13025 Zbl 0757.13005

[Schenck and Seceleanu 2010] H. Schenck and A. Seceleanu, "The weak Lefschetz property and powers of linear forms in $\mathbb{K}[x, y, z]$", *Proc. Amer. Math. Soc.* **138**:7 (2010), 2335–2339. MR 2011b:13042 Zbl 1192.13013

[Stanley 1980] R. P. Stanley, "Weyl groups, the hard Lefschetz theorem, and the Sperner property", *SIAM J. Algebraic Discrete Methods* **1**:2 (1980), 168–184. MR 82j:20083 Zbl 0502.05004

[Sturmfels and Xu 2010] B. Sturmfels and Z. Xu, "Sagbi bases of Cox–Nagata rings", *J. Eur. Math. Soc.* (*JEMS*) **12**:2 (2010), 429–459. MR 2011f:13009 Zbl 1202.14053

[Watanabe 1987] J. Watanabe, "The Dilworth number of Artinian rings and finite posets with rank function", pp. 303–312 in *Commutative algebra and combinatorics* (Kyoto, 1985), edited by M. Nagata and H. Matsumura, Adv. Stud. Pure Math. **11**, North-Holland, Amsterdam, 1987. MR 89k:13015 Zbl 0648.13010

[Zanello 2006] F. Zanello, "A non-unimodal codimension 3 level h-vector", *J. Algebra* **305**:2 (2006), 949–956. MR 2007h:13026 Zbl 1110.13012

Communicated by Craig Huneke
Received 2010-09-13 Revised 2011-06-09 Accepted 2011-07-19

migliore.1@nd.edu *Department of Mathematics, University of Notre Dame, 255 Hurley Hall, Notre Dame, IN 46556, United States*

miro@ub.edu *Facultat de Matemàtiques, Departament d'Algebra i Geometria, University of Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain*

uwe.nagel@uky.edu *Department of Mathematics, University of Kentucky, 715 Patterson Ofiice Tower, Lexington, KY 40506-0027, United States*

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

Resonance equals reducibility for A-hypergeometric systems

Mathias Schulze and Uli Walther

Classical theorems of Gel'fand et al. and recent results of Beukers show that nonconfluent Cohen–Macaulay A-hypergeometric systems have reducible monodromy representation if and only if the continuous parameter is A-resonant.

We remove both the confluence and Cohen–Macaulayness conditions while simplifying the proof.

1. Introduction

In a series of seminal papers of the 1980s, Gel'fand, Graev, Kapranov and Zelevinskiĭ introduced *A-hypergeometric systems* $H_A(\beta)$, a class of maximally overdetermined systems of linear PDEs. These systems, today also known as *GKZ-systems*, are induced by an integer $d \times n$-matrix A and a parameter vector $\beta \in \mathbb{C}^d$.

A-hypergeometric structures are nearly ubiquitous, generalizing most classical differential equations. Indeed, toric residues, generating functions for intersection numbers on moduli spaces, and special functions (Gauß, Bessel, Airy, etc.) may all be viewed as solutions to GKZ-systems, and the same is true for varying Hodge structures on families of Calabi–Yau toric hypersurfaces as well as the space of roots of univariate polynomials with undetermined coefficients.

Uli Walther was supported by the NSF under grant DMS 0901123.
MSC2010: primary 13N10; secondary 32S40, 14M25.
Keywords: toric, hypergeometric, Euler–Koszul, D-module, resonance, monodromy.

We shall identify A with its set of columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. A parameter β is *nonresonant* if it is not contained in the locally finite subspace arrangement of *resonant* parameters

$$\operatorname{Res}(A) := \bigcup_{\tau} (\mathbb{Z}A + \mathbb{C}\tau)\,, \tag{1-1}$$

the union being taken over all linear subspaces $\tau \subseteq \mathbb{Q}^n$ that form a boundary component of the rational polyhedral cone $\mathbb{Q}_+A$.

Assuming that the toric ring $\mathbb{C}[\mathbb{N}A] = \mathbb{C}[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ is Cohen–Macaulay and standard graded (the latter is equivalent to the classical notion of nonconfluence; see [Schulze and Walther 2008]), Gel'fand et al. [1989; 1990] proved the following fundamental theorems:

(I) $H_A(\beta)$ is holonomic.

(II) The rank (dimension of the solution space) of $H_A(\beta)$ equals the degree of $\mathbb{C}[\mathbb{N}A]$ for generic β.

(III) If β is nonresonant, the monodromy representation of the solutions of $H_A(\beta)$ in a generic point is irreducible.

More recent research has shown that statements (I) and (II) hold true irrespective of whether $\mathbb{C}[\mathbb{N}A]$ is Cohen–Macaulay or standard graded, [Adolphson 1994; Saito et al. 2000; Matusevich et al. 2005]. In Theorems 4.1 and 5.1, we prove the same of statement (III) while providing a converse inspired by [Beukers 2011].

The crucial tool for the proof of (III) in [Gel'fand et al. 1990, Theorem 2.11] is the Riemann–Hilbert correspondence of Kashiwara and Mebkhout, relating regular holonomic D-modules to perverse sheaves. Confluence (i.e., irregularity) of $M_A(\beta)$ rules out the use of the Riemann–Hilbert correspondence in the general case.

A powerful way of studying $H_A(\beta)$ is to consider the corresponding D-module $M_A(\beta)$ on $\mathbb{C}^n$ as a 0-th homology of the *Euler–Koszul complex* $K_\bullet(\mathbb{C}[\mathbb{N}A], \beta)$. This idea can be traced back to [Gel'fand et al. 1989] and was developed into a functor in [Matusevich et al. 2005]. Results from [Matusevich et al. 2005] show that $K_\bullet(\mathbb{C}[\mathbb{N}A], \beta)$ is a resolution of $M_A(\beta)$ if and only if β is not in the *A-exceptional arrangement* $\mathscr{E}_A$ (see Remark 2.2), a well-understood (finite) subspace arrangement of $\mathbb{C}^n$ comprised of the parameters β for which the solution space of $H_A(\beta)$ is unusually large.

Surprisingly, the Euler–Koszul technique combined with the D-module/representation-theoretic description of GKZ-systems from [Schulze and Walther 2009] serves as a replacement for the Riemann–Hilbert correspondence in the proof of (III). This provides an approach that is simultaneously conceptually simpler and more widely applicable.

2. Hypergeometric system and Euler–Koszul homology

Hypergeometric D-module. Let $A = (a_{i,j}) \colon \mathbb{Z}^n \to \mathbb{Z}^d$ be an integer $d \times n$-matrix, which we view both as a map, and as the finite subset $\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\}$ of columns. We assume that the additive group $\mathbb{Z}A$ generated by the columns of A is the free Abelian group $\mathbb{Z}^d$, but we do not assume that A is positive, i.e., we do allow nontrivial units in the semigroup $\mathbb{N}A$ (see Remarks 2.1 and 2.4).

Let $x_A = x_1, \dots, x_n$ be coordinates on $X := \mathbb{C}^n$, and let $\partial_A = \partial_1, \dots, \partial_n$ be the corresponding partial derivative operators on $\mathbb{C}[x_A]$. Then the *Weyl algebra*

$$D_A = \mathbb{C}\langle x_A, \partial_A \mid [x_i, \partial_j] = \delta_{i,j},\ [x_i, x_j] = 0 = [\partial_i, \partial_j]\rangle$$

is the ring of algebraic differential operators on $\mathbb{C}^n$. With $\boldsymbol{u}_+ = (\max(0, u_j))_j$ and $\boldsymbol{u}_- = \boldsymbol{u}_+ - \boldsymbol{u}$, write $\Box_{\boldsymbol{u}}$ for $\partial^{\boldsymbol{u}_+} - \partial^{\boldsymbol{u}_-}$, where here and elsewhere we freely use multiindex notation. The *toric relations of A* are then

$$\Box_A := \{\Box_{\boldsymbol{u}} \mid A\boldsymbol{u} = 0\} \subseteq R_A := \mathbb{C}[\partial_A],$$

and generate the *toric ideal* $I_A = R_A \cdot \Box_A$, whose residue ring is the *toric ring*

$$S_A := R_A/I_A \cong \mathbb{C}[\mathbb{N}A] = \mathbb{C}[\boldsymbol{a}_1, \dots, \boldsymbol{a}_n].$$

The *Euler vector fields* $E = E_1, \dots, E_d$ induced by A are defined as

$$E_i := \sum_{j=1}^{n} a_{i,j} x_i \partial_j.$$

Then, for $\beta \in \mathbb{C}^d$, the *A-hypergeometric ideal* and *D-module* are, by [Gel'fand et al. 1987; 1989], the left D_A-ideal and D_A-module

$$H_A(\beta) = D_A \cdot \{E - \beta\} + D_A \cdot \Box_A \quad \text{and} \quad M_A(\beta) = D_A/H_A(\beta).$$

The structure of the solutions to $H_A(\beta)$ is tightly interwoven with the combinatorics of the pair $(A, \beta) \in (\mathbb{Z}A)^n \times \mathbb{C}A$ [Sturmfels and Takayama 1998; Cattani et al. 1999; Matusevich and Miller 2006; Okuyama 2006; Berkesch 2011].

Remark 2.1. Suppose we were to weaken the condition $\mathbb{Z}A = \mathbb{Z}^d$ to "the rank of $\mathbb{Z}A$ is d". Pick a basis B for $\mathbb{Z}A$, interpreted as elements of $\mathbb{Z}^d$. In terms of B, A takes the form of the $d \times n$ matrix A' (say) which satisfies $A = BA'$ and $\mathbb{Z}A' = \mathbb{Z}^d$. Choose $\beta \in \mathbb{C}A = \mathbb{C}A'$. The hypergeometric systems $H_A(\beta)$ and $H_{A'}(B^{-1}\beta)$ are equivalent since $\ker_{\mathbb{Z}^n}(A) = \ker_{\mathbb{Z}^n}(A')$.

Torus action. Consider the algebraic d-torus $T := \operatorname{Spec}(\mathbb{C}[\mathbb{Z}A]) \cong (\mathbb{C}^*)^d$ with coordinate functions $t = t_1, \dots, t_d$. The columns $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ of A can be viewed as characters $\boldsymbol{a}_i(t) = t^{\boldsymbol{a}_i}$ on T, and the parameter vector $\beta \in \mathbb{C}^d$ as a character on its Lie algebra via $\beta(t_i\partial_{t_i}) = -\beta_i + 1$. These characters define an action of T on

$X^* := \operatorname{Spec}(\mathbb{C}[\mathbb{N}^n])$, interpreted as the cotangent space T_0^*X of X at 0, by

$$t \cdot \partial_A = (t^{\boldsymbol{a}_1}\partial_1, \ldots, t^{\boldsymbol{a}_n}\partial_n).$$

The toric ideal I_A is the ideal of the closure of the orbit $T \cdot \mathbf{1}_A$ of $\mathbf{1}_A = (1, \ldots, 1)$ in X^*, whose coordinate ring is S_A.

The contragredient action of T on the coordinate ring R_A of X^* is given by

$$(t \cdot P)(\partial_A) = P(t^{-\boldsymbol{a}_1}\partial_1, \ldots, t^{-\boldsymbol{a}_n}\partial_n)$$

for $P \in R_A$. It yields a $\mathbb{Z}A$-*grading* on R_A on the coordinate ring $\mathbb{C}[x_A, \partial_A]$ of T^*X:

$$-\deg \partial_j = \boldsymbol{a}_j = \deg x_j. \tag{2-1}$$

In particular, $\deg \partial^{\boldsymbol{u}} = A\boldsymbol{u}$, and $E - \beta$ and $\Box_A$ are homogeneous.

The following description of $M_A(\beta)$ was given in [Schulze and Walther 2009]. Consider the algebraic $\mathcal{D}_T$-module

$$\mathcal{M}(\beta) := \mathcal{D}_T/\mathcal{D}_T \cdot \langle \partial_t t + \beta \rangle,$$

where $\partial_t t := \partial_1 t_1, \ldots, \partial_d t_d$. It is $\mathcal{O}_T$-isomorphic to $\mathcal{O}_T$ but equipped with a twisted $\mathcal{D}_T$-module structure expressed symbolically as

$$\mathcal{M}(\beta) = \mathcal{O}_T \cdot t^{-\beta-1}$$

on which $\mathcal{D}_T$ acts via the product rule. The orbit inclusion

$$\phi\colon T \to T \cdot \mathbf{1} \hookrightarrow X$$

gives rise to a (derived) direct image functor $\phi_+\colon \mathcal{D}_T\text{-mods} \to \mathcal{D}_X\text{-mods}$. On X one has access to the *Fourier transform*: $\mathcal{F}(x_i) = \partial_i$, $\mathcal{F}(\partial_i) = -x_i$. By [Schulze and Walther 2009, Proposition 2.1], $\mathcal{F} \circ \phi_+\mathcal{M}(\beta)$ is represented by the Euler–Koszul complex $K_\bullet(S_A[\partial_A^{-1}], \beta)$. Thus, the latter is quasiisomorphic to $K_\bullet(S_A, \beta)$ if $\beta \notin \operatorname{Res}(A)$ by [Schulze and Walther 2009, Theorem 3.6], and hence Corollary 3.8 of [Schulze and Walther 2009] yields

$$M_A(\beta) = \mathcal{F} \circ \phi_+\mathcal{M}(\beta) \quad \text{if } \beta \notin \operatorname{Res}(A). \tag{2-2}$$

Euler–Koszul functor. We say that $\beta \in \mathbb{Z}A$ is a *true degree* of the graded R_A-module M if β is the degree of a nonzero homogeneous element of M. The *quasidegrees* of M are the points $\operatorname{qdeg} M$ in the Zariski closure of $\operatorname{tdeg} M \subseteq \mathbb{Z}A \subseteq \mathbb{C}A$.

A graded R_A-module M is called a *toric module* if it has a finite filtration by graded R_A-modules such that each filtration quotient is a finitely generated S_A-module. The toric modules with $\mathbb{Z}A$-homogeneous maps of degree zero form a category that is closed under subquotients and extensions. For every toric module the

quasidegrees form a finite subspace arrangement where each participating subspace is a shift of a complexified face of $\mathbb{Q}_{\geq 0}A$ by a lattice element.

For all $\beta \in \mathbb{C}^d$ and for any toric R_A-module M one can define a collection of d commuting D_A-linear endomorphisms denoted $E_i - \beta_i$, $1 \leq i \leq d$, on the D_A-module $D_A \otimes_{R_A} M$ which operate on a homogeneous element $m \in D_A \otimes_{R_A} M$ by $m \mapsto (E_i - \beta_i) \circ m$, where

$$(E_i - \beta_i) \circ m = (E_i - \beta_i - \deg_i m) \cdot m.$$

There is an exact functor $K_\bullet(-, \beta) = K_\bullet(-, E - \beta)$ from the category of graded R_A-modules to the category of complexes of graded D_A-modules; it sends M to the Koszul complex defined by all morphisms $E_i - \beta_i$. On toric modules, the functor returns complexes with holonomic homology. A short exact sequence

$$0 \to M' \to M \to M'' \to 0$$

of graded R_A-modules with homogeneous maps of degree zero induces a long exact sequence of *Euler–Koszul homology*

$$\cdots \to H_i(M'', \beta) \to H_{i-1}(M', \beta) \to H_{i-1}(M, \beta) \to H_{i-1}(M'', \beta) \to \cdots$$

where $H_i(-, \beta) = H_i(K_\bullet(-, \beta))$. If $M = S_A$ then $H_0(M, \beta) = M_A(\beta)$.

We refer to [Matusevich et al. 2005; Schulze and Walther 2009] for more details.

Rank (jumps) and monodromy reducibility. We shall write $D_A(x_A)$ for the ring of $\mathbb{C}$-linear differential operators on $\mathbb{C}(x_A)$; note that $D_A(x_A) = \mathbb{C}(x_A) \otimes_{\mathbb{C}[x_A]} D_A$ as left D_A-module. We further set $M(x_A) := \mathbb{C}(x_A) \otimes_{\mathbb{C}[x_A]} M$ for any D_A-module M.

The *rank* $\operatorname{rk}(M)$ of a D_A-module M is the $\mathbb{C}(x_A)$-dimension of $M(x_A)$. By Kashiwara's Cauchy–Kovalevskaya theorem [Saito et al. 2000, Theorem 1.4.19], it equals the $\mathbb{C}$-dimension of the *solution space* $\operatorname{Sol}(M) = \operatorname{Hom}_{D_A}(M, \mathbb{C}\{x_A - \varepsilon\})$ of M with coefficients in the convergent power series near the generic point $x_A = \varepsilon$ in (the analytic space associated to) X.

Remark 2.2. By [Adolphson 1994, Theorem 5.15] and [Matusevich et al. 2005, Theorems 2.9, 7.5],

$$\operatorname{rk} M_A(\beta) \geq \operatorname{vol}_A(A)$$

with equality for generic $\beta \in \mathbb{C}^n$. Here $\operatorname{vol}_A(G)$ denotes, for any $G \subseteq \mathbb{Z}A$, the simplicial volume of the convex hull of G taken in the lattice $\mathbb{Z}A$. More precisely, equality is equivalent to $\beta \notin \mathscr{E}_A$, where

$$\mathscr{E}_A := \sum_{j=1}^{n} \boldsymbol{a}_j - \bigcup_{i=0}^{d-1} \operatorname{qdeg} \operatorname{Ext}_{R_A}^{n-i}(S_A, R_A)$$

is the *exceptional arrangement*.

Definition 2.3. We say that a D_A-module M has *irreducible monodromy* if $M(x_A)$ is an irreducible $D_A(x_A)$-module (i.e., it has no nontrivial $D_A(x_A)$-quotients).

By [Walther 2007, Theorem 3.15], monodromy irreducibility of $M(\beta)$ is a property of the equivalence class $\beta \in \mathbb{C}A/\mathbb{Z}A$.

The nomenclature is based on the Riemann–Hilbert correspondence: $D_A(x_A)$-quotients of $M(x_A)$ correspond to monodromy-invariant subspaces of $\mathrm{Sol}(M)$ in nonsingular points of M. (Analytic continuations of an analytic germ satisfy the same differential equations as the germ itself).

Remark 2.4. Careful reading of [Matusevich et al. 2005] reveals that all fundamental results obtained through Euler–Koszul technology do not require $\mathbb{N}A$ to be a positive semigroup. As a matter of fact, $\mathscr{E}_A$ was defined in [Matusevich et al. 2005] in terms of local cohomology with supports at the origin of X^*; the translation between this definition and ours here can only be done if A is pointed. On the other hand, it is the Ext-based definition that is (implicitly) used in all proofs in loc. cit.

In consequence, the main theorems in [Walther 2007] and [Schulze and Walther 2009] remain true in the absence of positivity since the only ingredients in their proofs that are specific to the hypergeometric situation are those of [Matusevich et al. 2005].

3. Pyramids and resonance centers

Definition 3.1. For any subset F of the columns of A we write $\bar{F}$ for the complement $A \smallsetminus F$.

A *face of* A is any subset $F \subseteq A$ subject to the condition that there be a linear functional $\phi_F : \mathbb{Z}A \to \mathbb{Z}$ that vanishes on F but is positive on $\bar{F}$. This includes $F = A$ as possibility. Every face contains all units of $\mathbb{N}A$, and A is positive if and only if the empty set is a face of A.

For a given face F, we set

$$I_A^F := I_A + R_A \cdot \partial_{\bar{F}}$$

and note that $R_A/I_A^F = S_F$ as R_A-module.

Definition 3.2. Let F be a face of A. The parameter $\beta \in \mathbb{C}^d$ is *F-resonant* if $\beta \in \mathbb{Z}A + \mathbb{C}G$ for a proper subface G of F.

If β is G-resonant for all faces G properly containing F, but not for F itself, we call F a *resonance center for* β.

A resonance center is a minimal face F for which $\beta \in \mathbb{Z}A + \mathbb{C}F$. Every parameter β has a resonance center; A is a (and then the only) center of resonance for β if and only if β is nonresonant in the usual sense (i.e., $\beta \notin \mathrm{Res}(A)$, defined in (1-1)).

On the other hand, for positive A, the empty face is a resonance center for β if and only if $\beta \in \mathbb{Z}A$.

Example 3.3. It is easy to have several resonance centers for β. For example, consider $\beta = (\frac{1}{2}, 1)$ on the quadric cone $A = \left(\begin{smallmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{smallmatrix}\right)$; β has both extremal rays as resonance centers.

Definition 3.4. We say that A is a(n iterated) *pyramid over the face* F if $d = \dim_{\mathbb{Z}}(\mathbb{Z}A)$ equals $|\bar{F}| + \dim_{\mathbb{Z}}(\mathbb{Z}F)$.

The following equivalences are trivial or follow from [Walther 2007, Lemma 3.13].

Lemma 3.5. *The following statements are equivalent.*

(1) F *is a face and* A *is a pyramid over* F.

(2) $\boldsymbol{a}_j \notin \mathbb{Q}(A \smallsetminus \{\boldsymbol{a}_j\})$ *for any* $j \notin F$.

(3) $\mathbb{Z}A = \mathbb{Z}\boldsymbol{a}_j \oplus \mathbb{Z}(A \smallsetminus \{\boldsymbol{a}_j\})$ *for any* $j \notin F$.

(4) $\mathrm{vol}_F(F) = \mathrm{vol}_A(A)$.

(5) *For every* $\beta \in \mathbb{C}A$, *the coefficients* c_j *in the sum* $\beta = \sum_A c_j \boldsymbol{a}_j$ *are uniquely determined by* β *for* $j \notin F$.

(6) *The generators* $\Box_A$ *of* I_A *do not involve* ∂_j *for any* $j \notin F$.

(7) $S_F \otimes_{\mathbb{C}} \mathbb{C}[\partial_{\bar{F}}] = S_A$ *as* R_A*-modules.*

Notation 3.6. Suppose F is any nonempty face of A, and let X_F, X_F^*, T_F, $H_\bullet^F$, etc. be defined as in Section 2 with A replaced by F (cf. Remark 2.1 for the case where $\mathbb{Z}A/\mathbb{Z}F$ has torsion). Write $E^F = E_1^F, \ldots, E_d^F$ where $E_i^F := \sum_{j \in F} a_{i,j} x_j \partial_j$ is the part of E_i supported in F. Then, in particular,

$$M_F(\beta) = D_F/(D_F \cdot \langle E^F - \beta \rangle + D_F \cdot I_F) \quad \text{for } \beta \in \mathbb{C}F.$$

Suppose now that A is a pyramid over the face F, and let $\beta \in \mathbb{C}A$. The splitting in Lemma 3.5(3) corresponds to a splitting of tori $T_A = T_F \times \prod_{\boldsymbol{a}_j \in \bar{F}} T_{\boldsymbol{a}_j}$ which in turn gives a splitting of the spaces of Lie algebra characters $\mathbb{C}A = \mathbb{C}F \oplus \bigoplus_{\boldsymbol{a}_j \in \bar{F}} \mathbb{C}\boldsymbol{a}_j$. Then β decomposes correspondingly as

$$\beta = \beta^F + \sum_{j \in \bar{F}} \beta_j^{\bar{F}}.$$

Let $\iota_F \colon X_F^* \hookrightarrow X_A^*$ be the inclusion. By [Matusevich et al. 2005, Lemma 4.8], for $\beta \in \mathbb{C}F$,

$$\begin{aligned}(\mathcal{F} \circ \iota_{F,+} \circ \mathcal{F}^{-1}) M_F(\beta) &= \mathbb{C}[x_{\bar{F}}] \otimes_{\mathbb{C}} M_F(\beta) \\ &\cong H_0(S_F, \beta) = D_A/(D_A \cdot \langle E^F - \beta \rangle + D_A \cdot I_A^F) \quad (3\text{-}1)\end{aligned}$$

as D_A-modules. In the following lemma, (9) follows from (8) and (3-1) above.

Lemma 3.7. *If A is a pyramid over F then the following conditions hold*:

(8) *The ideal $H_A(\beta)$ contains $x_j\partial_j - \beta_j^{\overline{F}}$ for $j \notin F$.*

(9) *$M_A(\beta)(x_A) = \mathbb{C}(x_A) \otimes_{\mathbb{C}[x_F]} M_F(\beta)$ for $\beta \in \mathbb{C}F$.*

(10) *The solutions of $M_A(\beta)$ are the solutions of $M_F(\beta^F)$, multiplied with the unique solution to the system*

$$\{x_j\partial_j \bullet f = \beta_j^{\overline{F}} \cdot f\}_{j\in\overline{F}}.$$

In particular, $\beta \in \mathscr{E}_A$ if and only if $\beta^F \in \mathscr{E}_F$.

Proposition 3.8. *If $\beta \in \mathbb{C}A$ has a resonance center F over which A is a pyramid, then F is the only resonance center for β.*

Proof. Let G be a second resonance center for β and suppose G meets the complement of F; pick $\boldsymbol{a}_k \in G \cap \overline{F}$. Since $\mathbb{Z}\boldsymbol{a}_k$ is a direct summand of $\mathbb{Z}A$, it is also a direct summand of $\mathbb{Z}G$. It follows that $G \smallsetminus \{\boldsymbol{a}_k\}$ is a face G' of A.

As F and G are resonance centers,

$$\beta = z_k\boldsymbol{a}_k + \sum_{j\in\overline{F}\smallsetminus\{k\}} z_j\boldsymbol{a}_j + \sum_{j\in F} c_j\boldsymbol{a}_j, \quad \beta = c'_k\boldsymbol{a}_k + \sum_{j\in\overline{G'}\smallsetminus\{k\}} z'_j\boldsymbol{a}_j + \sum_{j\in G'} c'_j\boldsymbol{a}_j$$

where $z_k, z_j, z'_j \in \mathbb{Z}$ and $c'_k, c_j, c'_j \in \mathbb{C}$. By Lemma 3.5(5), the coefficients for $\boldsymbol{a}_k$ in these sums are identical, $c'_k = z_k \in \mathbb{Z}$. It follows that

$$\beta = \left(z_k\boldsymbol{a}_k + \sum_{j\in\overline{G'}\smallsetminus\{k\}} z'_j\boldsymbol{a}_j\right) + \sum_{G'} c'_j\boldsymbol{a}_j \in \mathbb{Z}A + \mathbb{C}G'.$$

This contradicts G being a resonance center. Thus $G \cap \overline{F} = \varnothing$ and so $G \subseteq F$. But then F can only be a resonance center if $F = G$. □

4. Resonance implies reducibility

The following result generalizes Theorem 3.4 in [Walther 2007] and Theorem 1.3 in [Beukers 2011].

Theorem 4.1. *Let F be a resonance center for $\beta \in \mathbb{C}A$. If A is not a pyramid over F then $M_A(\beta)$ has reducible monodromy.*

Proof. By hypothesis, we have $\beta - \gamma \in \mathbb{Z}A$ for some $\gamma \in \mathbb{C}F$. We first dispose of the case $F = \varnothing$. In that case, A is positive, $\gamma = 0$, $\beta \in \mathbb{Z}A$ and, by [Walther 2007, Theorem 3.15], we may assume $\beta = 0$. Then $\mathbb{C}(x_A)$ is a rank-1 quotient of $M_A(\beta)(x_A)$. But A is not a pyramid over F, so

$$\operatorname{rk}(M_A(\beta)) \geq \operatorname{vol}_A(A) > \operatorname{vol}_F(F) = 1 = \operatorname{rk}(\mathbb{C}(x_A))$$

by Remark 2.2 and Lemma 3.5. So $\mathbb{C}(x_A)$ is a proper quotient of $M_A(\beta)(x_A)$, and hence $M_A(\beta)$ has reducible monodromy. We can hence assume that F is not empty, and by [Walther 2007, Theorem 3.15], we need to show the reducibility of $M_A(\gamma)$.

Consider the surjection

$$M_A(\gamma) = H_0(S_A, \gamma) \twoheadrightarrow H_0(S_F, \gamma)$$

induced by the surjection $S_A \twoheadrightarrow S_F$. Therefore, it suffices to show that $0 < \mathrm{rk}(H_0(S_F, \gamma)) < \mathrm{vol}_A(A)$ by Remark 2.2. Since F is a resonance center for β, and hence for γ as well, γ is a nonresonant parameter for the GKZ-system

$$M_F(\gamma) = D_F/(D_F \cdot \langle E^F - \gamma \rangle + D_A \cdot I_F).$$

Then, by Remark 2.2, $\mathrm{rk}(M_F(\gamma)) = \mathrm{vol}_F(F) > 0$ and $\mathrm{rk}(M_A(\gamma)) \geq \mathrm{vol}_A(A)$. As A is not a pyramid over F, $\mathrm{vol}_F(F) < \mathrm{vol}_A(A)$ by Lemma 3.5. Finally, $\mathrm{rk}(M_F(\gamma)) = \mathrm{rk}(H_0(S_F, \gamma))$ by (3-1). Combining the above (in)equalities yields the claim. □

5. Resonance follows from reducibility

We now generalize Theorem 2.11 in [Gel'fand et al. 1990].

Theorem 5.1. *Let F be a resonance center for β. If A is a pyramid over F then $M_A(\beta)$ has irreducible monodromy.*

Proof.

First consider the case $F = A$. Then $\beta \notin \mathrm{Res}(A)$ and hence $M_A(\beta) = \mathcal{F} \circ \phi_+(\mathcal{M}_\beta)$ by (2-2). As in the proof of [Schulze and Walther 2009, Proposition 2.1], factor $\phi = \varpi \circ \iota$ into the closed embedding of tori

$$\iota : T \hookrightarrow \mathrm{Spec}(\mathbb{C}[\mathbb{Z}^n]) = Y^* \cong (\mathbb{C}^*)^n \tag{5-1}$$

induced by $\mathbb{Z}A \subseteq \mathbb{Z}^n$, followed by the open embedding

$$\varpi : Y^* = X^* \setminus \mathrm{Var}(\partial_1 \cdots \partial_n) \hookrightarrow X^*. \tag{5-2}$$

By Kashiwara equivalence, ι preserves irreducibility. The same holds for ϖ, because D-affinity of both the target and the source of the inclusion map allows to detect submodules on global sections. But global sections on Y^* and X^* agree because we are looking at an open embedding. Since $\mathcal{M}(\beta)$ is clearly irreducible, $\phi_+\mathcal{M}(\beta)$ is as well. As Fourier transforms preserve composition chains, $M_A(\beta)$ is irreducible. It follows that $M_A(\beta)$ has irreducible monodromy.

Suppose now that F is a proper face. Choose $\gamma \in \mathbb{C}F$ with $\beta - \gamma \in \mathbb{Z}A$. Then $M_F(\gamma)$ is irreducible by the first part of the proof, and the claim follows from Lemma 3.7(9) and [Walther 2007, Theorem 3.15]. Finally, if $F = \varnothing$ then A is positive and Lemma 3.7(8) shows that $M_A(\beta)(x_A) = \mathbb{C}(x_A)$ which has clearly irreducible monodromy. □

Acknowledgments

We are grateful to the referees for their comments, and for informing us that Mutsumi Saito has an article in press with *Compositio Mathematica* that also discusses reducibility of GKZ-systems (in much greater detail). We would also like to thank Alan Adolphson for raising a relevant question.

References

[Adolphson 1994] A. Adolphson, "Hypergeometric functions and rings generated by monomials", *Duke Math. J.* **73**:2 (1994), 269–290. MR 96c:33020 Zbl 0804.33013

[Berkesch 2011] C. Berkesch, "The rank of a hypergeometric system", *Compos. Math.* **147**:1 (2011), 284–318. MR 2012f:16022 Zbl 1214.33009

[Beukers 2011] F. Beukers, "Irreducibility of A-hypergeometric systems", *Indag. Math.* (*N.S.*) **21**:1-2 (2011), 30–39. MR 2832480 Zbl 1229.33023

[Cattani et al. 1999] E. Cattani, C. D'Andrea, and A. Dickenstein, "The $\mathcal{A}$-hypergeometric system associated with a monomial curve", *Duke Math. J.* **99**:2 (1999), 179–207. MR 2001f:33018 Zbl 0952.33009

[Gel'fand et al. 1987] I. M. Gel'fand, M. I. Graev, and A. V. Zelevinskiĭ, "Holonomic systems of equations and series of hypergeometric type", *Dokl. Akad. Nauk SSSR* **295**:1 (1987), 14–19. In Russian; translated in *Soviet Math. Dokl.***36**:1(1988), 5-10. MR 88j:58118

[Gel'fand et al. 1989] I. M. Gel'fand, A. V. Zelevinskiĭ, and M. M. Kapranov, "Hypergeometric functions and toric varieties", *Funktsional. Anal. i Prilozhen.* **23**:2 (1989), 12–26. In Russian; translated in *Funct. Anal. Appl.* **23**:2 (1989), 94–106. MR 90m:22025 Zbl 0721.33006

[Gel'fand et al. 1990] I. M. Gel'fand, M. M. Kapranov, and A. V. Zelevinsky, "Generalized Euler integrals and A-hypergeometric functions", *Adv. Math.* **84**:2 (1990), 255–271. MR 92e:33015

[Matusevich and Miller 2006] L. F. Matusevich and E. Miller, "Combinatorics of rank jumps in simplicial hypergeometric systems", *Proc. Amer. Math. Soc.* **134**:5 (2006), 1375–1381. MR 2006j:33016 Zbl 1090.33005

[Matusevich et al. 2005] L. F. Matusevich, E. Miller, and U. Walther, "Homological methods for hypergeometric families", *J. Amer. Math. Soc.* **18**:4 (2005), 919–941. MR 2007d:13027 Zbl 1095.13033

[Okuyama 2006] G. Okuyama, "$\mathcal{A}$-hypergeometric ranks for toric threefolds", *Int. Math. Res. Not.* **2006** (2006), Art. ID 70814, 38. MR 2007h:14076

[Saito et al. 2000] M. Saito, B. Sturmfels, and N. Takayama, *Gröbner deformations of hypergeometric differential equations*, Algorithms and Computation in Mathematics **6**, Springer, Berlin, 2000. MR 2001i:13036 Zbl 0946.13021

[Schulze and Walther 2008] M. Schulze and U. Walther, "Irregularity of hypergeometric systems via slopes along coordinate subspaces", *Duke Math. J.* **142**:3 (2008), 465–509. MR 2009b:13067 Zbl 1144.13012

[Schulze and Walther 2009] M. Schulze and U. Walther, "Hypergeometric D-modules and twisted Gauß-Manin systems", *J. Algebra* **322**:9 (2009), 3392–3409. MR 2010m:14028 Zbl 1181.13023

[Sturmfels and Takayama 1998] B. Sturmfels and N. Takayama, "Gröbner bases and hypergeometric functions", pp. 246–258 in *Gröbner bases and applications* (Linz, 1998), edited by B. Buchberger and F. Winkler, London Math. Soc. Lecture Note Ser. **251**, Cambridge Univ. Press, 1998. MR 2001c:33026 Zbl 0918.33004

[Walther 2007] U. Walther, "Duality and monodromy reducibility of A-hypergeometric systems", *Math. Ann.* **338**:1 (2007), 55–74. MR 2008e:32043 Zbl 1126.33006

Communicated by Bernd Sturmfels
Received 2010-10-01 Revised 2011-01-04 Accepted 2011-02-22

mschulze@math.okstate.edu *Department of Mathematics, Oklahoma State University, Stillwater, OK 74078, United States*

walther@math.purdue.edu *Department of Mathematics, Purdue University, 150 North University Street, West Lafayette, IN 47907-2067, United States*

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

The Chow ring of double EPW sextics

Andrea Ferretti

A conjecture of Beauville and Voisin states that for an irreducible symplectic variety X the subring of $CH^*(X)$ generated by divisors goes injectively into the cohomology of X, via the cycle map. We prove this for a very general double Eisenbud–Popescu–Walter sextic.

1. Introduction

A difficult problem in algebraic geometry is to characterize the kernel (and the image) of the cycle map

$$c : CH^*(X) \to H^*(X, \mathbb{Z})$$

for a smooth projective variety X over $\mathbb{C}$. When X is an *irreducible symplectic variety* there is this general conjecture:

Conjecture 1 [Beauville 2007]. *Let X be an irreducible symplectic variety, and let $DCH(X) \subset CH^*_{\mathbb{Q}}(X)$ be the subalgebra generated by the divisors. Then the cycle map*

$$c : CH^*(X)_{\mathbb{Q}} \to H^*(X, \mathbb{Q})$$

is injective when restricted to $DCH(X)$.

We refer to the original article for the motivation of the conjecture and its link with the conjectures of Bloch and Beilinson; we just remark that the conjecture was known to hold when X is a K3 surface from [Beauville and Voisin 2004]. Conjecture 1 explicitly means the following: any polynomial relation

$$P(D_1, \ldots, D_k) = 0$$

in the fundamental classes of divisors which holds in $H^*(X)$ already holds inside $CH^*_{\mathbb{Q}}(X)$.

This has been extended by Claire Voisin:

I'd like to thank K. O'Grady for his constant support, advice, and encouragement during my Ph.D..
MSC2010: primary 53C26; secondary 14C15, 14J35.
Keywords: EPW sextics, Chow ring, hyperkähler, irreducible symplectic varieties.

Conjecture 2 [Voisin 2008]. *Let X be an irreducible symplectic variety. Any polynomial relation*

$$P(D_1, \ldots, D_k, c_i(X)) = 0$$

in the fundamental classes of divisors and in the Chern classes of X *that holds in $H^*(X)$ already holds in $CH^*_{\mathbb{Q}}(X)$.*

Theorem [Voisin 2008]. *Conjecture 2 holds when*

- *$X = S^{[n]}$, for some K3 surface S, and $n \leq 2b_2(S)_{\mathrm{tr}} + 4$, where $b_2(S)_{\mathrm{tr}}$ is the rank of the transcendental part of $H^2(S)$, that is, the orthogonal of the Néron–Severi lattice, or*
- *X is the Fano variety of lines on a cubic fourfold $Y \subset \mathbb{P}^5$.*

To our knowledge, no other cases of the conjecture have been verified. We prove it to be true for double EPW sextics, a class of irreducible symplectic varieties introduced in [O'Grady 2008a] and named after Eisenbud, Popescu and Walter. (We will review their construction shortly.) More precisely:

Theorem 1.1. *Let X be a double EPW sextic and $f : X \to Y$ its associated double covering. Let*

$$h = f^*\mathcal{O}_Y(1)$$

be the natural polarization. Then every polynomial relation between h and the Chern classes of X which holds in $H^*(X, \mathbb{Q})$ already holds in $CH^*(X)_{\mathbb{Q}}$.

In particular, if X is very general, Conjecture 2 holds for X.

Theorem 1.1 is the main result of my Ph.D. thesis [Ferretti 2009]. Some facts that are only cited in the present paper are described there in more detail.

EPW sextics. In this section we recall some known facts about EPW sextics and we fix the notation that we shall use. The results here are from [O'Grady 2006; 2008a]; see also [Ferretti 2009] for a detailed introduction.

We start with a six-dimensional vector space V over the field $\mathbb{C}$. The space $\bigwedge^6 V$ is one-dimensional, so we choose once and for all an isomorphism

$$\mathrm{vol} : \bigwedge^6 V \to \mathbb{C}.$$

This endows $\bigwedge^3 V$ with a symplectic form, given by

$$(\alpha, \beta) = \mathrm{vol}(\alpha \wedge \beta),$$

for $\alpha, \beta \in \bigwedge^3 V$, so $\bigwedge^3 V$ becomes a symplectic vector space of dimension 20.

For each nonzero $v \in V$ we can consider the Lagrangian subspace

$$F_v = \{v \wedge \alpha \mid \alpha \in \bigwedge^2 V\}.$$

This is clearly isotropic, and the isomorphism

$$\begin{aligned} \varphi_v : F_v &\xrightarrow{\cong} \bigwedge^2(V/\langle v\rangle) \\ v\wedge\alpha &\longmapsto [\alpha] \end{aligned} \tag{1-1}$$

shows that $\dim F_v = \binom{5}{2} = 10$.

Since the subspace F_v only depends on the class $[v]\in\mathbb{P}(V)$, the subspaces F_v fit together, giving rise to a *Lagrangian subbundle* F of the trivial symplectic bundle $\mathbb{P}(V)\times\bigwedge^3 V$. The maps in (1-1) then yield an isomorphism

$$F \cong \mathcal{S}\otimes\bigwedge^2\mathcal{Q},$$

where $\mathcal{Q}$ is the tautological quotient bundle on $\mathbb{P}(V)$ and $\mathcal{S}$ the tautological subbundle. From this a standard computation gives

$$c_1(F) = c_1(\bigwedge^2\mathcal{Q}) + \mathrm{rk}(F)c_1(\mathcal{S}) = -6H, \tag{1-2}$$

where $H = c_1(\mathcal{O}(1))$ is the hyperplane class on $\mathbb{P}(V)$.

We are now ready to define the EPW sextics, as follows. Fix a Lagrangian subspace $A\subset\bigwedge^3 V$. Note that the symplectic form gives a canonical identification

$$\bigwedge^3 V/A \cong A^\vee.$$

Let

$$\lambda_A : F \to \mathcal{O}_{\mathbb{P}(V)}\otimes A^\vee \tag{1-3}$$

be the inclusion $F\hookrightarrow\mathcal{O}_{\mathbb{P}(V)}\otimes\bigwedge^3 V$ followed by the projection modulo A. The map λ_A is a map of vector bundles of equal rank 10.

Definition 1.2. We set

$$Y_A = Z(\det\lambda_A),$$

the zero locus of the determinant of λ_A. This is a subscheme of $\mathbb{P}(V)$; when it is not the whole of $\mathbb{P}(V)$, Y_A is called an *EPW sextic*. (It is indeed a sextic by (1-2).)

The support of the scheme Y_A is by definition the locus

$$\{[v]\in\mathbb{P}(V) \mid \dim(F_v\cap A)\geq 1\}.$$

We then set

$$Y_A[k] = \{[v]\in\mathbb{P}(V) \mid \dim(F_v\cap A)\geq k\},$$

so that $Y_A = Y_A[1]$, at least set-theoretically. The loci $Y_A[k]$ also have a natural scheme structure, given by the vanishing of the determinants of the $(11-k)\times(11-k)$ minors of λ_A.

The natural parameter space for EPW sextics is the Lagrangian Grassmannian $\mathbb{LG}(\bigwedge^3 V)$, or more precisely the Zariski open set parametrizing those A for which

$Y_A \subsetneq \mathbb{P}(V)$. We recall that $\mathbb{LG}(\bigwedge^3 V) \subset \mathrm{Gr}(10, \bigwedge^3 V)$ is the subvariety of Lagrangian subspaces; it is a smooth variety of dimension 55.

Double EPW sextics. Assume that Y_A is not the whole of $\mathbb{P}(V)$. The map of vector bundles λ_A in (1-3) is an *injective homomorphism of sheaves*, whose cokernel is supported on Y_A. If we denote by

$$i_A : Y_A \to \mathbb{P}(V)$$

the inclusion, then we have an exact sequence

$$0 \longrightarrow F \longrightarrow \mathcal{O}_{\mathbb{P}(V)} \otimes A^\vee \longrightarrow i_{A*}(\xi_A) \longrightarrow 0 \tag{1-4}$$

for some sheaf ξ_A on Y_A. For a generic Lagrangian subspace A the locus

$$Y_A[2] = \left\{ [v] \in \mathbb{P}(V) \mid \dim(F_v \cap A) \geq 2 \right\}$$

is properly contained in Y_A; it follows that ξ_A is generically free of rank 1. If

$$\zeta_A = \xi_A^\vee(3),$$

there is a natural *multiplication* map

$$m_A : \zeta_A \otimes \zeta_A \to \mathcal{O}_{Y_A}.$$

More precisely:

Lemma 1.3 [O'Grady 2008a]. *The map m_A is symmetric and associative, and gives an isomorphism between $\zeta_A \otimes \zeta_A$ and $\mathcal{O}_{Y_A}$.*

Thanks to the lemma we see that the sheaf

$$\mathcal{O}_{Y_A} \oplus \zeta_A$$

has the structure of an $\mathcal{O}_{Y_A}$-algebra, so we have an associated double covering.

Definition 1.4. We denote by X_A this double covering; the scheme X_A is called a *double EPW sextic*. We denote by

$$f_A : X_A \to Y_A$$

the covering map.

The scheme X_A is endowed with a polarization $h_A = f_A^* \mathcal{O}_{Y_A}(1)$.

Remark 1.5. The ramification locus of the map f_A is $Y_A[2]$. To see this we just need to observe that by construction the ramification locus is the locus where the sheaf ζ_A, or equivalently the sheaf ξ_A, is not locally free. Since $i_{A*}(\xi_A)$ is the cokernel of the map

$$\lambda_A : F \to \mathcal{O}_{\mathbb{P}(V)} \otimes A^\vee,$$

we see that the rank of ξ_A jumps exactly along $Y_A[2]$, hence our claim.

Definition 1.6 [O'Grady 2010]. We let

$$\Sigma = \left\{A \in \mathbb{LG}(\bigwedge^3 V) \mid \bigwedge^3(W) \subset A \text{ for some } W \subset V,\ \dim W = 3\right\}.$$

In other words, Σ is the set of Lagrangian subspaces of $\bigwedge^3 V$ containing a decomposable form. We also let

$$\Delta = \left\{A \in \mathbb{LG}(\bigwedge^3 V) \mid Y_A[3] \neq \varnothing\right\} \subset \mathbb{LG}(\bigwedge^3 V).$$

Finally we define

$$\mathbb{LG}(\bigwedge^3 V)^0 = \mathbb{LG}(\bigwedge^3 V) \setminus (\Sigma \cup \Delta).$$

We recall the following characterization of smooth double EPW sextics:

Theorem 1.7 [O'Grady 2010]. *The double covering X_A is smooth if and only if*

$$A \in \mathbb{LG}(\bigwedge^3 V)^0.$$

The relevance of these double coverings stems from the following result.

Theorem 1.8 [O'Grady 2010]. *Let $A \in \mathbb{LG}(\bigwedge^3 V)^0$. Then X_A is an irreducible symplectic variety. The Beauville–Bogomolov form on $H^2(X, \mathbb{Z})$ is the same as that of $S^{[2]}$, where S is a $K3$ surface, and the Fujiki constant of X_A is 3.*

Let $Z_A = f_A^{-1}(Y_A[2])$; this is the branch locus for the 2:1 covering, hence it is isomorphic to $Y_A[2]$ itself. Since the covering involution is antisymplectic, the symplectic form restricts to 0 on Z_A, that is, Z_A is isotropic. Under mild assumptions Z_A is a surface, hence a Lagrangian surface inside X_A. More precisely:

Proposition 1.9 [O'Grady 2010]. *Let $A \in \mathbb{LG}(\bigwedge^3 V)^0$. Then $Y_A[2]$ is a smooth connected surface of degree 40, with $\chi_{\text{top}}(Y_A[2]) = 192$.*

We will need the following relation in the Chow group. It appears in [O'Grady 2008b, Proposition 4.9], with a different proof from the one given below.

Proposition 1.10. *Let $A \in \mathbb{LG}(\bigwedge^3 V)^0$, $Z = Z_A$. The canonical class of Z satisfies*

$$2K_Z = \mathbb{O}_Z(6)$$

in $CH^*(Z)$.

Remark 1.11. The proposition determines K_Z only up to 2-torsion. Namely we can rewrite it as

$$K_Z = \mathbb{O}_Z(3) + \kappa,$$

where κ is a 2-torsion class. One can use the deformation argument in [Ferretti 2012] and the results of [Welters 1981] to show that the class κ is really nonzero.

Proof. For simplicity let $W = f(Z)$ denote the singular set of Y. We know that on W the map λ has constant rank 8, so we get the following exact sequence of *vector bundles* on W:

$$0 \longrightarrow \mathcal{H} \longrightarrow F \xrightarrow{\lambda|_W} \mathcal{O}_W \otimes (\bigwedge^3 V/A) \longrightarrow \zeta|_W \longrightarrow 0. \tag{1-5}$$

Here $\mathcal{H}$ is defined to be the kernel of $\lambda|_W$; it has rank 2. Identifying W with its preimage $Z \subset X$, we claim that the following isomorphisms hold:

$$\zeta|_W \cong \mathcal{N}_{Z/X}, \tag{1-6}$$

$$\mathcal{H} \cong \mathcal{N}^\vee_{Z/X}. \tag{1-7}$$

Assuming (1-6) and (1-7) for a moment, the exact sequence in (1-5) gives

$$c_1(\mathcal{N}^\vee_{Z/X}) - c_1(F) - c_1(\mathcal{N}_{Z/X}) = 0,$$

hence

$$2c_1(\mathcal{N}_{Z/X}) = -c_1(F) = \mathcal{O}_Z(6).$$

Since X has trivial canonical class, it follows that

$$2K_Z = 2c_1(\mathcal{N}_{Z/X}) = \mathcal{O}_Z(6),$$

as desired.

So we now turn to the proof of (1-6) and (1-7). Let $p \in Z$; then the covering involution φ fixes p, so φ^* acts on T_pX. This gives a decomposition

$$T_pX = (T_pX)_+ \oplus (T_pX)_-$$

in eigenspaces for φ^*, with eigenvalues ± 1. Since Z is the fixed locus of φ,

$$(T_pX)_+ = T_pZ.$$

On the other hand, since

$$X = \mathrm{Spec}(\mathcal{O}_Y \oplus \zeta),$$

we can identify

$$(T_pX)_- \cong \zeta_{f(p)}.$$

It follows that

$$(\mathcal{N}_{Z/X})_p \cong \zeta_{f(p)};$$

this fiber-wise identification is easily seen to globalize, hence yielding the isomorphism in (1-6).

For the other, we show that $\mathcal{H} \cong \zeta|^\vee_W$. Indeed observe that over W we have

$$\mathcal{H}_v = F_v \cap A \quad \text{and} \quad \zeta_v = \bigwedge^3 V/(F_v + A).$$

The symplectic form identifies $\mathcal{K}_v^\vee$ with the quotient $\bigwedge^3 V/(F_v \cap A)^\perp$, and since both A and F_v are Lagrangian we have

$$(F_v \cap A)^\perp = F_v^\perp + A^\perp = F_v + A,$$

thereby proving the isomorphism (1-7). □

Corollary 1.12. *For $A \in \mathbb{LG}(\bigwedge^3 V)^0$ the surface $Z_A \cong Y_A[2]$ is of general type.*

Plan of the paper. Before turning to the proof of Theorem 1.1, we explain the overall plan. Let $X = X_A$ be a smooth double EPW sextic, and denote its polarization $h = h_A$. The symplectic form gives an isomorphism

$$T_X \cong \Omega_X^1,$$

hence the odd Chern classes vanish. So we only need to consider $c_2(X)$ and $c_4(X)$. Moreover if A is very general in $\mathbb{LG}(\bigwedge^3 V)$, the group $\mathrm{Pic}(X_A)$ is cyclic, generated by h_A, so the second conclusion of Theorem 1.1 follows from the first.

The only relations in cohomology can be in degree 4, 6, or 8. Lemma 2.3 excludes the existence of relations of degree 4, hence we are left with relations in degree 6 or 8; these are listed in Propositions 2.1 and 2.2.

Since h^4, $c_2(X)\cdot h$, $c_2(X)^2$, and $c_4(X)$ are all proportional in cohomology, there must be some distinguished 0-cycle θ on X, such that all these classes are multiples of θ in $CH^4(X)$. We shall define θ as the class of any point on a suitable surface inside X_A; actually it will be easier to work with Y_A and pull back everything to X_A later.

Hence we look for a surface $S \subset X$ such that $CH^2(S)$ is trivial, so each point on S is rationally equivalent to each other. For instance, in the proof of the conjecture in the case where X is the Fano variety of a cubic fourfold, Voisin [2008] used a rational surface. In that case there is a family of Lagrangian surfaces on X, which are simply the Fano varieties of hyperplane sections of the cubic; if the section is singular enough, its Fano variety turns out to be rational.

In our case this construction is a delicate point: the analogue of S is an Enriques surface, but exhibiting it is complicated. The construction that we need is provided in [Ferretti 2012].

So the plan is as follows. In Section 2 we carry out the cohomology computations on X. In Section 3 we use the surface constructed in [Ferretti 2012] to define the class θ. In the rest of the paper we find enough relations in the Chow ring to finish the proof of the main theorem.

2. Cohomology computations

Let $X = X_A$ be a smooth double EPW sextic. In this section we compute the cohomological invariants of X, partly following [O'Grady 2008b]. We shall find

all relations in cohomology between h and the Chern classes of X. In later sections we shall show that these relations hold in the Chow ring.

Let σ be the symplectic form on X. Since the canonical of X is trivial

$$H^{4,0}(X) = H^0(X, \Omega_X^4)$$

is generated by σ^2. Moreover it is known that $H^3(X) = 0$, so we can compute the Euler characteristic

$$\chi(X, \mathcal{O}_X) = h^{0,0}(X) + h^{2,0}(X) + h^{4,0}(X) = 3.$$

The symplectic form on X gives an isomorphism

$$T_X \cong \Omega_X^1,$$

hence the odd Chern classes vanish. The Hirzebruch–Riemann–Roch theorem for X simplifies to

$$3 = \chi(X, \mathcal{O}_X) = \tfrac{1}{240}\left(c_2(X)^2 - \tfrac{1}{3}c_4(X)\right). \tag{2-1}$$

We introduce some more notation. Let us call

$$q \in \mathrm{Sym}^2(H^2(X, \mathbb{Q})^\vee)$$

the Beauville–Bogomolov form of X. Since it is nondegenerate, it allows us to give an identification

$$H^2(X, \mathbb{Q}) \cong H^2(X, \mathbb{Q})^\vee;$$

hence we obtain a dual quadratic form

$$q^\vee \in \mathrm{Sym}^2(H^2(X, \mathbb{Q})).$$

From [Verbitsky 1996, Theorem 1.5] (together with the computation of the Betti numbers in [Göttsche 1990]) we know that the cup product yields an isomorphism between $\mathrm{Sym}^2(H^2(X, \mathbb{Q}))$ and $H^4(X, \mathbb{Q})$, so we can regard $q^\vee$ as an element of $H^4(X, \mathbb{Q})$.

O'Grady [2008b] proves that we have the relation

$$q^\vee = \tfrac{5}{6}c_2(X), \tag{2-2}$$

and that for any $\alpha, \beta \in H^2(X, \mathbb{Q})$ we have

$$q^\vee \cdot \alpha \cdot \beta = 25q(\alpha, \beta). \tag{2-3}$$

We now work out the relations in the cohomology of X. Let

$$h = c_1(f^*\mathcal{O}_Y(1)) \in H^2(X).$$

Proposition 2.1. *In the cohomology ring $H^*(X, \mathbb{Q})$ we have*

$$h^4 = 12, \quad h^2 \cdot c_2(X) = 60, \quad c_2(X)^2 = 828, \quad c_4(X) = 324.$$

Proof. The first and the last relations are easily handled. Indeed,

$$h^4 = 2\deg(Y) = 12.$$

As for the last one we have

$$c_4(X) = \chi(X),$$

and since X is a deformation of $S^{[2]}$, where S is a K3, we have

$$\chi(X) = \chi(S^{[2]}) = 324.$$

By O'Grady's computations (2-3) and (2-2) we also have

$$c_2(X) \cdot h^2 = \tfrac{6}{5} q^\vee \cdot h^2 = \frac{25 \cdot 6}{5} q(h, h) = 60.$$

Finally we can use (2-1) to obtain $c_2(X)^2 = 828$. □

In degree 6 the only possible relation is a linear dependency between h^3 and $c_2(X) \cdot h$, and indeed we have:

Proposition 2.2. *There is a relation*

$$c_2(X) \cdot h = 5h^3$$

$H^6(X, \mathbb{Q})$.

Proof. From O'Grady's relation (2-3) we get

$$6q^\vee \cdot h \cdot \alpha = 6 \cdot 25 q(h, \alpha)$$

for all $\alpha \in H^2(X)$. On the other hand, by polarization of Fujiki's relation we obtain

$$25h^3 \cdot \alpha = 25 \cdot 3 \cdot q(h, h) q(h, \alpha) = 6 \cdot 25 q(h, \alpha).$$

So Poincaré duality implies that

$$25h^3 = 6q^\vee \cdot h$$

modulo torsion, and using (2-2) we get the thesis. □

We can instead exclude relations in degree 4:

Lemma 2.3. *The classes h^2 and $c_2(X)$ are linearly independent inside $H^2(X)$.*

This fact appears inside the proof of [O'Grady 2008b, Claim 3.1], but we add a short proof for completeness.

Proof. We can substitute $c_2(X)$ with its multiple $q^\vee$. Assume that we have a relation

$$h^2 + \lambda q^\vee = 0$$

for some $\lambda \in \mathbb{C}$. Then we get

$$h^2\alpha^2 = -25\lambda q(\alpha, \alpha)$$

for all $\alpha \in H^2(X)$. By polarization of the Fujiki formula we also obtain

$$h^2\alpha^2 = q(\alpha, \alpha)q(h, h) + 2q(h, \alpha)^2.$$

So if $q(\alpha, \alpha) = 0$ we obtain $q(h, \alpha) = 0$. This means that q is degenerate (the quadric defined by q would be contained in a hyperplane of $\mathbb{P}H^2(X)$), which is a contradiction. □

Finally, it will be useful to write out the explicit form of Hirzebruch–Riemann–Roch, using the above computations for the characteristic classes of X. We let

$$\mathcal{O}_X(1) = f^*\mathcal{O}_Y(1).$$

Then $\mathcal{O}_X(n)$ is ample on X, and since K_X is trivial, Kodaira vanishing yields

$$\chi(X, \mathcal{O}_X(n)) = h^0(X, \mathcal{O}_X(n)).$$

The formula of Hirzebruch–Riemann–Roch then reads

$$h^0(X, \mathcal{O}_X(n)) = \frac{h^4}{24}n^4 + \frac{c_2(X)\cdot h^2}{24}n^2 + \chi(\mathcal{O}_X) = \tfrac{1}{2}n^4 + \tfrac{5}{2}n^2 + 3. \qquad (2\text{-}4)$$

3. Definition of the class θ

Let $X = X_A$ be a double EPW sextic, $f : X \to Y$ the double covering. Our first task is to define a class

$$\theta \in CH^4(X)$$

of degree 1. Then we will show that the relations

$$h^4 = 12\theta, \quad h^2c_2(X) = 60\theta, \quad c_2(X)^2 = 828\theta, \quad c_4(X) = 324\theta$$

hold.

It will actually be easier to work on Y, so we should find the relationship between $CH(X)$ and $CH(Y)$.

Remark 3.1. The map $f : X \to Y$ induces a push-forward morphism

$$f_* : CH(X) \to CH(Y),$$

because f is proper (for the construction of Chow rings and morphisms between them see [Fulton 1984, Chapter 1]). On the other hand f^* is usually defined for flat

maps with fibers of constant dimension, or when the target is smooth, and neither is the case.

Following [Fulton 1984, Example 1.7.6] we can define f^* in our situation. Indeed, Fulton shows that if

$$Y = X/G$$

is the quotient of X by the action of a finite group G, we have a canonical isomorphism

$$CH(Y)_{\mathbb{Q}} \cong CH(X)^G_{\mathbb{Q}},$$

where as usual $CH(Y)_{\mathbb{Q}} = CH(Y) \otimes \mathbb{Q}$. So if f is the quotient map we can define f^* by the composition

$$CH(Y)_{\mathbb{Q}} \xrightarrow{\cong} CH(X)^G_{\mathbb{Q}} \hookrightarrow CH(X)_{\mathbb{Q}}.$$

Fulton also shows that the composition

$$CH(Y)_{\mathbb{Q}} \xrightarrow{f^*} CH(X)_{\mathbb{Q}} \xrightarrow{f_*} CH(Y)_{\mathbb{Q}}$$

is the multiplication map by $\sharp G$.

In our situation $G = \langle \varphi \rangle$, where φ is the covering involution, and the composition above is multiplication by 2.

Moreover we observe that

$$CH(Y)_{\mathbb{Q}} \cong CH(X)^{\mathbb{Z}/2\mathbb{Z}}_{\mathbb{Q}}$$

is a subring of $CH(X)_{\mathbb{Q}}$, so we can multiply cycle classes on Y even if it is singular.

We now recall the main result in [Ferretti 2012]. There we produce a surface $S \subset Y$ such that $CH^2(S) = \mathbb{Z}$; this will be the starting point for our investigation of $CH^*(Y)$. Let us fix a Lagrangian subspace $A \in \mathbb{LG}(\bigwedge^3 V)$. If B is another Lagrangian subspace such that

$$\dim(A \cap B) = 9, \tag{3-1}$$

it follows from the Grassmann formula that $Y_B[2] \subset Y_A$.

Theorem [Ferretti 2012]. *Let A be a Lagrangian subspace in $\mathbb{LG}(\bigwedge^3 V)^0$. Then there exists another Lagrangian subspace B such that* $\dim(A \cap B) \geq 9$, *and such that $S := Y_B[2]$ is either a (singular) Enriques surface, or a degeneration of such surfaces. In either case we have*

$$CH^2(S) \cong \mathbb{Z}.$$

Moreover we have the relation

$$[S] = \big[Y_A[2]\big] \tag{3-2}$$

in $CH^2(X_A)$.

So for such a choice of B we know that all points on $S = Y_B[2]$ are rationally equivalent. We define $\bar{\theta}$ as the class of any such point. We need to do some checks in order to show that this is actually well-defined. We also define

$$\theta = \tfrac{1}{2} f^*(\bar{\theta}) \in CH^4(X_A)_{\mathbb{Q}}.$$

Lemma 3.2. *Let* $B, B' \in \mathbb{LG}(\bigwedge^3 V)$ *such that* (3-1) *holds. Then*

$$Y_B[2] \cap Y_{B'}[2] \neq \varnothing. \tag{3-3}$$

Proof. It is enough to show that

$$Y_B[2] \cdot Y_{B'}[2] \neq 0$$

in $CH^*(Y_A)$. Thanks to (3-2) it will be enough to prove that

$$Y_A[2]^2 \neq 0.$$

By the definition of the ring structure on $CH^*(Y_A)$ we need to prove that

$$Z_A^2 \neq 0 \quad \text{in } CH^*(X_A).$$

But actually $Z_A^2 \neq 0$ already in cohomology. Indeed, using the fact that Z_A is Lagrangian, we have

$$Z_A^2 = c_2(\mathcal{N}_{Z_A/X_A}) = c_2(\Omega^1_{Z_A}) = c_2(Z_A) = \chi_{\text{top}}(Z_A) = 192$$

by Proposition 1.9. □

By the previous lemma we see that the class of $\bar{\theta} \in CH^4(Y_A)$ is actually independent of the chosen Lagrangian subspace B such that (3-1) holds.

4. Relations coming from vector bundle morphisms

In this section we omit A from the notation, for clarity. Fix a Lagrangian subspace $A \in \mathbb{LG}(\bigwedge^3 V)^0$ and denote $X = X_A$, $Y = Y_A$, and $Z = Z_A$. We shall exhibit a number of relations coming from exact sequences of sheaves on X and Y.

Lemma 4.1. *The following relation holds in* $CH(X)$:

$$3Z = 15h^2 - c_2(X).$$

Proof. We consider f as a map $X \to \mathbb{P}^5$, so it induces a morphism of vector bundles over X:

$$df : T_X \to f^* T_{\mathbb{P}^5}.$$

We notice that df is injective outside Z, so we can see Z as a degeneracy locus for this morphism. We then apply the Thom–Porteous formula in the form stated in [Fulton 1984, §14.4]. In their notation we have $e = 4$, $f = 5$, and $k = 3$.

This yields a cycle class

$$\mathbb{D}_3(df) \in CH^2(Z)$$

whose support is Z, and such that the image of $\mathbb{D}_3(df)$ in $CH^2(X)$ is

$$\Delta_2^{(1)}(c(f^*T_{\mathbb{P}^5} - T_X)) = c_2(f^*T_{\mathbb{P}^5} - T_X).$$

Here the total Chern class

$$c(f^*T_{\mathbb{P}^5} - T_X)$$

is defined formally in such a way that Whitney's formula holds, that is,

$$c(T_X) \cdot c(f^*T_{\mathbb{P}^5} - T_X) = c(f^*T_{\mathbb{P}^5}).$$

From the last equation and the fact that $c_1(T_X) = 0$ (since X is symplectic) we can obtain

$$c_2(f^*T_{\mathbb{P}^5} - T_X) = f^*c_2(T_{\mathbb{P}^5}) - c_2(T_X) = 15h^2 - c_2(X).$$

Since $\mathbb{D}_3(df)$ has support on Z, which is irreducible, we find that

$$kZ = 15h^2 - c_2(X) \tag{4-1}$$

for some $k \in \mathbb{Z}$. To find the right k, we observe that, again by [Fulton 1984, Theorem 14.4(c)], we have

$$\mathbb{D}_3(df) = [D_3(df)],$$

where $D_3(df)$ is the degeneracy locus of df. In other words $D_3(df)$ is just Z, with the scheme structure given by the vanishing of all 4×4 minors of df.

The map

$$f : X \to Y \subset \mathbb{P}^5$$

has, in suitable analytic coordinates around a point of Z, the local form

$$f(x, y, z, t) \underset{\text{loc}}{=} (x^2, xy, y^2, z, t).$$

The differential of f is then

$$df \underset{\text{loc}}{=} \begin{pmatrix} 2x & 0 & 0 & 0 \\ y & x & 0 & 0 \\ 0 & 2y & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

equating to 0 the determinants of its 3×3 minors yields

$$D_3(df) = \underset{\text{loc}}{V}(x^2, xy, y^2).$$

So we see that $D_3(df)$ has multiplicity 3 at each point of Z, hence $k = 3$.

Alternatively we could multiply (4-1) by h^2 to find

$$kZ \cdot h^2 = 15h^4 - c_2(X) \cdot h^2.$$

If we look at this relation in cohomology it becomes, thanks to Proposition 2.1,

$$40k = 15 \cdot 12 - 60,$$

so $k = 3$. □

We take a closer look at the differential of

$$f : X \to \mathbb{P}^5.$$

As a map of vector bundles, this is not injective exactly on Z. Hence it is always injective on stalks; in other words

$$df : T_X \to f^* T_{\mathbb{P}^5}$$

is an injective map *of sheaves*. Let $\mathcal{R}$ denote its cokernel; this is locally free of rank 1 outside Z. So we have the exact sequence

$$0 \longrightarrow T_X \longrightarrow f^* T_{\mathbb{P}^5} \longrightarrow \mathcal{R} \longrightarrow 0. \tag{4-2}$$

We now dualize it applying $\mathcal{H}om(\cdot, \mathcal{O}_X)$. We remark that

$$\mathcal{H}om(\mathcal{R}, \mathcal{O}_X)$$

is torsion-free, of rank one, and one can check in local coordinates that it is a line bundle. By (4-2) we get $c_1(\mathcal{R}) = 6h$, hence

$$\mathcal{H}om(\mathcal{R}, \mathcal{O}_X) \cong \mathcal{O}_X(-6).$$

Then we note that

$$\mathcal{E}xt^1(f^*(T_{\mathbb{P}^5}), \mathcal{O}_X) = 0,$$

because both sheaves are locally free. So if we let

$$\mathcal{Q} = \mathcal{E}xt^1(\mathcal{R}, \mathcal{O}_X),$$

the dual of (4-2) becomes

$$0 \longrightarrow \mathcal{O}_X(-6) \longrightarrow f^*(\Omega^1_{\mathbb{P}^5}) \xrightarrow{df^T} \Omega^1_X \longrightarrow \mathcal{Q} \longrightarrow 0. \tag{4-3}$$

We remark that $\mathcal{Q}$ is set-theoretically supported on Z, because both $\mathcal{R}$ and $\mathcal{O}_X$ are locally free outside Z. Actually the schematic support of $\mathcal{Q}$ is $2Z$, that is, the subscheme of X defined by the ideal $\mathcal{I}_Z^2$. This follows from:

Lemma 4.2. *Let $\mathcal{Q}$ be as above; then* $\mathrm{Ann}(\mathcal{Q}) = \mathcal{I}_Z^2$.

Proof. We only need to prove this locally. As in the proof of Lemma 4.1 we can choose local coordinates on X such that

$$f(x, y, z, t) \underset{\text{loc}}{=} (x^2, xy, y^2, z, t);$$

then df^T has the matrix

$$df^T \underset{\text{loc}}{=} \begin{pmatrix} 2x & y & 0 & 0 & 0 \\ 0 & x & 2y & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

hence we have the presentation

$$\mathcal{Q} \underset{\text{loc}}{=} \frac{\langle dx, dy\rangle}{\langle x\,dx, x\,dy + y\,dx, y\,dy\rangle}.$$

A given $h(x, y) \in \mathbb{C}[x, y]$ then annihilates $\mathcal{Q}$ if and only if both $h\,dx$ and $h\,dy$ belong to the $k[x, y]$-module generated by $x\,dx$, $x\,dy + y\,dx$, and $y\,dy$.

Let us make this more explicit. Assume that

$$h(x, y)dx = a(x, y)x\,dx + b(x, y) \cdot (x\,dy + y\,dx) + c(x, y)y\,dy.$$

This yields

$$\begin{aligned} h(x, y) &= xa(x, y) + yb(x, y), \\ 0 &= xb(x, y) + yc(x, y). \end{aligned}$$

The second equation implies $b(x, y) = yb'(x, y)$, so the first becomes

$$h(x, y) = xa(x, y) + y^2b'(x, y).$$

If h can be written this way, then we can choose c so that the second condition is satisfied. In short

$$h(x, y)\,dx \in \langle x\,dx, x\,dy + y\,dx, y\,dy\rangle_{k[x,y]}$$

if and only if $h \in (x, y^2)$.

We have the symmetric condition for $h(x, y)\,dy$, so we conclude that $h \in \mathrm{Ann}(\mathcal{Q})$ if and only if

$$h \in (x, y^2) \cap (x^2, y) = (x^2, xy, y^2).$$

The last equality between ideals can be proved, for instance, by the remark that both $(x, y^2) \cap (x^2, y)$ and (x^2, xy, y^2) consist of the polynomials h such that

$$h(0,0) = \frac{\partial h}{\partial x}(0,0) = \frac{\partial h}{\partial y}(0,0) = 0.$$

Finally, (x^2, xy, y^2) is exactly the square of the ideal (x, y) which locally defines Z. □

We now produce another exact sequence involving $\mathfrak{Q}$. Let

$$i : Z \hookrightarrow X$$

denote the inclusion. Recall that we have a canonical identification

$$\mathcal{I}_Z/\mathcal{I}_Z^2 \cong i_* \mathcal{N}_{Z/X}^\vee : \tag{4-4}$$

locally the function g vanishing on Z corresponds to the normal covector dg. Consider the natural projection

$$\pi : \Omega_X^1|_Z \to \mathcal{N}_{Z/X}^\vee;$$

we see this as a map on X:

$$\pi : \Omega_X^1 \to \mathcal{I}_Z/\mathcal{I}_Z^2.$$

Lemma 4.3. *We have* $\pi \circ df^T = 0$.

Proof. We keep the notation of the proof of Lemma 4.2. We need only to verify the thesis on Z. The image of df^T is generated by

$$x\,dx,\ x\,dy + y\,dx,\ y\,dy,\ dz,\ dt.$$

The first three elements vanish on Z, while the latter two are in the kernel of π. □

The above lemma and the exact sequence in (4-3) provide us a surjective map

$$\alpha : \mathfrak{Q} \to i_*(\mathcal{N}_{Z/X}^\vee).$$

Lemma 4.4. *The kernel of* α *is* $i_*(\det T_Z)$.

Proof. We can see this explicitly in local coordinates. Keeping the notation of the above proofs, $\mathfrak{Q}$ is locally generated, on Z, by dx, dy, and $x\,dy = -y\,dx$. The conormal bundle $\mathcal{N}_{Z/X}^\vee$ is generated by dx and dy, and α is the obvious projection.

The kernel of α is then generated by $x\,dy$. Under the identification in (4-4) this corresponds to the generator $dx \wedge dy$ of $\bigwedge^2 \mathcal{N}_{Z/X}^\vee$.

So

$$\ker \alpha = i_*(\det \mathcal{N}_{Z/X}^\vee) \cong i_*(\det T_Z),$$

since Z is Lagrangian. □

Thanks to the lemma we get the exact sequence we are looking for:

$$0 \longrightarrow i_*(\det T_Z) \longrightarrow \mathcal{Q} \longrightarrow i_* T_Z \longrightarrow 0. \tag{4-5}$$

We can now find new relations in the Chow ring of X.

Proposition 4.5. *In $CH(X)_{\mathbb{Q}}$ we have*

$$c_2(X) \cdot h = 5h^3$$

and $c_4(X)$ is a linear combination of h^4, $c_2(X) \cdot h^2$, and $c_2(X)^2$.

Proof. This is just a matter of putting together the relations that come from the exact sequences (4-3) and (4-5).

We start from (4-3), which yields

$$(1-6h) \cdot (1 + c_2(X) + c_4(X)) = (1-h)^6 \cdot (1 + c_1(\mathcal{Q}) + c_2(\mathcal{Q}) + c_3(\mathcal{Q}) + c_4(\mathcal{Q})).$$

Comparing the terms in degree up to 2 we get:

$$c_1(\mathcal{Q}) = 0, \quad c_2(\mathcal{Q}) = c_2(X) - 15h^2 = -3Z, \tag{4-6}$$

where the last equality is Lemma 4.1. Then in degree 3 we have

$$c_3(\mathcal{Q}) = 6h(c_2(\mathcal{Q}) - c_2(X)) + 20h^3 = 6h \cdot (-15h^2) + 20h^3 = -70h^3, \tag{4-7}$$

using the second of (4-6). Finally in degree 4 we get, using (4-6) and (4-7),

$$c_4(X) = 15h^4 + 15h^2 \cdot c_2(\mathcal{Q}) - 6h \cdot c_3(\mathcal{Q}) + c_4(\mathcal{Q}) = 15h^4 - 45h^2 \cdot Z + 420h^4 + c_4(\mathcal{Q}),$$

hence

$$c_4(\mathcal{Q}) = c_4(X) - 435h^4 + 45h^2 \cdot Z. \tag{4-8}$$

Next we look at the relations coming from (4-5). To do this we shall use Grothendieck–Riemann–Roch, which for the closed embedding

$$i : Z \hookrightarrow X$$

takes the form

$$\mathrm{ch}(i_* \mathcal{F}) = i_*(\mathrm{ch}(\mathcal{F}) \cdot \mathrm{td}(\mathcal{N}_{Z/X})^{-1}),$$

for any $\mathcal{F} \in \mathrm{Coh}(Z)$. This is because in our situation we have

$$R^k i_*(\mathcal{F}) = 0$$

for all such $\mathcal{F}$, thanks to [Hartshorne 1977, Corollary III.11.2].

Using that Z is Lagrangian we have $\mathcal{N}_{Z/X} \cong \Omega^1_Z$, so we can compute

$$\mathrm{td}(\mathcal{N}_{Z/X}) = 1 - \tfrac{1}{2} c_1(Z) + \tfrac{1}{12}(c_1(Z)^2 + c_2(Z)),$$
$$\mathrm{td}(\mathcal{N}_{Z/X})^{-1} = 1 + \tfrac{1}{2} c_1(Z) + \tfrac{1}{6} c_1(Z)^2 - \tfrac{1}{12} c_2(Z).$$

Then we have

$$\operatorname{ch}(\det T_Z) = 1 + c_1(Z) + \tfrac{1}{2}c_1(Z)^2,$$
$$\operatorname{ch}(T_Z) = 2 + c_1(Z) + \tfrac{1}{2}(c_1(Z)^2 - c_2(Z)).$$

So Grothendieck–Riemann–Roch for these sheaves becomes

$$\operatorname{ch}(i_* \det T_Z) = i_*\big(1 + \tfrac{3}{2}c_1(Z) + \tfrac{7}{6}c_1(Z)^2 - \tfrac{1}{12}c_2(Z)\big),$$
$$\operatorname{ch}(i_* T_Z) = i_*\big(2 + 2c_1(Z) + \tfrac{4}{3}c_1(Z)^2 - \tfrac{7}{6}c_2(Z)\big).$$

Next we use the fact that in $CH(Z)_{\mathbb{Q}}$ we have

$$c_1(Z) = -K_Z = -3i^*(h),$$

thanks to Proposition 1.10. So we obtain

$$\operatorname{ch}(i_* \det T_Z) = Z - \tfrac{9}{2}h \cdot Z + \tfrac{21}{2}h^2 \cdot Z - \tfrac{1}{12}Z^2,$$
$$\operatorname{ch}(i_* T_Z) = 2Z - 6h \cdot Z + 12h^2 \cdot Z - \tfrac{7}{6}Z^2.$$

We can use this to recover the Chern classes of $i_*(\det T_Z)$ and $i_*(T_Z)$. These are:

$$c_1(i_* \det T_Z) = 0, \qquad c_2(i_* \det T_Z) = -Z,$$
$$c_3(i_* \det T_Z) = -9h \cdot Z, \qquad c_4(i_* \det T_Z) = Z^2 - 63h^2 \cdot Z,$$

and

$$c_1(i_* T_Z) = 0, \qquad c_2(i_* T_Z) = -2Z,$$
$$c_3(i_* T_Z) = -12h \cdot Z, \qquad c_4(i_* T_Z) = 9Z^2 - 72h^2 \cdot Z.$$

Finally we use the exact sequence (4-5) to get the Chern classes of $\mathbb{Q}$. The first two are

$$c_1(\mathbb{Q}) = 0, \quad c_2(\mathbb{Q}) = -3Z,$$

in accordance with (4-6). Then we get

$$c_3(\mathbb{Q}) = -21h \cdot Z,$$

and comparing with (4-7) we obtain

$$-3h \cdot Z = -10h^3.$$

Using Lemma 4.1 this is equivalent to

$$c_2(X) \cdot h = 5h^3.$$

Finally we get

$$c_4(\mathbb{Q}) = 12Z^2 - 135h^2 \cdot Z;$$

comparing with (4-8) this yields

$$12Z^2 - 135h^2 \cdot Z = c_4(X) - 435h^4 + 45h^2 \cdot Z,$$

and using again Lemma 4.1 to write Z as a rational combination of $c_2(X)$ and h^2, we get the second claim of the thesis. □

5. Conclusion of the proof

First we recall that we have defined the class

$$\theta = \tfrac{1}{2} f^*(\bar{\theta}).$$

Here $\bar{\theta}$ is the class of any point on a surface $S \subset Y_A$. By (3-2) we know that

$$\big[Y_A[2]\big] = [S] \quad \text{in } CH^2(Y_A).$$

We also let $\bar{h} = \mathcal{O}_Y(1)$, so that $h = f^*(\bar{h})$.

Lemma 5.1. *There exists a line $L_0 \subset Y$ which meets $Y_B[2]$.*

Proof. Let V be the union of lines contained in Y.

Step 1: $\dim V \geq 2$. Let $R \subset \mathrm{Gr}(2, V)$ be the locus of lines $\ell \subset Y_A$. We can obtain R as follows. Let

$$Y_A = V(g),$$

where g is a degree-6 polynomial, and let $\mathcal{S}$ be the tautological subbundle on $\mathrm{Gr}(2, V)$, so that $\mathrm{Sym}^6(\mathcal{S}^\vee)$ is the fiber bundle whose fiber at ℓ is the vector space of homogeneous polynomials of degree 6 on ℓ.

Then we can define a section

$$s \in H^0(\mathrm{Gr}(2, V), \mathrm{Sym}^6(\mathcal{S}^\vee))$$

by the condition

$$s(\ell) = g|_\ell.$$

By definition R is the zero locus of s. It follows that

$$\dim R \geq \dim \mathrm{Gr}(2, V) - \mathrm{rk}\, \mathrm{Sym}^6(\mathcal{S}^\vee) = 8 - 7 = 1,$$

provided R is not empty. But we can show that $R \neq \varnothing$ by computing the fundamental class

$$[R] = c_7(\mathrm{Sym}^6(\mathcal{S}^\vee)) = 432 \cdot 134\sigma_{4,3}.$$

Here the notation is that of Schubert calculus, see for instance [Griffiths and Harris 1978, §1.5].

Since $V = \bigcup_{\ell \in R} \ell$ is birational to a $\mathbb{P}^1$-bundle over R, it follows that $\dim V \geq 2$.

Step 2: There exists B' such that $A \cap B = A \cap B'$ and $Y_{B'}[2]$ meets V. Let

$$U = A \cap B.$$

In [Ferretti 2012] we have shown that there exists a divisor $D_U \subset Y_A$ which is covered by the surfaces $[Y_{B'}[2]]$ as B' ranges through the Lagrangian subspaces which satisfy

$$A \cap B' \supset U.$$

In particular D_U has dimension 3; since two varieties of dimensions 2 and 3 in $\mathbb{P}^5$ always meet, it follows that

$$D_U \cap V \neq \varnothing.$$

So there exists a Lagrangian subspace B' such that $B' \cap A = U$ and

$$Y_{B'}[2] \cap V \neq \varnothing.$$

Step 3: B meets V. We lift everything to X_A, which is smooth, so intersection theory applies. Let

$$\widetilde{V}_1 = f^{-1}(V) \quad \text{and} \quad \widetilde{V}_2 = f^{-1}(Y_{B'}[2]).$$

One easily sees that on X

$$\widetilde{V}_1 \cdot \widetilde{V}_2 \neq 0.$$

Since $f^{-1}(Y_B[2])$ and $\widetilde{V}_2$ have the same homology class, it follows that

$$\widetilde{V}_1 \cdot f^{-1}(Y_B[2]) \neq 0;$$

in particular $\widetilde{V}_1$ must meet $f^{-1}(Y_B[2])$, and so

$$V \cap Y_B[2] \neq \varnothing. \qquad \square$$

Using Lemma 5.1 we can start proving that

$$h^4 = 6\theta \quad \text{in } CH(X). \tag{5-1}$$

Indeed let L_0 be any line meeting S and let Λ be any plane containing L_0. Then $\bar{h}^3$ is represented by the intersection

$$\Lambda \cdot Y = L_0 + C,$$

where C is a quintic on Λ. Multiplying by $\bar{h}$ we obtain

$$\bar{h}^4 = L_0 \cdot \bar{h} + C \cdot \bar{h}.$$

We claim that this is represented by a 0-cycle supported on L_0. This is clear for the first addend; for the second we represent $\bar{h}$ by a hyperplane containing L_0 and transverse to Λ. It follows that $C \cdot \bar{h}$ is supported on $C \cap L_0$.

Since L_0 is rational, $CH^1(L_0) \cong \mathbb{Z}$, so $\bar{h}^4$ is rationally equivalent to a multiple of a point of L_0. Finally $L_0 \cap S \neq \varnothing$, so we get

$$\bar{h}^4 = k\bar{\theta} \quad \text{in } CH^4(Y)_{\mathbb{Q}}$$

for some $k \in \mathbb{Q}$.

Pulling back this relation to X and using $f^*(\bar{h}) = h$, $f^*(\bar{\theta}) = 2\theta$ we see that $h^4 = 2k\theta$ in $CH^4(X)_{\mathbb{Q}}$. Since in cohomology we have $h^4 = 12$ we must have $k = 6$, and so (5-1) is proved.

Next we show that

$$h^2 \cdot c_2(X) = 60\bar{\theta}. \tag{5-2}$$

We start from Lemma 4.1; pushing forward that relation we get

$$3\big[Y_A[2]\big] = 15 \cdot 4\bar{h}^2 - f_* c_2(X) \quad \text{in } CH^2(Y). \tag{5-3}$$

Multiplying (5-3) by $\bar{h}^2$ we get

$$\bar{h}^2 \cdot f_* c_2(X) = 60\bar{h}^4 - 3\bar{h}^2 \cdot \big[Y_A[2]\big].$$

We already proved that $\bar{h}^4$ is a multiple of $\bar{\theta}$, and the cycle class

$$\bar{h}^2 \cdot \big[Y_A[2]\big] = \bar{h}^2 \cdot [S]$$

is supported on S, hence it is a rational multiple of $\bar{\theta}$ too.

We conclude that the relation (5-2) holds up to a multiple, that is,

$$\bar{h}^2 \cdot f_* c_2(X) = k\bar{\theta}.$$

As before, we pull back this relation to X in order to make computations in cohomology. We get

$$h^2 \cdot 2c_2(X) = 2k\bar{\theta}.$$

Since in cohomology we have $h^2 \cdot c_2(X) = 60$, we must have $k = 60$, and (5-2) is proved.

In a similar way, we can rewrite (5-3) as

$$f_* c_2(X) = 15\bar{h}^2 - 3\big[Y_A[2]\big]$$

and take squares to write $(f_* c_2(X))^2$ as a combination of $\bar{h}^4$ and a 0-cycle supported on $Y_B[2]$. This shows that $(f_* c_2(X))^2$ is a rational multiple of $\bar{\theta}$.

As usual a cohomology computation yields the precise form

$$c_2(X)^2 = 828\theta.$$

Now we can use Proposition 4.5 to conclude that $c_4(X) = k\theta$, and finally we get $k = 324$ by comparison with the analogous computation in cohomology. This takes care of all relations in degree 8.

The only relation in degree 6 comes from Proposition 2.2, and is

$$c_2(X) \cdot h = 5h^3.$$

We already proved that the same holds in $CH^*(X)$ in Proposition 4.5, so this ends the proof of the main theorem.

References

[Beauville 2007] A. Beauville, "On the splitting of the Bloch–Beilinson filtration", pp. 38–53 in *Algebraic cycles and motives*, vol. 2, edited by J. Nagel and C. Peters, London Math. Soc. Lecture Note Ser. **344**, Cambridge University Press, 2007. MR 2009c:14007 Zbl 1130.14006

[Beauville and Voisin 2004] A. Beauville and C. Voisin, "On the Chow ring of a $K3$ surface", *J. Algebraic Geom.* **13**:3 (2004), 417–426. MR 2005b:14011 Zbl 1069.14006

[Ferretti 2009] A. Ferretti, *The Chow ring of double EPW sextics*, thesis, La Sapienza Università di Roma, 2009. To appear in *Rend. Mat. Appl.* (7), **31**:3–4, http://www.mat.uniroma1.it/ricerca/rendiconti/rendiconti.html.

[Ferretti 2012] A. Ferretti, "Special subvarieties of EPW sextics", *Math. Z.* (2012). To appear.

[Fulton 1984] W. Fulton, *Intersection theory*, Ergebnisse der Mathematik (3) **2**, Springer, Berlin, 1984. MR 85k:14004 Zbl 0541.14005

[Göttsche 1990] L. Göttsche, "The Betti numbers of the Hilbert scheme of points on a smooth projective surface", *Math. Ann.* **286** (1990), 193–207. MR 91h:14007 Zbl 0679.14007

[Griffiths and Harris 1978] P. Griffiths and J. Harris, *Principles of algebraic geometry*, Wiley, New York, 1978. MR 80b:14001 Zbl 0408.14001

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

[O'Grady 2006] K. O'Grady, "Irreducible symplectic 4-folds and Eisenbud–Popescu–Walter sextics", *Duke Math. J.* **134**:1 (2006), 99–137. MR 2007e:14062 Zbl 1105.14051

[O'Grady 2008a] K. O'Grady, "Dual double EPW-sextics and their periods", *Pure Appl. Math. Q.* **4**:2 (2008), 427–468. MR 2009b:14076 Zbl 1152.14010

[O'Grady 2008b] K. O'Grady, "Irreducible symplectic 4-folds numerically equivalent to $(K3)^{[2]}$", *Commun. Contemp. Math.* **10**:4 (2008), 553–608. MR 2009f:32037

[O'Grady 2010] K. O'Grady, "EPW-sextics: taxonomy", preprint, 2010. arXiv 1007.3882

[Verbitsky 1996] M. Verbitsky, "Cohomology of compact hyper–Kähler manifolds and its applications", *Geom. Funct. Anal.* **6**:4 (1996), 601–611. MR 98a:53069

[Voisin 2008] C. Voisin, "On the Chow ring of certain algebraic hyper-Kähler manifolds", *Pure Appl. Math. Q.* **4**:3 (2008), 613–649. MR 2009m:14004 Zbl 1165.14012

[Welters 1981] G. E. Welters, *Abel–Jacobi isogenies for certain types of Fano threefolds*, Mathematical Centre Tracts **141**, Mathematisch Centrum, Amsterdam, 1981. MR 84k:14035 Zbl 0474.14028

Communicated by Ravi Vakil
Received 2010-10-20 Revised 2011-01-19 Accepted 2011-02-22

ferrettiandrea@gmail.com *Laboratoire Paul Painlevé, Université Lille1, 59655 Villeneuve d'Ascq CEDEX, France*
http://www.andreaferretti.it/

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

A finiteness property of graded sequences of ideals

Mattias Jonsson and Mircea Mustaţă

Given a graded sequence of ideals $(\mathfrak{a}_m)_{m\geq 1}$ on X, having finite log canonical threshold, we show that if there are divisors E_m over X computing the log canonical threshold of $\mathfrak{a}_m$, and such that the log discrepancies of the divisors E_m are bounded, then the set $\{E_m \mid m \geq 1\}$ is finite.

1. Introduction

Let X be a smooth algebraic variety over an algebraically closed field k of characteristic zero. The log canonical threshold of a nonzero ideal $\mathfrak{a}$ on X is a fundamental invariant of the singularities of the subscheme defined by $\mathfrak{a}$. Originally known as the *complex singularity index*, it shows up in many contexts related to singularities, and it has found a plethora of applications in birational geometry [Kollár 1997; Ein and Mustaţă 2006].

In this note we will be interested in the behavior of this invariant in certain sequences of ideals. Let $\mathfrak{a}_\bullet = (\mathfrak{a}_m)_{m\geq 1}$ be a *graded sequence of ideals* on X, that is, a sequence of ideals that satisfies $\mathfrak{a}_\ell \cdot \mathfrak{a}_m \subseteq \mathfrak{a}_{\ell+m}$ for every $\ell, m \geq 1$. We always assume that, in addition, some ideal $\mathfrak{a}_m$ is nonzero. The main motivating example is the graded sequence $\mathfrak{a}_\bullet^L$ associated to a line bundle L of nonnegative Iitaka dimension on a smooth projective variety X: the ideal $\mathfrak{a}_m^L$ defines the base-locus of the linear system $|L^m|$. Note that in this case the behavior of $\mathfrak{a}_\bullet^L$ is easy to understand if the section ring $\bigoplus_m \Gamma(X, L^m)$ is finitely generated over k. Indeed, in this case there is a positive integer p such that $\mathfrak{a}_{mp} = \mathfrak{a}_p^m$ for all m. The study of $\mathfrak{a}_\bullet^L$ is useful precisely when the section ring is not finitely generated (or at least, when this property is not known *a priori*).

Jonsson was partially supported by NSF grants DMS-0449465 and DMS-1001740. Mustaţă was partially supported by NSF grant DMS-0758454 and a Packard Fellowship.
MSC2010: primary 14F18; secondary 14B05.
Keywords: graded sequence of ideals, log canonical threshold.

To a graded sequence $\mathfrak{a}_\bullet$ as above, one can associate an asymptotic version of the log canonical threshold, by putting

$$\mathrm{lct}(\mathfrak{a}_\bullet) := \sup_{m;\, \mathfrak{a}_m \neq (0)} m \cdot \mathrm{lct}(\mathfrak{a}_m).$$

This can be infinite: for example, if $\mathfrak{a}_\bullet = \mathfrak{a}_\bullet^L$ as above, with L big, then $\mathrm{lct}(\mathfrak{a}_\bullet)$ is infinite if and only if L is nef (see Remark 2.2 below).

We will be concerned with the divisors that compute the log canonical thresholds of the elements of a graded sequence. We denote by $A(\mathrm{ord}_E)$ the log discrepancy of a divisor E over X (see Section 2 for the relevant definitions). Our main result, which gives a positive answer to a question of Mihai Păun, is Theorem A below. (Păun's question was motivated by [Siu 2009], in which Y.-T. Siu presents part of his arguments for the finite generation of the canonical ring. At the end of Section 6.3, he evokes a subtle point in his approach, involving the control of an infinite sequence of blow-ups. Although expressed in a different language, our main result shows that the infinite blow-up process in Siu's approach can be "stopped", provided that the log discrepancies of the divisors computing the log canonical thresholds are bounded.)

Theorem A. *Let $\mathfrak{a}_\bullet$ be a graded sequence of ideals on a smooth variety X such that $\mathrm{lct}(\mathfrak{a}_\bullet) < \infty$. If $I \subseteq \mathbb{Z}_{>0}$ is a subset such that for all $m \in I$ we have a divisor E_m over X that computes $\mathrm{lct}(\mathfrak{a}_m)$ such that $\{A(\mathrm{ord}_{E_m}) \mid m \in I\}$ is bounded, then the set $\{E_m \mid m \in I\}$ is finite.*

Corollary B. *Under the hypothesis in Theorem A, suppose that the set I is infinite. Then there is a divisor E over X that computes $\mathrm{lct}(\mathfrak{a}_m)$ for infinitely m. In particular, E computes $\mathrm{lct}(\mathfrak{a}_\bullet)$.*

In fact, since our proof will require replacing X by a suitable blow-up, we will need to prove a stronger version of the above theorem, in which we replace the log canonical threshold by the possibly higher jumping numbers, in the sense of [Ein et al. 2004] (see Theorem 4.1 below for the precise statement).

Here is a sketch of the proof. Let Z_m be the image of E_m on X, and let W be the Zariski closure of $\bigcup_{m \in I} Z_m$. We may assume that W is irreducible, and we first show that since $\mathrm{lct}(\mathfrak{a}_\bullet) < \infty$, the asymptotic order of vanishing $\mathrm{ord}_W(\mathfrak{a}_\bullet)$ is positive. In particular, W is a proper subset of X. If W has codimension at least two in X, then blowing-up X along W decreases the log discrepancies of the divisors E_m, and since these are bounded above, we reduce to the case when W is a hypersurface. In this case, we use the following result, which we believe is of independent interest.

Theorem C. *Let H be a hypersurface in X, and $\mathfrak{a}$ a nonzero ideal. Suppose that E is a divisor over X that computes $\mathrm{lct}(\mathfrak{a})$. If the image Z of E on X is a proper*

subset of H, and if H is smooth at the generic point of Z, then

$$\operatorname{ord}_Z(\mathfrak{a}) \geq \operatorname{ord}_H(\mathfrak{a}) \cdot \left(1 + \frac{\operatorname{ord}_E(I_Z)}{A(\operatorname{ord}_E)}\right),$$

where I_Z is the ideal defining Z.

Of course, as we have already mentioned, we need in fact a version of this result that applies also to higher jumping numbers (see Theorem 3.1 below for this more general version of the theorem). Using Theorem C, we show that if there were infinitely many Z_m that were properly contained in W, then the ideals in $\mathfrak{a}_\bullet$ would vanish along W more than they should. Therefore all but finitely many of the E_m are equal to W (note that at this point we are on some blow-up of our original variety).

In the following section we review some basic facts about log canonical thresholds and higher jumping numbers. The proofs of the stronger versions of Theorems C and A are given in Section 3, and respectively, Section 4.

2. Jumping numbers and valuations

In this section we recall some definitions and results concerning the invariants of singularities that we will use, and set the notation for the rest of the paper. We work over a fixed algebraically closed field k of characteristic zero. Let X be a smooth variety over k (in particular, we assume that X is connected and separated). All ideal sheaves on X are assumed to be coherent.

By a divisor E over X we mean a prime divisor on a normal variety Y that has a proper birational morphism $\pi\colon Y \to X$. This induces a discrete valuation of the function field $K(Y) = K(X)$, that we denote by ord_E. As usual, we identify two such divisors if they induce the same valuation. In particular, it follows from Hironaka's theorem on resolution of singularities that we may assume that both Y and E are nonsingular. If we denote by $K_{Y/X}$ the relative canonical divisor, then the *log discrepancy* of ord_E is given by $A(\operatorname{ord}_E) := 1 + \operatorname{ord}_E(K_{Y/X})$. Note that this depends on the variety X, and whenever the variety is not clear from the context, we will write $A_X(\operatorname{ord}_E)$. The *center* of E on X is the image $c_X(E) := \pi(E)$ of E. We always consider on $c_X(E)$ the reduced scheme structure. If $\mathfrak{a}$ is a nonzero ideal sheaf on X, we put

$$\operatorname{ord}_E(\mathfrak{a}) := \min\{\operatorname{ord}_E(f) \mid f \in \mathfrak{a} \cdot \mathcal{O}_{X,c_X(E)}\} \in \mathbb{R}_{\geq 0}.$$

If Z is the subscheme defined by $\mathfrak{a}$, we also denote this by $\operatorname{ord}_E(Z)$.

Given an irreducible closed subset Z of X, we define the order of vanishing along Z as follows. Consider the normalized blow-up of X along Z, and put $\operatorname{ord}_Z := \operatorname{ord}_E$, where E is the unique irreducible component of the exceptional

divisor that dominates Z. It is clear that in this case $c_X(E) = Z$. Note also that $\mathrm{ord}_Z(\mathfrak{a}) = \min_{x \in Z} \mathrm{ord}_x(\mathfrak{a})$.

Let us recall the definition of multiplier ideals. For details and proofs we refer to [Lazarsfeld 2004, Section 9]. Suppose that $\mathfrak{a}$ is a nonzero ideal on X. Let $\mu\colon X' \to X$ be a log resolution of $(X, \mathfrak{a})$, that is, π is proper and birational, X' is nonsingular, $\mathfrak{a} \cdot \mathcal{O}_{X'} = \mathcal{O}_{X'}(-F)$ for an effective divisor F, and $F + K_{X'/X}$ has simple normal crossings. For every $\lambda \in \mathbb{R}_{\geq 0}$, the multiplier ideal of $\mathfrak{a}$ of exponent λ is given by

$$\mathcal{J}(\mathfrak{a}^\lambda) := \pi_* \mathcal{O}_{X'}(K_{X'/X} - \lfloor \lambda F \rfloor).$$

The definition is independent of the choice of log resolution.

It is clear from the above definition that if $\lambda < \lambda'$, then $\mathcal{J}(\mathfrak{a}^{\lambda'}) \subseteq \mathcal{J}(\mathfrak{a}^\lambda)$. Furthermore, for every λ there is $\varepsilon > 0$ such that $\mathcal{J}(\mathfrak{a}^\lambda) = \mathcal{J}(\mathfrak{a}^t)$ for every $t \in [\lambda, \lambda + \varepsilon]$. One says that $\lambda > 0$ is a *jumping number* of $\mathfrak{a}$ if $\mathcal{J}(\mathfrak{a}^\lambda) \neq \mathcal{J}(\mathfrak{a}^{\lambda'})$ for every $\lambda' < \lambda$. It follows from the definition that if we write $F = \sum_i a_i E_i$, then for every jumping number λ there is i such that λa_i is an integer. In particular, the jumping numbers form a discrete set of rational numbers.

For basic properties of the jumping numbers and applications, we refer to [Ein et al. 2004]. The most important jumping number is the smallest one, known as the *log canonical threshold* and denoted by $\mathrm{lct}(\mathfrak{a})$. This is the smallest λ such that $\mathcal{J}(\mathfrak{a}^\lambda) \neq \mathcal{O}_X$ (note that $\mathcal{J}(\mathfrak{a}^0) = \mathcal{O}_X$).

It is convenient to index the jumping numbers as follows (see [Jonsson and Mustaţă 2010]). Let $\mathfrak{q}$ be a nonzero ideal on X. We put

$$\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}) := \min\{\lambda \mid \mathfrak{q} \not\subseteq \mathcal{J}(\mathfrak{a}^\lambda)\}.$$

Note that $\mathrm{lct}^{\mathcal{O}_X}(\mathfrak{a})$ is the log canonical threshold $\mathrm{lct}(\mathfrak{a})$ of $\mathfrak{a}$. It follows from the definition that if $\mathfrak{a} \neq \mathcal{O}_X$, then $\bigcap_{\lambda \geq 0} \mathcal{J}(\mathfrak{a}^\lambda) = (0)$, hence $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a})$ is finite. When $\mathfrak{a} = \mathcal{O}_X$, we make the convention $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}) = \infty$. We will also use the notation $\mathrm{Arn}^{\mathfrak{q}}(\mathfrak{a}) := 1/\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a})$ (where Arn stands for *Arnold multiplicity*). It follows from the definition that we have

$$\mathrm{Arn}^{\mathfrak{q}}(\mathfrak{a}) = \max_E \frac{\mathrm{ord}_E(\mathfrak{a})}{A(\mathrm{ord}_E) + \mathrm{ord}_E(\mathfrak{q})}, \tag{1}$$

where the maximum can be taken either over all divisors over X, or just over those lying on a log resolution of $(X, \mathfrak{a})$. We say that E computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a})$ (or $\mathrm{Arn}^{\mathfrak{q}}(\mathfrak{a})$) if the maximum in (1) is achieved by E.

The most interesting of the jumping numbers is the log canonical threshold. However, as the following lemma shows, the other jumping numbers appear naturally when we consider higher birational models.

Proposition 2.1. *Let $\pi : X' \to X$ be a proper birational morphism, with X' smooth, and $\mathfrak{a}$ and $\mathfrak{q}$ nonzero ideals on X. If $\mathfrak{a}' = \mathfrak{a} \cdot \mathcal{O}_{X'}$, and $\mathfrak{q}' = \mathfrak{q} \cdot \mathcal{O}_{X'}(-K_{X'/X})$, then*

$$\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}) = \mathrm{lct}^{\mathfrak{q}'}(\mathfrak{a}').$$

Proof. This is an immediate consequence of (1), and of the fact that for every divisor E over X, we have $A_X(\mathrm{ord}_E) = A_{X'}(\mathrm{ord}_E) + \mathrm{ord}_E(K_{X'/X})$. □

Suppose now that $\mathfrak{a}_\bullet$ is a graded sequence of ideals on X, and let $S = \{m \mid \mathfrak{a}_m \neq (0)\}$. Note that S is closed under addition. In this case we have the following asymptotic version of the jumping numbers:

$$\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet) := \sup_{m \in S} m \cdot \mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_m) = \lim_{m \to \infty, m \in S} m \cdot \mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_m) \tag{2}$$

(see [Jonsson and Mustaţă 2010, Section 2]). We put $\mathrm{Arn}^{\mathfrak{q}}(\mathfrak{a}_\bullet) = 1/\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet)$. When $\mathfrak{q} = \mathcal{O}_X$, we simply write $\mathrm{lct}(\mathfrak{a}_\bullet)$ and $\mathrm{Arn}(\mathfrak{a}_\bullet)$. Note that $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet) \in \mathbb{R}_{>0} \cup \{\infty\}$. One can show that $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet) = \infty$ if and only if $\mathrm{lct}(\mathfrak{a}_\bullet) = \infty$ (see [Jonsson and Mustaţă 2010, Corollary 6.10]).

Remark 2.2. If X is a smooth projective variety, L is a big line bundle on X, and $\mathfrak{a}_\bullet = \mathfrak{a}_\bullet^L$ is the graded sequence of ideals defining the base loci of the powers of L (see Introduction), then [Ein et al. 2006, Corollary 2.10] shows that $\mathrm{lct}(\mathfrak{a}_\bullet) = \infty$ if and only if L is nef.

If $\mathfrak{a}_\bullet$ is as above and E is a divisor over X, we will also consider the following asymptotic version of the order of vanishing along E:

$$\mathrm{ord}_E(\mathfrak{a}_\bullet) := \inf_m \frac{\mathrm{ord}_E(\mathfrak{a}_m)}{m} = \lim_{m \to \infty, m \in S} \frac{\mathrm{ord}_E(\mathfrak{a}_m)}{m}.$$

We have the following extension of (1)

$$\mathrm{Arn}^{\mathfrak{q}}(\mathfrak{a}_\bullet) = \sup_E \frac{\mathrm{ord}_E(\mathfrak{a}_\bullet)}{A(\mathrm{ord}_E) + \mathrm{ord}_E(\mathfrak{q})}. \tag{3}$$

For these facts, we refer to [Jonsson and Mustaţă 2010, Section 2]. We say that E *computes* $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet)$ if the supremum in (3) is achieved by E. Note however that unlike in the case of one ideal, there may be no divisor E that computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet)$; see [Jonsson and Mustaţă 2010, Example 8.5].

We will use the following Izumi-type estimate [Izumi 1985; Ein et al. 2003].

Proposition 2.3. *If E is a divisor over X with $c_X(E) = Z$, then*

$$\mathrm{ord}_E(\mathfrak{a}) \leq A(\mathrm{ord}_E) \cdot \mathrm{ord}_Z(\mathfrak{a})$$

for every nonzero ideal sheaf $\mathfrak{a}$ on X.

Proof. We may replace X by an affine open subset of the generic point of Z, and therefore assume that X is affine. In this case we may assume that $\mathfrak{a}$ is principal. If $\mathrm{ord}_Z(\mathfrak{a}) = m$, then for a general $p \in Z$ we have $\mathrm{ord}_p(\mathfrak{a}) = m$. By [Kollár 1997, Lemma 8.10], there is an open neighborhood U of p such that $\mathrm{lct}(\mathfrak{a}|_U) \geq 1/m$, and we get the assertion in the proposition since $U \cap Z \neq \varnothing$ implies

$$\frac{A(\mathrm{ord}_E)}{\mathrm{ord}_E(\mathfrak{a})} \geq \mathrm{lct}(\mathfrak{a}|_U). \qquad \square$$

3. An inequality between orders of vanishing

We keep the notation and the conventions from Section 2. The following is the main result in this section. Note that in the special case $\mathfrak{q} = \mathcal{O}_X$, this recovers Theorem C in the Introduction.

Theorem 3.1. *Let H be a hypersurface in X, and $\mathfrak{a}$, $\mathfrak{q}$ nonzero ideals on X. Suppose that E is a divisor over X that computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a})$. If the center Z of E on X is a proper subset of H, and if H is smooth at the generic point of Z, then the following inequality holds*

$$\mathrm{ord}_Z(\mathfrak{a}) \geq \mathrm{ord}_H(\mathfrak{a}) \cdot \left(1 + \frac{\mathrm{ord}_E(Z)}{A(\mathrm{ord}_E)(1 + \mathrm{ord}_H(\mathfrak{q}))}\right). \tag{4}$$

We start by recalling a basic estimate for the log discrepancy of a valuation. For a proof, see for example [Lazarsfeld 2004, page 157].

Lemma 3.2. *Let E be a divisor over X with $c_X(E) = Z$, and let ξ be the generic point of Z. If $x_1, \ldots, x_r$ form a regular system of parameters of $\mathcal{O}_{X,\xi}$, then*

$$A(\mathrm{ord}_E) \geq \sum_{i=1}^{r} \mathrm{ord}_E(x_i).$$

Corollary 3.3. *If H is a hypersurface in X, and E is a divisor over X such that $Z := c_X(E)$ is a proper subset of H, and H is smooth at the generic point of Z, then*

$$A(\mathrm{ord}_E) \geq \mathrm{ord}_E(H) + \mathrm{ord}_E(Z). \tag{5}$$

Proof. Let ξ be the generic point of Z. Since H is smooth at ξ, we may choose a regular system of parameters $x_1, \ldots, x_r$ of $\mathcal{O}_{X,\xi}$ such that H is defined at ξ by (x_1). Note that by assumption $r \geq 2$. By definition, we have $\mathrm{ord}_E(Z) = \min_j \mathrm{ord}_E(x_j)$. Let i be such that $\mathrm{ord}_E(x_i) = \mathrm{ord}_E(Z)$. If $i \geq 2$, then by the lemma

$$A(\mathrm{ord}_E) \geq \mathrm{ord}_E(x_1) + \mathrm{ord}_E(x_i) = \mathrm{ord}_E(H) + \mathrm{ord}_E(Z).$$

On the other hand, if $i = 1$, then using again the lemma we get

$$A(\mathrm{ord}_E) \geq \mathrm{ord}_E(x_1) + \mathrm{ord}_E(x_2) \geq 2 \cdot \mathrm{ord}_E(x_1) = \mathrm{ord}_E(H) + \mathrm{ord}_E(Z). \qquad \square$$

Proof of Theorem 3.1. Let us put $m = \mathrm{ord}_H(\mathfrak{a})$ and $p = \mathrm{ord}_H(\mathfrak{q})$. We can write $\mathfrak{a} = \mathcal{O}_X(-H)^m \cdot \tilde{\mathfrak{a}}$, and we get

$$\mathrm{ord}_E(\mathfrak{a}) = m \cdot \mathrm{ord}_E(H) + \mathrm{ord}_E(\tilde{\mathfrak{a}}), \quad \mathrm{ord}_Z(\mathfrak{a}) = m + \mathrm{ord}_Z(\tilde{\mathfrak{a}}) \tag{6}$$

(note that $\mathrm{ord}_Z(H) = 1$ since H is smooth at the generic point of Z). Since E computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a})$, it follows from (1) that

$$\frac{\mathrm{ord}_E(\mathfrak{a})}{A(\mathrm{ord}_E) + \mathrm{ord}_E(\mathfrak{q})} \geq \frac{\mathrm{ord}_H(\mathfrak{a})}{A(\mathrm{ord}_H) + \mathrm{ord}_H(\mathfrak{q})} = \frac{m}{1+p}. \tag{7}$$

Corollary 3.3 gives $\mathrm{ord}_E(H) \leq A(\mathrm{ord}_E) - \mathrm{ord}_E(Z)$, and combining this with (7) we deduce

$$\begin{aligned} m &\leq (1+p) \cdot \frac{\mathrm{ord}_E(\mathfrak{a})}{A(\mathrm{ord}_E) + \mathrm{ord}_E(\mathfrak{q})} \leq \frac{\mathrm{ord}_E(\tilde{\mathfrak{a}}) + m(A(\mathrm{ord}_E) - \mathrm{ord}_E(Z)) + p \cdot \mathrm{ord}_E(\mathfrak{a})}{A(\mathrm{ord}_E) + \mathrm{ord}_E(\mathfrak{q})} \\ &= m + \frac{\mathrm{ord}_E(\tilde{\mathfrak{a}}) + p \cdot \mathrm{ord}_E(\mathfrak{a}) - m(\mathrm{ord}_E(\mathfrak{q}) + \mathrm{ord}_E(Z))}{A(\mathrm{ord}_E) + \mathrm{ord}_E(\mathfrak{q})}. \end{aligned} \tag{8}$$

Therefore $\mathrm{ord}_E(\tilde{\mathfrak{a}}) \geq m(\mathrm{ord}_E(\mathfrak{q}) + \mathrm{ord}_E(Z)) - p \cdot \mathrm{ord}_E(\mathfrak{a})$. Using one more time the first equation in (6), this implies

$$(1+p) \cdot \mathrm{ord}_E(\tilde{\mathfrak{a}}) \geq m(\mathrm{ord}_E(\mathfrak{q}) + \mathrm{ord}_E(Z)) - pm \cdot \mathrm{ord}_E(H). \tag{9}$$

On the other hand, by Proposition 2.3 we have $\mathrm{ord}_E(\tilde{\mathfrak{a}}) \leq A(\mathrm{ord}_E) \cdot \mathrm{ord}_Z(\tilde{\mathfrak{a}})$, while clearly $\mathrm{ord}_E(\mathfrak{q}) \geq p \cdot \mathrm{ord}_E(H)$. Putting these together with (9) gives

$$\begin{aligned} (1+p)A(\mathrm{ord}_E) \cdot \mathrm{ord}_Z(\tilde{\mathfrak{a}}) &\geq (1+p) \cdot \mathrm{ord}_E(\tilde{\mathfrak{a}}) \\ &\geq m(\mathrm{ord}_E(\mathfrak{q}) + \mathrm{ord}_E(Z)) - pm \cdot \mathrm{ord}_E(H) \\ &\geq m \cdot \mathrm{ord}_E(Z). \end{aligned}$$

Combining this with the second equality in (6), we obtain

$$\mathrm{ord}_Z(\mathfrak{a}) = m + \mathrm{ord}_Z(\tilde{\mathfrak{a}}) \geq m \cdot \left(1 + \frac{\mathrm{ord}_E(Z)}{A(\mathrm{ord}_E)(1+p)}\right),$$

which completes the proof of the theorem. □

Remark 3.4. In Theorem 3.1 one can replace ord_E by any real valuation of $K(X)$, having center on X and computing $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a})$. The proof goes through if one uses the definition of $A(v)$ from [Jonsson and Mustaţă 2010, Section 5]. In this case, the assertion in Lemma 3.2 follows from Corollary 5.4 of that reference.

Example 3.5. The inequality in Theorem 3.1 is optimal, at least in an asymptotic sense. Indeed, let us consider the ideal $\mathfrak{a} = x^m(x, y^{m+1})$ in $k[x, y]$, where m is a positive integer. Since this is a monomial ideal, one can use Howald's theorem [Howald 2001] to compute its log canonical threshold. It is easy to check that

$\mathrm{lct}(\mathfrak{a}) = (m+2)/(m+1)^2$, and this log canonical threshold is computed by the (toric) divisor E over $X = \mathbb{A}^2$ such that

$$\mathrm{ord}_E\Big(\sum_{i,j\geq 0} c_{i,j}x^i y^j\Big) = \min\{(m+1)i + j \mid c_{i,j} \neq 0\}.$$

Note that $A(\mathrm{ord}_E) = m+2$, and the center of E on X is the origin. If we take $\mathfrak{q} = \mathcal{O}_X$ and $H = (x = 0)$, then

$$\frac{\mathrm{ord}_Z(\mathfrak{a})}{\mathrm{ord}_H(\mathfrak{a})\Big(1 + \frac{\mathrm{ord}_E(Z)}{A(\mathrm{ord}_E)}\Big)} = \frac{m+1}{m\left(1+\frac{1}{m+2}\right)} = \frac{(m+1)(m+2)}{m(m+3)},$$

and this converges to 1 when m goes to infinity.

Remark 3.6. The right-hand side of the inequality (4) is bounded above by

$$\mathrm{ord}_H(\mathfrak{a}) \cdot \left(1 + \frac{1}{\mathrm{lct}(I_Z) \cdot (1 + \mathrm{ord}_H(\mathfrak{q}))}\right),$$

where I_Z is the ideal defining Z. One could ask whether this expression is $\leq \mathrm{ord}_Z(\mathfrak{a})$, improving in this way the assertion in Theorem 3.1. However, this is not the case: let us consider the special case $m = 3$ in Example 3.5, that is, $\mathfrak{a} = x^3(x, y^4)$. With $\mathfrak{q} = \mathcal{O}_X$ and $H = (x = 0)$, we have $\mathrm{ord}_Z(\mathfrak{a}) = 4$, while

$$\mathrm{ord}_H(\mathfrak{a}) \cdot \left(1 + \frac{1}{\mathrm{lct}(I_Z)}\right) = 3\left(1 + \tfrac{1}{2}\right) = \tfrac{9}{2} > 4.$$

4. The main result

In this section we prove the generalized version of Theorem A in the Introduction. We work in the same setting as in Section 2.

Theorem 4.1. *Let $\mathfrak{a}_\bullet$ be a graded sequence of ideals on X, and $\mathfrak{q}$ a nonzero ideal on X such that $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet) < \infty$. If $I \subseteq \mathbb{Z}_{>0}$ is a subset such that for all $m \in I$ we have a divisor E_m over X that computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_m)$ such that $\{A(\mathrm{ord}_{E_m}) \mid m \in I\}$ is bounded, then the set $\{E_m \mid m \in I\}$ is finite.*

Corollary 4.2. *Under the same hypothesis as in Theorem 4.1, suppose that the set I is infinite. Then there is a divisor E over X that computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_m)$ for infinitely many m. In particular, E computes $\mathrm{lct}^{\mathfrak{q}}(\mathfrak{a}_\bullet)$.*

Proof of Theorem 4.1. Note that the hypothesis implies, in particular, that $\mathfrak{a}_m$ is nonzero for every $m \in I$. We assume that I is an infinite set, that $E_i \neq E_j$ for all $i \neq j$ in I and aim to derive a contradiction. Let $Z_m = c_X(E_m)$. We argue by induction on $M := \max\{A(\mathrm{ord}_{E_i}) \mid i \in I\}$. This is finite by assumption. Note that M is a positive integer, and $M = 1$ if and only if all the E_i's are divisors on X. At

several stages in the proof we will replace I by an infinite subset. Note that this can only decrease the value of M.

We start with the following lemma.

Lemma 4.3. *With the above notation, suppose that there is an infinite subset $J \subseteq I$ such that $W := \overline{\bigcup_{j\in J} Z_j}$ is irreducible, and $Z_j \neq W$ for all $j \in J$. In this case*

$$\operatorname{ord}_W(\mathfrak{a}_\bullet) \geq \operatorname{Arn}(\mathfrak{a}_\bullet) \geq \operatorname{Arn}^{\mathfrak{q}}(\mathfrak{a}_\bullet) > 0.$$

Proof. We only need to prove the first inequality. Let $C = \operatorname{Arn}(\mathfrak{a}_\bullet)$, so that $\operatorname{Arn}(\mathfrak{a}_m) \geq Cm$ for every m. If $j \in J$, then by Proposition 2.3 we have $\operatorname{Arn}(\mathfrak{a}_j) \leq \operatorname{ord}_{Z_j}(\mathfrak{a}_j)$.

We need to show that $\operatorname{ord}_W(\mathfrak{a}_m) \geq Cm$ for every $m \geq 1$. We may, of course, assume that $\mathfrak{a}_m$ is nonzero. By hypothesis, we can find $0 \leq \ell \leq m-1$ such that the set

$$\bigcup_{\substack{j\in J\\ j\equiv \ell \,(\mathrm{mod}\, m)}} Z_j \tag{10}$$

is dense in W. Since all Z_j are proper subsets of W, this implies that if in (10) we only take the union over those $j \in J$ with $j \equiv \ell \pmod m$ and with $j \geq N$, for some N, then the union is still dense in W. Let us fix $j_0 \in J$ with $j_0 \equiv \ell \pmod m$, and let $C' := \max_{x\in W} \operatorname{ord}_x(\mathfrak{a}_{j_0}) < \infty$ (recall that $\mathfrak{a}_{j_0}$ is nonzero). If $mp + j_0 \in J$, then the inclusion $\mathfrak{a}_m^p \cdot \mathfrak{a}_{j_0} \subseteq \mathfrak{a}_{mp+j_0}$ implies

$$p\cdot\operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_m)+\operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_{j_0}) \geq \operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_{mp+j_0}) \geq \operatorname{Arn}(\mathfrak{a}_{mp+j_0}) \geq C(mp+j_0).$$

Therefore $\operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_m) \geq Cm - C'/p$. Since we have arbitrarily large such p, and since the union of the corresponding Z_{mp+j_0} is dense in W, we conclude that $\operatorname{ord}_W(\mathfrak{a}_m) \geq Cm$, as required. □

A first consequence of the lemma is that if W is the closure of $\bigcup_{i\in I} Z_i$, then $W \neq X$. In particular, this shows that when $M = 1$, we have a contradiction.

Arguing by Noetherian induction on W, we may assume that W is minimal in X with the property that there is an infinite family of divisors $(E_i)_{i\in I}$ as above, with $\max\{A(\operatorname{ord}_{E_i}) \mid i \in I\} \leq M$. This implies first that W is irreducible. Indeed, if we consider the irreducible decomposition $W = W_1 \cup \cdots \cup W_r$, then there is j such that $Z_i \subseteq W_j$ for infinitely many $i \in I$. Since we may replace I by $\{i \in I \mid Z_i \subseteq W_j\}$, it follows from the minimality assumption on W that $W = W_j$.

A second consequence of the minimality of W is that for every infinite subset $J \subseteq I$, the union $\bigcup_{j\in J} Z_j$ is dense in W. In particular, if U is an open subset of X that meets W, then there are infinitely many $i \in I$ such that U meets Z_i (and the union of these $Z_i \cap U$ is dense in $W \cap U$). Therefore in order to deduce a

contradiction we may replace X by U and each $\mathfrak{a}_m$ by its restriction to U. We may thus assume that W is nonsingular.

We claim that the induction hypothesis on M implies that W is a hypersurface in X. Indeed, suppose that $c = \operatorname{codim}(W, X) \geq 2$, and let $\pi : X' \to X$ be the blow-up of X along W. If E is the exceptional divisor of π, then $K_{X'/X} = (c-1)E$. Since $c_X(E_i) \subseteq W$ for every $i \in I$, it follows that $c_{X'}(E_i) \subseteq E$, hence

$$A_{X'}(\operatorname{ord}_{E_i}) = A_X(\operatorname{ord}_{E_i}) - \operatorname{ord}_{E_i}(K_{X'/X}) \leq A_X(\operatorname{ord}_{E_i}) - (c-1).$$

If $\mathfrak{a}'_m = \mathfrak{a}_m \cdot \mathcal{O}_{X'}$ and $\mathfrak{q}' = \mathfrak{q} \cdot \mathcal{O}_{X'}(-K_{X'/X})$, then by Proposition 2.1 we have $\operatorname{lct}^{\mathfrak{q}}(\mathfrak{a}_i) = \operatorname{lct}^{\mathfrak{q}'}(\mathfrak{a}_i)$, and it follows from hypothesis and (1) that E_i computes $\operatorname{lct}^{\mathfrak{q}'}(\mathfrak{a}'_i)$ for every $i \in I$. Since $\max\{A_{X'}(\operatorname{ord}_{E_i}) \mid i \in I\} \leq M-1$, we have a contradiction by induction on M.

Therefore W is a smooth hypersurface in X. If $Z_i = W$, then $E_i = W$, hence this can be the case for at most one i. After discarding this i, we may assume that each Z_i is a proper subset of W. In particular, we may apply Theorem 3.1 to get

$$\operatorname{ord}_{Z_i}(\mathfrak{a}_i) \geq \operatorname{ord}_W(\mathfrak{a}_i) \cdot \left(1 + \frac{\operatorname{ord}_{E_i}(Z_i)}{A(\operatorname{ord}_{E_i})(1 + \operatorname{ord}_W(\mathfrak{q}))}\right). \tag{11}$$

Note that $\operatorname{ord}_{E_i}(Z_i) \geq 1$ for all $i \in I$. Let $\alpha = \operatorname{ord}_W(\mathfrak{a}_\bullet)$. We have $\alpha > 0$ by Lemma 4.3. Let us fix $\varepsilon > 0$ with

$$\varepsilon < \frac{1}{M(1 + \operatorname{ord}_W(\mathfrak{q}))}.$$

If we show that $\operatorname{ord}_W(\mathfrak{a}_m) \geq \alpha m(1+\varepsilon)$ for every $m \geq 1$, then

$$\alpha = \operatorname{ord}_W(\mathfrak{a}_\bullet) \geq \alpha(1+\varepsilon),$$

a contradiction. We now argue as in the proof of Lemma 4.3. Let $0 \leq \ell \leq m-1$ be such that the set in (10) is dense in W. We fix $j_0 \in I$ such that $j_0 \equiv \ell \pmod{m}$, and let $C' := \max_{x \in W} \operatorname{ord}_x(\mathfrak{a}_{j_0})$. It follows from the inclusion $\mathfrak{a}_m^p \cdot \mathfrak{a}_{j_0} \subseteq \mathfrak{a}_{mp+j_0}$ and from (11) that for every p such that $mp + j_0 \in I$ we have

$$p \cdot \operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_m) \geq \operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_{mp+j_0}) - \operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_{j_0}) \geq \operatorname{ord}_W(\mathfrak{a}_{mp+j_0})(1+\varepsilon) - C'.$$

Therefore for every such p we have $\operatorname{ord}_{Z_{mp+j_0}}(\mathfrak{a}_m) \geq \alpha m(1+\varepsilon) - C'/p$. Since there are arbitrarily large such p, and the union of the corresponding Z_{mp+j_0} is dense in W, we conclude that $\operatorname{ord}_W(\mathfrak{a}_m) \geq \alpha m(1+\varepsilon)$. As we have seen, this leads to a contradiction, and thus completes the proof of the theorem. □

Acknowledgement

We are grateful to Mihai Păun for the question that led to our main result.

References

[Ein and Mustaţă 2006] L. Ein and M. Mustaţă, "Invariants of singularities of pairs", pp. 583–602 in *International congress of mathematicians*, vol. 2, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2006. MR 2007m:14050 Zbl 1096.14030

[Ein et al. 2003] L. Ein, R. Lazarsfeld, and K. E. Smith, "Uniform approximation of Abhyankar valuation ideals in smooth function fields", *Amer. J. Math.* **125**:2 (2003), 409–440. MR 2003m:13004 Zbl 1033.14030

[Ein et al. 2004] L. Ein, R. Lazarsfeld, K. E. Smith, and D. Varolin, "Jumping coefficients of multiplier ideals", *Duke Math. J.* **123**:3 (2004), 469–506. MR 2005k:14004 Zbl 1061.14003

[Ein et al. 2006] L. Ein, R. Lazarsfeld, M. Mustaţă, M. Nakamaye, and M. Popa, "Asymptotic invariants of base loci", *Ann. Inst. Fourier* (*Grenoble*) **56**:6 (2006), 1701–1734. MR 2007m:14008 Zbl 1127.14010

[Howald 2001] J. A. Howald, "Multiplier ideals of monomial ideals", *Trans. Amer. Math. Soc.* **353**:7 (2001), 2665–2671. MR 2002b:14061 Zbl 0979.13026

[Izumi 1985] S. Izumi, "A measure of integrity for local analytic algebras", *Publ. Res. Inst. Math. Sci.* **21**:4 (1985), 719–735. MR 87i:32014 Zbl 0587.32016

[Jonsson and Mustaţă 2010] M. Jonsson and M. Mustaţă, "Valuations and asymptotic invariants for sequences of ideals", preprint, 2010. To appear in *Ann. Inst. Fourier*. arXiv 1011.3699

[Kollár 1997] J. Kollár, "Singularities of pairs", pp. 221–287 in *Algebraic geometry* (Santa Cruz, 1995), vol. 1, edited by J. Kollár et al., Proc. Sympos. Pure Math. **62**, Amer. Math. Soc., Providence, RI, 1997. MR 99m:14033 Zbl 0905.14002

[Lazarsfeld 2004] R. Lazarsfeld, *Positivity in algebraic geometry, II: positivity for vector bundles, and multiplier ideals*, Ergeb. Math. Grenzgeb. (3) **49**, Springer, Berlin, 2004. MR 2005k:14001b Zbl 1093.14500

[Siu 2009] Y.-T. Siu, "Techniques for the analytic proof of the finite generation of the canonical ring", pp. 177–219 in *Current developments in mathematics, 2007*, edited by D. Jerison et al., Int. Press, Somerville, MA, 2009. MR 2011c:32030 Zbl 1183.14012

Communicated by János Kollár
Received 2010-11-27 Revised 2011-05-23 Accepted 2011-06-30

mattiasj@umich.edu *Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043, United States*

mmustata@umich.edu *Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1043, United States*

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

On unit root formulas for toric exponential sums

Alan Adolphson and Steven Sperber

Starting from a classical generating series for Bessel functions due to Schlömilch, we use Dwork's relative dual theory to broadly generalize unit-root results of Dwork on Kloosterman sums and Sperber on hyperkloosterman sums. In particular, we express the (unique) p-adic unit root of an arbitrary exponential sum on the torus $\mathbb{T}^n$ in terms of special values of the p-adic analytic continuation of a ratio of A-hypergeometric functions. In contrast with the earlier works, we use noncohomological methods and obtain results that are valid for arbitrary exponential sums without any hypothesis of nondegeneracy.

1. Introduction

The starting point for this work is the classical generating series

$$\exp \tfrac{1}{2}(\Lambda X - \Lambda/X) = \sum_{i\in\mathbb{Z}} J_i(\Lambda)X^i$$

for the Bessel functions $\{J_i(\Lambda)\}_{i\in\mathbb{Z}}$ due to Schlömilch [1857], which was the foundation for his treatment of Bessel functions (see [Watson 1944, page 14]). Suitably normalized, it also played a fundamental role in Dwork's construction [1974] of p-adic cohomology for $J_0(\Lambda)$. Our realization that the series itself (suitably normalized) could be viewed as a distinguished element in Dwork's relative dual complex led us to the present generalization (which also generalizes unit-root results of Sperber [1975] on hyperkloosterman sums).

Let $A \subseteq \mathbb{Z}^n$ be a finite subset that spans $\mathbb{R}^n$ as real vector space and set

$$f_\Lambda(X) = \sum_{a\in A} \Lambda_a X^a \in \mathbb{Z}[\{\Lambda_a\}_{a\in A}][X_1^{\pm 1}, \dots, X_n^{\pm 1}],$$

where the Λ_a and the X_i are indeterminates and where $X^a = X_1^{a_1}\cdots X_n^{a_n}$ for $a = (a_1, \dots, a_n)$. Let $\mathbb{F}_q$ be the finite field of $q = p^\epsilon$ elements, p a prime, and let

MSC2010: 11T23.
Keywords: exponential sums, A-hypergeometric functions.

$\bar{\mathbb{F}}_q$ be its algebraic closure. For each $\bar{\lambda} = (\bar{\lambda}_a)_{a\in A} \in (\bar{\mathbb{F}}_q)^{|A|}$, let

$$f_{\bar{\lambda}}(X) = \sum_{a\in A} \bar{\lambda}_a X^a \in \mathbb{F}_q(\bar{\lambda})[X_1^{\pm 1}, \dots, X_n^{\pm 1}],$$

a regular function on the n-torus $\mathbb{T}^n$ over $\mathbb{F}_q(\bar{\lambda})$. Fix a nontrivial additive character $\Theta : \mathbb{F}_q \to \mathbb{Q}_p(\zeta_p)$ and let $\Theta_{\bar{\lambda}}$ be the additive character $\Theta_{\bar{\lambda}} = \Theta \circ \mathrm{Tr}_{\mathbb{F}_q(\bar{\lambda})/\mathbb{F}_q}$ of the field $\mathbb{F}_q(\bar{\lambda})$. For each positive integer l, let $\mathbb{F}_q(\bar{\lambda}, l)$ denote the extension of degree l of $\mathbb{F}_q(\bar{\lambda})$ and define an exponential sum

$$S_l = S_l(f_{\bar{\lambda}}, \Theta_{\bar{\lambda}}, \mathbb{T}^n) = \sum_{x\in\mathbb{T}^n(\mathbb{F}_q(\bar{\lambda},l))} \Theta_{\bar{\lambda}} \circ \mathrm{Tr}_{\mathbb{F}_q(\bar{\lambda},l)/\mathbb{F}_q(\bar{\lambda})}(f_{\bar{\lambda}}(x)).$$

The associated L-function is

$$L(f_{\bar{\lambda}}; T) = L(f_{\bar{\lambda}}, \Theta_{\bar{\lambda}}, \mathbb{T}^n; T) = \exp\biggl(\sum_{l=1}^{\infty} S_l \frac{T^l}{l}\biggr).$$

It is well-known that $L(f_{\bar{\lambda}}; T) \in \mathbb{Q}(\zeta_p)(T)$ and that its reciprocal zeros and poles are algebraic integers. We note that among these reciprocal zeros and poles there must be at least one p-adic unit: if $\mathbb{F}_q(\bar{\lambda})$ has cardinality q^κ, then S_l is the sum of $(q^{\kappa l} - 1)^n$ p-th roots of unity, so S_l itself is a p-adic unit for every l. On the other hand, a simple consequence of the Dwork trace formula will imply (see Section 3) that there is at most a single unit root, and it must occur amongst the reciprocal zeros (as opposed to the reciprocal poles) of $L(f_{\bar{\lambda}}; T)^{(-1)^{n+1}}$. We denote this unit root by $u(\bar{\lambda})$. It is the goal of this work to exhibit an explicit p-adic analytic formula for $u(\bar{\lambda})$ in terms of certain A-hypergeometric functions.

Consider the series

$$\exp f_\Lambda(X) = \prod_{a\in A} \exp(\Lambda_a X^a) = \sum_{i\in\mathbb{Z}^n} F_i(\Lambda) X^i \tag{1.1}$$

where the $F_i(\Lambda)$ lie in $\mathbb{Q}[\![\Lambda]\!]$. Explicitly, one has

$$F_i(\Lambda) = \sum_{\substack{u=(u_a)_{a\in A} \\ \sum_{a\in A} u_a a = i}} \frac{\Lambda^u}{\prod_{a\in A}(u_a!)}. \tag{1.2}$$

The A-hypergeometric system with parameter $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{C}^n$ (where $\mathbb{C}$ denotes the complex numbers) is the system of partial differential equations consisting of the operators

$$\Box_\ell = \prod_{\ell_a > 0} \biggl(\frac{\partial}{\partial \Lambda_a}\biggr)^{\ell_a} - \prod_{\ell_a < 0} \biggl(\frac{\partial}{\partial \Lambda_a}\biggr)^{-\ell_a}$$

for all $\ell = (\ell_a)_{a\in A} \in \mathbb{Z}^{|A|}$ satisfying $\sum_{a\in A} \ell_a a = 0$ and the operators

$$Z_j = \sum_{a\in A} a_j \Lambda_a \frac{\partial}{\partial \Lambda_a} - \alpha_j$$

for $a = (a_1, \ldots, a_n) \in A$ and $j = 1, \ldots, n$. Using Equations (1.1) and (1.2), it is straightforward to check that for $i \in \mathbb{Z}^n$, $F_i(\Lambda)$ satisfies the A-hypergeometric system with parameter i.

Fix π satisfying $\pi^{p-1} = -p$ and $\Theta(1) \equiv \pi \pmod{\pi^2}$. It follows from (1.2) that the $F_i(\pi\Lambda)$ converge p-adically for all Λ satisfying $|\Lambda_a| < 1$ for all $a \in A$. Let

$$\mathscr{F}(\Lambda) = F_0(\pi\Lambda)/F_0(\pi\Lambda^p).$$

The main result of this paper is the following statement. Note that we make no restriction (such as nondegeneracy) on the choice of $\bar{\lambda} \in (\bar{\mathbb{F}}_q)^{|A|}$.

Theorem 1.3. *The series $\mathscr{F}(\Lambda)$ converges p-adically for $|\Lambda_a| \le 1$ for all $a \in A$ and the unit root of $L(f_{\bar{\lambda}}; T)$ is given by*

$$u(\bar{\lambda}) = \mathscr{F}(\lambda)\mathscr{F}(\lambda^p)\mathscr{F}(\lambda^{p^2}) \cdots \mathscr{F}(\lambda^{p^{\epsilon d(\bar{\lambda})-1}}),$$

where λ denotes the Teichmüller lifting of $\bar{\lambda}$ and $d(\bar{\lambda}) = [\mathbb{F}_q(\bar{\lambda}) : \mathbb{F}_q]$.

Remark. Historically, expressing the unit root of a zeta- or L-function in terms of special values of p-adic hypergeometric functions has been accomplished by studying the action of Frobenius on the associated p-adic cohomology. Hypergeometric functions arise because the variation of p-adic cohomology of a parametrized family of varieties or of exponential sums is described by (p-adic) hypergeometric differential equations. A systematic listing of the correspondence between such parametrized families and classical hypergeometric equations is given in the appendix to [Dwork and Loeser 1993].

The first result of this type was Dwork's formula [1969] for the unit root of a nonsupersingular elliptic curve $y^2 = x(x-1)(x-\bar{\lambda})$ in terms of the Gaussian hypergeometric function $F(\frac{1}{2}, \frac{1}{2}, 1; \lambda)$. Later he established the corresponding result for the unit root of the family of Kloosterman sums $x + \bar{\lambda}/x$ using the p-adic Bessel function [Dwork 1974]. Since then, a number of authors have proved similar results.

We have systematically avoided the use of cohomology in this article. The cohomology spaces associated to the exponential sums $S_l(f_{\bar{\lambda}}, \Theta_{\bar{\lambda}}, \mathbb{T}^n)$ may not be well behaved for all $\bar{\lambda}$. In any case, it would require substantially more work to describe the action of Frobenius on cohomology (although, of course, this would give information about more than just the unit root).

2. Analytic continuation

We begin by proving the analytic continuation of the function $\mathscr{F}$ defined in the introduction.

Let $C \subseteq \mathbb{R}^n$ be the real cone generated by the elements of A and let $\Delta \subseteq \mathbb{R}^n$ be the convex hull of the set $A \cup \{(0, \dots, 0)\}$. Put $M = C \cap \mathbb{Z}^n$. For $\nu \in M$, define the *weight* of ν, $w(\nu)$, to be the least nonnegative real (hence rational) number such that $\nu \in w(\nu)\Delta$. There exists $D \in \mathbb{Z}_{>0}$ such that $w(\nu) \in \mathbb{Q}_{\geq 0} \cap \mathbb{Z}[1/D]$. The weight function w is easily seen to have the following properties:

(i) $w(\nu) \geq 0$, and $w(\nu) = 0$ if and only if $\nu = 0$.

(ii) $w(c\nu) = cw(\nu)$ for $c \in \mathbb{Z}_{\geq 0}$.

(iii) $w(\nu + \mu) \leq w(\nu) + w(\mu)$, with equality holding if and only if ν and μ are cofacial, that is, lie in a cone over the same closed face of Δ.

(iv) If $\dim \Delta = n$, let $\{\ell_i\}_{i=1}^N$ be linear forms such that the codimension-one faces of Δ not containing the origin lie in the hyperplanes $\{\ell_i = 1\}_{i=1}^N$. Then

$$w(\nu) = \max\{\ell_i(\nu)\}_{i=1}^N.$$

Let Ω be a finite extension of $\mathbb{Q}_p$ containing π and an element $\tilde{\pi}$ satisfying $\operatorname{ord} \tilde{\pi} = (p-1)/p^2$ (we always normalize the valuation so that $\operatorname{ord} p = 1$). Put

$$R = \Big\{ \xi(\Lambda) = \sum_{\nu \in (\mathbb{Z}_{\geq 0})^{|A|}} c_\nu \Lambda^\nu \;\Big|\; c_\nu \in \Omega \text{ and } \{|c_\nu|\}_\nu \text{ is bounded} \Big\},$$

$$R' = \Big\{ \xi(\Lambda) = \sum_{\nu \in (\mathbb{Z}_{\geq 0})^{|A|}} c_\nu \Lambda^\nu \;\Big|\; c_\nu \in \Omega \text{ and } c_\nu \to 0 \text{ as } \nu \to \infty \Big\}.$$

Equivalently, R is the ring of formal power series in $\{\Lambda_a\}_{a \in A}$ that converge on the open unit polydisk in $\Omega^{|A|}$, and R' the ring of those that converge on the closed unit polydisk. Define a norm on R by setting $|\xi(\Lambda)| = \sup_\nu\{|c_\nu|\}$. Both R and R' are complete in this norm. Note that (1.2) implies that the coefficients $F_i(\pi\Lambda)$ of $\exp \pi f_\Lambda(X)$ belong to R.

Let S be the set

$$S = \Big\{ \xi(\Lambda, X) = \sum_{\mu \in M} \xi_\mu(\Lambda) \tilde{\pi}^{-w(\mu)} X^{-\mu} \;\Big|\; \xi_\mu(\Lambda) \in R \text{ and } \{|\xi_\mu(\Lambda)|\}_\mu \text{ is bounded} \Big\}.$$

Let S' be defined analogously with R replaced by R'. Define a norm on S by setting

$$|\xi(\Lambda, X)| = \sup_\mu\{|\xi_\mu(\Lambda)|\}.$$

Both S and S' are complete under this norm.

Define $\theta(t) = \exp(\pi(t - t^p)) = \sum_{i=0}^{\infty} b_i t^i$. By [Dwork 1962, Section 4a],

$$\operatorname{ord} b_i \geq \frac{i(p-1)}{p^2}. \tag{2.1}$$

Let

$$F(\Lambda, X) = \prod_{a \in A} \theta(\Lambda_a X^a) = \sum_{\mu \in M} B_\mu(\Lambda) X^\mu.$$

Lemma 2.2. *One has $B_\mu(\Lambda) \in R'$ and $|B_\mu(\Lambda)| \leq |\tilde{\pi}|^{w(\mu)}$.*

Proof. From the definition,

$$B_\mu(\Lambda) = \sum_{\nu \in (\mathbb{Z}_{\geq 0})^{|A|}} B_\nu^{(\mu)} \Lambda^\nu,$$

where

$$B_\nu^{(\mu)} = \begin{cases} \prod_{a \in A} b_{\nu_a} & \text{if } \sum_{a \in A} \nu_a a = \mu, \\ 0 & \text{if } \sum_{a \in A} \nu_a a \neq \mu. \end{cases}$$

It follows from (2.1) that $B_\nu^{(\mu)} \to 0$ as $\nu \to \infty$, which shows that $B_\mu(\Lambda) \in R'$. We have

$$\operatorname{ord} B_\nu^{(\mu)} \geq \sum_{a \in A} \operatorname{ord} b_{\nu_a} \geq \sum_{a \in A} \frac{\nu_a (p-1)}{p^2} \geq w(\mu) \frac{p-1}{p^2},$$

which implies $|B_\mu(\Lambda)| \leq |\tilde{\pi}|^{w(\mu)}$. □

By the proof of Lemma 2.2, we may write $B_\nu^{(\mu)} = \tilde{\pi}^{w(\mu)} \tilde{B}_\nu^{(\mu)}$ with $|\tilde{B}_\nu^{(\mu)}| \leq 1$. We may then write $B_\mu(\Lambda) = \tilde{\pi}^{w(\mu)} \tilde{B}_\mu(\Lambda)$ with $\tilde{B}_\mu(\Lambda) = \sum_\nu \tilde{B}_\nu^{(\mu)} \Lambda^\nu$ and $|\tilde{B}_\mu(\Lambda)| \leq 1$. Let

$$\xi(\Lambda, X) = \sum_{\nu \in M} \xi_\nu(\Lambda) \tilde{\pi}^{-w(\nu)} X^{-\nu} \in S.$$

We claim that the product $F(\Lambda, X)\xi(\Lambda^p, X^p)$ is well-defined. Formally we have

$$F(\Lambda, X)\xi(\Lambda^p, X^p) = \sum_{\rho \in \mathbb{Z}^n} \zeta_\rho(\Lambda) X^{-\rho},$$

where

$$\zeta_\rho(\Lambda) = \sum_{\substack{\mu, \nu \in M \\ \mu - p\nu = -\rho}} \tilde{\pi}^{w(\mu) - w(\nu)} \tilde{B}_\mu(\Lambda) \xi_\nu(\Lambda^p). \tag{2.3}$$

To prove convergence of this series, we need to show that $w(\mu) - w(\nu) \to \infty$ as $\nu \to \infty$. By property (iv) of the weight function, for a given $\nu \in M$ we may choose

a linear form ℓ (depending on ν) for which $w(\nu) = \ell(\nu)$ while $w(\mu) \geq \ell(\mu)$. Since $\mu = p\nu - \rho$, we get

$$w(\mu) - w(\nu) \geq \ell(\mu - \nu) = \ell((p-1)\nu) - \ell(\rho) = (p-1)w(\nu) - \ell(\rho). \quad (2.4)$$

As $\nu \to \infty$, $(p-1)w(\nu) \to \infty$ while $\ell(\rho)$ takes values in a finite set of rational numbers (there are only finitely many possibilities for ℓ). This gives the desired result.

For a formal series $\sum_{\rho \in \mathbb{Z}^n} \zeta_\rho(\Lambda) X^{-\rho}$ with $\zeta_\rho(\Lambda) \in \Omega[\![\Lambda]\!]$, define

$$\gamma'\Big(\sum_{\rho \in \mathbb{Z}^n} \zeta_\rho(\Lambda) X^{-\rho}\Big) = \sum_{\rho \in M} \zeta_\rho(\Lambda) X^{-\rho}$$

and define for $\xi(\Lambda, X) \in S$

$$\begin{aligned}\alpha^*(\xi(\Lambda, X)) &= \gamma'(F(\Lambda, X)\xi(\Lambda^p, X^p)) \\ &= \sum_{\rho \in M} \zeta_\rho(\Lambda) X^{-\rho}.\end{aligned}$$

For $\rho \in M$ put $\eta_\rho(\Lambda) = \tilde{\pi}^{w(\rho)} \zeta_\rho(\Lambda)$, so that

$$\alpha^*(\xi(\Lambda, X)) = \sum_{\rho \in M} \eta_\rho(\Lambda) \tilde{\pi}^{-w(\rho)} X^{-\rho} \quad (2.5)$$

with

$$\eta_\rho(\Lambda) = \sum_{\substack{\mu, \nu \in M \\ \mu - p\nu = \rho}} \tilde{\pi}^{w(\rho) + w(\mu) - w(\nu)} \tilde{B}_\mu(\Lambda) \xi_\nu(\Lambda^p). \quad (2.6)$$

Since $w(\rho) \geq \ell(\rho)$ for $\rho \in M$, (2.4) implies that

$$w(\rho) + w(\mu) - w(\nu) \geq (p-1)w(\nu), \quad (2.7)$$

so by (2.6), $|\eta_\rho(\Lambda)| \leq |\xi(\Lambda, X)|$ for all $\rho \in M$. This shows $\alpha^*(\xi(\Lambda, X)) \in S$ and

$$|\alpha^*(\xi(\Lambda, X))| \leq |\xi(\Lambda, X)|.$$

Furthermore, this argument also shows that $\alpha^*(S') \subseteq S'$.

Lemma 2.8. *If* $\xi_0(\Lambda) = 0$, *then* $|\alpha^*(\xi(\Lambda, X))| \leq |\tilde{\pi}|^{(p-1)/D} |\xi(\Lambda, X)|$.

Proof. This follows immediately from (2.6) and (2.7) since $w(\nu) \geq 1/D$ for $\nu \neq 0$. □

From (2.6), we have

$$\eta_0(\Lambda) = \sum_{\nu \in M} \tilde{B}_{p\nu}(\Lambda) \xi_\nu(\Lambda^p) \tilde{\pi}^{(p-1)w(\nu)}. \quad (2.9)$$

Note that $\tilde{B}_0(\Lambda) = B_0(\Lambda) \equiv 1 \pmod{\tilde{\pi}}$ since $\operatorname{ord} b_i > 0$ for all $i > 0$ implies $\operatorname{ord} B_\nu^{(0)} > 0$ for all $\nu \neq 0$. Thus $B_0(\Lambda)$ is an invertible element of R'. The following lemma is then immediate from (2.9).

Lemma 2.10. *If $\xi_0(\Lambda)$ is an invertible element of R (resp. R'), then so is $\eta_0(\Lambda)$.*

Put

$$T = \{\xi(\Lambda, X) \in S \mid |\xi(\Lambda, X)| \leq 1 \text{ and } \xi_0(\Lambda) = 1\}$$

and put $T' = T \cap S'$. Using the notation of (2.5), define $\beta : T \to T$ by

$$\beta(\xi(\Lambda, X)) = \frac{\alpha^*(\xi(\Lambda, X))}{\eta_0(\Lambda)}.$$

Note that $\beta(T') \subseteq T'$.

Proposition 2.11. *The operator β is a contraction mapping on the complete metric space T. More precisely, if $\xi^{(1)}(\Lambda, X), \xi^{(2)}(\Lambda, X) \in T$, then*

$$|\beta(\xi^{(1)}(\Lambda, X)) - \beta(\xi^{(2)}(\Lambda, X))| \leq |\tilde{\pi}|^{(p-1)/D} |\xi^{(1)}(\Lambda, X) - \xi^{(2)}(\Lambda, X)|.$$

Proof. We have (in the obvious notation)

$$\begin{aligned}
&\beta(\xi^{(1)}(\Lambda, X)) - \beta(\xi^{(2)}(\Lambda, X)) \\
&\qquad = \frac{\alpha^*(\xi^{(1)}(\Lambda, X))}{\eta_0^{(1)}(\Lambda)} - \frac{\alpha^*(\xi^{(2)}(\Lambda, X))}{\eta_0^{(2)}(\Lambda)} \\
&\qquad = \frac{\alpha^*(\xi^{(1)}(\Lambda, X) - \xi^{(2)}(\Lambda, X))}{\eta_0^{(1)}(\Lambda)} - \alpha^*(\xi^{(2)}(\Lambda, X)) \frac{\eta_0^{(1)}(\Lambda) - \eta_0^{(2)}(\Lambda)}{\eta_0^{(1)}(\Lambda)\eta_0^{(2)}(\Lambda)}.
\end{aligned}$$

Since $\eta_0^{(1)}(\Lambda) - \eta_0^{(2)}(\Lambda)$ is the coefficient of X^0 in $\alpha^*(\xi^{(1)}(\Lambda, X) - \xi^{(2)}(\Lambda, X))$, we have

$$|\eta_0^{(1)}(\Lambda) - \eta_0^{(2)}(\Lambda)| \leq |\alpha^*(\xi^{(1)}(\Lambda, X) - \xi^{(2)}(\Lambda, X))|.$$

And since the coefficient of X^0 in $\xi^{(1)}(\Lambda, X) - \xi^{(2)}(\Lambda, X)$ equals 0, the proposition follows from Lemma 2.8. □

Remark. Proposition 2.11 implies that β has a unique fixed point in T. And since β is stable on T', that fixed point must lie in T'. Let $\xi(\Lambda, X) \in T'$ be the unique fixed point of β. The equation $\beta(\xi(\Lambda, X)) = \xi(\Lambda, X)$ is equivalent to the equation

$$\alpha^*(\xi(\Lambda, X)) = \eta_0(\Lambda)\xi(\Lambda, X).$$

Since α^* is stable on S', it follows that

$$\eta_0(\Lambda)\xi_\mu(\Lambda) \in R' \quad \text{for all } \mu \in M. \tag{2.12}$$

In particular, since $\xi_0(\Lambda) = 1$, we have $\eta_0(\Lambda) \in R'$.

Put $C_0 = C \cap (-C)$, the largest subspace of $\mathbb{R}^n$ contained in C, and put $M_0 = \mathbb{Z}^n \cap C_0$, a subgroup of M. For a formal series $\sum_{\mu\in\mathbb{Z}^n} c_\mu(\Lambda)X^\mu$ with $c_\mu(\Lambda) \in \Omega[\![\Lambda]\!]$ we define

$$\gamma\Big(\sum_{\mu\in\mathbb{Z}^n} c_\mu(\Lambda)X^\mu\Big) = \sum_{\mu\in M_0} c_\mu(\Lambda)X^\mu$$

and set

$$\zeta(\Lambda, X) = \gamma(\exp(\pi f_\Lambda(X))).$$

Of course, when the origin is an interior point of Δ, then $M_0 = \mathbb{Z}^n$ and $\zeta(\Lambda, X) = \exp(\pi f_\Lambda(X))$. In any case, the coefficients of $\zeta(\Lambda, X)$ belong to R.

Since $\exp(\pi f_\Lambda(X)) = \prod_{a\in A} \exp(\pi\Lambda_a X^a)$, we can expand this product to get

$$\zeta(\Lambda, X) = \gamma\Big(\prod_{a\in A}\sum_{\nu_a=0}^{\infty} \frac{(\pi\Lambda_a X^a)^{\nu_a}}{\nu_a!}\Big) = \sum_{\mu\in M_0} G_\mu(\Lambda)\tilde{\pi}^{-w(\mu)}X^{-\mu},$$

where

$$G_\mu(\Lambda) = \sum_{\nu\in(\mathbb{Z}_{\geq 0})^{|A|}} G_\nu^{(\mu)}\Lambda^\nu,$$

with

$$G_\nu^{(\mu)} = \begin{cases} \tilde{\pi}^{w(\mu)} \prod_{a\in A} \dfrac{\pi^{\nu_a}}{\nu_a!} & \text{if } \sum_{a\in A} \nu_a a = -\mu, \\ 0 & \text{if } \sum_{a\in A} \nu_a a \neq -\mu. \end{cases}$$

Since $\operatorname{ord}\pi^i/i! > 0$ for all $i > 0$, it follows that $G_\mu(\Lambda) \in R$, $|G_\mu(\Lambda)| \leq |\tilde{\pi}|^{w(\mu)}$, and $G_0(\Lambda)$ is invertible in R. This implies that $\zeta(\Lambda, X)/G_0(\Lambda) \in T$. Note also that since $F(\Lambda, X) = \exp(\pi f_\Lambda(X))/\exp(\pi f_{\Lambda^p}(X^p))$, it is straightforward to check that

$$\gamma'(F(\Lambda, X)) = \gamma(F(\Lambda, X)) = \gamma\left(\frac{\exp\pi f_\Lambda(X)}{\exp\pi f_{\Lambda^p}(X^p)}\right) = \frac{\zeta(\Lambda, X)}{\zeta(\Lambda^p, X^p)}.$$

It follows that if $\xi(\Lambda, X)$ is a series satisfying $\gamma(\xi(\Lambda, X)) \in S$, then

$$\begin{aligned}\alpha^*(\gamma(\xi(\Lambda, X))) &= \gamma'(F(\Lambda, X)\gamma(\xi(\Lambda^p, X^p))) = \gamma(F(\Lambda, X))\gamma(\xi(\Lambda^p, X^p)) \\ &= \frac{\zeta(\Lambda, X)\gamma(\xi(\Lambda^p, X^p))}{\zeta(\Lambda^p, X^p)}. \end{aligned} \tag{2.13}$$

Remark. In terms of the A-hypergeometric functions $\{F_i(\Lambda)\}_{i\in M}$ defined in (1.1), we have $\exp(\pi f_\Lambda(X)) = \sum_{i\in M} F_i(\pi\Lambda)X^i$, so for $i \in M_0$ we have the relation

$$F_i(\pi\Lambda) = \tilde{\pi}^{-w(-i)}G_{-i}(\Lambda). \tag{2.14}$$

Proposition 2.15. *The unique fixed point of β is $\zeta(\Lambda, X)/G_0(\Lambda)$.*

Proof. By (2.13), we have

$$\alpha^*\left(\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}\right) = \frac{G_0(\Lambda)}{G_0(\Lambda^p)}\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}, \tag{2.16}$$

which is equivalent to the assertion of the proposition. □

By the remark following Proposition 2.11, $\zeta(\Lambda, X)/G_0(\Lambda) \in T'$. This gives the following result.

Corollary 2.17. *For all* $\mu \in M_0$, $G_\mu(\Lambda)/G_0(\Lambda) \in R'$.

In the notation of the remark following Proposition 2.11, one has $\xi(\Lambda, X) = \zeta(\Lambda, X)/G_0(\Lambda)$ and $\eta_0(\Lambda) = G_0(\Lambda)/G_0(\Lambda^p)$, so (2.12) implies the following result.

Corollary 2.18. *For all* $\mu \in M_0$, $G_\mu(\Lambda)/G_0(\Lambda^p) \in R'$.

In view of (2.14), this implies that the function $\mathcal{F}(\Lambda) = F_0(\pi\Lambda)/F_0(\pi\Lambda^p)$ converges on the closed unit polydisk, which was the first assertion of Theorem 1.3.

3. *p*-adic theory

Fix $\bar{\lambda} = (\bar{\lambda}_a)_{a\in A} \in (\bar{\mathbb{F}}_q)^{|A|}$ and let $\lambda = (\lambda_a)_{a\in A} \in (\bar{\mathbb{Q}}_p)^{|A|}$, where λ_a is the Teichmüller lifting of $\bar{\lambda}_a$. We recall Dwork's description of $L(f_{\bar{\lambda}}; T)$. Let $\Omega_0 = \mathbb{Q}_p(\lambda, \zeta_p, \tilde{\pi})$ $(= \mathbb{Q}_p(\lambda, \pi, \tilde{\pi}))$ and let $\mathbb{O}_0$ be the ring of integers of Ω_0.

We consider certain spaces of functions with support in M. We will assume that Ω_0 has been extended by a finite totally ramified extension so that there is an element $\tilde{\pi}_0$ in Ω_0 satisfying $\tilde{\pi}_0^D = \tilde{\pi}$. We shall write $\tilde{\pi}^{w(\nu)}$ and mean by it $\tilde{\pi}_0^{Dw(\nu)}$ for $\nu \in M$. Using this convention to simplify notation, we define

$$B = \Big\{ \sum_{\nu\in M} A_\nu \tilde{\pi}^{w(\nu)} X^\nu \;\Big|\; A_\nu \in \Omega_0,\ A_\nu \to 0 \text{ as } \nu \to \infty \Big\}. \tag{3.1}$$

Then B is an Ω_0-algebra which is complete under the norm

$$\Big| \sum_{\nu\in M} A_\nu \tilde{\pi}^{w(\nu)} X^\nu \Big| = \sup_{\nu\in M} |A_\nu|.$$

We construct a Frobenius map with arithmetic import in the usual way. Let

$$F(\lambda, X) = \prod_{a\in A} \theta(\lambda_a X^a) = \sum_{\mu\in M} B_\mu(\lambda) X^\mu,$$

i.e., $F(\lambda, X)$ is the specialization of $F(\Lambda, X)$ at $\Lambda = \lambda$, which is permissible by

Lemma 2.2. Note also that Lemma 2.2 implies

$$\operatorname{ord} B_\mu(\lambda) \geq \frac{w(\mu)(p-1)}{p^2},$$

so we may write $B_\mu(\lambda) = \tilde{\pi}^{w(\mu)} \tilde{B}_\mu(\lambda)$ with $\tilde{B}_\mu(\lambda)$ p-integral.

Let

$$\Psi(X^\mu) = \begin{cases} X^{\mu/p} & \text{if } p \mid \mu_i \text{ for all } i, \\ 0 & \text{otherwise.} \end{cases}$$

We show that $\Psi \circ F(\lambda, X)$ acts on B. If $\xi = \sum_{\nu \in M} A_\nu \tilde{\pi}^{w(\nu)} X^\nu \in B$, then

$$\Psi\biggl(\Bigl(\sum_{\mu \in M} \tilde{\pi}^{w(\mu)} \tilde{B}_\mu(\lambda) X^\mu\Bigr)\Bigl(\sum_{\nu \in M} A_\nu \tilde{\pi}^{w(\nu)} X^\nu\Bigr)\biggr) = \sum_{\omega \in M} C_\omega(\lambda) \tilde{\pi}^{w(\omega)} X^\omega$$

where

$$C_\omega(\lambda) = \sum_\nu \tilde{\pi}^{w(p\omega-\nu)+w(\nu)-w(\omega)} \tilde{B}_{p\omega-\nu}(\lambda) A_\nu.$$

For any positive constant K there are only finitely many values of ν for which $w(\nu)$ and $w(p\omega - \nu)$ are $< K$. This implies that the series $C_\omega(\lambda)$ converges.

We have $pw(\omega) = w(p\omega) \leq w(p\omega - \nu) + w(\nu)$, so that

$$\operatorname{ord} C_\omega(\lambda) \geq \inf_\nu \{\operatorname{ord} \tilde{\pi}^{(p-1)w(\omega)} A_\nu\} = \frac{(p-1)^2 w(\omega)}{p^2} + \inf_\nu \{\operatorname{ord} A_\nu\}. \tag{3.2}$$

This implies that $\Psi(F(\lambda, X)\xi) \in B$.

Let $d(\bar{\lambda}) = [\mathbb{F}_q(\bar{\lambda}) : \mathbb{F}_q]$, so that $\lambda^{p^{\epsilon d(\bar{\lambda})}} = \lambda$. Put

$$\alpha_\lambda = \Psi^{\epsilon d(\bar{\lambda})} \circ \biggl(\prod_{i=0}^{\epsilon d(\bar{\lambda})-1} F(\lambda^{p^i}, X^{p^i})\biggr).$$

For any power series $P(T)$ in the variable T with constant term 1, define

$$P(T)^{\delta_{\bar{\lambda}}} = P(T)/P(p^{\epsilon d(\bar{\lambda})} T).$$

Then α_λ is a completely continuous operator on B and the Dwork trace formula [Dwork 1962; Serre 1962] gives

$$L(f_{\bar{\lambda}}, \Theta_{\bar{\lambda}}, \mathbb{T}^n; T)^{(-1)^{n+1}} = \det(I - T\alpha_\lambda \mid B)^{\delta^n_{\bar{\lambda}}}. \tag{3.3}$$

By Equation (3.2), the (ω, ν)-entry of the matrix of α_λ [Serre 1962, Section 2] has $\operatorname{ord} > 0$ unless $\omega = \nu = 0$. The formula for $\det(I - T\alpha_\lambda)$ [Serre 1962, Proposition 7a)] then shows that this Fredholm determinant can have at most a single unit root. Since $L(f_{\bar{\lambda}}; T)$ has at least one unit root (Section 1), it follows from (3.3) that $L(f_{\bar{\lambda}}; T)$ has exactly one unit root.

4. Dual theory

It will be important to consider the trace formula in the dual theory as well. The basis for this construction goes back to [Dwork 1964] and [Serre 1962]. We define

$$B^* = \left\{ \xi^* = \sum_{\mu \in M} A_\mu^* \tilde{\pi}^{-w(\mu)} X^{-\mu} \;\middle|\; \{A_\mu^*\}_{\mu \in M} \text{ is a bounded subset of } \Omega_0 \right\},$$

a p-adic Banach space with the norm $|\xi^*| = \sup_{\mu \in M}\{|A_\mu^*|\}$. We define a pairing $\langle\ ,\ \rangle : B^* \times B \to \Omega_0$: if $\xi = \sum_{\mu \in M} A_\mu \tilde{\pi}^{w(\mu)} X^\mu$, $\xi^* = \sum_{\mu \in M} A_\mu^* \tilde{\pi}^{-w(\mu)} X^{-\mu}$, set

$$\langle \xi^*, \xi \rangle = \sum_{\mu \in M} A_\mu A_\mu^* \in \Omega_0.$$

The series on the right converges since $A_\mu \to 0$ as $\mu \to \infty$ and $\{A_\mu^*\}_{\mu \in M}$ is bounded. This pairing identifies B^* with the dual space of B, i.e., the space of continuous linear mappings from B to Ω_0 [Serre 1962, Proposition 3].

Let Φ be the endomorphism of the space of formal series defined by

$$\Phi\left(\sum_{\mu \in \mathbb{Z}^n} c_\mu X^{-\mu} \right) = \sum_{\mu \in \mathbb{Z}^n} c_\mu X^{-p\mu},$$

and let γ' be the endomorphism

$$\gamma'\left(\sum_{\mu \in \mathbb{Z}^n} c_\mu X^{-\mu} \right) = \sum_{\mu \in M} c_\mu X^{-\mu}.$$

Consider the formal composition $\alpha_\lambda^* = \gamma' \circ \left(\prod_{i=0}^{\epsilon d(\bar{\lambda})-1} F(\lambda^{p^i}, X^{p^i}) \right) \circ \Phi^{\epsilon d(\bar{\lambda})}$.

Proposition 4.1. *The operator α_λ^* is an endomorphism of B^* which is adjoint to $\alpha_\lambda : B \to B$.*

Proof. As α_λ^* is the composition of the operators $\gamma' \circ F(\lambda^{p^i}, X) \circ \Phi$ and α_λ is the composition of the operators $\Psi \circ F(\lambda^{p^i}, X)$, $i = 0, \ldots, \epsilon d(\bar{\lambda}) - 1$, it suffices to check that $\gamma' \circ F(\lambda, X) \circ \Phi$ is an endomorphism of B^* adjoint to $\Psi \circ F(\lambda, X) : B \to B$. Let $\xi^*(X) = \sum_{\mu \in M} A_\mu^* \tilde{\pi}^{-w(\mu)} X^{-\mu} \in B^*$. The proof that the product $F(\lambda, X)\xi^*(X^p)$ is well-defined is analogous to the proof of convergence of the series (2.3). We have

$$\gamma'(F(\lambda, X)\xi^*(X^p)) = \sum_{\omega \in M} C_\omega(\lambda) \tilde{\pi}^{-w(\omega)} X^{-\omega},$$

where

$$C_\omega(\lambda) = \sum_{\mu - p\nu = -\omega} \tilde{B}_\mu(\lambda) A_\nu^* \tilde{\pi}^{w(\omega)+w(\mu)-w(\nu)}. \tag{4.2}$$

Note that

$$pw(\nu) = w(p\nu) \le w(\omega) + w(\mu)$$

since $p\nu = \omega + \mu$. Thus

$$(p-1)w(\nu) \le w(\omega) + w(\mu) - w(\nu),$$

which implies that the series on the right-hand side of (4.2) converges and that $|C_\omega(\lambda)| \le |\xi^*|$ for all $\omega \in M$. It follows that $\gamma'(F(\lambda, X)\xi^*(X^p)) \in B^*$. It is straightforward to check that $\langle \Phi(X^{-\mu}), X^\nu \rangle = \langle X^{-\mu}, \Psi(X^\nu) \rangle$ and that

$$\langle \gamma'(F(\lambda, X)X^{-\mu}), X^\nu \rangle = \langle X^{-\mu}, F(\lambda, X)X^\nu \rangle$$

for all $\mu, \nu \in M$, which implies the maps are adjoint. □

By [Serre 1962, Proposition 15] we have $\det(I - T\alpha_\lambda^* \mid B^*) = \det(I - T\alpha_\lambda \mid B)$, so (3.3) implies

$$L(f_{\bar\lambda}, \Theta_{\bar\lambda}, \mathbb{T}^n; T)^{(-1)^{n+1}} = \det(I - T\alpha_\lambda^* \mid B^*)^{\delta_{\bar\lambda}^n}. \tag{4.3}$$

From Equations (2.14) and (2.16), we have

$$\alpha^*\left(\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}\right) = \mathscr{F}(\Lambda)\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}.$$

It follows by iteration that for $m \ge 0$,

$$(\alpha^*)^m\left(\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}\right) = \left(\prod_{i=0}^{m-1} \mathscr{F}(\Lambda^{p^i})\right)\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}. \tag{4.4}$$

We have

$$\frac{\zeta(\Lambda, X)}{G_0(\Lambda)} = \sum_{\mu \in M_0} \frac{G_\mu(\Lambda)}{G_0(\Lambda)}\tilde\pi^{-w(\mu)}X^{-\mu},$$

so by Corollary 2.17 we may evaluate at $\Lambda = \lambda$ to get an element of B^*:

$$\left.\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}\right|_{\Lambda=\lambda} = \sum_{\mu \in M_0} \left.\frac{G_\mu(\Lambda)}{G_0(\Lambda)}\right|_{\Lambda=\lambda} \tilde\pi^{-w(\mu)}X^{-\mu} \in B^*.$$

It is straightforward to check that the specialization of the left-hand side of (4.4) with $m = \epsilon d(\bar\lambda)$ at $\Lambda = \lambda$ is exactly $\alpha_\lambda^*((\zeta(\Lambda, X)/G_0(\Lambda))|_{\Lambda=\lambda})$, so specializing (4.4) with $m = \epsilon d(\bar\lambda)$ at $\Lambda = \lambda$ gives

$$\alpha_\lambda^*\left(\left.\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}\right|_{\Lambda=\lambda}\right) = \left(\prod_{i=0}^{\epsilon d(\bar\lambda)-1} \mathscr{F}(\lambda^{p^i})\right)\left.\frac{\zeta(\Lambda, X)}{G_0(\Lambda)}\right|_{\Lambda=\lambda}. \tag{4.5}$$

Equation (4.5) shows that $\prod_{i=0}^{\epsilon d(\bar\lambda)-1} \mathscr{F}(\lambda^{p^i})$ is a (unit) eigenvalue of α_λ^*, hence by (4.3) it is the unique unit eigenvalue of $L(f_{\bar\lambda}; T)$.

Acknowledgement

We are grateful to D. Wan, whose interest in unit root questions encouraged us to write up our ideas on this topic.

References

[Dwork 1962] B. Dwork, "On the zeta function of a hypersurface", *Inst. Hautes Études Sci. Publ. Math.* **12** (1962), 5–68. MR 28 #3039 Zbl 0173.48601

[Dwork 1964] B. Dwork, "On the zeta function of a hypersurface, II", *Ann. of Math.,(2)* **80** (1964), 227–299. MR 32 #5654 Zbl 0173.48601

[Dwork 1969] B. Dwork, "p-adic cycles", *Inst. Hautes Études Sci. Publ. Math.* **37** (1969), 27–115. MR 45 #3415 Zbl 0284.14008

[Dwork 1974] B. Dwork, "Bessel functions as p-adic functions of the argument", *Duke Math. J.* **41** (1974), 711–738. MR 52 #8124 Zbl 0302.14008

[Dwork and Loeser 1993] B. Dwork and F. Loeser, "Hypergeometric series", *Japan. J. Math.* (*N.S.*) **19**:1 (1993), 81–129. MR 95f:33013 Zbl 0796.12005

[Schlömilch 1857] O. Schlömilch, "Ueber die Bessel'sche Funktion", *Zeitschrift für Math. und Phys.* **2** (1857), 137–165.

[Serre 1962] J.-P. Serre, "Endomorphismes complètement continus des espaces de Banach p-adiques", *Inst. Hautes Études Sci. Publ. Math.* **12** (1962), 69–85. MR 26 #1733 Zbl 0104.33601

[Sperber 1975] S. I. Sperber, *p-adic hypergeometric functions and their cohomology*, Ph.D. thesis, University of Pennsylvania, 1975. MR 2625626

[Watson 1944] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge University Press, 1944. MR 0010746 Zbl 0063.08184

Communicated by Hendrik W. Lenstra
Received 2010-12-07 Revised 2011-03-28 Accepted 2011-05-08

adolphs@math.okstate.edu *Department of Mathematics, Oklahoma State University, Stillwater, OK 74078, United States*

sperber@math.umn.edu *School of Mathematics, University of Minnesota, Minneapolis, MN 55455, United States*

ALGEBRA AND NUMBER THEORY 6:3(2012)

msp

Symmetries of the transfer operator for $\Gamma_0(N)$ and a character deformation of the Selberg zeta function for $\Gamma_0(4)$

Markus Fraczek and Dieter Mayer

The transfer operator for $\Gamma_0(N)$ and trivial character χ_0 possesses a finite group of symmetries generated by permutation matrices P with $P^2 = \mathrm{id}$. Every such symmetry leads to a factorization of the Selberg zeta function in terms of Fredholm determinants of a reduced transfer operator. These symmetries are related to the group of automorphisms in $\mathrm{GL}(2, \mathbb{Z})$ of the Maass wave forms of $\Gamma_0(N)$. For the group $\Gamma_0(4)$ and Selberg's character χ_α there exists just one nontrivial symmetry operator P. The eigenfunctions of the corresponding reduced transfer operator with eigenvalue $\lambda = \pm 1$ are related to Maass forms that are even or odd, respectively, under a corresponding automorphism. It then follows from a result of Sarnak and Phillips that the zeros of the Selberg function determined by the eigenvalue $\lambda = -1$ of the reduced transfer operator stay on the critical line under deformation of the character. From numerical results we expect that, on the other hand, all the zeros corresponding to the eigenvalue $\lambda = +1$ are off this line for a nontrivial character χ_α.

1. Introduction

In the transfer operator approach to Selberg's zeta function for a Fuchsian group Γ this function gets expressed in terms of the Fredholm determinant of a transfer operator constructed from the symbolic dynamics of the geodesic flow on the corresponding surface of constant negative curvature. Though this approach has been carried out, up to now, only for certain groups, like modular subgroups of finite index [Chang and Mayer 2000; 2001a; 2001b] and Hecke triangle groups [Mayer and Strömberg 2008; Mayer et al. 2012; Mayer and Mühlenbruch 2010], it has led to new points of view on the Selberg zeta function [Zagier 2002] and the theory

This work was supported by the Deutsche Forschungsgemeinschaft through the DFG Research Project "Maass wave forms and the transfer operator approach to the Phillips–Sarnak conjecture" (Ma 633/18-1).

MSC2010: primary 11M36; secondary 35J05, 35B25, 37C30, 11F72, 11F03.

Keywords: transfer operator, Hecke congruence subgroups, Maass wave forms, character deformation, factorization of the Selberg zeta function.

of period functions [Lewis and Zagier 2001]. Another application of this method is a precise numerical calculation of the Selberg zeta function [Strömberg 2008], which seems to be impossible by other means at the moment.

In this paper we discuss the transfer operator approach to the Selberg zeta function for Hecke congruence subgroups with a character. Of special interest is the behavior of its zeros for $\Gamma_0(4)$ under singular deformation by Selberg's character [Selberg 1990].

As found numerically in [Fraczek 2010], certain symmetries of the transfer operator for these groups play an important role in this process. These symmetries lead to a factorization of the Selberg zeta function for the full modular group $\mathrm{SL}(2,\mathbb{Z})$, as known. There it corresponds to the involution $Ju(z) = u(-z^*)$ of the Maass forms u for this group [Efrat 1993; Lewis and Zagier 2001]. Obviously the corresponding element $j = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{GL}(2,\mathbb{Z})$ generates the normalizer group of $\mathrm{SL}(2,\mathbb{Z})$ in $\mathrm{GL}(2,\mathbb{Z})$. It tuns out also that the symmetries of the transfer operator for $\Gamma_0(N)$ correspond to automorphisms of the Maass forms from its normalizer group in $\mathrm{GL}(2,\mathbb{Z})$.

For the group $\Gamma_0(4)$ with a character χ_α introduced in [Selberg 1990] and discussed also in [Phillips and Sarnak 1994], there is only one such nontrivial symmetry of the transfer operator. It corresponds to the generator of $\Gamma_0(4)$'s normalizer group in $\mathrm{GL}(2,\mathbb{Z})$ leaving invariant the character χ_α. The results of Phillips and Sarnak imply that the zeros on the critical line of one factor of Selberg's function stay on this line under the deformation of the character, and hence the corresponding Maass wave forms for the trivial character remain Maass wave forms. Numerical results [Fraczek 2010], on the other hand, imply that the zeros on the critical line of the second factor of this function should all leave this line when the deformation is turned on. A detailed discussion of these numerical results and their partial proofs is in preparation [Bruggeman et al. 2012].

The paper is organized as follows: in Section 2 we recall briefly the form of the transfer operator

$$\mathbf{L}_{\beta,\rho_\pi} = \begin{pmatrix} 0 & \mathscr{L}^+_{\beta,\pi} \\ \mathscr{L}^-_{\beta,\pi} & 0 \end{pmatrix}$$

for a general finite index subgroup Γ of the modular group $\mathrm{SL}(2,\mathbb{Z})$ and unitary representation π, and introduce the symmetries

$$\tilde{P} = \begin{pmatrix} 0 & P \\ P & 0 \end{pmatrix}$$

of this operator defined by permutation matrices P. Any such symmetry leads to a factorization of the Selberg zeta function in terms of the Fredholm determinants of

the reduced transfer operator $P\mathscr{L}^+_{\beta,\pi}$. The eigenfunctions with eigenvalues $\lambda = \pm 1$ of this reduced transfer operator then fulfill certain functional equations.

In Section 3 we discuss the generators $J_{n,-}$ of the group of automorphisms in $\mathrm{GL}(2,\mathbb{Z})$ of the Maass forms u for $\Gamma = \Gamma_0(N)$ and $\pi = \chi_0$ the trivial character. We introduce their period functions $\underline{\psi}$ and derive a formula for the period function $J_{n,-}\underline{\psi}$ of the Maass form $J_{n,-}u$.

In Section 4 we introduce Selberg's character χ_α and the nontrivial automorphism $J_{2,-}$ of the Maass forms for $\Gamma_0(4)$. We derive again a formula for the period function $J_{2,-}\underline{\psi}$ of the Maass form $J_{2,-}u$ leading to a permutation matrix $P_{2,-}$ which defines a symmetry $\tilde{P}_{2,-}$ of the transfer operator $\mathbf{L}_{\beta,\rho_{\chi_\alpha}}$. From this we conclude that the eigenfunctions with eigenvalues $\lambda = \pm 1$ of the operator $P_{2,-}\mathscr{L}^+_{\beta,\pi}$ correspond to Maass forms that are even or odd, respectively, under the involution $J_{2,-}$. Results of Phillips and Sarnak then imply that the zeros of the Selberg function on the critical line corresponding to the eigenfunctions with eigenvalue $\lambda = -1$ of this operator stay on this line under the deformation of the character.

2. The transfer operator and the Selberg zeta function for Hecke congruence subgroups $\Gamma_0(N)$

The starting point of the transfer operator approach to the Selberg zeta function for a subgroup Γ of the modular group $\mathrm{SL}(2,\mathbb{Z})$ of index $\mu = [\mathrm{SL}(2,\mathbb{Z}) : \Gamma] < \infty$ is the geodesic flow $\Phi_t : SM_\Gamma \to SM_\Gamma$ on the unit tangent bundle SM_Γ of the corresponding surface $M_\Gamma = \Gamma \backslash \mathbb{H}$ of constant negative curvature. Here

$$\mathbb{H} = \{z = x + iy : y > 0\}$$

denotes the hyperbolic plane with hyperbolic metric $ds^2 = (dx^2 + dy^2)/y^2$, on which the group Γ acts via Möbius transformations: $gz = (az+b)/(cz+d)$ if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. In the present paper we mostly work with the Hecke congruence subgroup

$$\Gamma_0(N) = \left\{ g \in \mathrm{SL}(2,\mathbb{Z}) : g = \begin{pmatrix} a & b \\ cN & d \end{pmatrix} \right\},$$

with index $\mu_N = N\prod_{p|N}(1+1/p)$, where p is a prime number. If $\rho : \Gamma \to \mathrm{end}(\mathbb{C}^d)$ is a unitary representation of Γ then the Selberg zeta function $Z_{\Gamma,\rho}$ is defined as

$$Z_{\Gamma,\rho}(\beta) = \prod_{\gamma}\prod_{k=0}^{\infty} \det\bigl(1 - \rho(g_\gamma)\exp(-(k+\beta)l_\gamma)\bigr), \tag{2.0.1}$$

where l_γ denotes the period of the prime periodic orbit γ of Φ_t and $g_\gamma \in \Gamma$ is hyperbolic with $g_\gamma(\gamma) = \gamma$. In the dynamical approach to this function it gets expressed in terms of the so-called transfer operator, well-known from D. Ruelle's thermodynamic formalism approach to dynamical systems. For general modular

groups Γ with finite index μ and finite-dimensional representation π this operator $\mathbf{L}_{\beta,\pi} : B \to B$ was determined in [Chang and Mayer 2000; 2001b] as

$$\mathbf{L}_{\beta,\pi} = \begin{pmatrix} 0 & \mathscr{L}^+_{\beta,\rho_\pi} \\ \mathscr{L}^-_{\beta,\rho_\pi} & 0 \end{pmatrix}, \tag{2.0.2}$$

where $B = B(D, \mathbb{C}^\mu) \bigoplus B(D, \mathbb{C}^\mu)$ is the Banach space of holomorphic functions on the disc $D = \{z : |z-1| < \frac{3}{2}\}$, and ρ_π denotes the representation of $\mathrm{SL}(2,\mathbb{Z})$ induced from the representation π of Γ. The operator $\mathscr{L}^\pm_{\beta,\rho_\pi}$ is given for $\operatorname{Re}\beta > \frac{1}{2}$ by

$$(\mathscr{L}^\pm_{\beta,\rho_\pi} \underline{f})(z) = \sum_{n=1}^{\infty} \frac{1}{(z+n)^{2\beta}} \rho_\pi(ST^{\pm n}) \underline{f}\Big(\frac{1}{z+n}\Big), \tag{2.0.3}$$

where $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. In the following we restrict ourselves to one-dimensional unitary representations π, hence unitary characters, which we denote as usual by χ. In this case the following theorem was proved in [Chang and Mayer 2001b].

Theorem 2.0.1. *The transfer operator* $\mathbf{L}_{\beta,\chi} : B \to B$ *with*

$$\mathbf{L}_{\beta,\chi} = \begin{pmatrix} 0 & \mathscr{L}^+_{\beta,\chi} \\ \mathscr{L}^-_{\beta,\chi} & 0 \end{pmatrix} \quad \text{and} \quad (\mathscr{L}^\pm_{\beta,\chi} \underline{f})(z) = \sum_{n=1}^{\infty} \frac{1}{(z+n)^{2\beta}} \rho_\chi(ST^{\pm n}) \underline{f}\Big(\frac{1}{z+n}\Big)$$

extends to a meromorphic family of nuclear operators of order zero in the entire complex β *plane with possible poles at* $\beta_k = (1-k)/2$, $k = 0, 1, 2, \ldots$. *The Selberg zeta function* $Z_{\Gamma,\chi}$ *for modular group* Γ *and character* χ *can be expressed as* $Z_{\Gamma,\chi}(\beta) = \det(1 - \mathbf{L}_{\beta,\chi}) = \det(1 - \mathscr{L}^+_{\beta,\chi}\mathscr{L}^-_{\beta,\chi}) = \det(1 - \mathscr{L}^-_{\beta,\chi}\mathscr{L}^+_{\beta,\chi})$.

This shows that the zeros of the Selberg function are given by those β-values for which $\lambda = 1$ belongs to the spectrum $\sigma(\mathbf{L}_{\beta,\chi})$, or equivalently to the spectrum $\sigma(\mathscr{L}^-_{\beta,\chi}\mathscr{L}^+_{\beta,\chi}) = \sigma(\mathscr{L}^+_{\beta,\chi}\mathscr{L}^-_{\beta,\chi})$. From Selberg's trace formula one knows that there are two kinds of such zeros: the trivial zeros at $\beta = -k$, $k = 1, 2, \ldots$, and the so-called spectral zeros. The former correspond to eigenvalues $\lambda = \beta(1-\beta)$ of the automorphic Laplacian with $\operatorname{Re}\beta = \frac{1}{2}$ or $\frac{1}{2} \le \beta \le 1$, and the latter to resonances of the Laplacian, that is, poles of the scattering determinant with $\operatorname{Re}\beta < \frac{1}{2}$ and $\operatorname{Im}\beta > 0$ [Hejhal 1983; Venkov 1990]. For arithmetic groups like the congruence subgroups with trivial or congruent character χ one knows that these resonances lie on the line $\operatorname{Re}\beta = \frac{1}{4}$, corresponding to the nontrivial zeros $\zeta_R(2\beta) = 0$ of the Riemann zeta function ζ_R in the trivial case and to the zeros $L(2\beta, \chi_\alpha) = 0$ of other Dirichlet L-functions in the congruent case, assuming the generalized Riemann hypothesis, as well as on the line $\operatorname{Re}\beta = 0$. For general Fuchsian groups and congruence subgroups with noncongruent character, however, these resonances can be anywhere in the half-plane $\operatorname{Re}\beta < \frac{1}{2}$.

2.1. ***Symmetries of the transfer operator for*** $\boldsymbol{\Gamma_0(N)}$**.** It turns out that there exists for any N a finite number h_N of $\mu_N \times \mu_N$ permutation matrices P with $P^2 = \mathrm{id}_{\mu_N}$ such that the matrix

$$\tilde{P} = \begin{pmatrix} 0 & P \\ P & 0 \end{pmatrix}$$

commutes with the transfer operator $\mathbf{L}_{\beta,\chi}$ and hence

$$P\mathscr{L}^{+}_{\beta,\chi} = \mathscr{L}^{-}_{\beta,\chi} P. \tag{2.1.1}$$

Thereby $P = (P_{ij})_{1\le i,j\le \mu_N}$ acts in the Banach space $B(D, \mathbb{C}^{\mu_N})$ as $(P\underline{f})_i(z) = \sum_{j=1}^{\mu_N} P_{ij} f_j(z)$ if $\underline{f}(z) = (f_i(z))_{1\le i\le \mu_N}$. We call such a matrix $\tilde{P}$ a symmetry of the transfer operator. As an example consider the group $\Gamma_0(4)$ and Selberg's character χ_α, $0 \le \alpha \le 1$, which will be described later. Its transfer operator $\mathbf{L}_{\beta,\chi_\alpha}$ has the following form:

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{+1} = \sum_{q=0}^{\infty} f_{-3}\big|_{2\beta}\tilde{S}T^{1+4q} + f_{-4}\big|_{2\beta}\tilde{S}T^{2+4q} + f_{-5}\big|_{2\beta}\tilde{S}T^{3+4q} + f_{-2}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{+2} = \sum_{q=0}^{\infty} e^{2\pi i(1+4q)\alpha} f_{-1}\big|_{2\beta}\tilde{S}T^{1+4q} + e^{2\pi i(2+4q)\alpha} f_{-1}\big|_{2\beta}\tilde{S}T^{2+4q} + e^{2\pi i(3+4q)\alpha} f_{-1}\big|_{2\beta}\tilde{S}T^{3+4q} + e^{2\pi i(4+4q)\alpha} f_{-1}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{+3} = \sum_{q=0}^{\infty} e^{-2\pi i\alpha} f_{-2}\big|_{2\beta}\tilde{S}T^{1+4q} + e^{-2\pi i\alpha} f_{-3}\big|_{2\beta}\tilde{S}T^{2+4q} + e^{-2\pi i\alpha} f_{-4}\big|_{2\beta}\tilde{S}T^{3+4q} + e^{-2\pi i\alpha} f_{-5}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{+4} = \sum_{q=0}^{\infty} e^{-2\pi i\alpha(1+4q)} f_{-6}\big|_{2\beta}\tilde{S}T^{1+4q} + e^{-2\pi i\alpha(2+4q)} f_{-6}\big|_{2\beta}\tilde{S}T^{2+4q} + e^{-2\pi i\alpha(3+4q)} f_{-6}\big|_{2\beta}\tilde{S}T^{3+4q} + e^{-2\pi i\alpha(4+4q)} f_{-6}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{+5} = \sum_{q=0}^{\infty} e^{2\pi i\alpha} f_{-4}\big|_{2\beta}\tilde{S}T^{1+4q} + e^{2\pi i\alpha} f_{-5}\big|_{2\beta}\tilde{S}T^{2+4q} + e^{2\pi i\alpha} f_{-2}\big|_{2\beta}\tilde{S}T^{3+4q} + e^{2\pi i\alpha} f_{-3}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{+6} = \sum_{q=0}^{\infty} f_{-5}\big|_{2\beta}\tilde{S}T^{1+4q} + f_{-2}\big|_{2\beta}\tilde{S}T^{2+4q} + f_{-3}\big|_{2\beta}\tilde{S}T^{3+4q} + f_{-4}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{-1} = \sum_{q=0}^{\infty} f_{+5}\big|_{2\beta}\tilde{S}T^{1+4q} + f_{+4}\big|_{2\beta}\tilde{S}T^{2+4q} + f_{+3}\big|_{2\beta}\tilde{S}T^{3+4q} + f_{+2}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{-2}=\sum_{q=0}^{\infty}e^{-2\pi i\alpha(1+4q)}f_{+1}\big|_{2\beta}\tilde{S}T^{1+4q}+e^{-2\pi i\alpha(2+4q)}f_{+1}\big|_{2\beta}\tilde{S}T^{2+4q}$$
$$+e^{-2\pi i\alpha(3+4q)}f_{+1}\big|_{2\beta}\tilde{S}T^{3+4q}+e^{-2\pi i\alpha(4+4q)}f_{+1}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{-3}=\sum_{q=0}^{\infty}e^{-2\pi i\alpha}f_{+4}\big|_{2\beta}\tilde{S}T^{1+4q}+e^{-2\pi i\alpha}f_{+3}\big|_{2\beta}\tilde{S}T^{2+4q}$$
$$+e^{-2\pi i\alpha}f_{+2}\big|_{2\beta}\tilde{S}T^{3+4q}+e^{-2\pi i\alpha}f_{+5}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{-4}=\sum_{q=0}^{\infty}e^{2\pi i\alpha(1+4q)}f_{+6}\big|_{2\beta}\tilde{S}T^{1+4q}+e^{2\pi i\alpha(2+4q)}f_{+6}\big|_{2\beta}\tilde{S}T^{2+4q}$$
$$+e^{2\pi i\alpha(3+4q)}f_{+6}\big|_{2\beta}\tilde{S}T^{3+4q}+e^{2\pi i\alpha(4+4q)}f_{+6}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{-5}=\sum_{q=0}^{\infty}e^{2\pi i\alpha}f_{+2}\big|_{2\beta}\tilde{S}T^{1+4q}+e^{2\pi i\alpha}f_{+5}\big|_{2\beta}\tilde{S}T^{2+4q}$$
$$+e^{2\pi i\alpha}f_{+4}\big|_{2\beta}\tilde{S}T^{3+4q}+e^{2\pi i\alpha}f_{+3}\big|_{2\beta}\tilde{S}T^{4+4q},$$

$$\mathbf{L}_{\beta,\chi_\alpha}\tilde{f}_{-6}=\sum_{q=0}^{\infty}f_{+3}\big|_{2\beta}\tilde{S}T^{1+4q}+f_{+2}\big|_{2\beta}\tilde{S}T^{2+4q}+f_{+5}\big|_{2\beta}\tilde{S}T^{3+4q}+f_{+4}\big|_{2\beta}\tilde{S}T^{4+4q},$$

where $\tilde{f}\in B(D,\mathbb{C}^\mu)\bigoplus B(D,\mathbb{C}^\mu)$ is given by $\tilde{f}=(\underline{f}_+,\underline{f}_-)$, $\underline{f}_\pm=(f_{\pm i})_{1\le i\le 6}$, and $\tilde{S}z=1/z$. The induced representation ρ_{χ_α} of the character χ_α on $\Gamma_0(4)$ is defined in terms of the coset decomposition of $\mathrm{SL}(2,\mathbb{Z})$

$$\mathrm{SL}(2,\mathbb{Z})=\bigcup_{i=1}^{6}\Gamma_0(4)R_i \tag{2.1.2}$$

as

$$\rho_\chi(g)_{ij}=\delta_{\Gamma_0(4)}(R_i g R_j^{-1})\chi_\alpha(R_i g R_j^{-1}),\quad 1\le i,\quad j\le 6. \tag{2.1.3}$$

Thereby we have chosen the following representatives $R_i\in\mathrm{SL}(2,\mathbb{Z})$ of the cosets $\Gamma_0(4)R_i$

$$R_1=\mathrm{id}_2,\qquad R_i=ST^{i-2},\quad 2\le i\le 5,\quad \text{and}\quad R_6=ST^2S. \tag{2.1.4}$$

It turns out that the two permutation matrices P_1 and P_2 corresponding to the permutations

$$\sigma_1=\begin{matrix}1&2&3&4&5&6\\1&2&5&4&3&6\end{matrix} \tag{2.1.5}$$

and

$$\sigma_2=\begin{matrix}1&2&3&4&5&6\\6&4&3&2&5&1\end{matrix} \tag{2.1.6}$$

fulfill (2.1.1) for $\alpha = 0$ and hence the corresponding matrices $\tilde{P}_i$, $i = 1, 2$, commute with the transfer operator $\mathbf{L}_{\beta,\chi_0}$ where χ_0 is the trivial character. The matrix $\tilde{P}_2$, on the other hand, commutes even with the operator $\mathbf{L}_{\beta,\chi_\alpha}$ for all α. The matrix $\rho_{\chi_0}(S)$ is given by the permutation σ_S where

$$\sigma_S = \frac{1\ 2\ 3\ 4\ 5\ 6}{2\ 1\ 5\ 6\ 3\ 4}, \tag{2.1.7}$$

and an easy calculation shows that $P_i\rho_{\chi_0}(S) = \rho_{\chi_0}(S)P_i$, $i = 1, 2$. The matrix $\rho_{\chi_0}(T)$, on the other hand, is given by the permutation σ_T with

$$\sigma_T = \frac{1\ 2\ 3\ 4\ 5\ 6}{1\ 3\ 4\ 5\ 2\ 6}. \tag{2.1.8}$$

One then checks that $P_i\rho_{\chi_0}(T) = \rho_{\chi_0}(T^{-1})P_i$, $i = 1, 2$. Therefore $P_i\rho_{\chi_0}(ST^n) = \rho_{\chi_0}(ST^{-n})P_i$ for all $n \in \mathbb{N}$ and $i = 1, 2$. For the character χ_α analogous relations hold for P_2.

For the trivial character χ_0 one can determine for the group $\Gamma_0(N)$ the number h_N of matrices P_i with the above properties and hence the defining symmetries of the transfer operator as follows:

Theorem 2.1.1. *For the Hecke congruence subgroup $\Gamma_0(N)$ and trivial character $\chi_0 \equiv 1$ there exist h_N matrices $\tilde{P} = \left(\begin{smallmatrix} 0 & P \\ P & 0 \end{smallmatrix}\right)$ commuting with the transfer operator $\mathbf{L}_{\beta,\chi_0}$ where P is a $\mu_N \times \mu_N$ permutation matrix satisfying $P^2 = \mathbf{1}_{\mu_N}$,*

$$P\rho_{\chi_0}(S) = \rho_{\chi_0}(S)P \quad \textit{and} \quad P\rho_{\chi_0}(T) = \rho_{\chi_0}(T^{-1})P,$$

and hence

$$P\mathcal{L}^+_{\beta,\chi_0} = \mathcal{L}^-_{\beta,\chi_0}P.$$

Thereby $h_N = \max\{k : k|24 \text{ and } k^2|N\}$. The permutation matrices P are determined by the h_N generators j of the normalizer group $\mathcal{N}_N$ of $\Gamma_0(N)$ in $\mathrm{GL}(2, \mathbb{Z})$. The Selberg zeta function Z_{Γ,χ_0} can be written as

$$Z_{\Gamma,\chi_0} = \det(1 - P\mathcal{L}^+_{\beta,\chi_0}) \det(1 + P\mathcal{L}^+_{\beta,\chi_0}).$$

Remark 2.1.2. For $\Gamma_0(4)$, obviously $h_N = 2$. According to Theorem 2.1.1, there exist two permutation matrices P_1 and P_2 given by the permutations σ_1 and σ_2 above. Since $P_1P_2 = P_2P_1$ and $P_i\mathcal{L}^+_{\beta,\chi_0} = \mathcal{L}^-_{\beta,\chi_0}P_i$, $i = 1, 2$, we find

$$P_1P_2P_1\mathcal{L}^+_{\beta,\chi_0} = P_1P_2\mathcal{L}^-_{\beta,\chi_0}P_1 = P_1\mathcal{L}^+_{\beta,\chi_0}P_2P_1 = P_1\mathcal{L}^+_{\beta,\chi_0}P_1P_2,$$

and the operators P_1P_2 and $P_1\mathcal{L}^+_{\beta,\chi_0}$ commute, where the operator P_1P_2 corresponds to the permutation

$$\sigma = \frac{1\ 2\ 3\ 4\ 5\ 6}{6\ 4\ 5\ 2\ 3\ 1}. \tag{2.1.9}$$

We find also $P_1P_2\mathscr{L}^+_{\beta,\chi_0} = \mathscr{L}^+_{\beta,\chi_0}P_1P_2$. But $(P_1P_2)^2 = \mathrm{id}_6$, hence this operator has only the eigenvalues $\lambda = \pm 1$ and the Banach space $B(D,\mathbb{C}^6)$ decomposes as $B(D,\mathbb{C}^6) = B(D,\mathbb{C}^6)_+ \oplus B(D,\mathbb{C}^6)_-$ with $P_1P_2\underline{f}_\pm = \pm\underline{f}_\pm$ for $\underline{f}_\pm \in B(D,\mathbb{C}^6)_\pm$. Therefore the elements $\underline{f}_\epsilon \in B(D,\mathbb{C}^6)_\epsilon$, $\epsilon = \pm$ have the form $(\underline{f}_\epsilon)_i = f_i$, $1 \le i \le 3$ and $(\underline{f}_\epsilon)_{\sigma(i)} = \epsilon f_i$, $1 \le i \le 3$. Denote by

$$\mathscr{L}^+_{\beta,\chi_0,\pm} : B(D,\mathbb{C}^6)_\pm \to B(D,\mathbb{C}^6)_\pm \quad \text{and} \quad P_1\mathscr{L}^+_{\beta,\chi_0,\pm} : B(D,\mathbb{C}^6)_\pm \to B(D,\mathbb{C}^6)_\pm,$$

the restrictions of the operators $\mathscr{L}^+_{\beta,\chi_0}$ and $P_1\mathscr{L}^+_{\beta,\chi_0}$, respectively, to the subspace $B(D,\mathbb{C}^6)_\pm$, which obviously is isomorphic to the space $B(D,\mathbb{C}^3)$. Then

$$\det(1 \pm P_1\mathscr{L}^+_{\beta,\chi_0}) = \det(1 \pm P_1\mathscr{L}^+_{\beta,\chi_0,+}) \det(1 \pm P_1\mathscr{L}^+_{\beta,\chi_0,-}),$$

where the operator $P_1\mathscr{L}^+_{\beta,\chi_0,\epsilon} : B(D,\mathbb{C}^3) \to B(D,\mathbb{C}^3)$ can be written as

$$P_1\mathscr{L}^+_{\beta,\chi_0,\epsilon} = \begin{pmatrix} 0 & \epsilon\mathscr{L}_{\beta,2}+\mathscr{L}_{\beta,4} & \epsilon\mathscr{L}_{\beta,1}+\mathscr{L}_{\beta,3} \\ \mathscr{L}_\beta & 0 & 0 \\ 0 & \mathscr{L}_{\beta,1}+\epsilon\mathscr{L}_{\beta,3} & \epsilon\mathscr{L}_{\beta,2}+\mathscr{L}_{\beta,4} \end{pmatrix}, \tag{2.1.10}$$

with $\mathscr{L}_{\beta,k}f = \sum_{q=0}^\infty f\big|_{2\beta}\tilde{S}T^{1+kq}$, $1 \le k \le 4$, and $\mathscr{L}_\beta = \sum_{k=1}^4 \mathscr{L}_{\beta,k}$. The operator $\mathscr{L}^+_{\beta,\chi_0,\epsilon}$ in the space $B(D,\mathbb{C}^3)$, on the other hand, has the form

$$\mathscr{L}^+_{\beta,\chi_0,\epsilon} = \begin{pmatrix} 0 & \epsilon\mathscr{L}_{\beta,2}+\mathscr{L}_{\beta,4} & \epsilon\mathscr{L}_{\beta,1}+\mathscr{L}_{\beta,3} \\ \mathscr{L}_\beta & 0 & 0 \\ 0 & \epsilon\mathscr{L}_{\beta,1}+\mathscr{L}_{\beta,3} & \mathscr{L}_{\beta,2}+\epsilon\mathscr{L}_{\beta,4} \end{pmatrix}. \tag{2.1.11}$$

To relate the Fredholm determinants of the operators $(P_1\mathscr{L}^+_{\beta,\chi_0,\epsilon})^2$ and $(\mathscr{L}^+_{\beta,\chi_0,\epsilon})^2$ we use the following simple lemma:

Lemma 2.1.3. *Let α, β, and γ be complex numbers and $\epsilon = \pm 1$. Then λ is an eigenvalue of the matrix*

$$\mathbb{L}_1 = \begin{pmatrix} 0 & \alpha & \beta \\ \gamma & 0 & 0 \\ 0 & \beta & \epsilon\alpha \end{pmatrix}$$

if and only if $\epsilon\lambda$ is an eigenvalue of the matrix

$$\mathbb{L}_2 = \begin{pmatrix} 0 & \alpha & \beta \\ \gamma & 0 & 0 \\ 0 & \epsilon\beta & \alpha \end{pmatrix}.$$

Proof. The proof follows from comparing the characteristic polynomials of the two matrices. □

This shows that, for all $n \in \mathbb{N}$,

$$\operatorname{trace}\mathbb{L}_1^n = \sum_{k=1}^3 (\mathbb{L}_1^n)_{k,k} = \epsilon^n \operatorname{trace}\mathbb{L}_2^n = \epsilon^n \sum_{k=1}^3 (\mathbb{L}_2^n)_{k,k}.$$

But then is not too difficult to see that also $\operatorname{trace}(\mathcal{L}^+_{\beta,\chi_0,\epsilon})^n = \epsilon^n \operatorname{trace}(P_1\mathcal{L}^+_{\beta,\chi_0,\epsilon})^n$ for all $n \in \mathbb{N}$ and hence $\det(1-(P_1\mathcal{L}^+_{\beta,\chi_0,\epsilon})^2) = \det(1-(\mathcal{L}^+_{\beta,\chi_0,\epsilon})^2)$ for $\epsilon = \pm$. Therefore the Selberg zeta function $Z_{\Gamma_0(4),\chi_0}(\beta)$ for the group $\Gamma_0(4)$ with trivial character χ_0 can be written as

$$\begin{aligned} Z_{\Gamma_0(4),\chi_0}(\beta) &= \det(1-(P_1\mathcal{L}^+_{\beta,\chi_0})^2) = \det(1-(\mathcal{L}^+_{\beta,\chi_0})^2) \\ &= \det(1-\mathcal{L}^+_{\beta,\chi_0})\det(1+\mathcal{L}^+_{\beta,\chi_0}). \end{aligned} \tag{2.1.12}$$

Furthermore, this function factorizes in this case also as

$$\begin{aligned} Z_{\Gamma_0(4),\chi_0}(\beta) = \det(1-P_1\mathcal{L}^+_{\beta,\chi_0,+})\det(1-P_1\mathcal{L}^+_{\beta,\chi_0,-}) \\ \times \det(1+P_1\mathcal{L}^+_{\beta,\chi_0,+})\det(1+P_1\mathcal{L}^+_{\beta,\chi_0-}). \end{aligned} \tag{2.1.13}$$

To prove Theorem 2.1.1 we relate the matrices P to the generating automorphisms in $\mathrm{GL}(2,\mathbb{Z})$ of the Maass wave forms for $\Gamma_0(N)$. We can determine this way the explicit form of these matrices P. For this we derive, in a first step, a Lewis-type functional equation for the eigenfunctions of the operator $P\mathcal{L}^+_{\beta,\chi}$ with eigenvalues $\lambda = \pm 1$.

2.2. ***A Lewis-type functional equation.*** Consider any finite index modular subgroup Γ and any unitary character $\chi : \Gamma \to \mathbb{C}^\star$, together with the induced representation ρ_χ of $\mathrm{SL}(2,\mathbb{Z})$. Assume there exists a symmetry $\tilde{P} = \left(\begin{smallmatrix} 0 & P \\ P & 0 \end{smallmatrix}\right)$, with P a permutation matrix with properties analogous to Theorem 2.1.1, and commuting with the transfer operator

$$\mathbf{L}_{\beta,\chi} = \begin{pmatrix} 0 & \mathcal{L}^+_{\beta,\rho_\chi} \\ \mathcal{L}^-_{\beta,\rho_\chi} & 0 \end{pmatrix}$$

of Γ. If $\underline{f}$ is an eigenfunction of the operator $P\mathcal{L}^+_{\beta,\chi}$ with eigenvalues $\lambda = \pm 1$ then one can show:

Proposition 2.2.1. *If $P\mathcal{L}^+_{\beta,\chi}\underline{f}(\zeta) = \lambda\underline{f}(\zeta)$ with $\lambda = \pm 1$ then the function $\underline{\Psi}(\zeta) := P\rho_\chi(T^{-1}S)P\underline{f}(\zeta-1)$ fulfills the functional equations*

$$\underline{\Psi}(\zeta) = \lambda\zeta^{-2\beta}P\rho_\chi(S)\underline{\Psi}\Big(\frac{1}{\zeta}\Big) \tag{2.2.1}$$

and

$$\underline{\Psi}(\zeta) - \rho_\chi(T^{-1})\underline{\Psi}(\zeta+1) - (\zeta+1)^{-2\beta}\rho_\chi(T'^{-1})\underline{\Psi}\Big(\frac{\zeta}{\zeta+1}\Big) = \underline{0}, \tag{2.2.2}$$

where $T' = ST^{-1}S$. On the other hand, every solution $\underline{\Psi}$ of (2.2.1) *and* (2.2.2) *holomorphic in the cut β-plane $(-\infty, 0]$ satisfying $\Psi_i(z) = o(z^{-\min\{1,2\,\mathrm{Re}\,s\}})$ as $z \downarrow 0$ and $\Psi_i(z) = o(z^{-\min\{0,2\,\mathrm{Re}\,s-1\}})$ as $z \to \infty$ determines an eigenfunction $\underline{f}$ with eigenvalues $\lambda = \pm 1$ of the operator $P\mathcal{L}^+_{\beta,\chi}$.*

Proof. Let $\operatorname{Re}\beta > \frac{1}{2}$. If $P\mathscr{L}^+_\beta \underline{f}(\zeta) = \lambda \underline{f}(\zeta)$, $\lambda = \pm 1$, then obviously

$$P\rho_\chi(STS)PP\mathscr{L}^+_\beta \underline{f}(\zeta+1) = \lambda P\rho_\chi(STS)P\underline{f}(\zeta+1).$$

Subtracting the two equations leads to

$$\lambda \underline{f}(\zeta) - \lambda P\rho_\chi(STS)P\underline{f}(\zeta+1) - (\zeta+1)^{-2\beta}P\rho_\chi(ST)\underline{f}\Big(\frac{1}{\zeta+1}\Big) = \underline{0},$$

and hence the function $\underline{\psi}(\zeta) := P\underline{f}(\zeta-1)$ fulfills the equation

$$\underline{\psi}(\zeta) - \rho_\chi(STS)\underline{\psi}(\zeta+1) - \lambda\zeta^{-2\beta}\rho_\chi(ST)P\underline{\psi}\Big(\frac{\zeta+1}{\zeta}\Big) = \underline{0}. \tag{2.2.3}$$

Replacing ζ by $\dfrac{1}{\zeta}$ and multiplying the resulting equation by $\zeta^{-2\beta}\rho_\chi(STS)P\rho_\chi(T^{-1}S)$ gives

$$\zeta^{-2\beta}\rho_\chi(STS)P\rho_\chi(T^{-1}S)\underline{\psi}\Big(\frac{1}{\zeta}\Big) - \zeta^{-2\beta}\rho_\chi(STS)P\rho_\chi(S)\underline{\psi}\Big(\frac{\zeta+1}{\zeta}\Big) - \lambda\rho_\chi(STS)\underline{\psi}(\zeta+1) = \underline{0}.$$

Since $\rho_\chi(S)P = P\rho_\chi(S)$, one finds, comparing with (2.2.3),

$$\underline{\psi}(\zeta) = \lambda\zeta^{-2\beta}\rho_\chi(STS)P\rho_\chi(T^{-1}S)\underline{\psi}\Big(\frac{1}{\zeta}\Big).$$

Hence the function $\underline{\tilde{\psi}} := \rho_\chi(T^{-1}S)\underline{\psi}$ fulfills (2.2.1). The same equation is then fulfilled also by the function

$$\underline{\Psi}(\zeta) := P\underline{\tilde{\psi}}(\zeta) = P\rho_\chi(T^{-1}S)P\underline{f}(\zeta-1), \tag{2.2.4}$$

that is,

$$\underline{\Psi}(\zeta) = \lambda\zeta^{-2\beta}P\rho_\chi(S)\underline{\Psi}\Big(\frac{1}{\zeta}\Big). \tag{2.2.5}$$

Inserting finally $\underline{\psi}(\zeta) = \rho_\chi(ST)P\underline{\Psi}(\zeta)$ into (2.2.3) and using (2.2.1) leads to

$$\underline{\Psi}(\zeta) - P\rho_\chi(T)P\underline{\Psi}(\zeta+1) - (\zeta+1)^{-2\beta}P\rho_\chi(T')P\underline{\Psi}\Big(\frac{\zeta}{\zeta+1}\Big) = \underline{0}.$$

But by assumption $P\rho_\chi(T)P = \rho_\chi(T^{-1})$; hence $P\rho_\chi(T')P = \rho_\chi(T'^{-1})$ and thus

$$\underline{\Psi}(\zeta) - \rho_\chi(T^{-1})\underline{\Psi}(\zeta+1) - (\zeta+1)^{-2\beta}\rho_\chi(T'^{-1})\underline{\Psi}\Big(\frac{\zeta}{\zeta+1}\Big) = \underline{0}. \tag{2.2.6}$$

Hence for $\operatorname{Re}\beta > \frac{1}{2}$ the first part of the proposition holds. By analytic continuation in β one proves the general case.

To prove the second part we follow the arguments of [Deitmar and Hilgert 2007, Lemma 4.1]: if $\underline{\Psi}(\zeta)$ is a solution of the Lewis equation (2.2.2) with $\beta \notin \mathbb{Z}$ then

$\underline{\Psi}$ has the asymptotic expansions

$$\underline{\Psi}(\zeta) \sim_{\zeta\to 0} \zeta^{2\beta} Q_0\Big(\frac{1}{\zeta}\Big) + \sum_{l=-1}^{\infty} \underline{C}_l^* \zeta^l, \quad \underline{\Psi}(\zeta) \sim_{\zeta\to\infty} Q_\infty(\zeta) + \sum_{l=-1}^{\infty} \underline{C}_l^{*\prime} \zeta^{-l-2\beta},$$

where $Q_0, Q_\infty : \mathbb{C} \to \mathbb{C}^\mu$ are smooth functions such that $Q_0(\zeta+1) = \rho_\chi(T')Q_0(\zeta)$ and $Q_\infty(\zeta+1) = \rho_\chi(T)Q_\infty(\zeta)$, and the constants $\underline{C}_l^*$ and $\underline{C}_l^{*\prime}$ are determined by the Taylor coefficients $\underline{C}_m = (1/m!)\underline{\Psi}^{(m)}(1)$. The functions Q_0 and Q_∞ are defined as follows for general β with $-2\operatorname{Re}\beta < M \in \mathbb{N}$:

$$\begin{aligned} Q_0(\zeta) := \zeta^{-2\beta}\underline{\Psi}\Big(\frac{1}{\zeta}\Big) &- \sum_{m=0}^{M} \zeta_{\rho_\chi}(m+2\beta, z)\underline{C}_m \\ &- \sum_{n=0}^{\infty} (n+\zeta)^{-2\beta} \rho_\chi(T'^{-n}T^{-1}) \Big(\underline{\Psi}\Big(1+\frac{1}{n+\zeta}\Big) - \sum_{m=0}^{M} \frac{\underline{C}_m}{(n+\zeta)^m} \Big), \end{aligned}$$

$$\begin{aligned} Q_\infty(\zeta) := \underline{\Psi}(\zeta) &- \sum_{m=0}^{M} \zeta'_{\rho_\chi}(m+2\beta, \zeta+1)\underline{C}_m \\ &- \sum_{n=0}^{\infty} (n+\zeta)^{-2\beta} \rho_\chi(T^{-(n-1)}T'^{-1}) \Big(\underline{\Psi}\Big(1-\frac{1}{n+\zeta}\Big) - \sum_{m=0}^{M} \frac{\underline{C}_m}{(n+\zeta)^m} \Big), \end{aligned}$$

where

$$\zeta_{\rho_\chi}(a,\zeta) = \frac{1}{N^a} \sum_{k=0}^{N-1} \rho_\chi(T'^{-k}T^{-1}) \zeta\Big(a, \frac{k+\zeta}{N}\Big)$$

and

$$\zeta'_{\rho_\chi}(a,\zeta) = \frac{1}{N^a} \sum_{k=0}^{N-1} \rho_\chi(T^{-k}T'^{-1}) \zeta_H\Big(a, \frac{k+\zeta}{N}\Big),$$

with $\zeta_H(a,\zeta)$ the Hurwitz zeta function. According to [Deitmar and Hilgert 2007, Remark 4.2] any solution $\underline{\Psi}$ of (2.2.2) with $\underline{\Psi}(\zeta) = \underline{o}(\zeta^{-\min\{1,2\beta\}})$ for $\zeta \to 0$ fulfills the equation

$$\underline{\Psi}(\zeta) = \zeta^{-2\beta} \sum_{n=0}^{\infty} (n+\zeta^{-1})^{-2\beta} \rho_\chi(T'^{-n}T^{-1}) \underline{\Psi}\Big(1+\frac{1}{n+\zeta^{-1}}\Big)$$

and moreover $\underline{C}_{-1}^* = 0$. But if $\underline{\Psi}(\zeta)$ fulfills also (2.2.1) then one finds

$$\lambda\zeta^{-2\beta} P\rho_\chi(S)\underline{\Psi}\Big(\frac{1}{\zeta}\Big) = \zeta^{-2\beta} \sum_{n=0}^{\infty} (n+\zeta^{-1})^{-2\beta} \rho_\chi(T'^{-n}T^{-1}) \underline{\Psi}\Big(1+\frac{1}{n+\zeta^{-1}}\Big),$$

and hence

$$\lambda P\rho_\chi(S)\underline{\Psi}(\zeta+1)=\sum_{n=1}^{\infty}(n+\zeta)^{-2\beta}\rho_\chi(T'^{-(n-1)}T^{-1})\underline{\Psi}\Big(1+\frac{1}{n+\zeta}\Big). \quad (2.2.7)$$

According to (2.2.4) $\underline{\Psi}(\zeta+1)=P\rho_\chi(T^{-1}S)P\underline{f}(\zeta)$, and hence we get

$$\lambda\rho_\chi(ST^{-1}S)P\underline{f}(\zeta)=\sum_{n=1}^{\infty}(n+\zeta)^{-2\beta}\rho_\chi(T'^{-(n-1)}T^{-1})P\rho_\chi(T^{-1}S)P\underline{f}\Big(\frac{1}{\zeta+n}\Big).$$

Inserting $T'^{-(n-1)}=ST^{(n-1)}S$ one arrives at

$$\lambda\underline{f}(\zeta)=\sum_{n=1}^{\infty}(n+\zeta)^{-2\beta}P\rho_\chi(ST^n)\rho_\chi(ST^{-1})P\rho_\chi(T^{-1}S)P\underline{f}\Big(\frac{1}{\zeta+n}\Big).$$

Since $\rho_\chi(ST^{-1})P=P\rho_\chi(ST)$ we get finally

$$\lambda\underline{f}(\zeta)=\sum_{n=1}^{\infty}\frac{1}{(n+\zeta)^{2\beta}}P\rho_\chi(ST^n)\underline{f}\Big(\frac{1}{n+\zeta}\Big).$$

Hence any solution $\underline{\Psi}$ of the Lewis equations (2.2.1) and (2.2.2) with the asymptotics at the cut $\zeta=0$ determines an eigenfunction $\underline{f}$ of the transfer operator $P\mathscr{L}^+_{\beta,\chi}$ with eigenvalues $\lambda=\pm1$. □

3. Automorphism of the Maass forms and their period functions for $\Gamma_0(N)$

The Maass forms $u=u(z)$ of a cofinite Fuchsian group Γ and unitary character χ are real analytic functions $u:\mathbb{H}\to\mathbb{C}$ with

- $\Delta u(z)=\lambda u(z)$,
- $u(gz)=\chi(g)u(z)$ for all $g\in\Gamma$, and
- $u(g_jz)=O(y^C)$ as $y\to\infty$ for some constant $C\in\mathbb{R}$ and all cusps $z_j=g_j(i\infty)$ of Γ.

The cusp forms are those forms which decay exponentially fast at the cusps. If $u\in L_2(M_\Gamma)$ we call u a Maass wave form.

Definition 3.0.1. An element $j\in\mathrm{GL}(2,\mathbb{Z})$ defines an automorphism J of the Maass wave form u for the group Γ and character χ if Ju with $Ju(z):=u(jz)$ is a Maass form for Γ and character χ.

Obviously j defines an automorphism J if and only if j is a normalizer of the group Γ and the character χ is invariant under j, that is, $\chi(jgj^{-1})=\chi(g)$ for all $g\in\Gamma$. Thereby $jz=(az^*+b)/(cz^*+d)$ if $\det g=ad-bd=-1$. We have to show that the function $Ju(z)=u(jz)$ has at most polynomial growth

at the cusps $z_i = \tau_i(i\infty)$ of Γ, where $\tau_i \in \mathrm{SL}(2,\mathbb{Z})$. If $\det j = -1$, one has $u(j\tau_i(z)) = u(j\tau_i j_{0,-} j_{0,-}(z))$ where $j_{0,-} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Then $j\tau_i j_{0,-} \in \mathrm{SL}(2,\mathbb{Z})$ and hence $j\tau_i j_{0,-} = \gamma_i R_i$ for some $\gamma_i \in \Gamma$ and some representative R_i of the cosets $\Gamma \setminus \mathrm{SL}(2,\mathbb{Z})$. But $R_i = \eta\tau_{\sigma(i)}$ for some $\eta \in \Gamma$ and some index $\sigma(i)$. Hence $u(j\tau_i(z)) = u(\tau_{\sigma(i)}(-z^*))$ which has at most polynomial growth at the cusps. The same argument applies if $\det j = 1$, and it shows also that Ju is a Maass wave form or a cusp form if u is one.

3.1. ***The group of automorphisms of Maass forms for $\Gamma_0(N)$ and trivial character χ_0.*** We restrict ourselves now to the case $\Gamma = \Gamma_0(N)$ and assume $\chi = \chi_0$. Denote by $\mathcal{N}_N$ the normalizer group $\{\Gamma_0(N)j : j \text{ normalizer of } \Gamma_0(N) \text{ in } \mathrm{GL}(2,\mathbb{Z})\}$. Using results of [Lehner and Newman 1964; Conway and Norton 1979], we find:

Proposition 3.1.1. *For $h_N = \max\{r : r|24 \text{ and } r^2|N\}$ and $k_N := N/h_N$ the normalizer group $\mathcal{N}_N$ is given by*

$$\mathcal{N}_N = \left\{ \Gamma_0(N) j_{n,\pm},\ j_{n,\pm} = \begin{pmatrix} 1 & 0 \\ nk_N & \pm 1 \end{pmatrix},\ 0 \le n \le h_N - 1 \right\}.$$

Proof. Using the fact that the divisors k of 24 are exactly the numbers for which $a \cdot d = 1 \bmod k$ implies $a = d \bmod k$ one shows that the normalizer group of $\Gamma_0(N)$ in $\mathrm{SL}(2,\mathbb{Z})$ is $\Gamma_0(N) \setminus \Gamma_0(N/\nu)$ [Lehner and Newman 1964] with $\nu = 2^{\min\{3,[\epsilon_2/2]\}} \cdot 3^{\min\{1,[\epsilon_3/2]\}}$, $\epsilon_2 = \max\{l : 2^l|N\}$, and $\epsilon_3 = \max\{l : 3^l|N\}$. But obviously $\nu = h_N$ and $[\Gamma_0(k_N) : \Gamma_0(N)] = h_N$ and hence $\mathcal{N}_N = \Gamma_0(N) \setminus \left(\Gamma_0(k_N) \bigcup \Gamma_0(k_N) j_{0,-}\right)$. Since $j_{n,\pm} \neq j_{m,\pm} \bmod \Gamma_0(N)$ for $n \neq m$, this group has just the $2h_N$ elements $\Gamma_0(N) j_{n,\pm}$, $0 \le n \le h_N - 1$. The normalizer group $\mathcal{N}_N$ is therefore generated by the h_N generators $\{\Gamma_0(N) j_{n,-}, 0 \le n \le h_N - 1\}$. □

3.2. ***The period functions of $\Gamma_0(N)$ and character χ.*** For u a Maass form with $\Delta u = \beta(1-\beta)u$ and $\Gamma_0(N) \setminus \mathrm{SL}(2,\mathbb{Z}) = \{\Gamma_0(N)R_i, 1 \le i \le \mu_N\}$ its vector-valued period function $\underline{u}$ is defined by

$$\underline{u} = (u_i(z))_{1 \le i \le \mu_N} \text{ where } u_i(z) = u(R_i z). \tag{3.2.1}$$

Then one has, as shown for instance in [Mühlenbruch 2006]:

- $\underline{u}(gz) = \rho_\chi(g)\underline{u}(z)$ for all $g \in \mathrm{SL}(2,\mathbb{Z})$ and ρ_χ the representation of $\mathrm{SL}(2,\mathbb{Z})$ induced from the character χ on $\Gamma_0(N)$ and
- $\Delta u_i(z) = \beta(1-\beta)u_i(z)$, $1 \le i \le \mu_N$.

Given two eigenfunctions $u = u(z)$ and $v = v(z)$ of the hyperbolic Laplacian with identical eigenvalue $\lambda = \beta(1-\beta)$, one knows [Lewis and Zagier 2001] that the 1-form $\eta = \eta(u, v)$, with

$$\eta(u,v)(z) := \left[v(z)\partial_y u(z) - u(z)\partial_y v(z)\right] dx + \left[u(z)\partial_x v(z) - v(z)\partial_x u(z)\right] dy$$

is closed. If $u = u(z)$ is a Maass wave form for $\Gamma_0(N)$ with eigenvalue $\lambda = \beta(1-\beta)$ and $R_\zeta(z) = y/((\zeta - x)^2 + y^2)$ denotes the Poisson kernel, the vector-valued period function $\underline{\psi} = (\psi_j(\zeta))_{1\le j\le \mu_N}$ is defined as

$$\psi_j(\zeta) := \int_0^\infty \eta(u_j, R_\zeta^\beta)(z). \tag{3.2.2}$$

The following result has been shown for trivial character χ_0 in [Mühlenbruch 2006]. That proof can be extended, however, immediately to the case of a nontrivial character χ.

Proposition 3.2.1. *The period function $\underline{\psi} = \underline{\psi}(\zeta)$ of a Maass wave form $u = u(z)$ for $\Gamma_0(N)$ and unitary character χ is holomorphic in the cut ζ-plane $\mathbb{C} \setminus (-\infty, 0]$ and fulfills there the Lewis functional equation* (2.2.2):

$$\underline{\psi}(\zeta) - \rho_\chi(T^{-1})\underline{\psi}(\zeta+1) - (\zeta+1)^{-2\beta}\rho_\chi(T'^{-1})\underline{\psi}\Big(\frac{\zeta}{\zeta+1}\Big) = \underline{0},$$

where ρ_χ denotes the representation of $\mathrm{SL}(2,\mathbb{Z})$ *induced from the character χ of $\Gamma_0(N)$.*

On the other hand, it follows from [Deitmar and Hilgert 2007] that the solutions of the above equation holomorphic in the cut ζ-plane with certain asymptotic behavior at the cut 0 and at ∞ are in one-to-one correspondence with the Maass wave forms. That paper treats only the trivial character but it can be extended also to the case of the nontrivial character χ. Since the function $\underline{\Psi}(\zeta) = P\rho_\chi(T^{-1}S)P\,\underline{f}(\zeta-1)$ with $\underline{f}$ an eigenfunction of the operator $P\mathscr{L}^+_{\beta,\chi}$ with eigenvalues $\lambda = \pm 1$ is such a solution of (2.2.2), these eigenfunctions are in one-to-one correspondence with the Maass wave forms. As in the case of the full modular group $\mathrm{SL}(2,\mathbb{Z})$ treated in [Chang and Mayer 1998; Lewis and Zagier 2001] one can extend this result to arbitrary Maass forms, that is, also to the real analytic Eisenstein series for $\Gamma_0(N)$ and unitary character χ.

3.3. ***Automorphisms of the period functions.*** We have seen that the group of automorphisms in $\mathrm{GL}(2,\mathbb{Z})$ of the Maass forms u of $\Gamma_0(N)$ and trivial character χ_0 is generated by the matrices

$$j_{n,-} = \begin{pmatrix} 1 & 0 \\ nk_N & -1 \end{pmatrix}, \quad 0 \le n \le \mu_N - 1.$$

Denote by $J_{n,-}u$ the Maass form $J_{n,-}u(z) := u(j_{n,-}z)$ and by $J_{n,-}\underline{\psi}$ its period function. Then one shows

Theorem 3.3.1. *The period function $J_{n,-}\underline{\psi} = (J_{n,-}\psi_j(\zeta))_{1\le j\le\mu_N}$ is given by*

$$J_{n,-}\psi_j(\zeta) = \zeta^{-2\beta}\psi_{\lambda_{n_-}\circ\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big), \tag{3.3.1}$$

where the permutations λ_{n_-}, σ and δ are determined through the coset representatives R_j of $\Gamma_0(N) \setminus \mathrm{SL}(2,\mathbb{Z})$ as follows:

$$j_{n,+}R_j = \theta_j R_{\lambda_{n,-}(j)}, \quad j_{0,-}R_j j_{0,-} = \gamma_j R_{\sigma(j)}, \quad \textit{and} \quad R_j S = \eta_j R_{\delta(j)},$$

with $\theta_j, \gamma_j, \eta_j \in \Gamma_0(N)$ for $1 \le j \le \mu_N$.

Proof. For $u = u(z)$ a Maass form for $\Gamma_0(N)$ and trivial character χ_0 and $\underline{u} = \underline{u}(z)$ its vector-valued Maass form consider the Maass forms $J_{n,\pm}u(z) = u(j_{n,\pm}z)$ and $J_{n,\pm}\underline{u}(z) = (J_{n,\pm}u_j(z))_{1\le j\le \mu_N}$, with $J_{n,\pm}u_j(z) = u(j_{n,\pm}R_j z)$. Since $j_{n,+}R_j = \theta_j R_{\lambda_{n,-}(j)}$ for some unique $\theta_j \in \Gamma_0(N)$ and permutation $\lambda_{n,-}$ of $\{1, 2, \ldots, \mu_N\}$ one gets for $J_{n,+}u_j$

$$J_{n,+}u_j(z) = u(R_{\lambda_{n,-}(j)}z) = u_{\lambda_{n,-}(j)}(z). \tag{3.3.2}$$

For $J_{n,+}u_j(-z^*) = u(j_{n,+}R_j(-z^*)) = u(j_{n,+}R_j j_{0,-}z)$, on the other hand, one finds

$$J_{n,+}u_j(-z^*) = u(j_{n,-}j_{0,-}R_j j_{0,-}z) = u(j_{n,-}R_{\sigma(j)}z),$$

since $j_{0,-}R_j j_{0,-} = \gamma_j R_{\sigma(j)}$ for some unique $\gamma_j \in \Gamma_0(N)$ and permutation σ of $\{1, 2, \ldots, \mu_N\}$. Hence

$$J_{n,+}u_j(-z^*) = J_{n,-}u_{\sigma(j)}(z). \tag{3.3.3}$$

Consider next $J_{n,+}u_j(Sz) = J_{n,+}u(R_j Sz)$. Since $R_j S = \eta_j R_{\delta(j)}$ for unique $\eta_j \in \Gamma_0(N)$ and permutation δ of $\{1, 2, \ldots, \mu_N\}$, one has

$$J_{n,+}u_j(Sz) = J_{n,+}u(R_{\delta(j)}z) = J_{n,+}u_{\delta(j)}(z).$$

Hence by (3.3.2)

$$J_{n,+}u_j(Sz) = u_{\lambda_{n,-}\circ\delta(j)}(z). \tag{3.3.4}$$

On the other hand, one gets for $J_{n,+}u_j(S(-z^*)) = J_{n,+}u_j(-Sz^*)$ by using (3.3.3):

$$J_{n,+}u_j(S(-z^*)) = J_{n,-}u_{\sigma(j)}(Sz) = u(j_{n,-}R_{\sigma(j)}Sz),$$

and therefore

$$J_{n,+}u_j(-Sz^*) = u(j_{n,-}\eta_{\sigma(j)}R_{\delta\circ\sigma(j)}(z)) = J_{n,-}u_{\delta\circ\sigma(j)}(z).$$

But $\sigma \circ \delta = \delta \circ \sigma$ and therefore

$$J_{n,+}u_j(S(-z^*)) = J_{n,-}u_{\sigma\circ\delta(j)}(z). \tag{3.3.5}$$

Define next

$$v_{\pm,j}(z) := J_{n,+}u_j(z) \pm J_{n,+}u_j(-z^*).$$

Then by (3.3.2) and (3.3.3) one has

$$v_{\pm,j}(z) = u_{\lambda(j)}(z) \pm J_{n,-}u_{\sigma(j)}(z)$$

and hence, if $\Delta u(z) = \beta(1-\beta)u(z)$,

$$\Delta v_{\pm,j}(z) = \beta(1-\beta)v_{\pm,j}(z) \tag{3.3.6}$$

and

$$v_{\pm,j}(-z^*) = \pm v_{\pm,j}(z). \tag{3.3.7}$$

Equations (3.3.4) and (3.3.5), on the other hand, show

$$v_{\pm,j}(Sz) = v_{\pm,\delta(j)}(z). \tag{3.3.8}$$

Set $\underline{\psi}'_{\pm}(\zeta) := \int_0^{i\infty} \eta(\underline{v}_{\pm}, R_\zeta^\beta)(z)$. Then, since $v_{\pm,j}(-z^*) = \pm v_{\pm,j}(z)$, one finds (see [Lewis and Zagier 2001])

$$\psi'_{+,j}(\zeta) = 2\beta \int_0^\infty \frac{t^\beta v_{+,j}(it)}{(\zeta^2+t^2)^{\beta+1}}\,dt, \tag{3.3.9}$$

$$\psi'_{-,j}(\zeta) = -\int_0^\infty \frac{t^\beta \partial_x v_{-,j}(it)}{(\zeta^2+t^2)^{\beta}}\,dt. \tag{3.3.10}$$

Using next the identity (3.3.8) one easily shows

$$\psi'_{\pm,j}(\zeta) = \pm\zeta^{-2\beta}\psi'_{\pm,\delta(j)}\Big(\frac{1}{\zeta}\Big). \tag{3.3.11}$$

But $v_{\pm,j}(z) = u_{\lambda(j)}(z) \pm J_{n,-}u_{\sigma(j)}(z)$ and hence

$$\psi'_{\pm,j}(\zeta) = \psi_{\lambda_{n,-}(j)}(\zeta) \pm J_{n,-}\psi_{\sigma(j)}(\zeta).$$

Therefore

$$\psi_{\lambda_{n,-}(j)}(\zeta) \pm J_{n,-}\psi_{\sigma(j)}(\zeta) = \pm\zeta^{-2\beta}\Big(\psi_{\lambda_{n,-}\circ\delta(j)}\Big(\frac{1}{\zeta}\Big) \pm J_{n,-}\psi_{\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big)\Big). \tag{3.3.12}$$

Adding these two equations leads finally to

$$\psi_{\lambda_{n,-}(j)}(\zeta) = \zeta^{-2\beta} J_{n,-}\psi_{\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big), \tag{3.3.13}$$

and therefore to the equation

$$J_{n,-}\psi_j(\zeta) = \zeta^{-2\beta}\psi_{\lambda_{n,-}\circ\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big), \tag{3.3.14}$$

which was to be proven. □

Remark 3.3.2. As can be seen from their action on the coset representatives R_j the permutation δ commutes with the permutations $\lambda_{n,-}$ and σ. Furthermore one has $\sigma^2 = \delta^2 = (\lambda_{n,-} \circ \sigma)^2 = \mathrm{id}$, where id denotes the identity permutation. This shows also that the automorphisms $J_{n,-}$ are involutions both of the Maass forms and the period functions, a special case of these involutions for all groups $\Gamma_0(N)$ being $J_{0,-}u(z) = u(-z^*)$.

Denote by $Q_{n,-}$, $0 \le n \le h_N - 1$, the $\mu_N \times \mu_N$ permutation matrix corresponding to the permutation $\lambda_{n,-} \circ \sigma \circ \delta$.

Theorem 3.3.3. *The permutation matrices* $P_{n,-} := \rho_{\chi_0}(S) Q_{n,-}$, $0 \le n \le h_N - 1$, *define symmetries*

$$\tilde{P}_{n,-} = \begin{pmatrix} 0 & P_{n,-} \\ P_{n,-} & 0 \end{pmatrix}$$

for the transfer operator

$$\mathbf{L}_{\beta,\chi_0} = \begin{pmatrix} 0 & \mathscr{L}^+_{\beta,\chi_0} \\ \mathscr{L}^+_{\beta,\chi_0} & 0 \end{pmatrix}$$

for $\Gamma_0(N)$ *and trivial character* $\chi_0 \equiv 1$, *with* $P^2_{n,-} = \mathrm{id}_{\mu_N}$, $P_{n,-}\rho_{\chi_0}(S) = \rho_{\chi_0}(S)P_{n,-}$, *and* $P_{n,-}\rho_{\chi_0}(T) = \rho_{\chi_0}(T^{-1})P_{n,-}$, *and therefore* $P_{n,-}\mathscr{L}^+_{\beta,\chi_0} = \mathscr{L}^+_{\beta,\chi_0}P_{n,-}$. *The permutation matrix* $P_{n,-}$ *is determined by the permutation* $\lambda_{n,-} \circ \sigma$ *and hence by the coset representatives* $j_{n,-}R_j j_{0,-}$.

Proof. The matrix $P_{n,-}\rho_{\chi_0}(S)$ is determined by the coset representatives $j_{n,-}R_j S j_{0,-}$ whereas $\rho_{\chi_0}(S)P_{n,-}$ is determined by the coset representatives $j_{n,-}R_j j_{0,-}S$ and $j_{0,-}S = Sj_{0,-}$. Hence $P_{n,-}\rho_{\chi_0}(S) = \rho_{\chi_0}(S)P_{n,-}$. On the other hand $Tj_{0,-} = j_{0,-}T^{-1}$ and therefore $P_{n,-}\rho_{\chi_0}(T) = \rho_{\chi_0}(T^{-1})P_{n,-}$. □

Obviously Theorem 2.1.1 follows from Theorem 3.3.3. For the automorphisms $j_{n,+} = j_{n,-}j_{0,-}$ one gets the symmetry

$$\tilde{P}_{n,+} = \begin{pmatrix} P_{n,+} & 0 \\ 0 & P_{n,+} \end{pmatrix},$$

with $P_{n,+}$ the permutation matrix corresponding to the permutation $\lambda_{n,-} \circ \sigma \circ \lambda_{0,-} \circ \sigma$ determined by the coset representatives $j_{n,+}R_j$.

Remark 3.3.4. The symmetry $P_{0,-}$ is given by $\rho_{\chi_0}(SM)$ where $M = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and ρ_{χ_0} denotes the representation of $\mathrm{GL}(2,\mathbb{Z})$ induced from the trivial character χ_0 of $\Gamma_0(N)$. The transfer operator $\mathscr{L}^{MM}_\beta$ of Manin and Marcolli [2002] for $\Gamma_0(N)$ turns out to coincide with the operator $\rho_{\chi_0}(S)P_{0,-}\mathscr{L}^+_{\beta,\chi_0}\rho_{\chi_0}(S)$ and appears as a special case of our operators $P_{n,-}\mathscr{L}^+_{\beta,\chi_0}$.

Corollary 3.3.5. *The permutation matrices $P_{n,-}$, $0 \le n \le h_N - 1$, generate a finite group consisting of the permutation matrices $\{P_{n,\pm}, 0 \le n \le h_N - 1\}$ and isomorphic to the normalizer group $\mathcal{N}_N$ of $\Gamma_0(N)$ in $\mathrm{GL}(2,\mathbb{Z})$. The symmetries $\{\tilde{P}_{n,\pm}, 0 \le n \le h_N - 1\}$ of the transfer operator $\mathbf{L}_{\beta,\chi_0}$ for $\Gamma_0(N)$ and trivial character χ_0 define a finite group isomorphic to the group $\mathcal{N}_N$.*

4. Selberg's character χ_α for $\Gamma_0(4)$

The group $\Gamma_0(4)$ is freely generated by the two elements $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 1 & 0 \\ -4 & 1 \end{smallmatrix}\right)$. Hence any $g \in \Gamma_0(4)$ can be written as $g = \prod_{i=1}^{N_g} T^{m_i} B^{n_i}$. If $\Omega(g) = \sum_{i=1}^{N_g} m_i$ then Selberg's character χ_α [Selberg 1990] is defined as

$$\chi_\alpha(g) = \exp(2\pi i \alpha \Omega(g)), \quad 0 \le \alpha \le 1. \tag{4.1}$$

Denote by z_i, $1 \le i \le 3$, the inequivalent cusps of $\Gamma_0(4)$ and by T_i the generators of their stabilizer groups Γ_{z_i} with $T_i z_i = z_i$. They can be taken as $z_1 = i\infty$, $z_2 = 0$, $z_3 = -\frac{1}{2}$ and $T_1 = T$, $T_2 = B$, $T_3 = T^{-1}B^{-1}$. The character χ_α is singular in the cusp z_i if and only if $\chi_\alpha(T_i) = 1$. Otherwise the character is nonsingular in z_i. It is well known that the multiplicity $\kappa(\chi_\alpha)$ of the continuous spectrum of the automorphic Laplacian Δ with character χ_α is given by $\kappa(\chi_\alpha) = \#\{i : \chi_\alpha(T_i) = 1\}$. Therefore $\kappa(\chi_\alpha) = 3$ for $\alpha = 0$ whereas $\kappa(\chi_\alpha) = 1$ for $\alpha \ne 0$ and hence the multiplicity of the continuous spectrum of the Laplacian changes from 3 to 1 when the trivial character is deformed to χ_α with $\alpha \ne 0$. It is known [Phillips and Sarnak 1994] that the character χ_α is congruent (or arithmetic) if and only if $\alpha \in \{k\frac{1}{8}, 0 \le k \le 4\}$. Since the Selberg zeta function given in (2.0.1) has the property $Z_{\Gamma_0(4),\chi_\alpha} = Z_{\Gamma_0(4),\chi_{-\alpha}}$ and obviously $\chi_\alpha = \chi_{\alpha+1}$ we can restrict the deformation parameter α to the range $0 \le \alpha \le \frac{1}{2}$.

Lemma 4.1. *The Selberg character χ_α is invariant under the map defined by $j_{2,-}z = z^*/(2z^* - 1)$, and $J_{2,-}u(z) := u(j_{2,-}z)$ is a Maass form for $\Gamma_0(4)$ and character χ_α if $u = u(z)$ is such a Maass form.*

Proof. We only have to show that χ_α is invariant under the map $j_{2,-}z = z^*/(2z^* - 1)$. For $g = T$ we find $j_{2,-}Tj_{2,-} = TB$ and hence

$$\chi_\alpha(j_{2,-}Tj_{2,-}) = \chi_\alpha(TB) = \chi_\alpha(T),$$

whereas for $g = B$ one finds $j_{2,-}Bj_{2,-} = B^{-1}$ and hence

$$\chi_\alpha(j_{2,-}Bj_{2,-}) = \chi_\alpha(B^{-1}) = \chi_\alpha(B).$$

Therefore $\chi_\alpha(j_{2,-}gj_{2,-}) = \chi_\alpha(g)$ for all $g \in \Gamma_0(4)$. □

If $u = u(z)$ is a Maass form for $\Gamma_0(4)$ with character χ_α and $\underline{\psi} = (\psi_j(\zeta))_{1 \le j \le 6}$ is its period function, denote by J_-u the Maass form given by $J_-u(z) := u(j_{2,-}z)$, and by $J_-\underline{\psi} = (J_-\psi_j(\zeta))_{1 \le j \le 6}$ its period function. Then one shows:

Theorem 4.2. *The period function $J_-\underline{\psi}$ of the Maass form J_-u is given by*

$$J_-\psi_j(\zeta) = \zeta^{-2\beta}\chi_\alpha(\eta_{\sigma\circ\delta(j)})\psi_{\lambda_{2,-}\circ\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big), \tag{4.2}$$

where the permutations $\lambda_{2,-}$, σ, and δ, as well as the $\eta_j \in \Gamma_0(4)$, are determined through the coset representatives R_j by

$$j_{2,+}R_j = \theta_j R_{\lambda_{2,-}(j)}, \quad j_{0,-}R_j j_{0,-} = \gamma_j R_{\sigma(j)}, \quad R_j S = \eta_j R_{\delta(j)},$$

with θ_j, γ_j, and $\eta_j \in \Gamma_0(4)$ for $1 \le j \le 6$.

Proof. Set $j_\pm := j_{2,\pm}$ and $J_\pm u(z) := u(j_\pm z)$. Then J_-u is a Maass form for $\Gamma_0(4)$ and character χ_α whereas J_+u is a Maass form for $\Gamma_0(4)$ and character $\chi_{-\alpha}$. The vector-valued Maass form $J_+\underline{u} = (J_+u_j)_{1\le j\le 6}$ is given by $J_+u_j(z) = u(j_+R_jz)$. We have chosen the representatives R_j of the cosets in $\mathrm{SL}(2,\mathbb{Z}) = \bigcup_{1\le j\le 6}\Gamma_0(4)R_j$ as follows:

$$R_1 = \mathrm{id}_2, \qquad R_j = ST^{j-2}, \quad 2 \le j \le 5, \qquad R_6 = ST^2S.$$

But $j_+R_j = \theta_j R_{\lambda_{2,-}(j)}$ for some $\theta_j \in \Gamma_0(4)$ and some permutation $\lambda_{2,-}$ of the set $\{1, 2, \ldots, 6\}$ and hence $J_+u_j(z) = \chi_\alpha(\theta_j)u(R_{\lambda_{2,-}(j)}z)$. It turns out that $\theta_j = B^{-1}$ for $1 \le j \le 3$ and $\theta_j = \mathrm{id}_2$ for $4 \le j \le 6$. Hence $\chi_\alpha(\theta_j) = 1$ and

$$J_+u_j(z) = u_{\lambda_{2,-}(j)}(z), \quad 1 \le j \le 6, \tag{4.3}$$

with $\lambda_{2,-}$ the permutation

$$\lambda_{2,-} = \frac{1\ 2\ 3\ 4\ 5\ 6}{6\ 4\ 5\ 2\ 3\ 1}. \tag{4.4}$$

Consider next $J_+u_j(-z^*) = J_+u_j(j_{0,-}z)$. Then

$$J_+u_j(j_{0,-}z) = u(j_+R_jj_{0,-}z) = u(j_+j_{0,-}j_{0,-}R_jj_{0,-}z).$$

If $j_{0,-}R_jj_{0,-} = \gamma_jR_{\sigma(j)}$ then $J_+u_j(j_{0,-}z) = u(j_-\gamma_jj_-j_-R_{\sigma(j)}z)$. But it turns out that $j_-\gamma_jj_- = \mathrm{id}_2$ for $j = 1, 2, 6$ and $j_-\gamma_jj_- = B$ for $j = 3, 4, 5$, hence $\chi_\alpha(j_-\gamma_jj_-) = 1$ and therefore

$$J_+u_j(-z^*) = J_+u_j(j_{0,-}z) = J_-u_{\sigma(j)}(z). \tag{4.5}$$

Since, furthermore, $J_+u_j(Sz) = u(j_+R_jSz) = u(j_+\eta_jR_{\delta(j)}z)$, one finds

$$J_+u_j(Sz) = \chi_{-\alpha}(\eta_j)u_{\lambda_{2,-}\circ\delta(j)}(z), \tag{4.6}$$

where δ is the permutation

$$\delta = \frac{1\ 2\ 3\ 4\ 5\ 6}{2\ 1\ 5\ 6\ 3\ 4}, \tag{4.7}$$

$\eta_j = \mathrm{id}_2$ for $j = 1, 2, 4, 6$, and $\eta_3 = \eta_5^{-1} = T^{-1}B^{-1}$. For $J_+u_j(-Sz^*)$ one gets with (4.5) $J_+u_j(-Sz^*) = J_-u_{\sigma(j)}(Sz) = u(j_{2,-}R_{\sigma(j)}Sz)$ and

$$J_+u_j(-Sz^*) = u(j_{2,-}\eta_{\sigma(j)}R_{\delta\circ\sigma(j)}z) = \chi_\alpha(\eta_{\sigma(j)})J_-u_{\delta\circ\sigma(j)}(z).$$

Using the explicit form of the η_j one shows $\chi_\alpha(\eta_{\sigma(j)}) = \chi_{-\alpha}(\eta_j)$ and therefore

$$J_+u_j(-Sz^*) = \chi_{-\alpha}(\eta_j)J_-u_{\delta\circ\sigma(j)}(z). \tag{4.8}$$

Define next $v_{\pm,j} = v_{\pm,j}(z)$ as

$$v_{\pm,j}(z) := J_+u_j(z) \pm J_+u_j(-z^*). \tag{4.9}$$

Then $v_{\pm,j}(-z^*) = \pm v_{\pm,j}(z)$, and by (4.6) and (4.8) we have

$$v_{\pm,j}(Sz) = \chi_{-\alpha}(\eta_j)v_{\pm,\delta(j)}(z). \tag{4.10}$$

If therefore $\psi'_{\pm,j}(\zeta) := \int_0^{i\infty} \eta(v_{\pm,j}, R_\zeta^\beta)(z)$ one gets from relation (4.10)

$$\psi'_{\pm,j}(\zeta) = \pm\zeta^{-2\beta}\chi_{-\alpha}(\eta_j)\psi'_{\pm,\delta(j)}\Big(\frac{1}{\zeta}\Big) \tag{4.11}$$

and using the identity (4.9)

$$\begin{aligned}\psi_{\lambda_{2,-}(j)}(\zeta) &\pm J_-\psi_{\sigma(j)}(\zeta)\\ &= \pm\zeta^{-2\beta}\chi_{-\alpha}(\eta_j)\Big(\psi_{\lambda_{2,-}\circ\delta(j)}\Big(\frac{1}{\zeta}\Big) \pm J_-\psi_{\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big)\Big).\end{aligned} \tag{4.12}$$

Adding these two equations leads finally to

$$J_-\psi_j(\zeta) = \zeta^{-2\beta}\chi_\alpha(\eta_{\sigma\circ\delta(j)})\psi_{\lambda_{2,-}\circ\sigma\circ\delta(j)}\Big(\frac{1}{\zeta}\Big), \qquad \square$$

Inserting the explicit form of the permutations

$$\sigma\circ\delta = \begin{matrix}1&2&3&4&5&6\\2&1&3&6&5&4\end{matrix} \tag{4.13}$$

and

$$\lambda_{2,-}\circ\sigma\circ\delta = \begin{matrix}1&2&3&4&5&6\\4&6&5&1&3&2\end{matrix} \tag{4.14}$$

and the character values

$$\chi_\alpha(\eta_1) = \chi_\alpha(\eta_2) = \chi_\alpha(\eta_4) = \chi_\alpha(\eta_6) = 1 \quad \text{and} \quad \chi_\alpha(\eta_3) = \chi_\alpha(\eta_5)^{-1} = e^{-2\pi i\alpha},$$

one finds

$$\begin{aligned} J_-\psi_1(\zeta) &= \zeta^{-2\beta}\psi_4\Big(\frac{1}{\zeta}\Big), & J_-\psi_2(\zeta) &= \zeta^{-2\beta}\psi_6\Big(\frac{1}{\zeta}\Big),\\ J_-\psi_3(\zeta) &= \zeta^{-2\beta}e^{-2\pi i\alpha}\psi_4\Big(\frac{1}{\zeta}\Big), & J_-\psi_4(\zeta) &= \zeta^{-2\beta}\psi_1\Big(\frac{1}{\zeta}\Big),\\ J_-\psi_5(\zeta) &= \zeta^{-2\beta}e^{2\pi i\alpha}\psi_3\Big(\frac{1}{\zeta}\Big), & J_-\psi_6(\zeta) &= \zeta^{-2\beta}\psi_2\Big(\frac{1}{\zeta}\Big). \end{aligned} \tag{4.15}$$

Define the matrix $Q_{2,-}$ through the equation $J_{2,-}\underline{\psi}(\zeta) = \zeta^{-2\beta}Q_{2,-}\underline{\psi}(1/\zeta)$.

Proposition 4.3. *The permutation matrix $P_{2,-} := \rho_{\chi_\alpha}(S)Q_{2,-}$ defines a symmetry*

$$\tilde{P}_{2,-} = \begin{pmatrix} 0 & P_{2,-} \\ P_{2,-} & 0 \end{pmatrix}$$

of the transfer operator

$$\mathbf{L}_{\beta,\chi_\alpha} = \begin{pmatrix} 0 & \mathscr{L}^+_{\beta,\chi_\alpha} \\ \mathscr{L}^+_{\beta,\chi_\alpha} & 0 \end{pmatrix}$$

for $\Gamma_0(4)$ and character χ_α with $P_{2,-}^2 = \mathrm{id}_6$,

$$P_{2,-}\rho_{\chi_\alpha}(S) = \rho_{\chi_\alpha}(S)P_{2,-} \quad \textit{and} \quad P_{2,-}\rho_{\chi_\alpha}(T) = \rho_{\chi_\alpha}(T^{-1})P_{2,-},$$

and therefore $P_{2,-}\mathscr{L}^+_{\beta,\chi_\alpha} = \mathscr{L}^-_{\beta,\chi_\alpha}P_{2,-}$. The permutation matrix $P_{2,-}$ corresponds to the permutation $\lambda_{2,-}\circ\sigma$ and hence is determined by the coset representatives $J_{2,-}R_j\,j_{0,-}$.

Proof. For our choice of coset representatives R_j as given in (2.1.4) one finds for $\rho_{\chi_\alpha}(S)$

$$\rho_{\chi_\alpha}(S) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{-2\pi i\alpha} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & e^{2\pi i\alpha} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \tag{4.16}$$

and hence the matrix $Q_{2,-}$ is given by

$$Q_{2,-} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & e^{-2\pi i\alpha} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{2\pi i\alpha} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.17}$$

For $\rho_{\chi_\alpha}(T)$ one finds

$$\rho_{\chi_\alpha}(T) = \begin{pmatrix} e^{2\pi i\alpha} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-2\pi i\alpha} \end{pmatrix}. \tag{4.18}$$

A simple calculation then confirms that $P_{2,-}\rho_{\chi_\alpha}(S) = \rho_{\chi_\alpha}(S)P_{2,-}$ and $P_{2,-}\rho_{\chi_\alpha}(T) = \rho_{\chi_\alpha}(T^{-1})P_{2,-}$, with

$$P_{2,-} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{4.19}$$

and hence defines a symmetry of the transfer operator $\mathbf{L}_{\beta,\chi_\alpha}$. The matrix $P_{2,-}$ coincides with the permutation matrix P_2 corresponding to the permutation σ_2 in (2.1.6). □

Remark 4.4. For the trivial character χ_0, the map $j_{0,-}z = -z^*$ defines an automorphism of the Maass forms for the group $\Gamma_0(4)$. Indeed, this is an automorphism for all Hecke congruence subgroups $\Gamma_0(N)$. In this case the permutation $\lambda_{0,-}$ is the trivial permutation and the matrix $Q_{0,-}$ is determined by the permutation $\sigma \circ \delta$. For $\Gamma_0(4)$ this is given by (4.13). Using (4.16) with $\alpha = 0$ one obtains for $P_{0,-} = \rho_{\chi_0}(S)Q_{0,-}$ the permutation σ_1 as given in (2.1.5). The symmetry $\tilde{P}_1$ for $\Gamma_0(4)$ and trivial character χ_0 hence corresponds to the automorphism $z \to -z^*$ of the Maass forms for this group.

We have seen that for every eigenfunction $\underline{f} = \underline{f}(\zeta)$ of the operator $P_2\mathscr{L}^+_{\beta,\chi_\alpha}$ with eigenvalues $\lambda = \pm 1$ the function $\underline{\Psi} = \underline{\Psi}(\zeta) = P_2\rho_{\chi_\alpha}(T^{-1}S)P_2\underline{f}(\zeta - 1)$ fulfills the functional equation

$$\underline{\Psi}(\zeta) = \lambda\zeta^{-2\beta}\rho_{\chi_\alpha}(S)P_2\underline{\Psi}\Big(\frac{1}{\zeta}\Big) = \lambda J_-\underline{\Psi}(\zeta) \tag{4.20}$$

and hence is an eigenfunction of the involution J_- corresponding to the automorphism $j_- = j_{2,-}$ of the Maass forms for $\Gamma_0(4)$ and character χ_α. Hence this shows:

Proposition 4.5. *The eigenfunctions $\underline{f} = \underline{f}(\zeta)$ of the operator $P_2\mathscr{L}^+_{\beta,\chi_\alpha}$ with eigenvalues $\lambda = \pm 1$ correspond to Maass forms which are even or odd, respectively, under the involution $J_- = J_{2,-}$.*

For a conjugate character $\hat{\chi}_\alpha$, Phillips and Sarnak [1994] have shown that the Maass cusp forms that are odd under the corresponding conjugate involution $\hat{J}$ are still cusp forms under the deformation of this character. Hence:

Corollary 4.6. *The zeros of the Selberg zeta function for the group* $\Gamma_0(4)$ *and character* χ_α *corresponding to eigenfunctions of the operator* $P_2\mathcal{L}^+_{\beta,\chi_\alpha}$ *with eigenvalue* $\lambda = -1$ *which for* $\alpha = 0$ *are on the critical line* $\operatorname{Re}\beta = \frac{1}{2}$ *stay, for all* α*, on this line.*

Remark 4.7. The operator $P_2\mathcal{L}^+_{\beta,\chi_\alpha}$ can be used to calculate numerically the Selberg zeta function for small values of $\operatorname{Im}\beta$ and arbitrary $0 \le \alpha \le \frac{1}{2}$. These numerical calculations confirm the above corollary and let us expect that all the zeros of the Selberg function corresponding to the eigenvalue $\lambda = 1$ of the operator $P_2\mathcal{L}^+_{\beta,\chi_\alpha}$ for $\alpha = 0$ leave the critical line when α becomes positive. A detailed discussion of the numerical treatment of the behavior of the zeros of the Selberg function under character deformation will appear elsewhere [Bruggeman et al. 2012].

References

[Bruggeman et al. 2012] R. W. Bruggeman, M. Fraczek, and D. Mayer, "Perturbation of zeros of the Selberg zeta function for $\Gamma_0(4)$", preprint, 2012. arXiv 1201.2324v1

[Chang and Mayer 1998] C.-H. Chang and D. Mayer, "The period function of the nonholomorphic Eisenstein series for PSL(2, **Z**)", *Math. Phys. Electron. J.* **4** (1998), no. 6. MR 99k:11070 Zbl 0927.11027

[Chang and Mayer 2000] C.-H. Chang and D. Mayer, "Thermodynamic formalism and Selberg's zeta function for modular groups", *Regul. Chaotic Dyn.* **5**:3 (2000), 281–312. MR 2001k:37045 Zbl 0979.37014

[Chang and Mayer 2001a] C.-H. Chang and D. H. Mayer, "Eigenfunctions of the transfer operators and the period functions for modular groups", pp. 1–40 in *Dynamical, spectral, and arithmetic zeta functions* (San Antonio, TX, 1999), edited by M. L. Lapidus and M. van Frankenhuysen, Contemp. Math. **290**, Amer. Math. Soc., Providence, RI, 2001. MR 2003h:11056 Zbl 1037.11032

[Chang and Mayer 2001b] C.-H. Chang and D. H. Mayer, "An extension of the thermodynamic formalism approach to Selberg's zeta function for general modular groups", pp. 523–562 in *Ergodic theory, analysis, and efficient simulation of dynamical systems*, edited by B. Fiedler, Springer, Berlin, 2001. MR 2002k:37046 Zbl 1211.37030

[Conway and Norton 1979] J. H. Conway and S. P. Norton, "Monstrous moonshine", *Bull. London Math. Soc.* **11**:3 (1979), 308–339. MR 81j:20028 Zbl 0424.20010

[Deitmar and Hilgert 2007] A. Deitmar and J. Hilgert, "A Lewis correspondence for submodular groups", *Forum Math.* **19**:6 (2007), 1075–1099. MR 2008j:11058 Zbl 1211.11063

[Efrat 1993] I. Efrat, "Dynamics of the continued fraction map and the spectral theory of SL(2, $\mathbb{Z}$)", *Invent. Math.* **114**:1 (1993), 207–218. MR 94h:11052 Zbl 0811.11037

[Fraczek 2010] M. Fraczek, *Character deformation of the Selberg zeta function for congruence subgroups via the transfer operator*, thesis, Clausthal Institute of Technology, 2010.

[Hejhal 1983] D. A. Hejhal, *The Selberg trace formula for* PSL(2, **R**), vol. 2, Lecture Notes in Mathematics **1001**, Springer, Berlin, 1983. MR 86e:11040 Zbl 0543.10020

[Lehner and Newman 1964] J. Lehner and M. Newman, "Weierstrass points of $\Gamma_0(n)$", *Ann. of Math. (2)* **79** (1964), 360–368. MR 28 #5045 Zbl 0124.29203

[Lewis and Zagier 2001] J. Lewis and D. Zagier, "Period functions for Maass wave forms, I", *Ann. of Math. (2)* **153**:1 (2001), 191–258. MR 2003d:11068 Zbl 1061.11021

[Manin and Marcolli 2002] Y. I. Manin and M. Marcolli, "Continued fractions, modular symbols, and noncommutative geometry", *Selecta Math. (N.S.)* **8**:3 (2002), 475–521. MR 2004a:11039 Zbl 1116.11033

[Mayer and Mühlenbruch 2010] D. Mayer and T. Mühlenbruch, "Nearest λ_q-multiple fractions", pp. 147–184 in *Spectrum and dynamics*, edited by D. Jakobson et al., CRM Proc. Lecture Notes **52**, Amer. Math. Soc., Providence, RI, 2010. MR 2011m:11148 Zbl 05907544

[Mayer and Strömberg 2008] D. Mayer and F. Strömberg, "Symbolic dynamics for the geodesic flow on Hecke surfaces", *J. Mod. Dyn.* **2**:4 (2008), 581–627. MR 2009j:37049 Zbl 1161.37031

[Mayer et al. 2012] D. Mayer, T. Mühlenbruch, and F. Strömberg, "The transfer operator for the Hecke triangle groups", *Discrete Contin. Dyn. Syst. A* **32**:7 (2012), 2453–2484. arXiv 0912.2236

[Mühlenbruch 2006] T. Mühlenbruch, "Hecke operators on period functions for $\Gamma_0(n)$", *J. Number Theory* **118**:2 (2006), 208–235. MR 2007i:11062 Zbl 1122.11022

[Phillips and Sarnak 1994] R. Phillips and P. Sarnak, "Cusp forms for character varieties", *Geom. Funct. Anal.* **4**:1 (1994), 93–118. MR 94k:11061 Zbl 0804.11038

[Selberg 1990] A. Selberg, "Remarks on the distribution of poles of Eisenstein series", pp. 251–278 in *Festschrift in honor of I. I. Piatetski-Shapiro on the occasion of his sixtieth birthday, Part 2* (Ramat Aviv, 1989), edited by S. Gelbart et al., Israel Math. Conf. Proc. **3**, Weizmann, Jerusalem, 1990. MR 93c:11035 Zbl 0712.11034

[Strömberg 2008] F. Strömberg, "Computation of Selberg's zeta functions on Hecke triangle groups", preprint, 2008. arXiv 0804.4837

[Venkov 1990] A. Venkov, "The theory of the Selberg zeta-function", pp. 47–54 in *Spectral theory of automorphic functions and its applications*, Math. Appl. (Soviet Ser.) **51**, Kluwer, Dordrecht, 1990.

[Zagier 2002] D. Zagier, "New points of view on the Selberg zeta function", in *Proceedings of Japanese–German Seminar: explicit structures of modular forms and zeta functions* (Hakuba, 2001), edited by T. Ibukiyama, Ryushi-Do, Nagano, 2002.

Communicated by Peter Sarnak
Received 2011-01-25 Accepted 2011-06-30

markus.fraczek@tu-clausthal.de *Institut für Theoretische Physik, Technische Universität Clausthal, Leibnizstraße 10, D-38678 Clausthal-Zellerfeld, Germany*

dieter.mayer@tu-clausthal.de *Institut für Theoretische Physik, Lower Saxony Professorship, Technische Universität Clausthal, Leibnizstraße 10, D-38678 Clausthal-Zellerfeld, Germany*

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. 3 2012

1937-0652(2012)6:3;1-C